AGARD-CP-595

AGARD-CP-595

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

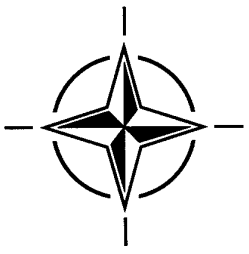
AGARD CONFERENCE PROCEEDINGS 595

Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar

(les Systèmes multi-senseurs et le fusionnement des données pour les télécommunications, la télédétection et les radars)

Papers presented at the Sensor and Propagation Panel Symposium, held in Lisbon, Portugal, 29 September - 2 October 1997.

Approved for public release;
Distribution Unlimited



NORTH ATLANTIC TREATY ORGANIZATION

Published April 1998

Distribution and Availability on Back Cover

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

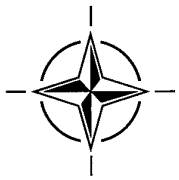
AGARD CONFERENCE PROCEEDINGS 595

Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar

(les Systèmes multi-senseurs et le fusionnement des données pour les télécommunications, la télédétection et les radars)

Papers presented at the Sensor and Propagation Panel Symposium, held in Lisbon, Portugal, 29 September - 2 October 1997.

19980611 117



North Atlantic Treaty Organization
Organisation du Traité de l'Atlantique Nord

DTIC QUALITY INSPECTED 6

The Mission of AGARD*

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

* AGARD merged with the Defence Research Group of NATO (DRG) on 1 January 1998 to form the Research and Technology Organization (RTO) of NATO. However, both AGARD and DRG will continue to issue publications under their own names in respect of work performed in 1997.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.



Printed on recycled paper

Published April 1998

Copyright © AGARD 1998
All Rights Reserved

ISBN 92-836-0051-7



Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Les systèmes multi-senseurs et le fusionnement des données pour les télécommunications, la télédétection et les radars

(AGARD CP-595)

Synthèse

Le fusionnement des données est une science qui a intrigué l'homme depuis ses origines. Au cours des siècles, différentes théories et techniques ont été développées, mais la cadence de ce développement n'a jamais atteint celle de la dernière décennie, que l'on ne peut que qualifier d'explosive. Ce symposium SPP, tenu à Lisbonne, au Portugal, a permis de présenter des méthodes et des techniques novatrices qui témoignent de la large mise en application de la science du fusionnement des données dans bon nombre de systèmes militaires.

Les communications présentées lors du symposium ont couvert les quatre grands domaines techniques suivants:

— les multisenseurs et leurs applications

— les méthodes de fusionnement des données

— les systèmes de détection et les techniques de fusionnement

— la validation opérationnelle et expérimentale et les technologies associées

Le communications présentées ont témoigné des progrès spectaculaires réalisés dans la compréhension des sciences, techniques et technologies qui sous tendent le fusionnement des données. L'importance des bases de données et des banques de données a été soulignée par plusieurs conférenciers. Un certain nombre de paramètres stratégiques, déterminants pour le fusionnement des données, ont été considérés comme étant d'une importance particulière, à savoir:

— l'opportunité du choix de senseurs, leur nombre et l'obligation d'indépendance liée aux calculs qu'ils effectuent;

— la façon de constituer la base de données et

— l'importance de bien comprendre les limitations des observations effectuées par les différentes catégories de senseurs

Suite aux discussions qui ont eu lieu, il est apparu très clairement que ces questions ne sont pas encore résolues et qu'il reste beaucoup à faire avant de pouvoir définir les domaines d'application. A titre d'exemple, ni les théories classiques comme la limite Cramer-Rao pour le traitement des signaux, ni la limite de Shannon dans la théorie de l'information ne paraissent adéquates pour la définition a priori de l'optimisation d'une base de données.

Une fois que la base de données est constituée, le fusionnement de données peut commencer. Le symposium a conclu qu'en général, chacune des méthodes présentées donnait des résultats satisfaisants mais non pas optimaux. Des méthodes classiques, convenablement adaptées et perfectionnées, ont été appliquées avec succès. Ces méthodes comprennent la théorie des probabilités et l'approche Bayesienne, en plus de la théorie des ensembles flous et l'analyse des séries temporelles. Certaines méthodes plus récentes présentent de l'intérêt aussi. Par exemple, ajouter une deuxième mesure à la théorie des probabilités offre de nouvelles possibilités. De la même façon, la théorie des ondelettes permet de faire des mesures locales en généralisant le filtrage morphologique. Ces deux théories ont été présentées et les différents degrés de réussite obtenus ont été démontrés.

Il se pourrait que d'autres, telles que la méthode globale qui est citée dans la communication, ou l'analyse multifractale, qui pourrait être considérée comme la continuation de la théorie des ondelettes, ou bien une toute nouvelle théorie encore inconnue, viennent enrichir la panoplie d'approches possibles. Ces méthodes sont à découvrir, implicitement, parmi les outils utilisés pour le fusionnement des données, comme par exemple dans le cas des systèmes neuronaux.

Tout au long du symposium, les conférenciers ont décrit des méthodes novatrices qui ont conduit à des applications réussies, débouchant sur des systèmes qui ont atteint ou qui sont en passe d'atteindre leur capacité opérationnelle optimale. Ces applications concernent quatre domaines très importants, à savoir:

— l'imagerie

— les systèmes radar, l'identification et la poursuite

— les télécommunications

— les aides à la prise de décisions

Comme il est souvent le cas dans un symposium consacré à des questions relevant des derniers développements d'une science dynamique, bon nombre des participants auront été amenés à considérer ce sujet d'une façon nouvelle, à la lumière des commentaires approfondis des auteurs. Dans la mesure où il est possible de juger déjà de la réussite du symposium, nous sommes convaincus qu'il sera considéré comme très fructueux.

Les résultats présentés à Lisbonne montrent très clairement que cette science prometteuse s'affine, et la diversité de ses applications indique qu'elle est d'un intérêt universel. En particulier, en ce qui concerne les applications militaires, la surveillance, la détection et l'optimisation des aides à la décision sont des domaines qui suscitent de plus en plus d'intérêt. L'influence de cette science sur la formulation des stratégies futures au sein du secteur militaire et en dehors est désormais incontestable.

Nous espérons que ce symposium SPP conduira à l'organisation d'autres conférences afin de permettre de poursuivre l'exploration de ce sujet riche en technologie. Pour le moment, nous nous bornerons à exprimer notre satisfaction devant la qualité des communications et des discussions qui ont animées ce symposium, et à souligner sa pertinence pour les pays membres de l'OTAN.

Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar

(AGARD CP-595)

Executive Summary

Data fusion is a science that has intrigued mankind since his origins. Over time, theories and techniques have continued to develop but never at a pace which, in the past decade, can be described only as explosive. This SPP symposium, held in Lisbon, Portugal, has revealed innovations in methods and technologies leading to broad implementation of the science in a number of systems, particularly for this symposium, in the military area.

This symposium consisted of papers which were classified into four broad technical areas:

- Multisensors and their applications.
- Data fusion methods
- Sensors systems and fusion techniques.
- Operational and experimental validations with their associated technologies.

The papers presented demonstrated the dramatic progress recently in the our understanding of the underlying science, technologies, and techniques of data fusion. The importance of data bases and data banks were underscored by many speakers. Of particular importance were the several strategic parameters critical to data fusion:

- The appropriateness of the choice of the sensors, their number and the necessity for independence that their measurements represent;
- The manner in which the data bank is constituted; and
- The importance of an understanding of the limitations of observations from classes of sensors.

It was clear from the discussions that these topics are not yet resolved and that much remains to be done before the areas for application can be fully realized. For example, neither the classical theories as a bound of Cramer-Rao in processing of the signals or the bound of Shannon in the theory of the information appear sufficient to define, a priori, the optimization of a database.

Once the data base constituted, fusion of the data can be undertaken. A general theme of the symposium was that, in general, the methods presented each gave satisfactory, though not optimum, results. Classical methods, adapted and perfected, have been applied with a much success. These included the theory of probabilities and the Bayesian approach in addition to the theory of the fuzzy sets and the analysis of temporal series. More recent methods also seem attractive. For example, the theory of probabilities, by the introduction of a second measure, offers new possibilities, and wavelets theory that allows local measures by generalizing the morphological filtering, were presented and in each case, demonstrated varying degrees of success.

One can wonder if other methods such as the method of aggregates that has been mentioned in communication, or the multifractal analysis that could be seen as a continuation of the wavelets theory, or even some yet undisclosed new theories, might not, in the near future, enlarge the panoply of approaches. Implicitly, one discovers these methods among the tools used in data fusion, as for example, in the case of neural systems.

Throughout the symposium, the innovative methods described by the authors have led to applications that allowed for successful demonstrations and systems that are nearing or are at full operational capability. These applications focused on four very important areas:

- imagery.
- radar systems, identification, and tracking.
- telecommunications.
- assistance to decision making.

As is often the case of a symposium dedicated to a topic on the cutting edge of a dynamic science, many of the attendees will have been challenged to think in new ways by the thoughtful comments of the authors. To the extent that early estimates of this can be made, we are confident this symposium will be proven to have been particularly fruitful.

The results presented in Lisbon clearly show the extent this promising science is maturing and the diversity of applications shows that it's relevance is universal. Specifically in the area of military applications, surveillance, detection, and the optimization of decision-making are domains that are receiving increasing attention. The influence of this science on future strategies in and out of the military arena is no longer in doubt.

It is our hope that this SPP symposium will lead to other conferences in which the exploration of this technologically-rich subject will continue. For the present symposium, it is sufficient to admit satisfaction in the quality of presentations and discussions, and the relevance that it represents for the NATO.

Contents

	Page
Executive Summary	iii
Synthèse	iv
Theme/Thème	ix
Sensor and Propagation Panel	x
	Reference
TA10 - Advanced Information Processing for Multisensory Systems I.E.P.G. Cooperative Technological Programs Technological Area 10 by C. de Leeuw and V. Chalmeton	1
Aircraft Sensor Data Fusion: An Improved Process and the Impact of ESM Enhancements by C.A. Noonan and M. Pywell	2
Multi-Sensor Remote Sensing for Military Cartography by A.C. van den Broek, P. Hoogeboom and M. van Persie	3
Design of a Multi-Sensor System for 3D Site Model Acquisition and Exploitation by I. Tannous, S. Gobert, T. Laurençot, J.-M. Dulac and O. Goretta	4
Airborne Dual Sensor mmW-Signatures of Maritime Targets and Sea-Clutter by R. Makaruschka and H. Essen	5
Data Fusion for Long Range Target Acquisition by P. Verlinde, D. Borghys, C. Perneel and M. Acheroy	6
Paper 7 Withdrawn	
Two-Frequency Millimeter Wave SAR by H. Schimpf and H. Essen	8
Système Multi-Senseurs Radar et Optronique pour l'Attaque au Sol by E. Mignot, J.P. Mestre and D. Levaillant	9
IR-Signatures of Ships: A Classification Feature Within a Sensor Network by H.D. vom Stein, J. Günther and K. Haese	10
Advanced Electro Optic Targeting System by R.W. Jacobs and H.S. Lapp	11
Paper 12 Withdrawn	
Méthode d'Evaluation des Performances d'une Poursuite Multicapteur by M. Desbois	13

Reconnaissance de scènes dynamiques à partir de données issues de capteurs: le projet PERCEPTION 14
by C. Tessier

Multisensor Data Fusion for Automatic Recognition of High Value Surface Targets 15
by K.-H. Bers, H. Essen, K. Jäger and H. Schimpf

A KF-Based Integration System for Land Vehicle Tracking from Real DGPS and INS Data 16
by C.S. Regazzoni, A. Teschioni and G. Tacconi

ARTAS: Multisensor Tracking in an ATC Environment 17
by R.A. Hogendoorn and W.H.L. Neven

Low Altitude Wave Propagation Effects in a Transmitter Independent Receiver Network 18
by N.J. Farsaris and P.P. Stavroulakis

Non-Linear Prediction of Ionospheric Parameters Important for Forecasting the Battlespace Environment 19
by N.M. Francis, A.G. Brown, A. Akram, P.S. Cannon and D.S. Broomhead

Fusion de données et fiabilité des transmissions HF 20
by C. Goutelard, J. Caratori and C. Pautot

Fusion de données et sondeur à extrêmement faible puissance pour télécommunications 21
by C. Goutelard and C. Pautot

HF Transmitter Location using Super-resolution DF and Ionospheric Sounding 22
by H.J. Strangeways

Smart Antenna Array Processing for Spread Spectrum Wireless Communication Systems 23
by Y.F. Huang, S. Kapoor, S. Gollamudi and S. Nagaraj

Propagation Diversity Constraints on Reception of GPS Signals at High and Equatorial Latitudes 24
by J. Aarons and B. Lin

Meteorological Data Fusion to Assess EM/EO Propagation Conditions in Coastal Environments 25
by J. Rosenthal, R. Helvey, P. Greiman and M. McGovern

Utilisation d'Informations Contextuelles dans des Algorithmes de Pistage Multicapteur 26
by V. Nimier

SAGESSE: un modèle de représentation de données pour la Fusion de Données Symboliques 27
by Y. Pollet and S. Robidou

Adaptive Intensity Matching Filters: A New Tool for Multi-Resolution Data Fusion 28
by S. de Béthune, F. Muller and M. Binard

The Fusion of Organic and Non-Organic Identity Information Sources Using Evidential Reasoning 29
by É. Bossé, J. Roy and S. Paradis

Pixel Fusion for Roads Detection and Classification 30
by S. Fabre, X. Briottet, P. Marthon and A. Appriou

Sensor Synergetics: The Design Philosophy of the Sensor Fusion Demonstrator and Testbed (SFD) at TNO-FEL 31
by A.J. van der Wal

The Nemesis Identification Data Fusion Demonstrator 32
by P.N. Griffith, L.A. Hooper and W.M. Everitt

Communications Management in Battlespace Data Fusion 33
by R.H. Deaves, D. Nicholson, P. Greenway and P. Vangasse

NATO Alliance Ground Surveillance Interoperability 34
by P.J. Lenk and G. Retzer

Imagery and Signals Intelligence Data Fusion: Issues and Methods 35
by E.L. Waltz and J. Hart

NOSTRAMARINE: Un Concept de Détection Multistatique Adapté à la Surveillance des Cibles Basse Altitude 36
by M. Lesturgie and M. Flécheux

The NATO Data Fusion Demonstrator Project 37
by J.M. Skinner

Pixelless Infrared Imaging Using Quantum Wells 38
by G.J. Brown and F. Szmulowicz

Nonlinear Optical Frequency Conversion Materials for Multi-Band Remote Sensing and Surveillance 39
by S. Caracci, N. Fernelius, M. Ohmer and D. Zelmon

Fusion d'informations et réseaux physiques pour les systèmes multi-senseurs sol-air 40
by C. Nahum

Development of a Wideband Airborne Laser Data Link 41
by R.A. Gill and R.J. Feldmann

Fusion of Visible and Thermal Imagery Improves Situational Awareness 42
by A. Toet, J.K. IJspeert, A.M. Waxman and M. Aguilar

Theme

Future weapon systems, whether ground, air or space based, will have to operate efficiently in a highly complex, hostile and dynamic battlespace environment. In this environment, a broad array of deception techniques (ECM) may be employed, and the need to fully discriminate among a dense array of objects will be increasingly difficult. In the past, emphasis has been placed on optimizing individual sensors for specific battlespace functions. Present and future threats will demand more capabilities than can be delivered by these individual sensors.

However, it is likely that when data from many individual sensors are combined in an intelligent and perhaps adaptive manner such a combination may provide the commander information needed to counter these threats. By using data from multiple sensors, operating in independent domains which are often sensitive to widely different physical phenomena, the threats posed by deception techniques may be better neutralized. Benefits to military systems can be anticipated in the areas of communication, target acquisition, target identification and delivery accuracy, together with overall system integrity and robustness.

In recent years, significant advances have been made in data fusion techniques and algorithms. Combined with the exponential growth of available computing power and memory density, real-time multisensor data fusion is now becoming feasible and affordable. Integrated C3I will permit an architecture applicable to efficient data exchange and may exploit the dramatic increase in computing power now available.

Multisensor data fusion can be applied to a variety of systems appropriate to communications, surveillance and target recognition. Data from multiple sensors will lead to an improved and more complete picture of the battlespace environment – e.g., location, heading, strength, etc. of both friendly and hostile forces. Early applications of data fusion techniques in communications have been employed for some time. For example, in cellular mobile radio or monofrequency broadcasting systems, diversity is used to improve reliability. At a more advanced level, the reliability of data transmissions requires a characterization of transmission channels which results from measurements (ground and space based) of electron density, magnetic fields, meteorological conditions, solar activity, etc. Additional information can be extracted from waveforms using data fusion. This may provide a robust system while increasing data flow.

Thème

Les systèmes d'armes futurs, qu'ils soient aéroportés, spatioportés ou basés au sol, devront être exploités dans un théâtre de bataille dont l'environnement sera hostile, évolutif et extrêmement complexe. Une large gamme de techniques de déception (ECM) pourra être mise en œuvre et la discrimination de la cible dans une concentration dense d'objets deviendra de plus en plus difficile. Dans le passé, l'accent était mis sur l'optimisation de senseurs individuels destinés à remplir certaines fonctions spécifiques du théâtre d'opérations. Or, contrer la menace actuelle et future exige plus de capacités que les senseurs individuels ne peuvent fournir.

Cependant, il est vraisemblable que lorsque les données fournies par un certain nombre de senseurs individuels seront combinées de façon intelligente et, peut être adaptative, la combinaison qui en résultera fournira aux commandeurs militaires les informations demandées pour contrer ces menaces. Il se pourrait qu'en exploitant les données fournies par de multiples senseurs, fonctionnant dans les domaines indépendants, sensibles à des phénomènes physiques très diverses, les menaces posées par les différentes techniques de déception soient plus faciles à neutraliser. Des avantages pour les systèmes militaires sont envisageables dans les domaines suivants: les communications, l'acquisition et l'identification de la cible, la précision du tir, la robustesse et l'intégrité globale du système.

Ces dernières années, des avancées remarquables ont été réalisées dans le domaine des algorithmes et des techniques de fusionnement des données. Avec la croissance exponentielle de la puissance de calcul et de la densité mémoire, le fusionnement des données multisenseur en temps réel devient faisable et abordable. Le C3I permettra la réalisation d'architectures qui autoriseront un échange de données performant et il fournira sans doute le moyen d'exploiter la croissance spectaculaire de la puissance de calcul qui est actuellement disponible.

Le fusionnement des données multisenseur s'applique à un éventail de systèmes qui relèvent des communications, de la surveillance et de la reconnaissance de la cible. En ce qui concerne la surveillance et la reconnaissance de la cible, l'apport de données multisenseur se traduira par une représentation plus complète de l'environnement du champ de bataille, c'est-à-dire position, direction et importance des forces amies et ennemies. Les premières applications des techniques de fusionnement des données sont en service depuis un certain temps. Par exemple, dans les systèmes de radiodiffusion monofréquence ou dans les radios mobiles cellulaires la technique de diversité est utilisée pour accroître la fiabilité. A un niveau plus avancé, la fiabilité de la transmission des données passera par la caractérisation des voies de transmission, qui sera à son tour tributaire de la mesure (terrestre et spatiale), de la densité des électrons, des champs magnétiques, des conditions météorologiques, de l'activité solaire etc. Des informations complémentaires pourront être extraites des formes d'onde par le fusionnement des données. Cette approche permettra de réaliser un système robuste, tout en augmentant le flux de données.

Sensor and Propagation Panel

Chairman: Mr. F. Christophe
Dept. Micro-Ondes
ONERA-CERT Toulouse
BP 4025
2 Avenue E. Belin
31055 Toulouse Cedex
France

Deputy Chairman: Dr. P.S. Cannon
Tactical Communications Dept.
D705 Building
Defence Research Agency
St. Andrews Road
Malvern, Worcs. WR14 3PS
UK

TECHNICAL PROGRAMME COMMITTEE

Co-Chairmen:

Prof. C. Goutelard (FR)
Prof. A.K. Hyder (US)

Programme Committee Members:

Prof. M. Acheroy (BE)
ICA P. Fuerxer (FR)
Dr. R. Klemm (GE)
Mr. G. Wyman (UK)
Prof. C. Brown (US)
Dr. D. Yavuz (NATO C3 Agency)

PANEL EXECUTIVE

Lt. Colonel G. Del Duca, IAF

Mail from Europe:
RTA-NATO
Attn: SPP Executive
7, rue Ancelle
92200 Neuilly-sur-Seine
France

Phone: 33 (0) 1 55 61 22 68

From North America:
RTA-NATO/SPP
PSC 116
APO AE 09777

Fax: 33 (0) 1 55 61 22 99

TA10 - ADVANCED INFORMATION PROCESSING FOR MULTISENSORY SYSTEMS

I.E.P.G. COOPERATIVE TECHNOLOGICAL PROGRAMS TECHNOLOGICAL AREA 10

Charles de Leeuw
Hollandse Signaalapparaten B.V.
Applied Systems Research Department
P.O. Box 42
7550 GD Hengelo, The Netherlands
e-mail: deleeuw@signaal.nl

Vincent Chalmeton
Thomson-CSF / TCAR
Fusion Competence Centre
6 Rue Nieuport
78852 Velizy cedex, France
e-mail: chalmeto@thomson-csf.fr

SUMMARY

Real-time interpretation of large amounts of battlefield information by both operator and system is increasingly difficult due to steadily growing complexity, and integration of systems. Therefore data fusion, i.e. combining data from several sources in order to obtain a global and coherent view on the battlefield, both at sensor level (multi sensor data fusion) and at abstract, strategic information level (information fusion), becomes more important.

The main elements of the TA10 demonstrator project are (1) the fusion of data from different platforms and different sensors, (2) the management and the allocation of the platforms, drones, helicopters, and sensors, (3) situation assessment and (4) the determination of the requirements for real-time data fusion. To perform the fusion task, the required input consists of plots/tracks from individual sensors (including false alarms) and information from other sources, e.g. intelligence, and geographical information. TA10 focuses on the integration of information from different platforms (various non-colocated platforms and drones) and different sensors (radar, optronic and ESM), instead of using only the sensors on a single platform. TA10's sensor management process uses the results from the fusion processes in order to allocate resources to improve the quantity and the quality of the information. The result of the project is a demonstration of the fusion process, the sensors management and the real-time capabilities using simulated sensors. The TA10 demonstrator project therefore covers topics like: Reasoning about hypotheses, alternatives, uncertainty, time and space; Statistical analyses; Implementation techniques, such as functional programming and knowledge-based technology.

In this paper we will describe the different concepts used for the multi sensor data fusion processes, and the distributed asynchronous software architecture chosen. Furthermore, the interaction between Data Fusion, Situation Assessment and Sensor/Platform Management will be explained in order to illustrate the applied technologies for advanced information processing for multisensory systems. In the paper some of the first results from this on-going project will be presented.

THE PARTICIPANTS

The project is set up as an international cooperation between industries and research organisations from France and The Netherlands. The contributing partners are:

- Thomson-CSF/TCAR (Vélizy, France)
- Matra Systemes & Information (Val de Reuil, France)
- Hollandse Signaalapparaten (Hengelo, The Netherlands)
- TNO-FEL (The Hague, The Netherlands)
- NLR (Amsterdam, The Netherlands)

1. SYSTEM OVERVIEW

The primary goal of this project is to design and implement a demonstrator showing the capabilities of multi sensor data fusion through advanced information processing. The TA10 demonstrator (TAD) is a system that demonstrates the core element of a battlefield ground surveillance system by simulation in near real-time. The core element of the battlefield ground surveillance system consists of the functional parts: Data Fusion, Situation Assessment, Sensor/Platform Management and Sensor/Platform simulation. The surveillance system observes the real world through a non-collocated heterogeneous multisensory system. The perceived world is organized into a world model by using appropriate data fusion techniques in order to represent it in a synthetic way to the operator. The real world and the multiple sensors are simulated in a non-interactive way. So, the TAD works on simulated data instead of real sensor signals. It is the aim of the study to define and implement a demonstrator able to show the use of advanced data processing techniques on data fusion under (near) real-time constraints.

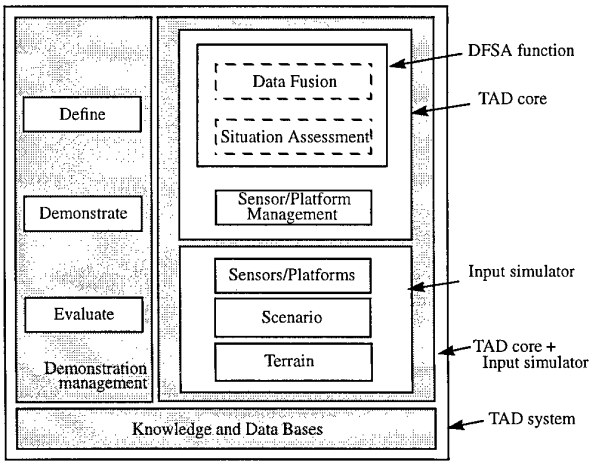


Figure 1 - Functional partitioning of TAD system

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

Figure1 identifies the main functionality of the TAD. The architecture itself, based upon generic principles, is described in the next sections and will show that the functions comprising the TAD core are distributed across a number of data fusion and sensor management nodes due to the geographical distribution of the sensor platforms:

1. The Input Simulator generates sensor reports, the perceived truth, as result of battlefield and sensor/platform simulation.
2. The sensor reports are processed by the Data Fusion and Situation Assessment process in order to obtain an inferred representation, i.e. a world model.
3. Based on the inferred representation, the Data Fusion and Situation Assessment process issues Sensor/Platform requests to the Sensor/Platform Management.
4. These requests result in the update of the sensor management plan, considering sensor performance characteristics.
5. The sensor management plan will be the basis for generation of sensor/platform commands to the Input Simulator, thereby closing the control and information loop.

1.1 GENERIC ARCHITECTURE

The TAD relies on an architecture which is a derivation from a generic distributed multi sensor data fusion architecture (figure 2 **[6]**). From the resulting definition of the generic architecture, the *demonstrator architecture* (figure 3) is derived. The following properties emerge from this approach:

1. The generic architecture can be tailored to other applications and domains. The demonstrator architecture (figure 3 and 4) represents such a specific application for battlefield surveillance.
2. The generic architecture is reconfigurable, which means that the architectural elements (nodes) and interfaces are defined in such way that new architectures (corresponding to specific fusion/processing schemes) can be easily generated and their performance evaluated.
3. The generic architecture is flexible in adopting new or other sensors and new or other platforms. The demonstrator architecture relies on a specific set of platforms and sensors.

These considerations imply that the demonstrator architecture is a specific instance of the generic architecture. Throughout the design and development of the demonstrator, this generality and flexibility is pursued. The demonstrator design pursues generic interfaces, processes and algorithms as much as possible and is only specific there where it is unavoidable. In particular, the fusion nodes have a common structure, which is tailored to the specific function they perform.

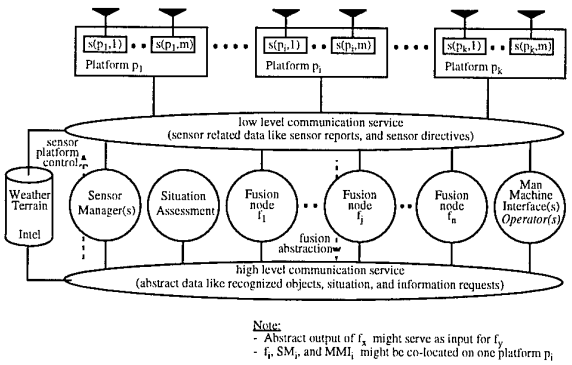


Figure 2 - Generic distributed multi-sensor data fusion architecture

The generic DFSA system architecture consists of the following elements:

- A number of platforms p_1... p_k. Each platform has mounted a number of sensors s(p,1)... s(p,m).
- A number of data fusion nodes f_1... f_n that fuse data (from sensors, i.e. sensor/platform reports, and/or from other fusion nodes, i.e. fusion reports).
- A situation assessment node, that takes as input fusion reports from one fusion node and performs situation assessment. It also makes directives for the management of the fusion process and the sensors/platforms.
- One or more, probably hierarchically structured, sensor/platform management nodes (hierarchical control: from general, high level sensor or platform directives to specific, low level sensor or platform commands).
- A low level communication service that provides data flow between the processes and platforms. Examples of data are sensor (detection) reports and platform or sensor commands. It primarily transports sensor or platform related data messages.
- A high level communication service that provides data flow between the processes themselves. Examples of data are situation reports and high level control commands to sensor manager.
- One or more man-machine-interfaces, which in general have access to all data stored in the sensors/platforms, nodes or data bases.

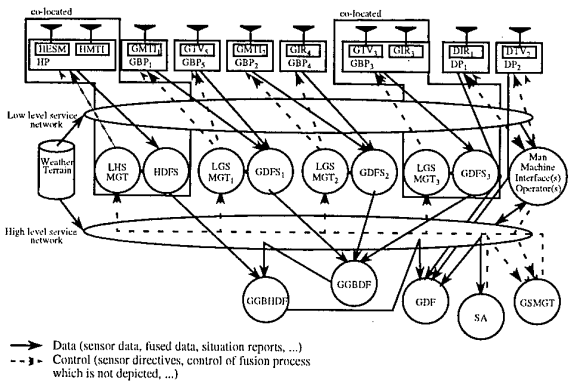


Figure 3 - TA10 Demonstrator Architecture (as a specific instance of generic architecture)

The data fusion nodes, situation assessment node and the sensor/platform management nodes and their interaction patterns (i.e. processing scheme) make up a specific architecture such as the TAD core (i.e. demonstrator architecture). These elements are configured into a specific architecture, processing the data from a specific set of platforms and sensors. A specific instance of the generic architecture for the TAD can be found in figure 3.

The distinction between low level and high level communication services is somewhat specific, suggesting the existence of a more generic architecture. However, this distinction is important, because it introduces the difference between two levels of data exchange, which also enhances the construction of generic (invariant) interfaces. In practice, these two services make use of the same underlying network.

The reconfigurability and flexibility with respect to sensors,

platforms, processing scheme and application domain require generic interfaces of the architectural elements, and a common (internal) structure for the elements. In this case, a specific architecture with specific elements is nothing more than an instance of the generic architecture tailored to the application with respect to available sensors and platforms, application domain and selected processing scheme. The demonstrator architecture is an instance of such a generic architecture, with a fixed fusion network.

1.2 DEMONSTRATOR ARCHITECTURE

The TAD core is based on a network of processing nodes (fusion nodes, assessment node, and sensor/platform management nodes) distributed on static or moving platforms connected by a communication network. The TAD is not focused on the simulation of the communication network but shall take into account the existence of a communication network. A way to do so, is to express the communication requirements of the TAD core on a fully networked asynchronous environment:

a. not local (distinct locations),
b. dynamically configurable (availability of processing nodes),
c. transport protocol independent (diverse and evolving communication resources),
d. dynamically link configurable (availability of communication resources).

The a) and c) requirements are right for the anticipated future evolution of the TAD (distribution of processing nodes on distinct hardware resources with diverse communicating capabilities) and are not considered as demanding requirements for the TAD to be developed in the project. The points a), c) and d) are to be fulfilled by the communication network services. The d) point is not required for the TAD. The TAD uses a transport level connection oriented reliable byte stream service for the communication between processing nodes.

The design of the TA10 demonstrator is based on the specific instance of the generic architecture as described above. This specific demonstrator architecture defines the selected sensor/platform configuration and the processing scheme being the TAD core. The demonstrator architecture is depicted in figures 4 and 5.

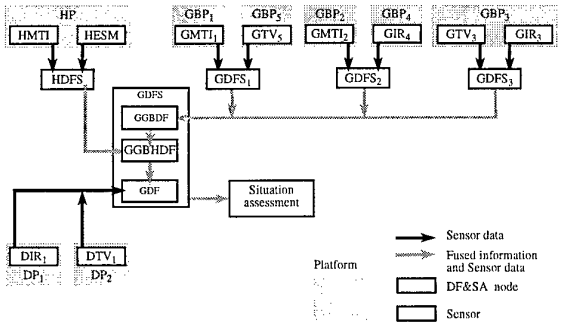


Figure 4 - Demonstrator Architecture: Fusion nodes

Figure 4 shows the data flow between the sensor platforms and fusion nodes as a result of the processing of incoming sensor data. Figure 5 shows the flow of control between the sensor management nodes and sensor platforms as a result of the Data Fusion and Situation Assessment process.

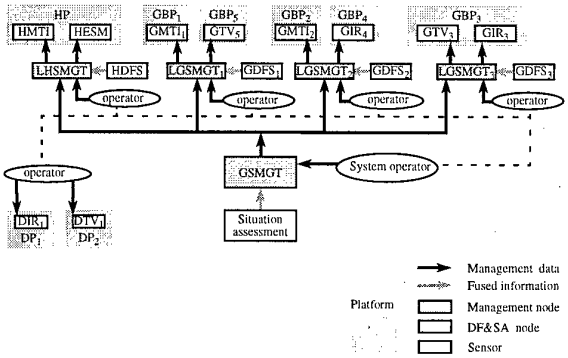


Figure 5 - Demonstrator Architecture: Sensor management

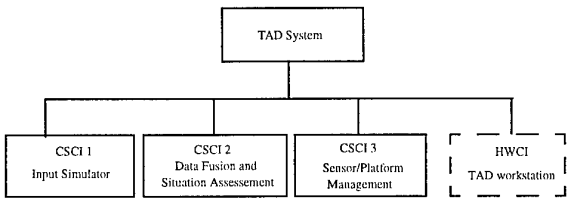
The following parameters define the specific TAD architecture (see also list of abbreviations):

- Application domain
  - Battlefield surveillance (defined by scenario)
- Platforms
  - Heliborne Platform (HP) with sensors HESM and HMTI
  - Ground Based Platform 1 (GBP_1) with sensor $GMTI_1$
  - Ground Based Platform 2 (GBP_2) with sensor $GMTI_2$
  - Ground Based Platform 3 (GBP_3) with GTV_3 and GIR_3
  - Ground Based Platform 4 (GBP_4) with sensor GIR_4
  - Ground Based Platform 5 (GBP_5) with sensor GTV_5
  - Drone Platform (DP_1) with sensor DIR_1
  - Drone Platform (DP_2) with sensor DTV_2
- Fusion nodes (comprising TAD core Data fusion function)
  - HDFS fusing HMTI and HESM (fusion of different types of sensors on the same platform)
  - $GDFS_1$ fusing $GMTI_1$ and GTV_5 (fusion of different types of sensors on different platforms)
  - $GDFS_2$ fusing $GMTI_2$ and GIR_4 (fusion of different types of sensors on different platforms)
  - $GDFS_3$ fusing GTV_3 and GIR_3 (fusion of different types of sensors on the same platform)
  - GDFS: This composed node receives the fused data coming from the previous nodes (HDFS, $GDFS_1$, $GDFS_2$ and $GDFS_3$). This system contains three hierarchical nodes:
    - GGBDF fusing $GDFS_1$, $GDFS_2$ and $GDFS_3$. This node has to fuse data coming from the three ground based data fusion nodes;
    - GGBHDF fusing GGBDF and HDFS. This node has to fuse data coming from the GGBDF outputs and the Heliborne Data Fusion System (HDFS) outputs;
    - GDF fusing GGBHDF, DIR_1 and DTV_2. This node has to fuse data coming from the GGBHDF outputs with the information extracted from the data collected by the drone(s).
- Situation assessment nodes (comprising TAD core Situation Assessment function)
- Sensor and platform management nodes (comprising TAD core Sensor/Platform Management function)
  - GSMGT taking as input orders from operator, (local) data fusion nodes and SA and prepare orders for LHSMGT, $LGSMGT_1$, $LGSMGT_2$, $LGSMGT_3$ (control

of DP_1 and DP_2 is also supported by the GSMGT).
  . LHSMGT taking as input orders from GSMGT, HDFS and possibly a local operator (on board), and prepare orders for HP (i.e. pilot, flight path), HESM and HMTI.
  . $LGSMGT_1$ taking as input orders from GSMGT, $GDFS_1$ and possibly local operator and prepare orders for GBP_1, $GMTI_1$, GBP_5 and GTV_5.
  . $LGSMGT_2$ taking as input orders from GSMGT, $GDFS_2$ and possibly local operator and prepare orders for GBP_2, $GMTI_2$, GBP_4 and GIR_4.
  . $LGSMGT_3$ taking as input orders from GSMGT, $GDFS_3$ and possibly local operator and prepare orders for GBP_3, GTV_3, and GIR_3.
- Man Machine Interface
  In general and from an operational point of view, the man machine interface which addresses the interaction needs of one or more operators has access to all data and information, whenever useful. Possibly, MMI parts can be co-located on platforms or other distant places, having immediate access to the local data and providing operator control to platform and sensors. E.g., an MMI part might be located on the heliborne platform that enables the local operator to examine HMTI and HESM data and the results from fusion node HDFS, as well as to control sensors / platform.

2. MAIN FUNCTIONAL COMPONENTS

Apart from supporting knowledge and databases the TAD consists mainly of one HardWare Configuration Item (HWCI) and three Computer Software Configuration Items (CSCI):
- 'TAD Workstation' HWCI
  The TAD workstation is the processing resource which supports the computing and operator interaction needs of TAD.
- Input Simulator CSCI
  The CSCI simulates the battlefield and sensors/platforms with the generation of sensor reports as the objective.
- Data Fusion and Situation Assessment CSCI
  This CSCI implements the data fusion and situation assessment nodes with as main functions to provide the operator a synthetic picture of the battlefield and guidance to Sensor/ Platform Management functions through requests.
- Sensor/Platform Management CSCI
  This CSCI implements the sensor/platform management nodes. Their main function is to provide steering commands to sensors/platforms based on data fusion, situation assessment or operator requests in order to have an efficient and effective deployment and use of these sensors/platforms.



2.1 SCENARIO AND SIMULATION

OPERATIONAL AREA

The area where the scenario will take place is a 20 x 20 km square area. This terrain is digitized taking into account (both during simulation and during the execution of the Data Fusion, Situation Assessment and Sensor/Platform Management processes) relief (hills), civil engineering, and vegetation aspects, in order to address intervisibility and background problems.

ENEMY AND NEUTRAL TARGETS

The agreed represented scenario includes 100 potential targets:
- 50 civilian ones which can move randomly on road network
- 20 military vehicles (cars, light vehicles, tanks) which move in accordance with military doctrine. In opposition to cars, tanks are not limited to roads. The scenario will only let them leave the roads if in accordance with military doctrine.
- 30 military units (Vehicles, troops) moving in accordance with military doctrine.

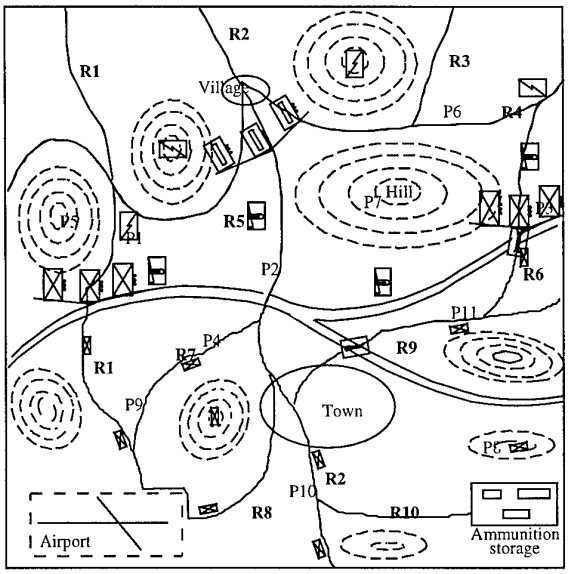


Figure 6 - Snapshot battlefield simulation at To+150min.

OWN SENSORS

The sensors are all friends. The enemy sensors are considered only as emitters. The friendly sensors include:
- 1 single airborne ESM (bearing, sensor characteristics) + Long range MTI radar (range, bearing, doppler),
- 2 fixed ground medium range MTI radar,
- 1 fixed ground IR + TV camera (bearing, bearing radial velocity, elevation, elevation radial velocity),
- 1 fixed ground IR camera,
- 1 fixed ground TV camera,
- 1 drone equipped with IR camera,
- 1 drone equipped with TV camera.

All landbased sensors are fixed and distributed along the FLOT (Forward Limit of Own Troops) in order to assure a coverage overlapping (see figure 7). As there is no battle, FLOT is used instead of FEBA (Forward Edge of Battle Area).

The drone platforms are sent at the appropriate time for a 30 to 120 minutes mission. That mission is planned by the system and validated by the operator, as deducted from the data fusion system and the current assessed situation especially about the areas with low quality information and/or potential risks.

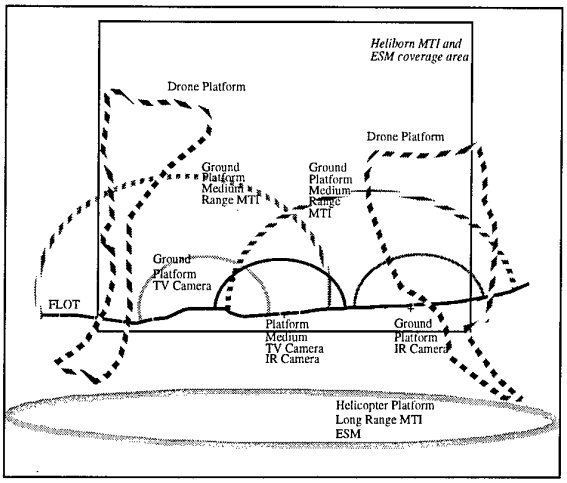


Figure 7 - Own sensors deployment

The airborne MTI radar is installed on a helicopter. The helicopter flight path has the form of a race course (60x10 km) at a 40 km stand-off distance from the frontier. It delivers three global radar pictures of the area per track.

Another important point is to distinguish the theoretical detection range of the sensors and their practical detection range. The theoretical range is the range obtained in the best conditions for the meteo and without any masking nor by the relief nor by the terrain coverage.The practical range will be limited more or less by the meteorological conditions, and mainly by the terrain relief and coverage. The ranges shown in the scenario figure above are the average practical ranges in operational conditions.

SIMULATION REMARKS

1. The battlefield is chosen in the south of France in a zone with moderate relief and few forests, so the masking effects will not dramatically decrease data fusion possibilities.
2. Due to microrelief, the real detection, even in the detection range, will not be equal to 100%.
3. The battlefield includes several possible military objectives to get an anticipated estimation of the situation assessment.
4. The enemy is assumed to be limited to the battlefield area and does not to have interactions with other troops outside. It will be thus a punctual action near the frontier between two stabilized lines.
5. The operator is assumed to know what the potential military objectives in the observed zone are.
6. At time t=0 the operator is assumed to have no other information then:
   - the battlefield map,
   - the potential military objectives in the observed zone,
   - the enemy military doctrine,
   - the information on his own sensors.
7. The enemy does not react to the observation.
8. During the demonstration there is no reaction from friendly troops.
9. For the operator (or observant) the starting state of the external world is totally unknown. After 15 minutes of evolution of civil targets, new targets of the military type can be injected into the field:
   - Military vehicles up to 1 per minute, with max. of 20.
   - Military units up to 1 each 5 minutes, with max. of 30.

   These injected targets constitute the initial total of 50 military objects (vehicles or units). Injection of targets will only happen at the borders of the 20 x 20 km battlefield.

2.2 KNOWLEDGE AND DATABASES

This section describes the data and knowledge in the TAD. The data and knowledge needed for the TAD consists of world model data/knowledge. The world model contains the data and knowledge that describes the battlefield. An object-oriented approach is followed by defining world model objects in a way that has an equivalent on the battlefield. Each object encapsulates three types of data or knowledge:

1. *Static data:* static attributes of objects of the world. Static data only consists of data that does not change during a demonstration, e.g. the typical size of a civil target.
2. *Dynamic data:* dynamic attributes of objects of the world. Dynamic data change during a demonstration and are defined as a function of time, e.g. the position of a military target.
3. *Behaviour knowledge:* knowledge that describes the behaviour of objects in the world, e.g. how the terrain conditions influence the speed of a military target.

The Input Simulator uses an instance of the world model to perform a simulation according to a specific scenario. The TAD core uses and maintains one or more instances of the world model with as goal to reconstruct the simulated scenario on basis of unreliable sensor reports. In this case, the objects in the world model are track hypotheses. The different instances of the world model are implemented in each CSCI or Computer Software Component (CSC), according to its specific needs. Several CSCs might share the same instance.

2.2.1 WORLD MODEL

The world model consists of track hypotheses. Each hypothesis contains data and/or knowledge about the natural object as found in the battlefield. It is assumed that this set of objects contains the basic data and/or knowledge for constructing the world model. The following objects are identified in the world model: Scenario, Battlefield, Atmosphere, Terrain, Entity, Sensor, Platform, Military Target, Military Unit, and Civil Target.

Figure 8 describes the hierarchical relationships in terms of 'HAS A' and 'IS A' relations. The relation 'HAS A' indicates that an object is composed of its parts (which are again objects), e.g. the object 'battlefield' consists of the objects 'atmosphere', 'terrain', and a number of 'entities'. The relation 'IS A' indicates a subtype/supertype relationship,e.g. the objects 'platform', 'sensor', 'military unit', 'military target' and 'civil target' are all a subtype of the object 'entity'. This implies that certain attributes of 'entity', such as 'location of the object', are inherited by the subtypes.

The object 'Battlefield' is regarded as having an atmosphere (e.g. weather conditions), terrain characteristics and a set of entities, such as military targets and sensors. A 'scenario' is regarded as multiple instances of 'battlefield' as a function of time in order to capture the dynamic aspect of the battlefield (note that each object will have a time line, i.e. history, present and possibly future). The 'HAS A' relation between 'platform' and 'sensor' indicates that a platform consists of one or more

sensors. The 'HAS A' relation of 'military unit' to itself models the fact that a military unit can be composed of subunits (e.g. a battalion is composed of companies), where the smallest military unit (i.e. platoon) is composed of 'military targets' such as tanks and personnel.

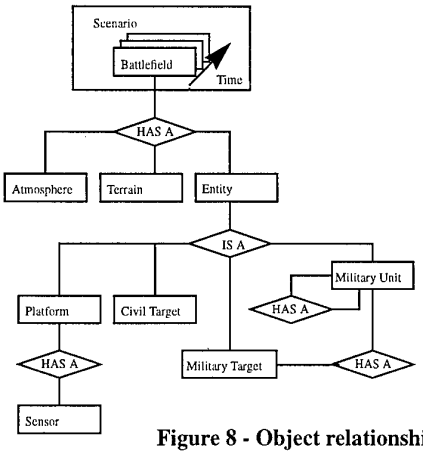


Figure 8 - Object relationships

2.2.2 OBJECT IDENTIFICATION

For each object the name, description and examples of the static and dynamic data is given. The behaviour knowledge associated with each object is only named.

Name	Description	Static data	Dynamic data	Behaviour
Scenario	The Scenario object specifies the battlefield as a function of time. The relation of objects with time is necessary in order to capture the dynamic aspects of the battlefield. This means that objects and their attributes are defined as function of time.	Time reference	Time	Incorporated in the behaviour of each object as a function of time.
Battlefield	The Battlefield object specifies the real world at an instant of time. Main aspect to identify this object is to represent a common frame of reference for the objects that are being modelled: atmosphere, and terrain, and entity	Coordination grid	Defined by the objects: Atmosphere, Terrain and Entity	Defined by the objects: Atmosphere, Terrain and Entity
Atmosphere	The Atmosphere object specifies atmospheric aspects of the battlefield. This object consists of objects, such as: daylight object and weather object		Humidity, temperature, Location	Environment behaviour
Terrain	The Terrain object specifies information about the battlefield terrain. Two types of terrain objects exist: natural objects, such as mountains and lakes and man-made objects, such as roads, bridges, and canals.	Location of lakes	Bridges might be destroyed	Environment behaviour
Entity	The Entity object is an abstract representation of the objects: Platform, Sensor, Civil target, Military target, and Military unit. Entity object provides generic attributes which are inherited by its subtypes. For example, typical the attributes are: speed, location, and size.	Size and maximum speed	Speed and location	Defined by the objects: Platform, Sensor, Civil Target, Military Target, and Military Unit.
Platform	The Platform object specifies characteristics of a platform.	Set of sensors located on a platform	Platform altitude	Sensor/Platform Behaviour
Sensor	The Sensor object specifies sensor characteristics.	Frequency (maximum and minimum)	Sensor altitude	Sensor/Platform Behaviour
Civil Target	The Civil target object specifies the characteristics of targets such as buses and cars.		Number of civil vehicles involved	Civil Target Behaviour
Military Target	The Military target object specifies military enemy targets.	Armed	Threat indication	Military target Behaviour
Military Unit	Military Unit object specifies which targets form a unit and it also specifies the characteristics of this unit.	Typical structure (composition)	Structure	Unit Behaviour

2.2.3 BEHAVIOUR KNOWLEDGE

The behaviour knowledge, as listed in the previous section for each object, consists of knowledge of the behaviour of single object and possibly in relation with other objects. The knowledge capturing the behaviour of the world model objects is identified in the following five behaviour items:

Name	Description	Involved objects
Environment Behaviour	Knowledge describing the interaction between the atmosphere and the terrain. E.g. the rain makes the terrain muddy.	Atmosphere, Terrain
Sensor/ Platform Behaviour	Knowledge describing the behaviour of sensors/platforms with respect to the type of sensor, the current atmosphere, terrain, and targets to be detected, tracked and identified. E.g. the weather influences the sensor performance.	Atmosphere, Terrain, Sensor/Platform, Military Target, Civil Target
Military Target Behaviour	Knowledge describing the behaviour of military targets with respect to terrain, atmosphere, type of military target, and military units (command and control). E.g. the terrain influences the speed of a military target.	Terrain, Atmosphere, Military Target, Military Unit
Civil Target Behaviour	Knowledge describing the behaviour of civil targets with respect to terrain, atmosphere, and type of civil target. E.g. the terrain influences the speed of a civil target.	Terrain, Atmosphere, Military Target
Unit Behaviour	Knowledge describing the behaviour of military units (ORBAT: Order of Battle) with respect to terrain, atmosphere, military units (structure and targets). E.g. a muddy terrain slows down the movement of troops.	Terrain, Atmosphere, Military Target, Military Unit

2.3 DATA FUSION AND SITUATION ASSESSMENT

Figure 4 shows that the TA10 data fusion occurs in a number of different distributed data fusion nodes. In the different fusion and assessment nodes sensor information or fused information is taken together to produce a more complete overview of the battlefield. The information is based on plots, tracks or already fused information **[1][2][3][5]**.

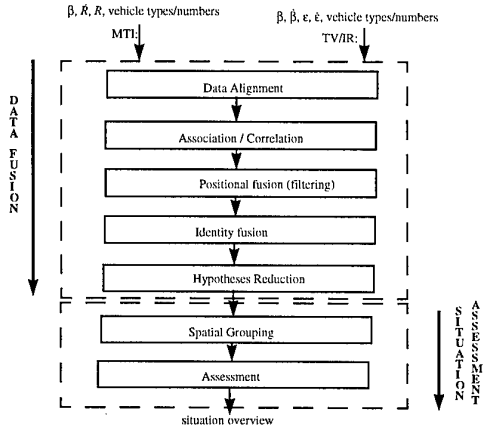


Figure 9 - Overview of fusion and assessment process

N.B.: In this paragraph the various items of the fusion process will be discussed. Although various combinations of sensors, each with their own peculiarities, are used, we will mainly limit ourselves to the example of the fusion node $GDFS_1$, in which MTI and TV data is fused.

2.3.1 DATA ALIGNMENT

Data Alignment can be split in alignment in time (Temporal alignment) and in space (Spatial alignment). The result of Temporal alignment is a set of measurements in the same time reference. The measurements from each sensor are expressed with respect to the sensor position. To avoid annoying covariance terms when a measurement is correlated with a track, the predicted position for this track is expressed in the coordinate system of the corresponding sensor. In general, measurements

from two different sensors can be combined as desired, if it is specified how to translate measurements from the two sensors in their own coordinate system, to and from a global coordinate system. In practice, more efficient methods will be used. E.g., in the current situation measurements will be translated from the TV coordinate system to the MTI coordinate system and visa versa. Only at the end of the fusion process, when the results are transferred, a translation to the global coordinate system will be made.

E.g., the MTI radar gives range and azimuth information of objects detected in a specific elevation domain. This is for each detected object an arc with as centre the position of the MTI radar. The TV camera and IR camera give azimuth and elevation information of objects detected within a specified range domain. This is for each detected object part of a straight line starting from the position of the camera.

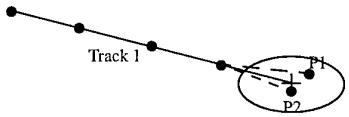
2.3.2 ASSOCIATION AND CORRELATION

Association and correlation involve sorting or correlating observations from multiple sensors into data sets, with each data set representing data related to a single distinct entity.

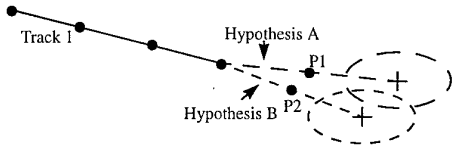
Now the data alignment processes have established that measurements can be compared with each other, it is possible to determine which observations may have been caused by the same object. In the Kalman filtering used in Positional fusion, for each combination of a sensor and a track, a correlation window is defined. Only measurements of the sensor in this correlation window, are correlated with the track. The current fusion step has to do association and correlation, thus, has to create this correlation window. Among other approaches, in TA10 Multiple Hypothesis Tracking is applied **[4]**.

The aim of Multiple Hypothesis Tracking is to postpone the final decision on difficult association situations until the receipt of more information. A number of candidate hypothesis will be generated and evaluated later on, as more data is received. The method is recursive, such that data sets only need to be processed on receival. Central to the MHT approach is the formation of a hypothesis tree. Each hypothesis in the tree contains a different series of plausible ways to partition information into tracks and false alarms. Each time a new set of observation data arrives, the tree is maintained by individually evaluating every existing hypothesis. A track occurring in one of more hypotheses is evaluated until it is eliminated (if the probability of it being a real track is too low), or turned into a definite track (if the probability of it being a real track is high enough). If a hypothesis contains a hypothesised track that is eliminated, this hypothesis cannot be true thus it is eliminated too. If a hypothesis contains a hypotheses track that is turned into a definite one, this part of the hypothesis is true, and only the remaining part has to be evaluated and becomes a 'new' hypothesis. A simple example will clarify the MHT principle:

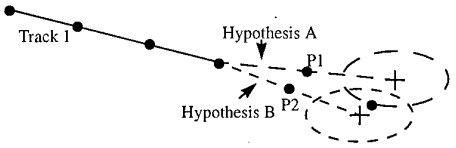
> The line connecting four plots below represents the movement of track 1. The '+' is the predicted position based on these four plots, and the ellipse shown is the correlation window belonging to this predicted position. Two received plots turn out to fall within this window:



The decision which one, if any, to assign to track 1 can be postponed. MHT supports 8 hypotheses, among which A and B, and evaluates them in parallel. Two predicted positions and corresponding gates are computed:

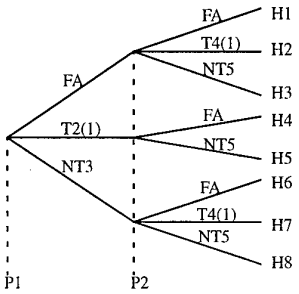


When the next data scan is received, one plot turns out to fall fairly centrally in the gate of the Hypothesis B track, and on the outskirts of the Hypothesis A track. Depending on the setup, the MHT algorithm may conclude that both hypothesis will be supported further, or accept Hypothesis B. Accepting Hypothesis B can for instance occur when Hypothesis A has become so unlikely that it is removed.



Below the corresponding MHT tree is depicted for the situation as above where two plots, P1 and P2, are received of which at most one can be assigned to the existing track. The tree reads from left to right and is structured in a series of layers that represent alternative assignments of the plots. The notation is:

FA : observation taken to be a False Alarm;
NT*i* : observation initiates the New Track nr *i*, hereafter referred to as T*i*;
T*j*(*i*) : observation associated with existing Track *i*, which results in the extended track T*j*.



The track renumbering T2(1) caused by the assignment of plot 1 to the existing track T1 is done to preserve the existence of T1 which may be needed later, so that the (hypothesised) track 2 contains all assignments of track 1 with the additional plot P1. Even in the simple example, identical tracks appear in more than one hypothesis. For instance, T4 corresponding to Hypothesis B shows up in both H2 and H7 (see combining paragraph 2.3.5).

2.3.3 POSITIONAL FUSION

In the TA10 positional fusion process, multiple observations of positional data of one object are combined in order to determine an estimate of the state vector using (extended) Kalman filtering [7]. Note that the data includes all available data of the state vector of the object, thus not only the position, but also for instance the (Doppler) velocity. This information in combination with the geographical information is used to select between different object behavioural models.

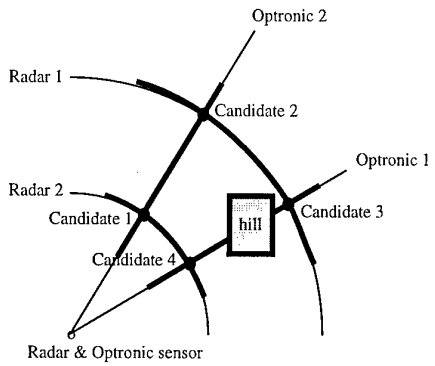


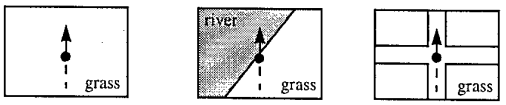
Figure 10 - Example of obstruction in terrain

Besides the information on the objects also information from the environment is taken into account (figure 10). Assume e.g. that two targets have been observed by both the radar and the optronic sensor and that the Optronic sensor (without range information) view in the directions of measurement Optronic 1 is hampered by an obstacle, like a hill. In principle there are four possible positions for commonly detected targets (candidate 1 - 4). Due to the hampered Optronic sensor view candidate 3 is a less likely object position than candidate 4 is in the Optronic 1 direction. If therefore candidate 4 is selected in the Optronic 1 direction, the only candidate for another commonly detected target is candidate 2.

If in a certain area the probability of the occurrence of clutter is high, the probability that a measurement in this area corresponds to a false alarm should be increased. In other words, this candidate is less likely than another candidate in another area, which regarding other aspects was equally likely. It has to be mentioned that whether or not a certain position is likely to be an object position, is different for the various kinds of objects. Thus the probability of a certain position being a real object position is, among others, dependent on the kind of object the detection is, or is expected to be. The probability that a tank is measured in a river is considerably lower than the probability that an amphibious vehicle is.

Furthermore, in the filtering process of some fusion nodes in TA10 the object behaviour related to the geographical and meteorological environment is taken into account. E.g. the terrain accessibility information:

1. If a military vehicle is travelling on a road, it is more likely that it will continue to travel on the road, then that it will leave the road and enter bushes, even though it is able to enter the bushes.



(A) (B) (C)

The position predicted by the Kalman filter is usually taken as the (one and only) expected next update position of the track. In most situations, like the one depicted in example (A) above, this is a logical and sane decision. However, it can happen that this movement involves a change to a less preferred terrain while this could be avoided. For example, in situation (B), the Kalman filter prediction is that the object moves into the river, while without further information it is more sensible that the object turns somewhat to the right and follows the border of the river. Therefore one might want to adopt a rule like:

'If the filter predicts the vehicle to move into terrain which is at least as good as the terrain the vehicle is currently moving on, assume that the vehicle is not going to manoeuvre to change its direction of movememt. If the terrain corresponding with the filter prediction is worse then the terrain the vehicle is currently moving on while this can be avoided, assume that the vehicle is possibly going to manoeuvre or is manoeuvring and account for a non-negligable (constant) vehicle acceleration in the filtering process.'

The first rule implies that when a vehicle is approaching a road junction, as in example (C), the probability of going straight on is the same as when there would not have been roads to the left and/or right. This kind of behaviour is desired when a vehicle is for example moving on grass or through bushes, but not when it is moving on a road. Therefore the next rule is added:

'If the vehicle is moving on a road and is approaching a road junction, assume that the vehicle is possibly going to manoeuvre or is manoeuvring and account for a non-negligable (constant) vehicle acceleration in the filtering process.'

2. The atmospherical and meteorological elements are divided in day/night conditions (either day or night), and weather conditions (clear weather, haze, light fog, dense fog, small rain, dense rain). For each of the possibilities an 'atmospheric and meteorological dependent' template will be made which is also input for the filtering process. Each such template defines the coverage area of all (relevant) sensors for the atmospheric and meteorological situation in the given terrain. It has to be decided which of the classes is most appropriate for the current atmospheric and meteorological situation, and the corresponding template is defined to be the actual one. The momentary valid template can change while the scenario is running, for example when it starts or stops raining.

The atmospheric and meteorological conditions do also have influence on the amount of clutter that is present. When the amount of clutter increases, the density of false targets measured by a radar increases. For each of the situations for which a sensor coverage area template is made, also a clutter map can be deduced (or measured). This has to be done for each sensor.

2.3.4 IDENTITY FUSION

Identity fusion is the process in which multiple observations of identity data are combined to determine a decision level identity declaration. In this stage of the fusion process, the simple rule will be adopted that MTI radar identity information (e.g. fixed wing, or helicopter) will only be used if it is the only identification information available for a certain object, otherwise TV camera identity information will be used.

Since the quality of TV camera identity information is considerably higher than that of the MTI radar, only using TV camera identity information for the fusion result is good enough at this level. Later on, in the assessment part of the process the various kinds of identity information that can be obtained from various sources (i.e. not only sensors) will be combined in a more advanced way. The use of templates is one of the methods that will be used.

2.3.5 HYPOTHESIS REDUCTION

The result of different processes is a number of possible world model hypotheses. To prevent a combinatorial explosion in the number of hypotheses, it is necessary to take measures to reduce the number of hypotheses, which is known as hypotheses reduction. Four techniques are used to manage the number of hypotheses:

1. Windowing
   This is done using the Kalman filtering information; only measurements that are within a gate around the expected update position are considered for correlation.
2. Clustering
   Observations in the area of interest are divided in clusters, such that the objects in distinct clusters do not interact with each other. Each cluster will subsequently be treated as an independent area.
3. Pruning
   The number of existing hypotheses is unlimited at the moment. In the step 4 it will be investigated whether hypotheses can be combined. For feasibility reasons we would like to have an upper bound on the number of hypotheses. This is achieved by simple pruning based upon a combination of breath control (maximum number of hypotheses) and adaptive pruning (when the sum of the probabilities from the best hypotheses exceeds a certain threshold, the remaining lower probability hypotheses are being pruned).
4. Combining
   It is investigated whether two or more of the hypotheses that survived the previous step have enough similarities so they can be combined.

2.3.6 SPATIAL GROUPING

An unit (e.g. company) which carries out an operational task (attack, moving in column) consists of vehicles which move in close proximity and have similar behaviour. The numerical processes produce a situation overview containing object tracks. Tracks which are close together and have a similar behaviour, have a (rather) high probability to belong to the same unit. Using the information in the situation overview, spatial groups are formed.

A spatial group is defined as a number of vehicles which are related because of their close proximity and similar behaviour. When the situation overview contains tracks indicating an attacking hostile unit, a spatial group is formed by tracks within a certain maximum distance from the centroid of the spatial group and which are moving in the same direction with similar velocity. The required radius for the circle to approximate the area which an unit covers depends on the selected smallest unit size which is required to be seen on the situation overview. In the case that the situation overview indicates units moving in column, a spatial group is formed by tracks within a rectangle positioned around the centroid of the spatial group and which are moving in the same direction with similar velocity. The size of the rectangle is determined by the smallest unit size which is required to be seen on the situation overview.

2.3.7 SITUATION ASSESSMENT

In general the goal of Situation Assessment is to produce combat intelligence concerning the enemy operating in the area of interest. Combat intelligence provides an understanding of the following aspects:

1. The composition of the enemy
   This describes the type and size of the enemy units and their co-operation in combat situations, and furthermore the identification of the observed enemy units.
2. The disposition of the enemy
   This indicates where and how the enemy has lined up its units and installations. Together with the composition this can reveal the (tactical) grouping of the enemy.
3. The combat effectiveness of the enemy
   This is the expression of the ability of a unit to pursue a combat activity in terms of the quality and the quantity of the available personnel and equipment.
4. The activities of the enemy
   This describes conclusions concerning the activities of the enemy based on enemy movements, establishment of assembly areas, the transportation of supplies etc., and the comparison of current and previous situation. Activities are compared with what is known of the doctrinal behaviour of the enemy.
5. The peculiarities and limitations of the enemy.
   This aspect denotes the strong and the weak elements of the enemy, that should get extra attention or should be exploited. It also describes features such as air support, special forces, deviations from doctrine etc.

In the TA10 project, Situation Assessment is aimed at uncovering composition, disposition, and activities of the enemy and at estimating enemy capabilities and intentions. Input to the Situation Assessment process will be reports produced by the Data Fusion process and information of possible enemy objectives. Enemy objectives involved in the TA10 project will be intentions to occupy a physical object or target (e.g. an airfield or ammunition storage) in the area of interest.

As mentioned before, the ultimate goal of Situation Assessment is to establish enemy intention, intention in this project being related to the occupation of a geographical target (objective). In order to do so, Situation Assessment will establish what units are active in the area of interest, estimate current activities and predict future activities in direct relation to the objective. The tasks that are performed in this process are the following:

1. Unit Classification
   This task deals with the classification of groups reported by

the Data Fusion process, i.e. determining the *composition* of the enemy. The organic composition of enemy units is compared with the composition of reported groups to classify them into units. The result of this task are units.

2. Unit Aggregation
   This task deals with the combination of units in to higher level units, sometimes referred to as higher level classification. This task uses known, organic models of the composition and behaviour of units, comparative to factors like classification, distance and spreading of component units. Since models are used to perform this task, the classification of a higher level unit is directly derived from the classification of the 'best fitting' model. The result of this task is (higher level) units.
3. Unit Association
   The Unit Association task is concerned with associating distant elements with known units. A distant element can be either a single vehicle or a subordinate unit. These associations can be used to support other situation assessment tasks, e.g. the association of an advanced reconnaissance vehicle with a unit gives an indication of the route to be taken by that unit. Knowledge of organic composition and disposition will be used to determine the certainty of associations. This task provides extended units.
4. State Estimation
   This task deals with the estimation of the current state of enemy units, i.e. determine the *disposition* and *activity* of enemy units. The location, velocity, and direction of a unit is derived from the known locations, velocities, and directions of its elements, an element being either a vehicle or a unit itself. This task also associates units with tracks. The activity of unit is derived by comparing the disposition of the elements of a unit with the organic disposition related to organic activities of the enemy. The result of this task is an estimated disposition and activity.
5. State Prediction
   This task deals with the estimation of the future state of units, relative to known possible targets or objectives, i.e. determine the future *disposition* and *activity* of enemy units by reasoning backward from potential targets. This way the probability of the *intention* to capture the different targets can be estimated from the path to be taken from the current position to a target, and factors like general direction of enemy movement, tracks and composition of units, and current activity.
6. Based upon the results in task 1-5, the Situation Assessment process determines a set of questions for additional information requirements, which are sent to the Sensor Management process. The questions include e.g. the extra need for information in certain areas, which could lead to e.g. sensor mode changes or setting up of drone plans by sensor management

2.4 SENSOR/PLATFORM MANAGEMENT

This process provides the TAD with the capabilities to perform Sensor/Platform Management on basis of the battlefield situation and desires of the operator. A sensor management plan is maintained and continuously updated which is the basis for platform and sensor management and control. The following types of sensor plus their platforms are managed: (1) Heliborne sensors, (2) Ground based sensors, and (3) Drone sensors.

The Sensor/Platform Management process provides the following capabilities:

1. Default sensor management plan
   The default sensor management plan is an initial plan for sensor management and reflects a long-term sensor management strategy, whereas the others below affect sensor management on short term. The default plan is off-line defined.
2. Prioritized requests for sensor information or direction
   The requests may be received from the Situation Assessment process and can be modified by the operator. They are prioritized indicating how demanding the wanted information is. A request is a query expressing a conjunction of the following elements:
   a. Battlefield area;
   b. Platform (which platform should be used for observation);
   c. Sensors (which types of sensor should be used for observation);
   d. Sensor mode (which sensor mode(s) should be used;
   e. Time interval (when should observation take place).
3. 'Hard' directives for sensor control from operator
   This data is treated as of highest priority and is a mean to override or by-pass the sensor management plan. The sensor management function adapts the sensor management plan accordingly.
4. Sensor performance data
   The sensor performance data is received from the (simulated) sensors/platforms. To maintain an optimal sensor management plan (i.e. allocation of sensors and selection of operational mode) and therefore to ensure maximum acquisition of relevant battlefield information, sensor performance is taken into account.
5. Meteorological and geographical data
   This data enables the sensor management system to constrain/adapt the plan to specific weather conditions (e.g. fog degrades IR) and terrain conditions (e.g. LOS constraints).
6. Interactive drone usage simulation/mission planning
   This capability provides the operator the support to plan for a drone mission during the TAD simulation. TAD will provide the operator with a provisional flight pattern / plan for the drone based on the results from the Data Fusion / Situation Assessment. The pattern will also be influenced by enemy threats. The enemy threats are limited to a small number and are considered to be static, i.e. no moving enemy threats (to meet calculation constraints). The operator is able to update or modify the drone flight plan. The plan contains: platform and sensors, list of trajectory points, speed on every leg, and start time

   After specifying the mission plan, this capability checks the mission plan on the following constraints:
   a. Availability of the drone.
   b. The observation points are within the battlefield area.
   c. Maximum speed
   d. Maximum mission time: 30 minutes

   If the drone mission plan is correct, it is input for and used by the drone mission control capability. This capability shall also provide means to store and retrieve drone mission plans under a unique name.
7. Drone mission control
   This capability controls the drone mission operations on basis of the drone mission plan provided by the drone mission planning capability. Drone control directives are issued (in near real-time) to the simulated drone platforms.

8. Sensor control
   This capability shall determine the time, place and mode of sensor/platform operations (excluding drones) based on the data as captured by the data processing capability. This will be done by means of construction and maintenance of a sensor management plan. Based on this plan, the sensor control capability shall distribute sensor/platform control directives to the (simulated) sensors and platforms. The sensor control directives support specification of changes at some time *t* to the following sensor/platform attributes: position, orientation and operational mode.
9. Sensor management plan presentation
   The system provides the operator with the capability to retrieve and display sensor management plan characteristics. It displays historical movements, aiming and ranges (i.e. sensor coverage) of sensor platforms, their current position (as known to the sensor management system) and operational mode, and their trajectories and aiming planned in the future. The system provides the operator with the capability to project (a selection of) this information on the battlefield area picture.
10. Drone mission plan presentation
    The system provides the operator with the capability to retrieve and display drone mission plan characteristics. It graphically displays the planned trajectory and the sensor coverage (as a function of time) and current position (as known by the sensor management system). The system provides the operator with the capability to project this graphical information on the battlefield area picture.

2.5 HARDWARE AND SOFTWARE ENVIRONMENT

Item	Used environments
Workstation Operating system Architecture	Sun SPARCstation Solaris 2.5 Splice '96
Standard Languages Functional Language KBS tool	C, C++ Haskell (Fuzzy)Clips
Graphic tool GIS	DtBuilder, Xforms Data base from MOD France
Word processor	FrameMaker on Sun
Documentation Knowledge engineering CASE Configuration management Project management	DOD-2167A (only significant parts) KADS TeamWork, method: Yourdon OMT Standard Unix tools MS-project
Quality	Procedures defined for TA10

3. PRESENT STATUS

The TA10 study is an on-going study project which will not be finished before summer 1998. In this paragraph the results and conclusions for the present status of the project is given. Today the following components are designed and implemented:

1. Architecture
2. World Model: databases and knowledge bases
3. Battlefield simulator & intervisibility calculations
4. Sensor simulator
5. Fusion nodes:
   - $GDFS_1$: mainly based on extended Kalman filtering and MHT techniques
   - $GDFS_3$: based on heuristics (rule based approach)
   - HDFS
6. Sensor/Platform Management

From the list above the Architeture, Sensor Simulation, and Sensor/Platform Management are already integrated. In the next step (before the end of 1997) the others will be added as well as the yet still unfinished modules. In the remaining period the TA10 study will concentrate on tests and evaluation of the system. Furthermore, the gap between the near real-time character of the TAD and a real-time system will be investigated.

ACKNOWLEDGEMENT

The contributions of the partners to the TA10 project and the involvement of the French and Dutch MOD are gratefully acknowledged.

REFERENCES

[1] S.S. Blackman, Multiple-Target Tracking with Radar Application, Artech House, 1986.
[2] Y. Bar-Shalom, On the Track-to-Track Correlation Problem, IEEE Transactions on Automatic Control, vol. AC-26, no. 2, april 1981.
[3] Y. Bar-Shalom, Multitarget-Multisensor Tracking: advanced applications, Artech House,1990
[4] D.B. Reid, An Algorithm for Tracking Multiple Targets, IEEE Transactions On Automatic Control, vol. ac-24, no. 6, december 1979
[5] E. Waltz & J.Llinas, Multisensor Data Fusion, Artech House, 1990
[6] R.G. Zuidgeest, Multi-sensor data fusion in commandand control and the merit of artificial intelligence, NLR Technical Publication, TP 94183 U, 1994
[7] R.G. Brown & P.Y.C. Hwang, Introduction to Random Signals and Applied Kalman Filtering, Wiley, 1992

LIST OF ABBREVIATIONS

BATS	BATtlefield Simulator
CSCI	Computer Software Configuration Item
DFS	Data Fusion System
DFSA	Data Fusion / Situation Assessment
DIR	Drone equipped with IR camera
DP	Drone Platform
DTV	Drone equipped with TV camera
ESM	Electronic Support Measures
FEBA	Forward Edge of the Battle Area
FLOT	Forward Lines of Own Troops
GBP_x	Ground Based $Platform_x$
GDF	Global Data Fusion
GDFS	Global Data Fusion System
$GDFS_x$	Ground based Data Fusion $System_x$
GGBDF	Global Ground Based Data Fusion
GGBHDF	Global Ground Based and Heliborne Data Fusion
GIR	Ground based IR camera
$GMTI_x$	Ground based medium range MTI $radar_x$
GSMGT	Global Sensors ManaGemenT system
GTV	Ground based TV camera
GTV+IR	Ground based IR + TV camera

HDFS	Heliborne Data Fusion System
HESM	Heliborne Electronic Support Measures
HMTI	Heliborne long range MT radar
HP	Heliborne Platform
IEPG	Independent European Program Group
IR	Infra Red
IS	Input Simulator
$LGSMGT_x$	Local Ground based Sensors Management $system_x$
LHSMGT	Local Heliborne Sensors Management system
LOS	Line-Of-Sight
MTI	Moving Target Indicator
NIS	Nato Identification System
OPFOR	OPposing FORces
SA	Situation Assessment
SM	Sensor Management
STANAG	STandard Nato AGreement
TAD	TA10 demonstrator study
UTM	Universal Transverse Mercator

AIRCRAFT SENSOR DATA FUSION: AN IMPROVED PROCESS AND THE IMPACT OF ESM ENHANCEMENTS

C.A. Noonan and M. Pywell
British Aerospace Military Aircraft and Aerostructures,
W392D, Warton Aerodrome, Warton, Lancashire, PR4 1AX, U.K.

SUMMARY

British Aerospace Military Aircraft and Aerostructures (BAe MA&A) has conducted various studies on the topic of Data Fusion. This paper highlights developments which, it is thought, offer a significant step towards optimum situation awareness. It examines sensor data quality, the fusion process and required improvements, and proposes a method for improving the quality of the resultant threat identification function. The paper also examines the key issues affecting the quality of track data fed into the fusion process by Electronic Support Measures systems (ESM), the prime contributor of threat identity data. Emitter recognition and location issues are discussed and potential routes are proposed to attain the necessary performance increases to support optimum situation awareness.

1 INTRODUCTION

Swift and unambiguous identification of hostile weapon system type and location in tactical scenarios is fundamental to aircraft mission success and survivability. To attain the highest probability of mission success the same identification and classification quality is required for all players in the scenario, whether hostile, friendly or neutral.

The fusing of data from multiple on-board multi-spectral sensors and other off-board data sources offers great potential for achieving high grade situation awareness. Example on-board sensors include radar, IFF, ESM, Forward-Looking and Search/Track Infra-Red (FLIR/IRST), LIDAR and Missile Warning Systems. This fusing corresponds to optimisation of own weapon targeting, threat evasion and countermeasures capabilities. These aspects can lead to improvements in aircraft lethality and survivability which, in turn, influence affordability, flexibility and availability. These five factors will be, arguably, the key product differentiators in the military aircraft market place of the future.

This paper defines 'Identification' in the military sense and discusses current identification processes and limitations. It describes the Identity Fusion Process and the contribution of various levels of ESM capability to that process. Potential improvements to identification are then described, comprising an improved fusion process and those relating to enhanced ESM performance. Identity Fusion Process simulation results are presented for different ESM capability levels, conclusions are drawn and a way ahead proposed. Although the issues addressed are applicable to all classes of military aircraft, the focus of work to date has been on fighter-sized aircraft of the present and future.

2 IDENTIFICATION

Reliable identification of own forces, threats, non-combatants and targets poses problems on any battlefield. This is particularly so in the air where participants in the battle may be highly dynamic and where own forces and hostile forces may be interspersed.

In order to avoid increasing the risk of being the perpetrator or the victim of fratricide, aircraft must improve their abilities to declare their identity to friendly forces and to recognise the declarations of others. To avoid collateral damage and casualties among non-combatants the ability to identify an intended target or a major threat positively before weapon release must be improved. The potential resources available to an identification system and the contributions they make to tactical situation awareness are summarised in Table 1.

When own forces are well separated from enemy forces and non-combatants, they may be identified by their actions conforming to some known mission plan. If a package encounters a friendly unit where the mission plan says it will be and at the right time, the task of the identification system will be relatively easy. By a similar argument, if forces are engaged in activities which do not correspond to the mission plan or any filed flight plan, they must be regarded as suspect. However, lack of adherence to a known plan is not a conclusive indicator of hostility.

Own forces may be identified by the ability to sign on to and exchange information with data communications networks and their ability to make the right responses to Co-operative Target Identification (CTI) systems. Again, the inability to make the appropriate communications and responses adds weight to the supposition of hostility.

For all participants and non-combatants, sensor data from Non Co-operative Identification (NCI) systems or from imaging systems may contribute to an eventual successful identification and, if they are making Radio/Radar Frequency (RF) emissions, these may be identifiable by the ESM. These latter approaches attempt to recognise the unit in question by its appearance or by the characteristics of the RF equipment it carries. However, the changing European political situation and the global armaments market increase the likelihood that similar units will fight on both sides of any conflict and so decrease the confidence which can be placed in these approaches.

When own forces are interspersed with enemy forces, the task of reliable identification becomes more difficult. Units which are identified by their participation in data communications networks can still be identified reliably if they are capable of declaring their position with high accuracy (*e.g.* by using the Global Positioning System), provided the observer knows its own position equally accurately. Interspersing of forces may pose difficulties for CTI and NCI systems if the sensor resolution of the system is poor and/or own and enemy forces lie in close proximity.

It is reasonable to assume that the ESM will be optimised to identify major threats from the RF emissions that they make. Traditionally, the ESM has produced coarse direction and estimates together with increasingly reliable identity statements and this combination of data qualities poses particular problems for the Sensor Fusion (SF) process. The identity statements are of increasingly high value but, in crowded scenarios, cannot currently be unambiguously associated with an individual track. These problems and potential solutions to them provide the subject matter for this paper.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

3 CURRENT PROCESSES AND LIMITATIONS

This section will look at the identification process from two points of view. Firstly, it considers how a conventional SF system brings together the identity information available to the aircraft. Secondly it considers current ESM and the ways in which they derive platform identity and location.

3.1 The Sensor Fusion Process

The SF process exists to gather together all the situation data arriving at the platform in question and to consolidate it into a single tactical picture. The platform which is of interest here is a fighter aircraft engaged in an air defence mission and the following assumes that application. SF may be regarded as a two stage process, see Figure 1. The first stage is tracking, during which all the sensor measurements referring to a particular platform are brought together over time. This gives the most complete and accurate estimate of the platform's position, motion and status that the measurements allow, along with some definition of the uncertainty in that estimate. A tracking process may use measurements from one or more sensors and the sensors may be distributed over multiple aircraft and depend on communications links.

If it were possible to gather all the sensor measurements at a single tracking process in a timely and reliable way, a single track database would be produced and the SF process would be complete at this stage. In practice, the constraints on the system are too great and this does not happen. So, at the end of the tracking stage several tracking processes will have produced their own database, each with its own view of the world, which then must be combined.

The second fusion stage, Track-to-Track Fusion, exists to combine these track databases into a single fused database. To do this it must perform the following three tasks:

- It must align the tracks to the same spatial axis set and time.

There is no guarantee that each tracker will use the same axis set. They may measure very different platform attributes and have very different points of view, especially when the sensors they serve are carried on separate aircraft. So the Track-to-Track Fusion process must perform any transformations necessary to align the platform data and corresponding uncertainty definitions to a common axis set. Also, each track will have received its last update at a different instant in time. So track data must be extrapolated to account for motion since it was last updated and each uncertainty definition must be modified to account for possible manoeuvres or changes in status since the update.

- It must deduce the number of targets giving rise to the data and the platform from which each track arose .

When data first arrives at the Track-to-Track Fusion process the number of targets is unknown. If two sensors each reported two targets, and if the reporting of false targets is sufficiently unlikely for us to ignore it, we must still consider the possibilities that there are two, three or four targets being detected, jointly, by the sensors. To do this we calculate the likelihood of each possibility and choose the most likely one. This process is commonly referred to association or correlation. Usually, this stage of the process will attempt to optimise some figure of merit for the association process which is linked to the likelihood of making the right choice. Firstly, an algorithm would mark each pair of tracks arising from different trackers as 'feasible' or 'not feasible' based on some statistical hypothesis test. Secondly, it would calculate the figure of merit for each 'feasible' pair. Finally, it would perform a search through the possible combinations of 'feasible' pairs and choose the one giving the best overall figure of merit. Some implementations shorten this process by performing these stages for new tracks only. For established tracks in such implementations, existing solutions would be retained until new data were received in direct contradiction to them.

- It must form joint tracks and joint identity statements for targets reported by more than one source.

The formation of joint tracks can range from a simple approach, which selects the best single track to represent the associated class, up to more complex approaches which calculate an optimal joint track using an algorithm based in estimation theory (*e.g.* minimum mean square error). Similarly, approaches to joint identity estimate formation can range from simple voting to algorithms based in statistical theory (*e.g.* Dempster's orthogonal sum).

A conventional approach to the implementation of Track-to-Track Fusion would adopt a process breakdown similar to the functional one described above and represented in Figure 1. However, this approach has limitations[19]:

- The ability to produce an unambiguous solution to the association question depends on the scenario. When the targets are dispersed and tracks from different targets are unambiguously separate, no problems arise. Similarly, when similar targets are grouped very tightly so that tracks from the same class may be interchanged without affecting the solution, no problems arise. However, when dissimilar targets lie close together incorrect associations may be made which affect the quality of the fused picture. This is discussed by Blackman [1]. In terms of the likelihoods described above, there would be conflicting feasible solutions with similar likelihood scores.
- The estimation process may be based on the assumption that data has been correctly associated and may attempt optimal combination of the data on that basis [1,2,3]. Processes of this kind, applied to incorrectly associated data, may produce meaningless results. Paradoxically, it is the 'optimal' algorithms which are the least robust in this respect because they rely most heavily on the assumption of correct track-to-track associations.
- When incorrect associations arise, they may change over time and with them the output of the estimation processes. This in turn leads to incorrect and changing information displayed in the cockpit with errors that move from platform to platform over time.

Data from the ESM is particularly prone to problems of this sort because, in present day systems, it tends to produce tracks with the coarsest positional accuracy. At the same time, the identity statements it produces may be the most specific and accurate available within the avionics system. Thus, there is a great incentive to use them. The result, when unfavourable scenarios are encountered, can be incorrect and changing identity statements displayed in the cockpit. There are several ways in which these limitations might be overcome. Two are considered in this paper:

- improvements in sensor accuracy and resolution, which would restrict the problem to more distant, tightly-grouped formations of targets.
- the use of algorithms which recognise the potential for ambiguity and take it into account, which would prevent the generation of incorrect and unstable solutions.

3.2 ESM, its Contribution and Limitations

Most threat weapon systems have a RF targeting and/or guidance component, usually in C-K band (0.5-40 GHz) and predominantly in E-J band (2-18 GHz), although there is now an increasing number of laser-only or laser-augmented systems. For the systems where RF is employed, the primary on-aircraft measurement and warning sensor is the ESM. In this paper the term ESM is taken to mean any level of Electronic Warfare (EW) antenna/receiver system capability, from simple Radar Warning Receiver (RWR), through conventional ESM, to the most complex (and costly) Electronic Intelligence (ELINT) equipment. A factor in common between RWR, ESM and ELINT is their function of detecting and .processing radar signals. The main differences are:

- RWRs are used primarily for threat warning.
- ESM is used for threat warning, the detection and identification of non-threat emitters, such as surveillance radars, and to determine emitter location. An ESM system designed primarily for emitter location is called an Emitter Location (EL) or Locator System, such as that fitted to the Electronic Combat and Reconnaissance (ECR) Tornado.
- ELINT is used for the detection, recording and analysis of radar/radio signals as well as locating emitters. It may be implemented by the addition of a recording and analysis capability to ESM, but often uses more sophisticated receiver systems. The emitter data resulting from ELINT analysis can be entered into the data bases which are needed for reprogrammable EW systems.

The data that ESM can contribute to the fusion process are:

- Azimuth Direction of Arrival (DOA) and, in some instances Elevation DOA,
- Emitter, mode and associated platform identification, each with a recognition confidence factor,
- Time of Arrival (TOA) and Range to the emitter, and
- Priority, if the emitter is a threat

The ESM, dependent upon its capability, may also provide measured emitter RF parameters (*e.g.* the ELINT fundamental parameters of frequency, Pulse Width and Repetition Interval, received power, and scan time/rate) and EL (from DOA and range). Derived parameters include RF type (fixed, hopper, deviations, *etc.*), PRI type (fixed, jitter, stagger, positions, elements, *etc.*) and Scan type (circular, conical, *etc.*). The accuracy of DOA, quality of and confidence in the above identification, and speed with which the ESM determines them, determine their importance to the fusion process. Probability of signal Intercept (POI), whilst crucial to the emitter recognition process, is not *per se* an input to the fusion process. POI is usually specified for an ESM; with typical values of approaching 100% for modern ESM. These factors are also arguably the key performance and cost differentiators between the sub-classes of ESM.

DOA Determination

ESM systems use DOA determination techniques which are based upon the measurement of some combination of amplitude, phase and time of arrival of an emitter's RF signal at a number of co-located and/or remotely located receive antennas on the airframe. Current techniques are listed in Table 2 and are adequately described in a number of texts, *e.g.* [4]-[6]. The resulting DOA accuracy is primarily a function of antenna type and locations, combined with receiver measurement accuracy of frequency, time and phase. Table 2 summarises typical current DOA accuracies for the main techniques. It should be noted that the absolute accuracy is often different for frequency sub-bands, dependent upon the technique and receiver combination(s) used. Newer DOA and EL techniques, *e.g.* differential Doppler, are discussed later.

Emitter/Platform Recognition

Significant commonality of RF parametrics of hostile and friendly radars limits the ability of current EW systems to provide the aircrew with unambiguous identification of the illuminating emitters. This is exacerbated by errors in RF parametric measurements and shortcomings of intelligence data programmed into the ESM. These issues, which are expanded upon in [7], have an adverse impact on situation awareness and the timely deployment of electronic countermeasures. There are four inter-related issues which need to be addressed if improved situation awareness is to be achieved and increased platform survivability ensured. These are the de-interleaving of incident RF pulse trains, the resolution of the fundamental problem of emitter ambiguity, the precise identification of platforms, and the potential benefits of the preceding items on countermeasures effectiveness. Of these four, emitter ambiguity is seen as the main issue and its resolution may lie in improved measurement and processing of intra-pulse modulation on signals.

Dense (>1 MPPS) RF environments can be achieved nowadays in certain frequency sub-bands, especially now high-PRF pulse-doppler radars are common. Typical current ESM can, within the very short time allowed to cater for countermeasure engagement/dispensing, only offer track file outputs as a list of potential solutions to which emitters it thinks it has seen, prioritised according to some pre-set rules *e.g.* hostile emitters are at the top of the list. Where association of these emitters to a platform can be made by the ESM, that too may be declared on the track file. The confidence of the ESM in its determination of the probability of a given emitter and platform declaration being correct is also flagged per emitter. For current systems this confidence factor is rarely unity for other than the simplest of RF scenarios.

Electronic 'fingerprints' of emitters and emitter types are discussed in a number of texts *e.g.* [8]-[10]. Although there are differing interpretations of the term 'fingerprinting', it means the use of a unique set of measurable parameters which enables either differentiation between emitter types (*e.g.* by features peculiar to the radar transmitter *type*), or between emitters of the same type, or indeed (and ideally) both. This topic is discussed later. A high level of emitter 'fingerprinting' can be achieved using current RWR/ESM (as opposed to ELINT systems), but only where the RF environment is relatively limited. [10] describes such an eight emitter, I-band scenario where, with one exception, the emitters could be unambiguously identified using today's ESM technology capabilities of: RF resolution (5 MHz), PW resolution (50 ns), TOA resolution (50 ns) and Scan resolution (4 ms).

Another problem of ESM capability limiting its usefulness to the determination of a real-time tactical picture is the update rate of track file information. The ideal ESM performance requirement is to see only the leading edge of the first RF signal (pulse or CW), and to instantly and unambiguously recognise the emitter, classify friend or foe and instigate crew notification, chaff dispensing and/or ECM engagement. In reality a few pulses and a few seconds are required to achieve the above with any degree of confidence. In order to provide sensible polar-type spokes on a CRT display the identified

emitter, its bearing and signal strength need to be displayed for a finite period of time. To cater for scanning or slow rotation rate emitters, where there may typically be up to 10 seconds between RF 'wipes' across the ESM antennas, the displays and appropriate countermeasures engaged may be kept on for the duration - until the emitter is definitely no longer illuminating the host aircraft. This can lead to the contents of the track file indicating emitter presence for some number of seconds after it has actually ceased to pose any form of threat, a limitation for fighter operations.

Emitter Location

The objectives and required accuracies of EL are summarised in Table 3. A number of conventional techniques exist for EL [5], [4], although only three are strictly applicable to military aircraft in the wingspan/length range 15 m (fighters) to 35 m (maritime patrol aircraft). These techniques are:

- Azimuth/Elevation: Ground emitters can be located instantaneously, assuming the use of aircraft altitude and reasonable DOA accuracy, either amplitude comparison (low cost) or interferometer (high cost) direction finding techniques. The down- and cross-range errors are a function of DOA errors, height accuracy and range to emitter. Table 4, adapted from [5] gives an indication of the down-range errors for two altitudes, using an early generation 3 azimuth, 3 elevation spiral antenna interferometer array. The ranging accuracy of this technique is best at high altitude (large depression angles) and degrades rapidly at low altitudes - an appreciable limitation for low level operations.
- Triangulation: The aircraft takes azimuth (or rather bearing) measurements at regular intervals of a few seconds, and uses triangulation and estimation algorithms to arrive at an accurate EL within a few seconds. This technique is more applicable to the ECR/ELINT role than to the fighter/bomber application, as it is not a forward-looking technique. It is, however, used on a number of current systems, *e.g.* that on the ECR Tornado, and requires very accurate DOA to achieve useful ranging against airborne emitters. Fig. 9 of [5] gives an indication of range uncertainty using this technique and two examples from that figure indicate the limitations of this technique for fighter applications as follows. To achieve a reasonable range uncertainty (say 3%) with an EL system with 1° DOA accuracy, at an own-aircraft speed of 400 Kts and at a nominal emitter to aircraft range of 15 n. miles, then one measurement would be required every 2 sec. for 60 sec. To reduce the required measurement time to that relevant to fighter operations (~5 sec.), a DOA accuracy of 0.1° would be required to achieve even 20% range uncertainty.
- Time Difference of Arrival (TDOA): Traditionally this has been a multi-platform technique, but can be implemented successfully on a large aircraft, now that TOA measurement systems with 1-5 ns resolution are available. It is a complex technique and technology, yielding high accuracy with high speed - but at high cost. It is unlikely to be feasible on a fighter-sized airframe due to the need for very wide spacing of antenna to form TDOA measurement baselines.

In each of the above techniques, various methods can be used to resolve location ambiguities and reduce overall emitter position error. One of the simplest methods is range estimation by the comparison of measured signal power against that stored in the ESM's emitter database for the Effective Radiated Power of that emitter. Paradowski [11] discusses EL techniques and algorithms, and includes a number of these methods.

4 POTENTIAL IMPROVEMENTS TO THE IDENTIFICATION PROCESS

This section will look at possible improvements to the identification process. Firstly, it considers how the SF process might better extract identity information from the ESM by allowing for ambiguity in its calculations. Secondly, it considers future ESM and the ways in which the quality of platform identity and location statements might improve, so improving the intrinsic capabilities of the system to resolve ambiguity.

4.1 The Improved Sensor Fusion Process

We have developed an approach to Identity Fusion which takes account of the ambiguity in the Track-to-Track association solution. The approach calculates the true probability of each identity for each track. When ambiguity is present it results in a solution which is offered with lower confidence than a conventional approach but which is stable and relatively free from error. In the absence of ambiguity, it produces a solution which is indistinguishable, numerically, from that produced by the conventional approach. We will refer to the new approach as Joint Probabilistic Identity Fusion. We propose two changes to the conventional SF process of section 3:

- The ESM 'tracks' will undergo a separate Track-to-Track Fusion process.

In particular, the alignment and association processes will be performed with respect to a partial fused picture comprising all non-ESM data. This may be expressed in terms of probabilities. Previously, we calculated a figure of merit which allowed us to maximise the likelihood of choosing the correct set of associations. The probability distribution, $P(\mathbf{X})$, that our fused picture is based on the true set of associations is written:

$$P(\mathbf{X}) = \prod_i P(\mathbf{x}_i|\mathbf{Z})$$

where $\mathbf{X}$ is the fused picture, $\mathbf{Z}$ represents the set of single source tracks and $\mathbf{x}_i$ is the i^{th} fused track in $\mathbf{X}$.

In the improved process we calculate:

$$P(\mathbf{X}') = \prod_i P(\mathbf{x}_i'|\mathbf{Z}_k)$$

and

$$P(\mathbf{X}) = \prod_i P(\mathbf{x}_i|\mathbf{X}', \mathbf{Z}_e)$$

where $\mathbf{X}'$ represents the partial fused picture, $\mathbf{x}_i'$ is the i^{th} partial fused track in $\mathbf{X}'$and $\mathbf{Z}_e$ and $\mathbf{Z}_k$ represent the set of ESM tracks and the set of other tracks respectively.

This change, in itself, does not improve matters greatly. The true value of $P(\mathbf{X})$ should not change. However, it imposes a processing structure into which our Joint Probabilistic approach fits neatly. Thus, it enables us to calculate a better approximation to $P(\mathbf{X})$.

Also, it is worth noting that situations leading to uncertainty about $P(\mathbf{X}')$ are relatively rare and we are able to obtain a good approximation to this distribution whilst avoiding a large amount of redundant processing. The key factor in the accuracy of the association process is the distribution of physical targets and this change delays ESM association until the most complete statement of this distribution is possible.

- Association and identity estimation will adopt a Joint Probabilistic approach [2].

This means that, during association, instead of associating ESM tracks with targets on a one-to-one basis, we will produce probabilities that each ESM track arose from all of the targets present in the scenario. We can calculate this with a high

degree of certainty using standard statistical theory. What we cannot do with any certainty is choose which one platform the ESM track arose from when these calculations yield similar probabilities for several tracks. The advantage of this approach is that we avoid this choice.

Previously we chose a set of associations, Θ, such that

$$\Theta = \arg\max_j \{P(\mathbf{X}|\Theta_j)\}$$

where Θ_j is the j^{th} feasible set of track-to-track associations. Then $P(\mathbf{X}|\Theta)$ was our (often poor) approximation to $P(\mathbf{X})$.

In the improved process, we continue to use this approach for non-ESM tracks to obtain the set of partial associations Θ' noting that $P(\mathbf{X}'|\Theta')$ is nearly always a good approximation for $P(\mathbf{X}')$.

For ESM tracks we calculate the probability:

$$\theta_{il} = p(\mathbf{x}_i \equiv \mathbf{z}_{el})$$

where $\mathbf{z}_{el}$ is the l^{th} ESM track.

During identity estimation, instead of combining associated identity statements under the assumption of correct association, we will combine all ESM identity statements into each platform, weighted by its probability of association.

Previously the probability of each identity was calculated for fused track i using Bayes rule:

$$p(y_m|\mathbf{Z},\Theta) = p(y_m)\times\prod_n p(y_m|\mathbf{z}_{in}) \div p(\mathbf{Z})$$

where y_m is the m^{th} feasible identity and $\mathbf{z}_{in}$ is an identity statement from sensor n associated with fused track i by Θ. Then the identity of track i was chosen

$$\mathbf{i}_i = \arg\max_m \{p(y_m|\mathbf{Z},\Theta)\}$$

and a declaration was made provided the associated probability exceeded some threshold. This worked well only when $P(\mathbf{X}|\Theta)$ proved to be a good approximation to $P(\mathbf{X})$.

In the improved process the probability of each identity is calculated:

$$p(y_m|\mathbf{Z},\Theta',\theta)$$

$$p(y_m)\times\prod_{n\neq e} p(y_m|\mathbf{z}_{in}) \times\sum_l [p(y_m|\mathbf{z}_{el},\theta_{il})] \div p(\mathbf{Z})$$

where $\mathbf{z}_{el}$ the l^{th} ESM track. Thus, the probability of each identity no longer depends on the unreliable ESM associations because we have not chosen ESM associations on the basis of questionable data. The most likely identity will be chosen :

$$\mathbf{i}_i = \arg\max_m \{p(y_m|\mathbf{Z},\Theta',\theta)\}$$

and a declaration made or withheld in the same way as before.

This change calculates the true probability of each identity class for each platform. When no ambiguity is present, this will produce the same clear identity statement as before. When ambiguity is present it will be reflected in a broader spread of probable identity classes for each affected platform and as a result it will be obvious that a confident statement of identity cannot be given. However, it may be possible to make more general statements of identity in these circumstances.

The improved SF process may be visualised, see Figure 2. Where a conventional process would produce incorrect and unstable identity statements, we assert that the Joint Probabilistic Identity Fusion process will allow correct and stable generalised statements to be made.

4.2 Improvements via Enhanced ESM Performance

The contribution of ESM to the data fusion process is currently undergoing a step improvement. Publications by ESM suppliers suggest that the ideal ESM performance of instantaneous unambiguous identification and exact spatial location of pulse/CW emitters may soon be feasible on any size of platform. Moreover, the technologies and techniques currently under development appear to eventually be applicable as relatively low cost retro-modification 'kits' to existing capability RWR/ESM. The three key development areas are:

- Advanced Combinational EL Techniques: Ongoing advances in processing speeds and processing technologies now enable combinations of classical DOA and EL techniques, in order to optimise EL performance, minimise errors and mitigate the shortfalls of individual techniques. A good example of this is the integrated ranging technique in the Litton Digital Receiver [12], developed under the U.S. Precision Location And IDentification (PLAID) programme [13]-[14], which combines long baseline (phase rate of change), TDOA, frequency Doppler and time Doppler. These combinations can also include novel techniques such as Differential Doppler [4], [11], which have only become realistic techniques for fighter aircraft with the advent of receivers capable of measuring frequencies to fractions of Hz [14].
- 'Fingerprinting': Current ESM have limited capability to quickly and unambiguously identify emitters, especially when the RF environment is dense. A number of agencies and research programmes have been addressing this fundamental limitation for many years. Only recently have technological developments occurred which now are believed to offer hope of achieving the above goal. During this time the proliferation of high performance and complex (multi-mode, high PRF) radars has continued, with a 1996 estimate of 4030 different radar types world-wide [15]. Close to ideal ESM capability is now believed to be feasible by using digital receivers (see below) together with combinations of analysis techniques such as classical parameter (frequency, PW, PRI *etc.*), clock de-interleaving, fine grain Intra-Pulse, Unintentional Modulation on Pulse and EL analysis [13]-[14]. Supplier claims [16] suggest that, by the year 2010, 100% of emitter ambiguity resolution may be resolvable as shown in Table 5, where SEI = Specific Emitter Identification.
- Digital Receivers: To enable the above, receiver/measurment system improvements have been required. A new generation of digital receivers have thus been developed, of which [12] and the Lockheed-Martin Passive Ranging Subsystem (PRSS) [17] are examples, approach readiness for in-service use. These have very high measurement accuracies/resolutions for frequency (<10 Hz), phase (few degrees), TOA (2 ns or better), PRI (sub-ns) and amplitude (<1 dB).

Other potential ESM enhancements include:

- higher receiver sensitivity (better than -60 dBm [10]), improved signal-to-noise and the use of more efficient spiral antennas, *e.g.* the spiral microstrip type [18]. These would equate to improved emitter detection range and could further assist in ambiguity resolution by comparison of measured signal power *vs.* emitter database effective radiated power.
- use of artificial intelligence (Knowledge-Based Systems), *cf.* Ch.8.7 of [6]. At the simplest level, ambiguity resolution could be aided by masking by logical aspects, *e.g.* a) ship radars don't fly; and b) if it's in front AND is above you AND has a high PRF AND is coming this way THEN it's very highly likely to be a threat!

5 SIMULATION RESULTS

Detailed simulations were performed and two scenarios were examined, see Figure 3. The first is an unambiguous Combat Air Patrol (CAP) scenario capable of resolution using a conventional SF algorithm and current ESM. The second is an ambiguous CAP scenario which is not. The scenarios are identical with the exception of the addition of a bomber formation to the latter, which is sufficient to introduce the ambiguities discussed previously. Each CAP aircraft is equipped with Radar, IRST and ESM, and they exchange track information using data links. The Radars are assumed to contribute identity information in the form of a size estimate and a Jet Engine Modulation measurement. Data link tracks contain the transmitting platform's estimate of identity based on locally sensed data. The ESM identity statements are assumed to be the most specific and accurate in the system.

Performance was measured using the conventional Identity Fusion process, as described in section 3, and the Joint Probabilistic Identity Fusion process and with 'current' and '2010' ESM. Results are presented as plots of % tracks a) correctly identified (*i.e.* Tornado, Hawk *etc.*), b) placed in the correct class (fighter, bomber *etc.*), c) not identified, and d) wrongly identified or classified. Basic identification as hostile, friend, neutral was not simulated but the same principles apply. The results indicate the performance of the system with respect to identification of hostile aircraft. In all cases, friends reported their identity and position (with GPS accuracy) *via* data links and were unambiguously identified.

Unambiguous scenario results:

- Figure 3.1 shows the performance of the conventional SF process with the 'current' ESM system. After ~30 seconds of the scenario every track was correctly identified or classified. After ~100 seconds it was possible to identify every platform.
- Figure 3.2 shows the performance of the improved SF process with the 'current' ESM. There is a slight improvement but the result is broadly similar to Figure 3.1.
- Figure 3.3 shows the performance of the conventional SF process with the '2010' ESM system. This combination of systems identifies the targets quickly and fully.

Ambiguous scenario results: The time windows, which are different for each scenario, were those during which sensor coverage of the hostile aircraft was or approached its maximum and thus most revealing of the SF process performance.

- Figure 3.4 shows the performance of the conventional SF process with the 'current' ESM system. This combination never fully identified the hostile targets. 20% or more of the identities were in error and the errors were unstable, moving from aircraft to aircraft.
- Figure 3.5 shows the performance of the improved SF process with the 'current' ESM. All aircraft are correctly identified after ~80 sec.
- Figure 3.6 shows the performance of the conventional SF process with the '2010' ESM system. Here, all targets were identified but the time taken (~120 seconds) to resolve all identities was longer than in Fig. 3.5. The simulation used did not Kalman filter the ESM tracks prior to Track-to-Track association. Had it done so, performance would have been better because the resulting sight-line angular velocity information would have made it easier to discriminate between the tracks and convergence to the fully identified state would have been quicker.

6 CONCLUSIONS AND WAY AHEAD

The simulations showed that the proposed Joint Probabilistic Identity Fusion process and the '2010' ESM system were both capable of resolving the kinds of ambiguity which would defeat conventional systems. Combined, they would be capable of dealing with greater and more complex ambiguities. They also showed that the Joint Probabilistic approach to Identity Fusion also promises improved performance in aircraft with an earlier generation ESM. Suggested future research paths are:

- examination and refinement of this approach using real sensor data, in a suitable rig and (subsequently) aircraft environment.
- further development and refinement of the apporach and algorithms *via* inclusion of a model of a year '2010' ESM.
- inclusion of 'Smart' sensor systems in the simulation. In our simulations, the opposing radars were dumb and noisy. Smart sensor systems which integrate data from multiple sensors and make Radar emissions only when it is absolutely necessary are feasible using today's technologies. Such systems will increase the uncertainty surrounding ESM data by reducing the data rate and will militate further for the use of an approach to Identity Fusion like the Joint Probabilistic one described here.

7 ACKNOWLEDGEMENTS

The authors wish to thank BAe MA&A for permission to publish and acknowledge input/comments from: Colin D. Hinds (EF2000 DASS Manager and EW Fellow Technologist), Gordon Slater (Manager, Nimrod 2000 EW), Jeff Green (EW Specialist, Nimrod 2000 EW) and Richard C. Freeman (formerly of the Data Fusion Group, Mission Systems R&D).

8 REFERENCES

1 **Blackman**, S.S. Multiple-Target Tracking with Radar Application. Artech House, 1986.

2 **Bar-Shalom**, Y., and **Fortmann**, T.E. Tracking and Data Association, Academic Press, 1988.

3 **Manyika**, J. and **Durrant-Whyte**, H.F. Data Fusion and Sensor Management a decentralised information-theoretic approach, Ellis Horwood Limited, 1994.

4 **Adamy**, D. EW101 Tutorials [on various aspects of emitter location techniques, ESM & receivers]. J. Electronic Defense 1994: Oct. pp.96-9; 1995 Jan. pp.60/1; Feb. pp.50-57; Mar. pp.67/8; Apr. pp.56/7; May pp.64/5; June pp.80/1; Aug. pp.64/5; Dec. pp.62/3; 1996 Jan. pp.58/9.

5 **Baron**, A.R., **Davis**, K.P. and **Hofmann**, C.P. Passive Direction Finding and Signal Location. Microwave Journal, September 1992, pp.59-74.

6 **Schleher**, D.C. Introduction to Electronic Warfare. Ch.2 *Electronic Support Measures.* (1990) ISBN 0-89006-142-4

7 **Wade**, L. The Significance of a Valid Threat Data Base to Effective Radar Warning. Tutorial, 1994 Association of Old Crows EW Tech. Symp. & Convention.

8 **Herskovitz**, D. The Other SIGINT/ELINT. J. Electronic Defense, April 1996. pp.36-40.

9 **Wiley**, R.G. Electronic Intelligence, The Analysis of Radar Signals. 2nd Ed. (1993) ISBN 0-89006-592-6. Ch. 7.8: *Incidental Intrapulse Shape - Uses and Causes.* pp.141-3.

10 **Self**, A.G. The Littoral - Some Implications for ESM Systems. Reference and Source Guide, Supplement to Jan. 1997 J. Electronic Defense, pp.60-66

11**Paradowski**, Col. L.R. Emitter Location Techniques and Algorithms: The Past, Current Status and Future Trends. Proc. 33rd Annual Association of Old Crows Int. EW Tech. Symp. and Convention, October 1996.

12**Baumann**, F.A. Litton Applied Technology's *Technical Update*, November 1995: Advanced Digital Receiver.

13**Linn**, P.A. (USAF Wright Labs.) Precision Location & Identification (PLAID). Presentation at 1995 Association of Old Crows Tech. Symp. and Convention, Washington.

14**Bass**, T. PLAID Software. Presentation at 1995 Association of Old Crows Tech. Symp. and Convention, Washington.

15**Pett**, M.C.D. (RACAL Radar Defence Systems) ESM in Today's High Density Signal Environment - the Surface View. Proc. Electronic Warfare Air, Land & Sea Conf. & Exhibition. 11-12 April 1996, London, U.K.

16**Baumann**, F.A. The Air Perspective. Proc. Electronic Warfare Air, Land & Sea Conf. & Exhibition. 11-12 April 1996, London, U.K.

17**Lockheed Martin Federal Systems** *Passive Ranging Subsystem*. Publicity leaflet 6-96.

18**Wang**, J.J.H. and **Tripp**, V.K. Spiral Microstrip Antenna Suits EW/ECM Systems. *Microwaves & RF*, December 1993, pp.147-150.

19**Watkins** F.M., **Noonan** C.A., **Roberts** K. and **Upton** N.K. Integration Issues in Modular Mission Management Aid Development. AGARD Conference Proceedings 555, Guidance and control for future air-defence systems, May 1994, paper 18.

Table 1: Identification System Resources

Resource:	Own Forces	Threats	Targets	Non-Combatant
Mission Plan	o	o	o	o
Communications with own forces.	*	*	*	*
IFF (CTI)	o	o	o	o
Long range imagery	o	o	o	o
RADAR(NCTI)	o	o	o	o
LIDAR(NCTI)	o	o	o	o
ESM	*	*	*	o
Self-defence systems		*		

* .. may provide conclusive identity statement.

o .. can contribute identity information.

Table 2: Direction of Arrival Technologies and Accuracies

(developed from [4] and [5])

Antenna Configuration	Typical r.m.s. Accuracies	Comment (regarding application to aircraft)
Amplitude Comparison	3-15°	Minimum capability of all sub-classes of ESM.
Phase interferometer	0.1-3°	Forward coverage, azimuth only (some elevation also).
Spinner	2-5°	Not appropriate to fast jets.
Multibeam	1.7-2°	Naval and ground vehicle applications.
Time Difference Of Arrival	<2°	Complex; needs large airframe for highest accuracy.

Table 3: Implications of Emitter Location Objectives

(from [4])

Objective	Enables	Required Accuracy
Electronic Order of Battle	Location of emitter types associated with specific weapons and units show enemy strength, deployment and mission.	Medium: 1km
Weapon sensor location (Self-Protection)	Focusing of jamming power or manoeuvre for threat avoidance	Low: general angle and range ~5km
Weapon sensor location (Protect Friends)	Threat avoidance by other friendly combatants	Medium: ~1km
Enemy asset location	Narrowed recce search or handoff to homing devices	Medium: ~5km
Precision target location	Direct attack by 'dumb bombs' or artillery	High: ~100m
Emitter differentiation	Sorting by location for separation of threats for identification processing	Low: general angle and range ~5km

Table 4: Typical Position/Range Determination Accuracies using conventional single-aircraft EL Techniques

(Adapted from [5])

Plan Range n. miles	Range Error			
	Altitude = 1 nm (6076 feet)		Altitude = 4 nm (24306 feet)	
	n. miles	%	n. miles	%
5	0.43	8.6	0.07	1.4
10	1.75	17.5	0.4	4
20	7	35	1.7	8.5
30	15.7	52.3	3.9	13
40	-	-	6.9	17.3
50	-	-	10.9	21.8

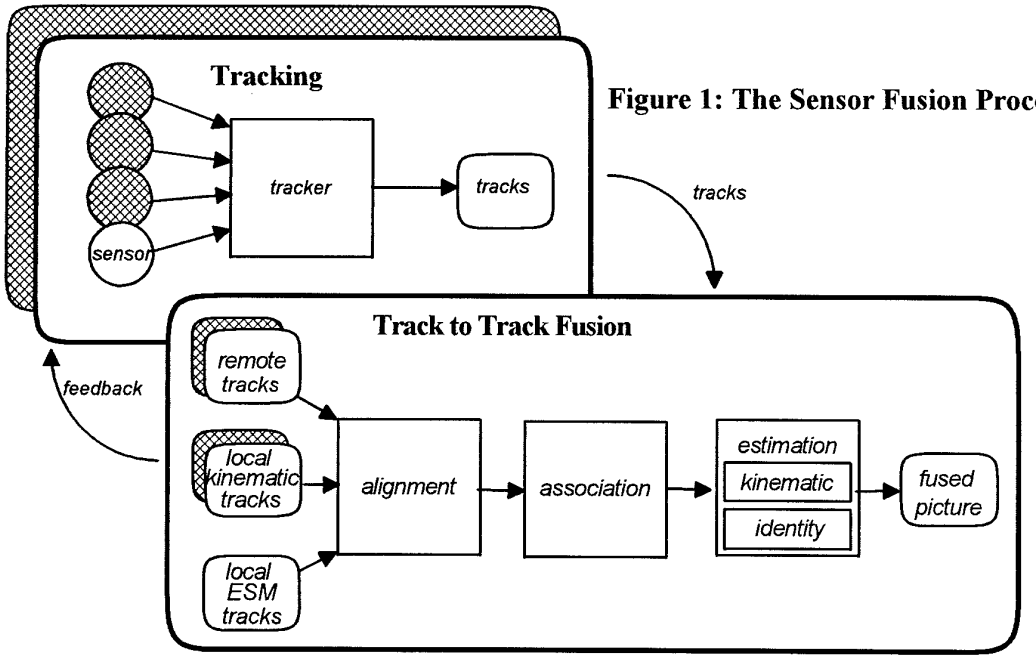


Figure 1: The Sensor Fusion Process

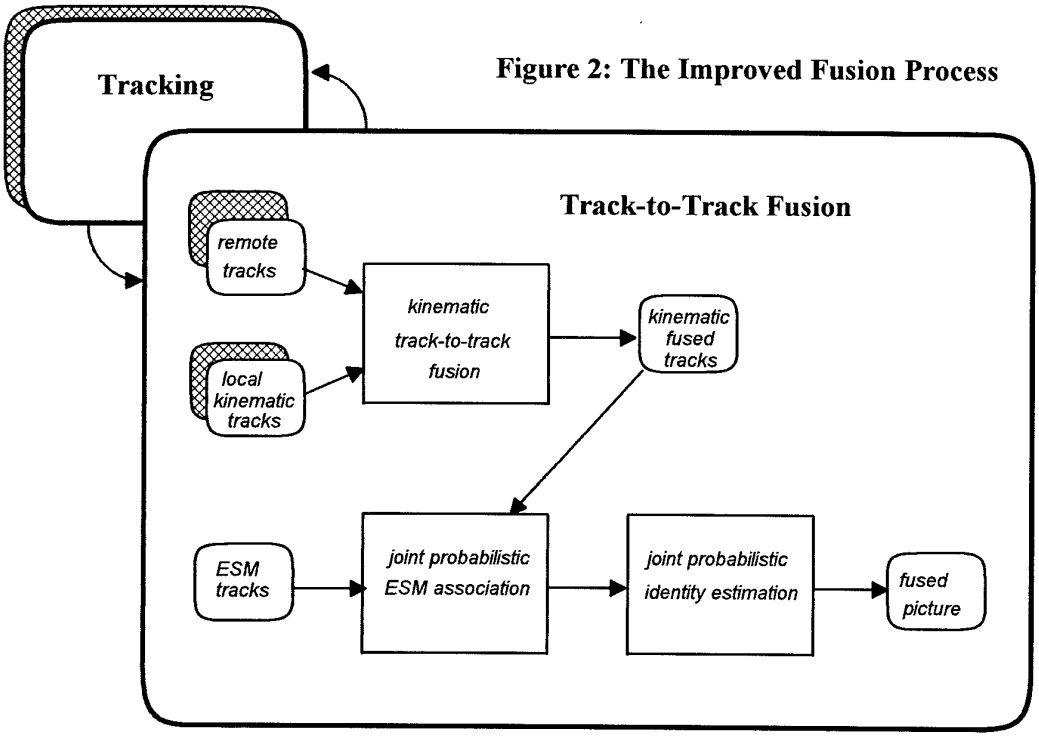


Figure 2: The Improved Fusion Process

Table 5: Year 2010 Scenario Ambiguity Resolution

(from [16])

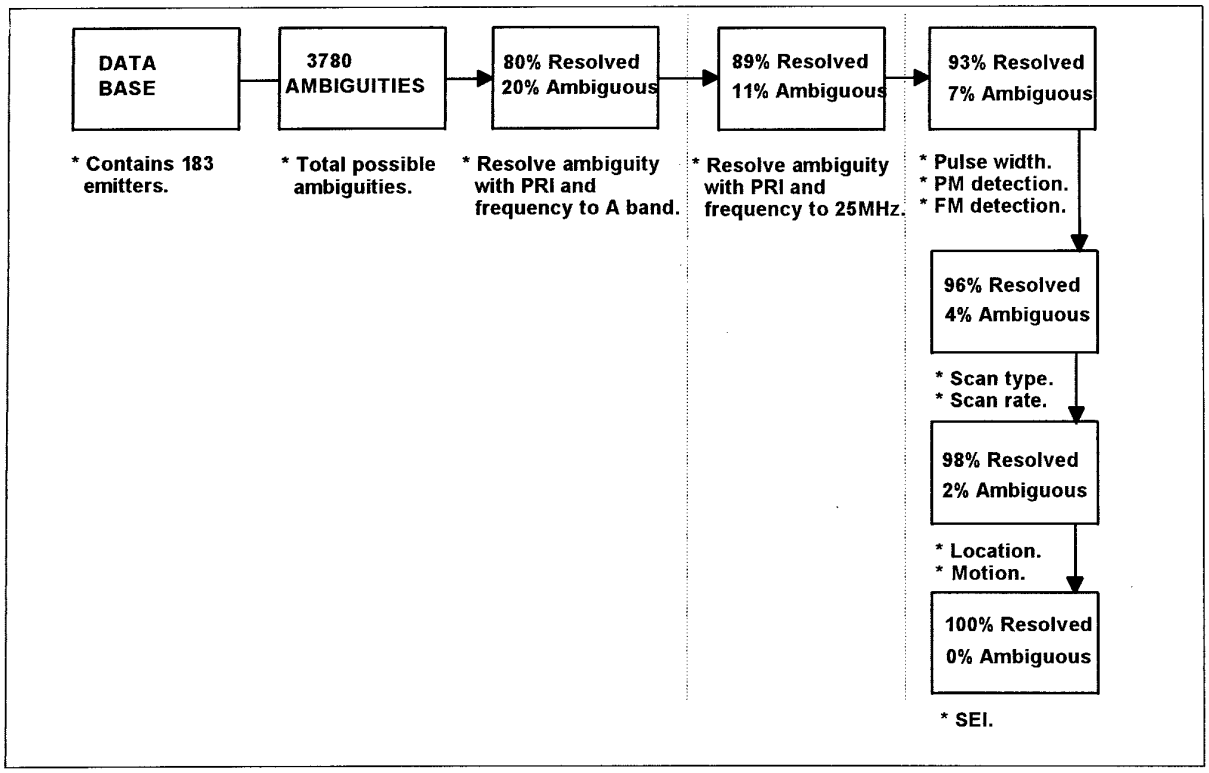


Figure 3: Test Scenarios

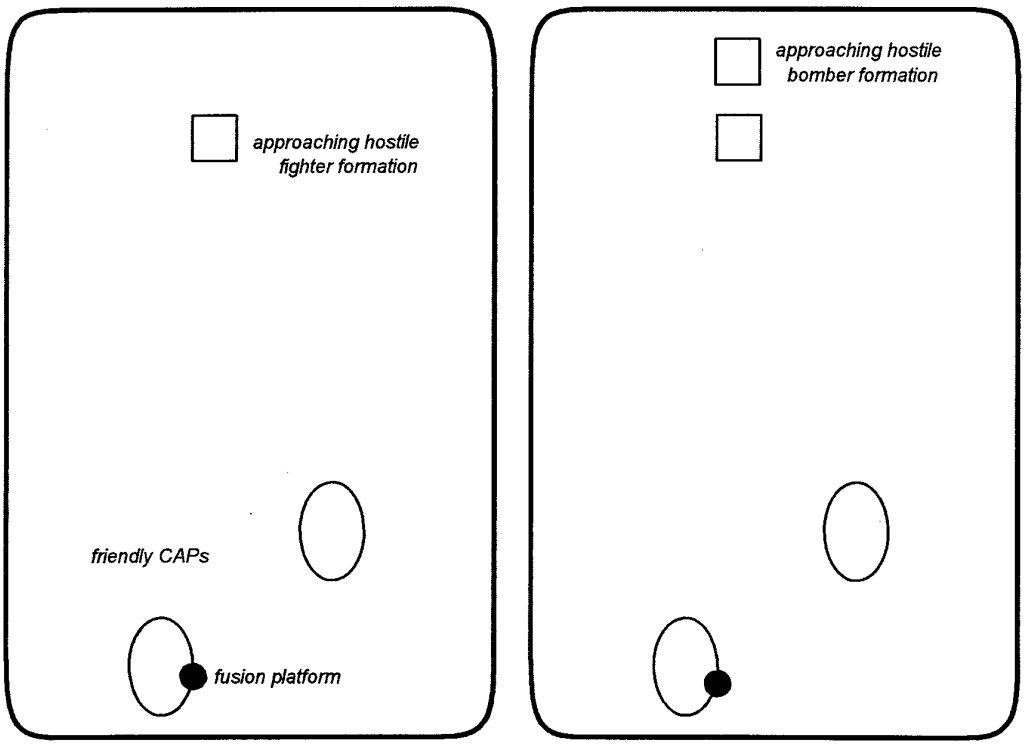


Figure 3.1

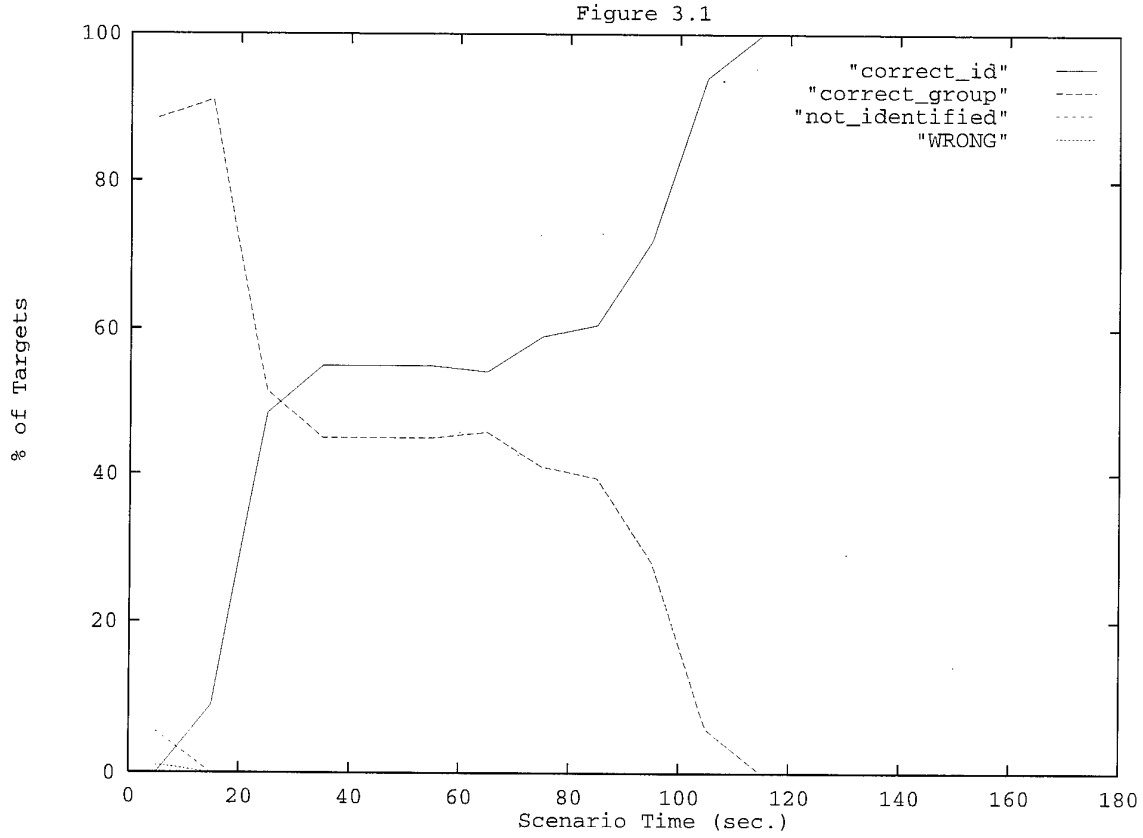


Figure 3.2

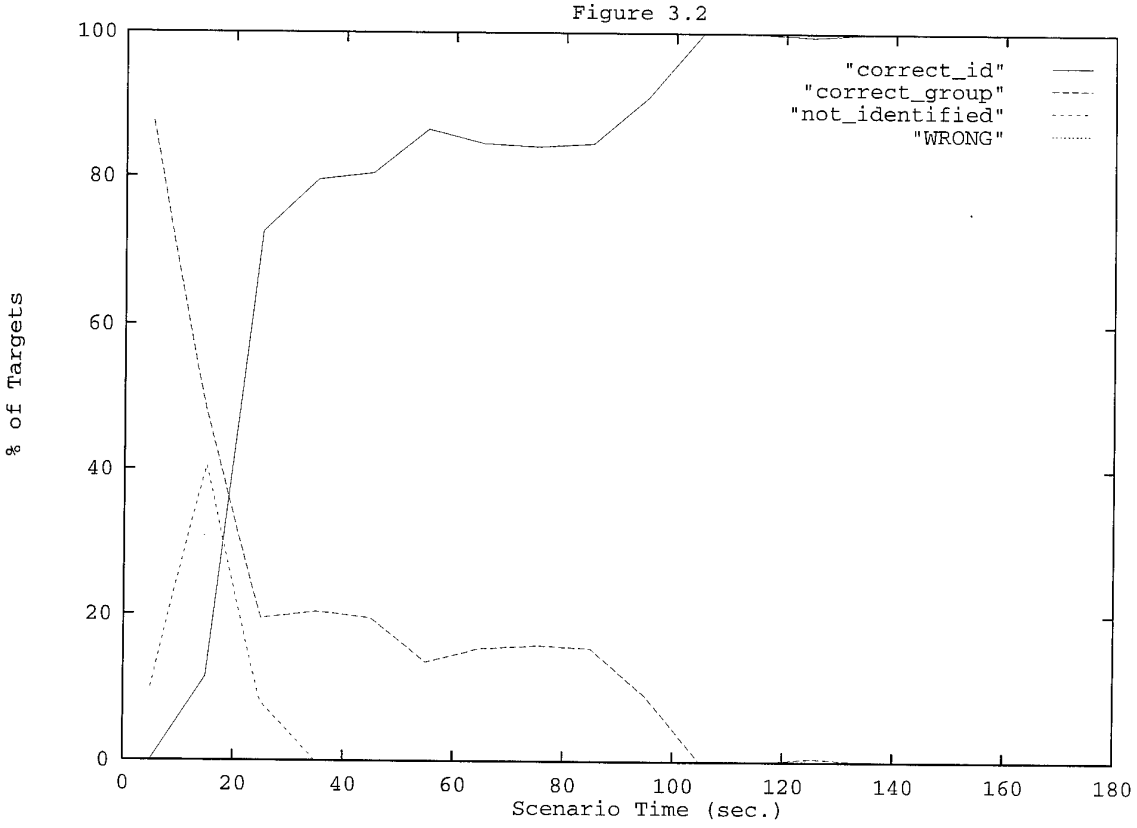


Figure 3.3

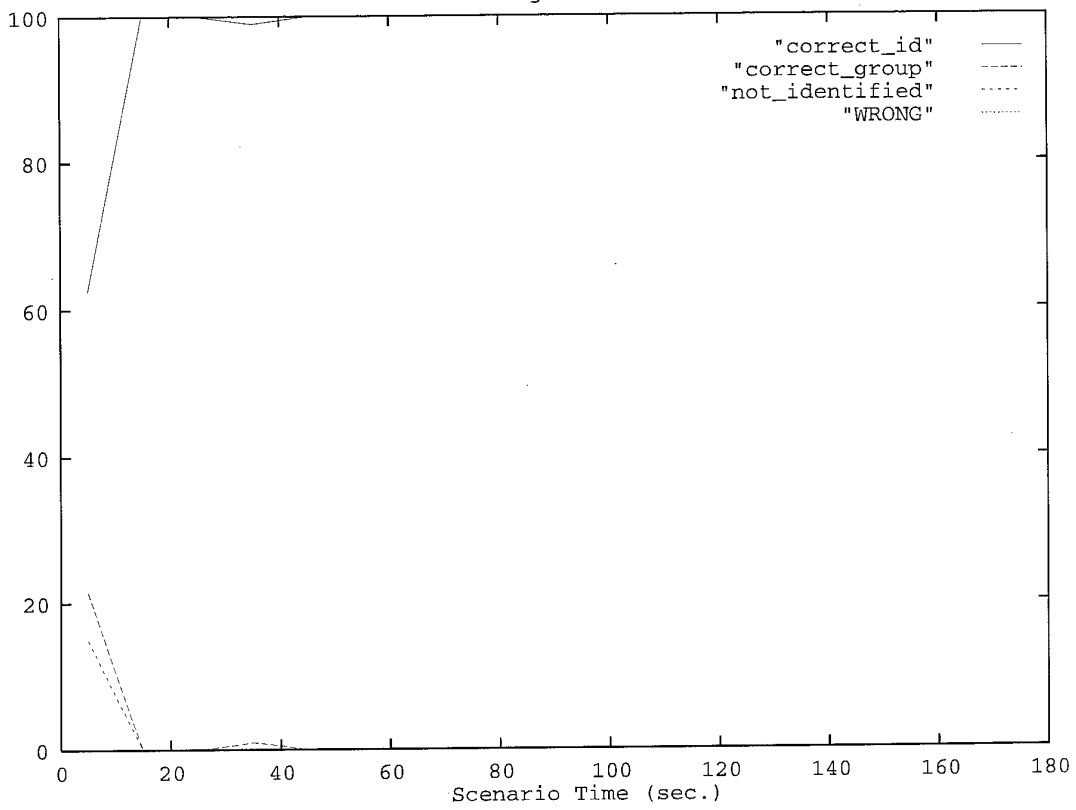


Figure 3.4

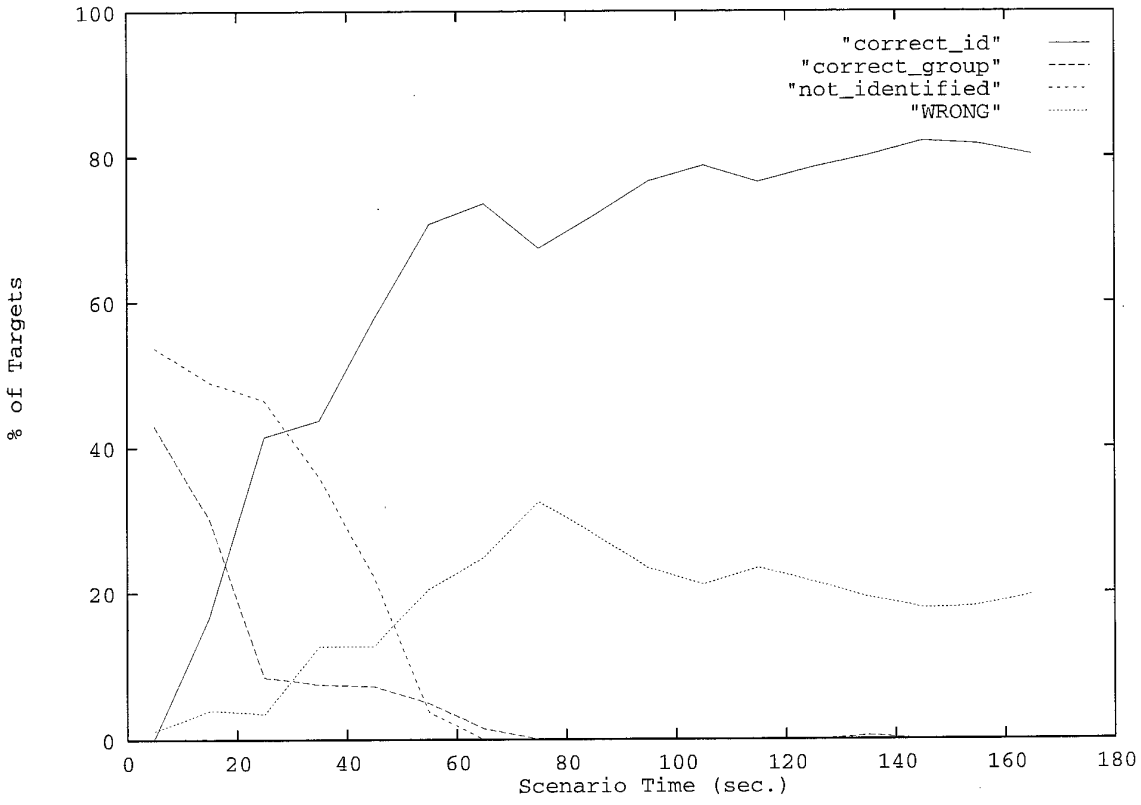


Figure 3.5

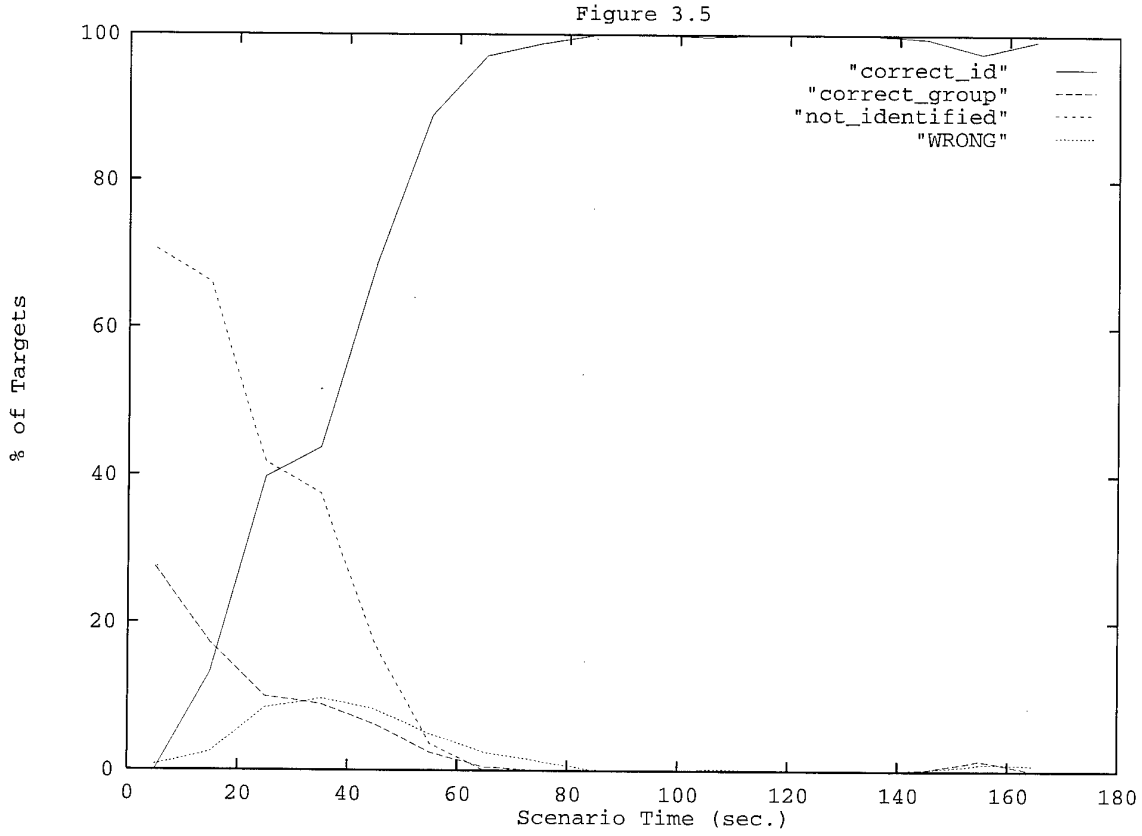
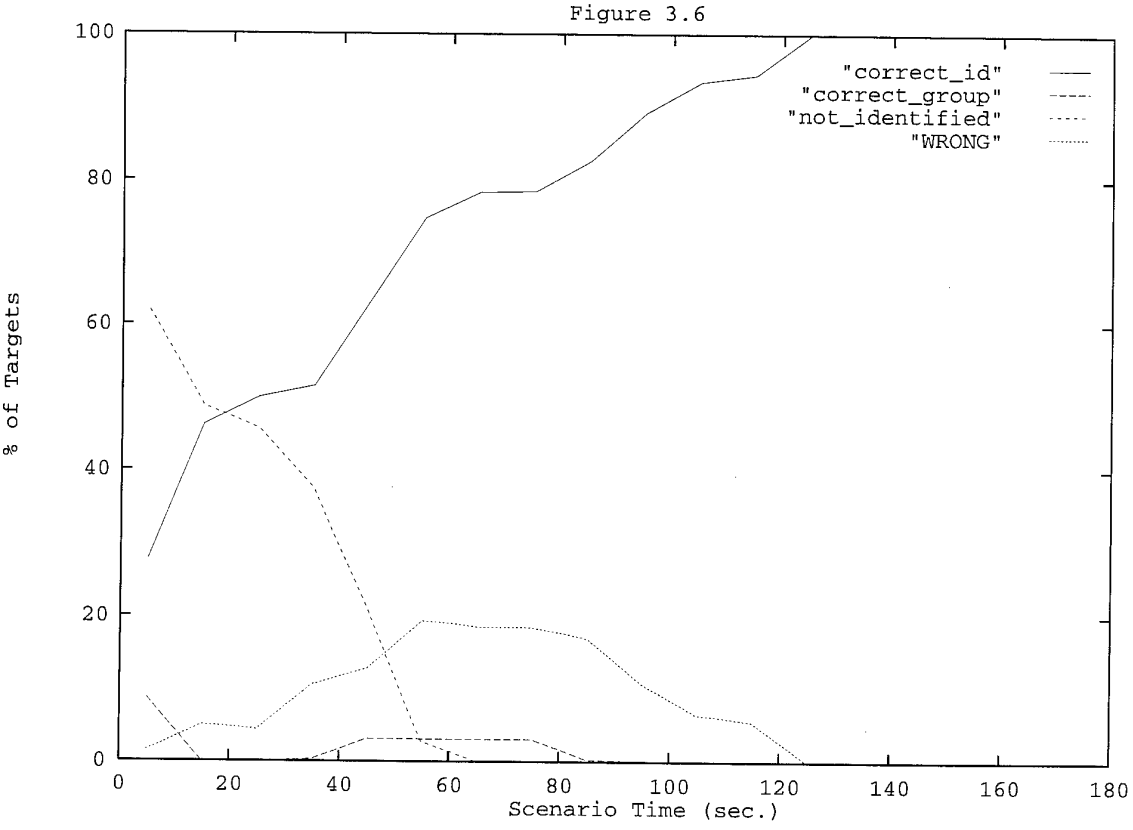


Figure 3.6



Multi-sensor remote sensing for military cartography

A.C. van den Broek and P. Hoogeboom
TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
The Netherlands

M. van Persie
National Aerospace Laboratory
P.O. Box 153
8300 AD Emmeloord
The Netherlands

1. SUMMARY

We have studied remote sensing data from sensors in different wavelength regions (optical, thermal infrared and microwave) and from different platforms (airborne and spaceborne) in order to extract geographical information. By comparing the extracted information with an existing geographical database of a test area in the Netherlands we find that to obtain military relevant cartographic information from remote sensing images resolutions of 5 meter or less are required. For appropriate classification of extended objects like agricultural fields multi-layer imagery is necessary.

2. INTRODUCTION

Due to the changing international situation after 1989 the tasks of the Royal Netherlands Army (RNLA) have changed considerably. Nowadays the RNLA considers and carries out operations outside the actual NATO area contrasting the situation during the Cold War. For NATO areas geographical information is sufficiently available and in case of allied operations this information is shared. For many other parts of the world geographical information is often sparse, and for non-NATO operations like UN operations each country is responsible for its own intelligence. In this context an independent and accessible source of geographical information is a necessity. Remote sensing data from satellites, but also from aeroplanes, UAV's, etc., offer such a source for geographical information.

The goal of the study presented here is to evaluate the potential of remote sensing for geographical information extraction. The extracted information can then be used to update outdated maps or to obtain basic information about an unknown site.

For this study we have collected remote sensing data from sensors in different wavelength regions (optical, thermal infrared and microwave) and from different platforms (spaceborne as well as airborne). The data have been collected for two test sites showing a variety of geographical features.

One test site is located in the Netherlands and shows no significant relief. It is called the 'Heerde' test site after the Dutch topographical map, which contains the test site. For this test site a digital geographical database is available, so that a detailed comparison between the remote sensing data and the geographical database is possible.
A complete collection of remote sensing images consisting of spaceborne TM, SPOT, KVR, ERS, JERS, and airborne CAESAR (optical, multi-spectral), PHARS (microwave , C-band) and TIR data is available.

The second test site, located in Germany, near Freiburg, comprises part of the Rhine Valley and the Black Forest showing moderate relief up to 1500 meter. For this site, called the Freiburg test site, TM, SPOT, ERS, JERS and KVR data, including a DEM (DTED) are available. In a following study it will be used to investigate the influence of relief on remote sensing images and the generation of a digital elevation model (DEM) from these images.

3. THE REMOTE SENSING DATA

3.1 Heerde test site

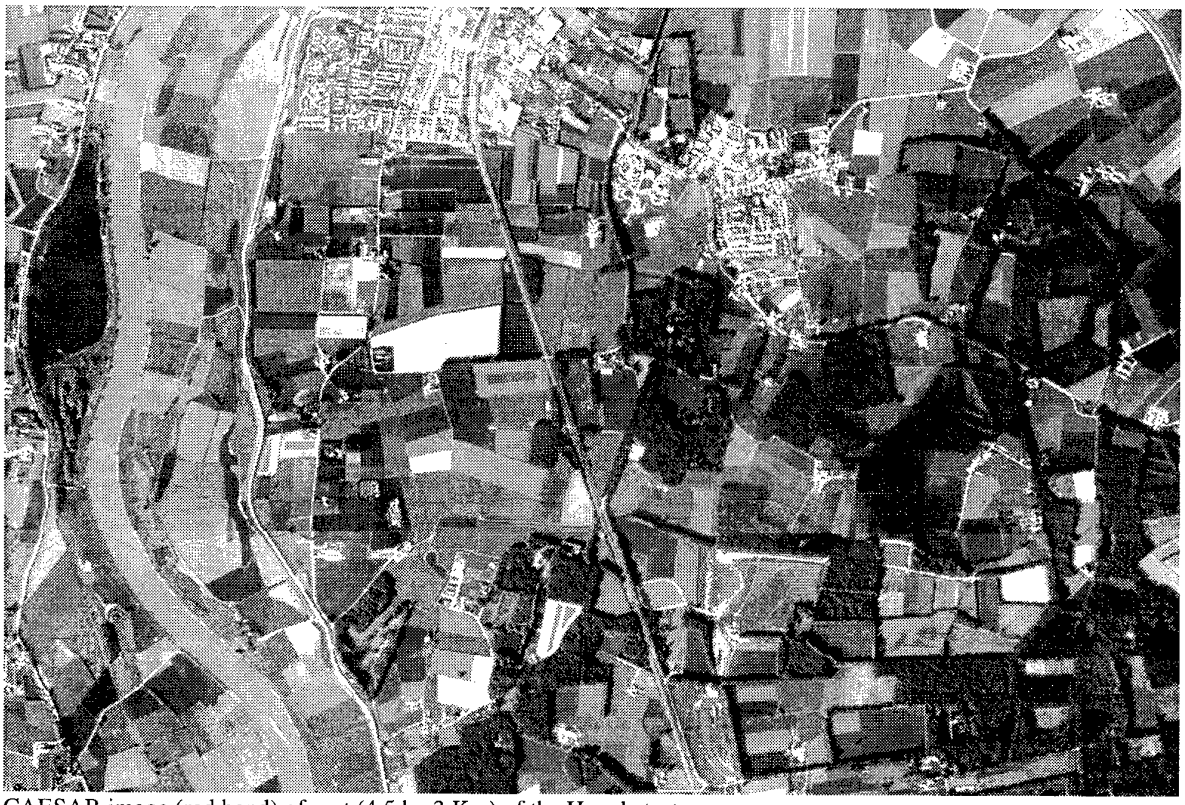
This test site is located on the TDN topographical map 27. The actual test site is oriented exactly east-west. and comprises an area of 10 by 20 km between the river IJssel and the Holterberg to the North of the city Deventer. The location is given by RD co-ordinates (x,y) 200000, 477500 m (South West corner) and (x,y) 220000, 487500 m (North East corner).
Low resolution satellite remote sensing data are available for the whole of map 27 East and West (RD 180000,475000 (South-west), RD 220000,500000 (North-east)), while high resolution airborne data are only available for the actual test site.
The test site includes the IJssel and its inundation (uiterwaarden) and shows landscapes like

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

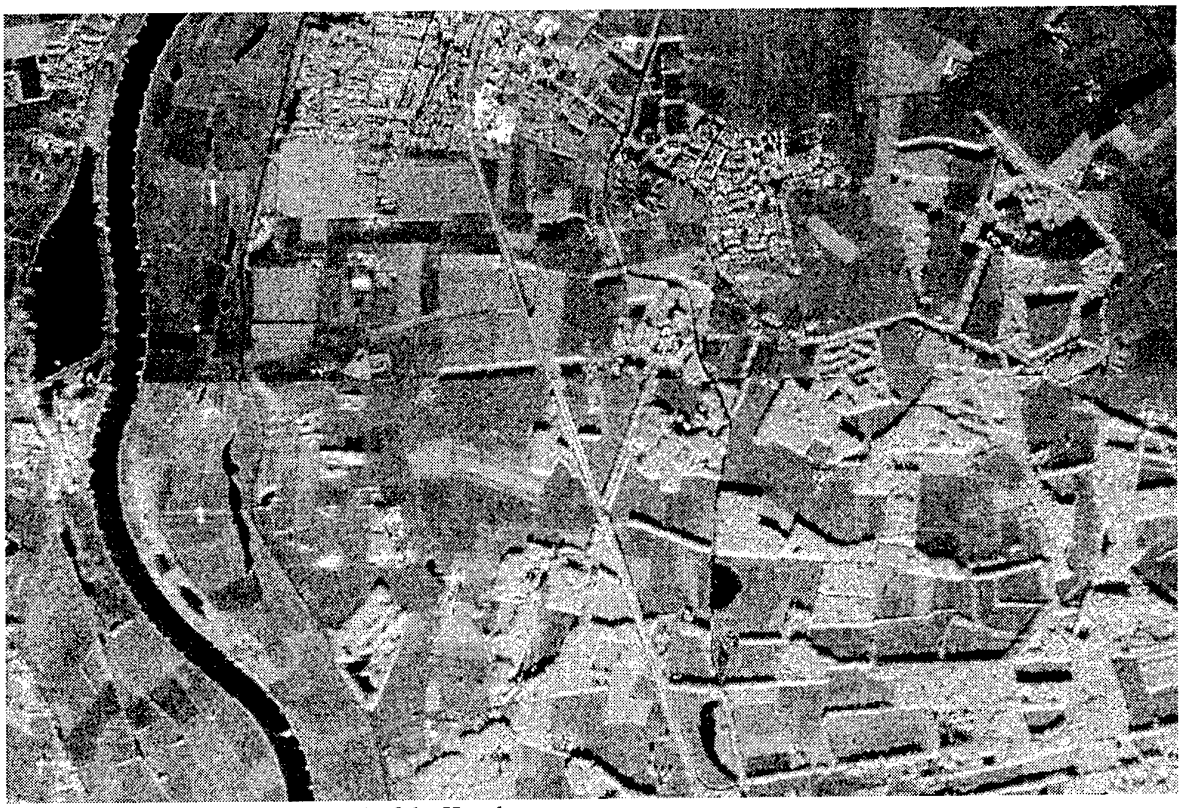
infrastructure, forest, agricultural areas etc. The terrain is flat up to a few meters. For this area the set of images shows resolutions ranging from 2 meters to 30 meters enabling a detailed comparison between the ground data and the extracted geographical information. Details of the data are given in Table 1.
All data were co-registered to the topographical maps 27 East and 27 West, with ground control points and using a stereographical projection and a Bessel ellipsoid as is usual for the Netherlands. The central point of the projection is located in the city of Amersfoort (RD co-ordinates (x, y) 155000, 463000 or 52.2° lat., 5.5° long.). The accuracy of the registration is determined by the geometrical accuracy of data and how accurately the ground control points can be determined with respect to the topographical map. This depends on the scale of the maps and on the resolution of the data. For the airborne data we used maps with a scale of 1:10,000, so that the data could be registered with an accuracy close to the dimension of the resolution cell. (i.e. 3 to 5 m). The satellite data could also be registered with accuracies within the dimensions of the resolution cell (i.e. 30 -10 m) using maps of 1:50,000 with exception of the KVR data which are accurately registered up to 10 m, while the resolution is 2 m. The is due to the lower geometrical accuracy of the KVR data.

Table 1. Remote sensing data for the Heerde test area.

sensor	*platform*	*resolution (m)*	*type*	*date (d-m-y)*
TM	spaceborne	30	optical/nir-multispectral	10-18-1993
SPOT XS	spaceborne	20	optical- multispectral	10-06-1992
SPOT PAN	spaceborne	10	optical- panchromatic	13-10-1992
KVR1000	spaceborne	2	optical- photographic	19-05-1992
ERS-1	spaceborne	25 (3 looks)	microwave- C-band/VV/23°	06-08-1992
JERS-1	spaceborne	25 (3 looks)	microwave- L-band/HH/35°	20-09-1993
CAESAR	airborne	3	optical- multispectral	22-09-1994
PHARS	airborne	6 (8 looks)	microwave -C-band/VV/45-65°	27-09-1994
TIR-camera	airborne	5	thermal infrared scanner	23-09-1994-14-10-1994



CAESAR image (red band) of part (4.5 by 3 Km) of the Heerde test area



PHARS image of part (4.5 by 3 Km) of the Heerde test area

4. THE TERAS GROUND DATABASE

In order to evaluate the remote sensing data for cartographic purposes we have used primarily the so-called TERAS (Terrein analyse systeem) database [1] for the Heerde test site, containing detailed geographical data in vector format.
These geographical data are comparable to the information on a topographical map with scale 1:50,000.
At first instance a collection of features and attributes have been selected on basis of their relevance for military terrain inventory. These features and attributes are extracted from a DIGEST (Digital Geographical Information Exchange Standard) code list [2]. About 60 features are selected covering 6 categories, which are listed here:

1. Culture; which includes typically man-made object like buildings, roads etc. (code A).
2. Hydrography; for example rivers, canals, ditches, lakes etc. (code B)
3. Relief portrayal; spot elevation, contour lines (code C)
4. Land forms; for example barren ground, depression (code D)
5. Vegetation; for example trees, crop land (code E)
6. Demarcation; e. g. an administrative boundary (code F)

Not all features in the list are suitable for comparison with remote sensing data. For example administrative boundaries (category Demarcation) are difficult to monitor with the current remote sensing data or are not present in the Heerde test area. Some features, like different types of towers, were combined, since distinction would involve identification which is not possible with the current remote sensing data set. The category relief portrayal is not considered here, but will be investigated using the second test site near Freiburg. Taking the features suitable for comparison three categories (Culture, Hydrography and Land forms/ Vegetation) are left in a compressed list, which is shown below in Table 2.

Table 2 Compressed list of features used in the comparison with remote sensing data.

Culture	
Code	name
AQ040	bridge
AQ065	culvert
AQ135	stopping area
AL240	tower

Hydrography	
code	name
BI030	lock/weir
BB140	jetty
BH020	canal
BI020	dam

Land forms/ Vegetation	
code	name
DA020	barren ground
EA010	cropland
EB020	heath
EA020	hedge row

AL015	building/complex
AT040	pylon
AL070	fence
AN010	railroad
AP030	road
AT030	high tension line
AD030	substation
AL020	build-up area

BH080	lake/ponds
BH090	inundation

EA040	orchard
EC030	trees

5. COMPARISON REMOTE SENSING DATA AND GROUND DATABASE.

We present here the comparison between vector data available in the TERAS data-set and remote sensing data collected for this study. The purpose of this comparison is to evaluate how accurately geographical information can be extracted from remote sensing sensors assuming that the TERAS data-set can be used as a complete reference data-set. In this way we try to determine the potential of remote sensing data for geographical information extraction.

In the comparison we have used the remote sensing images of the Heerde test area. For the satellite images we used TM, SPOT-PAN and KVR. The SPOT-XS image has been omitted in the comparison since it is expected that the results will average the results of PAN and TM. Also the ERS and JERS images are not considered here since the results are expected to be marginal compared to the other sensors. Only extracted point targets from these images have been compared with point features in the TERAS data-base. These results are shown and discussed in section 5.2. For the airborne images we used the CAESAR data, PHARS data and TIR data.

By comparing the various remote sensing images with the TERAS data-set (if needed also with the digitised topographical map) we have tried to determine for every feature how much of a feature was seen (i.e. *detected*) in the image. To do this different methods had to be used. For point data it is possible to count the number of points which are clearly related to objects seen in remote sensing images. The number of points which can be evaluated varies a lot for the different features. Usually the whole "Heerde" test area as imaged by PHARS and CAESAR has been taken for evaluation. In some cases the number of points is very large and a smaller area has been taken. In case of line or polygon objects also the length and area plays a role. A result is then obtained by inspection, for example by estimating the total detected length or area.

5.1 Results of the comparison

The results of the comparison are summarised in Table 3 for the six sensors mentioned above. The results are shown for the three classes: Culture, Hydrography and Land forms/ Vegetation. We discuss here the results for the three classes separately.

Culture

Some features like 'fences', 'towers' and high tension lines are generally too small to be detected by the sensors. For the detection of these features resolutions of less than 1 meter are required which are not available in the data-set. The low success rate of detection of towers is explained by the fact that towers are small when they are observed in vertical direction. For microwave sensors the viewing or look direction is not vertical (slant range geometry). The detection of towers is hindered in this case since the targets are confused with other targets in build-up areas where most of the towers are located.
The resolution of the TM sensors (30 m) causes this sensor to be less effective for detecting features in the category 'culture'. For SPOT-PAN the resolution of 10 meter makes this sensor 'intermediate' successful. Because KVR and CAESAR have relatively high resolutions (2-3 meter) most of the features are detected. Despite the somewhat lower resolution of CAESAR (3 m) compared to KVR (2 m) the first has a slightly higher success rate than the second because of the spectral information. Due to this information man-made objects and vegetation are easier to discriminate. The results for the TIR data are very sensitive to the emission and therefore to the temperature of the objects. The TIR data used here have been recorded during quite optimal conditions (late morning, clear sky). Man-made objects like bridges, roads and especially buildings are detected quite successfully, since they have been heated by solar radiation.
For the PHARS sensor the smaller objects like culverts and the smaller roads are difficult to detect because of the relatively low resolution (6 m) combined with the speckle.

Hydrography

The dimension of lakes and canals can vary substantially. For example the feature 'lakes' in the TERAS data-base contains many ponds sometimes overgrown by trees and the feature 'canals' contains many small ditches. This causes the detection rate to be quite low for the lower resolution sensors (TM and SPOT-PAN) despite the fact that these features are generally easy to detect.

The feature 'inundation' (IJssel Uiterwaarden) is difficult to observe directly since its main property is small scale relief over a large area. However the land use is different for the inundation compared to the surroundings so that observation of the land use makes detection indirectly possible. It appears that high resolution (< 5 meter) is advantageous for this purpose. The discrimination between water and vegetation is clearly more difficult for PAN recordings (SPOT, KVR) compared to multi-spectral recordings (CAESAR).
For the PHARS sensor the detection of canals is hindered by confusion with roads.
Despite the relatively high resolution of the TIR sensor (5 meter) the detection rate is substantially lower than that for the other high resolution systems (KVR, CAESAR) due to the fact that the contrast between water and the surroundings is not very high, especially when the water is surrounded by vegetation. However it should be noted that under certain circumstances, e.g. relatively warm water on a cold night water can be a dominant feature in the image.

Land forms/ Vegetation
In this case resolution is a less crucial parameter, since most objects are extended. For detection other information is important. Especially multi-spectral information plays a dominant role. This is due to the fact that vegetation reflects most of the light in the near-infrared compared to e.g. man-made objects. By observing in the near-infrared vegetation differences also become apparent.
It is therefore that the TM sensor has an even higher detection rate compared to the SPOT-PAN sensor despite the difference in resolution. The same is true for the CAESAR compared to the KVR sensor.
The detection rate of the PHARS sensors is lower on average compared to the multi-spectral optical sensors. Like in the optical, also in the microwave wavelength region vegetation differences are more easily observed when additional information is available, for example from multi-wavelength systems or polarimetric systems. The single channel PHARS sensor therefore shows a relatively low detection rate.
The TIR sensor is not ideal to discriminate between different types of vegetation, since vegetation attempts to suppress temperature differences by controlling the evaporation of plant moisture. The detection rate is consequently relatively low.

Table 3. Summary of the ability to detect features for the different sensors

	Name	TM	PAN	KVR	CAESAR	PHARS	TIR
Results for category culture.							
AQ040	bridge	10%	50%	80%	> 90%	50%	90%
AQ065	culvert	< 10%	20%	70%	80%	20%	20%
AQ135	stopping area	0%	30%	90%	90%	10%	50%
AL240	tower	< 10%	< 10%	10%	10%	< 10%	< 10%
AL015	building/complex	50%	90%	100%	> 90%	80%	100%
AT040	pylon	< 10%	10%	70%	90%	90%	< 10%
AL070	fence	0%	0%	< 10%	< 10%	0%	0%
AN010	railroad	70%	90%	100%	100%	100%	80%
AP030	road	20%	70%	> 90%	100%	50%	70%
AT030	high tension line	0%	0%	0%	0%	< 10%	0%
AD030	substation	< 10%	60%	90%	90%	90%	90%
AL020	build-up area	60%	80%	> 90%	100%	> 90%	100%
Culture	Average	21%	43%	67%	71%	50%	52%
Results for category hydrography							
BI030	lock/weir	0%	0%	30%	40%	40%	40%
BB140	jetty	< 10%	40%	90%	> 90%	> 90%	50%
BH020	canal	30%	20%	60%	80%	30%	50%
BI020	dam	20%	40%	50%	40%	40%	30%
BH080	lake/ponds	30%	30%	60%	80%	80%	50%
BH090	inundation	0%	30%	50%	50%	30%	50%
Hydrography	Average	15%	27%	57%	63%	52%	45%

Results for category land forms/vegetation

DA020	barren ground	90%	70%	70%	100%	20%	20%
EA010	cropland	60%	60%	70%	90%	80%	70%
EB020	heath	70%	80%	80%	80%	50%	40%
EA020	hedge row	50%	70%	80%	90%	80%	90%
EA040	orchard	70%	10%	50%	50%	20%	< 10%
EC030	trees	60%	80%	90%	100%	90%	100%
Land forms/ vegetation	Average	67%	62%	73%	85%	57%	55%

5.2 Other comparisons with remote sensing data

Roads

By displaying the remote sensing images on screen and digitising the detected roads (AP030) manually a database was created containing vector data for the different remote sensing sensors for the Heerde test area. The total length of the roads can be calculated and compared with the total length of the roads in the TERAS database. In this way a more quantitative comparison can be obtained compared with the inspection method discussed above. We show the total length (in meter) and some statistics in Table 4. Note that for PHARS two entries are available in Table 4. For PHARS I only the more obvious cases (dark lines) are extracted as roads. For PHARS II less restrictions are made and most of the dark lines are extracted as roads. Since roads (AP030) and canals (BH020) are easily confused in microwave images the result will be overestimated.

Point features

Microwave images are suitable for automatic point target extraction. Point targets can give increased backscatter which can easily be discriminated from the background with statistical means. The extracted points can be compared automatically with point features in the TERAS data-base, so that a quantitative comparison is possible. In this way ERS and JERS data, having improperly low resolutions for cartographic applications, may be used for the detection of point features. The comparison has been made for three ERS and one JERS images (see Table 5) and for the complete area covered by topographical maps 27 East and 27 West. A point was extracted from the images when the backscatter was 4.3 dB above the background giving more than 99% confidence that the point is not due to speckle.

About 3500 points representing mostly buildings and towers are selected from the TERAS database. A point feature in the TERAS data-set selection was said to be detected when it coincided within 50 m with a point extracted from the RS image. The result is that about 20% of the point features in the TERAS data-base are 'seen' by ERS and JERS (see Table). Another question is then how many of the 'detected' points are the same for all four cases, since the data differ not only in time but also in azimuth angle (different for descending and ascending passes) and in sensor (ERS/JERS). It appears that for cases 1 and 2 (difference in azimuth angle and time) 40% is the same, while for cases 2 and 3 70% is the same (difference in time only) and for cases 3 and 4 30% is the same (difference in sensor and time). The low value for cases 1 and 2 can be explained by the fact that the backscatter from point targets is quite dependent on the azimuth angle.

Table 4. Statistical data for AP030 (roads).

database	*no. of elem.*	*total length (m)*	*mean length (m)*	*s.d. length (m)*
TERAS	1646	433695	263	242
TM	1	107505	-	-
PAN	169	311750	1845	1621
KVR	2536	453285	178	201
CAESAR	807	619553	768	1256
PHARS I	64	147523	2305	2876
PHARS II	311	417945	1344	1638
TOP50	702	557823	774	1337

Table 5. Comparison results for the ERS/JERS and TERAS point data-set

Case	*Data*	*Date*	*Pass*	*Result*
1	ERS 1	23 06 92	ascending	22%
2	ERS 2	02 07 92	descending	20%
3	ERS 3	23 06 92	descending	24%
4	JERS	20 09 93	descending	20%

6. CONCLUSIONS

It is difficult to draw general conclusions for the whole set of features, since the characteristics of the features are quite diverse. In principle every feature needs to be considered on its own. By grouping features into classes some general conclusions can be drawn, since many of the 'culture' features are point targets, while the 'land forms/ vegetation' features are often polygons.
In general we can say that most of the features of the DIGEST list in Table 2 can be detected and recognised when the resolution and the wavelength is appropriately chosen. Man-made features like building or roads can be detected and often recognised quite adequately when the resolution is appropriate, i.e. less than 5 meter. The recognition of towers is difficult when the viewing direction is vertical. Direct detection of fences and power transmission lines is difficult due to their small dimensions. For resolutions less than 1 meter most features can be recognised. For extended features like crop land or forest multi-spectral or polarimetric information is more important than the resolution. Some features like inundation (BH090) and underground water (BH115) are difficult to observe directly with remote sensing sensors. Sometimes information can be obtained indirectly by observing other features which are related (i.e. context information; for example, green vegetation in case of available underground water).

Optical sensors make use of daylight and are weather dependent. For optical sensors high resolution images can be obtained, which facilitates the recognition of objects. Radiometric resolution and spectral information are also quite important to discriminate objects, especially between made-made objects and vegetation. Near-infrared information is very useful to discriminate different types of vegetation.

TIR sensors are very suitable to detect and to recognise man-made objects, like buildings, roads etc., when the circumstances are good (i.e. clear and sunny weather). However the circumstances are very crucial for TIR sensors. During a cloudy, misty or rainy day the contrast vanishes and objects (also man-made objects) are hardly visible in the image. For the discrimination of vegetation types TIR sensors are less suitable.

The main advantage of microwave sensors is that they are independent of most circumstances since it is an active sensor which can be used in principle during all weather conditions and at night. Microwave sensors can be appropriate for detecting objects due to specular reflections which give high returns in the image even when the resolution cell is significantly larger than the object (e.g. high tension pylons in ERS images). The result is a point feature which does not allow recognition. In general the ability of microwave sensors to recognise objects is rather low. Even when more resolution cells are covering the object recognition is difficult since then usually only a collection of points is seen. The contour of the object which is the main feature for recognition in the optical and TIR is in microwave image suppressed due the appearance of specular reflections and speckle. The low resolution of present-day spaceborne sensors is clearly disadvantageous for obtaining geographical information.

A quite important concept for detection and recognition is 'context information'. The context can provide information about a feature even when the feature cannot be observed directly. Since in many cases features show up as small objects in the image only a few pixels provide information limiting the recognition. However other independent information available for the interpreter, for example background information about location and the surroundings can provide the essential information so that recognition or even identification becomes possible. This so-called 'context information' will always play an important role in using remote sensing images.

7. ACKNOWLEDGEMENTS

This study has been partly sponsored by the Royal Netherlands Army under contract number A95KL756.

8. REFERENCES

[1] TERAS, Eindrapport Terrein Analyse Systeem, 1995, Werkgroep Militaire Geografie, Koninklijke Landmacht.
[2] DIGEST FACC (Digital Geographical Information Exchange Standard, Features and Attributes Coding Catalogue), 1992, Digital Geographical Information Working Group, Department of National Defence of Canada and the Defence Mapping Agency, USA.

PAPER No: 3

DISCUSSOR'S NAME: G. S. Brown

COMMENT/QUESTION:

What is your experience using radar data over foliage covered terrain?

AUTHOR/PRESENTER'S REPLY:

Foliage coverage from forests and tree-lines can be detected by radar, but the polarimetric radar performs much better in detection and classification of trees and tree types through their foliage than the single channel, fixed polarization. For the detection of structures under the foliage, penetration is required. Foliage penetration is only possible with low frequency radars, which are usually limited in their resolution.

Design of a Multi-sensor System for 3D Site Model Acquisition and Exploitation

I. TANNOUS*, S. GOBERT*, T. LAURENÇOT*, J.-M. DULAC**, O. GORETTA***

*THOMSON-CSF / SYSECA
Image and Geographic Information Unit
R&D Department
66, avenue Pierre Brossolette
92247 Malakoff - FRANCE
Tel: 33-01-41-48-03-84
Fax: 33-01-41-48-02-54
E-mail: issam.tannous@syseca.thomson.fr

**THOMSON TRAINING & SIMULATION
Tool & Database Product
5, rue Leonardo Da Vinci, BP 252
78703 Conflans Ste Honorine
FRANCE
Tel: 33-01-34-90-35-99
Fax: 33-01-34-90-36-02

***DGA/DSP/SPOTI/OER
18, rue du Dr Zamenhof
92131 Issy-les-Moulineaux
FRANCE
Tel: 33-01-41-46-23-03
Fax: 33-01-41-46-33-04

ABSTRACT

The increasing number of high resolution image sensors in a near future, as well as the need for realistic 3D site databases for applications like simulation, terrestrial mission preparation and training, has emphasized the need for multi-sensor image-based system dedicated to the production and exploitation of 3D site model. We present here the design of such a system, including the required functionnalities for this purpose. The system comprises a geometric reconstruction step, which aims at producing a geometrically reliable 3D model of the site, including the topography, from the aerial or satellite images available. Next step is the exploitation of other images, especially terrestrial photographs, for the mapping of real textures on the faces of the reconstucted objects, as well as the texturation of the terrain. The output of this stage is a realistic 3D site model that can be used by the exploitation module of the system, allowing the real time reconstruction of the scene from any point of view, interactively or according to a pre-defined scenario. This system has been prototyped by SYSECA and THOMSON TRAINING & SIMULATION (both subsidiaries of THOMSON-CSF) using their own software, and the results of an experimentation with real data have demonstrated the feasibility of our approach.

INTRODUCTION

The availability of an increasing number of high resolution imaging sensors (airborne and spaceborne) emphasizes the new capabilities of image-based systems for a lot of military (as well as civilian) applications relying on this means of observation. In addition to better coverage and availability of the image data due to their increased number, the high resolution of these sensors allows the characterization and reconstruction of human-sized objects from the images ; with the 1-meter class imaging satellites recently launched (or to be launched soon, see [2]), it will be possible to reconstruct an entire 3D site model from space, including the shape and description of buildings, roads and environment (trees, forests, relief).

The applications which benefit from these capabilities are typically the surveillance, monitoring and change detection on a site of interest, as well as terrestrial mission repetition and training. Such applications are based on the exploitation of 3D site models, and their efficiency rely on the quality of these models. The type of 3D site modelling needed for these applications must cope with two kinds of need :

— the site model must be geometrically reliable, which means that the location and dimensions of the objects must be close enough to the real ones ; such a constraint is emphasized by the need for high level of details required in applications like terrestrial mission preparation ;

— the site model must be photometrically reliable, which means that the model must include a realistic representation of the appearance of the faces (textures) of the objects of the site.

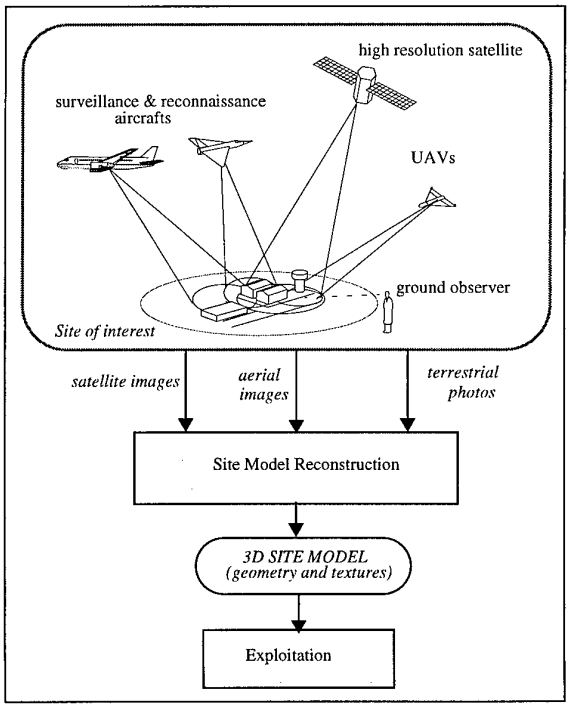


Figure 1 : Operational framework

With these constraints, it appears that a single source of images is not sufficient : it is clear that a single image cannot lead to a

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

reliable 3D reconstruction of a site ; futhermore, even if several images of a same sensor are exploited in stereoscopy for the geometric reconstruction, the angles of view cannot vary enough to give an exhaustive view of the site ; moreover, a single sensor offers generally one single spectral band, which is not sufficient for a realistic site reconstruction in terms of photometry. That's why it is necessary to acquire images from various sensors, from various angles of view, to be able to reconstruct a reliable 3D model site.

These considerations lay the stress on the fact that a 3D site model acquisition system must offer multi-sensor capabilities, insuring the optimal use of this amount of data for 3D information extraction. Consequently, an exploitation system must be able to cope with these complex data sets, including 3D geometry and textures.

Our paper deals with the design and experimentation of a system allowing the acquisition and updating of 3D models of sites of interest using sources like high resolution imaging satellites, aircrafts or UAVs cameras, and terrestrial photographies (see figure 1). The data provided by this system has been formatted to be exploited by existant software package dedicated to training and simulation. As a result, we demonstrated the feasibility of this approach and we validated our site model acquisition system design.

This paper is organized in 4 parts. First, we describe the workflow of 3D site model acquisition and exploitation, determining three phases : the geometric reconstruction, the rendering, and the exploitation. The second part describes in details the geometric reconstruction module. Then we focus on the rendering step. Next, we describe quickly the exploitation. Finally, we present the results obtained with our prototype.

1. DESCRIPTION OF THE SYSTEM'S WORKFLOW

Our system is divided in modules, and the typical workflow for 3D site modelling is represented on figure 2. The three main modules of the system are the *Geometric Reconstruction* module , the *Rendering* module, and the *Exploitation* module.

1.1. Geometric Reconstruction

The geometric reconstruction is based on photogrammetry techniques, where one can estimate the true geometry of the real objects from the images obtained with a sensor whose geometric configuration is known (the so-called geometric model of the images). In this step, the images that can be used are generally satellite or aerial images, because the photogrammetric exploitation of such sensors is possible, whereas terrestrial photographies are often unexploitable in this way.

The first step consists in a geometric pre-processing, which is compulsory for the 3D model extraction from the images. It comprises the so-called *aero/spatio-triangulation*, which aims at computing the geometric orientation parameters of each image, coherently with the other images. These parameters are necessary to be able to infer real geometric information from the images, while establishing the exact geometric correspondance between images objects and real-world objects. This module is based on an original multi-sensor triangulation algorithm developed by SYSECA, allowing the coherent geometric registration of heterogeneous images (e.g. SPOT and aerial images). In this step, a DTM (Digital Terrain Model) and an orthophoto are generated from the images, using an original automatic algorithm developed by SYSECA.

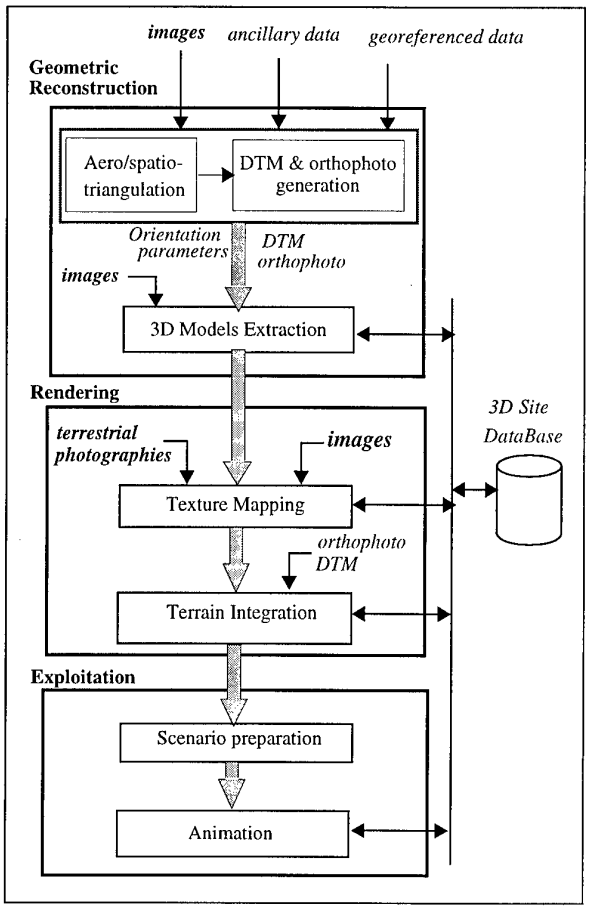


Figure 2 : System's workflow

Next step is the *3D models extraction*, where the system provides the tools for modelling the objects of the site from their views in the images. It includes two displays: one for the control of the tools and the visualization of the database that is being constructed, and another for the extraction of the objects from the images. The second display is a stereoscopic display, allowing the stereoscopic viewing of pair of images. In case of heterogenous images, this display is used in a split-screen mode. During this step, 3D models of the objects are extracted and stored in the 3D DataBase. This step uses an original software module developed by SYSECA (see [9]).

In order to allow the introduction of more complex objects in the 3D data base, or the addition of geometric details not extracted from the images (chemineas or balconies), the extraction module comprises a CAD-modeller environment, where the user can build complex objects from the objects of the database, add others objects on existing objects, or create new objects. This step can be used for entering known dimensions of objects (e.g. extracted from plans). The software used for this step is Ideal5 from THOMSON TRAINING & SIMULATION (TT&S).

1.2. Rendering

After having reconstructed a geometrically reliable 3D site model, the photometric reconstruction can be done. In this step, we map the textures extracted from the images on the faces of

the extracted geometric model. We make the distinction between the mapping of textures from terrestrial images, which must be done interactively, and the mapping of textures from the photogrammetrically exploited images (aerial or satellite).

We call the manual process of texturation from terrestrial images the *Texture Mapping*, and the step where textures are mapped from registered images is called the *Terrain Integration*. These steps are described with more details on section 3. It is based on a classical image software like PhotoShop for the manual extraction of textures from terrestrial images, and on the TT&S software ModTer for the Terrain Integration. In addition, the Terrain Integration is in charge of the formatting of the database and the DTM for the visualisation (triangulation).

1.3. Exploitation

Final step is the exploitation of this 3D model site. This exploitation consists in reconstructing a realistic view of the site in real time from any point of view. Such realistic simulation can be performed using several classes of observers or vehicles, whose behaviour can be accounted for in the deplacement through the site (pedestrian, tank, helicopter, aircraft, ...).

This exploitation comprises a preparation step, where a scenario can be defined. Then, the scenario is played in the Animation step. During the animation, the user is able to play a recorded trajectory, or interact in real time to move freely around the site.

The prototype is based on two software products from TT&S : TrackMaker for the scenario definition, and SpaceMagic for the Animation.

In the following, we describe more precisely the key parts of our system.

2. GEOMETRIC RECONSTRUCTION

This section decribes more precisely the geometric reconstruction step. After describing the geometric processing, we focus on the critical aspects of this step. We will insist on the problem of management of 3D data, and the solution we propose. Then, the extraction environment will be detailed, with the description of the Machine Man Interface configuration, the 3D extraction models and their acquisition modes, and finally the semi-automatic tools we implemented to support the extraction.

2.1. Geometric pre-processing

The *aero/spatio-triangulation* process is based on Ground Control Points introduced interactively in the system, on the images, and tie-points between couples of images found automatically, as well as the ancillary data provided with the images (ephemerids, interior ortientation of the sensor, ...). The ancillary data are used for automatically determining the interior orientation of the image, and to provide initial exterior orientation parameters for the triangulation. The exact parameters are then estimated using classical photogrammetric techniques (least-squares adjustments). The algorithm is fully multi-sensor, which means the coherent registration of many heterogeneous images (satellites and aerial) is possible.

The generation of the DTM is done automatically using the images. This DTM is computed on the whole area covered by the images (where at least two images are available). In order to prevent from the limitations of the correlation-based DTM generatio algorithm, it includes an erosion step, where the above-ground (artefacts due to superstructures) is automatically removed. This automatic step can be followed by an interactive erosion step for the residual above-ground.

The generation of the orthophoto is performed by using the orientation parameters of the images and the DTM to compute a geometrically rectified image (geocoded), where the effects of the sensor projection and the terrain induced deformations are removed. This orthophoto covers the whole imaged site by mosaicking the images.

2.2. Management of three dimensional data

The 3D model of a site is composed of many 3D primitives, which must be stored and managed specifically to be exploitable by each module of the system. Moreover, the coherence of the data seen by different modules must be insured. Such specifications show that the 3D data management problem in such a system has to be studied carefully.

Management of three dimensional data is traditionally a CAD world concern, whereas management of geographic data is a GIS domain concern. Though, current commercial GIS are not able to cope with 3D objects. The geographic data are essentially planimetric, with a z coordinate. The resulting dimension is called 2D5. A fully 3D data model is not implemented in GIS, because of the high complexity of 3D topology.

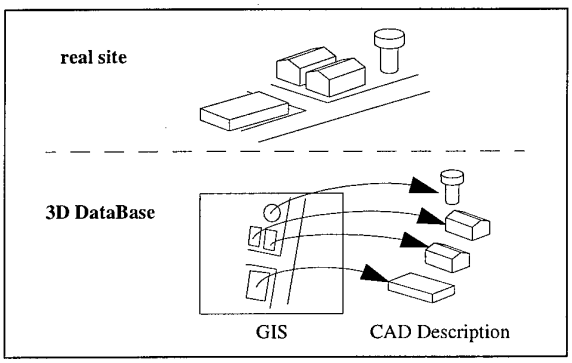


Figure 3 : Management of 3D data with GIS and CAD Modeller

Instead of trying to develop a heavy fully 3D GIS, we prefered to adopt a composite data model, taking the best of both technologies. The choice of this design is driven by the fact that the 3D-space indexing, necessary for some applications like geological survey (see reference [8]), is not essential for our application, which is dedicated to the above-ground extraction. So, the cartographic approach offered by a GIS is still relevant. To introduce the third dimension in the planimetric objects of a 2D5 GIS, the solution is to associate a 3D description to the footprint symbolizing a building in the GIS. Each building is then stored as a footprint associated with a CAD description (see figure 3). The association can be implemented with a Relational DBMS. Our approach, close to the one proposed by Förstner [1], offers a powerful solution at a moderate cost for development of 3D GIS devoted to 3D site modelling. Futhermore, the CAD approach for the 3D description of the objects allows a high level of detail, and can include a photorealistic texture for each facet of the objects.

2.3. Configuration of the extraction

The design of the Man-Machine-Interface dedicated to 3D model extraction from images is a difficult task, and our approach is the result of many concertations with operational staff.

Two displays are used:

— the *monitor screen* is a classic monitor, for controlling the interactive tools and the database that is being reconstructed,

— the *exploitation screen*, for viewing the images, is able to display the images in monoscopic and stereoscopic mode. The 3D objects that are being extracted are viewed as wire-frame models superimposed on the images.

For the nominal use of the system, we suppose a stereoscopic pair of images is available and is viewed on the exploitation screen.

In order to view and control the database, there are two specific windows on the monitor screen:

— a cartographic window, where the footprints of the objects are symbolized (similarly as in a GIS window); this window offers a synthetic view of the map of the site, showing the areas that have been reconstructed; it can be also useful for comparison of the reconstruction to an existing map.

— a perspective window, where the 3D objects of the site are viewed in perspective, from a point of view which can be interactively moved; in this window it is possible to view simultaneously the DTM as a wireframe; futhermore, it is also possible to drape the orthophoto on the DTM, which offers a good realism to the operator.

The three kinds of visualization of the database (exploitation screen, cartographic window and perspective window) are coherentely managed, which means every action on the objects from one window is taken into account by the others.

2.4. 3D Models

The type of objects to be extracted are divided in three classes: the *planimetric* objects (roads, rivers, areas, parkings, ...), the *volumetric* objects (buildings), and the *generic* objects (pylons, roadsigns, trees, antennas,...).

The planimetric objects use classic 2D5 primitives, like polylines, ribbons, or polygons. They are managed as usual in the GIS.

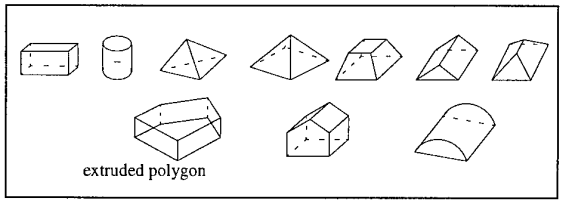


Figure 4 : Volumetric primitives

The volumetric objects entailed a specific process. Among the various approaches of 3D modelling in the CAD world, we decided to focus on Constructive Solid Geometry, where the only graphic primitives are volumetric. Objects are defined as a composition of elementary volumetric primitives. This ensures that the volumetric objects of the 3D database are geometrically coherent. The primitives used for volumetric objects acquisition can be defined by the users, and an example of such objects is given on figure 4.

The extraction is done from the images on the stereoscopic display. Each type of object implies an acquisition protocole. For example, a parallelepiped is extracted by four (3D) points to define the roof, and a pointing on the ground to define the altitude of the basis of the building. A peaked roof house can be extracted by only five points and a point on the ground. The selection of the type of object to be extracted is done through an ergonomic man-machine interface. Some examples of primitives acquisition are given on figure 5.

The generic objects (pylons, ...) are too complex to be fully extracted, but they can be represented by a small number of generic model, common to almost every site. Our approach is to extract from the images only the *3D bounding box* of the objects, whose CAD model is chosen in a generic model database. As a result, the extraction task is simple whereas the graphic model is highly complex and realistic.

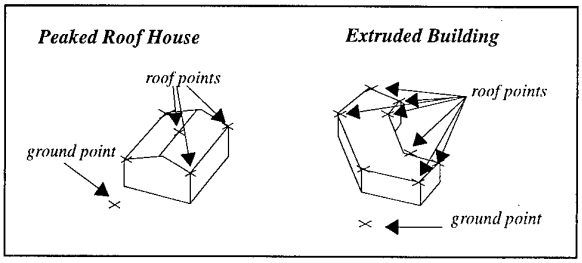


Figure 5 : Examples of acquisition modes

2.5. Semi-automatic extraction tools

In order to increase the productivity of the extraction, several tools are available in order to ease the acquisition of the objects. Because of the lack of maturity of most fully automatic extraction tools, we focus on *semi-automatic* extraction tools. The initial positioning of the primitives to be extracted is performed by the user, and the fine adjustment of the primitive to the images is in charge of the computer. The semi-automatic approach is probably the most promising way of increasing the productivity extraction by means of image processing algorithms (see reference [4] for a survey of the automated feature extraction methods on Digital Photogrammetric Systems).

One of this tools is the *semi-automatic road network caption*, as described in the reference [7], developed jointly by SYSECA and the French Geographic Institute (I.G.N.). This algorithm is interactive: the user sets the beginning point of the extraction, and the direction in which he wants to extract the road, and the algorithm finds the other points of the roads. When the algorithm fails to find the next point, it stops, and the user can correct the points extracted, or help the algorithm by giving manually some points (e.g.: where the road is interrupted by a bridge). The points extracted are given in a 3D referential.

For the *building extraction*, which is a tedious task in 3D site modelling, the algorithms we developed are based on a Model-Based Optimisation paradigm, and allow the extraction of the so-called extruded objects as well as the parametric objects (parallelepipeds, pyramids, prisms, cylinders, ...). These algorithms deal with the optimisation of the roofs of the objects on the images. For the extruded objects, a snake-like approach is used, similar to the one described by P. Fua in [3]. For the parametric objects, an original least-squares based parametric optimisation scheme is implemented. An example of use of the semi-automatic extruded building roof extraction is given on figure 6. One can see the automatic 3D refinement of the initial model jointly on both images of the stereoscopic pair.

Another very helpful tool is the "semi-automatic ground positioning". The user indicates a rough 3D initial and the fine 3D position is found automatically using a local correlation algorithm. This tool has proven to be reliable enough to take in charge for the user the measurment of the ground altitude need-

ed for the building acquisition, which increases in a significant way his productivity.

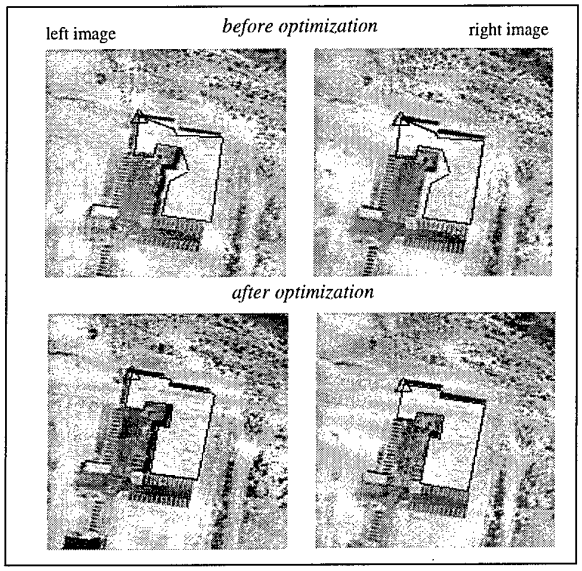


Figure 6 : a building's roof optimization

In addition, the acquisition modes of planimetric or volumetric objects are able to use *projections* of new objects on part of another object. It is also possible to build volumes by *copy-pasting* objects, which is useful for areas where several buildings are based on the same model (e.g. suburban areas with peaked roof houses).

The main advantage of these tools is the automatic estimation of the heights of the points, which appears to be a tedious task for non-photogrammetrist users. It can also be seen as a way of relaxing the operator's attention on buildings where the algorithm is supposed to work (good contrast), so that he can focus on "difficult" objects.

3. RENDERING

This stage is divided in two steps: the Texture Mapping and the Terrain Integration.

3.1. Texture mapping

With terrestrial photos, the user is supposed to extract by hand the textures from the image, e.g. by pointing the corners for a rectangular face, and make the association with a face of the 3D model. Whereas with the registered aerial or satellite images, which have been used as a support for the geometric reconstruction, this process can be automatic. Such an approach allows the complete texturation of the site, because the vertical or quasi-vertical satellite or aerial images bring the terrain and roofs textures, whereas the terrestrial photos give the textures of the walls of the buildings.

The texture mapping process is highly interactive : the operator must identify on the terrestrial photos the images patch corresponding to the faces of the 3D objects extracted in the geometric reconstruction stage. Such a human identification could be avoided if the photos were possible to register as the aerial or satellite images. The main obstacle to the photogrammetric exploitation of such photos is the type of camera generally used ; these cameras cannot be qualified of metric cameras, which means their geometric model cannot be known and estimated with a precision compatible with the required reconstruction quality. That's why we consider the texture mapping as an interactive step.

As it is symbolized on figure 7, this step consists in extracting the image patch to be mapped, eventually retouching this patch, and finally mapping this texture to the corresponding face of the object. The retouching step can be useful to correct the artefacts caused by trees, signs, or cars on face's texture, and to clean the edges of the extracted textures ; a classical software product like Photoshop is able to cope with these problems with efficiency. It has to be noticed that the texture mapping step includes the warping of the image patch to make it fit in the geometry of the face (a rectangle in the example of figure 7).

For faces not seen on photos, it is possible to map a synthetic texture, or a texture previously extracted on other images.

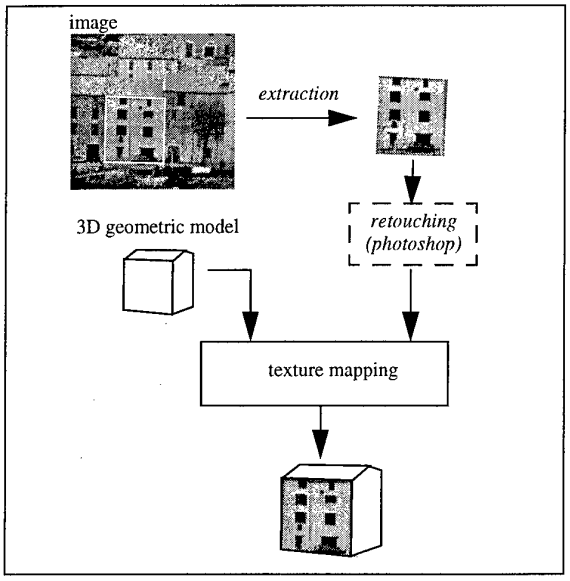


Figure 7 : Principles of texture mapping

3.2. Terrain integration

Terrain integration represents the formatting of the data collected to allow a realistic representation of the site. It uses the following steps:

— generic objects addition: a database of generic objects provides the full 3D description of the objects that have been extracted as bounding boxes ; the resizing of the 3D representation is done automatically using the bounding box;

— planimetry formatting : the planimetric objects (roads, rivers, ...) are formatted using pre-defined "gabarits" (templates for transversal representation of lineic primitives, like roads or rivers), insuring a realistic representation;

— terrain triangulation : the DTM must be formatted to be visualized; this formatting consists in computing a triangular patches set that represent the terrain in the most efficient manner; this triangulation is done coherentely with the extracted 3D objects, which insures the triangulated terrain is correctly connected to roads elements and buildings bases (Constrained Delaunay Triangulation);

— orthophoto mapping : the orthophoto is automatically mapped on triangular patches of the terrain, as well as on the

roofs not textured yet; in order to prevent from the lack of realism of the orthophoto textures at short distance (due to their resolution), we use a process of micro-texturation (high frequency textures) to insure a good realism of the animations when the point of view is too close to the ground.

The whole process of Terrain Integration is done with the TT&S software product called ModTer.

At this stage, the 3D site database can directly be used by a simulation software in order to provide realistic views and animation around the site.

4. EXPLOITATION

4.1. Scenario preparation

In this step, the operator determines the scenario he wants to play around the site. It is based on the TT&S software TrackMaker. The user can define the mobile object which is supposed to be used : helicopter, aircraft, tank, car, or even a pedestrian. Each mobile object is associated to a specific behaviour, insuring a realisitic animation. The user then defines the trajectory of the object, as well as the climatic conditions of the animation (weather, time).

As we will see below, the definition of trajectories is not indispensable for the animation, because the operator can decide to move in real time around the site using the keyboard and the mouse.

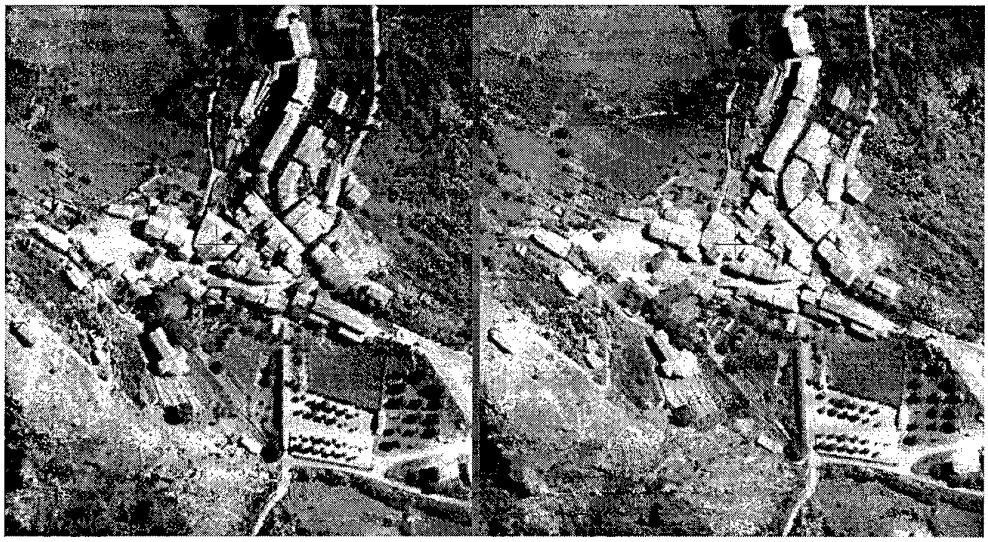
4.2. Animation

The animation step exploits the 3D site database to simulate with a great realism the scene that would be viewed from any point around the site. This exploitation is based on the TT&S software SpaceMagic, which is one of the world-wide references in the domain of simulation, offering high level performances in animation.

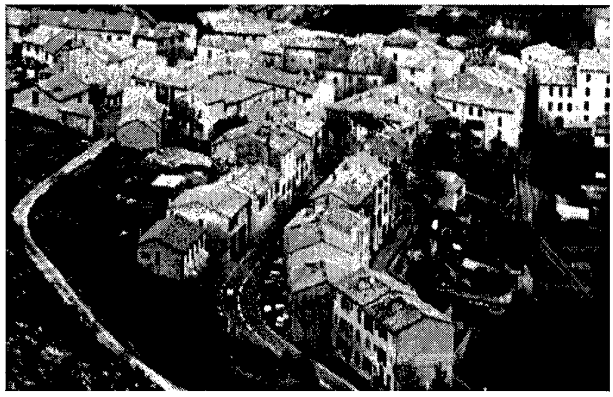
This software exploits the geometric models and the textures, to provide a totally fluent display of the reconstructed scene as the user moves around the site. The scene is computed and refreshed in real time. It is possible to use a predefined scenario, or the user can move interactively in real time.

This animation software can produce *fly-through* animations, simulating the scene viewed by a flying observer, but also the so-called *walk-through* animations, which can be a base for terrestrial mission preparation or training.

Note that the ouptut of the animation can be stored on videotape or CD-ROM to be played in an extern context.



aerial images pair (split view)



one of the helicopter photographs

Figure 8: source images

5. EXAMPLE OF RECONSTRUCTION

The example we present here is based on a data set collected on a village of the south-east of France. This data comprises :

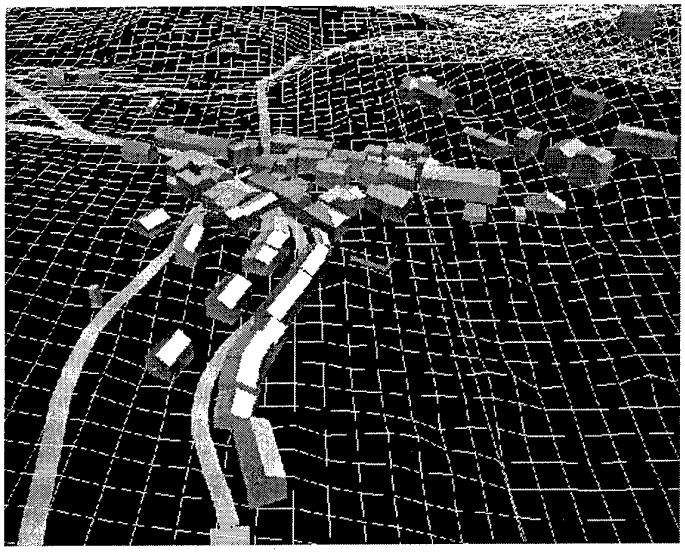
— a stereoscopic pair of high resolution aerial images (typically with a resolution of 1 meter), taken with a metric camera system (scanned black&white photos);

— several colour images, taken from a non metric camera fixed on an helicopter, flying at low altitude (allowing grazing angles of view).

Such a combination of images allow the full demonstration of the capabilities of our system. The aerial images are used to geometrically reconstruct the site, whereas the helicopter views allows the extraction and mapping of real textures on walls and roofs.

The figure 8 represents a split view of the stereoscopic pair, as well as one of the helicopter view (which is actually coloured).

On figure 9, one can see the result of the geometric reconstruction stage. The reconstructed site can be seen on the GIS window (which is a cartographic representation of the site) and in the perspective window. As it can be seen, the perspective window is able to display the 3D models with the DTM (represented with wireframes) or with the orthophoto drapped on the DTM. This latter representation gives a realistic view of the reconstructed site, but does not include the textures on buildings. This could be seen as the result of the geometric reconstructing stage.



perspective window with wireframe DTM



GIS window



perspective window with orthophoto drapped on the DTM

Figure 9: result of the the Geometric Reconstruction

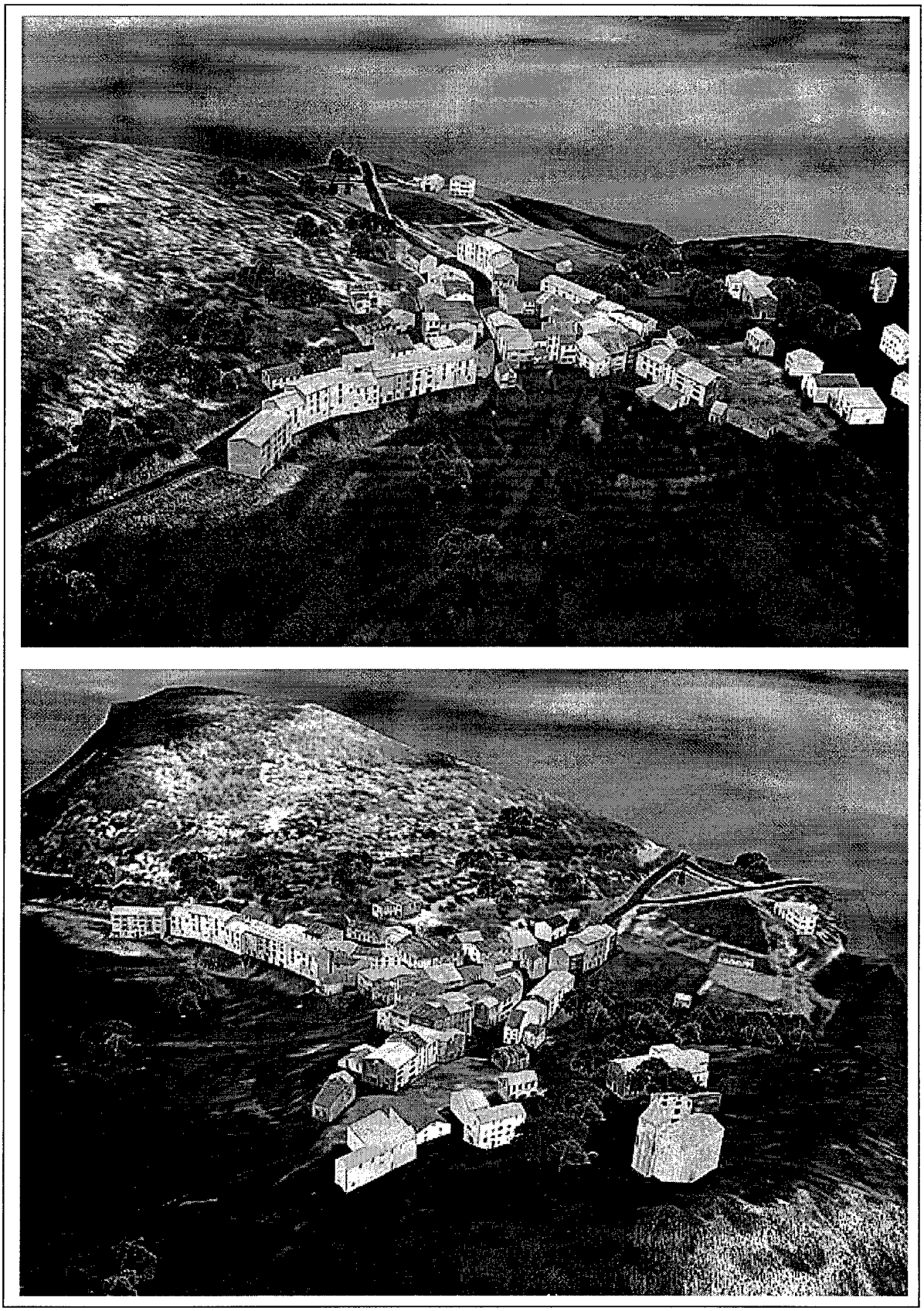


Figure 10: result of the 3D site reconstruction (geometry and textures)

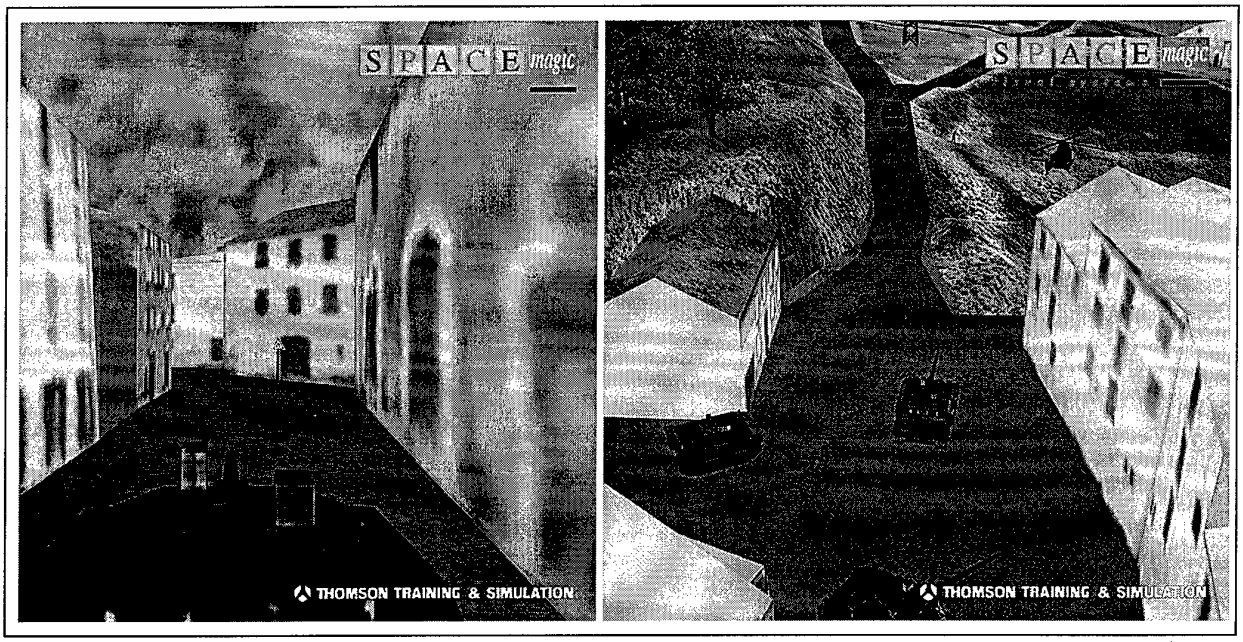


Figure 11: snaphots of a terrestrial mission preparation scenario

The final model site, including the wall textures can be seen on figure 10. Note that this view is presented in grey level in this article for printing constraints. In fact, the textures that have been mapped on the objects and the terrain are coloured, and give a realism even better to the reconstructed scene.

For the experimentation, we built a terrestrial mission repetition scenario, including helicopters, tanks and ground troops. Two snapshots of this scenario are shown on figure 1. On one of these snapshots, one can see a street of the village as it would be seen by a tank at ground level. The other snapshot represents a complex scene of the scenario, where tanks, troups and helicopters are evoluating on the site.

CONCLUSION

We described an image-based system for 3D site model acquisition and exploitation, dedicated to the production and exploitation of realistic 3D site databases.

This system comprises a geometric reconstruction module, allowing the extraction of 3D objects and terrain from a set of aerial or satellite high resolution images, and a rendering step, where textural information from any image are mapped on the extracted objects. These two steps provide 3D databases that can be used by the exploitation module of the system, based on a powerful simulation software. This system has been prototyped, and the results have demonstrated the capabilities of our approach.

With this prototype, we show that an autonomous 3D databases production line from images is feasible, and demonstrated its interest from an operational point of view. The availability in a near future of more and more sources of high resolution images will certainly emphasizes the need for such systems.

AKNOWLEDGEMENTS

Part of this work was supported by the french MoD (DGA/DSP/SPOTI/OER).

REFERENCES

[1] W. Förstner, "GIS - The Third Dimension", *Proceedings of the IUSM WG on GIS/LIS Workshop*, Hanover, september 1995.

[2] L.W. Fritz, "Commercial Earth Observation Satellites", *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXI, Part B4, Vienna, 1996, pp. 273-282.

[3] P. Fua, "Cartographic Applications of Model-Based Optimization", *Image Understanding Workshop*, 1996, pp. 409-419.

[4] O. Jamet, "Extraction Automatique d'Objets sur Stations Photogrammétriques numériques", *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXI, Part B3, Vienna, 1996, pp. 365-376.

[5] A. Heller, P. Fua, C. Connolly and J. Sargent, "The Site-Model Construction Component of the RADIUS Testbed System", *Image Understanding Workshop*, 1996, pp. 345-355.

[6] F. Lang and W. Förstner, "3D-City Modeling with a Digital One-Eye Stereo System", *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXI, Part B4, Vienna, 1996, pp. 261-266.

[7] F. Leymarie, N.Boichis, S. Airault, O. Jamet, "Towards the Automation of Road Networks Extraction Processes", *SPIE*, vol. 2960, Taormina, Italy, 1996, pp. 84-95.

[8] J.F. Raper and B. Kelk, "Three-Dimensional GIS", in *Three-Dimensional Applications in Geographical Information Systems*, J. Raper Ed., Taylor & Francis, London, 1989, pp. 299-317.

[9] I. Tannous, S. Gobert, T. Laurençot & P. Vorms, "SAPHIR-3D: a System for 3D Site Model Acquisition", to be published in *IAPRS*, Vol. 32, Part 3-4W2, "3D Reconstruction and Modeling of Topographic Objects", to be held in Stuttgart, september 17-19 ,1997.

PAPER No. 4

DISCUSSOR'S NAME: M. Desbois

COMMENT/QUESTION:

Utilisez-vous des moyens de désignation tri-dimensionels pour saisir l'information et l'intégrer dans la simulation?

(Do you use three-dimensional designation facilities to input data and incorporate it into the simulation?)

AUTHOR/PRESENTER'S REPLY:

Deux boules roulantes (x,y) (z) permettent la désignation 3D.

(Yes, we use two tracking balls (x,y) and (z) for 3D designation.)

Airborne Dual Sensor mmW-Signatures of Maritime Targets and Sea-Clutter

R. Makaruschka, H. Essen
FGAN - Forschungsinstitut für Hochfrequenzphysik
Neuenahrer Str. 20
D-53343 Wachtberg-Werthhoven
Germany

1. SUMMARY

In the framework of a NATO measurement campaign airborne signature measurements were conducted over sea with the dual frequency, polarimetric mmW Synthetic Aperture Radar MEMPHIS (Millimeterwave Experimental Multifrequency Polarimetric High Resolution Imaging System) with simultaneous operating front-ends at 35 GHz and 94 GHz onboard a cargo aircraft in side-looking configuration. Both front-ends are tied to the same system reference and are using the same IF pre-processing and radar waveform-generator. So as well a concise comparison between data at the two frequency bands, 35 GHz and 94 GHz, can be made as a use of the data for multichannel/multifrequency SAR processing.

The paper describes the system configuration and the mmW-SAR processing algorithm and gives representative results for the generated radar images for ship targets, chaff and the sea clutter with emphasis on the multiparameter evaluation.

2. INTRODUCTION

High resolution mmW-radar sensors are candidates for autonomous guidance applications for missiles directed against maritime targets. In a close-in phase of an attack it is essential for such seekers to discriminate between target and chaff. To be able to optimize self defence strategies as well as the performance of chaff against such sensors it is necessary to know the highly resolved signature of naval targets and chaff in a natural environment. In the final phase of the approach of an anti-ship missile seeker also steeper depression angles may be relevant. For this geometry the contrast between target, chaffs/decoys sea-clutter, including the impact imposed by the propagation medium (multipath effects) have to be known. Only experiments with airborne sensors can deliver this necessary information.

3. Measurement Set-Up

3.1 Experimental Radar

As mentioned above, the SAR system employs two radar front-ends, one at 35 GHz and the other at 94 GHz. Both are controlled by a common VME-Bus computer and tied to the system reference, from which all frequencies and trigger impulses used in the system are derived. Each receiver has four channels. The downconverted signals are quadrature demodulated to result in I- and Q-phase components and the logarithmically weighted amplitudes. The IF-signals from both front-ends are fed to the data acquisition and recording electronic. Additional information is retrieved from the inertial navigation system of the aircraft and a separate GPS-receiver unit. Fig. 3.1 shows a block diagram of the total set-up.

In dependence on the application, the systems can be used with polarimetric monopulse feeds, sensing elevation and transverse deviations or an interferometric pair of antennas with orthomode transducers to sense both polarimetric components. The elevation/azimuth-asymmetry of the beam, which is generally necessary for the SAR application is achieved by an aspheric lens in front of the feed horns.

A more detailed description of the radar was already given elsewhere [1]. MEMPHIS is, to our knowledge, the only SAR-system capable to gather full polarimetric data at both relevant mmW bands simultaneously.

The performance data of the front-ends are summarised in Tab. 3.1:

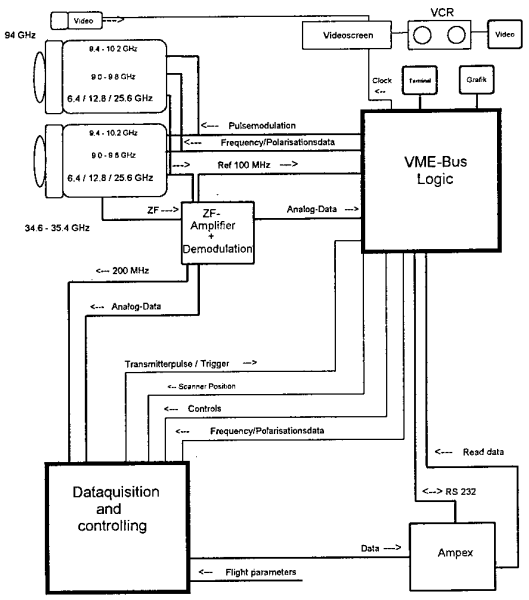


Fig. 3.1: Block-Diagram of the 35/94-GHz-SAR "MEMPHIS"

The measured data are recorded by means of a high speed digital recording machine AMPEX DCRSi. The recording speed is limited to 33 MByte/s allowing to record 660 range gates at a PRF of 6 kHz for both transmit frequencies.

In parallel to the radar data inertial data from the aircraft as well as a time code and GPS-data are recorded.

The calibration is based on pre- and postflight measurements against trihedral and dihedral precision corner reflectors.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

Transmitter:		
Centre Frequency	94 GHz	35 GHz
Output Power	1 kW	900 W
PRF	6 kHz (typ.)	
Pulse Width	80 ns - 2 µs	
Spectral Purity	> - 70 dB / Hz	
Phase Stability	> 10° RMS	
Polarization	switchable from Pulse to Pulse H-V or LC-RC	
	switchable manually linear or circular	
Waveform	Polyphase Franck Code	
Receiver:		
Four identical channels for co- and cross-polarisation, Monopulse or Interferometry		
Dynamic Range	60 dB	
System Noise Figure	15 dB (SSB)	
Antenna:		
Type	Monopuls or Interferometric Pair	
Dielectric Lense		
Diameter	30 cm	
3-dB-Beam Width		
Azimuth	1° (94 GHz)	2.5° (35 GHz)
Elevation	12°	16°
Gain	29 dB	27 dB
Data Acquisition:		
Data recording :	AMPEX DCRSi	
Data Rate	33 MByte / sec	
Channels	8 + flight parameters, time-code, GPS, etc.	
Capacity	48 G Byte	

Tab. 3.1: Performance Data of the Experimental Radar MEMPHIS

3.2 Geometry

The measurements discussed below were conducted in side-looking geometry.
Typical geometrical and flight parameters are tabulated in Tab. 3.2 .

Speed of Aircraft	130 kn.
Height above Ground	1000 ft
Depression Angle	20 °
Slant Range at Beam Centre	770 m
Swathwidth	406 m

Tab. 3.2 : Typical Measurement Parameters for mmW-SAR-Imaging

To be able to give the signature of ship for the main aspect angles of the ship - backboard and starboard, bow and stern - the aircraft flew along a track as indicated in Fig. 3..2.
The aircraft flew with a lateral displacement of 770 m from the target, while the ship moved slowly with a constant speed and held a steady and straight course throughout a trial run. To manoeuvre the aircraft into a stable straight course the aircraft had to fly loops as indicated in the sketch. One total measurement run on a ship took about 4 minutes.

Target tracking was achieved manually using a boresight video image of the target, within the radar's range gate.
Also for the characterization of chaff-clouds a straight flight path with a correct lateral displacement was necessary. Due to the limited swath-width of only about 400 m, the correct position of the chaff and the wind direction was important to adjust the lateral displacement of the aircraft track.

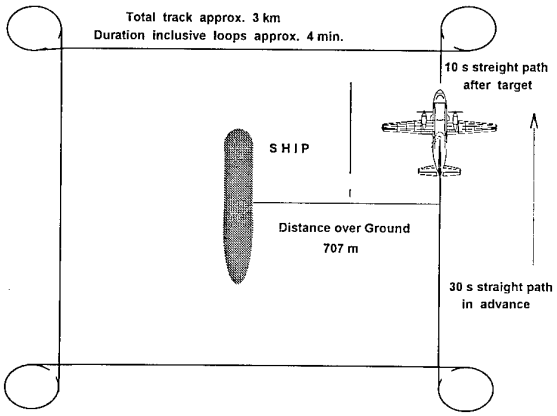


Fig. 3.2: Sketch of Flight Path during the Measurements

4. EVALUATION METHODS

The evaluation employed image based methods, namely a synthetic aperture algorithm. Based upon those methods, images areas of interest, containing clutter or targets were selected to apply integral evaluation methods. In addition to this approach circular flight patterns as suggested by Jones at Langley [2] were used to generate radar data in a reflectometer type of radar mode with the aim to gather clutter data for the full aspect angle range of 360° with respect to the direction of the sea swell.

4. 1 Image Generation

In a first step the radar data undergo off-line track, calibration, and reformatting procedures. Afterwards synthetic aperture processing is applied to these data to generate radar images for co- and cross-polarisation and both at 35 GHz and 94 GHz.
As already discussed elsewhere [3] SAR-processing at millimeterwavelengths requires a considerable lower amount of sophistication in comparison with algorithms applied at lower radar-frequencies. This can mainly be attributed to the short aperture length at mm-wave frequencies. Taking this into account, the SAR-algorithm used here is relatively simple although fully automatic autofocussing is applied, using only radar-data without supply of external inertial navigation information.

4.2 Integral Evaluation

The integral evaluation covered the determination of statistical characteristics of ships and clutter, and its dependence on the radar band, 35 GHz and 94 GHz, and also on aspect angle and environmental conditions. Special care was taken on polarimetric features, which are well recognised means to enhance image contrast [4] and to do classification [5] according to the scattering mechanisms involved.

5. Results

5.1 Sea Clutter

Figs. 5.1 to 5.3 show plots for the reflectivity of the sea surface versus aspect angle with respect to the wave direction for the radar frequencies 35 GHz and 94 GHz and polarization states VV and HH, VH and VV. The curves apply to sea state 5 and open sea conditions.

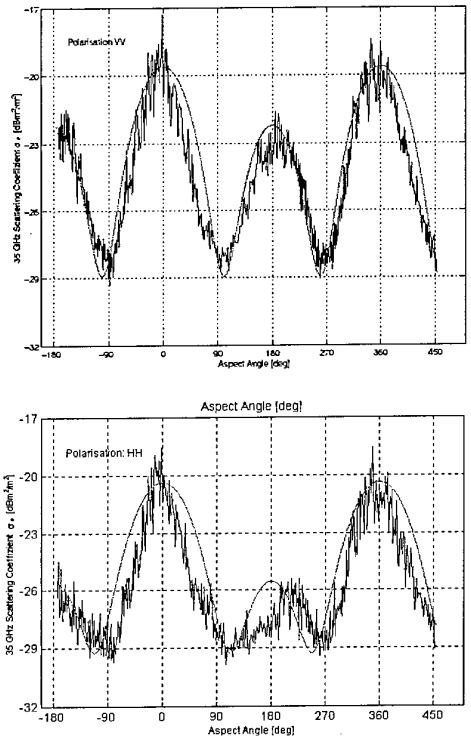


Fig. 5.1: Reflectivity versus Aspect Angle
Radar Frequency 35 GHz, T/R-Polarizations VV and HH.

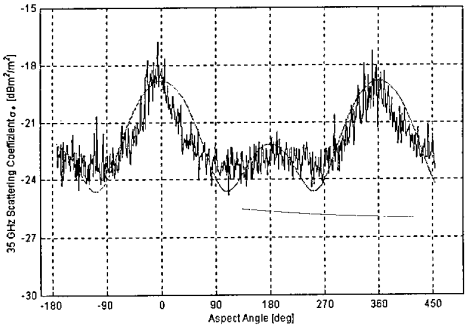


Fig. 5.2: Reflectivity versus Aspect Angle Radar Frequency 94 GHz, T/R-Polarization VV

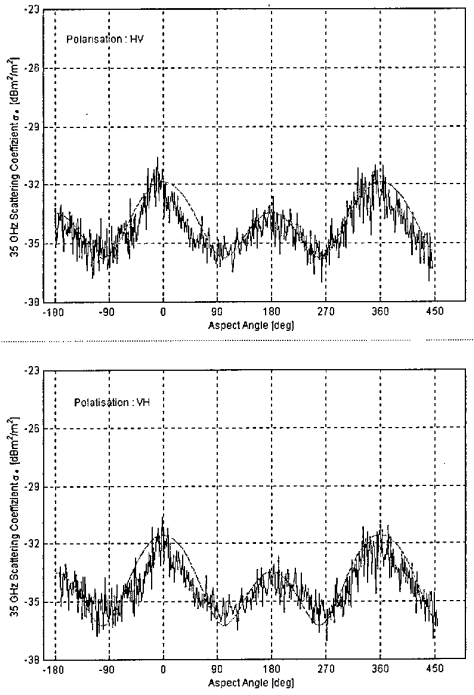


Fig. 5.3: Reflectivity versus Aspect Angle
Radar Frequency 35 GHz, T/R-Polarizations HV and VH.

The figures show the periodic pattern of growing and decreasing reflectivity dependent on the wave aspect, i.e. upwind, down-wind or cross-wind.
The experimental curves were modelled by the function [6]
$\sigma_0 = A + B \cos\phi + C \cos 2\phi$ [m^2/m^2],
where ϕ is the horizontal angle between the radar look direction and the upwind direction. The coefficients A, B and C are functions of the depression angle θ, wind speed u, radar wavelength λ and polarization p and can be expressed by means of a factor times the windspeed with the wind speed factor γ:
$A = a(\lambda,\theta,\phi,p)\, u^{\gamma a(\lambda,\theta,\phi,p)}$
$B = b(\lambda,\theta,\phi,p)\, u^{\gamma b(\lambda,\theta,\phi,p)}$
$C = c(\lambda,\theta,\phi,p)\, u^{\gamma c(\lambda,\theta,\phi,p)}$
The model curves were fitted to the experimental curves and are shown in the figures together with the experimental data using the relations between σ_0 at up-wind (u), down-wind (d) and cross-wind (c) which are easily derived :

$$A = \frac{\sigma^o{}_u + \sigma^o{}_d + 2\sigma^o{}_c}{4}$$

$$B = \frac{\sigma^o{}_u - \sigma^o{}_d}{2}$$

$$C = \frac{\sigma^o{}_u + \sigma^o{}_d - 2\sigma^o{}_c}{4}$$

The parameters were extracted for the available polarization states and radar frequencies and a depression angle of 37.5°. They are tabulated in Tab. 5.1.
σo at 94 GHz reproduces very well the number , which was measured for similar sea conditions during earlier measurements [7] for which the wind speed exponent was derived to be $\gamma = 2.3$.

T/R-Polarization	H-H	H-V	V-V	V-H
35 GHz				
A	0.0038	0.4125E-3	0.005	0.4125E-3
B	0.0038	0.1 E-3	0.0025	0.125 E-3
C	0.0025	0.1375 E-3	0.0035	0.1625 E-3
94 GHz				
A			0.0068	
B			0.0035	
C			0.0028	

Tab. 5.1: Model Factors for σo versus Aspect Characteristic at Sea State 5

To our knowledge these measurements are the only full polarimetric millimeterwave measurements of this kind, which allow to assess the frequency dependence of the model factors. If compared with respective X-band results [8] the principle characteristics, especially at 35 GHz, are very similar. At 94 GHz different dominant scattering processes lead to a significant change especially for the coefficient B, which is related to the up-wind/down-wind ratio.
General results were the following:

- sharp cross wind nulls are appearing
- the reflectivity at down-wind is always lower than at up-wind. The up-wind/down-wind ratio is dependent on the parameters λ,θ,ϕ,p.
- the co-polar signal at vertical polarization is stronger than the co-polar return at horizontal polarization. The ratio can range up to 10 dB. This different characteristic is more pronounced for low sea states than for rough sea.
- at 94 GHz the reflectivity at down-wind is close to that at cross wind.

All these results can be explained by a scattering model for the interaction between the sea surface and mm-waves, which is based on a scattering dominated by small facets located in the crests of the waves where the rise of the surface is steep and multiple scattering between facets is frequent.

5.2 Ships and Chaff

5.2.1 *SAR-Imaging of Ships and Chaff*

Measurements of ships and ships firing chaff have been done at 35 GHz and at 94 GHz with linear polarisation and with a fixed radar depression angle of 22.5°.
Fig. 5.4 and 5.5 show SAR images at 35 GHz and 94 GHz for a typical scene with bloomed chaff clouds and the corresponding ship.
The SAR images indicate where dominant scatterers on the ship are located. There exist cases, where due to multipath effects signal portions are reflected from the ship's body (dominant scatterers on the ship) into the water surface reproducing reflected radar images along the water line of the ship. This, however could be observed only for 35 GHz at calm sea conditions.

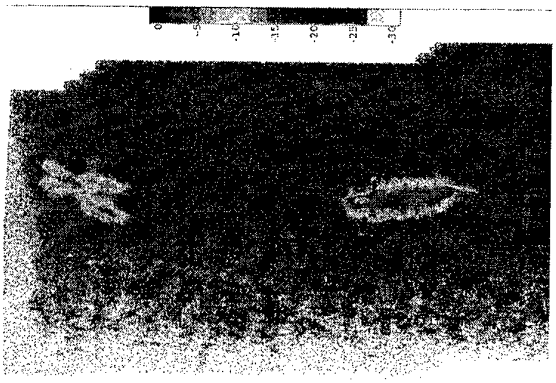


Fig. 5.4 SAR Image of Chaff and Ship at 35 GHz

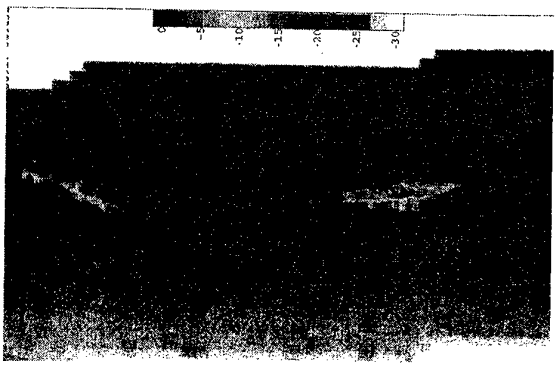


Fig. 5.5: As Fig. 5.4 at 94 GHz

Generally the measurements showed that the scattering centers appear more concentrated at 94 GHz, while at 35 GHz the scatterers produce wider lobes.
The signatures for chaff are considerably less wide spread at 94 GHz if compared with 35 GHz. It has, however to be noted, that the types of chaff measured here were not specially designed against mm-wave sensors.

5.2.2 *Statistical evaluation of Ship- and Chaff-Data*

Radar cross section evaluations have been carried out on different ship classes and different displacements of the ships by extracting statistical data from the SAR images..
The experimental results of radar cross sections at 35 GHz and 94 GHz are plotted in Fig. 5.6 versus ship type .
It can be seen that the measured average radar cross section of ships depends on the type of ship and also on the radar frequency.
For comparison reasons Fig. 5.7 presents the average RCS of ships versus the displacement, given in tons.
Usually the RCS of ships is expected to increase with frequency [8], thus at Ka-band the RCS should be lower than at W-band. However, the expected increase of RCS with frequency was not observed as in earlier ground based measurements [9].

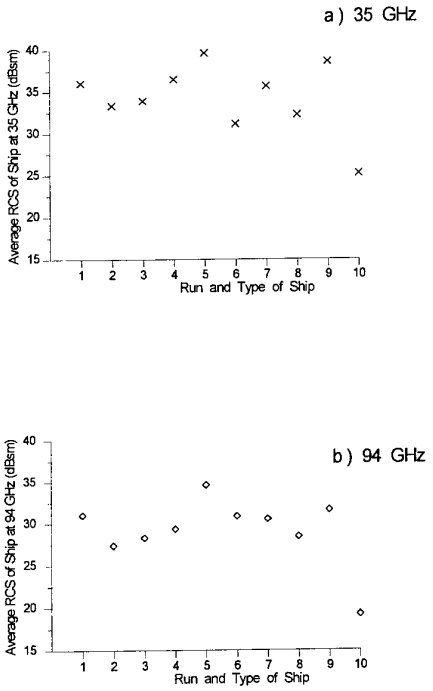


Fig. 5.6: Average RCS of different Ships of same Class at 35 GHz (a) and 94 GHz (b)

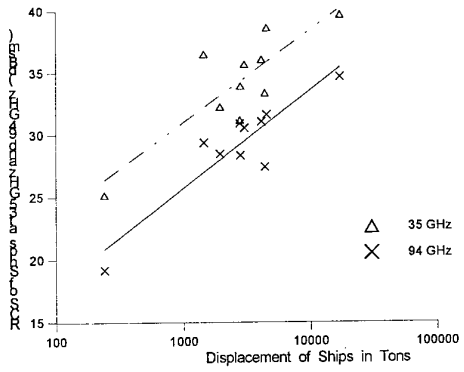


Fig. 5.7: Average RCS versus Displacement

In contrary, comparing both frequency bands, the radar cross section at 35 GHz is about 6 dB higher than the RCS at 94 GHz for almost all classes of ships considered here.
This can be interpreted by the following: Compared with the radar wavelength ($\lambda = 3$ mm and $\lambda = 8.6$ mm) flat plates and corners on a ship's body are not smooth and straight, therefore, they can not be looked at as basic dominant scatterers with ideal shape. Moreover at millimeter wavelength not well defined scatterers determine the signature but the whole community of numerous point source scatterers are considered to be the coherent average sum of their individual parts [8].
Fig. 5.8 presents the measured RCS of chaffs at 35 and 94 GHz at linear polarisation: HH. The highest cross section was measured to be about 35 dBsm at 35 GHz and at 94 GHz, respectively. At 35 GHz the average RCS of chaffs(27 dBsm to 30 dBsm) was about 5 dB higher than the RCS of chaff-clouds at 94 GHz.
In general the RCS of ship (broadside aspect) compared with the RCS of chaff-clouds was about 6 dB higher at both frequencies. For signal lock-transfer, the RCS of the chaff-clouds should be preferably higher than the ship's RCS. Therefore, lock- transfer with chaff as used here at millimeter waves would hardly be feasible.

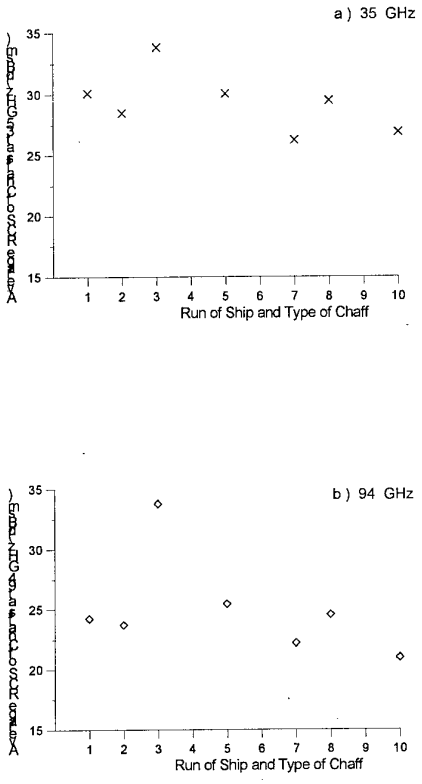


Fig. 5.8: Average RCS of different Types of Chaff for 35 GHz (a) and 94 GHz (b)

6. CONCLUSIONS

In the mmW radar bands (Ka-band and W-band) Synthetic Aperture Radar (SAR) systems offer attractive features for remote sensing applications.
The system used here is capable to do clutter measurements over sea and to supply the reflectivity data needed for the development of maritime mm-wave sensors, which, although typically not long range, will gain increasing importance in the future. The discussion of typical clutter data showed that the characteristics of 35-GHz sea -clutter seems to be very similar to that of classical radar bands and that at 94 GHz other scattering processes are dominant, which lead to a considerably reduced reflectivity of the sea surface in down-wind and cross-wind directions and an increase in up-wind direction.
SAR signature measurements on ships and chaff confirmed, what was already known from landbased measurements: The RCS of ships and chaffs depends on frequency, on the target geometry, and the type of chaff. It is especially remarkable that the RCS at 94 GHz, in contrary to theoretical

predictions, is lower than for 35 GHz as well for ships as for the chaff.
At millimeter wavelengths not distinct scatterers determine the signature of a target, but the mutual interaction of numerous scatterers of similar importance are dominantly contributing to the total radar cross section.
For self protection of a ship, RCS reduction of the ship will favour lock-transfer to chaff.
The images showed, that SAR at mmW exhibits certain characteristic features:
- due to the relatively rough surfaces on a ship, compared to the radar wavelength, dominant structures or scatterers exhibit no excessive backscatter cross-sections.
- the total radar cross section of ships is about 6 dB higher than the RCS on chaff clouds

6. REFERENCES

[1] S. Boehmsdorff, H. Essen, R. Makaruschka, "Target and Background Measurements over Land and Sea with a Polarimetric 35/94-GHz Synthetic Aperture Radar", SPIE Aerosense '97, Orlando Fl. 1997

[2] F.T. Ulaby, R.K.Moore, A.K.Fung, "Microwave Remote Sensing, Active and Passive", p. 1666, Artech House, Dedham, 1986

[3] H. Essen, H. Schimpf, "Remote Sensing with an Airborne 94-GHz Synthetic Aperture Radar", Proc. of EUSAR '96, Königswinter, Germany, Mar. 1996

[4] J.J. Van Zyl, "Unsupervised Classification of Scattering Behaviour using Radar Polarimetry Data", IEEE Trans. on Geoscience and Remote Sensing Vol. GE-27, pp. 36-45, 1989

[5] W.M. Boerner,"Wideband Interferometric Sensing and Imaging Polarimetry, and its Relevance to Wide Area Military Surveillance and Environmental Monitoring of the Terrestrial and Planetary Covers", SPIE Vol 2845, "Radar Processing, Technology and Applications"

[6] R.K.Moore, A.K.Fung, "Radar Determination of Winds at Sea", Proc IEEE 67, pp. 1504 - 1521

[7] H.H. Fuchs, "Measurements of Radar Backscatter from the Ocean Surface at 94 GHz as a Function of Windspeed, Direction and the Modulation by the Ocean Waves during the SAXON-FPN Experiment", AGARD Conference Proc. 543, P. 43, Rotterdam, Oct. 1993.

[8] M. Skolnik, "An Empirical Formula for the RCS of Ships at Grazing Incidence", IEEE Trans. on Aerospace and Electronic Systems, p. 292, Mar. 1978

[9] H.Essen, H.Schimpf, E.P. Baars, " Radar Cross Section Measurements on Ships at 94 GHz". AGARD Conf. Proc. EP- , London, 1984

[10] Technical Proc. AC/243(Panel3)TP/3 pp.75-90,Workshop on Military Applications of Millimetre Wave Imaging", May 1994

PAPER No: 5

DISCUSSOR'S NAME: Streng

COMMENT/QUESTION:

1. How did you handle the occurrence of ghost-scatterers? (In order to be able to automatically classify ships?)

2. Did you implement this?

AUTHOR/PRESENTER'S REPLY:

1. Compare with ships' model, several imges under various aspect angles.

2. May be done in the future.

DATA FUSION FOR LONG RANGE TARGET ACQUISITION

Patrick Verlinde, Dirk Borghys, Christiaan Perneel, Marc Acheroy
Signal and Image Centre
Royal Military Academy
Renaissancelaan 30,.
B1000 Brussels
Belgium

SUMMARY

An approach to the long range automatic detection of vehicles, using multi-sensor image sequences is presented. The algorithm we use was tested on a database of six sequences, acquired under diverse operational conditions. The vehicles in the sequences can be either moving or stationary. The sensors are stationary, but can perform a pan/tilt operation. The presented paradigm uses data fusion methods at four different levels (feature level, sensor level, temporal level and decision level) and consists of two parts.
The first part detects targets in individual images using a semi-supervised approach. For each type of sensor a training image is chosen. On this training image the target position is indicated. Textural features are calculated at each pixel of this image. Feature level fusion is used to combine the different features in order to find an optimal discrimination between target and non-target pixels for this training image. Because the features are closely linked to the physical properties of the sensors, the same combination of features also gives good results on the test images, which are formed of the remainder of the database sequences. By applying feature level fusion, a new image is created in which the local maxima correspond to probable target positions. These images coming from the different sensors are then combined in a multi-sensor image using sensor fusion. The local maxima in this multi-sensor image are detected using morphological operators. Any available prior knowledge about possible target size and aspect ratio is incorporated using a region growing procedure around the local maxima. A variation to this approach, that will also be developed in this paper, combines the previous feature and sensor level fusion, by extracting the features in each sensor as before but using the feature level fusion directly on the combination of all features from all sensors in what is sometimes called a « super feature vector ». Tracking is used in both cases to reduce the false alarm rate.
The second part of the algorithm detects moving targets. First any motion of the sensor itself needs to be detected. This detection is based on a comparison between the spatial co-occurrence matrix within one single image and the temporal co-occurrence matrix between successive images in a sequence. If sensor motion is detected it is estimated using a correlation-based technique. This motion estimate is used to warp past images onto the current one. Temporal fusion is used to detect moving targets in the new sub-sequence of warped images. Temporal and spatial consistency are used to reduce the false alarm rate.
For each sensor, the two parts of the algorithm each behave as an expert, indicating the possible presence of a target. The final result is obtained by using decision fusion methods in order to combine the decisions of the different experts. Several « k out of n » decision fusion methods are compared and the results evaluated on the basis of the 6 multi-sensor sequences.

1 INTRODUCTION

Long range automatic detection of vehicles is of great military importance to modern armed forces. The most critical factor of any system for automatic detection is its ability to find an acceptable compromise between the probability of detection (= 1 - probability of a miss) and the number of false alarms. This is the classical trade-off one finds in binary hypothesis testing between the two types of error one can make: the false rejection (FR : which corresponds here to a miss : there is a target, but it has not been found) and the false acceptance (FA : which is in this case the same as a false alarm : there is no target, but the system thinks there is one). In a single sensor detection system it is well known that if one reduces one type of error, the other type of error automatically increases. A possible way-out of this deadlock is to use more than one sensor and to combine the information coming from these different « experts ». This combination or (data) fusion can be done on different levels. In this paper, only the (common) case of a centralised fusion processor with all its sensors connected in parallel will be considered.
In the specific data fusion literature [1-5] one often distinguishes between the following (or equivalent) fusion levels: low level fusion (also called score or measurement level fusion), medium level fusion (which includes feature and sensor level fusion), high level fusion (also called decision level fusion) and temporal level fusion. As can be expected, in real (-time) applications, there is a trade-off to be made between the amount of information that can be combined and the bandwidth necessary to communicate all this information to the centralised fusion processor. The lower the level of fusion, the more information is available to be combined, but the larger becomes the bandwidth necessary to communicate with the centralised fusion processor (or for a fixed bandwidth, the slower becomes the fusion process). Vice versa one sees that when the level of fusion gets higher, the available information diminishes, but so does the necessary bandwidth. Furthermore not all data fusion levels are always applicable. For instance, if low level fusion is going to be used, care must be taken to combine only similar entities (scores, measurement results,...). It is therefore impossible to use low level fusion to combine the raw results coming from two (or more) totally different sensors (e.g. an imaging sensor and a range finder). But this constraint doesn't exist any longer on the decision level, where each sensor is considered as a separate « expert », who decides on his own. In the special case of target detection where the « hard » binary decision rule is

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

used (the « hard » decision is indeed bi-valued : target present (1) or not (0)), the central fusion processor contents itself to combine only the decisions (the 1's and the 0's) coming from different sensors, without considering the type of sensor. As a general conclusion concerning the different data fusion levels, one can state that all different fusion levels have their importance and their specific applicability domain.

Based on these considerations, we have tried to use data fusion on several levels to try to optimise the use of the available data. That is basically why this paper describes an approach to tackle the previously exposed problem using four different data fusion techniques related to several levels : feature level fusion, sensor level fusion, temporal level fusion and decision level fusion. The only fusion level that is not used in this paper is the low level fusion. This technique (in the form of pixel level fusion) is mainly used in remote sensing applications [6, 7]. In the main approach, we do however use two different medium level fusion techniques. In the following sections the use of these different data fusion techniques will be explained in more detail.

2. IMAGE DATABASE

For the development and testing of the algorithm, a database of 6 multi-spectral image sequences, numbered MS01 to MS06 [1] was compiled. The sequences correspond to two scenarios. The first scenario is a typical surveillance scenario in which the sensor watches a scene and tries to detect targets entering its field of view. In this scenario, the targets are moving. The second scenario is a reconnaissance scenario in which the sensor is mounted in a new terrain and it tries to detect the presence of vehicles which can now be stationary or moving. During image acquisition the sensor is stationary in both scenarios; it can only perform a pan and tilt operation. The following table presents some properties of the sequences.

Seq nr	Targets	Target Motion	Sc	Type of Sensors
MS01	Helicopter Truck	Across FOV	1	LW,TV
MS02	Truck	Towards Sensor	1	LW,TV
MS03	Helicopter	Across FOV	1	LW,TV
MS04	2 Tanks	None	2	LW,SW,TV
MS05	2 Tanks +Camoufl.	None	2	LW,SW,TV
MS06	Helicopter	Across FOV	1	LW,R,G,B

Table 1: Properties of sequences.

In the table the Sc column presents the scenario to which the sequence corresponds. In the sensor column the following abbreviations are used: LW and SW denote long-wave and short-wave infrared respectively. TV is B/W visual images. R,G,B are the components of a colour visual image. Each set of three subsequent sequences were acquired by the same sensor set. Sequence MS05 is the same as sequence MS04 except for the fact that in MS05 the targets are camouflaged.

[1] MS01-MS03: Courtesy of Defense Research Establishment Valcartier, Quebec, Canada ; MS04-MS05: Courtesy of Naval Air Warfare Center, China Lake, US; MS06: Courtesy of ASIAT-DTT, Peutie, Belgium

3. OVERVIEW OF THE APPROACH

The proposed algorithm consists of two independent parts. The first part searches for targets in single images while the second part uses multiple subsequent images in order to specifically find moving targets.

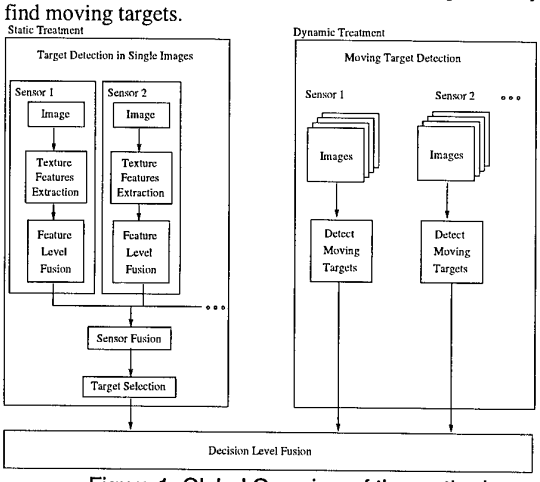


Figure 1: Global Overview of the method

For the first part of the algorithm we have chosen a semi-supervised approach based on texture feature extraction. Allthough we are not interested in explicitly modelling or measuring texture, these texture features are interesting because they are independent measurements of the local spatial distribution of grey values within an image and it is likely that some of these parameters will highlight the difference between targets and background. The texture parameters are even more appealing because it can easily be seen that features that are classically used for target detection such as intensity and gradient are just special cases of these texture parameters. Feature level fusion is used to combine the texture features from each image into a new image in which the grey value at each pixel is proportional to the probability that the pixel belongs to the target. These images from the different sensors are fused in a sensor fusion step.
The second part of the algorithm detects moving targets in sub-sequences from each sensor separately.
Each part of the algorithm behaves as an expert indicating the possible presence of vehicles in the scene. Decision fusion is used to combine the outcomes from all experts.

4. TARGET DETECTION IN SINGLE IMAGES (TDSI Module)

Introduction

For the detection of targets in single images, a semi-supervised approach based on texture features was chosen. For each sensor type, one image was selected to constitute the learning database. On these images the true targets were delimited. Then several texture parameters were calculated at each pixel of these learning images and logistic regression [4] was used to find a combination of the texture parameters that is proportional to the probability of finding a target at the corresponding image location.

The actual detection algorithm then applies the same combination to the texture features calculated on the remainder of the image database (test images). When this function is applied to the texture features calculated at each pixel of a test image, a new image, called feature-level-fused image, can be formed in which the maxima correspond to likely target positions. These feature-level-fused images, obtained from all the different sensors, are then fused again in a subsequent sensor fusion step.

To find the possible target positions, first the local maxima are determined in this sensor-fused image and then available prior knowledge about possible target size and aspect ratio is used to reject false targets.

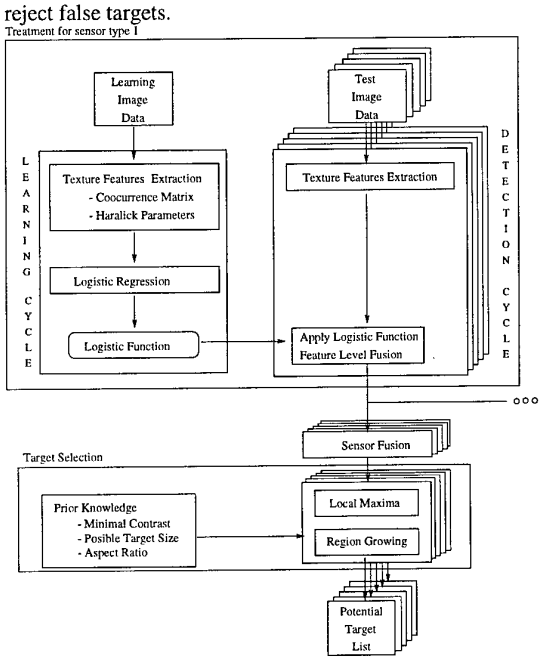


Figure 2: TDSI Module

Texture Parameters

The calculation of the texture features is based on the co-occurrence matrix. The co-occurrence matrix is defined as a function of a given direction and distance, or alternatively, as a function of a displacement (dx,dy) along the x and y direction in the image. For a given displacement (dx,dy), the (i,j) element of the co-occurrence matrix is the number of times the grey value G at the current position (x,y) is i when the value at the distant position (x+dx,y+dy) is j.

$$C^{dx,dy}(i,j) = P\left(G(x,y) = i \,\middle|\, G(x+dx, y+dy) = j\right)$$

The co-occurrence matrix can be calculated on the whole image. However, by calculating it in a small window scanning the image, a co-occurrence matrix can be associated with each image position. The centre of the window is denoted (x_c, y_c) and the corresponding co-occurrence matrix is $C^{dx,dy}_{x_c,y_c}(i,j)$

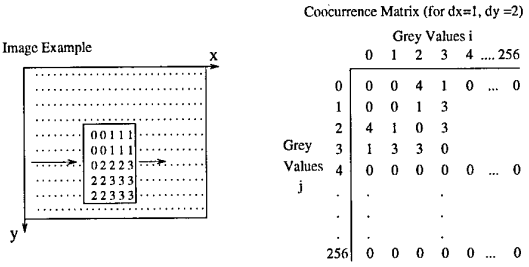


Figure 3: Co-occurrence matrix

In Figure 3 an example of a co-occurrence matrix is shown. The matrix corresponds to the small window of the image on the left and was calculated for a displacement of dx = 1, dy = 2. The textural features that were used, were introduced by Haralick [8-10] and are widely used in texture analysis. Based on the local co-occurrence matrix, the used parameters are defined as follows:

$$F_1(x_c,y_c) = \text{Energy} = \sum_i \sum_j C^{dx,dy}_{x_c,y_c}(i,j)^2$$

$$F_2(x_c,y_c) = \text{Contrast} = \sum_i \sum_j \left[(i-j) C^{dx,dy}_{x_c,y_c}(i,j)\right]$$

$$F_3(x_c,y_c) = \text{Max.Prob.} = \max\left[C^{dx,dy}_{x_c,y_c}(i,j)\right]$$

$$F_4(x_c,y_c) = \text{Entropy} = \sum_i \sum_j C^{dx,dy}_{x_c,y_c}(i,j) \log\left[C^{dx,dy}_{x_c,y_c}(i,j)\right]$$

$$F_5(x_c,y_c) = \text{Homogeneity} = \sum_i \sum_j \frac{\max\left[C^{dx,dy}_{x_c,y_c}(i,j)\right]}{\left[1+(i-j)^2\right]}$$

$$F_6(x_c,y_c) = \text{Variance} = \left[\sum_i (i-E_i)^2 \sum_j C^{dx,dy}_{x_c,y_c}(i,j)\right] \left[\sum_j (j-E_j)^2 \sum_i C^{dx,dy}_{x_c,y_c}(i,j)\right]$$

with $E_i = \sum_i i \sum_j C^{dx,dy}_{x_c,y_c}(i,j)$ and $E_j = \sum_j j \sum_i C^{dx,dy}_{x_c,y_c}(i,j)$

We are not interested in modelling or measuring texture but only in detecting a difference between target and background pixels. The "texture parameters" are only used as features of which we hope that some will highlight the difference between target and background. Because we do not intend to measure the texture within the target, the parameters are useful even for small targets and we can chose an arbitrary displacement (dx = 1, dy =1) for all calculations of the co-occurrence matrix.

The results for each texture feature can be converted into an image. Figure 4 shows the texture images corresponding to the first image set (IR and VIS) of sequence MS01.

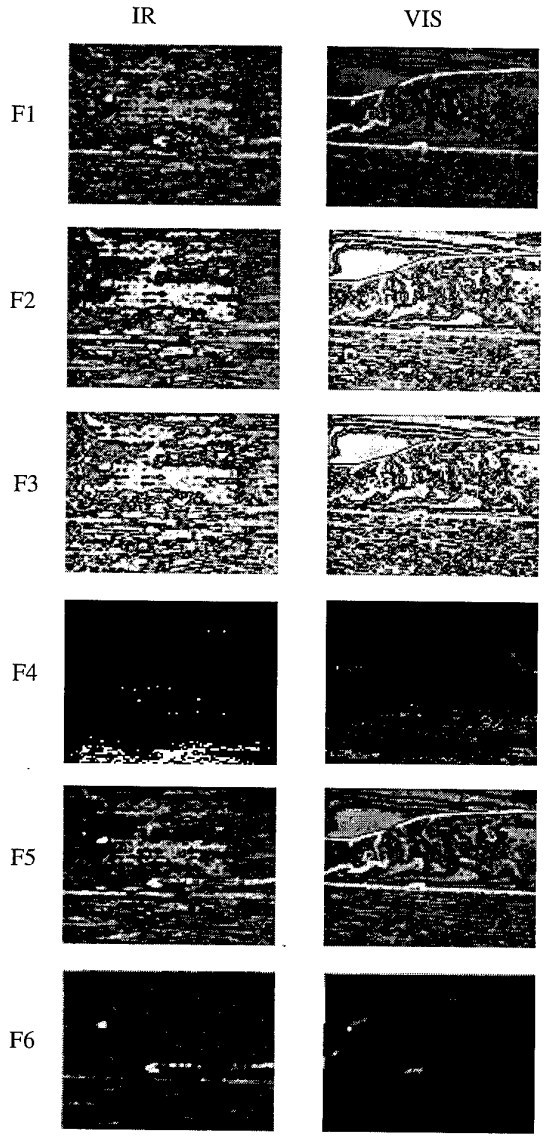


Figure 4: texture images

As can be noticed in Figure 4, the vehicles are clearly visible in some of the texture images. Hence the idea to combine the texture features to get an optimal discrimination between background and targets. When only two classes are involved, as is the case here (targets/ background), logistic regression offers an appropriate approach [4].

In the learning phase, at each pixel of the learning image(s), the texture features are calculated and stored in a table. Then the human operator interactively indicates the bounding rectangles surrounding the targets in the learning image(s) and a column is automatically added to the table assigning each measurement in the table to either class 0 (background), when it corresponds to an image pixel that falls outside the bounding rectangles, or class 1 (targets) when it is inside one of the rectangles. Logistic regression is then used to find a combination of the form :

$$p_{x,y}\left(\text{target}\middle|\vec{F}\right)=\frac{e^{b_0+\sum_i b_i F_i(x,y)}}{1+e^{b_0+\sum_i b_i F_i(x,y)}}$$

eq 1: Logistic Function

in which $p_{x,y}(\text{target}|\vec{F})$ is the conditional probability that a pixel (x,y) belongs to the class 1 (target class) given the vector of texture parameters $\vec{F}$ at the given pixel. The logistic regression was carried out using Wald's forward method. In this method, at each step, the most discriminant feature is added and the significance of adding it to the model is verified. This means that not all features will necessarily be included into the model.

Feature level fusion

If the learning images are representative for the images of a given sensor type, the most discriminating features for each sensor will have the highest weights b_i. Therefore, when using the same weights to combine the feature images of the remainder of the database into new images using equation eq 1, targets will appear as local maxima. This is the feature level fusion.

Sensor Fusion

The sensor fusion step combines the images obtained by the feature level fusion step. In the feature-level-fused images, for each sensor, targets appear as local maxima. Therefore it is possible to fuse these images by a simple multiplication. In the new images the targets will still appear as local maxima.

Region Growing around Local Maxima

In the sensor-fused image the local maxima will correspond to likely target positions. To detect the targets it is thus necessary to find these local maxima. A region growing procedure around the maxima is then used to incorporate available prior knowledge about possible target size and aspect ratio.

Local Maxima

The detection of local maxima is based on a succession of morphological operations [11, 12]. The basic operator is a dilation with a 2 by 2 structuring element.

Region Growing

To incorporate any available prior knowledge about the possible range of target size or aspect ratio, a region growing procedure is used. The initial regions for the region growing are the local maxima in the image. Surrounding pixels are added to these regions as long as their grey level differs less than a given threshold from the value at the local maxima. If the region becomes too large it is discarded. If the region growing of a given region stops before it reaches the upper size-limit, the other constraints are checked. If a constraint is not satisfied, the region is discarded.

Clutter rejection

To reduce the number of false targets, a simple clutter rejection scheme was implemented. A target is only declared if it was present in at least 7 out of 10 preceding images. More clever tracking methods [13, 14] could be used, but because our main interest is the exploration of data fusion methods, we did not implement this yet.

Modes of operation

The presented approach for the detection of targets in single images allows us to experiment with different levels of fusion. The target detection can be performed on the feature-level-fused images of each separate sensor (mode M1).
The second mode (M2) combines the feature-level-fused images from all sensors using the sensor fusion step described above.
In the third mode (M3), the logistic regression is applied to a superset of texture features, i.e. the feature-level-fused image is obtained by applying the logistic function to the set of features obtained from all sensors. This is only possible if the images from all sensors are registered.
A subdivision of modes 1 and 2 can be made according to whether the learning images used were acquired from the same sensor (i.e. the LW, SW and TV sensors for sequences MS04-05) or from the same generic class of sensors (i.e. using images from the LW and TV sensor of sequence MS01 to yield the weights for respectively all infrared-like sensors and all visual sensors).

5. MOVING TARGET DETECTION (MTD Module)

The second part of the algorithm focusses on the detection of moving targets. In order to detect moving targets, any sensor motion needs to be detected and its effects compensated first. Then, in a temporal fusion step, preceding images can be warped onto the current one. Moving objects will appear as a difference between the original image and the warped ones.

Detection of sensor motion

The detection of sensor motion is again based on co-occurrence matrices. This time the co-occurrence matrix is calculated between an image and the preceding one (temporal co-occurrence matrix).

$$C_{x,y}^{dx,dy,dt}(i,j) = P\left(G(x,y,t) = i \mid G(x+dx, y+dy, t+dt) = j\right)$$ If

no sensor motion occurred between the two images, ideally, for dx=dy=0 (i.e. no spatial displacement), all non-zero elements of the temporal co-occurrence matrix should lie on the diagonal. However, due to noise, there will be a small spread along the diagonal. If one calculates the spatial co-occurrence matrix for a small displacement, the spread along the diagonal is due to noise and to the fact that the image is not homogeneous.Therefore, when comparing this spatial co-occurrence matrix with the temporal co-occurrence matrix, the spread along the diagonal is expected to be the largest in the former one if no motion occurred between the two images that were used to calculate the temporal co-occurrence matrix. When motion is present, the spread along the diagonal quickly becomes larger. The measurement we used to detect sensor motion is based on the percentage of off-diagonal points in both co-occurrence matrices:

$$MC = \frac{\sum_j \sum_{i \neq j} C(i,j)}{\sum_j \sum_i C(i,j)}$$

This is calculated for both the temporal MC_{temp} and for the spatial co-occurrence matrix MC_{spat}. Sensor motion is said to be present if

$$\frac{MC_{temp} - MC_{spat}}{MC_{spat}} \geq 0.005$$

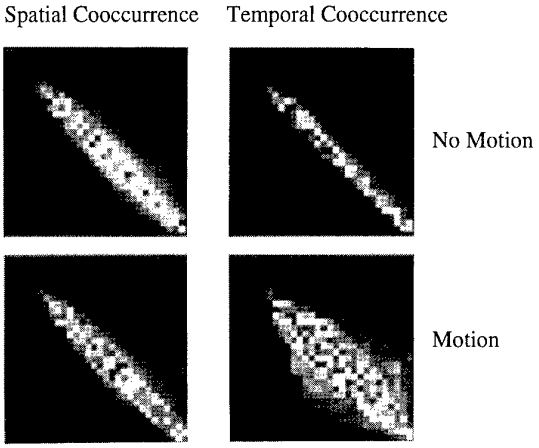


Figure 5: Detection of sensor motion

In Figure 5 the spatial and temporal co-occurrence matrix are shown. The upper images show the matrices for a part of a sequence where no sensor motion was present. The lower images show an example of both matrices calculated in a part of the same sequence where the sensor was moving.

Motion Estimation

If sensor motion is detected, we need to estimate it and compensate its effects on the images. Because it is known that the sensor is stationary and can only do a pan or tilt, the corresponding motion in the image will consist of a uniform translation. The motion is estimated by searching for the translation that optimises the correlation for a few horizontal and vertical lines. If the sensor is mounted in a moving vehicle or no apriori knowledge about the type of sensor motion is known, methods based on the model of a moving rigid planar patch [15] or optical flow techniques can be used [16].

Detection of moving targets

Once the sensor motion is estimated, preceding images are warped onto the current one. Then the original image is subtracted from the warped ones. If a moving object is present in the scene, we should find a large value at its position. The resulting images after subtraction are therefore thresholded and objects with acceptable size and aspect ratio are selected using a region growing procedure. Tracking is used to get the target list. Figure 6 shows the result of subtracting the original image from the warped ones.



Figure 6: Detection of Moving Targets

6. DECISION FUSION

The two parts of the algorithm each behave as experts indicating the possible position of targets in the scene. The final decision is reached by fusing the results of these experts. Because each expert only provides a binary decision - i.e. either a target is present or it is not - the decision fusion is implemented as a weighted "k out of N" voting-rule[1, 17]. The weights attributed to the decisions of each expert can depend upon several considerations.

For the detection of moving targets, each single sensor acts as an expert. For the detection of targets in single images, the decision was made after fusing the feature images from all sensors in Modes M2 and M3. In Mode M1, results from each separate sensor are passed to the decison level fusion step. The weights to be attributed to the decision of each expert will need to be adapted accordingly.

The weights may also depend on the type of scenario. In the surveillance scenario, the primary expert is the motion detector, whereas in the second scenario (reconnaissance), both types of expert are equivalent.

7. RESULTS AND DISCUSSION

In this section the results for the two parts of the algorithm are presented first. Then some results of decision fusion are presented.

Results of TDSI Module

Results of the feature-level fusion

For the feature level fusion a learning image set needs to be identified. For the so-called "generic sensor case"(GSC), the learning image set consists of the first Infrared (LW) and Visual image of MS01 on which the targets appear.

For the "sensor kind case"(SKC) three sets of learning images were identified. For sequences MS01-MS03 they are the same as the ones used in the "generic case". For sequences MS04-MS05 the first multi-sensor image set of MS04 is used, yielding weights for LW-,SW- and TV-type sensors. For sequences MS06 the first image set of MS06 is used. For the fusion of the superset of features (SSF) the learning images were the same as for the "sensor kind case". The following table presents the weights obtained for MS04-MS05.

Case	Sens	b0	b1	b2	b3	b4	b5	b6
GSC	LW SW	-15	2.4	0	0	4.7	0	0.01
	TV	-18	0	0	4.7	5.6	2.6	0
SKC	LW	-18	0	0.4	0	6.3	4.8	-0.008
	SW	-18	0	0	0	7.0	10	-0.06
	TV	-19	0	-0.06	4.6	7.6	3.5	-0.004
SSF	LW	-31	4.9	0	0	8.9	0	0
	SW	0	0	0	0	0	0	-0.05
	TV	0	0	0	0	6.8	8.8	-0.06

Table 2: Weights obtained by logistic regression for sequences MS04 and MS05.

Results for the single sensors (M1)

For each sequence, the probability of detection and the average number of false targets per image is given for both the "Generic Sensor Case" and the "Sensor Kind Case". Please note that for the first three sequences, the two cases are identical because the "generic sensor set" is the set of sensors that were used to acquire these sequences. For sequences MS04 - MS06 the results obtained for the SKC case are slightly better than those obtained for the GSC in most cases. However, the inverse is true for MS06LW. This is due to the choice of the learning image in the SKC case. The learning image is the first image on which the target appears. In the infrared image, the target happens to be white on a light grey background (clouds) and is very difficult to see. Because, in a part of the sequence, its background becomes a clear sky (dark), the weights are no longer appropriate and performance drops.

Sequence	Sensor	GSC		SKC	
		Pd	Nft	Pd	Nft
MS01	LW	85	10	85	10
	TV	32	4	32	4
MS02	LW	0	2	0	2
	TV	0.5	1.4	0.5	1.4
MS03	LW	84	11	84	11
	TV	21	16	21	16
MS04	LW	97	16	***98***	***13***
	SW	43	2	63	6
	TV	98	14	***93***	***13***
MS05	LW	81	12	***99***	***14***
	SW	0	1.16	9	4
	TV	51	14	***98***	***11***
MS06	LW	50	27	31	7
	RD	89	6	94	2.5
	GR	***96***	***0.4***	73	2.4
	BL	90	0.25	82	0.57

Table 3: Results for single sensors

In MS02 targets disappear in the large number of false targets caused by noise. They are only sporadically detected by the first part of the algorithm and rejected by the clutter rejection stage.

Results after sensor fusion (M2)

The following table presents the results after the sensor fusion step.

Sequence	GSC		SKC	
Sequence	Pd	Nft	Pd	Nft
MS01	83	3	83	3
MS02	20	5	20	5
MS03	16	16	16	16
MS04	95	2	***98***	***0.07***
MS05	94	0.2	***96***	***0.15***
MS06	23	7	47	0.8

Table 4: Results after sensor fusion

Results using superset of features (M3)

Sequence	Pd	Nft
MS01	**88**	10
MS02	0.5	1.7
MS03	**93**	31
MS04	0.4	12
MS05	0	20
MS06	40	6

Table 5: Results of superset of features

Results of MTD Module

For the moving target detection a threshold is defined as the number of subsequent images in which the target is detected. The maximal number of subsequent images is 9.
The following tables present the results as a function of the threshold T. Moving targets were only found in sequences MS01-MS03 and in MS06. For MS06 results for the three visual components were very similar, therefore only the red component is shown.

T	MS01				MS02			
	LW		TV		LW		TV	
	Pd	Nft	Pd	Nft	Pd	Nft	Pd	Nft
1	***88***	***0.06***	48	0.87	30	1.6	39	0.48
2	82	0.03	47	0.7	25	0.37	36	0.29
3	63	0.01	45	0.5	19	0.16	32	0.19
4	53	0	44	0.15	12	0.06	23	0.14
5	52	0	44	0.02	9	0.03	17	0.10
6	52	0	42	0.02	4	0.02	9	0.06
7	51	0	41	0	1	0.01	4	0.03
8	0	0	0	0	0	0	0	0

Table 6: Results of MTD for sequences MS01 and MS02

T	MS03				MS06			
	LW		TV		LW		RD	
	Pd	Nft	Pd	Nft	Pd	Nft	Pd	Nft
1	77	0.02	55	0.03	7	1.1	81	0.19
2	74	0.02	44	0.02	4	0.62	81	0.15
3	***73***	***0.01***	38	0.02	2	0.26	***80***	***0.15***
4	73	0.01	24	0.02	0	0.12	75	0.11
5	71	0.01	6	0.01	0	0.04	65	0.09
6	66	0.01	0	0.01	0	0	51	0.06
7	58	0.01	0	0.01	0	0	32	0.04
8	9	0	0	0	0	0	4	0

Table 7: Results of MTD for sequences MS03 and MS06

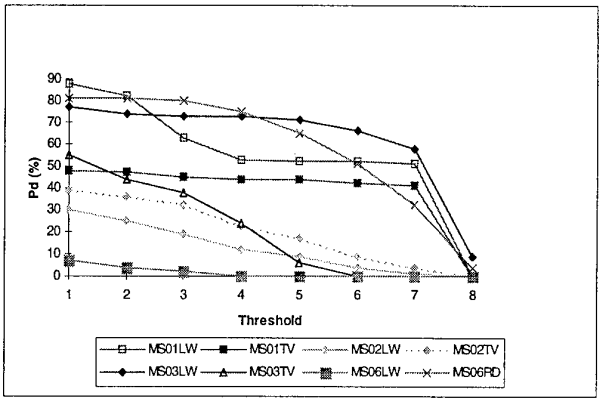


Figure 7: Probability of detection for MTD vs. threshold

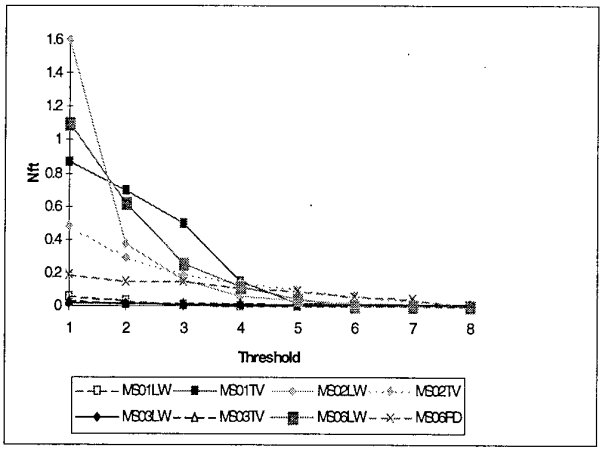


Figure 8: Average number of False Targets per image for MTD vs. threshold

In Figure 7 and Figure 8 it can be noted that by setting the threshold to 3 the number of false targets is greatly reduced while the probability of detection is hardly affected. Therefore we will use the threshold 3 in the decision fusion.
Another remark is that in sequences where moving targets are detected, the MTD module outperforms the TDSI module (cf. Table 3, Table 6 and Table 7)

Results after decision fusion

In sequences corresponding to scenario 1 (surveillance) the motion detector is the primary expert. In the decision fusion we should therefore attribute the highest weight to it.

In the reconnaissance scenario, the motion detector is equivalent to the target detector in single images and both experts should have the same weight.
The "K out of N" decision fusion only accepts a target if the number of experts $N_{experts}$ that have detected it is above a certain threshold T. Special cases for this threshold are:

- AND: $T = N_{experts}$
- OR: $T = 1$
- Majority voting: $T = \frac{N_{experts}}{2} + 1$

The following tables present the results for both scenarios at different thresholds. In Table 8 the results for sequences MS01-MS03 are shown after fusion of the outcomes for each individual sensor(mode M1). In Table 9 the results of fusing the sensor-fusion results with the motion data are shown for the same three sequences (M2). Please note that sequences MS01-MS3 were acquired using two sensors, the number of experts is 2 for the TDSI module (in mode M1) and 2 for the MTD module. In the decision fusion both parts of the algorithm have the same weight. In mode M2 only one expert is available for the TDSI module. However, it is counted twice to ensure that both parts of the algorithm have the same weight in the final decision.
Table 10 shows the results of the decision fusion for MS01-MS03 using the super-set of features (mode M3).
Table 11 shows the results of the decision fusion for sequence MS06 using the outcomes of the TDSI module applied to individual sensors (mode M1) and the results using the outcomes of the sensor fusion step. For both modes, the results are given for the "generic sensor case" and the "sensor kind case". Please note that, as in sequence MS06, we have 4 sensors (LW,RD,GR and BL), there are 8 independent experts.
In sequences MS04 and MS05 all targets are stationary and the MTD module does not report any targets. Therefore the decision fusion step is not necessary.

T	MS01		MS02		MS03	
	Pd	Nft	Pd	Nft	Pd	Nft
1	96	20	36	2	***85***	***11***
2	66	3	16	0.7	73	1
3	54	0.86	3	0.24	39	0.5
4	30	0.27	3	0.18	0	0.01

Table 8: MS01-MS03: Results of decision fusion using outcomes from individual sensors (mode M1)

T	MS01		MS02		MS03	
	Pd	Nft	Pd	Nft	Pd	Nft
1	65	3	***54***	***5***	79	20
2	49	0.9	17	0.4	63	0.8
3	3	0.11	3	0.1	19	0.18
4	0	0.01	0.5	0.01	6	0.01

Table 9: MS01-MS03: Results of decision fusion using outcomes from sensor fusion step (mode M2)

T	MS01		MS02		MS03	
	Pd	Nft	Pd	Nft	Pd	Nft
1	***95***	***11***	36	2	90	28
2	58	0.9	16	0.46	67	1.98
3	36	0.31	3	0.09	43	0.53
4	3	0.05	0.5	0.01	7	0.05

Table 10: MS01-MS03: Results of decision fusion using outcomes of the super set of features (mode M3)

T	MS06 (GSC)		MS06 (SKC)		MS06SF (GSC)		MS06SF (SKC)		MS06 (SSF)	
	Pd	Nft	Pd	Nft	Pd	Nft	Pd	Nft	Pd	Nft
1	98	34	98	13	98	0.79	***98***	***0.7***	84	6.5
2	98	3	94	2.5	69	0.59	64	0.3	55	1
3	86	1.4	73	1.1	29	0.41	19	0.3	6	0.2
4	58	0.7	32	0.7	0.9	0.02	0.9	0.1	4	5.9

Table 11: MS06: Results of decision fusion using the GSC and SKC for individual sensors and after sensor fusion and using the super set of features (SSF).

Discussion

For the discussion, the fusion method that gives the best results will be identified for each sequence separately. Then these "best results" will be analysed as a function of some of the properties of the sequences. The notion of best results depends on the application. For some applications the probability of detection is the critical factor while for others the ratio between probability of detection and number of false targets has to be maximised. To identify the "best fusion method" for each sequence, we have chosen the latter criterion (Pd/Nft). In the previous tables, these best result are highlighted with italic letters on a grey background. Please note that, for sequences where moving targets are present, the MTD algorithm used on a single sensor sometimes gives a higher ratio Pd/Nft than any of the fusion results. In all cases fusion will however increase the probability of detection.

Best results per sequence

MS01:
For MS01 the best results are obtained after decision fusion of the results obtained with the super-set of features (TDSI-Mode M3) and the results from the moving target detection.

MS02:
For MS02, the results of the TDSI module for single sensors (mode M1) and for the super-features (mode M3) are very poor. The targets are completely lost in the noise and rejected by the clutter rejection algorithm. For mode M2, results are slightly better. This is because the targets are enhanced by multiplying the feature-level-fused images while noise is tampered. The moving target detection gives better results. In the decision fusion, the best results are obtained by fusing TDSI-mode M2 with the MTD results.

MS03:
For this sequence the best results are obtained by the decision fusion of the results obtained by the TDSI module for single sensors and the MTD outcomes.

MS04 and MS05:
The MTD module does not report any targets. The best results are obtained by mode M2 of the TDSI module.

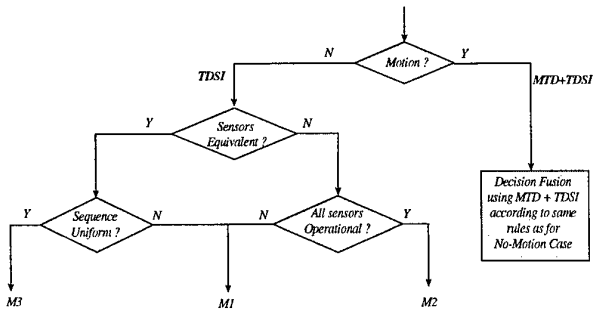
MS06:
The LW and TV sensors compete for the detection of the targets. The SW gives less good results. The best results are obtained after the decision fusion of the sensor fused data from the SKC case. It is interesting to note that the difference between the results of the SKC and the GSC case is very small.

Analysis of "best results":

We will now try to explain why a given fusion method gives the best results for each sequence. We will try to correlate subjective notions of image quality and sequence uniformity to the results.
For MS04 and MS05 the MTD doesn't find any targets, therefore it makes no sense to perform the decision fusion of TDSI results with MTD results. For these two sequences the LW and TV sensors give images that have a similar quality while the SW sensor gives a much lower contrast.
For MS01 both sensors provide images that have a similar quality and the sequence is uniform (targets have the same contrast with their surroundings throughout the sequence). For MS02 the images obtained from the LW sensor are less good than those obtained from the TV sensor.
For MS06 this is also the case. For MS03 the target is hardly visible in the beginning of the sequence while its contrast gradually increases.
We can make the following statements:

- If the sequence is not uniform, using M3 gives bad results because the test image is not characteristic for the rest of the database and using the super-set features causes features that allow detection for some sensors to be discarded.
- If the sensors are equivalent and the sequence is uniform, M3 gives good results.
- If one of the sensors provides images of lesser quality, both using the super-set of features (M3) and the single sensors (M1) give rise to too many false targets (due to noise). The sensor fusion step (M2) however reduces the potential false targets before the actual target detection.
- Because a simple multiplication is chosen for the sensor fusion step, the method based on sensor fusion (M2) is not applicable if one of the sensors is not operational.

The following figure presents these findings graphically.



Although these results seem logical, due to the fact that we only have 6 sequences to test this and because we did not use any objective measurements of image quality or sequence uniformity, this can not be generalised.

8. CONCLUSIONS

In this paper the use of data fusion at several levels is explored. The method was tested on a database of 6 multi-spectral image sequences. The approach consists of two main parts. The first part detects targets in single images (TDSI module) while the second part tries to detect moving targets (MTD module). The motion detection is performed for each sensor separately.
The TDSI module is based on texture features. Several texture features are calculated in each point of the images. These features are combined into a new image using feature fusion.
For the TDSI module three modes of operation are identified. In Mode M1 the detection is performed for each sensor separately. Mode M2 performs sensor fusion by combining the images obtained by feature-level-fusion from each sensor. Mode M3 determines the feature-level-fused images, using features from all sensors (super-features).
Decision level fusion is used to combine the results of the two parts of the algorithm.
Results show that the MTD module is very efficient when moving targets are present. In the TDSI module, the different sensors appear to be quite complementary: in some sequences the infrared sensors give the best results while in others, the visual sensor outperforms the infrared sensor.
The results show that the type of fusion that gives the best performance varies greatly from sequence to sequence. The performance is influenced by the presence of noise (in MS02), the presence of a less performant sensor (in MS02, MS04 and MS05), the type of scenario and the uniformity of the background. Although we do not pretend that there is a single optimal fusion paradigm that can solve all possible problems in all possible cases, for our test sequences, we did find a way to chose the optimal fusion method among the methods we tested, based on the criteria that are given above.

This paper has presented some of the advantages and/or disadvantages of using data fusion on different levels.
As a final conclusion one could state that different kinds of data fusion have different advantages and disadvantages and are therefore suited for solving different kinds of problems.

REFERENCES

[1] R. T. Antony, *Principles of Data Fusion Automation*: Artech House, 1995.
[2] D. L. Hall, *Mathematical Techniques in Multisensor Data Fusion*: Artech House, 1992.
[3] B. V. Dasarathy, *Decision Fusion*: IEEE Computer Society Press, 1994.
[4] L. A. Klein, "Sensor and Data Fusion Concepts and Applications," *SPIE Optical Engineering*, 1993.
[5] E. Waltz and K. Llinas, *Multisensor Data Fusion*: Artech House, 1990.
[6] K. Schutte, "Fusion of IR and Visual Images," FEL-TNO, The Hague, Research Report FEL-97-B046, 4 Feb 1997 1997.

[7] L. Kuntz-Sliwa, "Optimisation d'une Configuration Multicapteurs donnée - Fusion Pixel," in *Signal and Image dept.* Toulouse: Institut National Polytechnique de Toulouse, 1996, pp. 105.

[8] D. Hosmer and S. Lemeshow, *Applied Logistic Regression*: John Wiley & Sons, 1989.

[9] P. Verlinde, "Numerical Evaluation of the efficiency of a camouflage system in the thermal infrared," in *ESAT*. Leuven: K.U.L., 1989.

[10] P. Verlinde and M. Proesmans, "Global approach towards the evaluation of thermal infrared countermeasures," presented at Characterization, Propagation and Simulation of Source and Backgrounds, Orlando, USA, 1991.

[11] R. Haralick, "Statistical and structural approaches to texture," *IEEE Proc.*, vol. 67, pp. 786-804, 1979.

[12] C. Perneel, M. d. Mathelin, and M. Acheroy, "Detection of important directions on thermal infrared images with applications to target recognition.," presented at Forward Looking Infrared Image Processing, Orlando, USA, 1993.

[13] B. Borghys, P. Verlinde, C. Perneel, and M. Acheroy, "Long range target detection in a cluttered environment using multi-sensor image sequences.," presented at Signal Processing, Sensor Fusion and Target Recognition, Orlando, USA, 1997.

[14] Y. Bar-Shalom, *Multitarget-Multisensor Tracking: Advanced Applications*. Norwood MA: Artech House, 1990.

[15] F. G. J. Absil, "Implementation of a set of Tracking Algorithms," Koninklijke Militaire Academie, Amsterdam, Research report 94-21, 94.

[16] R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 29, pp. 1147-1152, 1981.

[17] R. Morris, "Image Sequence Restoration using Gibbs Distributions," in *Dept. Of Engineering*. Cambridge: Trinity College, 1995.

Two-frequency Millimeter Wave SAR

H. Schimpf, H. Essen

FGAN-Forschungsinstitut für Hochfrequenzphysik
Neuenahrer Str. 20
D-53343 Wachtberg-Werthhoven, Germany

Abstract

(NU) The FGAN–FHP MEMPHIS radar transmits simultaneously at 35 GHz and 94 GHz. Accordingly, when flown in a side-looking configuration, closely related SAR images can be generated for both frequencies. These two images can be combined for the purpose of target detection and discrimination.

(NU) It is analyzed how this combination can be performed, and how much one gains with respect to speckle reduction, target-to-background contrast, detection probability and false alarm rate. For this purpose, a simple algorithm is considered that consists of a threshold based prescreener and a subsequent feature based discrimination stage.

(NU) For multi-look images that are mapped to the ground it is found that the combination of two different frequencies with like polarization provides better speckle reduction and target-to-background enhancement than the combination of two orthogonal polarization channels at the same frequency. The detection/discrimination results depend on how precisely the two images are registered.

1. Introduction

(NU) During the last years, radar imaging has gained more and more importance for military applications. In multi-sensor drones for surveillance and reconnaissance applications as well as in dual mode seekers where radar signals have to be matched to IR images, it is necessary to process the returns of the radar sensor into images. Only on the basis of images extended high value targets like bridges, industrial complexes, bunkers, command posts or others can be detected and analyzed and later on fused with the corresponding output of the IR seeker, or matched with pregiven patterns. Also for smaller relocatable targets the improvement in resolution leads away from the former point target treatment to imaging methods for the purpose of detection and discrimination.

(NU) The MEMPHIS radar, used by FGAN–FHP for airborne measurements in a side-looking SAR configuration, is operated simultaneously at 35 GHz and 94 GHz. Both frequencies are derived from a common system reference. All system parameters like pulse shape and duration, PRF, transmit and receive polarization, phase code for high range resolution etc. are identical. The two antennas are boresighted to the the best possible precision. A detailed description can be found in [13].

(NU) With this interesting data at hand, the question arose whether by combining information from these two physically completely independent frequencies, an improvement in image quality, target/background contrast and detection algorithm performance can be achieved. As a basis for the analysis, a scene is chosen containing five different relocatable targets (missile launchers, air defence units) and several vehicles of opportunity (trucks, vans etc.). These targets are located in a grassy area with scattered isolated trees and bushes, surrounded by a wooded area with roads and several buildings. The detection algorithm is designed to discriminate the relocatable targets from the natural background and, if possible, also from the buildings.

(NU) Section 2 gives a description of the data that are used for the analysis, and of various ways how different SAR images can be registered to each other. In section 3, the effect of combining different images on speckle reduction and target/background contrast enhancement is described in detail. The influence of image combination on detection and discrimination algorithm performance is analyzed in section 4. The detection results are analyzed and summarized, leading to the conclusions in section 5.

2. Registration of images

The basic element of every SAR image is the local range-Doppler map (RDM), i.e. the outcome of the Doppler Fourier transform. This RDM is projected onto a ground map with predefined pixel spacing, thereby making use of all motion and aircraft attitude information. Consecutive RDMs may cause several returns to be projected to the same pixel

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

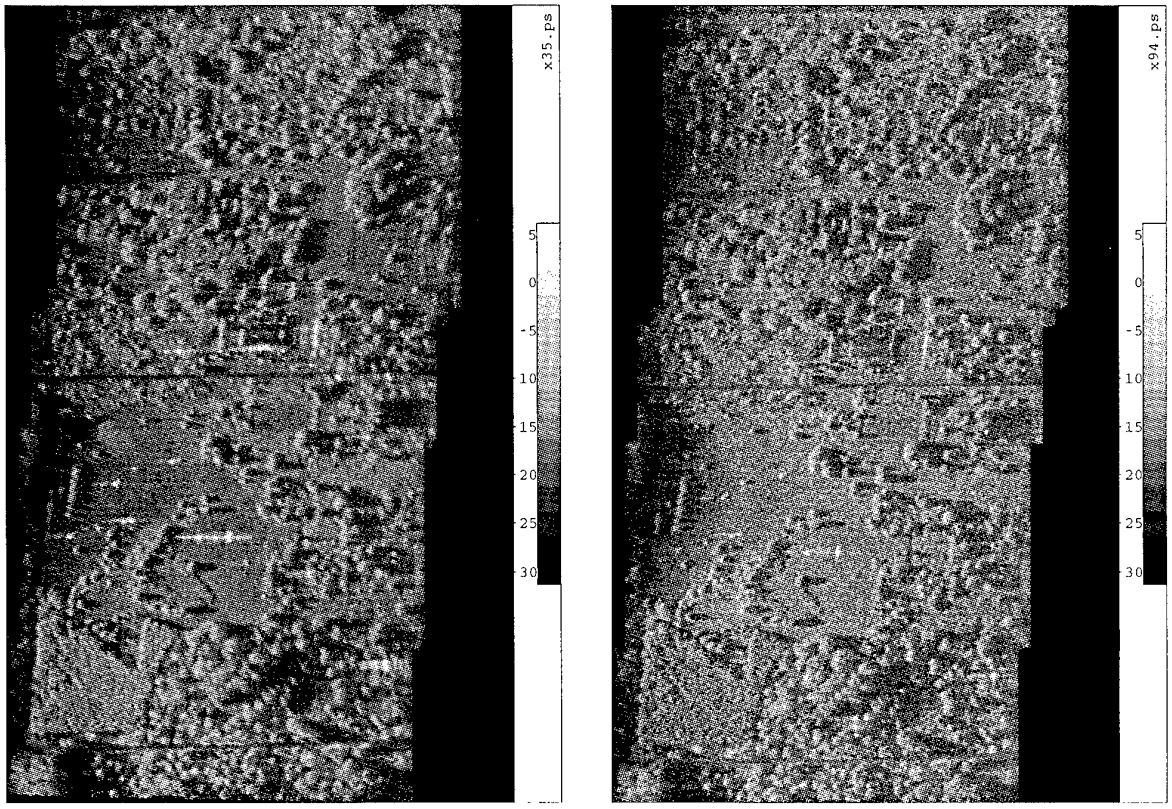


Fig. 1. SAR image of relocatable target area at 35 GHz (left) and 94 GHz, polarization LR

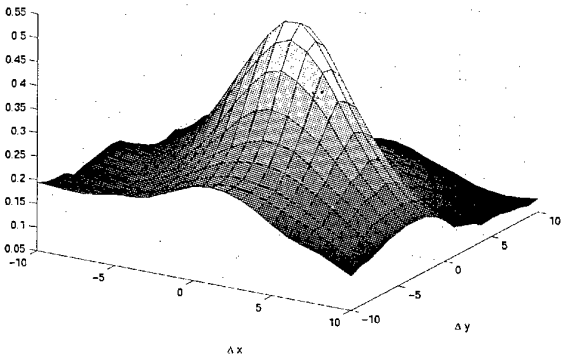


Fig. 2. Cross correlation coefficient between 35 GHz and 94 Ghz scene

on the ground. If one eliminates this overlap the final product is a single-look image which can still be complex, i.e. contain phase information. If one averages the signals within each pixel the result is a multi-look image which normally contains only intensity (one way to retain a certain phase information is Stokes averaging [2]).

(NU) Therefore, before combining two different images, one has to make two decisions. First, does one want to work with local RDMs or with ground maps? Second, does the resulting image have to contain (multichannel) amplitude plus phase or amplitude only?

(NU) When working with local RDMs one has to take into account the properties of the Doppler Fourier transform (DFT or, most often, FFT). One can use any aperture time 'T' up to the maximum possible to get an according Doppler resolution

$$\Delta = \frac{\lambda}{2\beta} = \frac{\lambda R}{2vT}$$

(v=flight velocity, R=slant range, λ=wave length, β=change in aspect angle). If one processes 'N' consecutive pulses then the resulting unambiguous range 'R_U' of the FFT is given by

$$R_U = N\Delta = \frac{\lambda R}{2v(T/N)} = \frac{\lambda R}{2v\Delta T} = \frac{\lambda R}{2v} \cdot PRF$$

If one wants to obtain equal Doppler resolution one has to fulfill the condition λ/T=const., hence the processing times 'T' must have the ratio $94/35 = 2.687 \approx 8/3$. As both radars are operated with the same PRF this means that the DFT lengths are different by this ratio, and this again means that

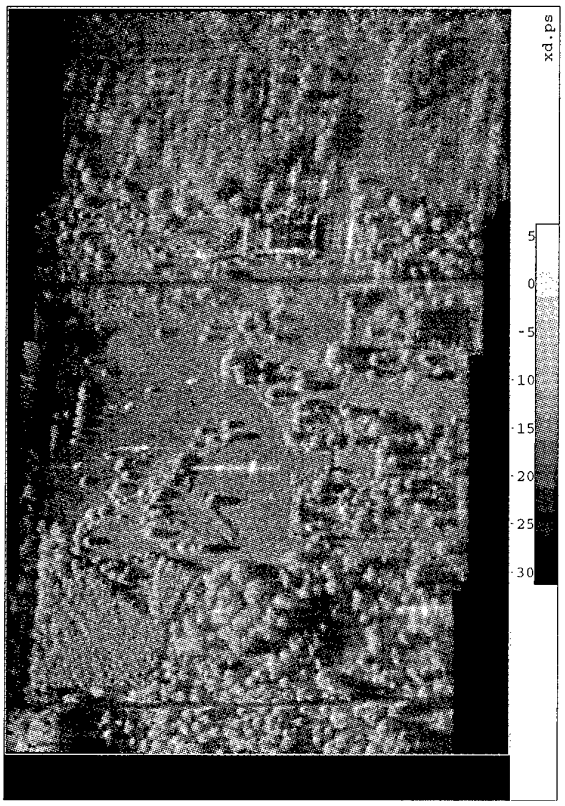


Fig. 3. combined SAR image 35+94 GHz

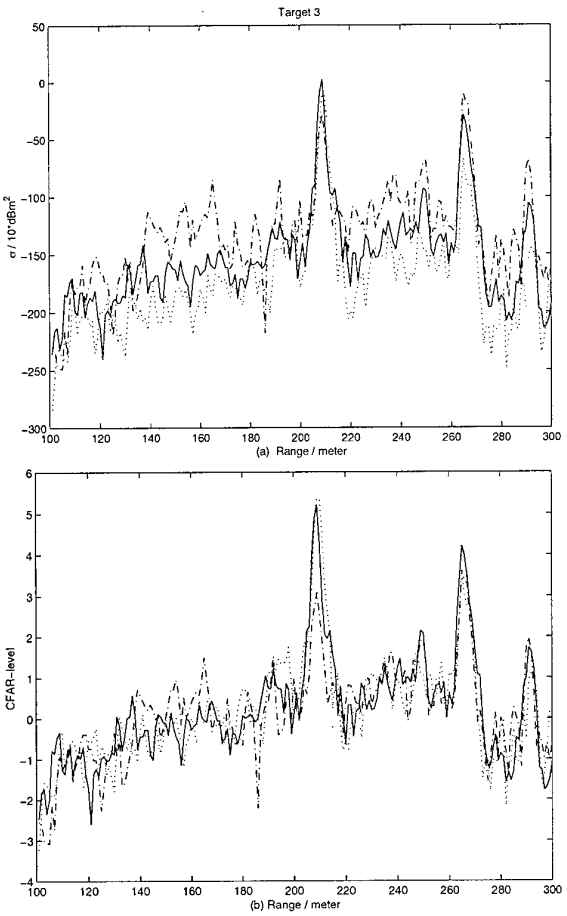


Fig. 4. Amplitude and CFAR profile for 35 GHz (dotted), 94 GHz (dashdot) and 35+94 GHz (solid)

one can use an FFT only for one case, but a general DFT for the other. This can be very processing time consuming and therefore is impractical.

(NU) The most natural thing would be to keep $\lambda \cdot PRF$=const., which would mean equal unambiguous range. In addition, one can obtain equal Doppler resolution by choosing the same FFT length which also facilitates the processing. On the other hand, as the aperture lengths and times would differ by the factor 94/35, this means that multi-look images cannot have the same degree of integration for both frequencies. However, as the MEMPHIS radar has one and the same PRF for 35 GHz and 94 GHz, the only way to fulfill the above condition would be to thin down the original PRF by using e.g. each 8th pulse for 35 GHz and each third pulse for 94 GHz. This is not realistic, of course, because one would throw away valuable data and, above all, the effective PRF would become so low that the unambiguous range is smaller than the beamwidth between nulls.

(NU) The simplest way to proceed, therefore, is to perform adequate FFTs for each frequency independently and match the resolution cell sizes by means of interpolation. In the case of equal FFT lengths, for example, 3 Doppler cells at 35 GHz would have approximately the same extent as 8 cells at 94 GHz. Consequently, 8 cells with values $b_1 \ldots b_8$ would be constructed out of the original 3 cells with values a_1, a_2 and a_3, using

$$
\begin{aligned}
b_1 &= b_2 = a_1 \\
b_3 &= \frac{2}{3}a_1 + \frac{1}{3}a_2 \\
b_4 &= b_5 = a_2 \\
b_6 &= \frac{1}{3}a_2 + \frac{2}{3}a_3 \\
b_7 &= b_8 = a_3
\end{aligned}
$$

either in real (intensity) or in complex space. The resulting interpolated 35 GHz image would then be matched to the 94 GHz image with the same resolution and the same number of Doppler cells (see below).

(NU) The methods described so far are adequate for any detection algorithm that works on individual RDMs or on small ensembles of consecutive RDMs. If, on the other hand, one works with radar maps of a somewhat larger extent that were created

by means of motion compensation and projection on a ground map, then the case is much easier as the two radars are mounted on the same platform. Here, the result of the processing are two maps with the same resolution and without (or with identical) geometrical distortions.

(NU) As there will always be a certain boresighting error between the 35 GHz antenna and the 94 Ghz antenna, the next step for both RDM or ground map is the matching of the two contributions by means of correlation. For this purpose it is usually sufficient to select only a part of the complete map, containing a number of dominant scatterers, to reduce the computational effort.

(NU) The following analysis was done with multilook ground maps consisting of two polarization channels. The measurements used for this purpose were performed with circular polarization. A detailed description can be found in [1] and [10]. Using circular transmit polarization leads to a natural splitting of the receive signal into one part that underwent an odd number of reflections (flat plate or trihedral corner), and another part that underwent an even number of reflections (dihedral type). Therefore, we use the terms 'odd' (LR, or LHC transmit and LHC receive) and 'even' (LL) channel in the following.

(NU) A polyphase coded wave form was used with a bandwidth of 100 MHz resulting in a high range resolution (HRR) of about 1.5 meter, the sampling interval was 5 ns providing a HRR cell spacing of 0.75m . The nominal depression angle was 17.5°, at a slant range of 800 m. The 3-dB two-way beamwidths of the lens antennas were 1° (2.7°) in azimuth and 12° (15°) in elevation for 94 GHz (35 GHz). The Doppler processing was performed in such a way that the along-track resolution was 1.0 meter. The resolution cells were projected on a ground map with a grid size of 1 m by 1 m, resulting in a number of 5 to 6 complex data values per pixel ([14], [15]).

(NU) The scene (cf. fig.1) is about 800 m long and 500 m wide. For the correlation a region of length 100 m and width 140 m containing 4 of the relocatable targets was chosen. The correlation was performed over the intensity values (in dB) of the LR channel. A relatively sharp maximum was found with an offset of 2 pixels both in 'x' and 'y' direction, the correlation coefficient at the maximum being ρ_{max}=0.523 (cf. fig.2). As the images (multilook two-channel amplitude-only) were given in dB-scale, also the combination of the two data sets was done by taking the mean (in dB) of the two contributions within each pixel. This corresponds to the geometric mean of the return powers. One could as well use the arithmetic mean, but this would have meant to linearize all values first, then average and transform back to dB-space, hence a much more tedious procedure. The final product are two-channel (odd and even) intensity images.

3. Target / background contrast enhancement

Fig.3 shows the result of combining the 35 GHz and the 94 Ghz image. When comparing this to the original images one notices essentially three effects: the speckle appears reduced, i.e. the image looks smoother, a slight loss in focus occurred, and the shadows are more pronounced, i.e. the contrast was enhanced.

(NU) When looking at the smeckle reduction, one sees a general dilemma. For a precise registration of the two images, a high correlation is needed. On the other hand, additional information is gained only, when the images are independent which means that their correlation is low. For the assessment of image smoothing, two subareas were chosen within the image. The first one contains mixed clutter, mainly grass with scattered bushes, the second one is a forest of deciduous trees. As a measure of smoothness we consider the standard deviation of the reflectivity calculated in dB-space:

	channel	mixed clt.	decid. forest
35 GHz	LR	3.77	4.21
	LL	4.00	4.46
	LR+LL	3.54	4.00
94 GHz	LR	3.85	4.41
	LL	3.54	4.12
	LR+LL	3.51	4.11
35+94 GHz	LR	3.29	3.78
	LL	3.25	3.79
	LR+LL	3.14	3.65

Table 1: Reflectivity standard deviations (dBm^2/m^2)

(NU) For comparison, this table contains also values for the case when both orthogonal receive channels are combined (summing the linearized backscatter powers and transforming the result back to dB). One sees that the combination LR+LL results in a smoothing with respect to LR or LL alone, that the combination 35+94 GHz, however, provides a more pronounced smoothing. This is in agreement with the correlation coefficient between the LR and the LL channels which is 0.583 for 35 GHz and 0.696 for 94 GHz, respectively.

(NU) The slight loss in focus is due to the slight boresight error between the two antennas. Indepen-

dent measurements show an azimuth misalignment of 0.4° which means that the crab angle correction must be different for both images. The correlation method when it is applied to the complete map gives the best results of course in the correlation region. Moreover, an optimal registration should allow not only for linear offsets but also for a slight rotation.

(NU) The deepening of the shadows is due to the fact that the pertinent cells contain only noise which ideally has zero correlation. Therefore, the effect of smoothing is strongest in the shadow areas.

(NU) The above results show that there exists a certain tradeoff. Combining two receive channels (amplitude only) of the same image preserves the original focusing but provides less smoothing due to a possible high correlation between the channels (depending on the clutter type). Combining 35 GHz and 94 Ghz, on the other hand, brings together more independent information, but with the possibility to deteriorate the focusing quality. This topic will be analysed further in the next paragraph.

(NU) The target/background contrast can be assessed by looking at range profiles across target positions. Fig.4a shows one example with the target located at a range of about 210 meters. At first sight, the result does not look very promising, although the 35+94GHz profile (solid line) shows less variation in the clutter region in front of and behind the target. The situation looks more favorable if one considers a CFAR-level profile instead of a mere range profile. As a simple example, the CFAR level 'x' for each pixel value 's' is calculated as $x = \frac{s-\mu}{\sigma}$ where μ and σ are the mean and standard deviation, respectively, of the reflectivity across the entire range profile with the 10 strongest values excluded in order to prevent other close-by targets from biasing the reference. This CFAR level makes use of the reduced standard deviation due to smoothing and yields an improvement in contrast (fig.4b).

4. Discrimination algorithm performance

(NU) The radar based discrimination of man-made structures embedded in a natural clutter background consists of several steps. First, the radar image layout has to be created by determining the pixel (resolution cell) size and performing the projection to a ground map. Next, the desired information contents of each resolution cell has to be chosen and the preprocessing done accordingly (([2]) and Novak et al. ([4], [6], [7])). The third step would be a prescreener ([10], [5], [8], [9]), i. e. some sort of a thresholding procedure which narrows down the number of possible candidate target sites and thus reduces the computational burden on the subsequent processing stages. Based upon the output of the prescreener, feature vectors are then constructed that help to discriminate between man-made targets and natural background.

(NU) For the prescreener, only single range profiles or a small number of range profiles must be used. The discriminator, too, should act on small areas that are adapted to the size of the target that one wants to find. This takes into account the fact, that in realistic systems, the incoming radar backscatter signals form a time sequence. That means that each detection algorithm has to start its work while the radar image is still being built up. The final product of the overall radar processing is an image that contains within each pixel the decision result "target yes/no". This image can subsequently be used, probably in connection with an original radar amplitude image, for further image processing techniques like pattern matching, comparison with photographical or IR images etc. ([3],[16]).

4.1 Prescreener

(NU) The prescreener used for this analysis might be called a shadow filter (SF). It acts on total power HRR profiles. The test variable T(n) for the n-th pixel within a HRR line is computed as:

$$T(n) = \sum_{i=n-4}^{n} (p_i - \bar{p}) - \sum_{i=n+1}^{n+5} (p_i - \bar{p})$$

(NU) Here $\bar{p}$ = total power average over one HRR line (500 to 600 pixels). The first term is positiv, if strong scatterers are present. The second term is negativ, if a shadow is present. Therefore, T(n) has a maximum for a strongly reflecting structure casting a shadow. The SF is adapted to the expected electrical target size of 5 m. T(n) has to fulfill the threshold condition $T(n) \geq 0.3\,m^2$ which was determined empirically. Of course, this SF does not only show small structures of the desired size. But if there is a large structure casting a shadow, then only its trailing edge passes the prescreener which may reduce considerably the amount of pixels that are left for subsequent discrimination.

4.2 Discrimination

(NU) For discrimination between man-made structures and natural clutter background, two features were used that were analyzed earlier ([10]). Only those cells that had passed the prescreener were subjected to the feature test. The features are:

F_1= slope of straight line fit to total power (8 strongest cells arranged in decreasing order)

F_2= $\frac{Std.dev.}{mean}(odd + even\,power)$

The features are computed over areas of 7 by 7 pixels centered at the cell under test.

(NU) For the feature test, one needs references for clutter and for vehicles, respectively. The clutter reference, consisting of 1332 pixels, was a mixed scene containing grass, scattered bushes and trees with their shadows, taken from former measurements in a different region ([10]). Thus, one avoids the use of one and the same scene for training and for testing.

(NU) For the target reference, data were taken from tower/turntable measurements performed with a similar radar (CORA 94) under the same depression angle. These are described in more detail in [11] and [12]. The data belong to a main battle tank covered by a camouflage net. This net reduced the overall cross section by less than 1 dB, but changed the geometry and signature of the scattering centers. This target reference is by no means optimized with respect to the specific relocatable targets that are in the scene. Rather it serves as a generic example for a "man-made object" reference for discrimination against a natural background.

(NU) The references are stored in the form of mean value and standard deviation of the respective feature, based on the feature distributions for the reference templates.

(NU) Next, one computes as described in [11] and [10] the distances of the feature vector under test to the target reference and to the clutter reference. These are the normalized Euclidian distances in feature space:

$$D_T = \sqrt{\sum \frac{(F_i - F_{Ti})^2}{\sigma_{Ti}^2}}$$

$$D_C = \sqrt{\sum \frac{(F_i - F_{Ci})^2}{\sigma_{Ci}^2}}$$

where F_{Ti} and F_{Ci} are the mean values and σ_{Ti} and σ_{Ci} are the standard deviations of feature F_i from the target and clutter reference distributions. The use of the above distance measure in feature space is justified when there is no correlation between the features. This was verified in the present case.

(NU) Now, any pixel under test is said to belong to the class of "relocatable target" if the condition

$$D_T < \gamma \cdot D_C$$

is fulfilled where the factor $\gamma \leq 1$ has to be chosen so as to minimize the number of false alarms while not affecting the detection of the real target. Here, $\gamma = 1$ was used throughout.

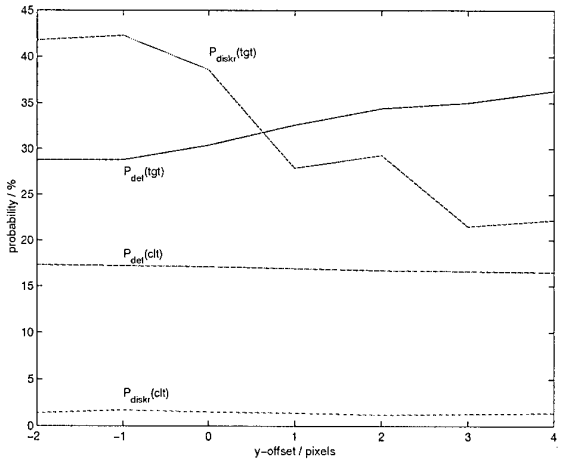


Fig. 5. Detection and discrimination probabilities for targets and clutter as a function of misregistration between 35 GHz and 94 GHz images

5. Results

(NU) The test was performed on the original 35 GHz and 94 GHz scene and for comparison on the combined 35+94 GHz scene with optimal correlation. Afterwards, the two images were offset with respect to each other by up to 6 pixels in the flight direction whereas in range the relative position was unchanged. This was done to study the influence of mis-registration on algorithm performance. For a quantitative assessment boxes were defined around each of the 8 targets. Within each box the number of pixels that passed the prescreener (NDET) were divided by the total number of pixels in this box (NPIX) to yield the detection probability P_{det}. Next, the discrimination probability is defined as P_{discr}=NDIS/NDET, where NDIS is the number of pixels that passed the prescreener and the feature test. Table 2 summarizes the P_{discr} values.

(NU) As can be seen from this table, the normal case seems to be that P_{discr} for 35+94 GHz lies between the values for the single frequencies (targets 1, 3, 4, 6 and 7). P_{discr} is lower than before in the case of targets 2 and 5 (where #5 even was lost completely), and only in one case (target 8) P_{discr} is higher than before.

(NU) If one looks at the effect of mis-registration, the behaviour is not uniform either. For half of

				pixel offset					
Tgt	35 GHz	94 GHz	35+94 GHz	1	2	3	4	5	6
1	82.9	62.1	78.8	44.4	93.8	46.7	4.8	–	–
2	80.0	70.0	64.4	64.0	34.6	2.0	–	–	–
3	38.1	91.8	65.4	28.6	–	2.0	–	–	–
4	43.8	84.1	45.3	56.9	57.9	66.1	43.9	15.2	34.4
5	7.0	41.1	–	18.0	–	7.4	–	1.5	–
6	49.0	73.0	60.0	44.9	75.0	53.2	59.7	75.0	77.8
7	72.3	49.0	60.5	50.4	61.4	44.6	65.9	56.3	54.1
8	76.7	66.0	84.8	53.1	5.9	–	69.0	9.4	–
ΣTgts	8	8	7	8	6	7	5	5	3

Table 2: Discrimination Probabilities (in %)

the targets (1,2,3,4) P_{discr} decreases obviously as a function of offset, in the case of targets 1 and 4 there is a slight maximum around an offset of 2 or 3. In one case (target 5) the originally lost target did reappear. Targets 6 and 7 are very little influenced, their P_{discr} remains more or less constant. P_{discr} of target 8 shows strong variability which may be due to 'splitting' (the target appears as two peaks) when the offset is too large.

(NU) If one combines all target detections and discriminations (from all 8 boxes) and likewise all clutter detections and discriminations (from outside the boxes) the overall behaviour as a function of mis-registration can be studied (fig.5). P_{det} for the targets increases because the targets are smeared out and more pixels pass the prescreener. P_{det} for the clutter decreases slightly because due to the smoothing effect less pixels cross the threshold. P_{discr}, on the other hand, decreases for the targets as expected and as seen in table 2. P_{discr} for the clutter remains almost constant. It must be kept in mind, though, that 'clutter' here means everything outside the target boxes. As there are numerous buildings in the area, which are also man-made objects, they are discriminated more or less like real targets and hence constitute true false targets, and not clutter outliers.

6. Conclusions

(NU) The present preliminary analysis is based on multi-look 2-channel amplitude-only images. It has been shown that the correlation between 35 GHz and 94 GHz is sufficient to warrant a precise registration of radar maps. On the other hand the correlation is lower than between the two orthogonal channels of a single frequency and therefore offers more independent information.

(NU) This becomes visible mainly in better smoothing or speckle reduction and also in enhancing the target/background contrast of a CFAR threshold detector. Concerning a discrimination algorithm which consists of a prescreener and a feature test, the advantage of combining two frequencies is not clearly evident in the present example. This may be due to the statistical, and not polarimetric, character of the two features that were used. These features were chosen because former experience has shown ([10]) that the available resolution of only 1 meter in range and cross-range is not sufficient for the optimum use of polarimetry.

(NU) The algorithm performance was studied as a function of mis-registration between the images of the two frequencies. An offset of one to two pixels is not necessarily critical as some of the targets show, but for larger shifts there is a rapid decrease of discrimination probability. On the other hand, there exist cases where P_{discr} is almost independent of mis-registrations up to 6 pixels. Obviously they consist of so many strong pixels that an additional smearing does not change the statistics that form the basis for the two features.

(NU) The analysis shows that a simple combination of amplitude only images does not fully demonstrate any advantages of combining 35 GHz and 94 GHz. In a next step, therefore, the images will be processed with enhanced Doppler resolution (the range resolution is limited due to limited bandwidth) and single-look with the full available polarimetric information (scattering matrix at 35 GHz, Stokes vector at 94 GHz). The combined image will then consist of six channels, and the discrimination features have to be constructed accordingly.

References

[1] Essen, H.; Schimpf, H.: Remote sensing with an airborne 94 GHz synthetic aperture radar. European Conference on Synthetic Aperture Radar, Königswinter, March 1996.

[2] Schimpf, H., Essen, H.: SAR measurements of extended targets at 94 GHz. AGARD Conference on Remote Sensing, Toulouse, April 1996.

[3] Bers, K., Jurkiewicz, K.: Model based Image Analysis for the Detection of High Value Surface Targets. AGARD-SPP Conf. on "Radar Signature Analy-

sis and Imaging of Military Targets", paper #35, Ankara, 7-10 Oct. 1996
[4] Novak, L.M. et al: Optimal Processing of polarimetric SAR imagery. Lincoln Lab. Journal Vol.3,2 (1990)
[5] Novak, L.M. et al: Studies of target detection algorithms that use polarimetric radar data. IEEE AES-25,2, p.150-165, 1989
[6] Novak, L.M. et al: Optimal speckle reduction in polarimetric SAR imagery. IEEE AES-26,2, p.293-305, 1990
[7] Novak, L.M. et al: Optimal polarimetric processing for enhanced target detection. IEEE AES-29,1, p.234-243, 1993
[8] Novak, L.M. et al: Performance of a high-resolution polarimetric SAR automatic target recognition system. Lincoln Lab. Journal Vol.6,1 (1993), p.11-24
[9] Kreithen, D.E. et al: Discriminating targets from clutter. Lincoln Lab. Journal Vol.6,1 (1993), p.25-52
[10] Schimpf, H., Essen, H.: Detection of Extended Targets in MMW Radar Imagery. AGARD-SPP Conf. on "Radar Signature Analysis and Imaging of Military Targets", paper #38, Ankara, 7-10 Oct. 1996
[11] Schimpf, H., Essen, H.: Evaluation of camouflage means on land targets at 94 GHz. AGARD Conf. Proc.542 (Supplement), Palma de Mallorca, Spain, 1993, p.19-1 to 19-10
[12] Essen, H.: Signaturmessungen bei 94 GHz an Panzern in unterschiedlichen Tarnzuständen. FGAN–FHP Techn. Bericht Nr.15-92(#332), Werthhoven, Aug.1992 (Geheim).
[13] Makaruschka, R., Essen, H.: Airborne Dual-Sensor MMW Signatures of Maritime Targets and Sea Clutter. AGARD-SPP Conf. on "Multi-Sensor Systems and Data Fusion...", paper #5, Lisbon, 29 Sept. - 2 Oct. 1997
[14] Schimpf, H., Pühl, M., Biegel, G.: SAR Measurements of Mines and Minefields at 35 GHz and 94 GHz. Proc. of SPIE Vol.3079, p.625, Orlando FL, April 1997
[15] Essen, H., Makaruschka, R., Baars, E.P.: Remote Sensing of Land Scenarios with an airborne 94 GHz SAR. AEROSENSE'96, Conf.2742, p.378, Orlando FL, April 1996
[16] Bers, K., Essem, H., Jäger, K., Schimpf, H.: Multisensor Data Fusion for Automatic Recognition of High Value Targets. AGARD-SPP Conf. on "Multi-Sensor Systems and Data Fusion...", paper #15, Lisbon, 29 Sept. - 2 Oct. 1997

Acknowledgement

This work was sponsored under contract of the German Ministery of Defence. We acknowledge the development of the SAR processing software and the generation of the SAR imagery by Alfred Wahlen.

PAPER No: 8

DISCUSSOR'S NAME: U. Lammers

COMMENT/QUESTION:

You mentioned that the 35 and 94 GHz signals are generated from a common reference. Was this done to look at differential phase information, and did you actually do this?

AUTHOR/PRESENTER'S REPLY:

Phase information is included in the polarimetric description (e.g. Stokes' parameters) of each return signal. These, of course, will be compared between 35GHz and 94GHz . However, this requires an absolutely precise (at least locally) registration of the two images.

SYSTEME MULTI-SENSEURS RADAR ET OPTRONIQUE POUR L'ATTAQUE AU SOL

E. MIGNOT DGA/SPAé
JP. MESTRE - D. LEVAILLANT
Thomson-CS Optronique, BP 55
Rue Guynemer, 78283 Guyancourt, France

Introduction

La variété des derniers conflits où furent engagées les forces alliées a montré les limitations des systèmes d'armes Air-Sol existants (fortes contraintes d'emploi, armements trop spécifiques, portées de tir insuffisantes face aux défenses Sol-Air,...).

La France a de ce fait l'objectif de se doter d'un armement tactique nouveau, complémentaire des missiles de croisière du type SCALP.

Cet armement de conception modulaire désigné AASM (Armement Air-Sol Modulaire), aura des portées supérieures à 15 km et devrait être capable de deux classes de précision métrique et décamétrique.

Avant le tir des armes, il est nécessaire de réaliser un certain nombre d'opérations telles que la détection, la reconnaissance et la localisation des cibles à traiter.

Le RAFALE va disposer de capacités nouvelles pour satisfaire ces fonctions grâce aux modes d'imagerie haute résolution de ses senseurs internes radar et optronique.

La DGA a confié à THOMSON-CSF, qui est responsable du développement de ces modes, l'étude:
- des possibilités du système multi-senseurs radar et optronique,
- du concept d'emploi du système multi-senseurs,
- des traitements nécessaires à l'exploitation temps réel visuelle ou assistée des images.

Ce travail s'appuie sur des simulations et des données réelles.

La présente communication a pour objet de présenter les résultats de l'étude en cours.

Liste des abréviations

AASM: Armement Air-Sol Modulaire
BDA: Battle Dammage Assessment
BFR: Basse Fréquence de Répétition
DDT: Domaine De Tir
DGA: Délégation Générale pour l'Armement
DRI: Détection Reconnaissance Identification
EM: ElectroMagnétique
GPS: Global Positionning System
HFR: Haute Fréquence de Répétition
HR: Haute Résolution
MFR: Moyenne Fréquence de Répétition
MTI: Moving Target Integration
PLC: Programmation Logique sous Contraintes
SACP: Sol-Air Courte Portée
SAR: Synthetic Aperture Radar
SER: Signature Equivalente Radar
TBA: Très Basse Altitude
THR: Très Haute Résolution

1- Aspect opérationnel

1-1 L'analyse technique du besoin opérationnel

Le nouvel armement tactique AASM doit permettre d'engager un objectif défendu par des missiles Sol-Air en restant hors de leurs domaines de tir, la portée de l'arme étant de 15 km à très basse altitude et plus de 50 km à haute altitude.
Parmi les autres exigences spécifiées pour cet armement on peut citer:
- tir de jour et de nuit
- traitement d'une grande variété de cibles: fixes (bâtiments, voies de communication,...), déplaçables (batterie S/A, bateau à quai,...) ou mobiles (véhicule, bateau en mer,...)
- capacité multicible
- tir toutes altitudes
- capacité "tiré et oublié".

Cet armement de conception modulaire devra être capable de deux classes de précision: métrique et décamétrique.
Ces exigences de précision nécessitent de disposer en amont de senseurs capables de détecter, de reconnaître et de localiser précisément les cibles à traiter.
Ces opérations peuvent être réalisées avant la mission à partir des informations fournies par les moyens de reconnaissance satellitaires ou aéroportés ou pendant la mission grâce aux senseurs de bord. Dans la plupart des cas les deux segments contribuent à la réussite de la mission.

Les fonctions techniques demandées aux senseurs sont donc:
- la détection,
- la reconnaissance,
- la localisation.

Les besoins de détection, reconnaissance et localisation pendant la mission sont différents selon l'arme et la cible considérés.

Pour un armement **de précision décamétrique** que l'on supposera à guidage inertie/GPS:

- le besoin de **détection/reconnaissance** est:
  - faible pour les cibles fixes, une reconnaissance de leur contexte est a priori suffisante,
  - fort pour les cibles déplacables pour lesquelles il est nécessaire de confirmer la présence, mais dans certains cas le contexte peut être une aide précieuse (par exemple les avions sont en général sur les aires de stationnement d'une base aérienne).
- le besoin de **localisation** précise est:
  - fort pour une cible fixe dont on ne connaît pas assez précisément les coordonnées absolues à l'issue de la préparation de mission,

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

- fort pour une cible déplaçable dont la prélocalisation sera dans la plupart des cas imprécise

(moyen d'écoute EM, par exemple, pour la localisation de radar sol)

Pour un armement **de précision métrique** qu'on supposera muni d'un senseur de guidage à imagerie:

- le besoin de détection/reconnaissance est:
  - fort ou faible, pour les cibles fixes, selon le concept d'armement considéré (contrôle actif ou pas avant le tir),
  - fort pour les cibles déplaçables pour lesquelles il est nécessaire de confirmer la présence,
  - fort pour les cibles mobiles,
- le besoin de **localisation** précise est:
  - faible ou moyen selon la qualité de la prélocalisation, la nature de la cible et le concept d'armement.

Ces considérations sur les besoins techniques montrent que l'analyse doit être menée au cas par cas. Ceci nous a conduit à considérer un ensemble de scénarios caractéristique des conditions d'emploi du système multi-senseurs pour analyser le besoin technique et les performances du système multi-senseurs.

1-2 La définition de scénarios de référence

La définition des scénarios de référence a été réalisée en trois étapes:

- analyse des cibles d'intérêt et choix de cibles de référence,
- positionnement de ces cibles dans des scénarios permettant de définir les conditions d'emploi du système,
- renseignement des caractéristiques des cibles, des fonds, des conditions météorologiques et climatiques pour le calcul des performances du système multi-senseurs.

Etant donné le grand nombre de cibles d'intérêt (>100), des critères de classement ont été définis:

- le type de cibles: Fixe/Déplaçable/Mobile,
- la dureté: Forte/Faible,
- la forme: Etendue/Ponctuelle

Nous avons ensuite sélectionné des cibles type dans chacune des classes en fonction:

- du cadre de l'étude, limité aux cibles fixes et déplacables,
- du nombre de cibles de la classe,
- de la priorité opérationnelle.

Les cibles sélectionnées ont ensuite été placées dans des scénarios.
Nous avons considéré, pour chaque scénario, un certain nombre d'éléments dimensionnants utiles pour le calcul des performances des senseurs et l'analyse de leur emploi dans la phase d'attaque.
Ces éléments dimensionnants sont:

- les défenses Sol-Air,
- le caractère monocible ou multicible de l'objectif,
- le type de mission,
- le profil d'attaque (BA: 200 ft, MA: 20000 ft, HA: 40000 ft),
- le profil de terrain,
- le type de milieu,
- le climat.

Les catéchistiques principales des cibles et des scénarios retenus sont présentées dans la planche ci-après.

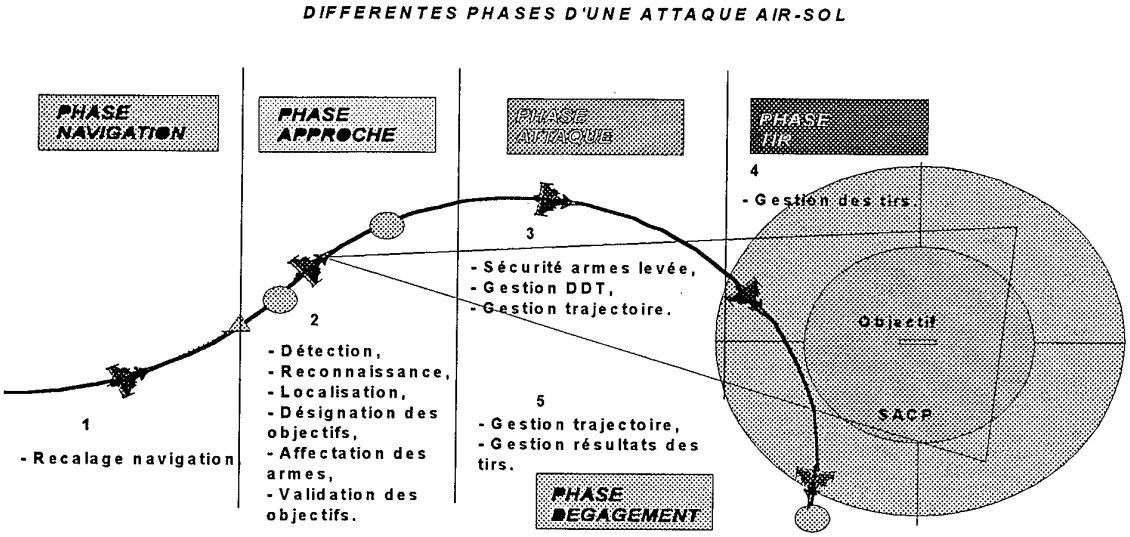
SYNTHESE DES OBJECTIFS ET DE L'ENVIRONNEMENT SOL-AIR DES SCENARIOS

SYNTHESE / SCENARIOS PROPOSES

Objectif		Caractéristiques Mobilité / Dureté / Environnement	Défenses sol- air	Profil	
1-1 Pont / Fleuve		Fixe / Forte / Urbain	SACP	BA-MA-HA	
1-2 Pont suspendu		Fixe / Forte / Rural	Non défendu	BA-MA-HA	
2-1 Bateau à quai		Semi-mobile / Forte / Urbain	SAMP + SACP	HA	
2-2 Bateau en mer		Semi-mobile / Forte / Marin	SACP + SATCP	MA - BA	
3-2 BASE AERIENNE	Piste	Fixe / Faible / Désertique	SACP + SATCP	MA - HA	
	Radar tactique	Fixe / Faible / Végétation			
	Hangarettes	Fixe / Forte / sable & béton			
	Soutes	Fixe / Faible / sable & béton			
	Avion	Semi-mobile / Faible / Béton			
	Véhicule S/A	Semi-mobile / Faible / Terre			
	Shelter	Semi-mobile / Faible / Sable			
	Bunker	Fixe / Forte / Sable & béton			

1-3 Les étapes d'une attaque air-sol

Nous terminons l'analyse de l'aspect opérationnel par une présentation des différentes étapes d'une attaque au sol.



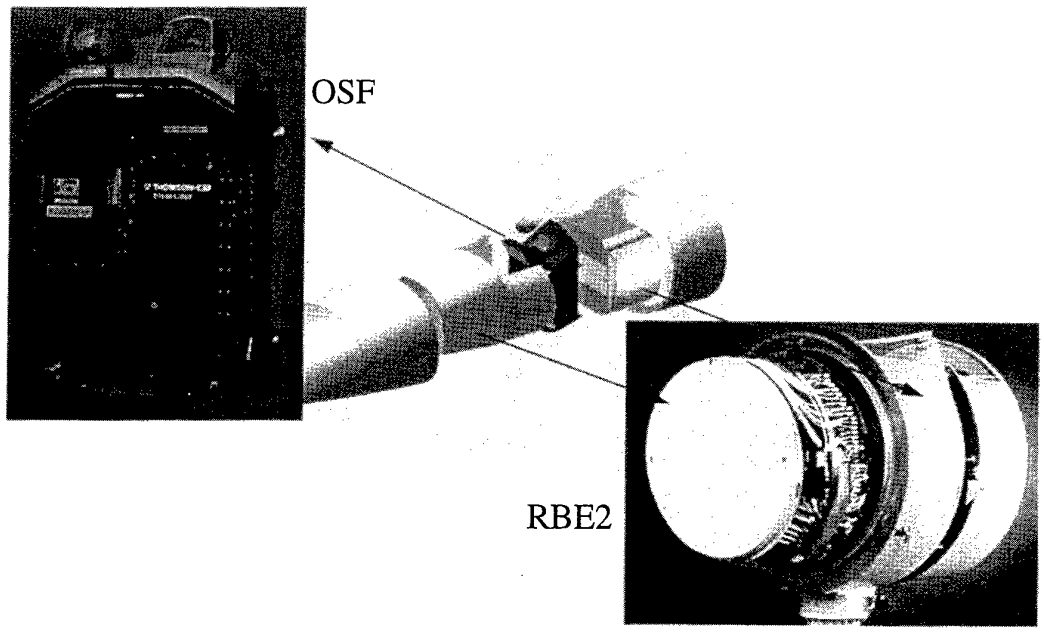
La mise en oeuvre des senseurs pour la détection, la reconnaissance, et la localisation s'effectue dans la phase approche.

2- Les senseurs

Le RAFALE dispose de deux senseurs de bord susceptibles de lui fournir des images:

- le radar RBE2 (Radar à Balayage Electronique 2 plans)
- l'optronique OSF (Optronique Secteur Frontal).

Le RBE2 et l'OSF sont les deux senseurs multifonctions du RAFALE couvrant le secteur frontal de l'avion. Ils sont tous les deux localisés dans la pointe avant de l'avion comme le montre l'image ci-dessous.



Le **RBE2** est un radar en bande X à balayage électronique deux plans, multi-formes d'onde (BFR, MFR et HFR) qui disposent de fonctions:
- air-air d'interception et de combat: recherche et pistage multicible, acquisition automatique et multipoursuite en combat,
- TBA fournissant une carte temps réel du sol pour le Suivi De Terrain tout temps de l'avion,
- air-sol: cartographie faisceau réel et imageries SAR.
- air-mer: détection et poursuite de cibles marines.

Le RBE2 dispose, dans sa fonction Air-Sol, de deux modes d'imagerie SAR:
- un mode HR de résolution décamétrique fournissant l'image d'une zone de terrain pouvant atteindre quelques km de côté,
- un mode THR de résolution métrique couvrant une zone plus petite.

L'**OSF** est un équipement multi-fonctions assurant:
- en air-air interception et combat: recherche, détection et multipistage 2D, mono-poursuite 3D et identification
- en air-sol: mono-poursuite 3D et identification
- en air-mer: détection et multi-pistage 2D, mono-poursuite 3D et identification

L'OSF disposera de deux modes d'imagerie:
- un mode visible avec une résolution de quelques dizaines de µrad,
- un mode IR avec une résolution moins bonne.

3- Les possibilités du système multi-senseurs

Pour juger des possibilités du système multi-senseurs des critères d'évaluation ont été définis. Ces critères sont de deux types:
- caractéristiques (paramètres qui ne dépendent, au premier ordre, que du mode senseur considéré, par exemple la résolution ou le domaine angulaire accessible),
- performances (paramètres qui dépendent du mode et du scénario considéré, par exemple la portée de reconnaissance).

Les définitions de ces critères ont été harmonisées afin d'être en mesure de comparer, sans ambiguïté, les résultats (définition unique et mêmes hypothèses de calcul pour les deux senseurs).

Les performances du système multi-senseurs sont ensuite estimées à l'aide des scénarios de référence présentés dans le §1.

3-1- Listes des critères retenus

Les critères d'évaluation retenus se répartissent en deux groupes:
- les critères quantitatifs
  - portées de détection,
  - portée de reconnaissance,
  - portée d'identification,
  - précision de localisation absolue,
  - précision de localisation relative,
  - taille de la zone imagée
  - temps d'obtention des informations,
  - cadence de rafraîchissement,
  - domaine angulaire accessible,
- les critères qualitatifs
  - sensibilités aux conditions climatiques et météorologiques,
  - utilisation jour/nuit,
  - discrétion,
  - dépendance vis à vis de la préparation de mission,
  - capacité d'évaluation du résultat de tir,
  - robustesse.

3-2- Les caractéristiques images

Les paramètres retenus pour caractériser les différents modes d'imagerie sont les suivants:
- dimension (2D ou 3D) et paramètres primaires correspondants (distance, gisement, site),
- résolution spatiale,
- taille de la zone imagée
- temps d'obtention des informations,
- cadence de rafraîchissement,
- domaine distance,
- domaine angulaire accessible,
- domaine d'emploi porteur.

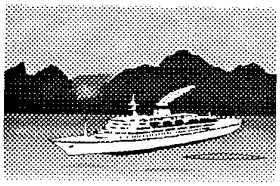
Afin de pouvoir comparer directement les caractéristiques, les paramètres qui décrivent spatialement l'image (taille et résolution) sont présentées dans le repère normal terrestre associé à l'avion en coordonnées sphériques et cartésiennes. Ces caractéristiques sont illustrées par des cas typiques de visée à des distances données différentes.

3-3- Les performances

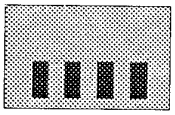
3-3-1 calcul des performances de DRI (Détection, Reconnaissance, Identification).

A- Evaluation des portées de DRI en optronique

La méthode d'évaluation des performances utilise les critères dits « de Johnson »..
La cible est modélisée par un rectangle de surface équivalente à sa surface, dans lequel on place une mire créneau dont la période dépend du critère considéré:
- 1 cycle pour la détection
- 3,5 cycles pour la reconnaissance
- 7 cycles pour l'identification



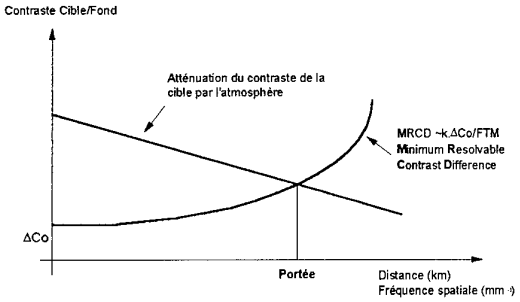
Cible réelle



Modèle fréquentiel de la cible sur le fond
Cas de la reconnaissance

Le contraste de la mire s'exprime en température pour l'infrarouge et en albedo pour le visible. Ce contraste est ensuite calculé après transmission dans l'atmosphère considérée (code LOWTRAN). On en déduit ainsi le contraste apparent, fonction de la distance d'observation.
L'équipement est modélisé par une courbe traduisant son aptitude à discriminer (« résoudre ») un écart de température (cas de l'infrarouge) ou d'albedo (cas du visible) pour une fréquence spatiale donnée. Cette dernière s'exprime directement en divisant la période de la mire par la distance d'observation.

Cette discrimination est proportionnelle au bruit de détection et inversement proportionnelle à la fonction de transfert de modulation (FTM). Cette dernière prend en compte la géométrie du détecteur, la qualité optique, le traitement électronique ainsi que la visualisation et les conditions d'observation par l'opérateur.
L'intersection de ces 2 courbes correspond à une distance pour laquelle la fréquence spatiale permet de rester au-dessus du seuil en température ou en albedo: c'est la portée recherchée.



Cette méthode a été éprouvée depuis 2 décennies par de très nombreux essais sur des cibles réelles et avec des équipements optroniques différant par leur bande spectrale et leur champ d'observation.

Cette méthode de calcul des portées de DRI est particulièrement efficace dans la comparaison d'équipements ou lorsque l'on fait varier des paramètres de nature technique. Elle doit ensuite être pondérée par le contexte opérationnel et le type de cible.

On ne traitera pas de la même manière la reconnaissance d'une base aérienne et celle d'une batterie sol/air. Dans le cas d'un objectif étendu comme une base aérienne, la détection d'un élément caractéristique (la piste par exemple) donnera accès instantanément à l'identification de l'objectif. Dans le cas d'une cible déplaçable dont le type est parfaitement déterminé (avion de combat par exemple) la reconnaissance sera effective, dans certains cas , dès sa détection.

On interprétera encore différemment un calcul dans le cas d'une mission préparée et dans le cas de l'attaque d'une cible d'opportunité. Dans le cas d'une mission préparée, on détectera souvent une cible fixe par la reconnaissance du contexte environnant.

Chaque cible peut donc relever d'une définition particulière conditionnée à la fois par sa nature propre et le type de mission.

Dans le domaine de la reconnaissance aéroportée, un STANAG (le 3769) s'attache a définir, pour chaque cible, ses dimensions ainsi que la dimension du détail permettant d'accéder à la détection, puis à la reconnaissance, etc..... Une analyse au cas par cas de ce STANAG, qui a été défini à l'origine pour les équipements dotés de film photographique (et donc opérant dans le domaine visible), se recoupe très bien avec les critères de Johnson.

B- Evaluation des portées de DRI en radar

La méthode de prédiction de performances utilisée s'appuie sur les travaux réalisés par THOMSON-CSF dans le domaine des SAR satellitaires.
Le principe général de cette méthode est de considérer que la distribution du signal dans l'image de la cible et des fonds est de Rayleigh et que la rétrodiffusion des cibles est principalement fonction de leur orientation.

On considère que la performance obtenue par un opérateur humain entraîné est identique à celle d'un détecteur:
- simple seuil lorsque l'opérateur ne prend pas en compte les détails de l'image de la cible mais effectue uniquement une analyse énergétique,
- double seuil lorsque l'opérateur analyse le contenu de l'image de la cible pour apprécier l'étendue de la tache de rétrodiffusion.

Lorsque l'image de la cible s'étend sur plus de cent pixels l'opérateur combine une analyse énergétique et statistique en préfiltrant l'image pour adapter le nombre de pixels à ses capacités d'analyse. Dans ce cas la distribution des macro-pixels obtenus après filtrage est proche d'une loi normale.

En photo-interprétation, on modélise la performance de visibilité des cibles sur un terrain homogène par une détection double seuil.

On étend ce principe à la reconnaissance et à l'identification en considérant que ces opérations nécessitent la visibilité d'un nombre de pixels supérieur.
On considère que les performances de reconnaissance et d'identification d'une cible peuvent être calculées par le produit:
- d'une probabilité de reconnaissance de l'objet isolé dans des conditions favorables (fond très faiblement rétrodiffusant et bruit thermique très faible),
- d'une probabilité de visibilité d'un nombre minimal de pixels p parmi n composant l'image de l'objet.

Le rapport p/n est de l'ordre de 50% pour la détection de petites cibles, il est supérieur pour reconnaître et identifier.

<u>Résultat en exploitation visuelle et en exploitation assistée</u>
En exploitation visuelle, le résultat est ajusté en fonction des délais d'analyse considérés et de la capacité de concentration des opérateurs concernés; le résultat final étant une interpolation entre les performances simple seuil et double seuil.
Dans le cas d'une exploitation assistée on fait l'hypothèse que le traitement fournit un résultat équivalent à celui d'un détecteur double seuil.

Etapes du calcul des portées de DRI

1- Pd et Pfa globale souhaitée sont fixées (même hypothèse pour le radar et pour l'optronique)
2- recherche de la Pd élémentaire (Pde) fonction de la Pd globale et du rapport p/n (pixels à observer), (calcul itératif)

$$Pd = \sum_{i=p}^{n} C_n^i \bullet Pde^i (1 - Pde)^{n-i}$$

3- recherche de la Pfa élémentaire (Pfae) fonction de la Pfa globale (calculée par simulation)
4- calcul du contraste nécessaire

C = log(Pfae) / log (Pde)

5- calcul du coefficient de rétrodiffusion équivalent au bruit radar $Ne\sigma_0$

$C = (1+ (1-k)\ \sigma_0 / Ne\sigma_0) \bullet (SER/S + Ne\sigma_0) / (\sigma_0 + Ne\sigma_0)$

avec:

- SER: surface équivalente radar de la cible
- S: surface apparente
- k: coefficient de pondération pour prise en compte de l'ombre (0<k<1)

6- calcul de la portée en utilisant l'équation du radar.

La suite des travaux, avec en particulier l'acquisition de nouvelles images réelles ou simulées, permettra d'ajuster cette méthode et de renseigner les paramètres (valeurs de SER, capacité de discrimination,...).

3-3-2 Calcul des performances de localisation

Le calcul des performances de localisation est réalisé à l'aide d'une méthode statistique basée sur une décomposition en erreurs primaires.
Les principales erreurs primaires considérées sont:

- la localisation porteur (erreurs en position, vitesse et angle)
- l'harmonisation des senseurs,
- les erreurs de formation de l'image dues au relief, à la direction de pointage,...
- les erreurs de désignation de la cible fonction de la résolution de mode.

3-4- Commentaires sur les caractéristiques et performances

Les premiers résultats obtenus permettent de dégager un certain nombre de tendances qui devront être confirmées par la suite des travaux.

L'examen des caractéristiques et performances des modes montrent de nombreux caractères de complémentarité:

1- les images de la scène fournies par radar et optronique correspondent à des plans différents:

- le radar donne une image plan horizontal: distance-gisement
- l'optronique une image orthogonale à l'axe d'observation, donc dans le cas des sites faibles une image plan vertical: gisement-site.

La fusion des deux images permet de reconstruire l'image 3D: distance-gisement-site. Il faut noter toutefois qu'il existe des possibilités d'acquisition de la troisième dimension pour chaque capteur:

- directement: par sitométrie, dans toute l'image, pour le radar, par télémétrie, au centre de l'image, pour l'optronique.
- indirectement par l'utilisation d'un modèle numérique de terrain.

2- l'optronique fournit une image continue de la scène (rafraîchie à haute cadence), le radar fournit un ensemble de « clichés » séparés au minimum par quelques secondes.

3- la variabilité des performances de DRI est essentiellement due:

- à la présentation de la cible en radar, quelques degrés d'écart et l'image de la cible peut être très différente,
- à la distance en optronique, celle ci joue à la fois directement sur la résolution métrique (effet zoom) et sur l'atténuation atmosphérique particulièrement sensible en optronique.

A contrario, les performances optroniques sont peu sensibles à la présentation de la cible tout comme celles du radar à la distance.

Pour le radar il est possible de dégager les indications suivantes:

- la *détection et la reconnaissance* **du contexte**, qui utilisent la visibilité des textures de terrain associées à des macro-objets (routes, fleuves, ...) ou des associations de points brillants (alignement de pylônes, de bâtiments, ...), sont bonnes car l'observabilité des éléments de contexte ne dépend pas, au premier ordre, de la présentation. Le mode SAR HR s'avère dans la plupart des cas suffisant pour cette opération.
- La *détection et la reconnaissance* **des objets étendus**, qui présentent quelle que soit la direction d'observation suffisamment de réflecteurs pour former une image enveloppe exploitable, sont possibles. La première des cibles caractéristiques, le pont à piles, est typique du phénomène. L'image du pont varie avec sa présentation, les piles signeront peu dans l'axe du pont, mais il subsiste toujours des éléments de structure tel que la rambarde observable dans toute les directions. Avec le mode SAR HR, il peut exister des directions d'observation (au voisinage de l'axe par exemple) où le pont est peu observable, en SAR THR la silhouette devrait être toujours visible.
- La *détection et la reconnaissance* **des petites cibles** de type véhicules ou avions de combat est un problème complexe en cours d'analyse. Le mode SAR HR n'est pas adapté à ce type de cibles et pour le mode SAR THR on peut indiquer certaines tendances qui devront être confirmer par des expérimentations:
- la détection des cibles est possible, le principal élément conditionnant semble être le fond sur lequel se trouve la cible: on voit à quelques dizaines de NM un avion sur une piste, on ne voit pas un véhicule S/A dans de l'herbe.
- La reconnaissance des cibles est possible, dans la mesure ou l'on peut exploiter le paramètre le plus discriminant: les dimensions. Par exemple, il est possible de discriminer un avion de combat d'un avion de transport. Il sera plus difficile de le discriminer avec un objet de dimensions proches: par exemple un camion citerne. Pour cette discrimination il est nécessaire d'utiliser le contenu

de l'image. Lorsqu'on atteint des résolutions métriques les points brillants sont résolus mais ne présentent pas un rayonnement isotrope. Par conséquent, si l'image de l'objet est suffisamment formée pour calculer ses dimensions, elle est aussi très fluctuante en fonction de sa présentation.

Avec un mode MTI il est possible de *détecter* des **cibles mobiles** dans l'image radar si celles-ci ont une vitesse radiale suffisante: supérieure à quelques m/s.

Pour l'optronique, dont la dépendance aux conditions météorologiques est bien connue, il est possible de dégager les indications suivantes:

- la *détection* et la *reconnaissance* **du contexte et d'objets étendus** sont possibles aussi bien en visible qu'en IR; le fait d'imager à site faible (S≤10°) favorisant plutôt les éléments contextuels ou les cibles qui présentent une extension verticale importante.
- la *détection* , la *reconnaissance* et *l'identification* **de petits objets** sont possibles aussi bien en visible qu'en IR.

La résolution étant angulaire, les portées sont très variables en fonctions des cibles auxquelles on s'intéresse. Les portées sont notablement plus importantes sur une base aérienne que sur une batterie sol/air par exemple, surtout si on les pondère par les éléments de contexte et de mission cités précédemment.

- L'image présentée permet une exploitation visuelle aisée. Ceci est valable dans tous les domaines spectraux relevant de l'optronique.

Sur toutes les cibles concernées par une arme tactique, la résolution et les champs considérés offrent un niveau d'information suffisant pour que le cerveau humain fasse son travail de corrélation et d'interprétation.
En particulier, l'exploitation visuelle de l'image est peu dépendante du site d'observation, les transformations géométriques nécessaires étant réalisées instinctivement par l'opérateur.

- Du fait de la contribution de la transmission atmosphérique, les portées sont beaucoup plus importantes à haute altitude qu'a basse altitude, l'épaisseur d'atmosphère traversée étant plus faible. Cette constatation s'applique à tout équipement optronique, et en particulier à ceux dédiés à la reconnaissance.
- Le débattement de la ligne de visée et la fréquence de renouvellement de l'image (fréquence vidéo) autorisent l'exploitation de l'image dans des conditions extrêmes de débattement angulaire et d'évolution du porteur, grâce en particulier à un tracking opérant à la fréquence vidéo.

3-5 Conclusions actuelles sur les possibilités du système multi-senseurs

Importance du contexte pour détecter et reconnaître.
Dans beaucoup de cas l'utilisation du contexte est une aide précieuse pour détecter et reconnaître:

- toujours pour les cibles fixes,
- souvent pour les cibles déplacables (avions, bateau à quai, convoi, trains,...).

Le radar offre de bonnes performances de détection et de reconnaissance, à grande distance, d'objets étendus tels que les routes ou les pistes , et a fortiori du contexte.

Le radar détecte à longue distance des objets de petite taille comme des avions sur une aire de stationnement, il peut avec son mode THR discriminer des objets de tailles différentes, par exemple, un avion de combat d'un avion de transport militaire.

L'optronique offre de bonnes performances de reconnaissance et d'identification sur toutes les cibles de petite taille, et a fortiori sur les cibles plus étendues.

Pour la localisation des cibles, les causes d'erreur principales sont différentes selon les capteurs mais dans tous les cas on constate **l'effet distance**: l'erreur de localisation croit linéairement avec la distance. Il existe toutefois un cas particulier où l'erreur n'est pas proportionnelle à la distance du point désigné, c'est celui où l'on peut utiliser un point d'appui de coordonnées connues présent dans l'image. Une double désignation point d'appui / cible permet alors de s'affranchir de l'effet distance. L'erreur résultante est alors principalement fonction des déformations de l'image.

Le travail à venir permettra de préciser les performances attendues du système en exploitation visuelle et en exploitation assistée.

4- Les traitements

La suite des travaux permettra de définir:

- le concept d'emploi des senseurs,
- les assistances nécessaires à l'exploitation des images.

4-1 Le concept d'emploi du système

Le but de cette tâche est de définir, à partir des possibilités du système multi-senseurs, les conditions d'emploi permettant de réaliser une utilisation optimale des différents modes d'imagerie radar et optronique ainsi que les traitements qui permettront d'effectuer cette mise en oeuvre.
Pour ce faire on utilisera une démarche analytique s'appuyant sur les scénarios de référence.

La mise en oeuvre des capteurs est conditionnée principalement par:

- les contraintes techniques d'emploi conséquences des caractéristiques des modes présentés dans le paragraphe précédent,
- les contraintes opérationnelles à respecter : nécessité de reconnaître la cible avant tir, affinage de localisation nécessaire, discrétion, temps sur zone à respecter, secteur d'attaque, distance de tir minimale ...
- le domaine de performances des capteurs

Le plein emploi du système dans l'environnement fortement évolutif du combat aérien amènera vraisemblablement à définir des logiques de pointage automatique ou de recalage des modes petit champ à partir d'une désignation effectuée dans un mode grand champ, a priori radar sauf contraintes d'approche très basse altitude ou de discrétion.

Vu le grand nombre de paramètres à prendre en compte pour réaliser une mise en oeuvre optimale du système, il sera sûrement nécessaire de définir des fonctions permettant d'aider à la mise en oeuvre du système multi-senseurs et d'en prédire les performances en fonction des utilisations proposées. Une de ces fonctions réaliserait la planification des ressources en préparation de mission, l'autre proposerait une replanification en vol à partir des événements imprévus qui ont conduit à ne pas respecter la préparation de mission. De telles fonctions pourraient être réalisées à l'aide des techniques de PLC classiques (en préparation de mission) et Anytimes (en vol), les algorithmes de PLC Anytimes permettent de disposer d'un résultat en contrôlant le temps de réponse. La PLC déjà largement utilisée pour l'allocation de ressources présente, par rapport à la programmation logique, l'avantage de l'évolutivité.

4-2 Les assistances à l'exploitation des images

Afin d'alléger la charge de travail du pilote, des traitements bord peuvent l'assister dans les fonctions de DRI et de localisation

Les fonctions de détection et de reconnaissance de contexte et d'objets étendus ont fait l'objet de travaux à THOMSON-CSF en radar comme en optronique. Elles sont basées sur l'utilisation de primitives graphiques décrivant des éléments observables du contexte (routes, bâtiments, pylônes,...). Les formes géométriques associées sont des segments, des ensembles de points et des chaînes.
Le traitement consiste à décrire des éléments du contexte sélectionnés et la cible à l'aide de ces formes géométriques. Cette description constitue le modèle de référence qui est établi en préparation de mission à l'aide de données cartographiques ou d'images. Pendant la mission le senseur image le sol et le traitement extrait de l'image ces éléments puis les met en correspondance avec ceux du modèle.

Le résultat du traitement est triple:
- recalage de l'image senseur (et donc de l'avion) par rapport au terrain,
- aide à l'identification du contexte par plaquage du modèle de référence sur l'image senseur recalée,
- reconnaissance de cibles étendues.

Si les coordonnées relatives d'un objectif sont connues par rapport au terrain décrit par le modèle de référence, le recalage image permet d'avoir automatiquement sa localisation par rapport à l'avion.

Ce principe de mise en correspondance devrait être utilisé pour la mise en correspondance des images radar et optronique.

Les traitements permettant de détecter et de reconnaître directement des cibles de petites tailles sont à étudier.
Il existe deux grandes familles pour la détection:
- les traitements qui reconnaissent le fond et qui par soustraction détectent les cibles,
- les traitements qui utilisent une description des cibles (formes,signatures) pour les détecter.

Les traitements du second type sont également utilisés pour la reconnaissance.

THOMSON-CSF a déjà étudié ou réalisé des applications (de classification de radars ou de cibles air menaçantes par exemple) utilisant des techniques neuronales, de segmentation statistique ou d'agrégation multi-critères.
Parmi les difficultés de l'application de détection et de reconnaissance de cibles sol imagées on peut citer:
- l'apprentissage qui devra englober un très grand nombre de cas (cibles, fonds, conditions climatiques et météorologiques,...),
- l'agrégation des informations partielles résultant des différents modes.

Les travaux à venir devront également considérer de quelle façon les informations de DRI issue d'un mode pourront être utilisées avec un autre et ceci dans deux hypothèses d'architecture:
- aval où le système agrège des informations de DRI issues de traitements locaux effectués par chaque senseur,
- amont où le système traite directement les images.

5- Les applications

Les principales applications considérées dans le cadre de ces travaux sont:
- le recalage de navigation,
- la désignation d'un objectif sol: détecter, reconnaître puis localiser l'objectif (si nécessaire),
- le traitement d'une cible par un AASM avec un senseur à imagerie initialisé à partir d'images acquises par les senseurs de bord.

Nous analyserons, dans le cadre de ces travaux, les possibilités du système en fonction des exigences de l'AASM rappelées dans le §1 (portées, précisions, emploi jour/nuit, toutes altitudes, capacité multicible,...).

6- Conclusions

Des travaux sont en cours pour disposer d'un armement tactique précis (métrique ou décamétrique) aux portées accrues afin d'améliorer la survivabilité et l'efficacité de l'avion dans un environnement complexe. La mise en oeuvre de cet armement nécessite de disposer en amont de capteurs permettant **de détecter, de reconnaître et de localiser** précisément les cibles. La disposition de ces capteurs à bord de **l'avion améliore grandement les capacités du système** en matière de **réactivité** (allégement de la préparation de mission) et de **traitements des cibles mobiles ou déplaçables.**

C'est pourquoi, en parallèle de travaux consacrés à l'armement, la DGA a lancé des travaux relatifs aux modes Air-Sol des senseurs du RAFALE.

Ces travaux ont deux finalités:
1- Avoir de modes d'imagerie radar et optronique très performants.

2- Etre en mesure d'exploiter leur complémentarité par des principes d'emploi et une assistance à l'exploitation pour détecter, reconnaître et localiser (système complexe et environnement dense et évolutif)

Il est très important de maîtriser la chaîne qui va de la collecte de l'information (la reconnaissance) au traitement de la cible (par l'arme) en passant par la confirmation et l'affinage des paramètres de désignation d'objectif (par les capteurs des avions du raid).

L'intérêt de cette maîtrise est double :

- garantir l'homogénéité du système d'armes et proposer la meilleure solution à coût minimal,
- rationaliser le traitement d'information capteurs pour les fonctions de détection et de reconnaissance qui existent au niveau reco, avion et arme.

IR-SIGNATURES OF SHIPS: A CLASSIFICATION FEATURE WITHIN A SENSOR NETWORK

H.D. vom Stein, J. Günther, K. Haese*

Allgemeine Nachrichtentechnik
Universität der Bundeswehr Hamburg
Holstenhofweg 85
D-22043 Hamburg
e-mail> h-d.vomstein@unibw-hamburg.de

*Now:
Deutsche Forschungsanstalt für
Luft- und Raumfahrt
Institut für Flugführung
D-38108 Braunschweig

SUMMARY

A system is described by which on the basis of IR-signatures ships are classified using methods of varying interactivity. A sequence of images taken with an IR-camera in a sea environment is superimposed so that pixels in corresponding frames match as closely as possible. With this motion compensation and a following long-term filtering the signal-to-noise ratio is significantly improved.
Subsequently the contour lines of the ship are extracted by a gradient-based or an area-based algorithm. For sequences with a good signal-to-noise ratio this can be done automatically in other cases an interactive component is necessary.
The contour of the "unknown" ship is the feature fo. a correlative or a neuronal classifier. For both classifiers the reliability and the computing time is strongly dependent upon the size of the reference database. It is of inestimable benefit to include all the available sensor data (e.g. sonar and radar) in the classification process thus quickly and significantly reducing the initially very large search space.

LIST OF SYMBOLS

π	image plane
π_0	object plane
x_π, y_π	point coordinates in image plane
x_0, y_0	point coordinates in object plane
f	focal length of camera
d	distance between origin on image and object plane
D_T	part of denominator of inverse perspective transformation
θ	slant
τ	tilt
s_0	length between characteristic points in π_0
s_π	length between characteristic points in π
s_{π_x}, s_{π_y}	x, y -component of length s_π
$\bar{s}$	normalized length s
$\bar{S}$	sum of normalized lengths s_j
ω_r	weight of neuron at location r
$\omega_{r'}$	weight of winner neuron at location r'
m	input vector
ε	learning coefficient
$h_{rr'}$	neighbourhood function
ς	width of neighbourhood function
P	'wavering' product
s_{kn}	n^{th} silhouette of the k^{th} reference image
$h_k(x,y)$	frequency of the k^{th} reference image at the spot (x,y)
N	number of silhouettes in the k^{th} group
h^k	reference image
s_1	database silhouette
s_2	silhouette of the object to be classified

1 INTRODUCTION

Even for a skillful operator it is a tedious and lifetime labor to classify unknown ships with the aid of a large database in the form of a book e.g. with Jane's catalog of Fighting Ships. Using the latest CD-ROM version of this catalog accelerates the classification process, but the interactive component of the operator is furthermore very high.
Therefore the question arises: Can the classification be done partly or fully self-acting? In this paper a method is proposed to classify "unknown" ships in a maritime scene. Depending on the signal-to-noise ratio of the IR-image the classification can be done automatically or an interactive component of an operator is necessary.

At first in chapter 2, a computer-aided interactive classifier will be described. Then in the subsequent chapter 3 motion-compensation and long-term filtering is discussed. This is the basis for the feature extraction, which is discussed in connection with the correlative classifier in chapter 4 and with the neuronal classifier in chapter 5. In both chapters it is easily seen that all available sensor data will accelerate and ameliorate the classification process. With a short conclusion in chapter 6 this paper will be terminated.

2 A COMPUTER AIDED INTERACTIVE CLASSIFIER

The problem of recognition of an unknown ship can be led back to the basic scheme of pattern recognition shown in Figure 2-1.



Figure 2-1:Basic scheme of a pattern recognition system

The object to be classified (OTBC) is filmed by a sensor (IR or Video). After extracting characteristic features these are compared with the corresponding items in the database and the unknown object is classified.
In many cases, due to the low signal-to-noise ratio the automatic feature extraction is impossible. Therefore, under these bad conditions, a human operator carries out the feature extraction.
To what extent a special feature is significant for the classification process cannot be answered easily. In this chapter the feature "ship parts" and their position to one another have proven significant. Figure 2-2 shows an infrared picture of a ship. The crosses mark the x and y coordinates of the bow, the masts, the bridge, the funnel and the stern in the image.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

The Euclidean distances between the different individual "ship parts" and the bow are scaled to the distances between bow and stern. These numbers are the features of the OTBC.
A database contains the same data for all models to be classified. Therefore, a classifier compares the data of the OTBC with the corresponding data of each individual model. The difference between two corresponding "ship parts" is a measure of the similarity (distance near zero) or dissimilarity

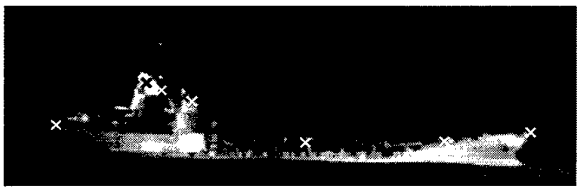


Figure 2-2: Infrared picture of a ship

(distance very great) of the OTBC and the chosen model of the database.
Having compared the OTBC with all models one has a number for each comparison (OTBC-model). Sorting these numbers in a rising series and using the similarity as a feature, the model with the lowest number has the highest probability to be the OTBC.
Furthermore the operator's statement that he has chosen a particular "ship part" is included in the classification process in a fault-tolerant way.
Confusion of a mast with a gun is much less probable than that of a gun with a rocket launcher. Therefore each pair of "ship parts" is allocated a confusion probability and derived from this probability a number that is added to the already calculated Euclidean distances. Improbable assignments result in great distances.
Thus the total distances imply the aberration of the geometrical distances and the confusion in the determination of corresponding "ship parts" in a fault tolerant way.
In accordance with the number of "ship parts" and models in the database the time necessary for the calculation of all possible distances between the OTBC and the models increases exponentially. (NP problem) With the so-called A* algorithm a perhaps suboptimal solution is found within acceptable time.
Furthermore leaving extraordinary meteorological situations out of account (pitch and roll-angles are negligible) the position angle is directly connected to the visibility of the individual "ship parts". Thus the information of visibility must also be stored in the database.
Preliminary tests with a database of 200 models and 7 to 8 marked "ship parts" showed, that the model corresponding to the OTBC was found on the average among the first 5 ships of the list arranged to decreasing similarities.
A visual comparison of the OTBC and the model images verifies and completes the recognition process.
Further details are described in (Ref. 1)

3 MOTION-COMPENSATION AND LONG-TERM FILTERING

The image quality of thermal imagers available today is bad with regard to those of the visual spectrum. The IR-images suffer from a low signal-to-noise ratio, low contrast at the object boundaries, the propagation conditions and sensor artifacts. Therefore the usual steps in image processing are not sufficient enough to guarantee almost error free results.
A good basis for preprocessing and extraction of the contour line as a feature relevant for classification is the so-called motion-compensated sequence of IR-images. These sequences have much more data at their disposal than a single image and therefore the evaluation security increases in a significant statistical sense.
With the aid of "time information" new evaluation algorithms adapted to these problems can be formulated.
As shown in Figure 2-1 and Figure 3-1 a sequence of the OTBC is taken from a sensor platform. OTBC and sensor platform are both in relative motion to one another. Thus it is not possible to take directly the mean of all images in the sequence, to ameliorate the signal-to-noise ratio.
At first both motions must be estimated and then compensated to get a stationary object that does not move relative to the image border. Then the mean value of the images of the sequences is formed.

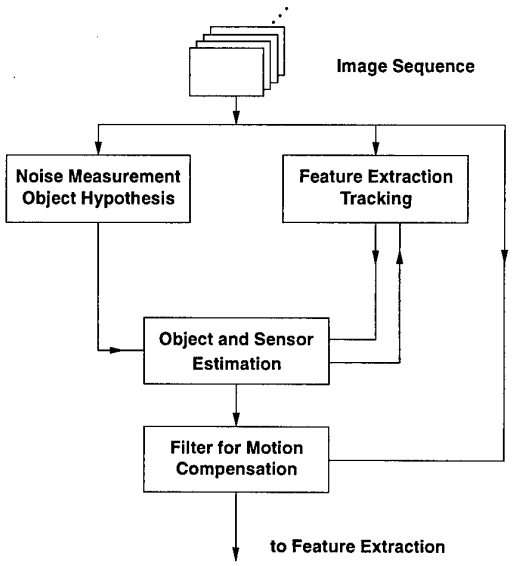


Figure 3-1: Motion compensation of image sequences

The main items of this procedure are explained in Figure 3-1.

In a first step a noise measurement is made in an image sequence. This is the basis for the adaptation of several modules to image quality, which changes significantly due to meteorological situations, object attributes and distance between sensor and OTBC.
Then the motion compensation of the OTBC and the sensor platform, which are subjected to three-dimensional motion, are carried out.
It is necessary to know scene area indicators, which can be clearly assigned to the object (OTBC) and to the sensor platform.
Hot-spots are well suited for this purpose and they belong to the OTBC. Therefore a hot-spot detector module was realized, which determines all hot-spots in the images of a sequence. In a spatio-temporal filter the hot-spots, e.g. sunglints, which do not belong to the OTBC, are eliminated.
A scene area indicator for the sensor-platform ego-motion is the horizon, which can be determined in nearly all images without great difficulty.
Both methods serve as the basis for an Extended Kalman filter, by which a motion-compensated calculation of a mean value image is realized, that means that the OTBC in all images of the

scene is shifted so that its position relative to the image border is fixed. (Ref. 2, 3, 4)
Thus the signal-to-noise ratio of the mean value image is significantly improved depending on the square root of the length of the sequence.

The result of motion-compensation is illustrated in the Figure 3-2, where a single image of the sequence and the mean value image are shown. To demonstrate the amelioration of the mean image details of the ships are magnified.
The basis for the following contour extraction are the mean value image and all the images of a scene which will be shown in the next chapter.

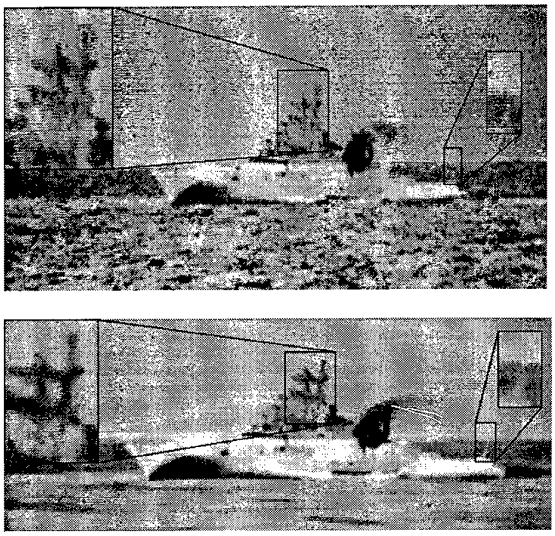


Figure 3-2: top: Single image of a sequence
bottom: Mean value image

4 FEATURE EXTRACTION AND THE CORRELATIVE CLASSIFIER

In the interactive classification procedure in chapter 2 the "ship parts" and their positions are used as features for the classifier. For the automatic and semiautomatic classification process the contour of the OTBC proved to be a powerful feature.
Therefore besides the interactive approaches two automatic techniques of object contour extraction have been studied. These differ from the wellknown procedures in literature, as they take into account the time information from motion compensated image sequences.
The first method uses primarily distinguishing criteria for different temporal and spatial local gray-level histograms, whereas the second method is especially based on homogeneity criteria.
Both procedures can be regarded as extensions of the contour-based or region-oriented segmentation algorithms known in literature for at least a couple of years.
The first method of contour extraction is divided into two moduls. In one modul stationary contour areas are extracted and in a second module an adaptive hierarchical contour extraction with a subsequent rule-based algorithm is performed.
With the aid of a comparison of the local and temporal variance of the compensated sequence, stationary structures corresponding to the OTBC and instationary structures corresponding to the water-surface texture are separated.
Stationary object contours appear only in regions where the local variance is significantly higher than the time variance. Therefore the quotient from the spatial variance of a local window of the time averaged image sequence and from the time variance of this image sequence locally averaged in this window is evaluated. Three classes can be roughly assigned to this quotient:
1. In the sky and in the internal object areas the gray levels are due to Gaussian noise and using a suitable scaling the variance quotient obeys a F-distribution.
2. The variance quotient in the area of the water surface also responds to a F-distribution, but the expected value is shifted towards smaller values.
3. In the neighbourhood of the contour the variance quotient is larger and its distribution is unknown.

With these conditions in mind appropriate threshold values for stationary contours can be found. The security of this procedure is improved by a hierarchical application to a spatial/time subsampled motion-compensated image sequence.

Subsequently an adaptive hierarchical contour extraction is carried out for the stationary contour areas. This helps to determine the fine-structure of the OTBC-contour. The result of the contour based segmentation in a motion-compensated and long-term filtered image sequence is shown in Figure 4-1.The

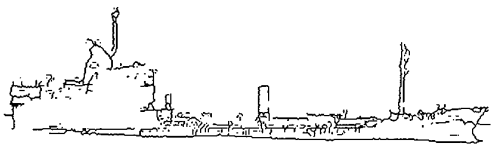


Figure 4-1: Result of the contour-based segmentation.

good performance of this technique is noteworthy.

The second method of contour extraction is a region-oriented segmentation approach, which is based on an extended split & merge algorithm. An extended set of characteristics is used in the split - as well as in the merge phase. It is also based on the inclusion of different spatial and temporal statistics of the motion-compensated image sequence.
To determine image-adapted decision thresholds an artificial neural network is implemented in this method. Furthermore the knowledge of the source is employed in the segmentation procedure.
In the split process the image is stepwise divided in details which can be regarded as homogeneous with regard to a special feature.
In the merge phase homogeneous neighbouring base segments are grouped together, if they are sufficiently similar.
Regionally global as well as regionally local criteria are used for similarity or dissimilarity of regions.The base segments and the regions are iteratively connected whereby the decision threshold is lowered step by step.
In a last step, called interpretation phase, the combination of regions is completed with regard to plausibility criteria. Then on the bases of rules and of a scene model the regions are assigned to background and object.
Finally the outer border of the OTBC is determined which corresponds to its contour.
The result of the region-based-algorithm in a motion-compensated and long-term filtered image sequence is shown in Figure 4-2.

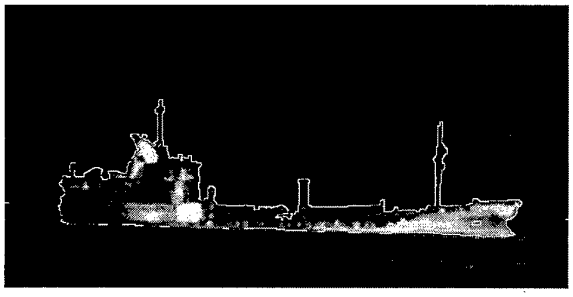


Figure 4-2: Result of the region-based segmentation

With this just mentioned method one receives always a closed contour, which is the direct or indirect basis for the different classifiers. A combination of both segmentation methods is under development.

4.1 Correlative Procedure

By means of the correlative classifier the object to be classified is assigned to available 3D ship models. In this approach the correlation coefficient serves as a resemblance criterion. By way of example, the structure of a classifier is presented with the aid of two correlation criteria (Equation 4-2 and Equation 4-3).

Figure 4-3 shows the structure of the correlative classifier. Classification is done in two steps designated as classifiers I and II in Figure 4-3.

4.2 Comparative Data

For the purpose of classification the object to be classified is compared with models from a database. The database is established according to the following scheme. On the basis of 3D-ship models, 2D-silhouettes (Figure 4-4 top) are produced.

The 3D-models are looked at under a specific role-, pitch- and position angle. Only the silhouettes that may appear in dependence upon the sensor location are stored in the database and supplied to a cluster algorithm.

The database silhouettes are roughly quantized. The gray levels are binarized and the length of the OTBC is normalized to 32, 48, or 64 points (pixels). The result is a binary matrix shown in Figure 4-4, centre.

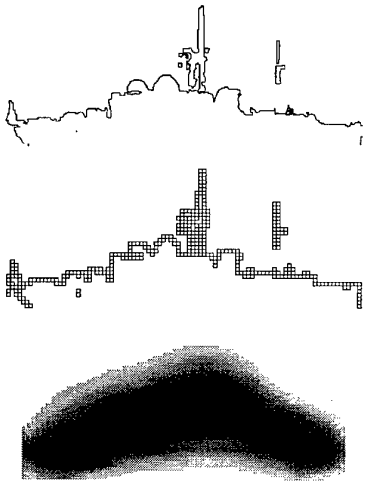


Figure 4-4: top: Silhouette
centre: Roughly quantized binary matrix
bottom: Gray-level reference image

From the binary silhouettes produced that way, a cluster algorithm determines a restricted number of reference images each of which represents a group of ship silhouettes similar in appearance (Figure 4-4 bottom). The relative frequencies of a pixel occurring at the spot (x, y) of all N silhouettes belonging to one group are stored in the matrix elements of these reference images according to Equation 4-1.

$$h_k(x,y) = \frac{\sum_{n=1}^{N} s_{kn}(x,y)}{N} \qquad \text{Equation 4-1}$$

The binary matrix and the reference image matrix are the comparative data used for classification.

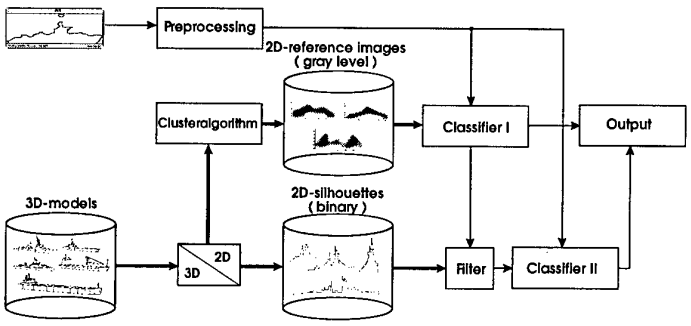


Figure 4-3: Structure of the classifier

4.3 Classification

The ship silhouette provided by means of feature extraction is also roughly quantized for classification. The silhouette of the object to be classified is now available as a binary matrix. Regarding classification it is of no importance whether an automatically produced or a manually produced silhouette is used.

In both cases the OTBC is available as a binary matrix.This matrix is classified by means of classifier I depicted in Figure 4-3. In a first step, the binary matrix of the OTBC is assigned to

one or more reference images stored in the database, that is, to one or more classes of silhouettes according to the maximum of Equation 4-2

$$NA_{h_k s_2} = \frac{\sum_{x=0}^{n-1}\sum_{y=0}^{m-1} h_k(x,y)\cdot s_2(x+v_x, y+v_y)}{\sum_{x=0}^{n-1}\sum_{y=0}^{m-1} s_2(x,y)} \qquad \text{Equation 4-2}$$

$$0 \le h_k(x,y) \le 1\,,\; s_2(x,y) \in (0,1)$$

After this first step, statements can be made and issued on the probable group and the position angle with regard to the observer. A fault tolerant filter selects the binary silhouettes corresponding to the group and to the position angle from the 2D-database. They are used for classification in classifier II.

In the second classifier it is attempted to assign the silhouette of the object to be classified unambiguously to a database silhouette. For this purpose an average-free normalized cross-correlation function is used (Equation 4-3). In some cases it is insufficient to calculate the correlation coefficient only with the exact pixels of the OTBC and the database silhoutte. An operator wich includes neighbouring pixels provides better results. For this purpose the second classifier uses the operator O_{KKFMFB} to broaden the silhouettes.

$$O_{KKFMFB} = \begin{bmatrix} 0 & 0.125 & 0 \\ 0.125 & 0.5 & 0.125 \\ 0 & 0.125 & 0 \end{bmatrix}$$

In both classifications the silhouette of the object to be classified (s_2) is moved by v_x and v_y on the corresponding comparative image until the maximum correlation coefficient is found.

The 'correlation coefficients' of Equation 4-2 and Equation 4-3 are between 0 and 1.

In the final step the correlation coefficents are sorted in a rising series. For unambiguous classification it is necessary to carry out a visual comparison of the object to be classified (OTBC) with the database silhouettes. Here, the operator is supported by the computer laying the database silhouette assigned to the OTBC and the OTBC silhouette on top of one another (Figure 4-5). The database silhouettes are represented in Figure 4-5 by gray squares.

$$KKFMF_{s_1 s_2}(v_x, v_y) = \frac{\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}\left\{s_1(x,y)-\overline{s_1(x,y)}\right\}\cdot\left\{s_2(x+v_x, y+v_y)-\overline{s_2(x,y)}\right\}}{\sqrt{\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}\left\{s_1(x,y)-\overline{s_1(x,y)}\right\}^2 \cdot \sum_{x=0}^{n-1}\sum_{y=0}^{m-1}\left\{s_2(x,y)-\overline{s_2(x,y)}\right\}^2}} \qquad \text{Equation 4-3}$$

$$s_1(x,y) \in (0,1)\,,\; s_2(x,y) \in (0,1)$$

4.4 Results

Extensive tests have been carried out with the classifiers described. Good results are obtained if the spatial resolution of the matrix (Figure 4-4) is well adapted to the dimension of the OTBC in the image. The loss of information increases with a decreasing matrix width. Too great a number of matrix columns is also pointless as very high accuracy requirements are to be established regarding the 3D-models and the extracted silhouettes. Furthermore, the amount of computation increases quadratically with increasing the matrix width and hight by the same factor.

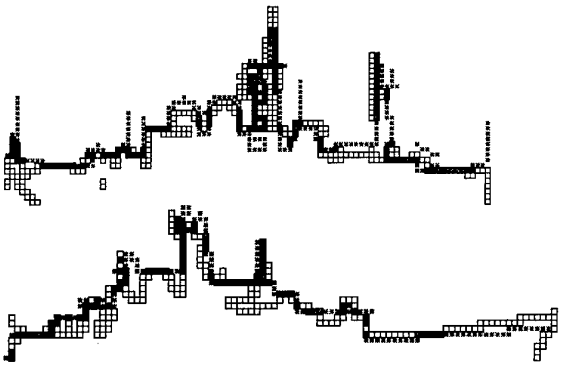


Figure 4-5: Results
top: Automatically extracted silhouette
bottom: Manually extracted silhouette

In conclusion it can be stated that classifier I ensures a silhouette assignment of almost 100% to the right silhouette groups and classifier II achieves an identification rate of 80% of the ship silhouettes. However, it is to be noted that these results refer to manually established silhouettes and to currently used databases comprising approximately 100 3D-models. The classification and identification results obtained with automatically extracted ship silhouettes are strongly dependent on the quality of the silhouettes provided.

5 NEURAL SHIP RECOGNITION FROM PERSPECTIVE IMAGES BASED ON THE CONTOUR

In chapter 3 the motion compensation of an infrared image sequence was explained. Motion compensation is necessary in order to extract the ship from the image background as shown in chapter 4. Then, finally, the contour of the ship is obtained, which is the feature to identify the ship.

This identification is performed by
- extracting characteristic points of the contour, which are found independent of the viewing position,
- calculating scale invariant features between the characteristic points for contours of many different viewing positions,
- training a self-organizing feature map with these features, which does the identification after the learning process.

This identification method then provides the feasibility of ship recognition on the basis of one contour only.

5.1 Extraction of Characteristic Points

The contour of a ship in the image plane π (In this chapter in the contrary to the previos one two different coordinate systems with the indices π and 0 are used) is an aggregate of points,

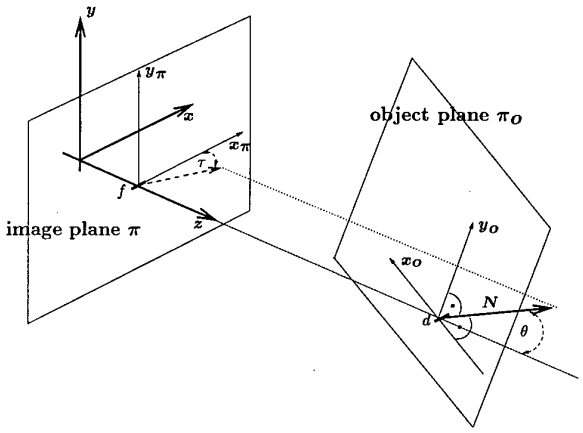


Figure 5-1: Relation between object and image plane

which are projected with slant θ (see Figure 5-1) from the ship in 3D space into the 2D image plane. Hence, the shape of the contour is dependent on the camera's viewpoint. Fortunately, this aggregate of points contains some points, which can be found independent of the viewing position at constant values of normalized y_π -coordinate, if some assumptions are fulfilled.
These assumptions are
- long distances d between the ship and the camera
- and the focal length $f >> x_\pi, y_\pi$.

Using these assumptions the formulas of inverse projection from image plane π into object plane π_0, as they read.

$$x_0 = \frac{d(x_\pi \cos(\tau) + y_\pi \sin(\tau))}{\cos(\theta) D_T} \qquad \text{Equation 5-1}$$

$$y_0 = \frac{d(-x_\pi \sin(\tau) + y_\pi \cos(\tau))}{D_T} \qquad \text{Equation 5-2}$$

with

$$D_T = (f + x_\pi \tan(\theta)\cos(\tau) + y_\pi \tan(\theta)\sin(\tau)),$$

are simplified to

$$x_0 = (x_\pi \cos(\tau) + y_\pi \sin(\tau))/\cos(\theta) \qquad \text{Equation 5-3}$$

$$y_0 = -x_\pi \sin(\tau) + y_\pi \cos(\tau). \qquad \text{Equation 5-4}$$

Furthermore, in standard recognition situations the sea is calm, so that tilt τ is small (see Equation 5-3 and Equation 5-4). Then, normalizing the y_π -coordinate to the height of the ship, characteristic 3D points are projected to image points lying on a special hull of the contour (see Figure 5-2) at constant normalized values of y_π.

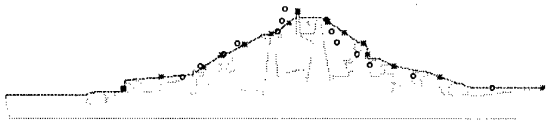


Figure 5-2: Characteristic points (*) of the contour under $\theta = 5°$ and inverse projected characteristic points (o) of contour in Figure 5-3

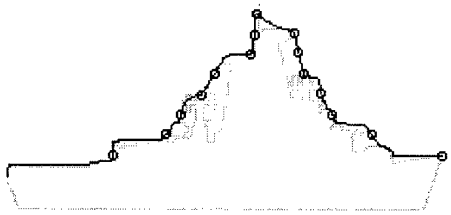


Figure 5-3: Characteristic points of contour under $\theta = 65°$

5.2 Scale Invariant Feature Calculation

Now, the length $s_0(\theta)$ between the characteristic points are modeled for all possible viewpoints between $0° < \theta < 65°$.

$$s_0(\theta) = \sqrt{s_{\pi_y}^2 + s_{\pi_x}^2 / \cos^2\theta}. \qquad \text{Equation 5-5}$$

If this length in the image as well as in the object plane is normalized by proper perimeter of the contour, then it can be approximated using Taylor expansion (Ref. 6):

$$\bar{s}_0(\theta) \cong \bar{s}_\pi + a_{\bar{s}} \cdot \theta^2 \qquad \text{Equation 5-6}$$

This approximation also holds for an integral length, $\bar{S}_i = \sum_{j=0}^{i} \bar{s}_j$, and can as well be deduced for the angles between the contour segments.
These relations between integral length (angles) and slant θ are used to generate the training data for the neural network on the basis of one contour per ship. The neural network is a self-organizing feature map, which has to identify later on contours obtained from 3D models of ships and noisy contours extracted from infrared image sequences.

5.3 Topographic Presentation with SOM

The neural network algorithm of KOHONEN (Ref.7, 8) produces some topographic presentation of the training data in a space of lower dimension than the input training data space. A topographic presentation of the training data is desired in order to classify even unknown ships or hardly identifiable ones.

KOHONEN's algorithm can be summarized describing the learning rule for the weight vector ω_r at location r by Equation 5-7 and Equation 5-8.

$$\omega_r(j) = \omega_r(j-1) + \Delta\omega_r(j) \qquad \text{Equation 5-7}$$

$$\Delta\omega_r(j) = \varepsilon(j)\, h_{rr'}(j)[-m(j) - \omega_r(j-1)] \qquad \text{Equation 5-8}$$

Equation 5-8 contains $h_{rr'}(j)$, which is a Gaussian neighbourhood function. Its maximum is at the location r' on the feature map.

$$h_{rr'}(j) = \exp\left(-\frac{|r-r'|^2}{2\cdot\varsigma^2(j)}\right) \qquad \text{Equation 5-9}$$

with

$$|r-r'| = \sqrt{(r_1-r_1')^2+\cdots+(r_n-r_n')^2}$$

This location is found by calculating the nearest neighbour to the input vector m:

$$|\omega_{r'} - m| = \min_{r\in A} |\omega_r - m| \qquad \text{Equation 5-10}$$

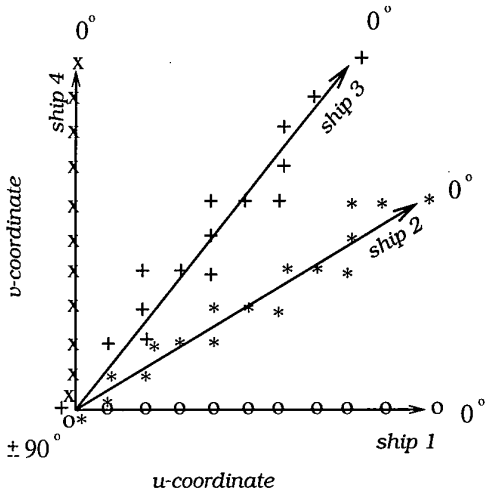


Figure 5-4: Ideal organization of ships on a feature map (0° < θ< ± 90°)

Analyzing this algorithm's behaviour as well as that of the training data, it can be deduced, that on a two-dimensional feature map ideally the training data should be organized in a way shown in Figure 5-4. In fact, a map of this kind (see Figure 5-5) is obtained including 4 ships in the data set. Selecting 18 ships topology, is best preserved on four dimensional maps. This is concluded from the wavering-product, which is a measure of topology preservation on a feature map (Ref. 6)

Neurons in each dimention	Wavering-product	3D-model contours (0°<θ<65°)	Infrared-contours
27 x 27	-0.0221	84%	55%
9 x 9 x 9	-0.0043	80%	45%
5 x 5 x 5 x 5	-0.0020	81%	30%
9 x 9 x 9 x 9	$2.5\cdot10^{-4}$	81%	35%
5 x 5 x 5 x 5 x 5	0.0255	81%	50%

Table 5-1: Identification results on various feature maps containing 18 ships

The wavering-products of trained higher dimensional feature maps are shown in Table 5-1. A value near zero indicates good topology preservation (Ref. 6). This is achieved in case of four dimensional feature maps. In case of four dimensional maps the identification rate of contours from 3D-models is 81% . Only 35% of the infrared contours are identified. Surprisingly, lower dimensional maps show better identification results. 84% of the 3D-model contours and 55% of infrared contours are identified. It is concluded, that after some parameter studies four dimensional maps can be trained to reach as good identification rates as in case of the two-dimensional map.

In summary, the presented contour based method to recognize ships from various viewpoints shows good results on the chosen data set of 18 ships. Its robustness is demonstrated by the infrared contour identification rate of 55% in a two-dimensional feature map. In contrast, an early proposed method of viewpoint invariant ship recognition using higher order neural networks (Ref. 9) is not as successful as this method, though it uses 9 different contours of the ships during the learning process.

The proposed neural identification method is able to identify ships on the bases of their contours. Because the shape of the

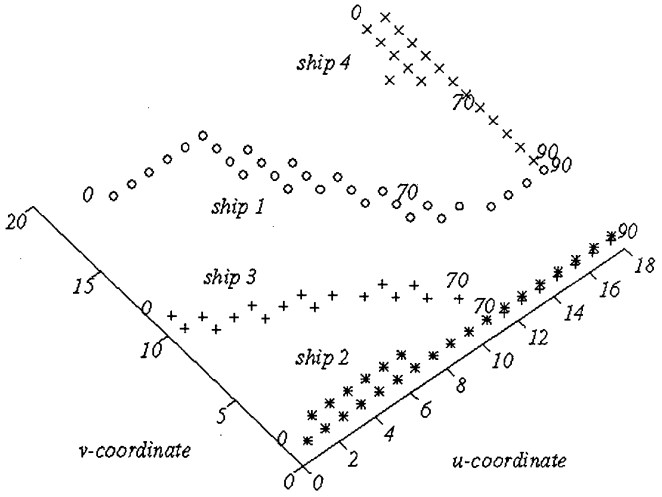


Figure 5-5: 2D-Feature map trained with 4 ships

contour is viewpoint dependent, it is worth mentioning, that the proposed method is feasible to identify a ship if only one of its contours is known. In contrast, other methods need large databases of possibly occurring contours, which have to be laboriously collected. This advantage is achieved by extracting characteristic points on a special defined hull of the contour invariant of the viewpoint. Afterwards, length and angles between these points are modeled for all viewpoints and then used to train a self-organizing feature map. Its identification rates reach 84% in case of good contours, 55% in case of blurred ones.

These results are very good in comparison to other approaches of viewpoint invariant ship identification (Ref. 9).

6 CONCLUSION

In this paper different methods of classifying IR-signatures of ships have been discussed. Depending on the signal-to-noise ratio of the IR-images the classification can be done automatically or an interactive component is necessary. In every case the reliability and the computing time of the classifier is strongly dependent upon the size of the reference database. Therefore it is necessary to include all the available sensor data (e.g. sonar, radar) in the classification process to reduce the amount of models in the database as much as possible. This is even more significant for the neuronal than for the correlative classifier. In the contrary the neuronal classifier needs an considerable smaller amount of silhouettes than the correlation classifier. The comparison of the classification error rates is difficult, as the number of models in the database is to different.

REFERENCES

1. Höft, O., Ein fehlertolerantes Verfahren zur Schiffsklassifikation unter Verwendung eines probabilistischen Modells, Dissertation 1994,Uni Bw Hamburg.

2. Meier, W., vom Stein, H. D., Infrared Image Enhancement with Nonlinear Spatio-Temporal Filtering. EUSIPCO-92 Brüssel, pp. 1397-1400.

3. Meier, W.; vom Stein, H. D., A Two-Step Kalman Process for Sensor and Object Motion Estimation. ISPRS Symposium: Spatial Information from Digital Photogrammetry and Computer Vision. München 1994 pp. 555-561.

4. Meier, W., Modellgestützte Bildverarbeitungsschritte zur Analyse von Infrarot-Bildfolgen maritimer Szenen, Dissertation 1995, Uni BwHamburg.

5. Smeets, W., Ein modellbasierter regionenorientierter Ansatz zur Segmentation von Infrarot-Bildsequenzen, Dissertation 1995, Uni Bw Hamburg.

6. Bauer, H.-U; Pawelzik, K.R., Quantifying the Neighbourhood Preservation of Self-Organizing Feature Maps. IEEE Transactions on Neural Networks, 3(4):570--579, 1992.

7. Kohonen, T., Self-organized Formation of Topologically Correct Feature Maps, Biological Cybernetics, 43:59--69, 1982.

8. Kohonen,T., Self-Organization and Associative Memory. Springer Series in Information Sciences 8, Heidelberg, 1984.

9. Meier, W.; Haese, K.; vom Stein, H.-D., An Integrated System for Infrared Image Sequence, Analysis and Classification. Proceedings IASTED International Conference on Signal and Image Processing, pp. 229--232, 1995.

10. Pizlo, Z.; Rosenfeld, A., Recognition of Planar Shapes from Perspective Images Using Contour-Based Invariants. Image Understanding, 56(3):330--350, November 1992.

PAPER No: 10

DISCUSSOR'S NAME: N.J. Farsaris

COMMENT/QUESTION:

This seems to be a very interesting recognition system. What is the maximum range of recognition/classification obtained using that system?

AUTHOR/PRESENTER'S REPLY:

Due to the focal length of our IR-system, it is in the order of 2-3 km for the moment.

DISCUSSOR'S NAME: P. Verlinde

COMMENT/QUESTION:

Did you include a "rejection option" into your automatic identification scheme?

AUTHOR/PRESENTER'S REPLY:

Up to this moment a "rejection option" is not included. The classification only makes a proposal which must be verified by the operator.

However, a "rejection option" can be added.

DISCUSSOR'S NAME: G.S. Brown

COMMENT/QUESTION:

Why must the human remain in the decision loop?

AUTHOR PRESENTER'S REPLY:

The human makes the definite decision; therefore he (or she) must remain in the loop!

PAPER No: 10

DISCUSSOR'S NAME: A.J. van der Wal

COMMENT/QUESTION:

1. Why the preference for the Kohonen revised network????

2. Is it because of its self-organisation aspect?

3. Did you also try other NN architectures?

AUTHOR/PRESENTER'S REPLY:

Multi-layer perceptions have been tried, but they didn't perform well. Perhaps because we didn't have enough being done. Kohonen images seemed to perform the best in identification.

Advanced Electro Optic Targeting System

Richard W. Jacobs/ Henry S. Lapp
US Air Force, Wright Lab
Targeting Systems Branch,
Electro-optic Technology Division
WL/AAJS, Bldg 622, 3109 P St.
Wright Patterson AFB, Ohio 45433-7700

Abstract

Future Aircraft targeting systems are integrating sensor inputs for automated target detection, location and identification with the goal to both lower system cost and increase system performance. This paper describes a common aperture electro optical (EO) targeting concept for an integrated air-to-ground EO targeting system using promising developments in multispectral target detection and in laser target identification. The integrated capability allows use of a minimum system aperture while increasing recognition and targeting ranges for targeting outside missile threat envelopes. The technical status and preliminary test results of the Wright Laboratory thrusts to develop both laser identification and thermal spectral detection capabilities will be discussed. These involve developments and demonstrations of multi-spectral detection using a few thermal bands and non-cooperative target identification using a burst illumination laser with eye-safe wavelength and a short wave infrared gated camera. These individual concepts will be mature enough to support an integration effort in the 1999/2000 timeframe. The cost, performance and level of automation supports use for both manned and unmanned air vehicle targeting and reconnaissance missions.

1. The Need for an Advanced EO Targeting System

For the past several years the USAF has been interested in building and demonstrating extended range EO systems that provide a targeting capability from stand off ranges outside the missile threat envelope, using a smaller aperture compatible with lower aircraft observability concerns. In the early 1990's work was sponsored to integrate the advantages of both the FLIR and Laser with the "FLASER" or laser plus flir. Flaser was designed to demonstrate detection of hot targets in the open with the new advanced flir detectors and coupled with a burst illumination laser for target ID. This work was down scoped due to funding problems but the work content continued in parts under other programs. The advantages of using the Laser as a burst illuminator of potential targets, similar to a flash lamp in photography, and then using the flir to detect targets in the open and filter out the clutter has shown great potential. Section 3 will provide more detail with the discussion of the Enhanced Recognition and Sensing Ladar (ERASER) program. Addtionally it is clear that future targeting should be

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

done at extended ranges whenever possible such that the ingressing aircraft platform can "stand-off" outside the ranges of surface to air missiles. This requirement also translates into being able to recognize and identify friend, foe, or neutral forces at the extended ranges. EO sensors or systems offer distinct advantages for recognition with high directionality and good spatial resolution. A paper by Eismann et al showed that a high confidence of target detection and identification can be achieved with a detection probability of 0.9 and a false alarm rate of 0.01 to 0.001 per square kilometer of area searched. The current state of the art in FLIR systems allows for good capability for targets in the open or which are partially concealed but have high target to background contrast. However, for targets in camouflage, concealment, and deception (CC&D), current capability is poor. The other concern is that the EO targeting sensor is not an all weather sensor so most attack aircraft will have radars that provide the capability to detect tactical sized targets in the open or moving. The radar is the baseline targeting sensor, with the EO system as the clearer weather capability that provides a cost effective capability to find deep hide targets and to guide precision weapons. This motivates us to strive to build a targeting system in the future that can be capable in these more demanding situations. Whatever solutions we recommend in the future must be additionally cost effective or affordable. One aspect to achieve that may be to share resources by integrating the best of both radar and EO systems to achieve enhanced recognition and ID. Alternatively, it may not be necessary to have a complete EO system on every offensive aircraft but rather to share the resources between a fleet of them or to use various assets between manned fighters, unmanned aircraft, or Joint Stars vehicles.

2. Standoff Capability

Recent operational requirements are driving us toward increased range ID and targeting. In particular, it can be shown that if one can do these functions at ranges of at least 13 km, one can stand off and survive all or most surface to air missile threats. The objectives of our programs are to extend this standoff range as far as possible, but recent developments show promise of doing this at about 20km range. One of the key contributors to this extended range is the ERASER program.

3. Enhanced Recognition and Sensing LADAR

The ERASER program has two objectives, the first being to demonstrate air to ground identification of ground targets using modified eye safe laser designators packageable into existing flir targeting systems. This will provide future pilots with positive, timely, and reliable target ID and Identification Friend or Foe (IFF) capability as an upgrade to existing operational flirs. The second objective is to provide detailed engineering performance measurements that will support flir upgrade decisions and performance estimations. ERASER will also evaluate the capability to achieve a limited air-to-air ID capability by taking

advantage of laser vibrometry and microdoppler developments and related laser sources being developed. Several key points about the ERASER program make it unique. The identification is done using a short laser pulse to illuminate the target of interest for an instant, then, a range gated image intensified Charge Coupled Device IR camera is used to both filter out the backscatter and receive the optical pulse for recognition. This offers a distinct advantage over scanning systems as the single pulse approach is nearly covert. This serves to minimize detection of the source. The ERASER backscatter reduction benefit is analogous to the effect of achieving greater visibility from regular headlights over high beam lights while driving an automobile in the high backscatter conditions of a snow or sand storm. Additionally ERASER will make use of the same laser system for recogition as other laser functions onboard and in particular the same laser source as the laser designator/ranger. The intent is thus to effect cost savings and prevent proliferation of sensors, apertures, and functions. An advantage of the ERASER approach over a normal FLIRsystem is the active illumination of the target, and, a decreased dependence on environmental factors and background irradiance by using an active system that uses target reflectivity. Also, ERASER expects to use eye safe laser wavelengths (~1.5 um) for operational safety. This is projected to be accomplished through either Optical Parametric Oscillator (OPO), or a Raman shifter, to a 1.06um pump laser. Several approaches to this wavelength shifting include a potassium titanyl phosphate (KTP) OPO, or an erbium doped fiber pre-amplifier. In addition, a new solid state laser designed for 1.5 micron as fundamental wavelength may be feasible soon.

The ERASER program has had a goal to use existing reflective common aperture FLIR designator system hardware to affect retrofit and affordability. In addtition ERASER could include a degree of Automatic Target Recognition (ATR) capability although it will also maintain the pilot in the loop prior to weapon release. Ideally ERASER would be used to provide positive target identification (x, y TV like picture) along a "soda straw" for the pilot prior to launching a laser guided weapon. Presently pilots prefer to positively confirm the target identity prior to actual weapon launch. The paper by Watson et al indicated that the desired long range of ERASER should be achievable with a transmitted energy of about 400 mJ for the eye safe wavelengths.

Progress on ERASER has been encouraging in that preliminary tests in 1996 with a breadboard system showed good results to detect at medium range and realistic desert conditions. These tests were performed at Camp Roberts in California during National Guard summer exercises. ERASER performed well at ranges of approximately 3-10 km and provided eye safe 2 Dimensional imagery. The brassboard ERASER was able to see through fog oil smoke intentionally generated around a tank as target. In the late Summer of 1997 an extended test is planned for White Sands Missile Range in New Mexico, to evaluate and demonstrate ERASER at up to 25km range.

4. Multispectral sensing

Multispectral sensing is a means to exploit the spectral information of a potential target by using several wavelengths of interest. Multispectral differs from hyperspectral in that in the former 2-3 key wavelengths are used rather than dozens with the latter, thus relieving the hardware and processing demands of the system. Desert Storm showed the need to automatically detect and identify military targets with very low false alarms as well as to locate targets with sufficient accuracy to perform precision strike. Additionally targets with Camouflage, Concealment, & Deception (CC&D) or in deep hide make detection more difficult. Multispectral systems offer the capability to perform limited area search from a passive sensor from low to high altitudes. Allthough limited by weather, thermal multispectral offers a day or night passive detection capability. Multispectral sensing requires significantly less spatial resolution so systems can be designed that image wider areas than single band spatial sensors.

By using multispectral detection in selected wavelength bands germane to the targets/backgrounds/atmospheres of interest, e.g., the wavelengths characteristic of vehicle paints, camouflage nets, etc., it is possible to exploit these characteristics and augment the detection capabilities of passive IR systems. Multispectral is used for target detection enhancements by a careful reduction in the background clutter. This is done by choosing band combinations for which the clutter data is highly correlated between bands. When done correctly, the suppressed clutter level is less sensitive to typical diurnal variations in environment of single band clutter. The few bands are selected to take advantage of the correlation of the background. Ideally two bands could be chosen that have all the background pixels very correlated (need definition of correlation) and the target pixels with lots of color relative to the background. A simple subtraction of the images would then leave only target pixels. Many years of tower experiments with both imaging and non-imaging thermal hyperspectral instruments against various backgrounds at various times of the day/night and seasons have shown that backgrounds consistently do exhibit a high degree of correlation. The general case however requires more than two bands to achieve high probabilities of target detection and more importantly low false alarms.

Results so far have shown the potential for multispectral systems using both tower and airborne imaging systems. In parallel, new detection algorithms have been developed for optimum performance and minimum false alarms. The paper by Eismann et al showed that detection at greater than 15 km range is feasible for both Mid Wave and Long Wave (MW and LW) IR systems. The thermal multispectral program is progressing well in the U.S. with tower tests of both imaging and non-imaging sensors completed and flight tests planned starting in the end of 1998. Programs are planned or in place to mature multispectral techniques so that they should be available for an advanced eo targeting system integration effort after the year 2000.

5. **Shared Resources**

Future weapons systems' costs need to be significantly lower. Lower costs can be achieved by integrating functions, reducing the cost of components, sharing subsystem components, increasing performance or increasing aircraft survivability. The proposed integrated targeting concept impacts the final system cost using all of the mentioned cost reduction options. The stabilization requirements for an advanced eo targeting system like this should be less demanding and lower cost because the pixel size previously required for a flir to do detection is now much larger since we are doing spectral detection not spatial detection. The pixel size could grow from a foot or two to a meter or two. The short pulse on the laser illuminator also minimizes the need for expensive stabilization since motion is stopped for the duration of the illumination pulse. The use of larger pixels for detection and the use of the shortest wavelength possible while still maintaining eye safe laser illumination allows minimum apertures and aircraft windows to be employed - again a cost saving over current systems. The integration of the radar and EO information provides another source of cost savings. It is possible to detect a potential target with a radar or other cueing sensor that hands off the target detection/location to a attack aircraft only equipped with the EO system. It is also possible to use the EO equipped aircraft for detection/ID/designation and use a third aircraft as the carrier of laser guided or other precision weapons. These targeting concepts could mean that not every airplane in a future wing has both radar and eo but rather a team approach is used. Or, other assets such as unmanned aerial vehicles could be used to cue the attack platform. These concepts should provide for cost effective identification, recognition and targeting.

6. **Risk Assessment**

The chief risks associated with an advanced EO targeting system built on the preceding technologies and concepts are now outlined. ERASER is built up from largely existing laser technologies and the new generation of reflective aperture flirs The primary laser is either a currently available1.06 laser designator which is shifted in wavelength to 1.54 or 1.56 u, or, emerging solid state lasers that offer more power and packaging potential while still providing the multifunctions of 1.06 u ranging, designation and at 1.54 or 1.56 u target illumination. There is low risk in using the available designator laser although medium risk for the wavelength shifting or adding more power needed to achieve longer ranges. The Infrared gated camera system with ERASER is judged low risk. ERASER itself will be demonstrated using largely retrofitted aircraft parts (e.g., apertures, optics) in the next three years which serves as risk mitigation for a system program. Thus ERASER is medium to low risk. Multispectral sensing is also being actively pursued with programs to develop and demonstrate both day only reflective multispectral and day/night thermal multispectral capabilities in brassboard applications. The thermal instrument is planned to fly at both low and high

altitudes to gather sufficient background, target and atmospheric data to support both the system utility assessment as well as system design trades. By 2000 the thermal spectral technology should be mature enough to be considered as a candidate for integration into a targeting system. Day only reflective multispectral technology is judged medium to low risk, but for overall (day and night) use the thermal multispectral capability is still high risk. The Automatic Target Recognition role remains to be mitigated and we expect it to require additional development for the ERASER and Mulitspectral technologies. It is judged medium to high risk. Finally, the integration and demonstration of the whole system is judged medium to high risk.

7. Application & Summary

We believe that the promising developments outlined above should lead to an advanced EO targeting system for a host of applications. In the USAF we believe that our plans will be of primary interest to Air Combat Command in support of such aircraft as the Joint Strike Fighter as well as various upgrades to existing fighter platforms such as F15 or F18. Part of the Air Force vision for future combat identification is for the combination of on and off board sensors used in a system of systems approach with interoperable data links. Additionally the system concepts outlined above should be of interest for use in Unmanned Aerial Vehicles. In summary we believe we are working toward the next generation eo targeting system and that in particular the efforts in ERASER and Multispectral sensing are key mitigators of the risk. We believe that the next generation, advanced EO targeting system can be built and demonstrated using these key technologies which should be available post 2000 time frame. We look forward to integrating these into an advanced system and demonstrating them for the purposes mentioned.

References

1. Enhanced Recognition and Sensing Ladar, E.A. Watson, T.P. Grayson, S.A. Mcdonald, M.D. Nelso, L.J. Barnes, P.F. McManamon, IRIS Symp. June '96 London
2. Enhanced Recognition and Sensing LADAR for Air-to-Ground Combat Indentification, F.R.Kile, R.M. Zumrick, B.K.Karch, Wright Lab, T. Willett, Northrop-Grumman, J. Pryce Veda, IRIS Symp. March '97 Tucson
3. Enhanced Recognition and Sensing LADAR(ERASER), Robert M. Zumrick and Frederick V. Heitkamp, Wright Lab, Combat ID Systems Conference April '97 San Diego
4. Application of Multispectral Infrared Sensing to Long Range Air-to-Ground Target Detection, M.T. Eismann, C.R. Schwartz, J.N. Cederquist, Wright Lab & ERIM, IRIS London, UK, June '96
5. Future Airborne Sensor Systems, R.A. Baker, GEC-Marconi Sensors Ltd., IRIS Symposium June '96 London
6. Navy EO Sensor Testbed Development for Infrared Search and Track, George R. Ax, Jr., BDM Federal, James R. Buss, Office of Naval Research, IRIS Symp. June '96 London

PAPER No: 11

DISCUSSOR'S NAME: R. Sabatini

COMMENT/QUESTION:

What kind of laser source are you using for the LADAR?

AUTHOR/PRESENTER'S REPLY:

Nd: YAG at λ = 1.064 μm.
Current research on Raman-shifted Nd: YAG and Er : Glass at λ = 1.53 μm.

METHODE D'EVALUATION DES PERFORMANCES D'UNE POURSUITE MULTICAPTEUR

M. Desbois

NATO C3 Agency
Command, Control & Sensor Division
P.O. Box 174
2501 CD s'Gravenshage, Nederland

SOMMAIRE

Le propos de cet exposé est de décrire la méthodologie qui a été développée pour l'évaluation des poursuites multicapteur au sein de l'Agence C3 de l'OTAN. Cette méthodologie a originellement été développée dans le but de quantifier les exigences de performances en termes applicables à l'établissement de la situation aérienne renseignée (*Recognized Air Picture RAP*) dans le cadre de la rédaction des spécifications du futur système de commandement et de contrôle des opérations aériennes (*Air Command and Control System, ACCS*) de l'OTAN. Un certain nombre de mesures de performances (MOPs) ont ainsi été définies. Un processus de segmentation des pistes et des trajectoires des cibles poursuivies s'impose comme condition préalable au calcul d'une partie de ces mesures: la description de cette segmen-tation s'inscrit également dans la méthodologie développée, tout comme la mise en œuvre de différents filtres permettant une étude ciblée des performances de poursuite.

La description fonctionnelle de l'architecture de ce système d'évaluation détaille les problèmes temporels et spatiaux ainsi que la manière dont ils ont été traités. De même, différentes approches sont étudiées: la simulation, l'utilisation de données réelles et le calcul analytique.

LISTE DES ACRONYMES

Les acronymes suivants sont utilisés à plusieurs reprises dans ce document.

ACCS	*Air Command & Control System* Système de commandement et de contrôle des opérations aériennes
MOP	*Measure of Performance* Mesure de performance
RAP	*Recognized Air Picture* Situation aérienne renseignée
SFP	*Sensor Fusion Post* Cabine de fusion de données
IDCP	*Identification Data Combining Process* Processus d'identification par combinaisons des données.

1. INTRODUCTION

L'approche retenue lors de l'élaboration de cette méthode était guidée par le désir de maintenir des liens étroits entre les exigences opérationnelles des utilisateurs et les diverses mesures analytiques de performance. Chaque mesure peut ainsi être pondérée selon l'importance qu'elle requiert auprès des différents utilisateurs: en effet, les besoins des "usagés" de la situation aérienne diffèrent considérablement, et il semblait important de ne pas chercher à attribuer une performance globale mais plutôt de présenter les résultats en fonction des différents profils d'utilisateurs.

Cet exposé se concentre sur l'évaluation des performances de poursuite multicapteur dites à "fusion de plots" plutôt qu'à "fusion de piste", mais la méthode s'étend aisément à tous les types de fusion de données. Les exemples choisis concernent l'étage de fusion des détections fournies par les différents capteurs du théâtre d'opération qui est matérialisé par le *"Sensor Fusion Post (SFP)"* dans l'architecture de l'ACCS.

2. METHODOLOGIE

Trois approches sont offertes pour l'évaluation de performances; la simulation, l'utilisation de données réelles et le calcul analytique. Elles correspondent à des besoins différents et peuvent être complémentaires. Indépendamment de l'approche choisie, la première étape consiste à établir un environnement opérationnel dans lequel des exigences de qualité et de détection peuvent être définies.

2.1. Exigences Opérationnelles

La plupart des utilisateurs ne sont pas préoccupés par les performances techniques de la fonction qui élabore la situation aérienne. Ils ont une notion confuse concernant la stabilité des pistes qu'ils reçoivent, le temps de réaction aprés une

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

manœuvre, la création de fausses pistes dans les zones de fouillis, etc. L'expression d'exigences opérationnelles est d'autant plus difficile que les utilisateurs intègrent souvent les problèmes auxquels ils sont confrontés avec un système donné[1]. Les spécifications d'un système reposent ainsi parfois sur la correction de problèmes rencontrés avec le système précédent et non pas à l'exigence opérationnelle originale. L'analyse des spécifications doit ainsi tenir compte du bais naturel qu'ont les opérateurs vis à vis du système en place.

La Figure 2 illustre le concept de l'évaluation de performance d'un système de poursuite multicapteur. La référence est constituée par le scénario tel qu'il a été généré et "joué" pour produire les détections des capteurs. L'objectif du concept et d'analyser quelles sont les performances affectées par quels facteurs, afin de mieux cerner la sensibilité du système à ces différents paramètres. Au cours du processus d'élaboration de la situation aérienne, les capteurs produisent d'abord leur vision de l'environnement, que l'on pourrait appeler la situation mesurée. Puis le pistage produit par filtrage des mesures une situation estimée (dont la qualité est en principe supérieure à celle de la situation mesurée). Enfin, une situation prédite est calculée pour être visualisée par les opérateurs. L'un des objectifs de l'évaluation est de quantifier les facteurs qui limitent ou affectent sensiblement les performances du système. Les méthodes mises en œuvre travaillent soit par comparaison entre les entrées et les sorties (situation prédite par rapport à la référence par exemple), soit par analyse des processus et des modèles impliqués.

2.2. Méthode par Simulation

La Figure 1 illustre la méthode par simulation qui traite la poursuite multicapteur comme une "boite noire", recevant les détections des capteurs en entrée et délivrant des pistes en sortie.

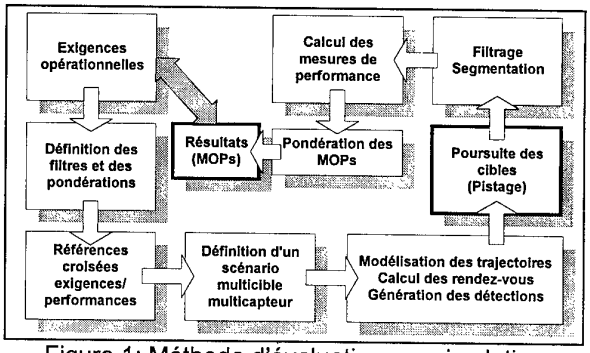


Figure 1: Méthode d'évaluation par simulation

2.2.1. *Scénario*

L' environnement de simulation ne requiert pas un réalisme tactique (représentation exacte d'une menace), mais doit comprendre un certain nombre d'éléments caractéristiques qui constituent les points d'intérêt d'un ensemble d'utilisateurs. Les exigences opérationnelles sont ensuite reliées à une ou plusieurs mesures de performance. Ces références croisées sont établies par l'étude de chaque caractéristique de la situation aérienne et par sélection des performances qui traduisent la qualité de cette caractéristique. L'évaluation de la performance finale sera conduite en fonction de ces références croisées aprés avoir appliqué les pondérations et filtrages définis par les utilisateurs.

De manière générale, un scénario se caractérise par trois composantes:

- Les moyens de détection, à savoir le nombre, le type et les caractéristiques des capteurs, leur position dans l'espace, le recouvrement des couvertures électromagnétiques, le terrain, les moyens de communications utilisés pour acheminer les informations vers le centre de fusion,
- La configuration aérienne des objets aériens (nombre d'aéronefs, type, vitesse, altitude, manœuvres, signature, émission électro-magnétique, séparation des appareils, etc.),
- L'environnement électronique et météorologique, tel que la température, les précipitations, le vent, le brouillage, etc.

Ce que l'on appelle communément un scénario consiste en fait en une famille de cas voisins dont certains paramètres de ces trois composantes varient (mêmes moyens de détection, mêmes aéronefs, conditions climatiques différentes par exemple). Ainsi, on pourra étudier les variations de performance en fonction de l'évolution de certains paramètres.

2.2.2. *Monte-Carlo*

Le processus complet d'élaboration de la situation aérienne est très complexe parce qu'il englobe des

[1] Certains opérateurs disposent de la visualisation des plots synthétiques et peuvent observer des manoeuvres avant que la piste ne rende compte de l'évolution des cibles. S'ils sont en liaison radio avec les pilotes, ils peuvent savoir qu'un chasseur manoeuvre avant même que le système ne le détecte.

sous-processus notoirement différents, tels que la détection, la transmission des données, le filtrage numérique, et la visualisation des pistes. Chaque niveau est une source d'erreur potentielle de telle sorte que l'analyse du processus complet n'est pas déterministe. Devant une telle complexité, il est nécessaire de recourir à des techniques de rejeu d'un même scénario, en appliquant des conditions initiales différentes, ce qui permet d'accroître le degrés de confiance que l'on peut attribuer aux résultats. Les techniques de rejeu permettent de rassembler un nombre plus important d'échantillons d'analyse et ainsi d'estimer, en plus des valeurs de performance, la dispersion des

- Il n'existe pas de données de référence (c'est à dire les trajectoires du scénario)
- Il n'est pas aisé d'effectuer des mesures statistiques (le volume des données reste limité et le rejeu reste trés limité).

Il est possible de pallier plus ou moins la première contrainte en utilisant des aéronefs équipés de moyens de localisation très précis, tels qu'une balise GPS (*Global Positioning System*) embarquée ou une centrale inertielle précise. Ces équipements fournissent des informations qui peuvent être utilisées pour recalculer les trajectoires a posteriori.

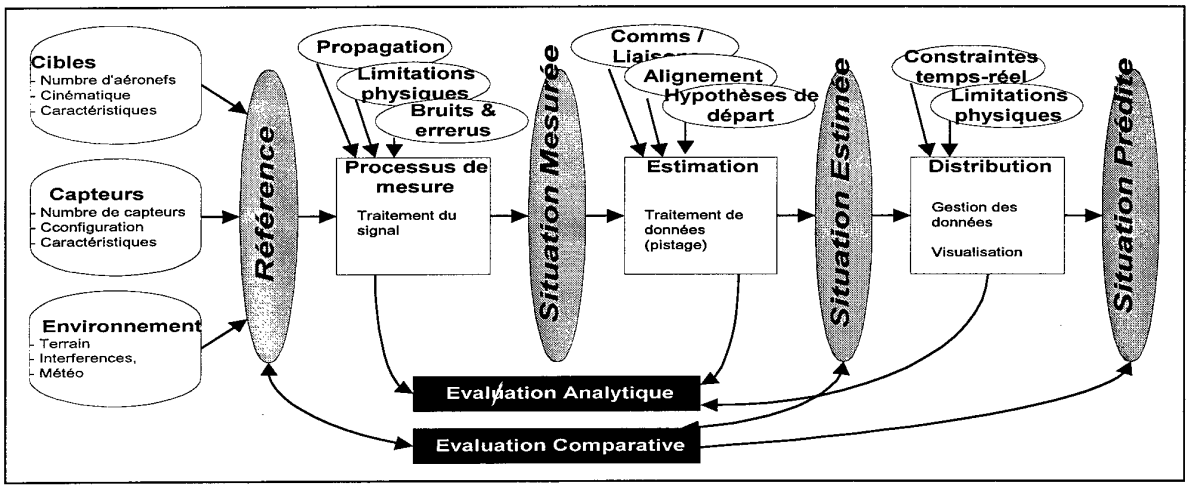


Figure 2: Concept de l'évaluation de performances de poursuite

mesures.

2.2.3. *Utilisation de cette méthode*

Cette méthode permet ainsi non seulement de comparer les performances de différents types de poursuites (voire de différents réglages), mais encore de mesurer à quel point les exigences opérationnelles sont satisfaites et comment la poursuite répond aux besoins des utilisateurs. Elle autorise également l'analyse de phénomènes aux limites et de comportements extrêmes, qui - bien que non réalistes - permet de mieux comprendre le comportement de l'ensemble matériel et logiciel que forme un système de poursuite multicapteur.

2.3. Exécution de données réelles

Les tests et vérifications des algorithmes de poursuites par des moyens de simulation est une étape nécessaire, cependant il est également utile de valider les performances vis à vis d'un environnement réel. La méthode par simulation peut s'appliquer à des données réelles avec les deux contraintes majeures suivantes:

Si l'on dispose d'enregistreurs synchronisés capable de stocker simultanément les détections des capteurs, puis de rejouer ces données, en les mixant éventuellement avec des données simulés (fouillis par exemple), on peut augmenter les échantillons d'analyse.

2.4. Méthode Analytique

La mise en place d'une simulation de poursuite multicapteur, avec rejeux, peut sembler lourde et coûteuse lorsque l'on cherche à obtenir des résultats rapidement ou à analyser un problème particulier. La prédiction des performances par une méthode analytique peut alors s'avérer souhaitable pour estimer les limites d'un ensemble donné de capteurs et d'aéronefs, et pour identifier les paramètres prépondérants. Ces limites ont l'avantage d'être génériques dans la mesure où elles ne dépendent pas d'un scénario particulier (avec les autres méthode, un doute persiste toujours quant à la pertinence du scénario ou des données enregistrées). Les performance obtenues permettent de délimiter le périmètre des possibilités d'un algorithme donné ou d'une configuration de capteurs spécifiques. Bien sûr, il n'y a aucune

garantie que les performances calculées puissent être effectivement obtenues par un système réel. Par contre, si les performances estimées ne satisfont pas les exigences, il est certain que le système ne sera pas conforme, qu'il y a un risque important, et que des solutions alternatives doivent être envisagées.

2.4.1. *Difficulté de la méthode*

L'évaluation des performances d'un système de surveillance est une tâche difficile parce qu'elle comprend des éléments conflictuels. Un système de poursuite est un système stochastique hybride qui traite deux types d'inconnues:

- Discrètes pour ce qui concerne la détection (faux plots, fouillis): la présence ou l'absence d'un écho, de même que l'apparition d'un faux plot peuvent changer radicalement la création et la mise à jour d'une piste.
- Continues quant aux erreurs dynamiques et aux bruits de mesure: les covariances d'erreur sont remarquablement traitées par les algorithmes de filtrage numérique.

2.4.2. *Champ d'application*

L'étude analytique supportée par une modélisation précise du système et des traitements mis en œuvre est souvent nécessaire préalablement à l'investigation d'un nouveau concept.

3. MESURES DE PERFORMANCE

Il n'est pas possible de sélectionner seulement une ou deux mesures pour évaluer les performances d'une poursuite. Les critères d'exigence se traduisent généralement au travers de plusieurs MOPs. Malgré l'interdépendance des mesures, il est souhaitable de définir des MOPs qui rendent compte des performances de certaines sous-fonctions comme par exemple la création automatique de pistes ou la corrélation. Il est important de mesurer le taux de bonnes ou mauvaises associations, mais il ne faut pas que de les erreurs d'associations condamnent l'analyse de la précision en position par exemple (voir le paragraphe 3.1 sur la segmentation).

Mesure	definition
Couverture	Pourcentage moyen de cibles représentées effectivement par un piste parmi l'ensemble des cibles sur chaque intervalle de temps
Vraisemblance	Pourcentage moyen entre le nombre de pistes représentant une cible réelle et le nombre total de pistes
Délai d'initialisation	Délai moyen entre la première détection d'une cible et la confirmation de la piste associée.
Taux de rafraîchissement	Période moyenne de mise à jour des pistes.
Précision en position	Moyenne des différences entre la position des cibles et de la piste qui leur est associée.
Précision en altitude	Moyenne des différences entre l'altitude des cibles et de la piste qui leur est associée.
Précision en cap	Moyenne des différences entre le cap des cibles et de la piste qui leur est associée.
Précision en vitesse	Moyenne des différences entre la vitesse des cibles et de la piste qui leur est associée.
Continuité	Pourcentage du temps pendant lequel une cible est représentée continuement par la même piste au long de son existence.
Stabilité	Moyenne des variations de cap dans le temps
Taux de fausses pistes	Moyenne du nombre de fausses piste crées et maintenues par le système.

Table 1: Liste de mesures de performance

L'une des gageures est de limiter le nombre des MOPs pour conserver une compréhension globale et intuitive de l'analyse. Un jeu de 15 à 20 MOPs permet une granularité d'introspection acceptable. La Table 1 illustre les MOPs les plus couramment utilisées.

Cette liste s'applique essentiellement aux performances de poursuite. La méthodologie générale peut être appliquée à l'examen de processus plus fin tels que la détection, la tenue d'une piste individuellement, les communications, ou la visualisation de la situation aérienne. Des mesures de performances similaires peuvent être analysées sous différents angles: par exemple, le délai d'initialisation peut être mesuré sous un angle opérationnel (délai entre la pénétration d'un aéronef dans une zone de défense et l'affichage de la piste correspondante sur l'écran d'un opérateur) ou technique (délai entre l'instant de détection d'une cible et la création de la piste qui la représente dans la table des pistes).

3.1. Segmentation

Avant d'analyser une piste, il est nécessaire de la décomposer en segments, chaque segment représentant une partie de la piste pendant laquelle celle-ci représente le même aéronef. Ce pré-traitement des pistes et des trajectoires est indispensable à l'étude objective des échanges piste/cible, du calcul des fausses pistes et de l'estimation des précisions en position, vitesse, cap et altitude.

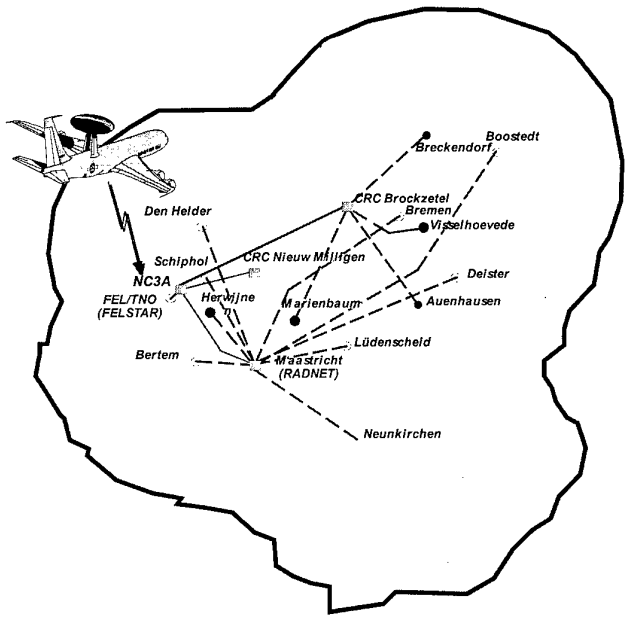


Figure 3: Environnement du Laboratoire ASEP

3.2. Sélection des échantillons par filtrage

Il est clair que différents utilisateurs ont des priorités dissemblables concernant la qualité des pistes. Ces priorités peuvent varier en fonction de paramètres tels que le cap, l'altitude, la vitesse ou l'identité des pistes. Par exemple, l'opérateur d'un système de missiles sol-air considérera les pistes rapides et approchantes comme les pistes les plus importantes en comparaison des pistes qui s'éloignent de sa zone de responsabilité.: ainsi, il pourra avoir des exigences de performance différentes pour ces pistes. Des profiles de filtrage ont donc été mis en œuvre pour permettre une sélection des éléments à prendre en compte pour l'évaluation de performances selon la position géographique des pistes, leur cap, vitesse, altitude, identité, etc.

3.3. Pondération

Les mesures de performance individuelles sont calculées pour chaque mise à jour de piste, puis sont globalisée en utilisant une moyenne simple plutot qu'une moyenne quadratique. Cette manière de procéder est plus représentative, principalement pour les scénarios comprenant des cibles trés manœuvrantes. L'utilisation d'une moyenne quadratique tend à rendre trop prépondérants les mauvais résultats obtenus pendant les virages et à fausser le résultat global. En effet, la brièveté des manoeuvres (du moins dans un environnement militaire) conduit à considérer celles-ci comme une perturbation après laquelle la piste doit rapidement reconverger vers la cible (la précision de la piste pendant la manœuvre n'a pas d'intérêt opérationnel).

Afin d'éviter que quelques mesures particulièrement mauvaises ne viennent dénaturer le résultat global, une pré-sélection est automatiquement appliquée avant d'effectuer la pondération. Habituellement, 5% à 10% des plus mauvaises mesures sont rejetées du calcul de la moyenne de chaque cycle. Cette élimination reflète la réalité dans la mesure où les opérateurs concentrent justement leur attention sur les parties difficiles (virages, formations sérrées, etc.) et peuvent agir manuellement pendant ces instants critiques.

4. LABORATOIRE ASEP

La Division commandement, contrôle et capteurs de l'Agence C3 de l'OTAN (NC3A) a développé une plate-forme ASEP (*Air Surveillance Exploratory Prototype*). Ce laboratoire dispose d'un environnement intégré permettant l'expérimentation des fonctions de surveillance et d'identification avec des données réelles ou simulées. Des lignes de communication acheminent les détections temps-réel en provenance de 15 capteurs civils ou militaires, les pistes délivrées par les moyens d'alerte aéroportés (NAEW), les situations aériennes et plans de vols de deux centres de

contrôle (CRC) et du centre d'Eurocontrol installé à Maastricht comme le montre la Figure 3.

Les moyens de simulation comprennent le générateur de scénario SPACE (*Scenario Preparation and Coverage Evaluation*) et le modèle MOCCA (*Model of Command and Control of Air Defence Systems*). Enfin, les logiciels MUTANT (*Multi-Sensor Tracking Analysis Tool*) et SMART (*Segmentation and MST Analysis and Reporting Tool*) de calcul des performances et l'outil de visualisation TSDS (*Track and Sensor Display System*) complètent la panoplie des outils informatiques utilisés pour l'évaluation des performances de poursuite qui composent le projet RAPACE (*RAP Analysis & Comparative Evaluation*) illustré par la Figure 4.

L'outil de génération de scénario SPACE permet de définir le réseau de capteurs en s'appuyant sur l'ensemble des couvertures électromagnétiques qui peuvent être calculées et visualisées, pour n'importe quelle tranche d'altitude, en tenant compte de la nature du terrain (zone urbaine, forêts, désert, mer, etc.) et du relief. Les trajectoires des aéronefs peuvent être ainsi crées et modifiées en ayant connaissance des capacités de détection du système.

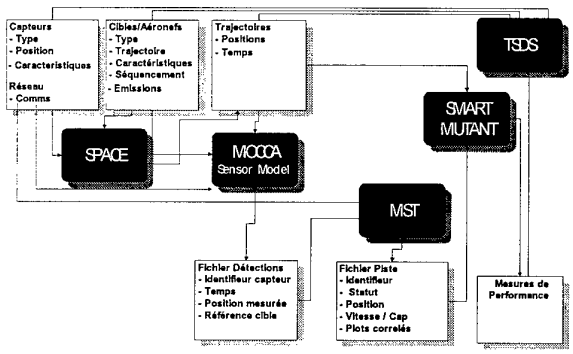


Figure 4: Projet RAPACE

Les scénarios créés par SPACE sont déroulés par l'outil MOCCA qui intègre des modèles fins pour un grand nombre de capteurs. Ces modèles simulent précisément les performances de capteurs à impulsion, impulsion Doppler, forme d'onde continue, 2D et 3D, couvrant l'ensemble des caractéristiques des radars primaires et secondaires, fixes et aéroportés, actifs et passifs, susceptible d'être réellement utilisés.

L'outil MOCCA intègre également un modèle de poursuite multicapteur. Les algorithmes mis en œuvre sont équivalents à ceux utilisés dans les systèmes modernes de contrôle aérien civil ou militaire.

Les fichiers de détection et ceux des pistes sont conservés dans une base de donnée accessible par l'ensemble des outils. Ils peuvent également être enregistrés sur des fichiers dans un format défini par l'Agence. Cette possibilité autorise l'exécution de scénarios réels ou simulés sur des plate-formes différentes, par exemple des bancs de test industriels, des systèmes en cours d'implémentation ou des systèmes existants. Les performances peuvent ainsi être évaluées à l'Agence ultérieurement.

L'usage d'une base de données permet l'utilisation d'une application graphique TDSD écrite en GSQL (*Graphical Sequential Query Language*) qui extrait et manipule directement les informations de la base. Cette approche facilite grandement la mise en œuvre de tris et filtres appliqués aux données.

Les mesures de performances peuvent être visualisées graphiquement par des histogrammes tels que ceux reproduits à la Figure 5. Ces résultats représentent les performances obtenues par une poursuite sur le même scénario aérien, avec des environnement électromagnétiques différents.

Un indice de satisfaction peut être attribué à chaque mesure de performance (de 1 à 10 par exemple) de telle sorte que l'ensemble des mesures peut être ensuite ramené à une note globale par pondération des indices en fonctions des critères de pertinence des performances vis à vis de chaque opérateur. Cette méthode permet notamment de comparer rapidement plusieurs techniques lorsque l'analyse détaillée de chaque mesure peut s'avérer extrèmement fastidieuse.

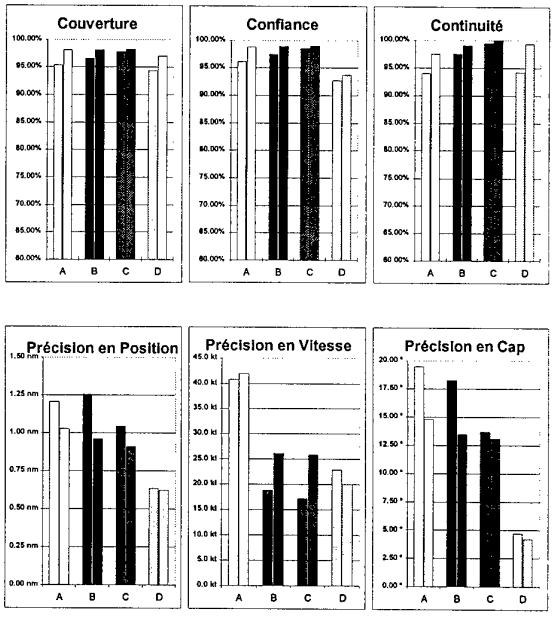


Figure 5: Histogrammes de mesures

5. UTILISATIONS RECENTES

Le projet RAPACE a été mis en œuvre pour l'étude des concepts modernes de fusion de données et d'identification. Le processus de sélection des soumissionnaires pour le projet d'évaluation et de démonstration de l'IDCP (*Identification Data Combining Process*) et pour l'appel d'offre de l'ACCS LOC1, ainsi que les études d'intégration multicapteur pour la rénovation de la capacité d'alerte aéroportée de l'OTAN sont des illustrations récentes des utilisations de ce projet.

6. DISCUSSION SUR LES PERFORMANCES

La question de l'intégration de capteurs multiples dans un même système de fusion de données est intimement liée avec la notion de performance. La fusion de données multicapteur apporte en général une solution là où un système monoradar ne répond plus aux exigences opérationnelles, ou du moins, lorsqu'il n'y répond plus sous certaines conditions. Un autre facteur d'intégration est de permettre une dégradation progressive en cas de panne ou de dysfonctionnement d'un élément du réseau, ou encore une meilleure résistance aux contre-mesures. On attend donc en général d'un système multicapteur de meilleures performances pour assurer les mêmes fonctions. Cependant, il est clair que l'intégration multicapteur autorise également de nouvelles fonctionnalités dont le gain est plus difficile à mesurer.

La fusion multicapteur illustre également la notion de "système". En fait, à une période où les coûts de dévelopement logiciel sont en baisse, alors qu'inversement les recherches et développements sur des capteurs nouveaux coûtent de plus en plus cher (l'amortissement des études est critique lorsque le nombre de capteur produits reste limité), il est tentant de pallier aux problèmes de détection par des techniques informatique. La mise à profit de la probabilité de détection cumulée multicapteur, la diminution de la résolution obtenue par le répartissement géographique des capteurs, où encore l'accroissement le la cadence d'échantillonage sont autant de problèmes de détection résolus en partie par la fusion de données. Le problème reste à traduire ces notions en terme de performances identifiables et mesurables.

Dans une élaboration traditionnelle, la poursuite multicapteur est décrite comme un processus purement passif à boucle ouverte, traitant successivement les mesures effectuées sur l'environnement. En fait, dans un contexte élargi, il convient d'envisager de fermer la boucle, c'est à dire d'agir sur les capteurs pour optimiser leur perception de l'environnement. Le contrôle intelligent des capteurs à partir des résultats de poursuite devrait permettre là encore d'améliorer la performance globale du système en concentrant l'énergie et en commandant des formes d'onde particulières sur les cibles prioritaires par exemple, et en ayant un système trés adaptatif à son environnement.

Des travaux sont donc nécessaires pour prototyper un système réactif et mettre en oeuvre un environnement de simulation qui autorise les traitements de poursuite à agir sur les modèles de capteurs. Dans un tel système, l'analyse de la performance aura une importance encore accrue puisque cette performance sera à la fois un but et un moyen d'optimisation du système.

7. REFERENCE

Blom "*TRAQUME, a Method of Measuring the Operational Quality of Tracks for ATC*", NLR, July 1989

Y.Bar-Shalom, "*Performance Evaluation of a cascaded logic for track formation in Clutter*", IEEE Trans. AES, AES-25, Nov 1989

J.Llinas and E.Waltz "*Multi-Sensor Data Fusion*", Artech House, Norwood MA, 1990

J.Bravin "*A Method of Evaluating Multi-Sensor Tracking Performance based on Requirements of the Recognized Air Picture Users*", STC-TN-283, June 1990 (NC)

S.Mori, K-C.Chang, C.Chong, K.Dunn "*Tracking Performance evaluation: Track Accuracy in Dense Target Environment*", SPIE's 1990 Technical Symposia on Aerospace Sensing, Orlando, 1990

J.Llinas "*Assessing the Performance of Multi-Sensor Fusion Systems*", SPIE Vol 1611, 1991

S.Mori, K-C.Chang, C-Y.Chong, K.Dunn "*Performance Analysis of Optimal Data Association with Applications to Multiple Target Tracking*", Chapter 7 of "Multitarget-Multisensor Tracking, Applications and Advances", Artech-House, 1992

J.Roy and É.Bossé "*Definition of a Performance Evaluation Methodology for sensor Data Fusion Systems*", DREV-R-9423, Sept. 1995 (NU)

PAPER No. 13

DISCUSSOR'S NAME: M. Balci

COMMENT/QUESTION:

Have you considered any civilian standards with a view to using them in military service too? As an example, the ASTERIX format can be mentioned.

AUTHOR/PRESENTER'S REPLY:

The ASTERIX format has been foreseen and is already in use for 60th ATC and military radars as far as surveillance is concerned. There is still some discussion on how to incorporate specific military requirements such as tactical data links, sensor control orders, etc..... Airborne sensor systems such as Joint STARS or AWACS do not have equivalent civilian systems to share data formats with.

DISCUSSOR'S NAME: M. Balci

COMMENT/QUESTION:

What is the relationship between the performance measurements and the overall system performance?

AUTHOR/PRESENTERS REPLY:

Means of performance can be either technical or operational, and can apply either one sub-function function, or on the whole system, according to the level of detail required.

DISCUSSOR'S NAME: R. Hogendoorn

COMMENT/QUESTION:

5 to 10% of the measurements are not taken into account in the performance measurements. Are they nevertheless considered as, e.g. a measure of robustness?

AUTHOR/PRESENTER'S REPLY:

The discarded fraction corresponds to targets performing steep manoeuvres that do not last very long, so they are not of great interest for an operator. Furthermore, this fraction consists of false or random tracks that have bad MOPs and, thus, disturb the MOP figures. However, these tracks are looked at to verify that they belong to the above two classes.

PAPER No. 13

DISCUSSOR'S NAME: J.L. Maillart

COMMENT/QUESTION:

1. J'imagine que la modélisation des performances dans la méthode analytique nécessite de connaître exactement le détail des algorithmes réels ou de la simulation: dans le système développé. Est-ce que ce sont les mêmes personnes qui mettent au point à la fois les algorithmes et leur modèle de performances, ou bien des equipes différéntes?

2. Avez-vous pu comparer les résultats obtenus, pour les différentes couches de traitement, entre la méthode de simulation et la méthode analytique?

(1. I imagine that performance modelling in the analytical method requires precise knowledge of the real algorithms or of the simulation: in the developed system. Do the same people design the algorithms and their performance models, or are there two different teams?

2. Have you been able to compare the results obtained, for different layers of treatment, by the simulation method and the analytical method?)

AUTHOR/PRESENTER'S REPLY:

1. Ce sont les mêmes personnes qui pilotent la méthode analytique sachant que les niveaux de détail réclament souvent une expertise du domaine concerné qui est, en général, fourni par des experts.

2. Les travaux conçernant l'approche analytique ne sont pas suffisament avancés pour avoir un recul suffisant concernant la possibilité de confronter les méthodes analytiques en confrontation avec les résultats de simulation ou ses données réelles.

(1. The same people work on both methods, although the detail often requires special expertise, which is provided by experts.

2. Work on the analytical approach is not advanced enough to give any perspective on the way in which analytical methods could be compared with the results of simulation or its real data.)

Reconnaissance de scènes dynamiques à partir de données issues de capteurs : le projet PERCEPTION

Catherine Tessier
Onera - Cert
2 avenue Édouard-Belin BP 4025
31055 Toulouse Cedex
France

20 août 1997

Résumé

L'objectif du projet Perception est d'étudier et développer des méthodes permettant, à partir de l'observation de l'environnement par différents capteurs, d'élaborer et de mettre à jour une représentation du monde réel propre à faciliter une prise de décision. Le projet contribue ainsi à répondre de manière générique aux besoins opérationnels de veille, surveillance, renseignement, reconnaissance, que ce soit dans le cadre de systèmes autonomes ou de systèmes d'aide à la décision, dont le fonctionnement global s'inscrit dans le cadre d'une boucle fermée perception – décision – action.

La fonction de perception nécessite une chaîne de traitements qui s'étend des capteurs jusqu'à une représentation sémantiquement riche de la situation. Cette chaîne est rebouclée pour permettre d'effectuer une gestion des ressources de perception, et elle fait appel à des techniques tant numériques que symboliques. Elle est organisée autour de quatre grands modules : les Traitements Numériques ont pour objectif de délivrer des objets reconnus dans l'environnement observé, à partir des signaux fournis par les capteurs ; les Traitements Symboliques ont pour but d'élaborer, à partir des objets délivrés par les Traitements Numériques, la ou les situations en cours dans l'environnement, en vue de les présenter à un niveau décisionnel ; le Modèle d'Environnement construit et met à jour tous les éléments pertinents pour la représentation de l'environnement perçu ; La Gestion du système de Perception répond aux requêtes de compléments d'informations émises par les Traitements Symboliques et planifie la prise de nouvelles informations.

1 Introduction

L'objectif du projet Perception [1][1] est d'étudier et développer des méthodes permettant, à partir de l'observation de l'environnement par différents capteurs, d'élaborer et de mettre à jour une représentation du monde réel propre à faciliter une prise de décision. Le projet contribue ainsi à répondre de manière générique aux besoins opérationnels de veille, surveillance, renseignement, reconnaissance, que ce soit dans le cadre de systèmes autonomes ou de systèmes d'aide à la décision, dont le fonctionnement s'inscrit dans le cadre d'une boucle fermée perception – décision – action.

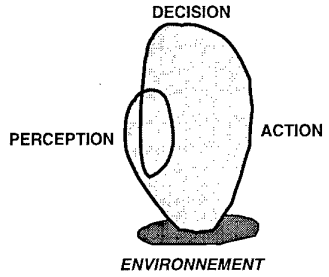


FIG. 1 – *la boucle perception – décision – action.*

La fonction de perception élabore et met à jour une représentation sémantiquement riche des données issues de différents capteurs (caméras noir et blanc et couleur, caméras infrarouges, radar, éventuellement capteurs "humains") disposés dans un environnement dynamique. Plus précisément, il s'agit de fournir des renseignements sur les *actions* des objets en mouvement, et éventuellement sur leurs intentions, en s'appuyant sur des modèles spatio-temporels de haut niveau [2].

1. consulter
http: //www.cert.fr/francais/dera/dera/PERCEPTION/perception.html

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

La chaîne de traitements mise en œuvre s'étend des capteurs jusqu'à une représentation symbolique de la situation, le rebouclage permettant d'effectuer une gestion des ressources de perception. Cette chaîne est organisée autour de quatre grands modules :

• les Traitements Numériques ont pour objectif de détecter les objets en mouvement dans l'environnement, à partir des signaux fournis par les capteurs, et de les reconnaître.

• les Traitements Symboliques ont pour but d'élaborer, à partir des objets délivrés par les Traitements Numériques, la ou les situations en cours dans l'environnement, en vue de les présenter à un niveau décisionnel.

• le Modèle d'Environnement construit et met à jour tous les éléments pertinents pour la représentation de l'environnement perçu, en fonction des résultats fournis par les Traitements Numériques et Symboliques.

• la Gestion du système de Perception (rebouclage) répond aux requêtes de compléments d'informations émises par les Traitements Symboliques, en utilisant le Modèle d'Environnement et les Traitements Numériques. Il planifie la prise de nouvelles informations.

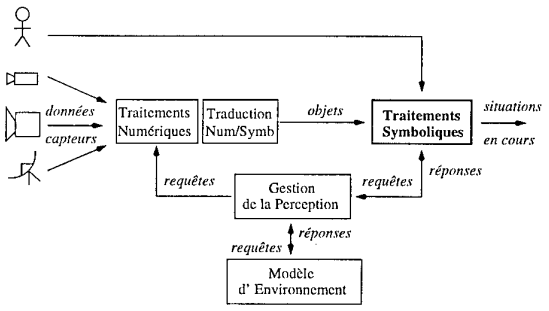


FIG. 2 – *la fonction de perception*

2 Les scénarios de référence : les données VIGILE

La base de données VIGILE, constituée dans le cadre du projet, se place dans le cadre général de la surveillance d'une zone semi-urbaine par un système autonome mettant en œuvre plusieurs capteurs hétérogènes. D'un point de vue pratique, des scénarios ont été joués sur un parking entouré de bâtiments et de végétation et les données ont été recueillies par les capteurs suivants : au sol, un radar 94 GHz, une caméra PtSi 3-5 μm en champ moyen, une caméra analogique noir et blanc en champ large et une caméra CCD couleur en champ large ; en surplomb de la scène observée, une caméra PtSi 3-5 μm en champ étroit, une caméra bi-bande 3-5 μm et 8-12 μm en champ large, deux caméras couleur en mode stéréoscopique et champ large, une caméra CCD noir et blanc en champ restreint. Les scénarios ont été joués de jour et de nuit.

Les scénarios sont de trois types :

- "feu" : ronde de surveillance d'un véhicule de pompiers, déclenchement d'un feu par un pyromane, extinction du feu par les pompiers ;

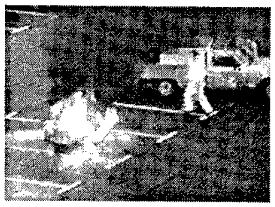


FIG. 3 – *une scène du scénario "feu"*

- "circulation" : interaction de véhicules et de piétons donnant lieu à différentes occultations ;

- "agression concertée" : encerclement d'un véhicule par plusieurs autres véhicules et piétons, mise en évidence d'un comportement collectif agressif.

L'objectif de ces scénarios est d'illustrer la capacité de la chaîne de traitements à prendre en compte des points durs (saturation de certains capteurs, occultations intermittentes, passagères ou définitives d'objets, conflits de plans, détection d'un comportement global à partir de comportements individuels...) tout en demeurant proche d'une certaine réalité opérationnelle. Même si VIGILE peut apparaître, à première vue, comme une base de données "jouet", un examen plus approfondi en regard de scénarios opérationnels montre qu'une majorité de ses caractéristiques sont pertinentes : nature des objets, types de scénarios, distance d'observation, dynamique d'évolution, masquage des objets par des éléments du décor ou par de la fumée. Le fait que les capteurs soient fixes est compensé par la simulation de prise d'informations partielles (sur des fenêtres d'images ou à partir de capteurs à champ restreint).

3 Les Traitements Numériques (TN)

Les Traitements Numériques ont pour objectif de détecter et de reconnaître les objets en mouvement dans l'environnement, à partir des signaux fournis par les capteurs. Dans un premier temps, le capteur radar n'est pas considéré.

3.1 Calibration

Les signaux provenant de plusieurs caméras hétérogènes et non colocalisées, la première étape, pour chaque scénario, est une étape de calibration géométrique qui permet d'associer tout point d'une image 2D à son homologue dans le repère 3D de la scène observée, dont on dispose d'un plan coté. Cette mise en correspondance est réalisée par la méthode de Tsai [3].

3.2 Détection des objets

On ne s'intéresse qu'aux objets en mouvement car eux seuls apportent, au cours du temps, une information nouvelle permettant de mettre à jour la connaissance déjà établie sur la scène.

La détection est effectuée par collaboration entre les algorithmes de détection d'objets en mouvement (algorithmes "dynamiques" , fondés sur le calcul du gradient spatio-temporel de plusieurs images successives) et les algorithmes classiques de détection (algorithmes "statiques" fondés sur la segmentation de l'image courante en régions de radiométrie homogène). L'intérêt de cette coopération est d'allier la bonne précision obtenue par le détecteur de contours avec la bonne caractérisation obtenue par la segmentation en régions, le but essentiel étant d'extraire les informations numériques qui vont permettre de reconnaître un objet parmi un ensemble prédéfini d'objets.

Une opération de détection, qu'elle soit en mode infrarouge ou couleur, est activée par une requête émise de GP vers TN. Cette requête contient la plage temporelle à explorer, la zone d'intérêt à scruter dans les images, les capteurs à sélectionner.

Détection en mode infrarouge

La première phase du traitement consiste à détecter un éventuel changement radiométrique dans l'image courante par rapport à l'image précédente ou à l'image suivante. Ce changement, appelé *phénomène*, est uniquement de nature radiométrique. Si un tel changement est décelé, on calcule le gradient spatio-temporel de trois images consécutives I_{t-1}, I_t, I_{t+1} , centrées sur l'image courante :
$GST(I_{t-1}, I_t, I_{t+1}) = inf(|I_{t-1} - I_t|, |I_{t+1} - I_t|)$.
Ce gradient est ensuite filtré par un seuillage par hystérésis (le seuil a une valeur haute et une valeur basse). Le résultat de ce seuillage est une image binaire qui est ensuite filtrée par une fermeture morphologique, qui permet de connecter des régions très voisines mais non connexes initialement à cause notamment du bruit dans les images. Chaque région ou composante connexe correspond à un phénomène détecté.
Dans le cas des images VIGILE, le contraste thermique est suffisant entre les corps froids des objets du décor et les corps chauds en mouvement pour que les phénomènes détectés correspondent avec une quasi certitude à des objets en mouvement.

La seconde phase consiste à affiner le résultat précédent en appliquant sur l'image binaire une méthode de croissance de régions. Cette étape permet de déterminer avec précision la forme des corps chauds et en mouvement.

Les régions ainsi déterminées sont étiquetées et caractérisées par un ensemble de paramètres radiométriques et morphologiques, en vue de l'étape de reconnaissance des objets.

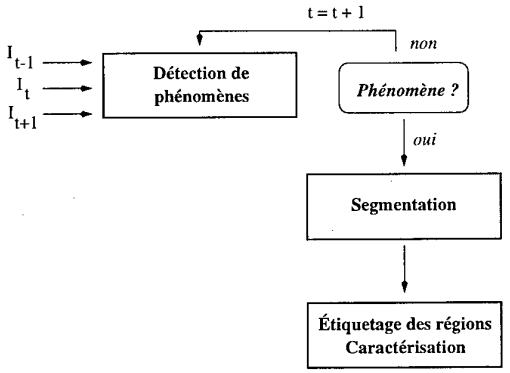


FIG. 4 – *détection en mode infrarouge*

Détection en mode couleur

La détection de phénomènes est réalisée selon le même principe que précédemment, avec un calcul de gradient spatio-temporel prenant en compte les trois composantes rouge, vert, bleu :
$GST(I_{t-1}, I_t, I_{t+1}) = sup_{c=R,V,B}[inf(|I^c_{t-1} - I^c_t|, |I^c_{t+1} - I^c_t|)]$.

En revanche, dans le cas des images couleur, on ne peut plus affirmer qu'un phénomène corresponde toujours à un objet en mouvement. La méthode appliquée est la méthode de détection des contours d'objets mobiles d'Orkitz [4], fondée sur le calcul d'un gradient spatio-temporel à partir d'un triplet d'images successives : si A_{t-1}, A_t et A_{t+1} sont les amplitudes respectives du gradient des trois images, les frontières en mouvement sont mises en évidence par le gradient $G_t(x, y)$:
$G_t(x, y) = Max[A_{t-1}(x, y); A_t(x, y); A_{t+1}(x, y)]$

$-Max[A_{t-1}(x,y); A_{t+1}(x,y)]$
Les contours sont ensuite fermés, puis l'étiquetage des composantes connexes et la fusion des régions adjacentes sont réalisés. Comme précédemment, ces régions sont étiquetées et caractérisées.

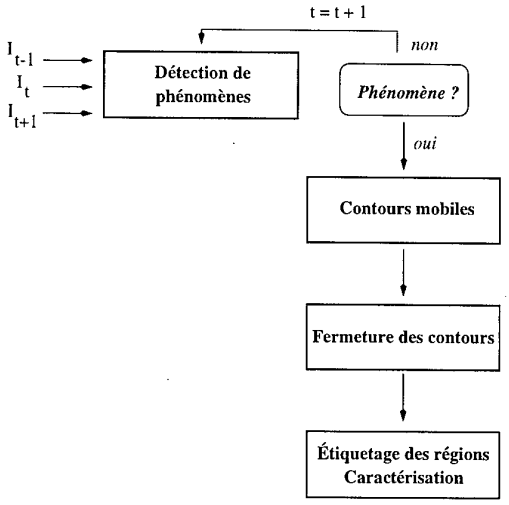


FIG. 5 – *détection en mode couleur*

Classification et pistage des objets

La *classification* est réalisée essentiellement sur les critères d'élongation et d'orientation des régions. Elle s'appuie sur des valeurs de référence, apprises sur la base d'images VIGILE et correspondant aux deux classes d'objets d'intérêt : les véhicules et les piétons.
De manière générale, les piétons sont bien reconnus en mode infrarouge ; en ce qui concerne les véhicules, la classification est meilleure en mode couleur (en l'absence d'occultations), quoique plus délicate à effectuer.
En sortie de classification, une *fiche* F_n est émise, avec les renseignements suivants : type d'objet, vraisemblance du type, date t_n de détection, position dans l'environnement, erreur sur la position, monochromacité, couleur, rotondité, élongation, orientation principale, amplitude et orientation de la vitesse.

Le *pistage* a pour but de chaîner dans le temps les objets observés : un chaînage d'objets est une *piste.* Chaque arrivée d'un nouvel objet permet soit de reconnaître la continuation d'une piste existante et donc de la préciser, soit d'initialiser une nouvelle piste.
Un objet est associé à une piste existante si ses attributs (les différents champs de la fiche F_n) sont compatibles avec ceux de la piste. Dans le cas particulier des images VIGILE, on tient compte du fait qu'un objet ne peut changer de type (un piéton ne se transforme pas en véhicule) – mais que le classifieur peut commettre des erreurs –, et que des attributs de type couleur ou forme peuvent varier (en fonction de l'éclairage par exemple). Les incertitudes d'association sont modélisées dans le cadre de la théorie des possibilités.
Exemple :

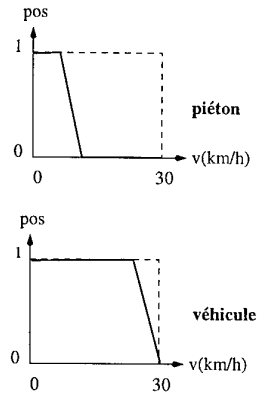


FIG. 6 – *distribution de possibilités sur la vitesse des objets*

Étant donné un objet o et une piste p, le degré de possibilité de l'association de o à p est le minimum des degrés de possibilité d'association compté sur l'ensemble des attributs. On retient les associations objet - piste dont la possibilité est supérieure à un seuil donné.

4 Le Modèle de l'Environnement (ME)

Le Modèle de l'Environnement du système de perception est constitué de l'ensemble des connaissances évolutives relatives à l'environnement observé dont la mise à jour incombe au système. C'est également une banque de données dotée d'outils d'interrogation et de mise à jour au service des autres modules du système.

Le *modèle géométrique* est la partie du modèle d'environnement qui décrit les attributs géométriques de position et de forme des objets observés. Dans le cas des scénarios VIGILE, il décrit les éléments du décor (ces connaissances peuvent être imprécises et donc être mises à jour au fur et à mesure que les capteurs acquièrent de l'information), ainsi que les véhicules et piétons. Chaque objet est constitué des éléments suivants : un identificateur (identique à celui donné par TN pour les objets identifiés par ce module), un type, des attributs géométriques (points tridimensionnels, arêtes, faces, position dans un repère donné) et

non géométriques (couleur, température), et un modèle prédictif qui permet de fournir une description de l'objet en l'absence d'informations de mise à jour. Un historique de l'évolution des objets et de leurs attributs au cours du temps est conservé.
Le modèle permet en particulier de calculer les *vues attendues*, c'est-à-dire l'estimation de ce qu'un capteur donné perçoit à un instant donné.

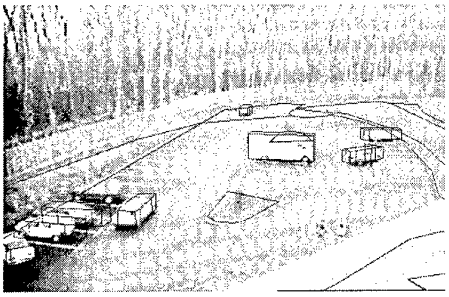


FIG. 7 – *le modèle géométrique superposé à une image*

En tant que base de données, le modèle d'environnement peut effectuer :

- des opérations de *mise à jour* (modification des attributs des objets existants, création de nouveaux objets) en fonction des fiches délivrées par TN ;
- des opérations de *d'interrogation* en fonction des requêtes émanant du module de Gestion de la Perception. L'interrogation peut porter sur tous les attributs d'un objet, un attribut particulier ou un attribut particulier daté.

5 Les Traitements Symboliques (TS)

À partir des fiches F_n issues de TN, les Traitements Symboliques [5], [6], [7] élaborent la ou les situations en cours dans l'environnement observé, c'est-à-dire émettent des hypothèses sur ce que *font* les objets. Les modèles de référence utilisés sont des *activités prototypes*, ensembles de propriétés attendues pour les objets, décrites par des cubes logiques (conjonction de formules logiques atomiques [8]), et des *plans prototypes* représentés par des réseaux de Petri interprétés [9] dont les places sont associées aux activités ; soit $\mathcal{P}$ l'ensemble des plans prototypes.

Une fonction de traduction numérique/symbolique transforme d'abord les attributs des objets des fiches F_n en

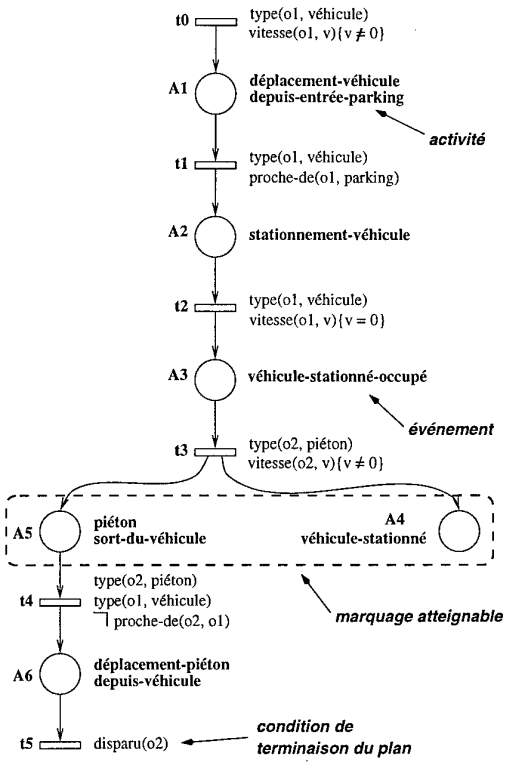


FIG. 8 – *le plan prototype "arrivée-en-véhicule"*

conjonction de propriétés logiques.

La *situation courante* S_n à l'instant t_n est un ensemble de *P-situations* $(P_{i,m_i,n})$, définies comme des réseaux de Petri marqués P_i de $\mathcal{P}$; un plan prototype P_i peut être marqué par le marquage m_i à l'instant t_n si les propriétés de certains objets de F_n correspondent à l'interprétation de certaines places de P_i.
Étant donné la fiche F_{n+1} et la situation courante S_n, l'élaboration de la situation courante S_{n+1} à l'instant t_{n+1} consiste à effectuer une mise en correspondance des propriétés des objets contenus dans F_{n+1} avec les marquages atteignables $m_i + k$ des P-situations $(P_{i,m_i,n})$ (qui correspondent à la continuation de P-situations existantes – $(P_{i,m_i+k,n+1})$) ; ou, si certains objets n'ont pas pu être mis en correspondance de cette façon, à créer de nouvelles P-situations $(P_{j,m_j,n+1})$.
Un objet donné pouvant être associé à plusieurs P-situations, S_{n+1} est l'ensemble des différentes hypothèses de plans susceptibles d'être en cours dans l'environnement à l'instant t_{n+1}.

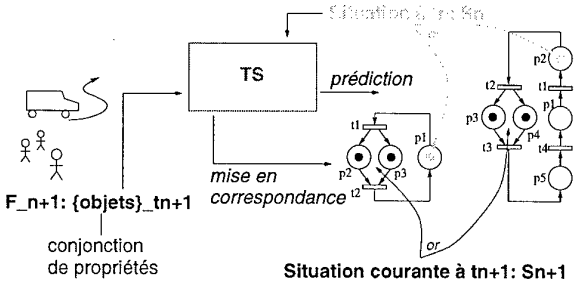


FIG. 9 – *le module de Traitements Symboliques*

Afin d'affiner l'élaboration des situations, des requêtes d'informations complémentaires à celles qui figurent dans les fiches sont émises de TS vers GP.

6 La Gestion de la Perception (GP)

Le module de Gestion de la Perception a un double rôle :

- il reçoit les requêtes de TS, interroge ME et redirige les réponses vers TS ; il est en fait l'interlocuteur unique des autres modules.
- il élabore les stratégies de prise d'informations pour TN, selon les besoins de TS et en fonction de la dynamique propre de l'environnement.

En ce qui concerne le second point, il s'agit de générer des *itinéraires de prise d'information* selon deux modes : un mode de réponse à des requêtes de TS pour lesquelles ME n'avait pas l'information ou une information trop ancienne ; un mode de surveillance, dans lequel GP cherche à mettre à jour ME en tenant compte des évolutions possibles de l'environnement et de l'*érosion des croyances* sur cet environnement (plus le temps passe, plus les parties de l'environnement qui ne sont pas observées sont susceptibles de changer).
Les itinéraires dépendent donc de l'état courant de l'environnement : une distinction explicite est faite entre un degré de confiance représentant une évaluation de la pertinence de l'estimation courante de l'état de l'environnement d'une part, et l'imprécision de cette estimation d'état, qualifiée par une mesure de probabilité, d'autre part. L'évolution des croyances au cours du temps est alors représentée formellement selon des principes de persistance maximale et d'érosion du degré de confiance. Les règles d'érosion permettent alors de définir des stratégies de perception.
Les zones d'intérêt possibles qui sont identifiées au niveau du modèle de l'environnement (calcul des vues attendues) permettent à GP de désigner des *fenêtres* d'intérêt dans le champ de vision des capteurs [10]

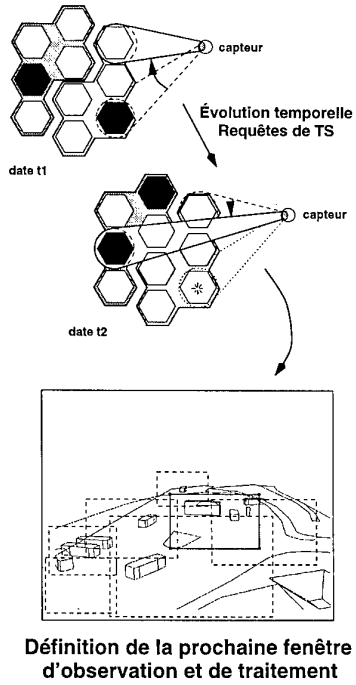


FIG. 10 – *gestion de la prise d'information*

7 Traitement d'un scénario

Supposons pour simplifier que l'ensemble $\mathcal{P}$ des plans prototypes ne contienne que trois éléments : *arrivée-en-véhicule* (voir Figure 8), *départ-en-véhicule* et *déplacement-piéton* (voir Figure 12).

Le modèle d'environnement est supposé connaître le décor et les objets immobiles (véhicules stationnés par exemple) avant le démarrage effectif du processus de surveillance.

• À la date t_1, TN fournit la fiche F_1, qui contient un nouvel objet $P1$ tel que *type(P1, piéton), vitesse(P1, 4km/h).*

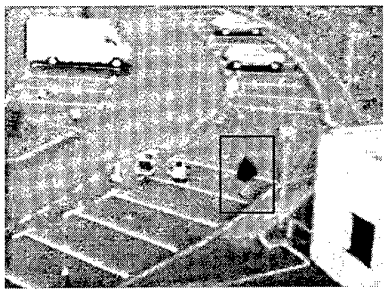


FIG. 11 – *image1, origine de la fiche* F_1

Aucune P-situation n'a encore été créée ; les plans candidats, étant donné cette seule information (piéton en déplacement) sont les trois plans de $\mathcal{P}$. Pour obtenir des informations complémentaires, TS envoie alors des requêtes vers GP pour demander s'il existe un véhicule dans la direction de marche du piéton ou proche du piéton (plan *départ-en-véhicule*) ou s'il existe un véhicule dont s'éloigne le piéton (plan *arrivée-en-véhicule*). ME, sollicité à son tour par GP, renvoie les réponses *type(V1, véhicule), vitesse(V1, 0), se-rapproche-de(P1, V1)* et *type(V2, véhicule), vitesse(V2, 0), se-rapproche-de(P1, V2)* à la première requête et "échec" aux deux autres. Les P-situations constituant la situation S_1 peuvent alors être construites : *arrivée-en-véhicule* est rejeté, tandis que *déplacement-piéton* et *départ-en-véhicule* sont marqués respectivement en A1 (activité "déplacement-piéton") et A2A1 (activités "déplacement-piéton-vers-véhicule" et "véhicule-stationné").

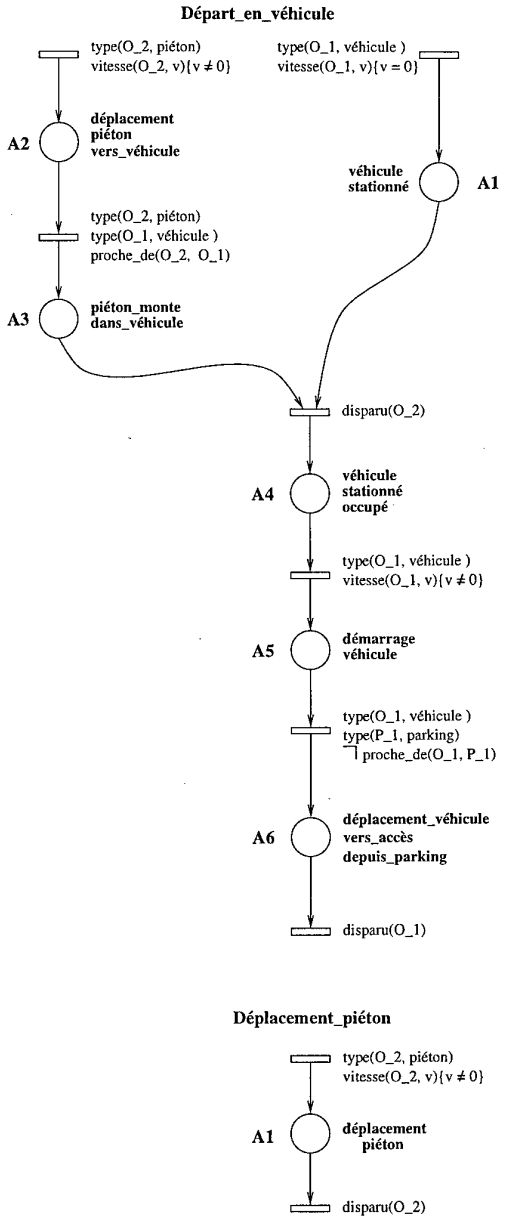


FIG. 12 – *les plans prototypes* départ-en-véhicule *et* déplacement-piéton

• À la date t_2, avec $t_2 = t_1 + 4s$, TN fournit la fiche F_2, qui contient toujours l'objet $P1$ (qui a été pisté) et qui est maintenant tel que *proche-de(P1, V1).*
La situation S_2 est donc la suivante : l'activité "déplacement-piéton" du plan *déplacement-piéton* est toujours vérifiée ; mais deux marquages du *départ-en-véhicule*

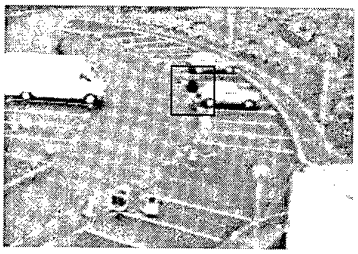


FIG. 13 – *image2, origine de la fiche F_2*

sont maintenant possibles: A2A1 (activités "déplacement-piéton-vers-véhicule" et "véhicule-stationné") et A3A1 (activités "piéton-monte-dans-véhicule" et "véhicule-stationné").

• À la date t_3, avec $t_3 = t_2 + 6s$, TN fournit une fiche F_3 vide (pas d'objet en mouvement observable).

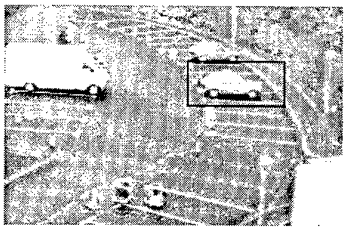


FIG. 14 – *image3, origine de la fiche F_3*

TS émet alors cinq hypothèses de P-situations constituant la situation S_3 : en ce qui concerne le plan *déplacement-piéton*, l'activité "déplacement-piéton" est soit terminée (la condition de terminaison du plan est vérifiée), soit occultée ; en ce qui concerne le plan *départ-en-véhicule*, trois marquage sont possibles : A2A1, avec l'activité "déplacement-piéton-vers-véhicule" occultée, A3A1,
avec l'activité "piéton-monte-dans-véhicule" occultée, et A4, correspondant à l'activité partiellement inobservable "véhicule-stationné-occupé".

• La dernière fiche F_4, délivrée à la date t_4, avec $t_4 = t_3 + 14s$, contient le véhicule $V1$ (pisté par TN) tel que *vitesse(V1, 10km/h)* et un nouvel objet en mouvement $V2$ identifié comme un véhicule tel que *type(V2, véhicule), vitesse(V2, 30km/h).*
Le véhicule *V1* permet de poursuivre le plan *départ-en-véhicule*, avec le marquage A5 (activité "démarrage-véhicule").
En ce qui concerne *V2*, un nouveau plan doit être

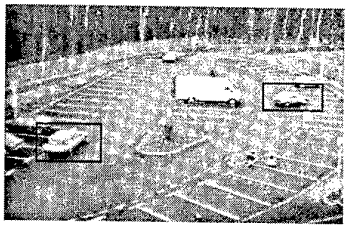


FIG. 15 – *image4, origine de la fiche F_4*

considéré: *arrivée-en-véhicule* est candidat, instancié en P-situation par le marquage A1 (activité "déplacement-véhicule-depuis-entrée-parking"). La situation S_4 met donc en jeu une conjonction de deux P-situations.

8 Conclusions

Les différents modules de la chaîne de traitements sont implantés sur différents calculateurs, dans des langages différents, et dialoguent entre eux par échange de messages Ascii par protocole sockets. GP, qui a un rôle central, est le seul module à devoir traduire les messages reçus et envoyés.
Les travaux en cours portent sur différents approfondissements à apporter :

- amélioration de la classification des objets (base d'apprentissage, fusion des classifieurs infrarouge et couleur) ;
- introduction de connaissances de TS dans le pistage ;
- modèle d'environnement "coopératif" (définition d'explications raisonnables à des objets non identifiés ou à l'absence ou la disparition d'objets ; établissement d'informations complémentaires pertinentes) ;
- robustesse de la reconnaissance de situation : définition d'un cadre algébrique pour le traitement de l'incertitude symbolique [7] ;
- pilotage effectif des TN à partir de GP.

Remerciements

Les résultats présentés ici sont issus des travaux de l'équipe de projet PERCEPTION : Claude BARROUIL, Charles CASTEL, Laurent CHAUDRON, Patrick FABIANI, Roger MAMPEY, Patrick SECCHI, Catherine TESSIER.

Références

[1] Perception. Rapports intermédiaire et final du projet. Technical Report 1-2/7995.02-3575.00, Cert, BP 4025, 31055 Toulouse Cedex 04, fev-oct 1996.

[2] E.D. Dickmanns. Machine perception exploiting high-level spatio-temporal models. *Agard Lecture Series 185*, pages 6–1 –6–17, Aug. 1992.

[3] R.Y. Tsai. A versatile camera calibration technique for high accuracy 3d nachine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, RA-3(4):323–344, Aug. 1987.

[4] M. Orkitz. Localisation d'objets mobiles dans des scènes filmées par une caméra fixe. *Traitement du Signal*, 9(4):325–346, 1992.

[5] Ch. Castel, L. Chaudron, and C. Tessier. What is going on? a high level interpretation of sequences of images. In *Workshop on Conceptual Descriptions from Images, ECCV'96*, pages 13–27, Cambridge, UK, April 1996.

[6] Ch. Castel, L. Chaudron, and C. Tessier. 1st order c-cubes for the interpretation of petri nets: an application to dynamic scene understanding. In *TAI'96, 8th International Conference on Tools with AI*, pages 366–373, Toulouse, France, Nov. 1996.

[7] L. Chaudron, C. Cossart, N. Maille, and C. Tessier. A purely symbolic model for dynamic scene interpretation. *International Journal on Artificial Intelligence Tools*, 1997. To be published.

[8] L. Chaudron. A model of symbolic uncertainty for knowledge fusion in the situation assessment process. In *NATO Symposium on coping with uncertainty in Defence Decision making (unclassified)*, volume 2, pages 15.1–10, The Hague, the Netherlands, Jan 1995.

[9] R. David and H. Alla. *Petri nets and Grafcet.* Prentice Hall, 1991.

[10] P.J. Fabiani. Strategy of perception and temporal representation of beliefs. In *TAI'96, 8th International Conference on Tools with AI*, pages 48–51, Toulouse, France, Nov. 1996.

Multisensor Data Fusion for Automatic Recognition of High Value Surface Targets

K.-H. Bers *, H. Essen **, K. Jäger *, H. Schimpf **
* FGAN–Forschungsinstitut für Informationsverarbeitung und Mustererkennung FIM
Eisenstockstr. 12, D-76275 Ettlingen, Germany
e-mail: ber@gate.fim.fgan.de
** FGAN–Forschungsinstitut für Hochfrequenzphysik FHP
Neuenahrer Str. 20, D-53343 Wachtberg, Germany
e-mail: schimpf@gaserv.fhp.fgan.de

SUMMARY

This paper describes a system for surveillance and target recognition for autonomous stand-off weapons based on the interpretation of multisensor data in combining the individual sensor channels to optimize the system performance by intelligent fusion techniques.

To meet this demand, we propose a model-based method for the automatic recognition of high value fixed or relocatable surface targets like bridges, airfields, industrial installations and command posts. A production net is used to represent the knowledge about target structures. The analysis is carried out by a knowledge based classification system.

Starting with primitive objects extracted separately from each individual sensor channel (low-level-processing), more complex parts are composed step by step until the target object is recognized (high-level-processing). The radar preprocessing comprises a thresholding prescreener and a subsequent feature based discrimination stage, the features making use of polarimetric and geometric properties, and of scattering center statistics. The IR-preprocessing is more related to geometric and structural features like line elements, corners and circles. To enhance the detection probability, in the low level stage a high false alarm rate is accepted as regards the primitive objects. The reduction of the false alarm rate is done in the high level stage by the spatial fusion of 2D- or 3D-structures and by using additional context information of the target environment.

The purpose of this paper is to address the state of the algorithms for automatic recognition of high value surface targets. For this, sensor data have been interpreted which were recorded with an imaging infrared sensor and a coherent, polarimetric, high range resolution radar. To enable the spatial fusion process, the raw sensor data were registered together with inertial data of the aircraft.

The efficiency of the analysis system is demonstrated by an example involving the detection of command posts.

The results show the suitability of the method for future autonomous stand-off weapon systems like drones and missiles with imaging sensors.

1. INTRODUCTION

Future military systems for surveillance and target recognition will increasingly be based on the interpretation of multisensor data. The amount and complexity of the data will exceed the human ability to interpret the data creating a demand for automation of all or part of the fusion process. In the case of stand-off weapons (missiles and drones [1,2]), human interaction is removed entirely from the decision loop and the processing has to be fully automated.

One of the main research activities of FGAN is the development and evaluation of methods for automatic target acquisiton in multisensor data.

For the detection of specific high value fixed or relocatable surface targets we propose a **model-based structural analysis method** using geometric and generic target information and additional context. In the *low-level processing stage* a high false alarm rate of primitive objects making up a target is accepted to enhance the detection probability. The reduction of the false alarm rate is done in the *high-level processing stage* by the spatial fusion of more complex 2D- or 3D-structures and the interpretation of context information like map and situation information. The process of building up more complex structures from less complex structures is implemented by a production system. In this way, the system performance is based on a proper scene analysis and it is more independent of an optimal selection of algorithms and the associated parameters than conventional system design.

The model-based approach has been applied succesfully to the detection of different high value fixed target classes like bridges and airfields [3,4,5,6]. This paper describes the efficiency of the analysis system for the detection of *command posts* which belong to the class of high value relocatable targets.

Due to the non-availability of the final sensor system, which is still under development (section 2.1) sensor data have been analysed which were taken by an imaging infrared sensor (8-12μm) and a co-

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

herent, polarimetric, high range resolution radar (94GHz) in a side looking synthetic aperture approach.

This work is closely connected to current experimental studies of future weapon systems as directed by the German MoD.

2. BACKGROUND

2.1 Dualmode Seeker System ABG

The development of an autonomous missile system for the detection and engagement of command posts (ABG) is identified as an important task by the military defense authorities. The studies are carried out by German companies. BGT and SIEMENS. The FGAN research institutes are involved in an advisory capacity concerning the definition of sensor parameters and the evaluation of data fusion algorithms for target detection.

The activities started with the definition of the expected scenario together with the inclusion of the availability of aerial reconnaissance information for target description and mission planning. Based on these requirements, the seekerhead system design and system parameters were determined for the construction of an experimental seekerhead system. In parallel, signal processing methods for sensor data fusion have been developed. During the next phase of the project, captive flight tests will be realized to demonstrate the suitability of the dualmode sensor and to evaluate the overall system performance.

The results of these studies lead to a concept for a seekerhead system, which provides a dualmode sensor consisting of an imaging infrared sensor with a focal plane array detector and a real aperture stepped frequency pulse radar. Both sensors are integrated within a common aperture and the system is provided with an antenna, scanning in azimuth and elevation direction. The system is also equipped with an Antiradiation (AR)-component which is not further considered in this paper.

The experimental seekerhead system is still under construction and will be evaluated by captive flight tests starting in 1998.

2.2 Dualmode Data Sources

Due to the lack of real ABG sensor data at present, the development of target detection algorithms is implemented with simulated data. Several measurement campaigns have taken place with different existing sensor systems to gather data of the quality similar to that of the later ABG sensor hardware.

Radar data was collected by using the experimental radar system *MEMPHIS* (= **M**illimeter wave **E**xperimental **M**ultifrequency **P**olarimetric **H**igh resolution **I**maging **S**ensor) of FGAN-FHP [7,8]. This is a coherent, polarimetric, high range resolution radar, operating at a nominal frequency of 94GHz, which has been installed on board an aircraft in a side-looking SAR configuration. In [9] the measurement system is described and some evaluation procedures with emphasis on SAR preprocessing are discussed. The radar is equipped with a dual channel receiver to register simultaneously the parallel and orthogonal polarization component of the backscattered signal. The radar data together with flight parameters (pitch, roll, heading, time code) were gathered by means of fast acquisition electronics and written onto a magnetic tape. For image analysis the resulting radar images were interpreted as 2-dimensional images.

Fig. 1 shows an example of the examined radar images (scene A). The scene contains a command post with various deployed elements, whereas the most obvious part of the target is located next to a big building in the upper part of the SAR image. It consists of a cluster of vehicles (jeeps, light and heavy trucks, power generators) covered by camouflage nets along the lower side and around the lower corners of the building. The surroundings of the target area consist of grass-land with many scattered trees and bushes. Moreover, there are numerous artificial man-made objects in the scene, like sheds, mounds, various small buildings and uncamouflaged vehicles.

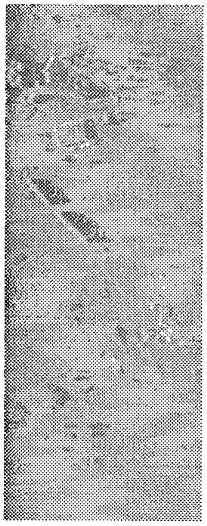


Fig. 1: SAR Image

The IR data were gathered with the experimental dualmode sensor system MSS [10]. This system is equipped with an imaging infrared sensor (8-12μm) and a FMICW monopulse radar (94GHz) and is provided with a scanning antenna. Because of the poor lateral resolution of the MSS radar data (versus the expected data of the ABG system), the

mmW channel is not considered. For the data collection the sensor frontend was fixed forward to a transport helicopter under a variable depression angle. Sensor and flight parameters were recorded during the measurements. The scene is scanned in a pushbroom way producing a sequence of overlapping images. Because of the narrow field-of-view of the sensor, a single image contains only a small section of the scene.

To get a complete coverage of the target area and to simulate the future ABG sensor characteristics it is necessary to register the individual IR-frames in a ground map. Fig. 2 shows the result of the registration process.

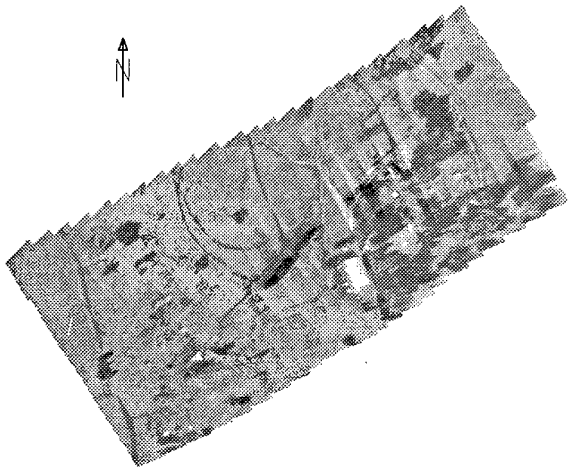


Fig. 2: IR Ground Map

2.2 Data Harmonization for ABG Simulation
IR- and SAR-data were recorded using different carriers with different flight parameters. As consequence, only a portion of the scene is covered by both sensors. To enable the data fusion process, both data sets have to be transformed into a common reference coordinate system. This is done by an affine transformation of the radar data into the infrared image representation neglecting special 3D-effects of the different image generation process (Fig. 3). These data are stored in the database of the simulation test bed (section 3.1) together with the data of temporal and geometrical harmonization and could be addressed for analytical purposes as originating from the real ABG sensor.

Therefore the harmonization process is not basic to the sensor fusion process, because the detection algorithms, described in section 3, are designed to take into consideration that in the real-time system the infrared information is gathered frame by frame and the radar data line by line, so that processing runs synchronized with data acquisition.

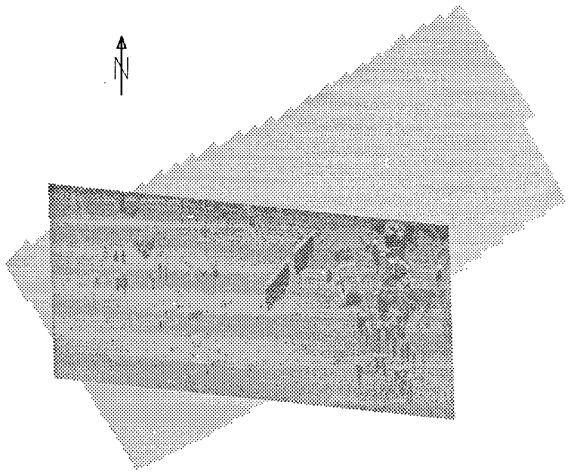


Fig. 3: Harmonization of IR and SAR Data

3. MODEL-BASED SENSOR FUSION

3.1 The FIM BPI-Sensor Fusion Test Bed
For the task of target detection in multispectral data the **B**lackboard-based **P**roduction system for **I**mage understanding (BPI) is used [11,12,13].

The BPI-System is a framework for model-based structure analysis of complex scenes and consists of three basic components: a global database (blackboard), a control unit and a set of processing modules (agents). Fig. 4 shows the blackboard architecture adapted for target classification in dualmode sensor data.

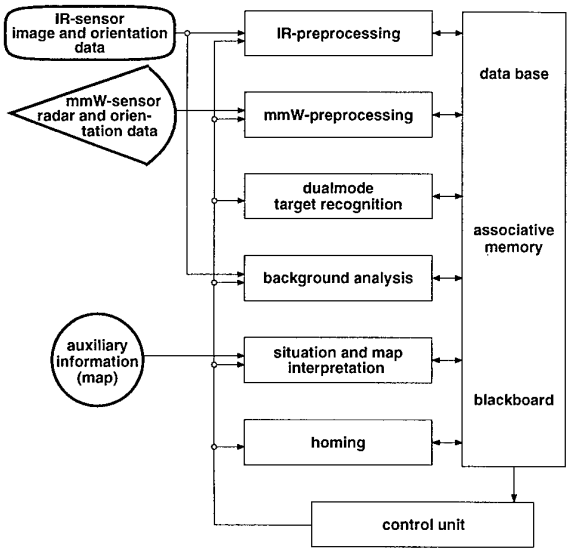


Fig. 4: System for Target Recognition

The analysis starts with a *low-level-processing* or preprocessing of the sensor raw data by extracting proper primitive objects independently in each sensor channel. Objects with higher complexity (partial objects) of the searched target will be created by applying production rules of the *medium-level-processing*. The inference process of the model-based sensor fusion analysis will combine these intermediate results to classify the final target by *high-level-processing* which uses production rules of higher complexity.

All objects, intermediate or final, represent certain structures of the target and are stored with suitable attributes in the database which functions as an associative memory. The productions ("rules") are implemented by processing modules which communicate exclusively via the blackboard. The control unit monitors the system activities. The analysis is *data-driven* and is designed for running in parallel. Details concerning the architecture and the dataflow of the BPI-System are described in [14].

In addition to the modules for the dualmode target recognition, modules for background analysis (e.g. texture analysis [15]) and the automatic interpretation of auxiliary information (situation and map analysis) are integrated into the BPI-System. It is expected that this additional context information (e.g. neighbourhood relations to objects of the background) will enhance the detection probability.

For a dualmode seeker system with the option of target engagement, modules for homing (e.g. tracking algorithms) have to be incorporated.

In general, the fusion of multisensor data is based on a spatial confirmation of geometrical target features in a world reference system. For this spatial fusion it is necessary to know the process of image generation, which enables the transformation of sensor data into the reference system (registration). In particular, the sensor parameters must be available which consist of the extrinsic (position, orientation) and intrinsic (focal length, range resolution...) sensor parameters.

3.2 Modelling

The general interactions of productions and the stepwise transfer of objects into objects of higher complexity can be displayed by a production net [16]. It represents the target model or the knowledge about the target.

For our basic research we define as an example a generic command post consisting of

- cluster of vehicles
- power generators
- containers
- antennas

which often are installed close to buildings (in urban areas) or which are partly obscured e.g. in clearings (in wooded areas).

Consequently, basic signatures that can be expected - with respect to the sensor resolution at the proper distance - are contours (lines), areas (edges), hotspots (cues) and similar structural parts of the target components depending on the signature in the individual sensor channel. For each part of these structures (object parts), a proper set of attributes is processed.

Maps, aerial photos (if available) or comparable information like military studies can be used as knowledge sources for the geometric arrangement of these primitive objects making up the target. By this the model generation process needs surveillance in advance.

The general procedure of modelling is illustrated by the production net of fig. 6 using the command post of test scene A (fig. 5) as an example. The productions determine how a given set of objects is transferred into a set of more complex objects.

In our example, the command post is composed of a cluster of vehicles and power generators situated near a building. The building is represented by its roof which appears in aerial images as a parallelogram.

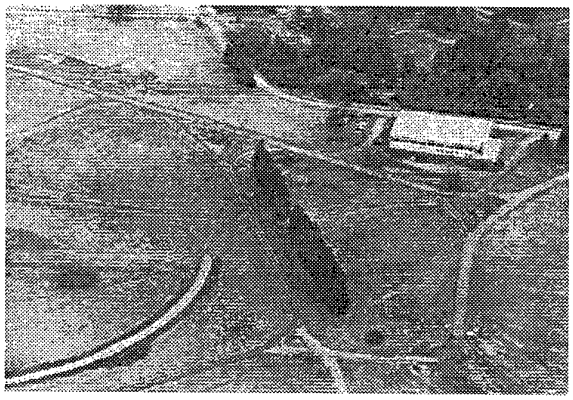


Fig. 5: Reference Image

Starting with the primitive objects LINE, the objects ANGLE are built by objects LINE enclosing an angle α. If two objects ANGLE form a structure like an open parallelogram they are combined to an object U_STRUCTURE. An object PARALLELOGRAM can be built if objects U_STRUCTURE and LINE are compatible. Objects PARALLELOGRAM are actually used to describe buildings, because in aerial images containing houses only their roofs can be recognized which often appear as parallelograms.

For the construction of vehicle clusters, primitive objects CUES are combined to PAIRS if their dis-

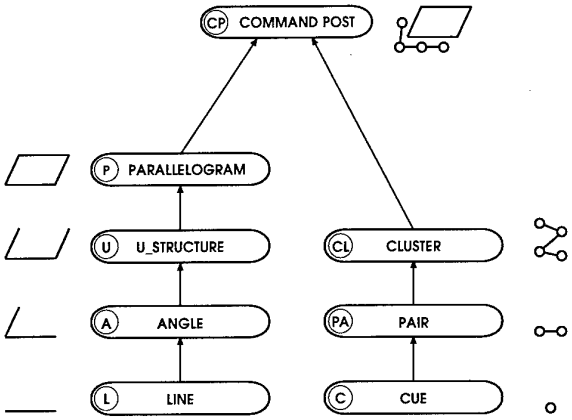


Fig. 6: Model for Object COMMAND POST

tance corresponds to the model parameter. Two objects PAIRS are grouped to an object CLUSTER if they have a common endpoint. In a similar way objects CLUSTER are produced by objects PAIRS and objects CLUSTER or objects CLUSTER and objects CLUSTER.

Finally the target object COMMAND POST is generated by an object CLUSTER situated in the neighborhood of an object PARALLELOGRAM.

With this model and the use of known flight and sensor parameters the geometric structure of the command post and its parametric description can be computed at the actual detection time.

3.3 Model-Based Multi Sensor Fusion

3.3.1 IR-Preprocessing

The basic elements for the structure analysis are the primitive objects LINE and CUE (fig. 6). They are created in the preprocessing stage by standard iconic image processing methods. In this way symbolic descriptions of the images are generated.

The method for generating the objects LINE is based on a contour approximation. These lines are extended to longer lines.

The choice of the cue detection algorithm is guided by the spatial resolution of the object. For the detection of point targets a hotspot operator is used based on a nonlinear filtering.

As we have a scanning sensor providing an image sequence, the IR-preprocessing algorithms are applied to each individual image. For the construction of more complex objects, lines and cues are registered in a world reference system and are stored in the database as primitive objects LINE and CUE respectively.

3.3.2 SAR-Preprocessing

Similar to the IR-preprocessing the generation of primitive objects in the radar channel is carried out by a radar specific method. As input for sensor fusion, the object CUE, as candidate for vehicle, and the object LINE, as a basic element of man-made objects are considered; they are extracted from the data in two steps.

The first step consists of an amplitude based thresholding prescreener to select potential target areas and to reduce the computational burden in the subsequent processing stages. Two prescreener types are defined which work on single range profiles. The first uses the total power contrast to determine areas of potential target (vehicle) sites. The second type of prescreener represents a shadow filter and detects strongly reflecting structures casting a shadow like vehicles and edges of buildings.

Based upon the output of the prescreener, feature vectors are then constructed in the second step to discriminate between cues and natural clutter background. For this purpose, several polarimetric and geometric features are computed over small areas centered at the cell under test. Only those cells that had passed the prescreener were subjected to the feature test. The size of the areas is adapted to the size of the target i.e. the discriminator acts on a small number of range profiles. This takes into account the fact, that in the real ABG-sensor system, the incoming radar backscatter signals form a time sequence. That means that each detection algorithm has to start its work while the radar image is still being constructed.

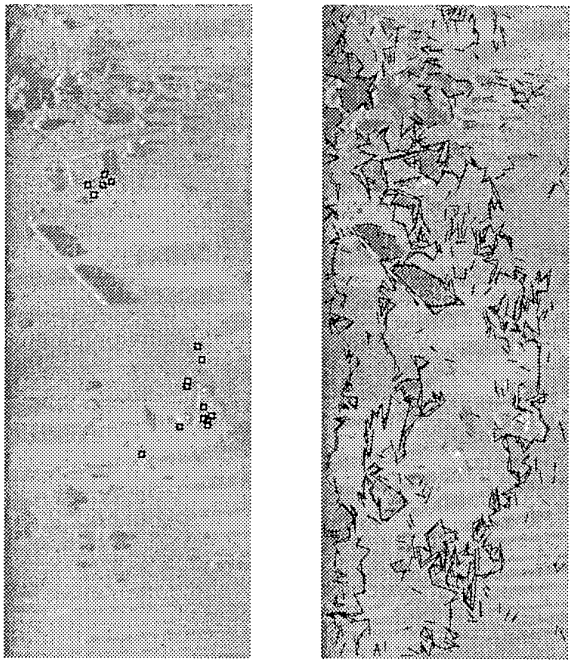
For the generation of the primitive objects LINE the same procedure is applied as described above for the extraction of lines in the IR channel. The lines are extracted in a moving window of the SAR map corresponding to the consecutive accumulation of range profiles. In combination with the results of the shadow filter, lines with high reflecting parts are candidates for man-made objects.

The final result is an image that contains within each pixel the decision CUE, if it is a candidate for vehicle and a symbolic description of objects LINE for all candidates of buildings. Details concerning this radar specific preprocessing are described in [3,9,17].

Fig. 7 shows the results which are stored as primitive objects to be processed in the following model-based data fusion.

3.3.3 Application to Dualmode Sensor Data

The general task of the ABG sensor is to analyse a certain area of interest, obtained via a priori information, and to identify autonomously command posts within that area and define an adequate aimpoint on that target. In doing this, special constraints defined by the different phases of mission (midcourse guidance, acquisition, recognition, engagement) have to be taken into consideration.



Cues (black squares) Lines (black lines)

Fig. 7: SAR Low-Level Results

During midcourse guidance, the missile is steered to the target area by INS with image based navigation update. This part of the mission is not further addressed in this paper.

By approaching the predefined target area, the acquisition phase is initiated, which is characterized by scanning the scene under a low depression angle. By that, a large search area is covered by the sensor with a medium spatial resolution of objects in the IR channel. At the beginning of the acquisition phase, the application of signal processing methods should result in cueing and selecting *areas of interest*, which are potential sites of command posts. This action can be accompanied by autonomous flight path correction to optimize the data gathering for target detection.

After the acquisition phase, the areas of interest are interpreted in more detail during the target recognition phase. Therefore, the spatial resolution of objects in the IR channel is enhanced by increasing the depression angle continously. At the same time, the size of the scanning area is reduced. At the end of this phase, locations of candidates for a command post are determined to initialize the engagement phase. The overall detection process can be interpreted as an *active vision* procedure.

Starting with cueing structures of poor complexity, areas of interest are determined, which are analysed in the next processing stage in more detail. This results in cueing objects with higher complexity and the definition of more precise areas of interest (e.g. with smaller size) to be analysed in the following stage. The analysis ends when the desired degree of complexity, precision of localisation and confidence of target identification is reached.

In every stage of the analysis process the complexity of modelling and the selection of the processing method have to be adapted to the degree of target resolution. Moreover, in the real system, the detection process can be supported by an active control of the sensor parameters (scan angle and rate, depression angle etc.) and the flight path of the missile.

Fig. 8 shows the result of the model-based structure analysis generated during a simulated acquisition phase of test scene A. The objects CLUSTER (white polygons) are built up by objects CUE (white squares), which were extracted in the IR- and radar channel. Candidates for buildings (black) are represented by objects U_STRUCTURE and objects PARALLELOGRAM produced by primitive objects LINE, which were extracted in the individual sensor channels. All objects are assessed according to a spatial confirmation of the associated infrared and radar structures.

The area of interest (white rectangle) is determined by the area surrounding the best assessed candidate for the object COMMAND POST (black) produced by the objects CLUSTER (black) situated near candidates for buildings (objects PARALLELOGRAM).



Fig. 8: Area Of Interest

In the next step, the target recognition phase is simulated by a detailed analysis of individual IR-frames belonging to the area of interest. Fig. 9 shows a sequence of 6 consecutive IR-frames (area

of interest) covering parts of the potential target while scanning. The primitive objects CUE (white squares) and LINE (black lines), detected in the individual frames, deduce the object PARALLELOGRAM (white) and the object CLUSTER (white polygon) and together form the object COMMAND POST, which is the highest level of complexity of the target model in the implemented sensor fusion process.

The results are visualized in an infrared panoramic representation.



Frame n_1-n_6 IR-Panorama Representation

Fig. 9: Detail Analysis

4. CONCLUSIONS

The FGAN research institutes FIM and FHP are involved in the development of dualmode sensor fusion technologies. A main task consists of advising the German MoD on the evaluation of system performance, sensor characteristics and algorithm quality. The design of systems, sensor hardware and the integration of algorithms for realtime application are tasks of the companies and are not part of our activities.

For the evaluation of multisensor data fusion algorithms, the BPI test bed has been developed. The examinations carried out with data obtained from a multisensor system have shown that knowledge-based structure analysis for autonomous detection, classification and identification of special targets is appropriate for seekerhead and drone applications.

The objective of the ongoing research is the extension of the system to a three-dimensional fusion and interpretation of multisensor data. In addition, procedures for texture analysis will be incorporated in the system for the interpretation of context information. It is assumed, that these extensions will improve detection and target localization performance.

5. REFERENCES

[1] Migaud P.: *Application Conjointe de l'Infrarouge passif et du Millimetrique actif au Guidage d'Armements.* Int'l Radar Conference, Paris, Mai 1994, p.1-4

[2] Nivelle F., Rudolph W., Winter F.: *ASTRID: Autonomous Search and Track with Radar and Infrared Dual-Mode IR/mmW Dual-Mode Seeker.* In: Proc. to MSP 4th Symposium on Technologies for Precision Air Strike Operations in Rapid-Reaction and Localized-Conflict Scenarios. Sevilla, Spain, 1995

[3] Bers K.-H., Jäger K., Jurkiewicz K.: *Model Based Image Analysis for the Detection of High Value Surface Radar Targets.* AGARD SPP Symposium on Radar Signature Analysis and Imaging of Military Targets, Ankara, 1996

[4] Bers K.-H., Jäger K., Lütjen K.: *High Value Target Detection by Multisensor Data Analysis.* 2nd NATO-IRIS Joint Symposium, NATO PANEL IV, London, 1996

[5] Bers K.-H., Jurkiewicz K.: *Model Based Structure Analysis for High Value Target Detection in Radar Images.* Radar Imaging and Classification Techniques, NATO AC/243(panel 10)TP/1, Wachtberg, Germany, May 1995

[6] Bers K.-H., Jurkiewicz K.: *Model-based Acquisition and Terminal Homing of High Value Targets for Precision Guided Weapons.* 34th DRG's Seminar on Precision Guided Weapons, NATO AC/243-TP/6, Brussels, 1993

[7] Boehmsdorff S., Essen H.: *Target and Background Measurements over Land and Sea with a polarimetric 35/94-Ghz Synthetic Aperture Radar.* SPIE 3062-17, Orlando, April 1997

[8] Makaruschka R., Essen H.: *Airborne Dual Sensor mmW Signatures of Maritime Targets and Sea Clutter.* AGARD SPP Symposium on Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar, Lisbon, Sept. 1997

[9] Schimpf H., Essen H.: *Detection of Extended Targets in Millimetre Wave Radar Imagery.* AGARD SPP Symposium on Radar Signature Analysis and Imaging of Military Targets, Ankara, 1996

[10] Rudolph W., Migaud P.: *Autodirecteurs bimodes.* Defense and Technology International Mars, 1992

[11] Erman L.D., Hayes-Roth F., Lesser V.R., Reddy R.: *The HEARSAY-II Speech-Understanding System.* Comp. Surveys 12, 1980

[12] Nii P.: *The Blackboard Model of Problem Solving.* AI Magazine 7(2): 38-53, 82-102, 1986

[13] Lütjen K.: *BPI: Ein Blackboardbasiertes Produktionssystem für die automatische Bildauswertung.* In Hartmann G. (ed) Mustererkennung 1986, 164-168, Berlin: Springer, 1986

[14] Stilla U.: *Map-Aided Structural Analysis of Aerial Images.* ISPRS Journal of Photogrammetry and Remote Sensing, 50(4): 3-10, 1995

[15] Bargel B.: *Automatische Klassifikation von Fernerkundungsdaten durch statistische und strukturelle Texturanalyse.* Dissertation, FIM-Bericht 114, 1983

[16] Stilla U., Michaelsen E.: *Semantic Modelling of Man-Made Objects by Production Nets.* In: Gruen A., Baltsavaries E.P., Henricsson O. (eds) Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Basel: Birkhäuser, 1997

[17] Schimpf H., Essen H.: *SAR Measurements of Extended Targets at 94Ghz.* AGARD Conf. Proc. No.582 "Remote Sensing", Toulouse, April 1996

A KF-BASED INTEGRATION SYSTEM FOR LAND VEHICLE TRACKING FROM REAL DGPS AND INS DATA

Carlo S. Regazzoni, Andrea Teschioni and Giorgio Tacconi

Department of Biophysical and Electronic Engineering (DIBE)
University of Genoa
Via all'Opera Pia 11A
16145 Genova (Italy)

e-mail: carlo@dibe.unige.it

SUMMARY

The present work is addressed to perform an estimation in an accurate and robust way of the trajectory of a land vehicle by using a Differential Global Positioning System (DGPS) and an Inertial System (INS).

The use of a Kalman Filter (KF) approach for integration, data-fusion and estimation tasks has been proved as able to providing precise and robust evaluation of cinematic variables (linear position and velocity) even in the case of long missions or under critical conditions of temporary incompleteness or unreliability of part of the acquired data.

From the state of the art, it can be seen that the DGPS is a very precise sensor providing 3D geographic position, but presents low output rate and temporary signal loss or accuracy degradation, while the INS provides continuous outputs of rotation angles and linear acceleration with high output rate but the inertial units have burdensome intrinsic errors which bring about a degradation of precision increasing with time.

Practically, the integration of DGPS and INS is forecast to provide continuous estimates over time, corrupted by small and almost unchanging errors. The system has been tested over an extensive set of real data providing good results both in precision and in robustness.

1. INTRODUCTION

In the last years, an increasing interest for the development of systems able to compute the precise position of an object in the ground has been shown.

In this work, a system for the determination of the position of a vehicle based on the integration of two different sensors, DGPS and INS, is proposed.

The integration and estimation approach is based on the use of the Kalman Filter.

The control of the position and the tracking of a vehicle equipped by a set of localization sensors and instruments for the estimation of cinematic variables are a very difficult problem.

In this context Kalman Filter (KF) has intensively been used in navigation systems since 1960. Most recent applications concern with the use of KF in the satellite navigation systems, such as, for example, the Global Positioning System (GPS).

One of the most interesting problems is the integration of inertial systems (gyroscopes, accelerometers, i.e. Inertial Navigation Systems – INS) with other Navigation Systems, for example GPS.

In fact, inertial systems are characterised by some drift phenomena that cause the increasing of the errors at the increasing of the time according to less or more complex laws.

Due to this fact, during long missions, inertial systems have to be used together with other navigation systems whose aim is to support INS in the global navigation system.

The present work aims at estimating, in an accurate and robust way, the trajectory of a land vehicle by using the Differential Global Positioning System (DGPS) and the Inertial Navigation System (INS).

In the most of cases, the estimation of a vehicle position achieved by KF cannot be optimal due to the fact that the noise corrupting the chosen status' variables is Non-Gaussian.

In this paper, a new dynamic model for KF equation has been introduced, allowing one to have only Gaussian variables corrupting the status' variables and consequently to obtain an optimal position estimation of the vehicle.

The proposed system is also able to manage the different reliability of the DGPS data, thanks to the adaptation of KF parameters to the different situations.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

The paper is organized as follows: first of all the characteristics of DGPS and INS are presented and possible integration approaches are described.

In the following, the architecture of the proposed system is presented and most important concepts concerning the new integration approach are shown. Finally, some experimental results obtained in the position estimation of the vehicle are illustrated.

2. THE DGPS SIGNAL

The Global Positioning System (GPS) [1] is a navigation system based on satellite communications use that allows the user to determine its position in every ground site which is not occluded by eventual obstacles.

GPS offers, with every atmospheric condition and at every instant of the day, an instantaneous and global vehicle positioning, with a few meters error, depending on the receivers' cost.
GPS is currently employed in many application fields: for example, industrial applications concern with surveillance, maps creation, photogrammetry, public security, telecommunications.

The precision error of the GPS may be reduced to less than one meter, by means of the introduction of differential correction techniques [2].

The Differential GPS (DGPS) is an improvement of GPS that uses errors introduced by GPS in the localisation of known points in order to correct the errors in the localisation of unknown positions. DGPS is a very precise sensor that gives 3D geographic positions, but that presents a low output rate ($\leq$ 1 Hz) and either temporary signal losses or accuracy degradation.

The main disadvantage of DGPS is the occlusion of the signal by buildings, tunnels or trees: in such cases the number of visible satellites, on which the DGPS signal is based, decreases and the positioning provided by the DGPS is not trustful.

3. THE INS SIGNAL

Inertial Navigation System (INS) is a system for the estimation of a vehicle position, which is based on the Newton laws of the classical mechanics.

The vehicle position is obtained by means of a double integration of the measured vehicle acceleration.

Three accelerometers and three gyroscopes, aligned with respect to axes mutually orthogonal, which constitute an inertial reference system, measure the linear acceleration and the angular speed of the vehicle [3].

The gyroscopes also provide the necessary information in order to stabilise the platform on which sensors have been mounted with respect to the chosen inertial reference system. The described stabilisation can be performed both physically, within a system called *gimbal* [4], both analytically, in the so-called *strap-down* [5] systems. After an initial start-up, the INS begins to produce a continuous output of the position, speed and direction of the vehicle, independently from every external agent and from environmental conditions.

The INS provides continuous outputs of rotation angles and linear accelerations at high output rates [10, 50] Hz, by reducing sampling errors and by improving the precision.

The INS advantage lies in its independence from external factors and in the fact that from the INS outputs position, speed and orientation angles are immediately available. Moreover, in the short time, INS is more precise than GPS, by a 10 factor [6]. Main disadvantages of INS are its relevant cost and the fact that, if it is not integrated by some other positioning methods, INS has an intrinsic time-varying error that causes a rapid deterioration in the precision of the position estimation. Such error corresponds to around 2 Km for each navigation hour.

4. WHY INTEGRATION OF GPS AND INS ?

The general motion of a vehicle in a three-dimensional space can be described by means of six parameters: three accelerations and three orientation angles.

A system for measuring the trajectory of a vehicle has to evaluate six independent measured sizes, from whose it has to derive the already mentioned parameters.

GPS signal integrated by INS signal can be used in order to achieve such a purpose.

It can be demonstrated that the GPS shows a significative superiority with respect to the INS in terms of accuracy and precision, given that inertial sensors are affected, in the long term, by relevant errors.

Moreover, GPS has an higher cost than INS and it does not require an initialisation procedure. Nevertheless, also GPS has certain limits and INS has some useful characteristics, in particular the continuous output of position, speed and direction of the vehicle: in fact, INS has outputs independent from external agents, such as the satellites visibility or the meteorological conditions.

Due to these facts, the integration of GPS and INS, is able to offer, as final result, a system which allows to maintain advantages of both navigation techniques and to remove many disadvantages which could be present if GPS and INS were singularly considered.

Hence the integration of DGPS and INS data is forecast to provide continuous estimates over time, corrupted by small and almost unchanging errors.

5. THE KF-BASED INTEGRATION APPROACHES

The main objective of the integration between DGPS and inertial data is to mix all the available information provided by sensors in order to obtain the estimation of the position, speed and direction of a vehicle: this will allow to reconstruct the trajectory of the vehicle.

The synchronisation of data coming from DGPS and the ones computed by the inertial navigation system is essential in order to correctly compare all the information concerning the moving vehicle.

In general, INS can be used in order to interpolate the position estimated by DGPS with a very high frequency and in order to replace the DGPS signal when it is obscured or in the case of bad satellites configuration.

Most commonly used technique for DGPS and INS data processing and fusion is Kalman Filtering.

However, this filtering technique is not the solution to all problems introduced by the DGPS/INS Integration: in fact, one that has to design such a system must take into account some specifications that have to be respected in order to obtain a system that is really operating.

These requirements mainly are computational efficiency, precision of the estimation, robustness to different kinds of noise, compatibility with different systems: however, the recursive characteristic of the Kalman Filter makes it efficient for real-time applications.
Also its noise robustness is a characteristic necessary for real-time navigation systems.

In the following paragraphs, the proposed approach based on the use of Kalman Filter will be presented: the peculiarities of the proposed approach will be highlighted and some experimental results obtained in the estimation of the position of a land vehicle.

6. SYSTEM DESCRIPTION

The KF-based approach used for the DGPS/INS integration is based on a *tight integration* scheme [7]: the structure of the system is shown in Figure 1.

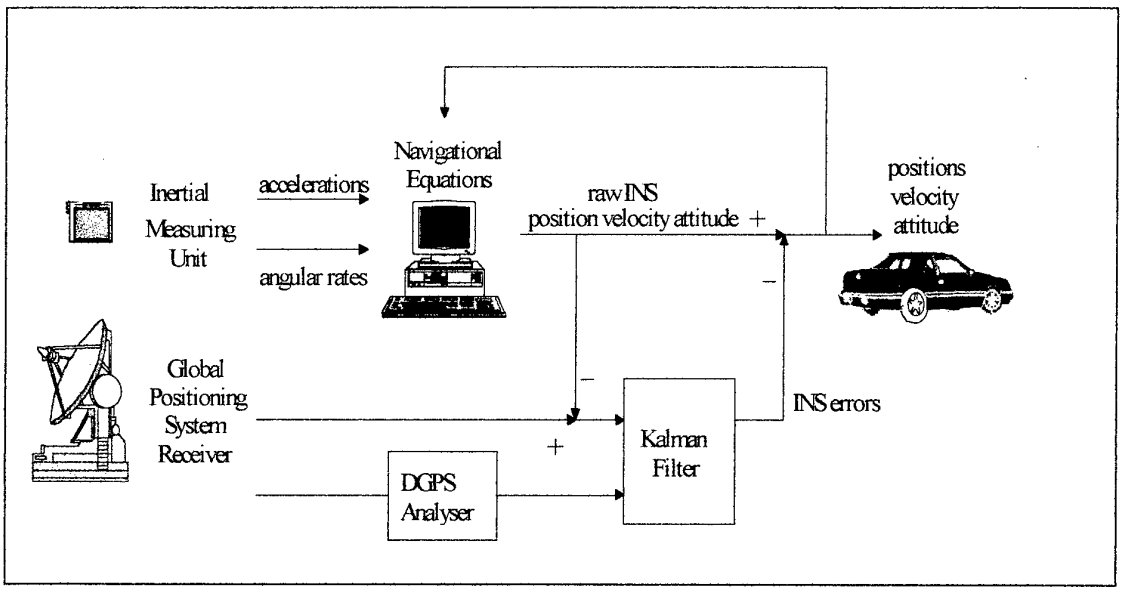


Figure 1. KF-based architecture for DGPS/INS tight integration

By observing Fig. 1 it is possible to see that:

- a unique Kalman Filter is used to process signals coming from satellites and inertial ones;
- GPS based positioning is directly integrated by the INS outputs;
- DGPS and INS are strictly integrated and the performances of the two systems become strictly dependent each other.

In fact, the capability of DGPS to "follow" the vehicle when the vehicle performs sharp manoeuvres will depend on the accuracy of the INS outputs. On the other part, the INS accuracy will depend on the frequency and on the quality of the updating produced by DGPS.

In such a way, if measures coming from DGPS are not trustful, they can be refused in a unique time, without compromising the performances in the position calculation, thanks to INS data.

On the counterpart, if the INS signal is wrong and the DGPS signal is absent or not trustful, the system is not able to establish the exact position of the vehicle, unless a dedicated Kalman Filter for only DGPS data processing is inserted.

Following the proposed approach, the position of the vehicle is estimated through the use of INS data to calculate the trajectory of reference and of DGPS data, if available, to update the trajectory itself.

Before describing the structure of the used Kalman Filters based architecture and of the real DGPS/INS integration, it could be useful to expose which is the main strategy of the DGPS/INS integration itself.

At a first approximation, dynamic variables in the status vector of the system model could be the three error variables obtained by means of the difference between the position, the speed and the acceleration estimated by DGPS and the ones calculated by INS, that is:

- $\bar{x} = (\delta\bar{r} \;\; \delta\bar{v} \;\; \delta\bar{a})$ (1)

Such error variables, which are estimated by KF, are then used to update the INS output and to generate the position of the vehicle. It is important to recall the different rate of DGPS data (1 Hz) and of INS data (10 Hz).

Because of the sensors configuration, system will have to read one DGPS datum every ten INS read data. The integration system produces as output an estimation of the vehicle position at a rate corresponding to 2 Hz.

Due to this fact, it will be necessary to insert a pre-processing step on DGPS data, by interpolating two successive samples in order to obtain a 2 Hz DGPS output and by making an INS data compression, which can be obtained through the averaging in a unique value five successive INS samples.

The proposed system, as previously said, uses DGPS produced data as a reference to continuously update the inertial navigation system.

In such a way, an eventual absence of DGPS signal, due to satellites problems or to obstacles placed between users and satellites themselves, could be compensated through the use of a INS signal which can be considered as trustful, at least in the short periods.

The estimation of the $(\delta\bar{r} \;\; \delta\bar{v} \;\; \delta\bar{a})$ vector, which is calculated by KF, will be subtracted from the current values of position, speed and acceleration calculated by INS and it allows to update the values of INS outputs.

Unfortunately, the updating offered by DGPS is sometimes not trustful or absent at all.

The KF-based system has been designed in order to make it possible to also adopt, besides to the basic fusion strategies (DGPS present, DGPS absent), other operating possibilities, in order to take into account the characteristics of the DGPS signal: DGPS quality can be characterized by a reliability degree depending on the value of a parameter called Position Dilution of Precision (PDOP) [7].

Such "Multi-Hypothesis" approach makes the proposed system able to commute between different operating possibilities, according to the reliability of DGPS data and consequently to the value of the PDOP parameter.

In such a way, PDOP becomes the control variable depending on which to apply the different available integration strategies.

The proposed integration system is also able to follow the vehicle even in the worst case, when DGPS is absent.

In this case, the status vector constituted by the error/updating variables estimated by KF is obtained through subtracting current the cinematic measures calculated by INS from the ones estimated by the system at the previous instant.

In the following paragraph, the proposed data fusion approach, that allows to find the optimal estimation of the vehicle position through the integration of DGPS and INS, is presented.

7. THE PROPOSED DATA FUSION APPROACH

The peculiarity of the employment of the Kalman Filter in the proposed approach mainly resides in the noise modeling approach.

In fact, it is known that a KF gives optimal results only if noise affecting status and measures vectors is Gaussian.

In this case, such assumption is not valid, due to the fact that the models of the noise acting on the accelerometers and on the gyroscopes are not linear and they have non-Gaussian statistics.

In fact, taking into account only the system dynamics and the final estimation targets, the variables in the status vector $\bar{x} = (\delta\bar{r} \;\; \delta\bar{v} \;\; \delta\bar{a})$, are three error variables carried out by the differences between the positions, velocities and accelerations provided by the DGPS and the INS.

Moreover, one of the main critical aspects of the DGPS/INS integration is the deviation from Gaussianity of some of the noise components corrupting the status variables' true values:

$$\bullet \qquad \begin{bmatrix} \delta\bar{r}' \\ \delta\bar{v}' \\ \delta\bar{a}' \end{bmatrix} = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \delta\bar{r} \\ \delta\bar{v} \\ \delta\bar{a} \end{bmatrix} + \begin{bmatrix} \bar{\xi}_g \\ \bar{\Delta}_a \\ \bar{\xi}_g \end{bmatrix} \qquad (2)$$

The term $\left[\delta\bar{r}' \quad \delta\bar{v}' \quad \delta\bar{a}'\right]^T$ is the time derivative of the differential status variables, I is a [3x3] identity matrix.

The term $\bar{\Delta}_a$, that is one introduced by inertial sensors, can be identified as the error affecting the INS measured speed and can be modelled as non-Gaussian noise.

In this case, the estimation which could be obtained from a system defined in such a way should be only sub-optimal, due to the fact that a KF is able to guarantee the optimal estimation only when additive Gaussian noises are present [8].

In order to obtain an estimation which could be approximately considered as optimal it is necessary to perform some modification on the status model: such modifications should have to take into account the non-Gaussian noise components corrupting the outputs of the system.

Therefore, assuming this term to be ideally Gaussian might degrade the system performances. [8]

In this paper, a new approach that allows to reach the optimality of the position estimation is proposed: in particular, through this approach, the optimality is achieved through the linear modelling of the error present on sensors through the techniques described in the following and the extension of the dimension of the Kalman status vector.

At this end, the dynamic variables which characterise the behaviour of the error present in the sensors, conveniently corrupted by white noise with known statistics, will be added to the status vector itself.

It has also to be pointed out that different strategies in order to perform the optimal estimation will be adopted according to the PDOP value, that is according to the reliability of DGPS data.

Finally, the estimated error will be then subtracted from the INS data, in order to update, on the basis of the more reliable DGPS data, the inertial navigation system.

In order to achieve optimal estimates under these critical conditions, a two-step algorithm is proposed.

The first step was an accurate error analysis, in order to establish the components having the higher influence in the cinematic status variables and to define their dynamic model.

Successively it has been possible to extend the dimensions of the status vector by adding the noise components and by linearly modelling their dynamics.

The dynamic system should be linear and corrupted by Gaussian noise: then, it should be possible to calculate, through the Kalman approach, an optimal estimation of the status vector.

Let us briefly expose the approaches that have been followed in order to achieve an optimal estimation of the status vector.

8. THE ERROR MODELING CHOSEN APPROACH

The first step to be covered in order to find the optimal estimation of the vehicle position is to define an appropriate model of the non-Gaussian noise corrupting the status' variables of the system.

The main components of this noise are due to accelerometer and gyroscope errors and they are called *drifts*.

The errors of accelerometers are modelled as being composed of *bias* ($\bar{b}$), *random ramp*, *random walk* and a *first-order Markov process* ($\bar{m}$) with a time constant $\bar{\tau}$, being $\bar{\tau}$ a vector composed by three components $\left[\tau_x \quad \tau_y \quad \tau_z\right]$.

The chosen model of the drift rate error $\bar{\vartheta}$ satisfies the following equations:

$$\bar{\vartheta} = \left[\bar{\Delta}_a \quad \bar{r} \quad \bar{s} \quad \bar{m}\right]^T ; \bar{\vartheta}' = A\bar{\vartheta} + \bar{\zeta}; \mathrm{A} = \begin{bmatrix} 0 & 0 & \mathrm{I} & \mathrm{I} \\ 0 & 0 & 0 & 0 \\ 0 & \mathrm{I} & 0 & 0 \\ 0 & 0 & 0 & -\mathrm{I}/\tau \end{bmatrix}; \bar{\zeta} = \begin{bmatrix} 0 & 0 & \bar{\zeta}_1 & \bar{\zeta}_2 \end{bmatrix} \tag{3}$$

where $\bar{\Delta}_a$ is the acceleration drift, $\bar{r}$ is the ramp slope, $\bar{s}$ is the sum of bias, random walk and ramp and $\bar{\zeta}$ is the white Gaussian noise [9].

The same model can be applied to the gyroscope drifts.

Actually, in the definition of (3), components modelled by the first-order Markov process have not been taken into account.

Because of this fact, the model of the error corrupting the outputs of the accelerometer can be defined in the following way.

Let $\bar{\Delta}_a$ be the error corrupting the output of the accelerometer *(drift)* and let $\bar{\Delta}_a{}'$ its derivative with respect the time *(drift rate)*; drift can be defined as a vector composed by components at constant error components $\left(\bar{b}\right)$ and by components given by the sum of bias, random walk and ramp $\left(\bar{s}\right)$, in the following way :

$$\bar{\Delta}_a = \begin{bmatrix} \bar{b} & \bar{s} \end{bmatrix} \tag{4}$$

The model of the error affecting accelerometers has then be modelled as:

$$\Delta_a' = \mathrm{B}\,\Delta_a + \bar{\xi} \tag{5}$$

where B is defined as:

$$B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{6}$$

while $\bar{\zeta}$ is a white Gaussian noise.

Once the dynamic behaviour of the Non-Gaussian noise corrupting the outputs of the accelerometers has been modelled, it is then possible to re-define the model of the status of the system, by increasing the dimensions of the status vector in order to take into account the behaviour of the modelled error components.

In such a way the status' variables vector will be only corrupted by a Gaussian noise and the optimal estimation of the vehicle position will be possible.

9. EXTENDED STATUS VECTOR FOR NON-GAUSSIAN NOISE MODELLING AND FILTERING

The status vector of the Kalman Filter is augmented by adding to it the status variables that model non-Gaussian noise components, so that the augmented final status vector will be

$$\bar{x} = \left(\delta\bar{r} \quad \delta\bar{v} \quad \delta\bar{a} \quad \bar{r} \quad \bar{s} \quad \bar{m}\right)^T$$

Now the system is assumed to be corrupted only by white Gaussian noise, so we can achieve the optimal estimate.

The dynamic variables added to the status vector concern only with the accelerometer drift model.

The drift rate of the gyroscope is evaluated by using a parallel standing-alone KF, whose resulting estimate, $\bar{\Psi}'$, is used to compute the rotation rate of the vehicle $\left(\bar{\omega}\right)$, needed by the integration filter to express the inertial acceleration $\bar{a}_i$, provided by the INS, in terms of a non-inertial acceleration $\bar{a}_{ni}$ comparable with the one provided by the DGPS (application of the 3D motion equations for a rigid body) [10].

The resulting status model is:

$$\begin{bmatrix} \delta\bar{r}' \\ \delta\bar{v}' \\ \delta\bar{a}' \\ \bar{r}' \\ \bar{s}' \\ \bar{m}' \end{bmatrix} = \begin{bmatrix} 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & I & I \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -I/\tau \end{bmatrix} \cdot \begin{bmatrix} \delta\bar{r} \\ \delta\bar{v} \\ \delta\bar{a} \\ \bar{r} \\ \bar{s} \\ \bar{m} \end{bmatrix} + \begin{bmatrix} \bar{\xi}_r \\ \bar{\xi}_v \\ \bar{\xi}_a \\ \bar{\xi}_{rs} \\ \bar{\xi}_s \\ \bar{\xi}_m \end{bmatrix} \tag{7}$$

where $\bar{\xi}_i$ are white Gaussian noise components.

The observation equation is $\bar{y} = H\bar{x} + \bar{\upsilon}$, $\bar{y}$ being the observation vector, H the measurement matrix and $\bar{\upsilon}$ the vector of Gaussian noise:

$$\bar{y} = \begin{pmatrix} \delta\bar{r} & \delta\bar{v} & \delta\bar{a} \end{pmatrix}; \; H = \begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

The estimated position, velocity and acceleration errors computed by the KF, which now represent an optimal estimation of the status variables, are then subtracted from the INS current position, velocity and acceleration, thus updating the values of the INS outputs.

In such a way, the optimal trajectory of the vehicle can be obtained through the integrated use of the DGPS and INS signals.

Unfortunately, this integration scheme is valuable only when the updating measures provided by the DGPS are reliable.

In order to solve this problem a multi-hypothesis approach for the integration has been introduced and in particular the KF estimates have been made dependent of the value of the PDOP parameter, which is the one that gives the reliability index of the DGPS data.

10. THE MULTI-HYPOTHESIS CHOSEN APPROACH

The Multi-hypothesis approach, which has been used in the proposed approach in the integration between DGPS and INS data, is a particular strategy that allows to take into account the variable goodness of the measure, provided by the DGPS.

In such a way, an eventual absence of the DGPS signal, due to satellites problems or to some obstacles located between the user and the satellites, could be compensated through the use of the INS signal, that, in the low or medium term, can be considered as trustable.

The estimation of the $\bar{x} = \begin{pmatrix} \delta\bar{r} & \delta\bar{v} & \delta\bar{a} & \bar{r} & \bar{s} & \bar{m} \end{pmatrix}^T$ vector, that is calculated by the KF, will be subtracted to the current values (e.g. position, speed and acceleration) given by INS and the values of the outputs of the inertial navigation system will be updated.

The KF-based system has been designed in such a way to be able to also adopt, besides to the two basic function (DGPS present, DGPS absent), some other operating characteristics, in order to make it possible to take into account the characteristics of the DGPS signal: in fact, as it has been previously anticipated, the parameter that allows to commute among the different modules is PDOP, that is the Position Dilution of Precision.

The architecture of the Kalman Filter will then be able to commute among different operating characteristics, depending on the PDOP value, that will become the control variable from which the different integration strategies will be applied.

In particular, PDOP is able to point out if the measure of the vehicle position calculated by DGPS is reliable or not: if PDOP is low (<3), data coming from DGPS have to be considered as reliable, while, if PDOP has an average value ($\in$[3,6]) or an high (>6) value, DGPS data will have a low reliability and this fact will influence the trustfulness of the measures and the trustfulness of the estimations.

In order to take into account of the PDOP value, and then of the error on the DGPS measures, the increasing error on the measures is automatically reflected on the increasing of the variance of the noise corrupting the status vector.

In particular, the adopted Multi-hypothesis strategy allows us to take into account the variability of the corrupting noise by adaptively modifying the covariance matrices of the error on the observations and on the status, depending on the PDOP characteristics.

In such a way, the KF is able to switch among different integration strategies, depending on the PDOP current value.

The variance of the observation is run-time changed according to this value: if the PDOP is low (<4) DGPS data are considered reliable and the variance of the observation is small.

This variance increases as the PDOP increases (from middle ($\in[4, 8]$) to high (> 8) values). As the variance of the measure increases, the reliability in current observations is reduced.

In order to do that, the different observation variances, applied as the PDOP varies, are assessed in a preliminary off-line phase on the basis of a suitable training set of data.

In particular, the different variances that are applied to the system measurements have previously been computed on an experimental basis, through a data training set in which PDOP, DGPS measures and corresponding real measures are known.

The developed integration system will be able to follow the vehicle even in the worst case, that is full absence of the DGPS signal.

Through the proposed approach, even if the DGPS is missing, the system can track the vehicle by using the status vector obtained through the subtraction of the estimated cinematic variables from the current values of the INS.

Being the *tight integration* strategy implemented in a closed loop configuration, position, speed and acceleration that have been estimated will become the feedback of the system.

11. EXPERIMENTAL RESULTS

In this paragraph, the results obtained through the proposed integration between DGPS and INS data are proposed.

The system has been tested on an extensive set of real data acquired during many vehicle missions on urban and suburban roads in North Italy.

The results of the DGPS/INS integration system proposed (Fig. 3) derive from the data acquired during a test mission in which DGPS data were missing for some periods (Fig. 2).

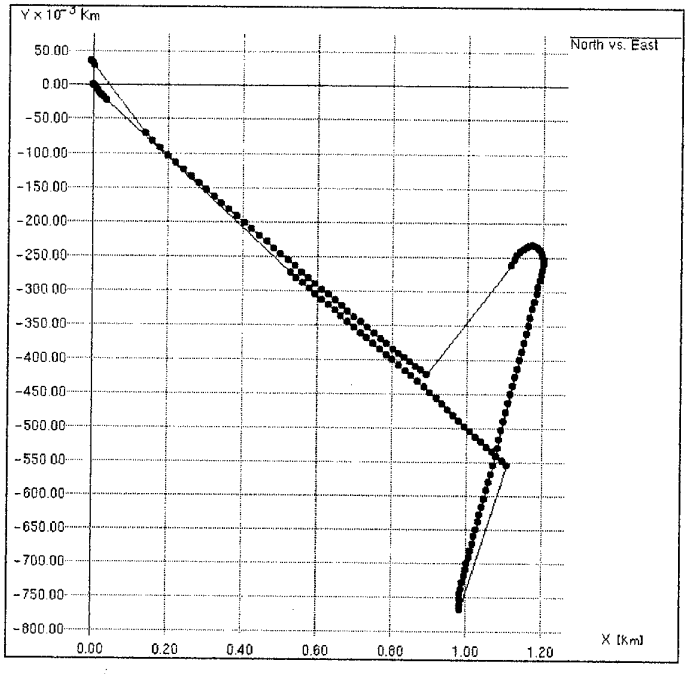


Figure 2. Example of DGPS data related to a 5 minute long mission. The DGPS data are missing four times (where spots are absent).

As it can be observed, the proposed system is able to restore the trajectory of the vehicle, by covering through the INS data the loss of the DGPS data: this integration allows one to have the complete trajectory of the vehicle even when the DGPS signal is missing.

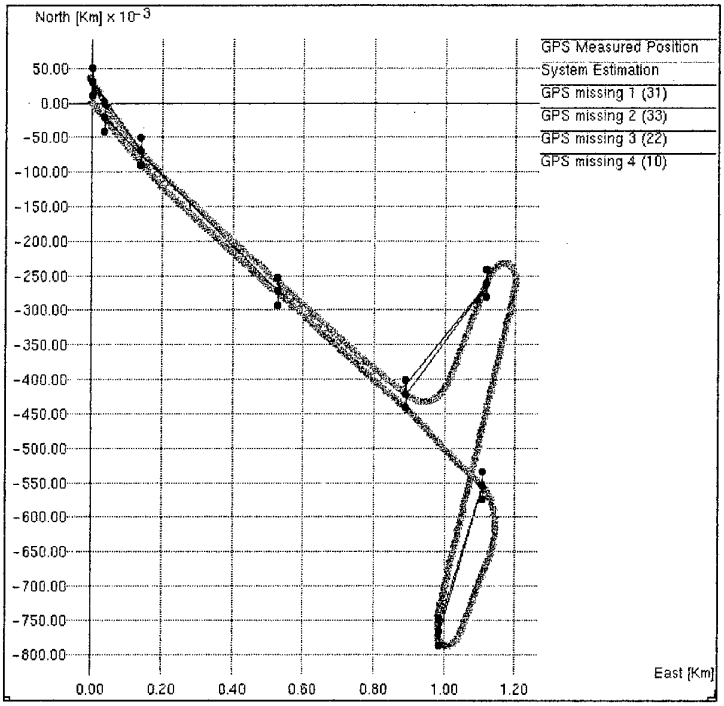


Figure 3. The result of the DGPS/INS integration referring to the mission in Fig.2: the trajectory of the vehicle is restored.

The per-cent error statistics of the test mission are shown in Table 1. The error was computed as the absolute Euclidean Distance between the reference and the estimated trajectory. The sample percentage and the mean error are reported.

Error (mt)	North		East	
	%	Mean	%	Mean
<2	65	0.48	91	0.58
[2, 5)	33	3.11	7	2.47
[5, 10)	2	5.98	2	5.43
≥10	0	0	0	0

Table 1 Per-cent error statistics of the restored trajectory.

Also these quantitative figures show that the error made in the position estimation trajectory remains quite small even in the cases when DGPS signal is absent.

CONCLUSIONS

In the present work the problem of the continuous estimation of the position of a vehicle in the ground has been addressed.
The data that have been used in order to deifne the current position of the vehicle and its displacements are a Different Global Positioning System (DGPS) and an Inertial Navigation System (INS).
The optimal estimation of the vehicle position has been obtained through the integration and fusion of the information coming form each one of the used sensors. The approach followed for the integration is a KF-based Multi-Hypothesis Approach.
The proposed solution has been shown able to provide a robust system able to compute in a precise way the position of the vehicle and resistant to corrupting noise. Experimental tests on the field have shown the goodness of the results obtained through the proposed algorithm.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Roberto Bozzano and Alessandra Tesei for the software implementation of the method and DIVITECH S.p.A. (Address: Strada Lombardore 207 10040 Leini' (TO) – Italy) for providing input data and for their cooperation.

REFERENCES

[1] Simon and El Sherief, "Real Time Navigation using the Global Positioning System", IEEE AES Magazine, pp 31-37, Jan 1995.
[2] Evers and Kasties, "Differential GPS in a Real Time Land Vehicle Environment. Satellite Base Van Carrier Location System", IEE AES Systems Magazine, pp. 26-32, Aug. 1994.
[3] Mattissek and Baumker, "Integration of a Fibre Optical Gyro Attitude and Heading Reference System with Differential GPS", Symposium Gyro Technology, pp. 6.0-6.18, Stuttgart, Germany, 1992.
[4] Harrison, Gallager and Grace, "An Algorithm Providing All-Attitude Capability for three-gimballed Inertial Systems", IEEE Trans. On AES, Vol. 7, No. 3, pp. 532-543, May 1991.
[5] O'Donnell, Slater, Dozier, Duncan and Streeter, "Inertial Navigation: Analysis and Design", McGraw-Hill Book Company, New York, 1994.
[6] Sohne, Heinze and Groten, "Integrated INS/GPS System for High Precision Navigation Applications", Proc. IEEE Position, Location and Navigation Symposium, pp. 310-313, 1994.
[7] Parkinson, "History and operation of NAVSTAR, the Global Positioning System", IEEE Trans. on AES, vol.30, no.4, Oct. 1994.
[8] Brown and Huang, "Introduction to random signals and applied Kalman Filtering", (John Wiley & Sons, 1985).
[9] Nash, Kasper, Crawford and Levine, "Application of Optimal Smoothing to the testing and evaluation of inertial navigation system and components", IEEE Trans on Automatic Control", vol, AC-16, No. 6, pp. 806-816, Dec. 1971.

PAPER No: 16

DISCUSSOR'S NAME: R. Sabatini

COMMENT/QUESTION:

Do you think that your system can be used successfully for Airport Traffic Management (on the ground)?

Do you think that a DGPS/Radar approach would be more suited for this application?

AUTHOR/PRESENTER'S REPLY:

Yes, the system can be used for tracking any vehicles on the ground, but due to integrity issues (related to DGPS) and considering that angular data are not strictly required for airport traffic management, I think the DGPS/Radar approach would be more suitable.

DISCUSSOR'S NAME: R. Sabatini

COMMENT/QUESTION:

Are you thinking about implementing "carrier phase" DGPS techniques together or an alternative to "pseudo-range" techniques?

AUTHOR/PRESENTER'S REPLY:

Not at the moment. However, an upgrade of the system to interpherometric DGPS techniques is foreseen.

DISCUSSOR'S NAME: R. Sabatini

COMMENT/QUESTION:

Which kind of integration scheme have you adopted?

AUTHOR/PRESENTER'S REPLY:

Tightly-coupled DGPS/INS. In particular, a closed-loop integrated DGPS/INS architecture.

PAPER No: 16

DISCUSSOR'S NAME: R. Sabatini

COMMENT/QUESTION:

Does the system process attitude information (i.e. angular data from the INS)?

Reply:

No, only position is computed.

Which kind of INS are you using (strapdown/platform)?

Reply:

Strapdown

DISCUSSOR'S NAME: G. S. Brown

COMMENT/QUESTION:

Do you have data that verifies that your noise is, indeed, Gaussian?

AUTHOR/PRESENTER'S REPLY:

In order to verify the Gaussian assumption made on the experimental data used, in particular concerning the dynamic behaviour of the error introduced by inertial sensors in speed computation, a test for Gaussian (Kolmogorov - Smirnov) has been performed.

In particular, it has confirmed the validity of the assumption

$$\bar{\Delta}'_a = f(\bar{\Delta}_a) + \bar{\xi}_q$$

shown during the presentation.

ARTAS: Multisensor Tracking in an ATC Environment

R.A. Hogendoom and W.H.L. Neven

National Aerospace Laboratory NLR
PO Box 153, 8300 AD EMMELOORD
The Netherlands
e-mail: hogend@nlr.nl, neven@nlr.nl

Summary
ARTAS (an acronym for ATC Radar Tracker and Server) is currently in pre-operational test at four different sites in France, Germany and the Netherlands. The ARTAS system consists of a tracker, responsible for maintaining up-to-date target state vectors, and a server, which handles client subscriptions (e.g. from the ATC display system) and delivers the target state vectors to these clients. An ARTAS system co-operates with adjacent ARTAS systems by exchanging target state vector information.

The main features of the ARTAS Tracker are
- tracking with up to thirty radars (PR, SSR or CMB)
- on-line estimation of the radar systematic errors
- on-line estimation of the radar accuracy and coverage
- high-accuracy position and velocity-vector estimation
- responsiveness to target manoeuvres
- insensitivity to clutter
- target classification

The tracking filters are interacting multiple-model (IMM)-based filters, a four-model filter for high-speed and highly manoeuvring targets and a two-model filter for low-speed targets [1]. The plot-to-track association is based on probabilistic data association (PDA), with special joint probabilistic data association (JPDA) algorithms in case of target close approach situations [2]. Track initiation is done by time-reversed multiple-hypothesis tracking. Target classification is based on Shafer-Dempster reasoning.

Introduction
ARTAS is designed as a track data server. Track data users can subscribe to a certain service and receive the track data in ASTERIX format via a local-area or wide-area network (LAN/WAN, figure 1). Users can be ATC centres, flightplan processing systems, air-traffic management units and so on. Each user can have a dedicated service, taking into account requirements with respect to data contents and update frequency. An ARTAS unit also receives its input data from the radars via the local-area or wide-area network. Furthermore, an ARTAS unit can communicate via the network with other, adjacent, ARTAS units in order to provide a continuous air-picture to its users. Track data from adjacent units is used to accelerate the initiation of tracks at the border of the unit's own domain of interest (DOI) and to smooth the transition of a track from one unit's DOI to another unit's DOI. Finally, when there is sufficient coverage of the own unit's DOI by adjacent ARTAS units, the adjacent ARTAS units can take over the surveillance in case of an own unit failure. Thus, enhancing the overall reliability of the surveillance.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

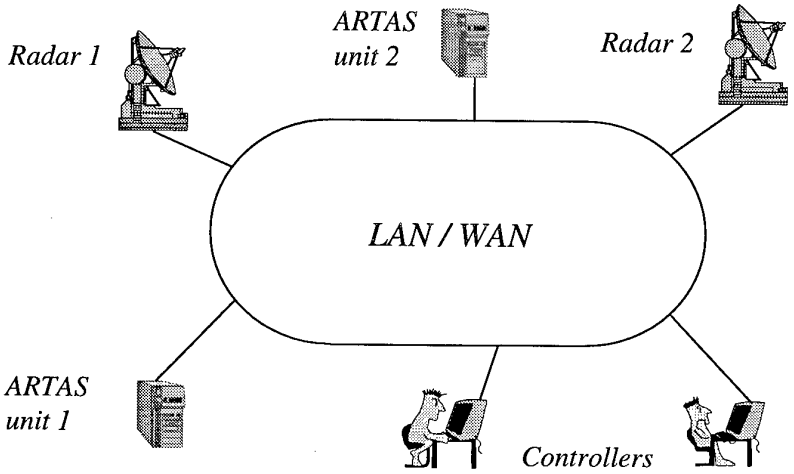


Figure 1.The ARTAS Environment

The internal structure of an ARTAS unit is shown in figure 2. The Router Bridge is the interface to the external network. It pre-processes the incoming radar data, i.e. it performs format checks and sectorisation of the plot data and keeps track of the operational status of the radars. The Server is responsible for the handling of ARTAS user requests and the distribution of the track data, according to the different user services. The most simple service that is provided is a regular broadcast of all track data. MMI/Supervision is the man-machine interface and supervision unit. It provides a basic display of the unit tracks and control functions for the ARTAS unit. The Tracker, finally, is responsible for keeping an up-to-date air picture. An ARTAS unit consists of two identical chains of a Router Bridge/Tracker/Server/MMI/Supervision subunits. All subunits operate in a multiple-computation redundancy mode; that is, there is a master and a slave subunit that both perform the same processing, except that the slave subunit does not provide any output. Instead, the slave performs some additional processing to keep master and slave in synchronisation.

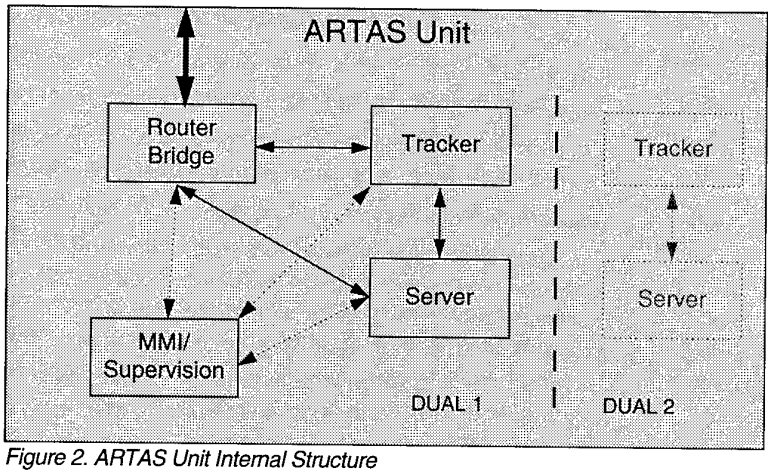


Figure 2. ARTAS Unit Internal Structure

All the ARTAS subunits run on off-the-shelf hardware and are programmed in ADA, except for the MMI, which is programmed in C++.

The ARTAS Tracker

Basically, the task of the tracker is to provide estimates of the state of aircraft in the domain of interest of the ARTAS unit. It makes use of maximum 30 sensors; present types are primary radar (PR) and secondary surveillance radar (SSR). Extensions to incorporate Mode-S and automatic dependent surveillance (ADS) data are foreseen for 1998.

A prime requirement for handling multisensor data is the ability to cope with sensor alignment errors, i.e. systematic radar errors like position bias, range- and azimuth bias, but also time-stamping bias and transponder-delay error. The latter is an example of a, so-called, micro-error; a systematic error that depends on the object being tracked. The former errors are macro-errors; they only depend on the sensor involved. Unfortunately, systematic errors may change in time, due to e.g. changing atmospheric conditions and radar maintenance. Therefore, the ARTAS Tracker contains an on-line systematic-error estimation module that is able to track varying systematic radar errors.

Another requirement for handling multisensor data is a proper treatment of coordinate transformations. This becomes a more obvious problem when the size of the system area becomes large. ARTAS uses WGS84 as a reference system. Measurement processing and track update processing are done in local Cartesian systems, such that the error, induced by coordinate transformations, is minimised. This implies that all sensors and all tracked objects have their own local Cartesian system that may change in time when objects move.

Track continuation uses the reports of all available sensors to estimate the state of a target. Each track extrapolation/update cycle is based on the reports of a single sensor, though. Subsequent cycles, however, may be of entirely different sensors. Prior to the track update, all the relevant reports are corrected for micro-errors (systematic errors that vary from target to target) and slant-range effects. Track continuation is discussed in more detail below.

Track initiation is done based on the reports of single sensors only. It is based on multiple-hypothesis tracking (MHT) and is done retrospectively [3]. Considering the fact that a new target generally enters the coverage of only a single radar, the gain of a shorter track initiation delay did not warrant the additional complexity of a multi-radar initiation in a civil ATC environment. This trade-off may not be valid in a military environment, though.

The ARTAS Tracker maintains aircraft and non-aircraft tracks since, in many cases, the best way of dealing with anomalies, like reflections and sidelobes, is to track them and to classify them as being non-aircraft. To that end, the ARTAS Tracker contains a track classification module, which classifies tracks using Shafer-Dempster reasoning [4]. The criteria, used in the classification, are based on radar environment characteristics, target behaviour and a set of models for specific anomalies, like reflections and sidelobes. An advantage of Shafer-Dempster reasoning is the ease with which additional criteria, like target signature information, can be incorporated into the classification process.

Track Continuation

For the ARTAS Tracker, a Bayesian approach to track continuation was adopted. This approach did prove to yield a high-performance tracker, as experience with the NLR JUMPDIF prototype tracker has shown [1].

Basically, there are four major problems that occur during track continuation

1. Non-linear aircraft dynamics during a turn
2. The association of measurements with existing tracks
3. The occurrence of outlier measurements (non-Gaussian measurement noise)
4. Sudden starts and stops of manoeuvres

For each of these problems, adequate solutions were already developed for the JUMPDIF prototype [1]; the result, an Interacting Multiple-Model Probabilistic Data-Association (IMMPDA) algorithm with Extended Kalman Filters (EKF) [6] was used in extensive performance tests. The results of these performance tests were used as a basis for the ARTAS Tracker performance requirement specification. A number of improvements, with respect to the JUMPDIF tracker, were made in the ARTAS Tracker, though.

For target resolution situations, new joint probabilistic data-association (JPDA) algorithms were developed [2] that perform considerably better than the probabilistic data-association (PDA) algorithm, that is used in JUMPDIF. These JPDA algorithms, however, require more computations than the PDA algorithm. In order to save CPU-load, these JPDA algorithms are only used when a target resolution situation is detected.

The ARTAS Tracker is required to track targets down to zero groundspeed. In general, it is not

necessary to track low-groundspeed targets with an advanced four-model (left turn, right turn, change of groundspeed, straight flight) IMMPDA filter to get a good tracking performance. Therefore, a simplified two-model (manoeuvring flight, straight flight) IMMPDA filter is used to track these targets.

JUMPDIF contained a two-model (climb/descent, level flight) IMMPDA filter for SSR mode-C measurements. In the ARTAS Tracker this filter was replaced by a three-model (climb, descent, level flight) IMMPDA filter in order to be more responsive to changes in the rate of climb/descent. Furthermore, two algorithms to estimate the target altitude in absence of SSR mode-C information were implemented. One algorithm, Triangulation, is discussed in more detail below. The other algorithm, Height-from-Coverage, uses the assessed coverage of each radar, that detects or does not detect the target, to calculate a height interval for the target. Although not very accurate, using the result of this algorithm is often better than using a default altitude.

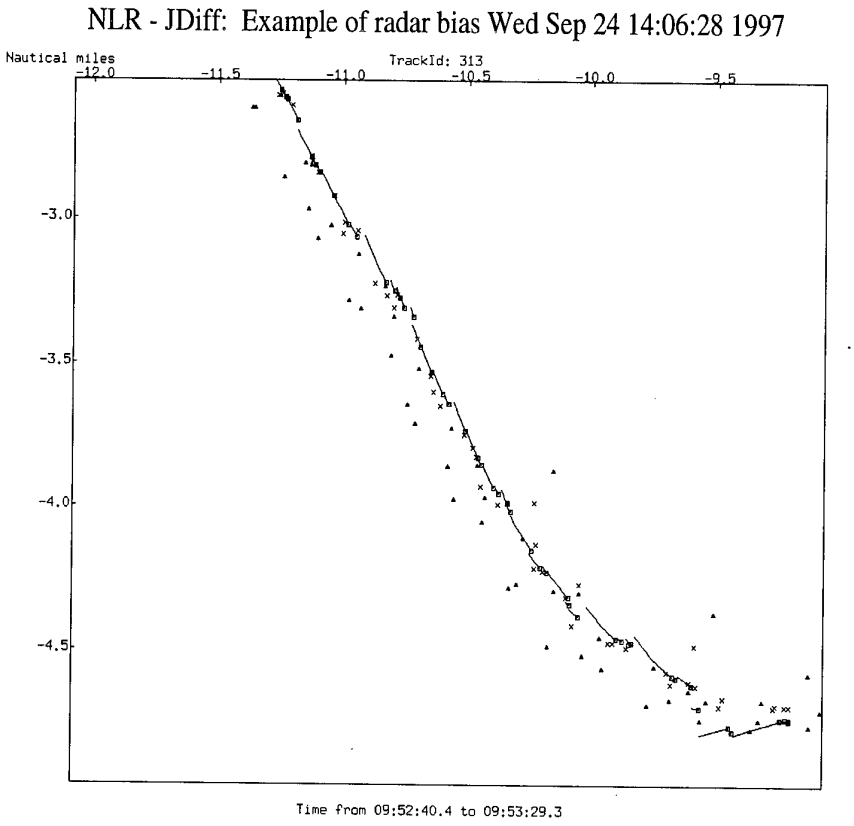


Figure 3. Example of biased plots

The key solution to the multisensor track continuation problem that is applied in the ARTAS Tracker is a proper correction for the macro- and micro-systematic errors of all involved measurements, before they are used within the track extrapolation/track update cycle. This essentially reduces the multisensor problem to a single-sensor problem. The time sequence of track extrapolation/track update cycles, obviously, contains track extrapolation/track update cycles for all the available sensors. The difference between cycles for different sensors is the use of a different measurement matrix for the Extended Kalman filters.

Figure 3 shows a track, departing from Schiphol airport, that uses biased measurements from three different radars. Triangles indicate the raw plots, crosses the nearest-neighbour plot positions

(corrected for the estimated radar biases) and the squares the updated track position. The vectors indicate the predicted flightpath up to the next measurement instant. Figures 4 and 5 show the ARTAS Tracker estimates of the groundspeed and SSR mode-C height of this track, respectively.

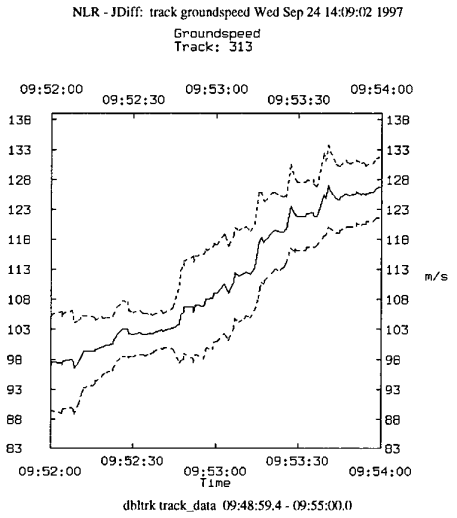


Figure 4. Track groundspeed estimate

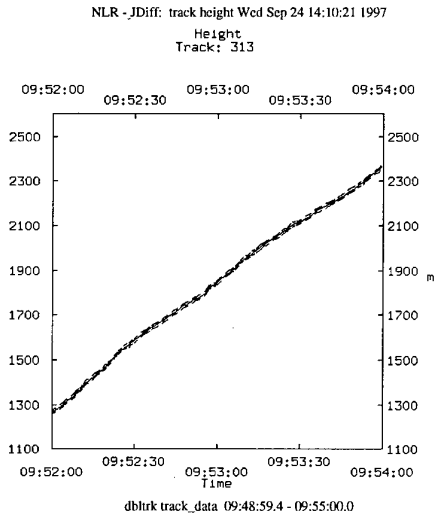


Figure 5. Track mode-C height estimate

Systematic Radar-Error Estimation

The ARTAS Tracker estimates the following (macro-)systematic errors:

- range bias
- azimuth bias
- range gain (a range bias that increases with increasing range)
- antenna squint (non-verticality of the plane of the radar beam)
- verticality error (antenna rotation axis not perpendicular)
- time-stamping bias

The problem with dynamic estimation of the (macro-) systematic errors is that, in principle, the filter equations are coupled with the track continuation equations of the individual tracks. It is, of course, very well possible to make a selection of a small number of well-behaved tracks and to solve the resulting set of equations. In [5], a different approach is taken, which decouples the equations for (macro-)systematic error estimation from the track continuation equations. Effectively, it comes down to integration of the innovations of all tracks and filtering these innovations with a Kalman filter. Due to the larger timeconstant of the systematic error process, the filtering equations become independent of the individual track maintenance equations. This algorithm is implemented in the ARTAS Tracker and uses a selection of non-manoeuvring tracks in order to save CPU-load and to increase the speed of convergence of the estimation process. Figures 6 and 7 show results of the (macro-) systematic-error estimation process on a 2-radar PR scenario.

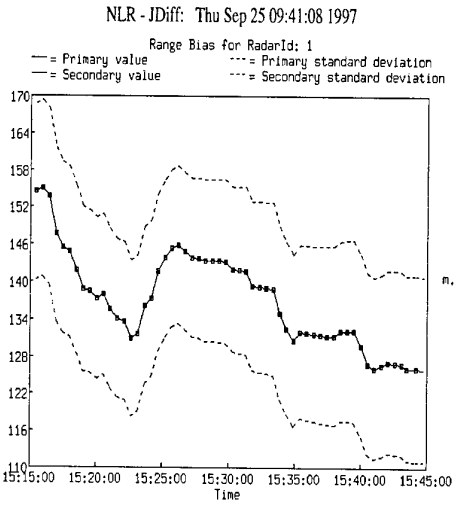


Figure 6. TAR estimated range bias

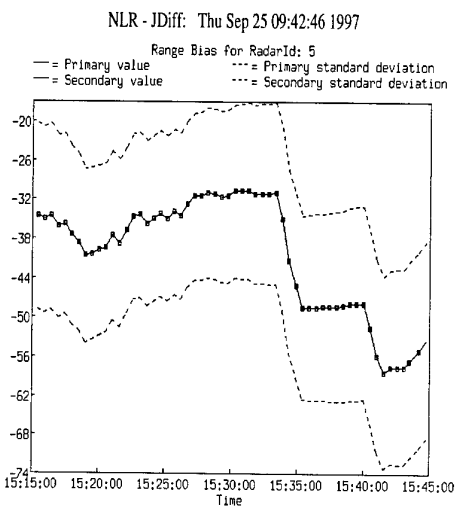


Figure 7. LAR estimated range bias

Triangulation

After estimation of the systematic radar errors that are radar-dependent only (macro errors), the track-related errors (micro errors) can be estimated. Within the ARTAS Tracker, these micro errors consist of the transponder delay error (i.e. the difference between the actual delay and the nominal value of 3 microsecond as specified by ICAO) and the geometric height, estimated from range-azimuth position measurements in a multi-radar environment.

A general solution to this problem is to extend the state vector of an object with these components and to extend the corresponding extended Kalman filter equations accordingly. Since this is a very costly solution (in terms of CPU), we have looked for a robust method that is not coupled with the track continuation equations. In situations where an SSR radar has a co-located primary radar, a robust method to estimate the transponder delay error is to average the difference in range measurements of the two radars. In other situations, the transponder delay error and geometric height estimations are coupled.

Consider the situation that two non-co-located radars observe an object at the same moment in time. To perform triangulation, we use the difference between the projections of the plots to a common 2-dimensional Cartesian coordinate system (the track-local coordinate system) as the innovation term in a Kalman-like filter update step for the estimation of the transponder delay error and the geometric height.

Since a simultaneous measurement of one object by two non-co-located radars is quite unusual, we perform a triangulation on the basis of a triplet of projected plot positions (under the condition that the track groundspeed and course are constant): the first and third projected position are interpolated to the time of the middle plot.

The performance of this algorithm depends, among others, on the geometric configuration of the radars involved: the middle plot should be from a different radar than the other two plots, with a line-of-sight opposite to that of the other radars, and as close to the object as possible.

In figure 8, we see a part of a track from a live data collection. The recording was made for 3 secondary and 2 primary radars, but the Tracker was run with only the primary plot data. The track is flying at FL 290 (8839.2 m); the plots are not corrected for systematic radar errors. The estimate of the geometric height and the 1-sigma margin are given in figure 9; the initial estimate is 6000 m.

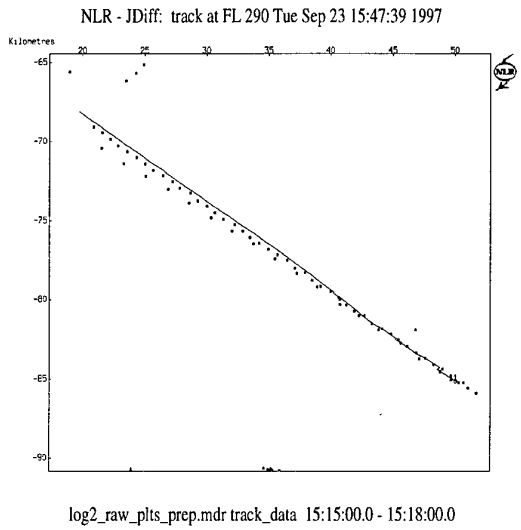


Figure 8. Track observed by 2 PR radars

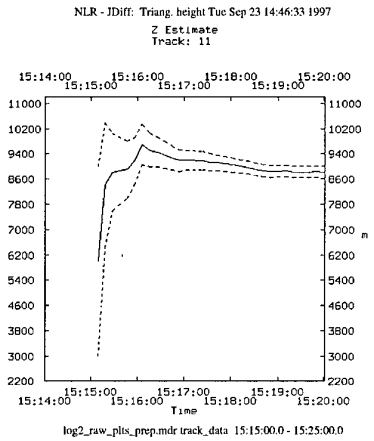


Figure 9. Triangulated height as function of time

Conclusion

Adequate systematic error estimation is a pre-requisite for accurate multisensor tracking. In the ARTAS Tracker, several powerful methods are employed for the on-line estimation of both macro- and micro-systematic errors. These methods provide accurate estimates of the systematic errors as shown by a number of examples. By having accurate systematic error estimates, the multisensor problem is essentially reduced to a time-sequential single-sensor problem, which is, obviously, much easier to solve.

References

[1] H.A.P. Blom, R.A. Hogendoorn and B.A. van Doorn, Design of a Multisensor Tracking system for Advanced Air-Traffic Control. In: Y. Bar-Shalom(ed.), Multi-Target-Multisensor Tracking: Applications and Advances, Vol. II, Artech House, 1992.

[2] E.A. Bloem and H.A.P. Blom, Joint Probabilistic Data Association Methods avoiding track coalescence, Proc. 34th IEEE Conf. on Decision and Control, December 1995, pp. 2752-2757.

[3] R.A. Hogendoorn, H.A.P. Blom, Bayesian Track Initiation by time-reversion of Trajectory Models, Proc. IEEE Int. Conf. on Control and Appl., April 1989, Jerusalem, paper WA-1-5.

[4] P.L.. Bogler, Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems, IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-17, No. 6, 1987, pp. 968-977

[5] B.A. van Doorn and H.A.P. Blom, Systematic Error Estimation in Multisensor Fusion Systems, SPIE Conf. on Signal and Data Processing of Small Targets, April 1993, Orlando

[6] H.A.P. Blom, A Sophisticated Tracking Algorithm for Air Traffic Control Surveillance Radar Data, Proc. Int. Conf. on Radar, pp. 393-398, Paris, May 1984

PAPER No: 17

DISCUSSOR'S NAME: P. Hoogeboom

COMMENT/QUESTION:

The desired accuracy you mentioned is 11 meters. The aircraft you measure are much larger, and can show glints. How does this influence the accuracy and the correction schemes?

AUTHOR/PRESENTER'S REPLY:

The radar measurements presented to the system are point measurements, probably based on the centroid of the reflected radar image. So the estimator cannot take into account the physical extent of the target. On the other hand, the estimated bias is based on the measurement of a large number of targets, both close by and far away, so the physical extent and the glints, resulting from that, do not affect the estimator very much.

DISCUSSOR'S NAME: M. Desbois

COMMENT/QUESTION:

Does the integration of mode S or GPS-equipped aircraft modify your approach concerning the selection of tracks for the debiassing.

AUTHOR/PRESENTER'S REPLY:

The method selected for ARTAS remains mainly on sensor selection and current traffic selection based on several criteria, among with the quality of tracks that takes into account mode S or GPS information.

Low Altitude Wave Propagation Effects in a Transmitter Independent Receiver Network.

Dipl. Eng. Nikos J. Farsaris and Prof. Peter P. Stavroulakis.

Telecommunication Systems Institute of Crete
37 Iroon Polytechneion Ave.
73100 Chania Crete Greece.
Tel: +30-821-28457, Fax: +30-821-28459
Mail: njf@tsinet.gr, peter@tsinet.gr

Abstract:

For the design and implementation of a Passive Multistatic Early Warning and / or Fire Control Radar (T.I.R.N., Transmitter Independent Receiver Network, [1]) many aspects include the environmental conditions that T.I.R.N's may encounter at operational conditions.

Even operating in a standard troposphere, the result of the refraction phenomenon is the curved propagation of the electromagnetic waves. The problem becomes more complicated at low altitudes, where tropospheric refractive index variability, multipath propagation and clutter (sea or ground) must be taken into consideration in order to avoid erroneous results for the location of a target as well as its velocity estimation. Although simplified models of the above environmental parameters are used in this analysis, (e.g. lateral refraction variability is not studied here, and Weibull Distribution Model is taken for granted as regards the sea clutter), useful results can be produced for the evaluation of multistatic systems.

In this paper a four-receiver model using the signal of a non-cooperative transmitter is evaluated for its sensitivity at the above conditions and is compared with a conventional monostatic radar operating with the same signal. It is shown that the T.I.R.N. proposed is a more efficient system than its monostatic equivalent. The results presented, despite of being strictly theoretical (not yet evaluated in real operational conditions) may help the design of multistatic radars.

1. Introduction.

As it has been discussed in [1] T.I.R.N.s operation depends strongly on the spatial separation of the receivers. Actually it is shown [2] that for any target case and mode of location (Triangulation, T.D.O.A., or Range Sum) on a 4-receiver model, the receivers must not be located at the same planar surface for 3-D location and dopller vector velocity measurement. (Later in this paper this constraint will be strengthened.)

That means that receivers must be placed in different altitudes also. So a realistic model includes receivers in an altitude range of 30 to a mere 2000 meters.

Also we must note that in this paper the radar transmitter is non-cooperative it is located in an ambiguous location, and is using an unknown antenna pattern of radiation. One of main purposes of T.I.R.N. design is the ability to use enemy radars as target illuminators.

2. Model Consideration.

It is understandable that if the transmitter and/or the target is located at a low altitude then tropospheric effects must be considered in order to avoid erroneous results. In this paper the model considered is consisted of four receivers: Two at an altitude of less than 100 m , one at an altitude of about 500m and one at altitude of about 1500m. The distances between the receivers vary from 10 to 40 km. (Fig.1) It is assumed that for adequate triangulation location the receivers are equipped with 2 or 4 element monopulse antennas. The reason of using monopulse techniques is that they give good angular discrimination even retaining wide beamwidths for best bistatic coverage. Their elements are assumed to have wide horizontal and narrow vertical beamwidths. (multiple horn antennas or interlaced phased arrays are considered).

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

For this configuration a useful T.I.R.N. range is about 120 km.

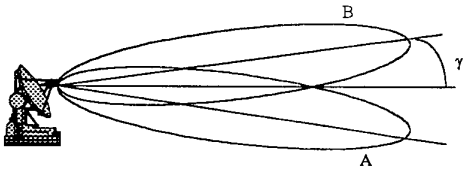


Fig 1.Two- Beam Monopulse antenna patterns. Targert is in boresight level when signal strengths at A, B are equal. 4-element can be used for both azimuth and elevation discrimination.

Then we can consider cartesian coordinate systems attached to the receivers all aligned to a central "point of interest" coordinate system. That point of interest could be any receiver, a command post of a weapon system in the area. In modest latitudes (x,y,z) can be North, West, Altitude corrected by first order approximation for the mean radius of the earth R=6378km. That means that altitudes must be corrected by the factor d^2/R and the angles between the axis by an order of $180d/\pi R$ degrees or about d/π mils if R,d in km. More accurately the axes must be corrected by the exact differences of Latittude and Longitude of the receiver and the "point of interest".

The main purpose of this paper is to correct radar measurements to real atmospheric conditions at least as they are theoretically modeled.

2.1. Some Free Space Calculations.

If transmission, scattering and reception were taking place to free space then the location and the velocity measurements would require simple algebraic calculations. The corrections mentioned earlier can be made during installation. When using triangulation method the location of the target or transmitter can be determined by the following set of calculations:

If φ_1 and φ_2 are the look angles (azimuths) and (x_1,y_1) (x_2,y_2) are the receiver locations then the transmitter location (x,y) is extracted by the linear system:

$$\begin{bmatrix}\cos\varphi_1 & \sin\varphi_1\\ \cos\varphi_2 & \sin\varphi_2\end{bmatrix}\cdot\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}x_1\cos\varphi_1+y_1\sin\varphi_1\\ x_2\cos\varphi_2+y_2\sin\varphi_2\end{bmatrix} \quad (1)$$

And for z

$$z = z_i + d\cot\vartheta_i$$

$$d=\sqrt{(x-x_i)^2+(y-y_i)^2}$$

(2)

In this case θ is the measured from the z axis of the ith transmitter. The reason z (altitude) is separeted will be discussed later in this paper.

A serious shortcoming of the triangulation method is the target "ghosts" that can be presented if there are more than one targets in the area. So additional method of location is required for target location verification and this can be T.D.O.A. or Range Sum Method [2],[3],[4] if the target is close or remote to the T.I.R.N. area respectively. Range Sum is also desirable for a single receiver target location. These methods are generally non-linear and so computing complexity makes them more slow.

An interesting feature of T.I.R.N's is the velocity vector estimation by measuring (where the signal is appropriate) the Doppler shift at each receiver. It is known that :

$$f_D = -\frac{f_t(v_t+v_r)}{c}$$

(3)

where f_D and f_t are the Doppler frequency shift and the transmitted frequency respectively, v_t and v_r are the radial velocities of the target relative to the transmitter and the receiver. Note that they are the signed magnitudes of the projections of the real target velocity vector on the transmitter-target (t-t) and receiver-target (r-t) directions.

Using a four-receiver T.I.R.N. with synchronized receivers after demodulation we get signals with frequencies proportional to

$$\delta = v_i + v_{ni} + v_e$$

(4)

or

$$\delta = v_{ei} + v_e$$

(5)

where v_i is the r_i-t signed magnitude (algebraic value) of the velocity vector projection (I=1,2,3,4), v_e is systematic error made because of v_t and a possible frequency demodulation error -the same for all synchronized receivers-, v_{ni} noise (random) errors which can be taken as uncorrelated and $v_{ei}= v_i+v_{ni}$ is the estimation of measurement after the systematic error removal.

For the velocity vector estimation any Cartesian or spherical target-centered coordinate system is convenient. If v(u,w,s) is the real target velocity vector and v_i its projection in a random direction

then:

$$\mathbf{v}\cdot\mathbf{v}_i = v_i^2 \qquad i=1,2,3,4 \tag{6}$$

Applying equations (4), (5) ,(6) at the coordinate system defined we get for the r_i-t directions the linear following system:

$$u\cos\phi_i \sin\theta_i + w\sin\phi_i \sin\theta_i + s\cos\theta_i + v_e = \delta_i \quad i=1,2,3,4 \tag{7}$$

where φ,θ are the angular coordinates of the receivers relative to the target and v_{ni} is embodied to u,v,s.

If the same methodology as the on-line solution is followed then this is equivalent to:

$$\mathbf{G}\bar{v} = \bar{\delta} \quad \bar{v} = [u \quad w \quad s \quad v_e]^T$$

$$\mathbf{G} = \begin{bmatrix} \sin\theta_1\cos\phi_1 & \sin\theta_1\cos\phi_1 & \cos\theta_1 & 1 \\ \sin\theta_2\cos\phi_2 & \sin\theta_2\cos\phi_2 & \cos\theta_2 & 1 \\ \sin\theta_3\cos\phi_3 & \sin\theta_3\cos\phi_3 & \cos\theta_3 & 1 \\ \sin\theta_4\cos\phi_4 & \sin\theta_4\cos\phi_4 & \cos\theta_4 & 1 \end{bmatrix}$$

$$\bar{\delta} = [\delta_1\delta_2\delta_3\delta_4]^T \tag{8}$$

If use of the receiver centered coordination systems is desirable then all the cosines of the above system must change sign.

Note that x,y,z,u,v,s then is the 6-D coordinates of the target at the point of interest.

Agtually in free-space a useful **G** could be:

$$\mathbf{G} = \frac{1}{r}\begin{bmatrix} x_1 - x & y_1 - y & z_1 - z & r \\ x_2 - x & y_2 - y & z_2 - z & r \\ x_3 - x & y_3 - y & z_3 - z & r \\ x_4 - x & y_4 - y & z_4 - z & r \end{bmatrix}$$

$$r = \sqrt{x^2 + y^2 + z^2} \tag{9}$$

3.Standard atmosphere case.

Another simple case for T.I.R.N. environmental conditions is when propagation is taking place in a Standard atmosphere with propagation factor K≠1 (usually K=4/3 but can raise to 1.4 at Mediterranean environments.) In that case all the radiowave measured parameters (distances, alltitudes, and elevation angles must be corrected)

The correction is simple and straightforward for z assuming that propagation factor bends the propagation line (ray) towards the earth's surface then equation (3) becomes:

$$z - z_i = d\cot\vartheta_i - \frac{K-1}{K}\frac{d^2}{2R} \tag{10}$$

or

$$z = z_{fs} - \frac{K-1}{K}\frac{d^2}{2R} \tag{11}$$

where z_{fs} is the free space z estimation.

That also means that for the angle of departure of this ray:

$$\cot\vartheta_{id} = -\frac{\partial z}{\partial d}_{d=d} = -\cot\vartheta_i + \frac{K+1}{K}\frac{d}{R} \tag{12}$$

From equation (10) comes the strengthening of the condition that not all the receivers must be in the same level plane. If D_{max} is the maximum distance of 2 receivers of the T.I.R.N. then a receiver must be located at an altitude at least order of magnitude larger than $[(K-1)D_{max}^2/KR]$ That means that for 40 Km receiver separation at least a pair of them must have 500 meters of altitude difference or higher if possible.

Note all θ's at equation (8) are actually angles of departure, so a correction must be done to them too, in order to avoid erroneous velocity estimation.

It must also be noted that even if T.D.O.A. or Range Sum is used for target location the same z and θ corrections must be made. This is because the propagation model is not changing by changing the method of location.

4.Tropospheric waveguide (duct) case

In environments over sea evaporation ducts may occur
at very low altitudes (less than 30 m) while over ground irregularities of the air temperature and humidity may cause irregular refractive index (or modified refractive index variability able to cause waveguide effects at altitudes from 100 to 300 m. These phenomena are more severe in environments with islands. In Mediterranean sea there are waveguides at 70% of time in the summer. If both the target and the receiver are in the waveguide then real elevation angle data are difficult to be obtained.

These ducts of course are really not disturbing the receivers located above them so it is better for the T.I.R.N. if the higher placed receivers act as height finders (z-finders to be more precise) in order to avoid evaporation ducts. The opposite

problem is taking place if the duct is taking place over ground. Then the lowest receivers are more likely to receive direct rays. The fact that only rays with elevation angles between ±1.5 degrees are waveguided [5] leads to a simple rule that is: Always trust the larger $|\theta|$ given by the receivers and if all $|\theta|$'s are smaller than 1.5 degrees trust the highest receiver. That rule

Being more precise in that case equation (1) can still be used (lateral refractive index variabilities are anyway smaller than vertical), but propagation in that case at the vertical level cannot be confirmed by all receivers so the one with the less altitude discrimination ambiguity can be used and corrected as described at (10). Bearing then in mind that ranges in low altitudes are measured in tenths of Kilometers and altitudes in hundrents of meters then a good approach in velocity estimation can be done using (10) with the corrected altitudes and assuming that :

$$\frac{1}{r} \cong \frac{1}{d} \cong \frac{1}{\sqrt{d^2 + z^2_{ambig.}}} \tag{13}$$

Actually the above approximation can be used for all range conversions at low altitudes. A simple calculation will show that in 150 Km range even 5Km altitude difference would give an error in range at about 84 m while in 30 Km , 1Km altitude difference would give a range difference of 17 meters (or the length of a modern fighter!) with that simplifications **G** for an approximate velocity estimation can be written as:

$$\mathbf{G}=\frac{1}{r}\begin{bmatrix} x_1-x & y_1-y & z_1-z_{fs}+\dfrac{(K-1)d_1^2}{2KR} & r \\ x_2-x & y_2-y & z_2-z_{fs}+\dfrac{(K-1)d_2^2}{2KR} & r \\ x_3-x & y_3-y & z_3-z_{fs}+\dfrac{(K-1)d_3^2}{2KR} & r \\ x_4-x & y_4-y & z_4-z_{fs}+\dfrac{(K-1)d_4^2}{2KR} & r \end{bmatrix} \tag{14}$$

5. Multipath propagation

During operation at low altitudes not only the direct ray but also the reflected ray must be taken into consideration. The fact that all the model consideration has been done with unknown transmitter location one useful approximation is to consider the target itself a secondary transmitter and so the problem leads to a point to point radiolink problem. [5],[6].

Dealing with multipath in a T.I.R.N. having monopulse antennas however the problem becomes more complicated because the reflected signal undertakes different gains at the lower and upper beams of the antenna or their sidelobes. That would give significant θ errors. It has been stated elsewhere in this paper that elevation measurement is the one which is more sensitive and so does vertical velocity s then. An attempt is going to be done of calculating these effects:

As shown in Fig.2 a monopulse antenna with beam offset angle γ of the foresight is aiming at a target at an angle of θ≈90 degrees It is clear then that the lower beam aims at θ+γ while the upper beam aims at θ-γ . In case of no multipath then the amplitudes of electric field E in both sidelobes lower (A) and upper (B) would be

$$E_A = E_B = E\sqrt{D(\gamma)} \tag{15}$$

where D(γ) is the directivity of the antenna at an offset γ. In case that the reflection point is located at distance d_m and altitude z_m and the reflection coefficient is ρ.exp(jχ) then:

$$E_{Am} = E\sqrt{D(\gamma)} + \xi E\sqrt{D(-\gamma+180-2\vartheta)}$$
$$E_{Bm} = E\sqrt{D(\gamma)} + \xi E\sqrt{D(\gamma+180-2\vartheta)}$$
$$\xi = \rho e^{[j\chi+(z-z_m)\sin(90-\vartheta+d/R)]} \tag{16}$$

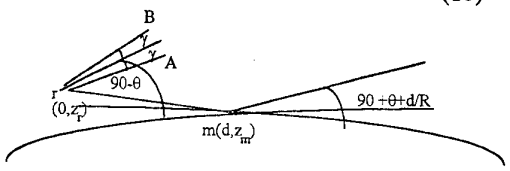


Fig 2. Geometry of the Reflection on curved earth

Then the difference of theese two terms is proportional to antenna directivities at these two angles It has been shown [6] that modified monopulse antennas can eliminate this error. Also is shown that the elevation error can be either positive or negative. Also simple Sum-Delta monopulse antennas with vertical sidelobe cancellers might be used. If multiple reflection points do exist then the vector sum must be taken and arbitarity in phase will most likely reduce the effect.

This effect is strongly reduced if the target is height-monitored by two antennas simultaneously for the same reason.

The fact is that in a T.I.R.N even at intense multipath environments there is a little possibility of deep target fading because of the simultaneous use of more than one receivers in any case.

6. Clutter environments.

It is generally known that little bistatic clutter data are available. The fact that antennas with wide horizontal beam angles are taken into cosideration leads to the fact that large clutter signals are to be expirienced during T.I.R.N.'s operation.

From some bistatic measurements though [7] , it has been stated that bistatic to monostatic clutter is about -3 to -20 dB lower than monstatic[8]. Dopller spread using either Rayleigh or Weibull distributions is not expected to exceed that of a monostatic radar.

Clutter rejection though can be better by combining MTI or Dopller shift measurements taken by the receivers and exclude (filter) measures that seem quasi-static to all the receivers.

Anyway, since no operational T.I.R.N. is in use currently, simulations (based on bistatic and monostatic measurements availlable) have shown that for Rayleigh and Weibull clutter distributions 4-receiver to monostatic clutter rejection is about 6 to 10 dB better than a monostaticradar with the same antenna pattern.

Conclusions.

The above results although mathematically modelled using simple Electromagnetic wave propagation theory with the appropriate aproximations, simmulation results with EREPS and RPOT shareware software of NOSC (U.S.Navy) for standard atmosphere and some waveguide cases have been run and comply well with the theoreticall results presented here.

Multipath for monopulse antennas and Clutter cases still require some research in order to be fully covered, mostly because they are signal-dependent and to this time only low PRF and CW have been examined.

Anyway the capability of 6-D detection of a T.I.R.N., although is effected by environmental parameters that effect other types of radar still remains better that a monostatic radar. Although computational and design complexity is higher the much promising performance of a T.I.R.N. and the fact that it is totally passive makes it an attractive system model for further development.

References

[1] N. J. Farsaris, Prof. P. P. Stavroulakis "Target Detection via Measurements Taken by a Transmitter Independent Receiver Network." AGARD CONFERENCE PROCEEDINGS 582 paper #29, 1996

[2] N. J. Farsaris, Prof. P. P. Stavroulakis, Prof. S. S. Kouris: "Neural Network Designs for T.I.R.N. Applications." International Conference in Communications and Control COMCON VI 1997

[3]R.J. Doviac, C.M. Weil "Bistatic Radar Detection of the Melting Layer" Journal of Applied Meteorology 11, 1012-1016 1972

[4]J. Wurman et. al. : "Design of a Bistatic Dual Dopller Radar for retrieving vector winds Using one transmitter and a Remote Low Gain Receiver." IEEE Proc vol. 82 pp 1861-1871 1994.

[5]Prof. S. S. Kouris: "Antennas and Electromagnetic wave propagation Theory" ("Στοιχεία Θεωρίας Κεραιών και Διαδόσεως Ηλεκτρομαγνητικών Κυμάτων") Zitis Publications 1985 (in Greek)

[6]Merrill Skolnik "Radar Manual" 2e McGraw Hil1990, Chapter 20: David Murrow "Heigth finding and 3-D radar"

[7] Nicholas Willis "Bistatic Radar" Arthech House 1991 Chapter 9 .

[8]A.M. Demvillle "The Bistatic ReflectionFrom Land And Sea of X-Band Radio Waves." Parts I,II Gec Electronics Ltd Stanmore England Memorandum SLM 1802 July 1967, 2116July 1968

NON-LINEAR PREDICTION OF IONOSPHERIC PARAMETERS IMPORTANT FOR FORECASTING THE BATTLESPACE ENVIRONMENT

N M Francis[1],
A G Brown[2],
A Akram[1],
P S Cannon[1],
D S Broomhead[3].

[1]Radio Science Propagation Group,
[2]Signal and Information Processing Group,
Defence Evaluation and Research Agency,
Malvern, Worcestershire, WR14 3PS, UK.

[3]Department of Mathematics, UMIST, P.O. Box 88,
Manchester, M60 1QD, UK.

Email: *nmfrancis@dera.gov.uk*

1.0 SUMMARY

Modelling results from a programme to improve the accuracy of the prediction of ionospheric parameters are presented. Ionospheric conditions impact upon communications, navigation / GPS, satellite operations and surveillance radar operations. As a result, ionospheric predictions have an important bearing upon the specification of the battlespace environment.

The paper presents a novel and robust technique that can cope with the problems of noise and non-contiguity that are endemic to solar-terrestrial data sets.

The method of using non-linear radial basis function (RBF) neural networks (NNs) to model the noon-day variation of the critical frequency of the F2 layer of the ionosphere, foF2, is investigated. A technique based upon singular value decomposition is also adopted for the purposes of noise reduction. The performance of the model is compared with the results obtained from the reference persistence model predictions. Consecutive noon-day foF2 values from 1957 to 1990 from the Slough ionosonde station (UK) are used to train and test the model. Predictions are made for timescales of one to thirty days ahead, using both the RBF and persistence models. Relative performance is quantified using root mean square error (RMSE) between the RBF and persistence prediction time series compared with the actual time series, over the testing interval.

It is found that RBF NNs offer a significant improvement, approximately 60%, over the performance of the reference persistence model.

2.0 INTRODUCTION

2.1 Military relevance

Ionospheric disturbances can effect both ground and space based communications, navigation and surveillance systems. Examples include military communications, GPS, counter stealth HF over-the-horizon radar and ballistic missile defence. Even well designed systems are vulnerable. Predictions of the ionised battlespace environment are, therefore, of paramount operational significance. The effects of the ionosphere on a number of military systems are shown in Table 1.

Table 1 The effects of the ionosphere on C3I systems

	Communications	Surveillance	Navigation	
Systems	HF communications and broadcasting UHF/SHF satellites LEO and MEO cellular & data SATCOM VLF-MF communications and broadcasting	UHF/SHF radars HF OTHR radar Spaced based SAR Geo-location	GPS LOREN/DECCA	
Effects	Data/symbol errors Sky-wave/ground-wave interference Loss of communications	Range and bearing errors Loss of target discrimination Spectral distortion Prohibits remote sensing Loss of SAR aperture phase coherence	Range errors Sky-wave contamination	Loss of phase lock and data
Severity	Can lead to complete HF blackout Up to 30Hz Doppler shift HF multi-path up to 8ms 30 dB fades at 400 MHz 20 dB fades at L-band	Many km of range uncertainty at HF Bearing errors up to 100 degrees Over 200 m at UHF	GPS single freq position errors up to 75 metres	GPS loss of position update
Causes	Ionospheric irregularities Multi-path Attenuation Doppler	Total electron content variations Ionospheric irregularities	Total electron content variations	Ionospheric irregularities

2.2 Ionospheric prediction

A long standing goal of solar-terrestrial physics has been the prediction of specific events, or indices of activity, that impact upon communications, satellite and utility operations.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

Correlation studies that relate solar wind energy input to various indices of geomagnetic activity have played a crucial role in understanding the sun - solar wind - magnetosphere - ionosphere coupled system. In addition, linear predictive techniques have been used to deduce relationships between input and output data [*Milsom et al.,* 1987, 1989]. Although linear techniques have achieved some measure of success, the limit of these approaches appears to have been reached. A number of studies have demonstrated the importance of non-linear behaviour within the sun-solar wind-magnetosphere-ionosphere system [*Baker et al.,* 1990]. These have led to the development of non-linear theoretical models [*Klimas et al.*, 1992] and to the use of non-linear time series methods [*Vassiliadis et al.,* 1992] to characterise the behaviour of these complex systems.

It is unlikely that linear models will ever adequately describe the complex non-linear magneto-hydrodynamic (MHD) systems that govern the behaviour of the solar-terrestrial environment. These coupled MHD equations are often insoluble using analytical techniques. In the absence of a tractable solution derived from the full MHD equations, it is an attractive option to look for non-linear features in experimental data in order to develop models using empirical methods. These methods, which have evolved from work begun a decade ago to find evidence of finite dimensional dynamics in fluid turbulence, have received recent attention. However, they have been developed for well-controlled experimental systems and, therefore, work best on large sets of clean, stationary data.

In practice, solar-terrestrial data sets are far from this ideal; generally, the signal to noise ratio is poor and the data is highly quantised, with large numbers of data drop-outs from the time series. The large range of physically significant time-scales also means that the time series are rarely stationary. Thus, it follows that naive application of these non-linear techniques is unlikely to succeed. In this study, we attempt to apply novel techniques derived from studies into artificial intelligence to derive more robust non-linear predictive models.

The current state of the art for geophysical forecasting [*Joselyn,* 1995] falls far short of the predictive precision that would be required for an acceptable model of future solar-terrestrial activity. Any prospective model must be compared with the reference persistence model to assess its relative performance. The latter predicts that the value of an observable parameter at some specified point in the future will be identical to the currently observed measurement of that parameter. This model performs well for quiet time conditions, when the terrestrial environment is relatively undisturbed. However, it cannot predict the onset of the short-lived impulsive disturbances that characterise periods of unusually high geomagnetic and ionospheric activity. A successful predictive model must offer a tangible increase in performance over the persistence technique to prove the value of the method employed.

Current theoretical and empirical models fail this stringent criterion. In particular, prediction of storm events and spurious disturbances are very poor. Even the overall performance can be inferior to the standard persistence model. This shortfall in the performance of existing models can be attributed to several causes. Firstly, there is an incomplete understanding of the propagation of disturbances from their solar origins to the near-earth environment. Secondly, the coupling between the magnetosphere, ionosphere and solar is not well understood. Finally, very few continuous time series exist that are indicative of solar wind activity and associated solar activity. The combination of these factors makes the adoption of knowledge independent modelling techniques highly desirable.

2.3 foF2 time series

The critical frequency of the F2 layer of the ionosphere (see Figure 1), foF2, is one of the most significant parameters of the ionised upper atmosphere for the purposes of military and civil applications. As this area often contains the highest electron concentration within the ionosphere, it generally determines the maximum usable frequency for HF communications. It also represents the usable portion of the ionosphere that is farthest from the ground, a fact that has direct bearing on the maximum usable range of a communication link. For a fixed elevation angle, transmission range increases as the maximum height of the point of reflection increases.

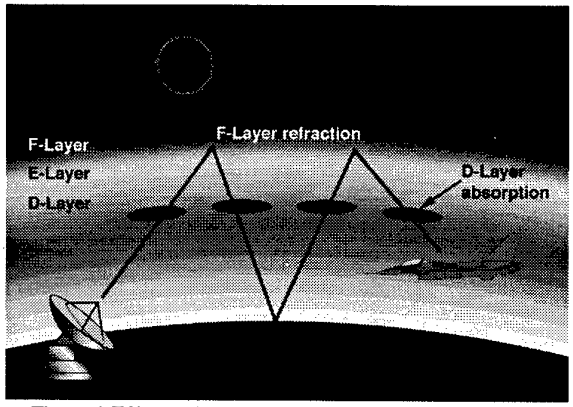


Figure 1 Effects of the ionosphere upon communications.

The foF2 time series is determined from ionograms (Figure 2), produced by ionosondes. These ionosondes sweep across the HF frequency spectrum and produce profiles of frequency as a function of virtual height, for the HF frequency band (3 to 30 MHz). The value of foF2 is the maximum frequency of the ordinary wave that is returned from the F2 layer of the ionosphere, at vertical incidence. A global network of such sounding stations record measurements of foF2 every fifteen minutes and compile these measurements to provide hourly and daily averages, in conjunction with monthly medians. These values are available from a number of World Data Centres (WDCs) that act as repositories for solar-terrestrial data. World maps, such as the example given in Figure 3, are then constructed to provide global contour plots of foF2 for the HF community. The WDCs also provide on-line predictions of geophysical activity as an additional service.

The present study utilises the foF2 time series corresponding to the ionosonde station at Slough, England. Daily noon values of foF2 for the period 1957 to 1990 were used to train the neural network model and test its effectiveness.

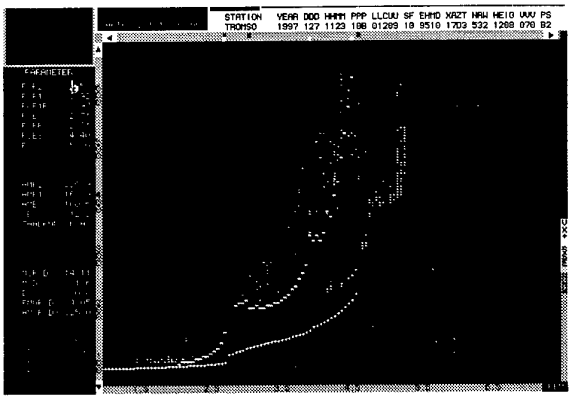


Figure 2 Vertical Ionogram.

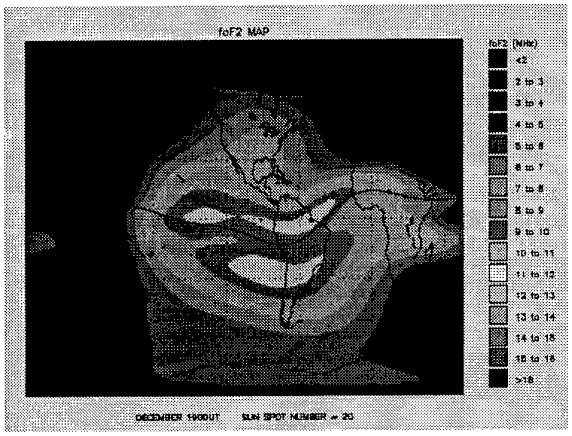


Figure 3 Global foF2 contour map.

The time series as a whole contained approximately 10% missing data points. These missing points are not randomly distributed through the time series, introducing a potential bias into the model. This poses a significant obstacle for prospective non-linear prediction schemes. Therefore, a procedure has to be defined for the processing of input training vectors that contain a missing data point. This scheme should minimise the detrimental effects upon the model of including training vectors that contain a missing point.

A simple, but not particularly effective, solution is to set all missing data points to a value of zero, subsequent to the normalisation process. This normalisation process is described in the discussion of the analysis techniques employed in this study. This minimises the average mean square interpolation error, for a constant value interpolation, over the de-meaned time series as a whole.

The next level of sophistication involves the use of adaptive linear interpolation techniques that examine the local behaviour of the time series in the region of the missing point. Such techniques use this information to construct a linearly optimal estimate of the value of the missing data point. However, use of a linear interpolation scheme is highly undesirable for time series that display significant non-linear behaviour, due to the fragility of the non-linear properties that have to be characterised.

This study has chosen to adopt non-linear modelling techniques with the aim of improving the prediction of ionospheric indices. In the absence of a suitably robust non-linear method of dealing with missing data points, it is necessary to discard those training vectors that contain missing data points.

3.0 MATHEMATICAL METHODS

3.1 Non-linear Analysis

Linear techniques have proved to be powerful tools in the characterisation and prediction of time series. However, there are situations in which these techniques are unsuitable and non-linear techniques need to be considered.

The fundamental principle behind non-linear prediction techniques is Takens theorem [*Takens,* 1981]. A corollary of which is that, given a dynamical system, possibly of high or infinite dimension that has a stable attracting set with dimension d, there generally exists a function, requiring at most $2d+1$ independent inputs, that can be used as a predictor for a time series arising from that system. Takens' result guarantees the existence of non-linear predictive functional models of the data, but provides no details about the dimension of the attractor, or how to calculate the underlying function.

3.2 Singular value decomposition

Solar-terrestrial data sets are typically noisy and singular value decomposition (SVD) can be employed to pre-process the time series to allow subsequent selection of principal components to optimise separation of the signal and noise subspaces. Removing the noise subspace improves the predictive performance of RBF NNs, as described in section 3.3.

SVD is a powerful technique associated with matrix based computations and analysis. SVD is used principally to provide an optimal solution for a linear least squares problem. In addition, SVD can also be used to devise filters for noisy signals. In this study, the matrices used in the calculations have been derived from the time series, using a moving window technique.

For an ideal error free measurement system (the data lies in a finite dimensional space), measurement data can be arranged in a matrix, where the matrix is known to be rank deficient. This means that some or all of the rows or columns of a matrix are linearly dependent upon each other. This linear dependency breaks down when any form of noise is present, as in almost all actual measurement systems. The resultant measurement matrix often becomes full rank in such cases. SVD can be used in these instances to determine a set of orthogonal bases that minimise the effects of this loss of linear dependency between the rows or columns of the matrix, due to noise. In essence, SVD provides the optimal linear separation of the signal and noise subspaces and can be used to minimise redundancy in the input vector.

To achieve this optimal estimate, a sliding window, of length n, is passed along the data set, one point at a time, to construct the matrix of delay vectors. The SVD of this matrix can be used to create a set of n orthogonal filters whose summed output would produce the original time series. Each filter corresponds to a principal component (PC). The magnitudes of these PCs can be plotted in order of decreasing size to assess

the relative linear importance of the output of each of these filters in characterising the variation of the time series. The efficiency of the predictive scheme can then be optimised by only including those PCs that have a positive effect upon the model output error. Those PCs that adversely affect the model accuracy are removed.

In summary, SVD provides a useful technique for filtering a noisy signal and provides an estimate of the number of inputs required to characterise the underlying attractor of a dynamical system. It is also the method of choice for solving linear least squares problems, a property we will exploit in the next section.

3.3 Radial basis function neural networks (RBF NNs)

Neural networks address the problem of fitting points (X_i,Y_i), where i=1..N, to a model of the form $Y_n=f(X_n)$. Clearly, for a finite data set, there will be a whole class of functions, F, such that any $f \in F$ satisfies these relationships. Matters are further complicated by presence of noise. By fitting the data exactly, we inadvertently fit this noise as well. As a result, the model is likely to perform sub-optimally on unseen data, as the unseen noise will not be predictable from that contained in the original data set. Any process that minimises the contribution of noise to the model will necessarily improve the predictive accuracy.

To address both of these problems simultaneously, it is usual to select the f that minimises the modelling error in a restricted class of functions. Increasing the size of this class will result in a decrease in the error calculated over the original training data. However, on previously unseen test data drawn from the same distribution the error will eventually begin to grow, as the noise in the original data is modelled more and more accurately. At this point the approximation is deemed optimal, since it models the general trends within the data without modelling the noise.

There are numerous linear and non-linear methods for determining such functional relationships between time series. The RBF neural network offers one approach to the solution of this problem. Given two sets of data points $\{X_n\}_{n=1}^N$ and $\{Y_n\}_{n=1}^N$, it is desirable to construct a function $f(X_n)$ that minimises the summation:-

$$E = \sum_{n=1}^{N} (f(X_n) - Y_n)^2 \qquad (1).$$

The RBF approach reduces the construction of $f(X_n)$ to that of a linear least squares problem. This can be solved using SVD, by assuming $f(X_n)$ to be the linear weighted sum of radially symmetric functions of X, which contain the non-linear adaptability of the network, that is,

$$f(X_n) = \sum_{i=1}^{p} \omega_i \varphi(|X_n - c_i|) \qquad (2).$$

A number of individual basis functions, φ_i where i=1,..,p are chosen with centres c_i and weights ω_i to construct the function f. Given this formulation of the problem, equation (1) reduces to a linear least squares problem that can be solved by applying SVD.

In the results presented, φ is a cubic function of the form,

$$\varphi(r) = r^3 \qquad (3).$$

Applying equation (3) reduces equation (1) to a linear least squares problem of the form **MW=Y**, where $\mathbf{W}=[w_1,.....,w_p]$, $\mathbf{Y}=[Y_1,.....,Y_n]$ and,

$$\mathbf{M} = \begin{bmatrix} \varphi|X_1 - c_1| & \cdots & \varphi|X_1 - c_p| \\ \vdots & \ddots & \vdots \\ \varphi|X_1 - c_1| & \cdots & \varphi|X_1 - c_p| \end{bmatrix} \qquad (11).$$

We require the weights w_i, so solving for **W** yields **W=NY**, where $\mathbf{N}= (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ (if $p \le n$) or $\mathbf{N}= \mathbf{M}^T(\mathbf{M}\,\mathbf{M}^T)^{-1}$ (if $p > n$), the Moore-Penrose inverse of a non-square matrix, and a '-1' indicates the matrix inversion operator.

This method has several advantages over the more commonly used Multi-Layer Perceptron (MLP) techniques. MLPs train through the use of gradient descent algorithms, which determine optima on a non-linear surface given a set of initial conditions. There is no guarantee that a particular optimum is global and various initial conditions must be tested to find the best optimum solution. Gradient descent also relies on an iterative process, the back-propagation of training errors through the network to adjust the network parameters to provide a better fit to the training data. Cross validation against a control data set determines when the network generalisation capabilities are optimum, for a certain set of initial conditions. Iterative processes are typically numerically intensive and arbitrary to a certain degree.

RBF NNs are assured of finding the global optimum solution in a single pass, because the training process is a linear least squares problem. Thus, RBF NNs are relatively cheap in computational terms and more straightforward to use than MLPs. The disadvantage of the RBF approach is that, because its form restricts the search to a linear hyper-surface in the function space, it can require more degrees of freedom (centres) to perform as well as the MLP. This can sometimes be an issue when the amount of data is limited. In general, however, the ease of fitting the RBF more than compensates for any necessary increase in the complexity of the class of models. In addition, RBF NNs provide a straightforward functional relationship between input and output vectors, which facilitates understanding of the model constructed by the network. Rule extraction for MLPs is a much more complicated matter and the information contained in the knowledge matrix is not amenable to intuitive decomposition.

4.0 ANALYSIS TECHNIQUE

The RBF model was trained and tested using noon day foF2 values from the Slough ionosonde covering the period from 01 January 1957 to 31 December 1990. The total length of the time series is 12,418 points. The first 10,000 were used to train

the RBF model while the remainder of the points were used to test the predictive performance of the model on unseen data.

The time series as a whole was reduced to zero mean and unit variance. Normalisation of the input data in this fashion removes the need for an amplitude or DC component parameter input for the radial basis functions that we have adopted for the purposes of this study.

A sliding window of length 90 days was passed along the time series, one point at a time, to construct the matrix of input vectors that are required to train the RBF model. This value was selected by assessing the relative performance of the one day ahead prediction model using a range of input window lengths. Input vectors that contain one or more missing data points are then removed from the training set. For this study, 10% missing data points gave rise to approximately 90% rejection of possible training vectors. The training vector set was then projected onto its principal axes using SVD. Principal components that adversely affected the prediction were removed.

The remaining filter outputs were then used to train the RBF model. The optimum number of centres (functions) was determined by assessing the relative performance of the one day ahead prediction model using a range of numbers of centres. A value of 140 centres was found to give the best results. Models were then constructed, using these parameters, to predict from 1 to 30 days ahead. The results were compared with the corresponding persistence model prediction, using root mean square error (RMSE).

5.0 DISCUSSION OF RESULTS

The results displayed in Figure 4 represent the relative performance of the RBF NN model in terms of RMSE, in comparison with the reference persistence model, for prediction timescales of one to thirty days ahead. For a one day ahead prediction, the RBF model (no noise filtering) produces an RMS error of 0.655 MHz, in comparison with an RMS error of 1.186 MHz for the standard persistence model. This represents a relative increase in performance of approximately 45% over the persistence model. Optimising the noise filtering process led to the removal of 36 of the PCs that adversely affected the accuracy of the model. For a one day ahead prediction, the RBF model (with SVD noise filtering) produces an RMS error of 0.484 MHz, a relative increase in performance of 59% over the standard persistence model.

The graph shows that the RBF model represents a significant improvement over the standard persistence model. This is particularly true with regard to predictions up to ten days ahead. In this region, the predictive accuracy stays roughly constant. The performance increase margin becomes narrower as the prediction time-scale increases to thirty days and the model error rises, but noise reduction mitigates this problem. The first effect can most likely be attributed to the fact that optimised model parameters were only derived for the one day ahead prediction. This was necessary due to limitations on available processing resources. These optimal values will become increasingly inaccurate as the prediction extends further into the future. Better results would be obtained if the model was optimised separately for each of the number of steps ahead required for the prediction.

Noise reduction provides further worthwhile improvements, particularly in regard to the long range and one step ahead predictions. The noise reduction process was also optimised with respect to the one day ahead prediction, hence the radical improvement over the unfiltered model for the one day ahead model. Much smaller improvements are observed for short time-scale predictions (2-10 days). Apparently, the noise reduction scheme produced for the one day ahead prediction seems to be too specific to have a pronounced effect on the other short term models. This indicates that optimising each model separately, with regard to noise reduction, would also provide larger improvements in terms of predictive accuracy. However, the short term prediction accuracy plateau now extends to 15 days, five days longer than for the corresponding unfiltered RBF model. Noise reduction provides tangible improvements for medium time-scale predictions (10-25 days), giving a fairly constant improvement in accuracy across the whole period. The effects of the noise reduction scheme upon the longer term predictions (25-30 days) are more pronounced. This arises from the fact that prediction error, due to the effects of noise, increases with the time-scale of the prediction. Hence, any noise reduction process will have a more pronounced effect upon longer term predictive models.

Figure 5 represents the relative difference between the RBF model and the persistence model for each of the types of prediction. This allows assessment of the similarities between the model produced by the RBF and the reference persistence model, for varying prediction timescales. The effects of noise reduction on these features can also be determined.

The divergence of the unfiltered RBF model from the persistence model increases markedly after twelve days. This point coincides with the narrowing of the performance increase margin noted in regard to Figure 4. This feature can also be observed for the filtered RBF model. However, in the latter instance, the relative difference undergoes a dramatic decrease after 25 days. This effect is coincident with the improvement in predictive accuracy over the unfiltered RBF model noted in Figure 4. In this graph, the filtered RBF model is seen to be superior to the unfiltered RBF model for prediction timescales greater than 25 days. These two observations lead to the deduction that the noise reduction process improves the ability of the RBF to successfully model structure that can be attributed to the recurrence period of the solar rotation (25-30 days).

Comparison between the testing data set and the prediction output time series reveals those areas where the RBF prediction is least affective. The predicted onset of periods of unusually high geomagnetic activity lags behind the actual time series. In addition, the model underestimates the peak magnitude of these events by an approximate factor of two. This gives rise to large error deviations in these regions that can account for much of the overall RMS error. However, the performance of the RBF models is still superior in this respect with regard to the reference persistence model.

Three factors could account for the relatively poor behaviour of the model during periods of elevated geophysical activity. Firstly, periods of unusually high ionospheric activity are rare in comparison with quiet time conditions. As a result, the training process is biased away from the prediction of these infrequently occurring events, in favour of accurate prediction

of quiet time conditions. We plan to rectify this problem by increasing the relative importance of non-quiet time conditions. Secondly, the timescales associated with the evolution of such disturbances are very much less than the resolution of the input time series. Therefore, the input time series does not contain sufficient information concerning the dynamics of these events to accurately model their rapid development. Use of additional higher resolution inputs may be required to address this issue. Finally, the current RBF model relies solely upon a single input time series to produce a prediction. Certain data sets are capable of providing precursive information that has a direct bearing upon the prediction of foF2. Including these time series as additional inputs to the model would provide a measure of forewarning that could improve the prediction accuracy for periods of high ionospheric activity.

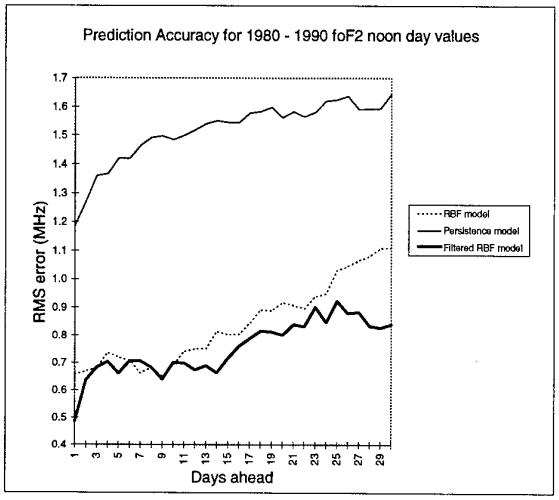


Figure 4 Prediction error versus no. of days ahead for prediction of foF2 noon day values (RMSE).

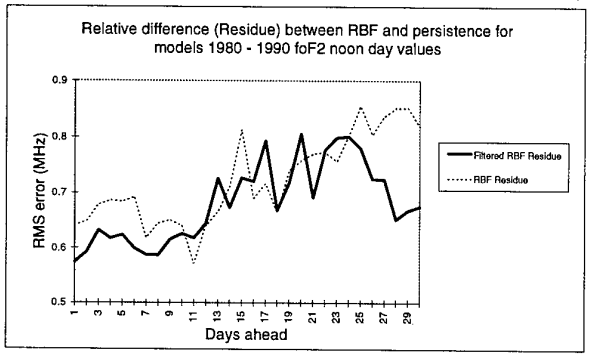


Figure 5 Relative difference between RBF and persistence models for prediction of foF2 noon day values (RMSE).

6.0 CONCLUSIONS

This study has presented a novel robust technique that can cope with the problems of noise and non-contiguity that are typical to solar-terrestrial time series.

The method of using non-linear radial basis function (RBF) neural networks (NNs) to model the noon-day variation of the critical frequency of the F2 layer of the ionosphere, foF2, was implemented. The benefits of SVD noise reduction were assessed and were found to offer a substantial improvement.

The performance of the model was compared with the reference persistence predictions. The unfiltered RBF model produces an RMS error of 0.655 MHz, which is approximately 45% more accurate than the corresponding persistence prediction, with an RMS error of 1.186 MHz. For a one day ahead prediction, the RBF model (with noise filtering) produces an RMS error of 0.484 MHz, a relative increase in performance of 59% over the standard persistence model. This accuracy is maintained for a ten to fifteen day period before the error starts to rise significantly. Optimising the models for each individual prediction time-scale will bring further improvements to the accuracy of the predictions.

In conclusion, the RBF NN model presented in this study is a significant improvement over the standard persistence model, which acts as the reference for the assessment of any proposed prediction scheme.

7.0 REFERENCES

D. N. Baker, A. J. Klimas, R. L. McPherron, J. Buchner, 'The Evolution from Weak to Strong Geomagnetic Activity: an Interpretation in terms of Deterministic Chaos', Geo. Res. Letters, Vol. 17, No. 1, p. 41-44, Jan 1990.

J. A. Joselyn, 'Geomagnetic Activity Forecasting: The State of the Art', Reviews of Geophysics, 33, 3, August 1995, P383-401, Paper number 95RG01304.

A. J. Klimas, D. N. Baker, D. A. Roberts, D. H. Fairfield, 'A Non-linear Dynamical Analogue Model of Geomagnetic Activity', J. Geophys. Res., Vol. 97, No. A8, P12,253-12,266, Aug. 1992.

J. D. Milsom, N. G. Riley, C. R., 'Recommendations for the Advancement of the MRC short-term Ionospheric Forecasting Service', Marconi Research Centre, MTR 87/94.

J. D. Milsom, S. J. Goldsmith, D. W. Green, C.R. Telfer, N. S. Wheadon, 'Research and Development of a short-term Ionospheric Forecasting Service', Marconi Research Centre, MTR 89/21.

F. Takens, 'Detecting Strange Attractors in Turbulence', Lecture Notes in Mathematics 898, Springer, Berlin (1981), P366-381.

Vassiliadis, A. S. Sharma, K. Papadopoulos, 'Time Series Analysis of Magnetospheric Activity using Non-linear Dynamical Methods', Chaotic dynamics: Theory and Practice, Plenum Press, New York, 1992.

PAPER No: Paper 19

DISCUSSOR'S NAME: P. Hoogeboom

COMMENT/QUESTION:

The prediction result for one day ahead is much better than for more days ahead. Also, the results seem to indicate a cycle at 4, 9, 14 days ahead of slightly improved results. Is there any explanation for these observations?

AUTHOR/PRESENTER'S REPLY:

Due to limitations on available processing resources, it was only possible to optimise the model parameters for the one day ahead prediction. Hence, these results show a much greater relative improvement over the persistence model, compared with the longer timescale predictions. The observed, 4, 9, 14 day cycles may well be an artifact of this inability to optimise each individual model. So, we are not prepared to comment upon these features, in terms of what physical processes they might represent, until the cause of uncertainty can be eliminated.

Fusion de données et fiabilité des transmissions HF

C. Goutelard, J. Caratori, C. Pautot
LETTI-UNIVERSITE PARIS-SUD
Bât. 214 – 91405 Orsay cedex
France

1. SUMMARY

High frequency reliable radiocommunications in decametric band is currently based on procedure such as ALE (Automatic Link Establishment), RTCE (Real Time Channel Evaluation) ... which purpose is to test channel transmission capacity and interferences level.

Establishment tests, made in real time, give interesting results by using long time experienced procedures with very good instantaneaous reliabilities. Though, procedure can be improved with short time prevision of transmission channel characteristics and with interferences forecasting which are numerous in decametric band.

Ionospheric forecasting methods allow to take into account ionospheric transmission characteristic. It is supposed that a very short time forecasting method (1 to 2 hours) or that a short time method (24 to 48 hours) is implemented. Interference forecasting is of prime importance in prevision of good establishment of links.

Electromagnetic spectrum observation in HF band shows that :

- Correlation from a day to another of interferences levels are often strong (figure 1).
- Spectral correlations are good until distance of about 400/500 km (figure 2).
- Interferences levels are dependent of observation azimuth (figure 3).
- Spectral occupancy is correlated with critical frequency FoF2.

Taking into account these characteristics lead to consider the establishment of interferences level forecasting and the critical frequency of F2 layer.

Jamming level characterization is introduced by considering the notion of clear frequency for a specified level. Frequency is supposed to be clear for a threshold S if the interferences level remains lower to this threshold. Clearness probability can be determined for each traffic band and we can plot surface of probability that a clear frequency one day remains clear the following days. These surfaces plotted as a function of the threshold and the forecast horizon are shown figure 5.

Correlation between spectral occupancy and FoF2 frequency is also modelised. Comparison between FoF2 frequency determined by zenithal sounding with measurement of occupancy – m_1F -has be proved very satisfying (figure 6).

A data fusion processing lead to a totaly passive method of real time network management.

2. GENERALITES

L'établissement de radio-communications fiables dans la gamme décamétrique est actuellement basé sur des procédures de type ALE (Automatic Link Establisment), RTCE (Real Time Channel Evaluation) qui visent à tester les capacités de transmission du canal et les niveaux d'interférences. La communication étant établie, on utilise une procédure AME (Automatic Maintain Establisment) qui permet de dégager la liaison de la fréquence dont la clarté se détériore lors la transmission de données. Les tests d'établissement sont effectués en temps réel et donnent des résultats intéressants en utilisant des procédures dont l'ancienneté leur confère d'excellentes fiabilités instantanées. Ces procédures présentent cependant l'inconvénient de nécessiter des mesures sur le canal et, par conséquent, l'utilisation d'émission d'ondes électromagnétiques. Par ailleurs, la capacité de transmission du canal n'est pas le seul paramètre à prendre en compte et l'occupation spectrale, particulièrement importante dans ce domaine de fréquences, impose d'en tenir compte.

On peut observer que le spectre des ondes décamétriques s'étendant de 3 à 30MHz est particulièrement encombré. On dénombre plusieurs centaines de milliers d'émissions radioélectriques simultanées sur le globe. Ces émissions constituent une banque de données particulièrement riche que l'on peut exploiter de deux façons différentes :

- Tout d'abord, on doit remarquer que ces émetteurs constituent des signaux parasites qui gênent les communications.
- On doit donc, pour l'établissement des communications, rechercher parmi les fréquences attribuées à un utilisateur, celles qui seront susceptibles d'être peu brouillées. Cette action consiste à rechercher les fréquences claires – c'est-à-dire celles pour lesquelles le niveau des interférences est inférieur à un seuil donné – mais il convient de connaître quelles sont les périodes de temps pendant lesquelles cette fréquence demeurera claire. Cette approche ne peut qu'être probabiliste et il est donc usuel de déterminer les probabilités de clarté de chacune des fréquences se trouvant dans le spectre.
- L'ensemble de ces fréquences peut être utilisé comme une banque de données pour déterminer les caractéristiques générales du milieu de propagation. Cette approche peut être justifiée par le fait que les utilisateurs s'adaptent aux conditions ionosphériques avec de plus en plus de précision et, par conséquent, suivent l'évolution du milieu. On a donc songé à utiliser les caractéristiques d'encombrement du spectre décamétrique pour en déduire un paramètre capital dans les télécommunications ionosphériques, à savoir la fréquence critique FoF2 de la région F2 de l'ionosphère.

Cette approche permet donc de réaliser, d'une part une prévision de l'état de l'ionosphère, d'autre part une prévision des interférences par une méthode totalement passive. Les résultats obtenus dans cette étude montrent que les interférences sont corrélées dans le temps et dans l'espace et que le spectre de la gamme décamétrique est corrélé avec l'état de l'ionosphère. Ces résultats permettent de réaliser une application opérationnelle.

3. CARACTERISATION DES INTERFERENCES

Une étude sur les interférences a été entreprise au LETTI dès le début des années 1980 afin de tenter de les caractériser. Des mesures permanentes ont été faites et ont permis de constituer

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

une banque de données s'étalant sur plus d'un cycle solaire et comprenant des centaines de millions de mesures. A partir de cette banque de données, une statistique a été faite pour déterminer les principales caractéristiques des interférences.

On s'est, dans un premier temps, orienter sur l'étude des interférences en Europe occidentale. On a, pour cela, utilisé plusieurs stations de mesures compatibles pour permettre des comparaisons entre plusieurs sites. Le premier objectif poursuivi a été d'examiner si ces corrélations existaient dans les paramètres caractérisant des interférences. On a pour cela utilisé plusieurs types de mesures.

3.1. Corrélation temporelle des interférences [1]

On a examiné si les interférences présentaient des cohérences dans le temps et des cohérences dans l'espace. La figure 1 représente quatre spectres obtenus en deux sites différents distants de 450km, l'un situé dans la région sud de Paris, l'autre situé dans le sud-ouest de la France (Les Eyzies de Tayac). Chaque spectre représente un ensemble de 15 spectres superposés mesurés sur une durée de 1 heure dans la gamme 1 à 30MHz. Deux spectres ont été relevés à 23h00, deux autres à 14h00 T.U. On peut constater que les 15 spectres superposés conservent une bonne cohérence ce qui signifie que les caractéristiques d'ensemble du spectre décamétrique sont relativement bien cohérentes. L'examen des spectres montre également l'emplacement des bandes d'émission de radio-diffusion. On peut noter que de jour le spectre est plus étendu vers les hautes fréquences, ce qui est provoqué d'une part le milieu et d'autre part par l'adaptation des utilisateurs aux évolutions du milieu. On peut remarquer également une absorption des basses fréquences plus importante en milieu de journée que la nuit. Ces enregistrements confirment les phénomènes connus d'absorption et de variations de l'ionisation du canal ionosphérique mais montrent également une bonne corrélation temporelle sur l'échelle d'une heure de la globalité des spectres ainsi qu'une grande cohérence spatiale qui est visible par la comparaison des spectres obtenus aux stations STUDIO1 (Paris) et STUDIO2 (sud-ouest de la France). L'analyse de la cohérence spatiale a été faite sur des distances variant de 0 à 1500km et l'on a pu constater que, jusqu'à des distances de l'ordre de 500km, on peut considérer que les spectres sont bien corrélés, comme le montre le résultat de la figure 1. Une analyse plus fine de la cohérence spatio-temporelle peut être faite en observant des bandes de fréquences étroites. Ce type d'observation a été fait sur de nombreuses émissions.

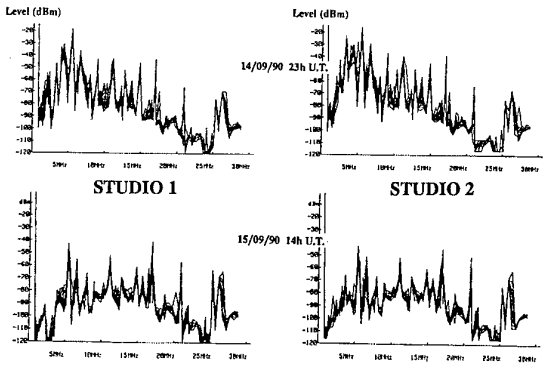


Figure 1 : *Corrélation spatiale ente spectres HF mseurée en des points distants de 450 km*

La figure 2 montre les évolutions sur deux jours et demi de signaux occupant des bandes de fréquences de largeur de 1kHz à quatre fréquences différentes. On peut noter l'existence très nette d'une cohérence temporelle avec un cycle de 24 heures ainsi que la bonne corrélation entre la station STUDIO1 et STUDIO2 distantes de 450km révélée par ces enregistrements.

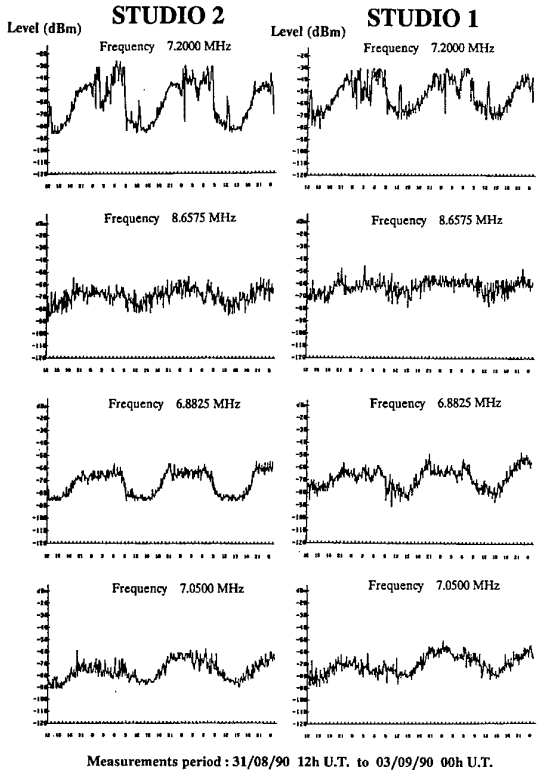
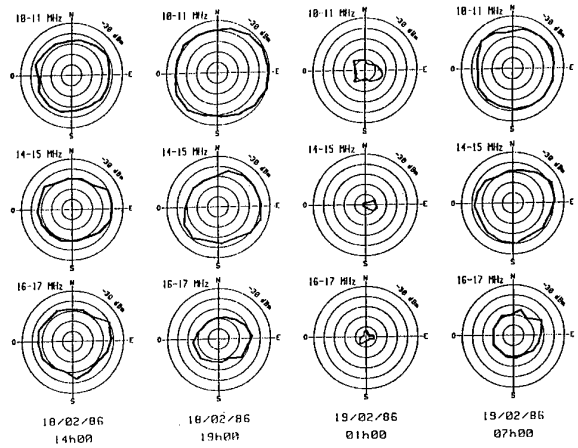


Figure 2 : *Corrélation temporelle et spatiale de spectres HF entre STUDIO1 et STUDIO2*

3.2. Variations azimutales [1]

Une étude de l'intensité des brouilleurs a été effectuée en fonction de leur angle d'arrivée. On a utilisé pour cela un réseau d'antennes composé de 32 antennes large bande couvrant la gamme décamétrique et dont le lobe est orientable dans toutes les directions. La résolution typique de ce réseau est de 15 degrés à 15MHz. On a fait une étude statistique de l'intensité des champs en fonction de l'angle azimutal d'arrivée des rayons pour toute la gamme de fréquences en établissant la statistique sur des bandes de 1MHz et en ramenant l'intensité du champ à une bande de 3kHz. Les résultats obtenus sont illustrés par les diagrammes de la figure 3 qui montrent les valeurs obtenues dans 3 bandes de fréquences différentes : 10-11MHz, 14-15MHz, 16-17MHz, pour quatre heures de la journée : 14h00, 19h00, 01h00 et 07h00. On peut, en premier lieu, noter l'effet journalier des variations du milieu de propagation particulièrement visible par le fait qu'à 01h00 T.U. le niveau des champs reçus est quasiment nul pour les fréquences comprises entre 14 et 17MHz et extrêmement faible pour les fréquences comprises entre 10 et 11MHz. On peut constater également l'effet de lever et coucher de soleil, particulièrement visible sur la bande 16-17MHz. Ces relevés révèlent clairement que l'intensité des interférences en provenance de la direction Est est notablement plus importante que celle qui provient de la région Ouest ou Nord-Ouest. Cette différence s'explique sans difficulté en observant que beaucoup plus d'émetteurs se trouvent évidemment localisés à l'Est de la région parisienne qui correspond à une zone terrestre, qu'à l'Ouest de la région parisienne qui correspond à une zone maritime.



Centre des cercles : -80dBm/3kHz, Différence entre 2 cercles : 10dBm

Figure 3 : *Distribution azimutale de puissance reçue dans une bande de 1 MHz et convertie dans une bande 3 kHz*

3.3. Analyse des résultats

Les résultats obtenus au cours de cette étude ont montré que la corrélation temporelle des interférences était bonne ce qui traduit le fait que les utilisateurs des transmissions haute fréquence par voie ionosphérique règlent leurs échanges de façon très ordonnée en s'attribuant en particulier des heures de vacations. Il va de soi que ce phénomène est particulièrement marqué dans les zones de radiodiffusion mais ce phénomène est également observé de façon extrêmement constante pour les liaisons à caractère professionnel.

La cohérence spatiale constatée laisse penser que l'échantillonnage spatial des mesures d'interférences peut être effectué avec un pas relativement large de l'ordre de 500km pour en avoir une cartographie complète sur une zone donnée.

La répartition azimutale des angles d'arrivée des interférences apporte la preuve que l'utilisation d'aériens directifs, voire adaptatifs, est particulièrement utile dans le domaine décamétrique.

Les spectres présentés concernent l'Europe occidentale. Il semble que pour cette région les résultats soient relativement cohérents. Il est certain que l'encombrement spectral varie selon le point d'observation que l'on se fixe sur le globe. On a effectué des mesures dans l'océan pacifique et on a constaté que l'encombrement spectral était beaucoup plus réduit. Les résultats de l'étude présentée dans cette communication sont donc à considérer comme valables sur la zone Europe occidentale et qu'il conviendrait, pour une autre zone géographique, de s'appuyer sur une banque de mesures différente.

4. MODELISATION DES INTERFERENCES

4.1. Modélisation des interférences [2]

Pour caractériser les évolutions temporelles de l'occupation spectrale, nous avons évalué pour une fréquence d'observation donnée la probabilité pour qu'à une heure donnée le niveau de l'interférence présente ne dépasse pas un seuil S. On a effectué ces observations sur des périodes minimales de 15 jours et sur des périodes maximales de deux mois et demi. Afin d'effectuer une étude quantitative on a effectué ces calculs pour les 95 bandes de fréquences allouées pour couvrir la bande décamétrique (bande de radio-diffusion, liaisons fixes, liaisons mobiles, liaisons maritimes ...). Pour chacune des bandes observées on a constitué des tableaux $A(N_j, N_t, N_f)$ composés d'éléments $a(j, t, f)$ donnant le niveau des interférences et où :

- N_j nombre de jours d'observation ($N_j = 15$)
- N_t nombre d'échantillons dans une heure ($N_t = 30$)
- N_f nombre de points de fréquence dans la bande analysée.

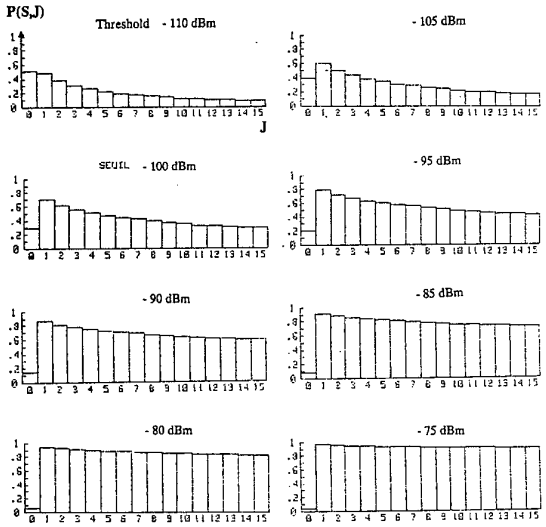
Pour un seuil fixé S on forme un second tableau $B(N_j, N_t, N_f)$ dont les éléments $b(j, t, f)$ prennent les valeurs :

$$b(j, t, f) = 0 \quad \text{si} \quad a(j, t, f) \geq S$$
$$b(j, t, f) = 1 \quad \text{si} \quad a(j, t, f) < S$$

La probabilité pour que la fréquence soit claire pendant J jours consécutifs pour le seuil S est alors donnée par :

$$P(S,J) = \frac{\sum_{j=1}^{N_j-J+1} \sum_{t=1}^{N_t} \sum_{f=1}^{N_f} \prod_{i=j}^{j+J-1} b(i,t,f)}{(N_j - J + 1) N_t N_f}$$

La figure 4 montre une évolution typique de la probabilité de clarté $P(S, J)$ en fonction de J et pour différentes valeurs de S.



Gamme de fréquence 11.975 à 12.330 MHz 12h00 U.T, Date des mesures : 24/06/89 au 09/07/89

Figure 4 : *Probabilité de clarté en fonction du nombre de jours J*

On peut noter que la valeur correspondant à $J = 0$ représente la probabilité de non clarté ce qui correspond au complément de la probabilité de clarté sur un jour. On observe d'autre part que pour les valeurs de J supérieures à 1 la décroissance de la courbe de probabilité est régulière, monotone et d'allure exponentielle. Cette constatation nous a conduit à approcher cette courbe par une loi de la forme :

$$P(S, J) = \exp(-A J - B)$$

où A et B sont des fonctions des seuils S. Une étude statistique a montré que cette approximation est tout à fait acceptable et que les coefficients A et B pouvaient être exprimés simplement en fonction du seuil S. On aboutit ainsi à l'expression de la probabilité de clarté donnée par la relation suivante :

$$P(S,J) = \exp\left[-\exp(-A_1 S - A_2) \cdot J - \exp(-B_1 S - B_2)\right]$$

Les coefficients A_1, A_2, B_1, B_2 permettent donc de décrire totalement la probabilité $P(S, J)$ et une étude de la valeur de ces coefficients a été entreprise. Ces coefficients varient en fonction de la bande analysée et en fonction du temps. La probabilité de clarté de $P(S, J)$ peut être représentée par une surface en fonction de J et S. La figure 5 montre une représentation de ces surfaces pour la bande de fréquence 12330kHz à 12401kHz pour la période du 25 février 1993 au 8 mars 1993.

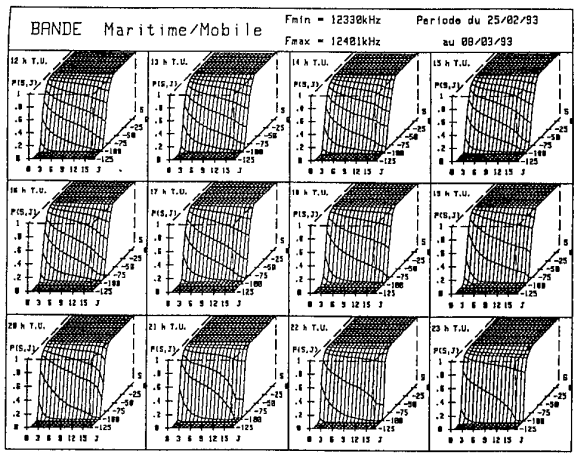


Figure 5 : *probabilité de clarté en fonction du temps et du seuil*

On peut noter sur ces figures que la surface comprend trois zones : celle correspondant au niveau S élevé pour laquelle la probabilité de clarté est voisine de 1, celle pour laquelle les seuils sont faibles et la probabilité de clarté est voisine de zéro et la zone intermédiaire où l'on peut voir que les variations de la probabilité de clarté ont bien des allures exponentielles. On peut noter qu'au delà d'une douzaines de jours les courbes semblent s'écarter d'allure exponentielle. Ce phénomène s'explique aisément en considérant que la population sur laquelle est faite la statistique décroît en fonction de la profondeur de prévision et qu'au delà de 12 à 15 jours, le nombre de fréquences claires devient trop faible pour que la statistique soit valable. C'est la raison pour laquelle il a été jugé normal de limiter les profondeurs de prévision à des valeurs de 7 jours qui correspondent à un compromis acceptable. Pour tester la validité de cette méthode, on a recherché sur les courbes réelles les valeurs des coefficients A_1, A_2, B_1, B_2 établies sur des valeurs de 7 jours et pour toutes les heures de la journée. A partir du calcul d'une surface de régression on a calculé les coefficients de corrélation R_a et R_b entre les valeurs de A et B obtenues par une mesure directe de chacune des courbes de probabilité de clarté et avec les coefficients déduits de l'ensemble de la surface.

Le tableau 1 montre que les coefficients de corrélation qui sont obtenus sont supérieurs à 0,96 et souvent voisins de 0,98 ou 0,99.

T.U	A1	A2	B1	B2	Ra	Rb
0	0,106	13,640	0,092	11,580	-0,964	-0,994
1	0,106	13,710	0,087	11,080	-0,959	-0,995
2	0,104	13,420	0,091	11,470	-0,966	-0,992
3	0,103	13,340	0,090	11,320	-0,967	-0,989
4	0,104	13,430	0,085	10,830	-0,956	-0,995
5	0,112	14,230	0,095	11,860	-0,976	-0,997
6	0,108	13,510	0,113	13,440	-0,989	-0,996
7	0,099	12,320	0,099	11,790	-0,991	-0,998
8	0,097	12,220	0,101	12,060	-0,988	-0,993
9	0,087	11,070	0,095	11,410	-0,992	-0,995
10	0,094	11,780	0,099	11,620	-0,978	-0,987
11	0,087	10,960	0,089	10,490	-0,989	-0,996
12	0,094	11,770	0,095	11,230	-0,979	-0,987
13	0,108	13,210	0,107	12,330	-0,957	-0,963
14	0,096	11,790	0,099	11,430	-0,980	-0,989
15	0,085	11,100	0,097	11,640	-0,988	-0,984
16	0,086	11,160	0,097	11,560	-0,990	-0,984
17	0,090	11,630	0,011	12,990	-0,988	-0,996
18	0,143	17,530	0,132	15,800	-0,994	-0,997
19	0,106	14,010	0,109	13,860	-0,969	-0,982
20	0,117	14,690	0,109	13,250	-0,980	-0,996
21	0,121	15,030	0,105	12,750	-0,989	-0,999
22	0,111	14,050	0,089	11,190	-0,977	-0,996
23	0,103	13,300	0,083	10,550	-0,956	-0,999

Tableau 1 : *Probabilité de clarté*

Ce résultat quantitatif prouve la validité de la méthode par l'excellente corrélation obtenue entre la modélisation et les mesures.

3.2. Application à la recherche de fréquences claires

La modélisation des interférences obtenues par cette méthode a conduit à la détermination d'une méthode de prévision à court terme des interférences. La méthode consiste à mesurer l'encombrement spectral pendant 6 jours consécutifs puis à utiliser ces informations pour estimer l'encombrement au jour J. Dans ce but, on a défini une probabilité de clarté $C(S, J)$ pour qu'une séquence claire 6 jours consécutifs soit également claire le $J+1^{ième}$ jour. Un calcul simple permet de montrer que $C(S, J)$ s'exprime par la relation :

$$C(S, J) = exp(- A) = exp[- exp(- A_1 S - A_2)]$$

Il apparaît ainsi que l'indice de clarté est indépendant du nombre de jours d'observation et ne dépend que du seuil S pour une fréquence donnée et une heure donnée. Cette méthode, appliquée aux cas de prévision réelle a donné des résultats tout à fait satisfaisants traduits par la figure 6 qui montre la comparaison entre les probabilités mesurées et les probabilités établies à partir des prévisions.

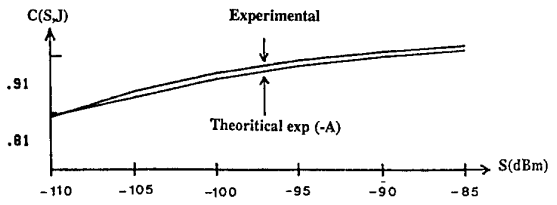


Figure 6 : *Courbes théorique et calculée de l'indice de clarté*

Les résultats obtenus montrent que la prévision des interférences à court terme peut être effectivement envisagée avec l'espoir d'obtenir des résultats satisfaisants. La méthode apparaît particulièrement simple à mettre en œuvre puisqu'elle nécessite qu'une station expérimentale légère constituée d'un récepteur piloté par un calculateur afin d'établir les statistiques et associé à une antenne de caractéristiques connues. Les prévisions peuvent être effectuées à partir d'une période d'observation de 6 jours qui permet de faire une projection sur le jour suivant. Cette prévision à 24 heures peut être affinée par une prévision plus fine qui prend en compte les fluctuations de la journée considérée par rapport aux journées précédentes. Cette méthode permet, en affinant les résultats, d'obtenir des prévisions d'une meilleure qualité.

5. CARACTERISATION DU CANAL IONOSPHERIQUE

L'allure générale du spectre HF varie dans le temps et, comme le montre les spectres représentés sur la figure 1, on peut constater que la corrélation spatiale est suffisamment bonne. On a donc songé à utiliser la forme de ce spectre pour réaliser une modélisation de l'ionosphère. Il est évident que la seule observation du spectre ne peut permettre de définir avec précision le profil de l'ionosphère sans émettre des hypothèses supplémentaires. On s'est attaché à examiner si l'allure générale de ce spectre était en relation avec la fréquence critique de la région F2 qui apparaît comme le paramètre principal de l'ionosphère. Pour procéder à cette étude on a volontairement séparé les spectres représentant les zones de radio-diffusion des autres spectres. En effet, les émissions dans les bandes de radio-diffusion sont de puissances importantes et ne peuvent être traitées avec les autres spectres. L'étude présentée peut être appliquée aux seuls spectres de radio-diffusion qui constituent un échantillonnage du spectre décamétrique ou à l'ensemble des autres émetteurs. Les résultats présentés concernent ceux établis sur l'ensemble des autres émetteurs. Il est possible de définir des modèles de spectres à partir de considérations théoriques. Ces considérations qui ont été faites ont finalement montré qu'elles s'accordaient assez pauvrement avec la réalité. C'est la raison pour laquelle on a bâti une méthode plus pratique qui s'avère donner de très bons résultats.

Cette méthode consiste à définir une densité de probabilité par bandes de fréquences allouées. Cette densité de probabilité est donnée par la relation :

$$P_f(fi) = \frac{P(fi)B(fi)}{\sum_i P(fi)B(fi)}$$

dans laquelle :

- $B(fi)$ représente la largeur de bande de la I^{ième} bande allouée centrée sur fi
- $P(fi)$ représente la puissance totale reçue dans la bande $B(fi)$

On définit alors le moment du premier ordre de cette fonction densité de probabilité à l'aide de la relation :

$$m_1F = \sum_I fi\, P_f(fi)$$

Ce paramètre m_1F se trouve étroitement corrélé avec la fréquence critique FoF2 de l'ionosphère.

Ce résultat est illustré (figure 7) par deux enregistrements effectués par le sondeur zénithal de la station française de Poitiers et par le relevé des spectres décamétriques effectué à la station STUDIO2 du LETTI située à 150 km de Poitiers. Sur ces enregistrements on a fait apparaître les variations du paramètre m_1F déduit des spectres décamétriques. On peut noter une excellente corrélation entre ce paramètre m_1F et la fréquence critique FoF2. Cependant, il doit être remarquer également que lors des transitions jour-nuit, nuit-jour les courbes de m_1F s'écartent sensiblement de la courbe FoF2. Cet écart s'explique par le fait que les émetteurs captés proviennent pour une part importante de la direction Est, comme le montre la figure 3, et par conséquent les utilisateurs se trouvent « en avance » par rapport au point d'observation situé en France. Cette différence est systématique et a été donc traitée comme un biais. Ce biais a été estimé de façon à pouvoir le corriger systématiquement.

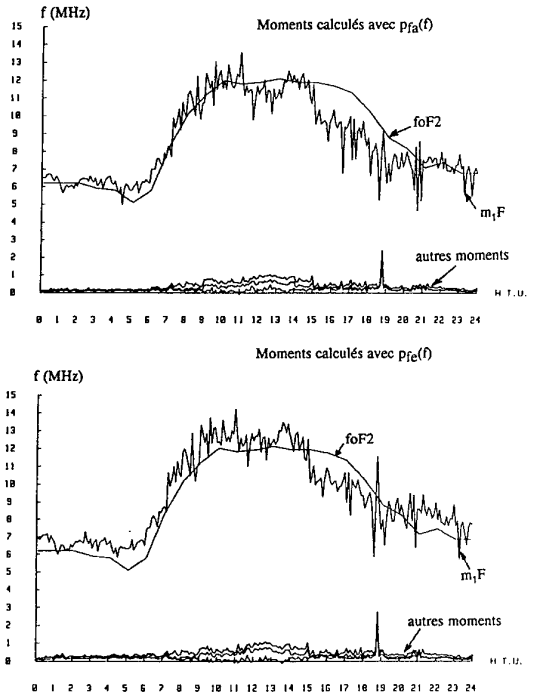


Figure 7 : *Corrélation entre les mesures FoF2 et F0F2 déduit du spectre HF*

Afin de pouvoir évaluer les performances de cette mesure de FoF2, on a introduit deux paramètres :

- L'un représentant l'écart entre la fréquence critique FoF2 vraie et le paramètre m_1F :

$$X = \text{foF2vraie} - m_1F$$

- L'autre représentant l'écart entre la fréquence critique FoF2 et la fréquence foF2 estimée par des prévisions à 90% :

$$Y = \text{FoF2vraie} - \text{foF2 à 90\%}$$

Pour l'ensemble des mesure effectuées, on a calculé l'écart moyen et l'écart quadratique moyen des deux quantités X et Y. Les résultats ont donné :

Sans correction du biais

- Ecart moyen de X : $\mu x = 0{,}313$MHz
- Ecart type moyen de X : $\sigma x = 0{,}807$MHz
- Ecart moyen de Y : $\mu y = 3{,}654$MHz
- Ecart type moyen de Y : $\sigma y = 1{,}346$MHz

Avec correction du biais

- Ecart moyen de X : $\mu' x = 0{,}071$MHz
- Ecart type moyen de X : $\sigma' x = 0{,}418$MHz

On peut remarquer la correction du biais introduit un écart moyen de X quasi nul et un écart type moyen de 0,418MHz qu'il faut comparer à l'écart moyen de Y et à l'écart type moyen

de Y. On peut noter que les résultats obtenus sont tout à fait bons pour ce genre de mesures. Un écart type de 0,4MHz peut être considéré comme la précision moyenne des sondeurs zénithaux. La méthode de prévision de la fréquence critique FoF2 apparaît donc tout à fait correcte.

6. FUSION DE DONNEES ET GESTION

L'analyse du spectre décamétrique conduit à la détermination de deux paramètres importants de l'ionosphère concernant la gestion des prévisions des interférences et la prévision du paramètre foF2. L'objectif de la méthode décrite est d'être totalement passive et donc utilisable par tous les systèmes pour lesquels une grande discrétion est requise. Le système fonctionne alors par fusion de données selon le modèle illustré sur la figure 8.

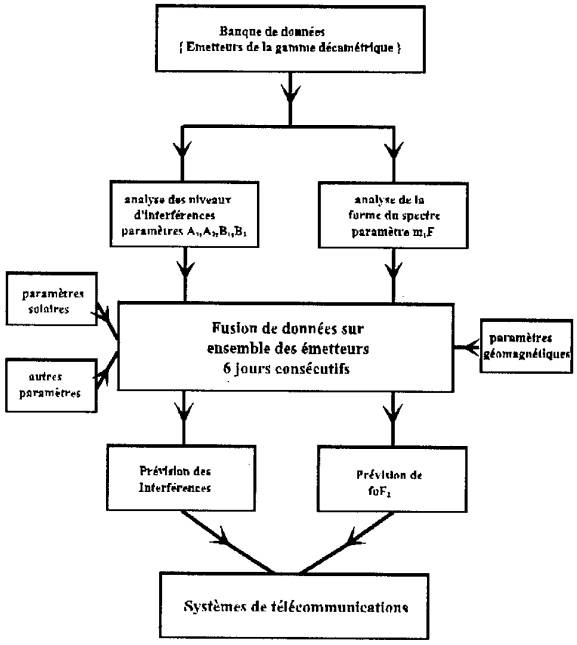


Figure 8 : *Fusion de données pour prévisions passives*

L'ensemble des émetteurs de la gamme décamétrique existant à la surface terrestre constitue la banque de données analysée par un récepteur panoramique qui fournit le spectre décamétrique, résultat de la fusion des données individuelles. Il faut distinguer dans ces données les bandes de radio-diffusion des autres bandes. L'analyse a montré que l'utilisation de ces dernières donne de meilleurs résultats que l'utilisation des bandes de radio-diffusion. Les analyses sont ensuite poursuivies par la segmentation du spectre dans les 95 bandes allouées auxquelles sont attachés 4 paramètres qui les représentent complètement. La fusion de données peut alors être complétée par l'insertion de paramètres solaires, de paramètres magnétiques et d'autres paramètres pouvant être obtenus par des méthodes totalement passives. Ces paramètres agissent comme variables exogènes dans le procédé de traitement des fusions de données. La fusion de données fournit alors les prévisions de la fréquence critique FoF2 de l'ionosphère et la prévision des niveaux des interférences. Pour effectuer la fusion de données il est préférable d'utiliser la méthode de la théorie des possibilités plutôt que la méthode de la théorie des probabilités car la théorie des possibilités introduit une variable supplémentaire qui permet de fiabiliser plus fortement la méthode de prévision.

Les résultats obtenus permettent d'assurer, sur chaque niveau d'interférences, avec un préavis de l'ordre de 4 à 5 heures, des précisions de niveau moyen d'interférences voisin de 7dB et une précision sur la fréquence foF2 de l'ordre de 0,4MHz. Ces résultats sont, la plupart du temps, satisfaisants pour résoudre les problèmes de gestion de systèmes ne pouvant utiliser des méthodes de prévisions actives telles que l'utilisation de sondages ionosphériques. Les performances obtenues fiabilisent fortement l'établissement et le maintien des télécommunications d'autant que ces systèmes peuvent être dans le cas de liaisons à courtes distances – inférieures à 1000km – implantés aux deux extrémités ce qui permet de faciliter, entre autres, les choix de fréquences identiques pour réaliser la liaison. En tout état de cause cette méthode apporte des avantages indéniables par rapport aux méthodes de prévision à long terme.

7. CONCLUSION

Les interférences constituent dans la gamme des ondes décamétriques une gêne très importante pour les télécommunications. Une modélisation spatio-temporelle des interférences montre qu'il est possible d'établir des prévisions pour prévoir au jour le jour leurs évolutions. Les résultats obtenus en matière de prévision des niveaux des interférences montrent qu'il est possible d'établir une modélisation suffisamment précise qui permet d'obtenir des prévisions à court terme avec une bonne fiabilité. L'opérateur dispose donc d'un système permettant de choisir a priori la fréquence qui restera claire pour la durée de la liaison qu'il a prévue. L'utilisation du spectre décamétrique pour l'établissement de prévision de propagation basée sur l'exploitation des évolutions de la forme du spectre apparaît possible et introduit une nouvelle méthode de connexion dans les systèmes opérationnels.

La méthode élaborée est simple et son expérimentation a donné d'excellents résultats. L'implantation de ce système de prévision passive de transmission des canaux ionosphériques apparaît possible et contribue à leur évaluation en temps réel. Testées sur des liaisons courtes distances – inférieures à 1000km – le succès de ces deux procédures a été constaté. Son extension à de grandes distances peut être envisagée mais il apparaît que la décorrélation spatiale des interférences doit être corrigée. Cette correction est accessible et il serait alors particulièrement intéressant d'introduire les résultats de ces mesures dans un processus de correction de prévisions à long terme ou de programmes universels qui permettrait d'obtenir une représentation de l'ionosphère et des interférences à l'échelle mondiale et par des méthodes totalement passives.

Références

1. Canat T., Caratori J., Goutelard C., « Modélisation spatio-temporelle des interférences électromagnétiques HF en Europe occidentale », in « Use or reduction of propagation and noise effects in distributed military systems », AGARD CP486, October 1990, pp 19.1-19.19

2. Caratori J., Goutelard C., « Compatibilité électromagnétique dans la bande décamétrique : modélisation, prévision, évaluation passive de l'ionosphère », in « Digital communication systems : propagation effects technical solutions », AGARD CP574, 18-21 september 1995, pp 23.1-23.9

Fusion de données et sondeur à extrêmement faible puissance pour télécommunications

C. Goutelard, C. Pautot
LETTI-UNIVERSITE PARIS-SUD
Bât. 214 – 91405 Orsay cedex
France

1. SUMMARY

Radiocommunications in decametric band are made difficult by important variability of transmission channel and by level and density of interferences which congest this part of the radioelectric spectrum.

Link establishment methods are currently based on a preliminary sounding of the channels and a scrutation of interference spectrum. This method is implemented by systems which works alternately with radiocommunication systems or use independent test system.

The disadvantages of these methods lie in an uneasy uses and a often insufficient periodic actualisation of estimation.

Implementation of system allowing to test channel with extremely low emitted power gives a new way very interesting to establish links and to keep them.

Electromagnetic soundings of the ionosphere is limited by the problem of electromagnetic pollution they create, which limits their use, and by the high sensitivity of the interference included in the band. The emitted power reduction solves the first point, on condition that sensitivity to the jammers remains acceptable.

This results is obtained for implementation of non linear signal processing which allows to choose the wave shape and coding by solving an optimisation problem without constraints. Results obtained allows to do soundings with very low power. Figure 1 presents a zenithal sounding measured with a emitted power of 10mW.

The establishment of this ionogram with the used method takes 26 seconds. Figure 2 and 3 show results obtained with a backscattering sounding and compare a classical method with the new method used, code time duration used and emitted power are identical in the two cases.

- On figure 2, an impulse response is represented.
- On figure 4, a scattering function is represented.

We can notice the protection given by the new method.

These news capacities allow a permanent ionospheric sounding, on all the frequencies. A data fusion can then be made taking into account available informations (figure 7), especially ionospheric disturbances.

This method can be applied in particular to the network management to improve their efficiency (figure 8).

2. GENERALITES

Les radiocommunications dans la gamme décamétrique sont rendues difficiles par les grandes variations des caractéristiques du canal et par le niveau et la densité des interférences qui encombrent cette partie du spectre radioélectrique.

Les méthodes d'établissement des liaisons sont actuellement basées sur un test préliminaire des canaux et une scrutation du spectre des interférences. Ces méthodes mettent en œuvre des systèmes de test séparés. Les inconvénients de ces méthodes sont liés aux lourdeurs d'utilisation et à la réactualisation périodique souvent insuffisante des estimations.

La mise en œuvre de systèmes permettant de tester les canaux avec des puissances d'émission extrêmement faibles, donne une dimension nouvelle particulièrement intéressante pour l'établissement des liaisons et leur maintien.

Les sondages électromagnétiques de l'ionosphère se heurtent au problème de la pollution électromagnétique qu'ils créent, ce qui limite leur utilisation, et à leur sensibilité aux interférences présentes dans la bande. La réduction de la puissance d'émission résout le premier point à condition que la sensibilité aux brouillages demeure acceptable.

Ce résultat est atteint par l'utilisation de traitement de signaux non linéaires qui permettent de choisir la forme d'onde et le codage par la résolution d'un problème d'optimisation sous contraintes.

Il est présenté, dans ce document, les résultats de méthodes de sondages électromagnétiques de l'ionosphère associant des durées d'analyse très courtes à des puissances émises très faibles. Partant de ces résultats, il a été envisagé une méthode permettant d'effectuer des sondages permanents de l'ionosphère sans entraîner de pollution électromagnétique du spectre décamétrique, ce qui permet d'envisager des méthodes d'analyse de canaux qui, associées à des méthodes de prévision d'interférences, permettent une gestion optimale beaucoup plus sûre que par les méthodes conventionnellement utilisées. Ces méthodes, particulièrement économiques car elles nécessitent des puissances très réduites, conduisent à des systèmes de sondage indétectables auxquels il ne peut donc être reproché de gêner les autres utilisateurs [1] [2].

Le système envisagé consiste à allier le sondage zénithal de l'ionosphère et une méthode de sondage par rétrodiffusion tout azimut qui permet de dresser, à partir d'un point unique situé au sol, une cartographie de l'ionosphère et de connaître ainsi sur une zone de 6000km de diamètre les modes de propagation pouvant exister entre deux points quelconques de la zone. L'information apportée par la mesure de la fonction de diffusion permet également de prendre en compte les effets dynamiques du milieu et d'assurer ainsi une meilleure fiabilité des systèmes de télécommunication haute fréquence par voie ionosphérique.
La méthode proposée s'applique à tous les types de sondages de l'ionosphère, verticaux, par rétrodiffusion ou obliques. Dans cette dernière utilisation, le principe de la gestion de réseau doit être considéré comme une extension des méthodes actuelles.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

3. SONDAGES ELECTROMAGNETIQUES DE L'IONOSPHERE AVEC DES SYSTEMES A EXTREMEMENT FAIBLE PUISSANCE

Il est bien connu que les résultats de sondages électromagnétiques qui s'appuient sur les techniques radars peuvent, par l'utilisation de séquences codées, fournir des résultats de grande qualité pourvu que l'énergie consacrée à une mesure soit suffisante.

La technique de compression d'impulsions, utilisée depuis longtemps, a permis de réduire les puissances des radars dans un facteur de plusieurs dizaines. L'utilisation de la technique de l'émission continue par modulation linéaire de fréquence - technique « CHIRP » - a permis d'effectuer des sondages avec des puissances très faibles mais cette technique nécessite des systèmes bistatiques.

En 1968, le LETTI a proposé une méthode dite d'impulsions longues codées qui a consisté à répartir des codes, non plus à l'intérieur d'une impulsion comme dans le cas des systèmes à compression d'impulsions, mais sur un ensemble d'impulsions consécutives de façon à pouvoir conserver les propriétés des systèmes monostatiques et à utiliser des séquences très longues. Cette méthode a donné rapidement de très bons résultats mais elle a été limitée par l'instabilité du milieu de propagation qui fixe une borne supérieure à la séquence et, par conséquent, en limite les performances. Les applications de cette méthode ont donné lieu à la mise en œuvre de la station **STUDIO** (**S**ystème de **T**raitement **U**niversel de **D**iagnostic **IO**nosphériques) du LETTI en 1968 et des applications de cette technique ont été reprises dans le radar transhorizon français **NOSTRADAMUS** (**N**ouveau **S**ystème **TRA**nshorizon **D**écamétrique **A**ppliquant les **M**éthodes **U**tilisées dans **S**tudio). Une station a également été construite au C.N.E.T. (Centre National d'Etude des Télécommunications) dans laquelle le codage et le traitement de signal de la station STUDIO ont été introduits dès 1985.

Les progrès dans l'efficacité des méthodes de sondage de l'ionosphère ne pouvaient plus dès lors s'appuyer sur ces techniques qui avaient été testées durant de longues années. Deux voies ont été développées au LETTI et ont conduit à des avancées significatives dans le domaine de la réduction des puissances utilisables. La première a consisté à rechercher des séquences mieux appropriées aux objectifs visés et la seconde à incorporer dans le traitement de signal des méthodes non linéaires.

3.1. Séquences adaptées aux techniques de sondage

L'utilisation des techniques de codage par l'utilisation de séquences émises pour obtenir des taux de compression élevés est basée essentiellement sur l'analyse des propriétés de la fonction de diffusion du signal utilisé. Les caractéristiques générales de la fonction de diffusion sont bien connues, le volume enserré par le carré du module est une constante et la fonction de diffusion est maximale pour une fréquence Doppler nulle et un décalage temporel nul. La fonction de diffusion permet de caractériser les propriétés d'un signal en présence de bruit gaussien mais en présence d'interférences, dont les caractéristiques ne sont pas gaussiennes, les caractéristiques de la fonction de diffusion ne sont plus suffisantes. La réduction de la puissance émise s'appuie sur la prise en compte des paramètres suivants :

- Les propriétés d'autocorrélation de la séquence utilisée qui doivent rester excellentes, c'est-à-dire proches de la fonction de corrélation parfaite qui prend des valeurs nulles en dehors du pic central.
- La fonction de diffusion doit présenter une forme non homogène en dehors du pic central afin d'apporter une protection maximale dans la zone correspondant au Doppler dont peut être affecté le signal utile.
- La séquence doit avoir des caractéristiques qui entraînent, vis-à-vis des interférences qui constituent les signaux les plus gênants de la gamme décamétrique des protections maximales.

Concernant la première caractéristique, il apparaît que les séquences classiquement utilisées, tels que les codes PN - ou codes de Huffman – n'apportent pas de protection suffisante sur la fonction de corrélation et entraînent des biais dans les signatures des sondages de l'ionosphère. Les séquences de Barker apparaissent trop courtes et les séquences de Golay nécessitent une émission impulsionnelle incompatible avec des taux de compression très élevés. Il est intéressant d'utiliser des séquences longues ayant des fonctions de corrélation aussi proches que possible que de la corrélation parfaite. Il a été démontré [3] qu'il n'existe pas de séquences binaires ayant des fonctions de corrélation parfaites et que les meilleures séquences devaient admettre, en dehors du pic central de corrélation, un pic secondaire situé au milieu de la plage nulle de la fonction de corrélation périodique de la séquence. Ces séquences qui ont été déterminées [3] ont été appelées séquences W.G. et il est montré que ces séquences existent pour de très nombreuses longueurs multiples de 4 qui ont été déterminées jusqu'à des longueurs 20 000. Pour chaque longueur il existe également une grande variété de séquences dont les propriétés sont décrites.

En ce qui concerne les séquences Q-aires, des séquences telles que les séquences de Franck ont été trouvées. Cependant, ces séquences apparaissent comme un cas particulier des séquences plus générales qui ont été publiées dans [4] et qui présentent des conditions d'orthogonalité particulièrement utiles en matière de protection vis-à-vis des brouilleurs et des problèmes de détection en général et une méthode de détermination de séquences plus générale a été proposée dans [5] et a conduit à une technique de détermination générale qui permet de prendre en compte des contraintes supplémentaires telles que les caractéristiques générales des interférences. Ces techniques ont été mises en œuvre pour déterminer un ensemble de séquences particulièrement adaptées au sondage électromagnétique de l'ionosphère.

3.2. Traitement des signaux

La seconde voie qui a été utilisée pour réduire les puissances d'émission concerne les traitements linéaires susceptibles d'être mis en œuvre. Des traitements non linéaires telle que la technique d'excision ont été proposés dès le début des années 1980 [6] et donnent des résultats satisfaisants dans le cas où l'occupation spectrale du brouilleur est faible devant le domaine fréquentiel occupé par le signal utile. D'autres techniques non linéaires plus puissantes peuvent être utilisées comme l'utilisation des moments d'ordre supérieur, mais leur mise en œuvre pratique se heurte au volume de calculs nécessaires pour les utiliser. Bien que des algorithmes visant à réduire la complexité des calculs aient été proposés, il apparaît encore très difficile de songer à les appliquer dans des systèmes opérationnels. Des concepts nouveaux peuvent être mis en œuvre en s'appuyant sur la définition d'un processus non linéaire défini par sa réponse impulsionnelle multidimensionnelle et qui permet de rendre compte de la totalité du processus et également de le définir. A partir de ce concept, il devient possible d'examiner la réponse d'un système non linéaire aux différents types d'interférences rencontrés dans

le spectre décamétrique. Le processus peut alors être optimisé par la résolution d'un problème d'optimisation sous contraintes. Le problème se réduit à rendre maximum un paramètre – qui peut être le rapport signal à bruit en sortie du processus – sous la contrainte de l'utilisation d'un signal utile constitué par une séquence ayant de bonnes caractéristiques en présence de signaux parasites constitués par les différents types d'interférences.

Les processus non linéaires peuvent alors être déterminés par la résolution informatique de ces problèmes d'optimisation. L'utilisation de processus non linéaires offre sur les processus linéaires l'avantage considérable d'offrir des structures extrêmement diverses qui permettent d'obtenir une optimisation plus efficace dans un cadre plus général.

4. RESULTATS OBTENUS AVEC LES SONDEURS A EXTREMEMENT FAIBLE PUISSANCE

Les performances obtenues par les techniques mises en œuvre sont illustrées par un ensemble d'exemples présentés dans ce paragraphe. Ces techniques sont actuellement utilisées dans les systèmes de sondages zénithaux, obliques bistatiques et par rétrodiffusion. Deux types de résultats sont présentés, l'un concernant le sondage zénithal, l'autre le sondage par rétrodiffusion.

Le sondage zénithal est une mesure qui est particulièrement bien connue et les techniques utilisées ont été fréquemment décrites. Dans les systèmes monostatiques les séquences de Golay sont couramment utilisées pour réaliser le sondage. On a comparé les résultats obtenus dans l'utilisation des séquences de Golay avec les résultats obtenus en utilisant la technique proposée. On a utilisé pour effectuer la comparaison deux mesures permettant d'effectuer des sondages avec la même durée. La figure 1 représente les résultats obtenus avec ces deux méthodes. L'ionogramme de la figure 1a a été obtenu en utilisant des séquences de Golay à 16 moments avec une puissance d'émission de 150W. La durée de mesure est de 26 secondes et on peut voir apparaître les échos obtenus sur la région E et sur la région F. On peut noter que le mode 2F est parfaitement visible. La figure 1b représente l'ionogramme obtenu le même jour avec la méthode proposée. La puissance d'émission a été réduite à 10mW et la durée de mesure, avec la technique proposée, est réduite à 22 secondes. On peut noter que la région E et la région F sont parfaitement visibles et que les interférences présentes sur l'enregistrement effectué avec des codes de Golay sont fortement atténuées. Ces résultats illustrent la possibilité d'effectuer des sondages zénithaux avec de très faibles puissances.

Figure 1 : *Comparaison des performances des ionogrammes zénithaux obtenus*

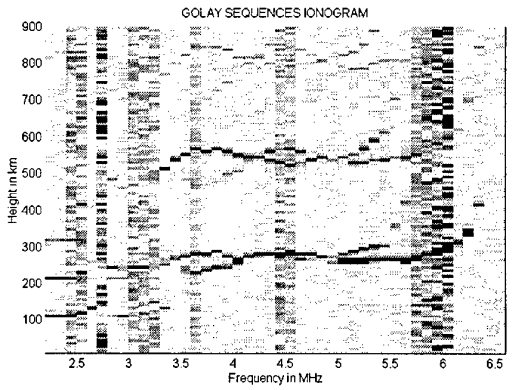


Figure 1.a : *150W Code de Golay, Temps de mesure : 26s*

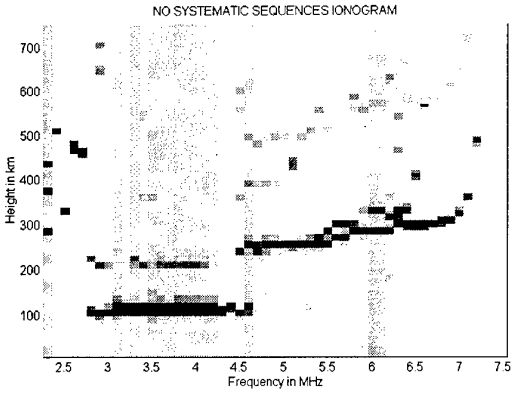


figure1.b : *Nouvelle méthode code non systèmatique 10 mW STUDIO 1997 Temps de mesure : 22s*

La même méthode est utilisée pour effectuer des sondages par rétrodiffusion qui exigent des puissances supérieures à celles utilisées dans les sondages zénithaux. Le principe du sondage par rétrodiffusion s'appuie en effet sur la technique radar appliquée à des distances de plusieurs milliers de kilomètres alors que le sondage zénithal peut être considéré comme une transmission point à point la réflexion sur l'ionosphère étant du type spéculaire. Les différences de bilans de propagation sont de l'ordre de 120 à 140dB ce qui nécessite, pour le sondage par rétrodiffusion, des puissances habituellement beaucoup plus élevées que dans les sondages zénithaux. Dans les sondages par rétrodiffusion les puissances ont progressivement diminué de plusieurs centaines de kilowatts initialement utilisées à quelques kilowatts dans les méthodes actuelles. La nouvelle méthode permet d'obtenir des résultats avec des puissances plus réduites.

La figure 2 illustre les performances obtenues en rétrodiffusion. La figure 2a montre un écho de rétrodiffusion s'étendant jusqu'à 7500km pour une puissance d'émission de 300W. La figure 2b représente une évolution de la réponse impulsionnelle du canal sur une durée de 30 minutes. Ces résultats ont été obtenus par une méthode déjà élaborée dans les années 1990 et qui a été depuis améliorée. La figure 2c représente une réponse impulsionnelle obtenue avec une puissance d'émission de 100milliwatts et la figure 2d représente l'évolution temporelle de cette réponse impulsionnelle sur une durée de 30 minutes. Ces résultats ont été obtenus avec l'utilisation d'une antenne log-periodic dont le gain, dans la direction du lobe principal est de l'ordre de 9dB. Il va de soi qu'associer à un réseau les résultats sont améliorés de façon sensible.

Figure 2 : *Mesures en rétrodiffusion*

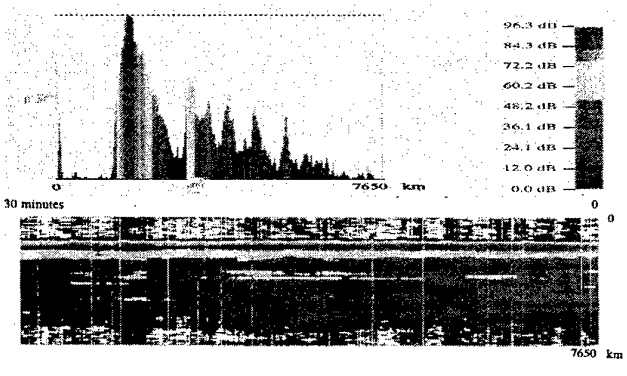


Figure2.a et 2.b : *Nouvelle méthode 300W 1988*

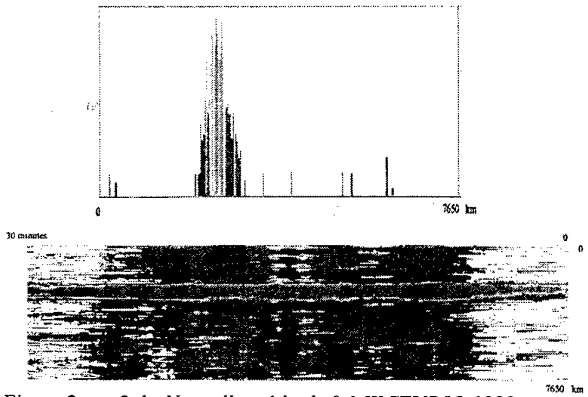


Figure 2.c et 2.d : *Nouvelle méthode 0.1 W STUDIO 1988*

La figure 3 permet de juger de l'efficacité des séquences mises en œuvre. Pour tester les séquences, on a effectué sur une période de 12 minutes deux essais consécutifs en conservant la puissance d'émission invariante et sans modifier les réglages de l'émetteur et du récepteur de la station. La fréquence d'émission a été choisie de façon à ce que la mesure soit perturbée par des interférences gênant la réception. Durant les six premières minutes, on a utilisé une séquence pseudo-aléatoire de longueur 4095. On peut constater que, durant cette période, le brouillage est complet et qu'aucune trace de l'écho de rétrodiffusion n'est décelable. Durant les six dernières minutes on a utilisé une séquence nouvelle appelée code non systématique. La longueur de la séquence utilisée est de 3937, c'est-à-dire extrêmement voisin de celle de la séquence pseudo-aléatoire bien qu'on puisse noter que les longueurs choisies avantage cette dernière. On peut noter, durant ces six dernières minutes, l'apparition de l'écho de rétrodiffusion et une réjection totale des interférences apparaissant dans cette période. Il convient de noter que cette illustration met en évidence l'importance du codage, les traitements de signaux étant strictement identiques pour la séquence pseudo-aléatoire et pour la séquence non systématique.

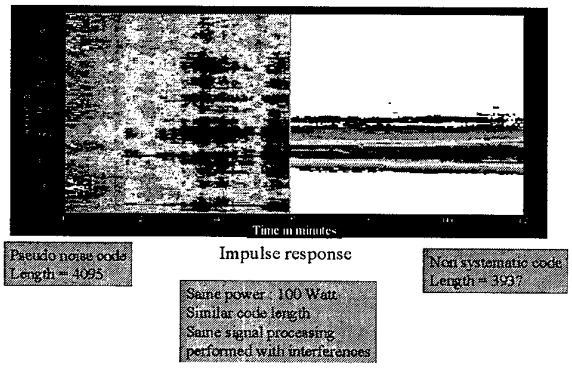


Figure 3 : *Protection contre les interférences avec méthode de codage classique et nouveau codage utilisant un code non systématiques*

La figure 4 illustre enfin les résultats obtenus pour la mesure de fonction de diffusion. On a mesuré à la même fréquence et de façon immédiatement consécutive deux fonctions de diffusion, l'une par l'utilisation d'une séquence pseudo-aléatoire, l'autre par l'utilisation d'une séquence non systématique. On peut noter la présence d'un brouilleur qui rend invisible l'écho de rétrodiffusion dans le cas de la détection par une séquence pseudo-aléatoire. Sur la figure inférieure, on peut noter que le changement de séquence permet de retrouver l'écho de rétrodiffusion. Il faut signaler encore que dans cette expérimentation, les réglages de l'émetteur et du récepteur ont été invariants.

Les résultats présentés donne une illustration des performances atteintes actuellement par les nouvelles méthodes mises en oeuvre.

Figure 4 : *Effets des interférences sur la fonction de diffusion*

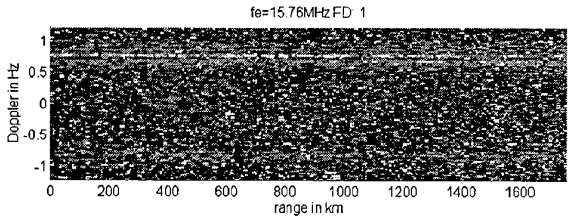


Figure 4.a : *Code pseudo-aléatoire*

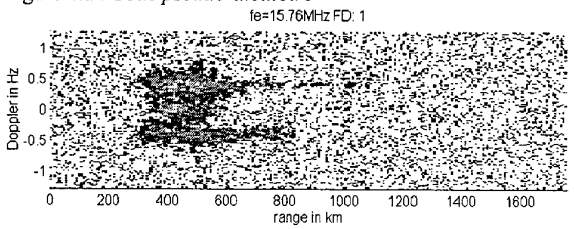


Figure 4.b : *Code non systématique*

Le tableau 1 résume une évolution des puissances utilisées habituellement pour effectuer différents types de sondages : vertical, oblique bistatique et par rétrodiffusion. Avec les nouvelles méthodes présentées, les puissances pouvant être utilisées dans des systèmes opérationnels sont de l'ordre du watt pour les sondages verticaux et obliques bistatiques et de l'ordre de 50 à 100 watts pour les sondages par rétrodiffusion. Ces ordres de grandeur ouvrent de nouvelles possibilités pour la réalisation de systèmes de gestion automatique des télécommunications dans la bande décamétrique.

YEARS	VERTICAL	OBLIQUE BISTATIC	BACK-SCATTERING
1960's	10 kW	10 kW	100 kW
1970's	1 kW	1 kW	100 kW
1980's	100 W	100 W	10 kW
1997	1 W (10mW)	1 W (10mW)	100 W (0,1 W)

Tableau 1 : *Evolution des puissances utilisées*

5. APPLICATION AUX TELECOMMUNICATIONS. GESTION DE RESEAUX

Compte tenu des résultats qui peuvent actuellement être utilisés, on a considéré que la gestion des réseaux pouvait être améliorée par une utilisation plus fréquente des sondeurs. Les hypothèses prises pour concevoir un système de gestion de réseaux en temps réel sont les suivantes :

- Les puissances utilisées par les sondeurs - typiquement 1W pour les sondages zénithaux et obliques bistatiques, 50 à 100W pour les sondages par rétrodiffusion – sont suffisamment faibles pour que les sondeurs puissent être utilisés sans apporter de nuisances aux autres utilisateurs et sans être détectables par ceux-ci. On peut remarquer que les sondages par rétrodiffusion qui utilisent traditionnellement des largeurs de bandes de plusieurs

dizaines de kHz conduisent à des densités de puissances extrêmement faibles par bandes de 3kHz. A titre d'exemple, l'utilisation d'une puissance de 60W avec un signal dont le spectre s'étend sur 30kHz conduit à une densité de puissance émise de 6W pour 3kHz ce qui rend son émission indétectable par des systèmes classiques.

- Le système de prévision est un système centralisé utilisant un sondage zénithal et/ou un sondage par rétrodiffusion.
- Des sondages obliques bistatiques peuvent être utilisés en complémentarité avec les sondages précédents ou bien de façon isolée à condition que ces sondages soient inclus dans les systèmes de télécommunication.

Nous nous intéresserons dans cette communication à la gestion centralisée d'un système utilisant les sondages zénithaux et par rétrodiffusion. On considérera que le sondage zénithal donne une information de bonne précision sur le profil d'ionisation à la verticale de la station. Le sondage par rétrodiffusion fournit les cinq types d'enregistrements (figure 5) : la fonction de diffusion du canal, de l'ionogramme de rétrodiffusion, une représentation PPI de la propagation et les propagations longues distances faisant apparaître, en particulier, les super-modes. Ce type de sondage offre beaucoup d'intérêt car il permet de mettre en évidence l'anisotropie du milieu ionosphérique, de faire apparaître directement les MUF et donner accès à l'évaluation des LUF ainsi que par l'intermédiaire de la fonction de diffusion et de la représentation PPI de la détection des perturbations ionosphériques, de leur évolution et de leur prévision.

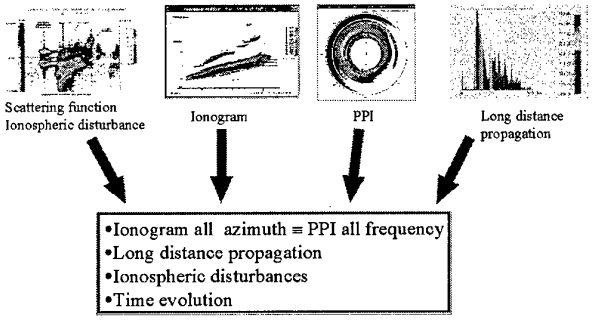


Figure 5 : *Données issues du sondage par rétrodiffusion*

L'ensemble de ces mesures est alors traité par une méthode de fusion de données dans laquelle les grandeurs d'entrée sont (figure 7) :

- Les sondages zénithaux, les sondages obliques bistatiques, les sondages par rétrodiffusion.
- Les données relatives à l'activité solaire.
- Les données relatives à l'activité magnétique.
- D'autres mesures, telles mesures pouvant provenir de satellites, de réseaux de sondeurs, de mesures ionosphériques diverses…

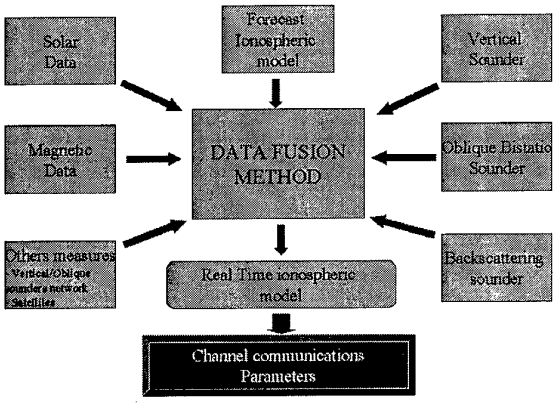


Figure 6 : *Fusion de données pour une station de sondage ionosphérique unique*

La fusion de données réalise alors une réactualisation du modèle ionosphérique par la prise en compte de toutes les données disponibles pour fournir un modèle ionosphérique temps réel sur une zone de 6000km de diamètre. Les résultats sont alors présentés par des cartes telle que celle apparaissant sur la figure 8 permettant de déterminer les propriétés du canal pour des liaisons s'effectuant soit entre un point voisin de la station et un correspondant situé dans la zone, soit entre deux correspondants situés dans cette zone. Le principe consiste, pour établir ces liaisons, en des tracés de rayons obtenus à partir du modèle ionosphérique temps réel réactualisé par la fusion des données disponibles. La fusion de données peut être traitée par différentes méthodes. L'une des méthodes les plus classiques consiste à utiliser la théorie des probabilités. Il est apparu, lors de l'analyse de ce problème, que de meilleurs résultats étaient obtenus en utilisant la théorie des possibilités. Cette théorie présente sur la théorie des probabilités l'avantage d'introduire une variable supplémentaire qui permet une solution plus complète et plus précise du problème.

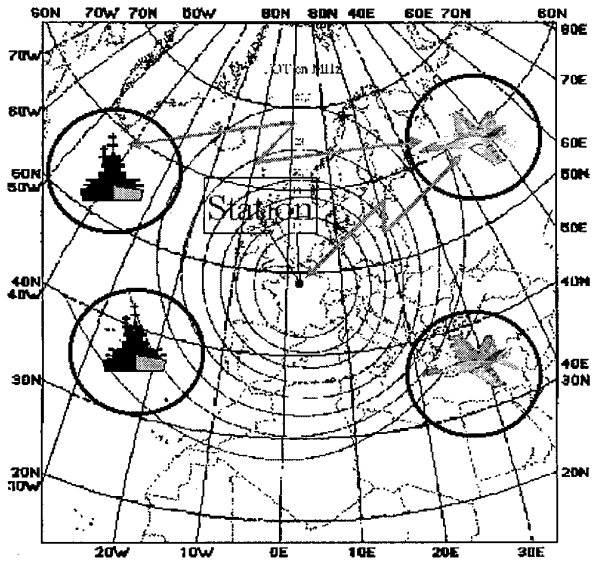


Figure 7 : *Exemple d'application à la gestion d'une réseau*

La méthode de la théorie des possibilités utilisée peut se résumer de la façon suivante :

1) On définit un ensemble flou par

$$\mu_A = U \rightarrow [0,1]$$

A est en ensemble flou défini sur le domaine A avec

a)

$$\forall u \in U$$
$$\mu_{A \cap B_{(u)}} = \max\left(\mu_{A_{(u)}}, \mu_{B_{(u)}}\right)$$
$$\mu_{A \cup B_{(u)}} = \min\left(\mu_{A_{(u)}}, \mu_{B_{(u)}}\right)$$
$$\mu_{\neg A_{(u)}} = 1 - \mu_{A_{(u)}}$$

où $\mu_{\neg A_{(u)}}$ désigne la mesure du contraire de $A_{(u)}$

b)

$A \subseteq B$ implique

$\forall \in U \quad , \quad \mu_{A_{(u)}} \leq \mu_{B_{(u)}}$

c)

$$\forall_u \in U \quad , \quad \forall_v \in V$$
$$\mu_{A \times B(u,v)} = \min\left(\mu_{A_{(u)}}, \mu_{B_{(v)}}\right)$$

2) μ est défini par une forme trapèzoïdale symétrique

- qui a pour support $\mu_{(u)} > 0$
- dont le noyau est défini par $\mu_{(u)} = 1$

Le support et le noyau définissent totalement $\mu_{(u)}$ qui est déterminé pour chaque événement.

3) La résolution du problème se réduit alors à la recherche de la probabilité maximale de la prévision effectuée.

L'application de cette méthode conduit à des prévisions très précises qui permettent, avec des temps de préavis de 1 à 2 heures, de déterminer les fréquences optimales avec des probabilités supérieures à 99%.

5. CONCLUSION

La possibilité d'effectuer des sondages de l'ionosphère avec des puissances extrêmement faibles conduit à un nouveau concept de gestion de réseaux en temps réel. Les techniques usuelles d'évaluation en temps réel des canaux (RTCE) sont souvent appliquées dans des systèmes de sondage point à point par l'utilisation des équipements de transmission. Par rapport à cette méthode de conception déjà ancienne, la méthode proposée offre l'avantage d'une scrutation permanente de l'ionosphère avec des systèmes non polluants, indétectables qui peuvent être, soit incorporés dans des équipements de télécommunication, soit fonctionner de façon autonome et parallèle. Les avantages liés à cette méthode tiennent essentiellement au fait que la scrutation peut être faite de façon adaptative par l'opérateur dans toutes les gammes de fréquences ce qui n'est pas toujours possible dans les systèmes RTCE. La dimension spatiale introduite par le sondage par rétrodiffusion qui devient une méthode légère compte tenu des faibles puissances mises en jeu et par la simplicité des aériens pouvant être utilisés – une simple antenne log-periodic suffit – permet une gestion à partir d'un site unique avec une prévision spatiale des perturbations qui transitent à travers l'ionosphère.

L'avance apportée par l'utilisation de puissances extrêmement faibles permet d'échapper aux critiques justement faites dans le passé de systèmes trop polluants ou trop onéreux pour être utilisés. Il apparaît désormais que ces systèmes peuvent être incorporés à faible coût et sans pollution dans les équipements futurs. Les progrès effectués dans l'avenir sur les traitements de signaux et sur les techniques de codage permettront probablement de rendre ces méthodes encore plus attractives et contribueront à augmenter la fiabilité des télécommunications.

Références

1. Goutelard C., Caratori, Joisel A., « Electromagnetic sounding technique using spatial and spectral sampling of the reception signals. Application to the study of inhomogeneities in ionosphere plasma », in « Aspects of electromagnetic waves scattering in radiocommunication », AGARD CP244, October 1977, pp13.1-13.21

2. Goutelard C., « NOSTRADAMUS : Projet français de radar transhorizon », in « Use or reduction of propagation and noise effects in distributed military systems », AGARD CP486, October 1990

3. Goutelard C., « Les séquences WG », in « Radiolocations techniques », AGARD CP528, Juin 1992

4. Goutelard C., « Les séquences GQ ; Séquences Q-aires orthogonales à corrélation parfaite », in « Digital communications systems : propagation effects technical solutions », AGARD CP574, 18-21 September 1995, pp 28.1-28.13

5. Goutelard C., « Coding and extremely low power sounder », in « HF radio systems and techniques », IEE, 7-10 July 1997

6. Goutelard C., Joisel A., « Protection adaptative des systèmes de détection fonnctionnant par étalement de spectre », in « Propagation factors affecting remote sensing by radio waves », AGARD CP345, 24-28 Mai 1983, pp 9.1-9.13

HF Transmitter Location using Super-resolution DF and Ionospheric Sounding

Hal J. Strangeways
Institute of Integrated Information Systems
School of Electronic and Electrical Engineering
The University of Leeds
Leeds LS2 9JT
U.K.

1. SUMMARY

A new method of SSL (Single Site Location) is presented which makes use of multi-sensor information. Also, unlike conventional SSL methods which generally ray-trace backwards using a single angle of incidence and azimuth, this method uses the incident angles of two or more simultaneous paths (e.g. 1 and 2 hop F2) to determine both an updated ionosphere model at the reflection point(s) and the location of the target transmitter. The angles of azimuth of arrival and elevation are determined using a super-resolution algorithm to decompose the multipath wavefield received on an antenna array. Analytic models are used to approximate the ionosphere and determine propagation paths. If vertical ionospheric sounder data is available, the obtained electron density profile can be fitted using quasi-parabolic segments (MQP model) and parameters of this model optimised in the minimisation procedure. A Nelder Mead [1] simplex algorithm is used for the function minimisation. Greater accuracy in unknown transmitter location can be achieved by including additional constraints in the system such as measured elevation angles of known transmitters.
The SSL determination is performed under MATLAB via a menu-driven graphical user interface. This employs pull-down menus for data input, program choice, ionospheric profile plotting and presentation of results.

2. INTRODUCTION

The SSL method enables the location of a distant HF transmitter to be determined from a single site. This avoids the need for employing two or more receiving sites and the additional complication of communications between them. Further, for the case when the base line for triangulation is short, it can even determine more accurate transmitter locations than triangulation. The main source of error in the conventional HF transmitter SSL technique has generally been found to be the characterisation of the ionosphere at the time of observation [2]. The method presented reduces errors from this source by using information about the ionosphere implicit in the received data from the target and, if available, vertical or oblique sounders and other HF transmitters. The elevation angles of simultaneously received paths from any given transmitter inherently contain information regarding the real-time ionosphere conditions. A method by which electron density profiles can be updated from arrival angle measurements of known transmitters has previously been presented [3]. The updated profiles could then be used to improve the accuracy of SSL determination. The present method differs in that a joint solution is made for both an updated ionosphere model and the distance to the unknown transmitter and also in the ease with which data from other sensors or additional parameters to be optimised can be included in the determination. When the input data just consists of the angles of arrival of 1 and 2 hop paths from the target transmitter, the Nelder-Mead simplex algorithm is used to minimize a multivariable function, this function expressing the difference between the two simultaneous paths as a function of the various variables of the real-time ionosphere that are optimised e.g. foE, foF2, hbF2, hmF2. The method can only be applied to the situation where two simultaneous paths exist from transmitter to receiver. However, this is a condition which is often satisfied and it has been found that super-resolution D/F methods can be used to successfully decompose the received HF wavefield, received on a suitable antenna array, into its constituent component, giving the DOA of each [4,5,6,7,8]. Thus such a scheme compliments more traditional SSL methods using interferometry which are generally only performed for single moded propagation.

3. ARRIVAL ANGLE DETERMINATION

A 6 channel HF receiver was deployed together with a 6 element 2D antenna array consisting of vertical monopoles. Signals received by the array were input to the 6 channel receiver and each mixed down to 20 kHz before A/D conversion at 16 kHz.. Super-resolution algorithms were developed to efficiently decompose the multi-component HF wavefield into its constituent paths [4,5,6,7,8].These included DOSE (Direction Of arrival by Signal Elimination)[4,5] , a null-steering algorithm, which operates on the data vector and thus can work on single snapshot data and fast versions of MLE which operates either in element [6,7] or beam space [8]. All these algorithms are based on MLE ensuring that they are robust, can work successfully with coherent sources (as arise in multipath) and also can also operate well at low SNR. The algorithms were found to very significantly outperform MUSIC and MVE (CAPON) both on real and simulated data [4,9].

The azimuth and elevation of one and two hop paths from the unknown transmitter were thus obtained for input to the range-finding program. The range-finding program does not require the propagation mode to be identified (e.g.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

low/high angle, F1 or F2 reflection); only the number of hops (ionospheric reflections). The ratio of the tangents of the observed elevation angles (tan 2-hop angle/tan 1-hop angle) can be determined to aid this investigation. This value will be 2 for a flat Earth and equal virtual reflection heights, will be larger than 2 for greater ranges where the curvature of the Earth is important and can be less than 2 for shorter ranges where the virtual height for the 2 hop path exceeds that of the 1 hop path. The array was not able to separate o- and x- modes but the analytical models used to model the ionosphere in the optimisation process are, in any case, unable to account for magnetic field effects. It would be possible, however, to develop a system which considered both magneto-ionic modes if an antenna system incorporating crossed loops were employed together with a numerical ray-trace program in the function minimisation procedure.

4. RANGE-FINDING PROGRAM

The program is operated via a menu-driven graphical interface as shown in figure 1. This is first used to input initial data: transmitter frequency, time and date of measurement, sunspot number, receiver location and angles of elevation of 1 and 2 hop paths and their azimuth. The sunspot number can be determined from the date if required. The method first determines expected parameters for the ionosphere at the reflection point(s). The simplest method is to assume that the ionospheric parameters are the same at the one or two hop reflection points as then it is only necessary to optimise for one fixed set of ionospheric parameters and the ray paths can be determined from analytic formulae. However, it is also possible to assume a difference in the ionosphere between the 1 hop and 2 hop reflection points as will occur if horizontal gradients are present. These are most likely to be important near dawn and dusk. In this case a numerical ray-tracing program or a segmented MQP model [10] is required to determine the ray paths. The entirety of both 1 and 2 hop paths are then determined rather than assuming the range of both hops of the 2 hop path are identical. In this case optimisation can still be performed for a single set of ionospheric parameters by introducing the difference between the ionospheric parameters at the different reflection points by way of a specified model. Alternatively, the optimisation can be performed to determine ionospheric gradients as well as vertical profiles but more input data from reference transmitters (see below) would be required. The optimisation program requires as good as possible starting parameters for the ionosphere at the reflection point(s). This can be provided either just on the basis of an ionospheric model or using real-time information from a vertical sounder or sounders situated at the location of the SSL system of elsewhere to update an ionospheric model. For the former method, ionospheric parameters are first obtained for the time of day, date and receiver location from the IRI90 (International Reference Ionosphere) (step 1). Any other ionosphere model could be used instead if required. An estimated distance is then determined using the one hop elevation angle and propagation distance formulae based on the MQP [11], (step 2). Using the azimuth angle and estimated 1-hop distance, a new estimated location for the one hop reflection point is determined (step 3). This new location is then used to update the ionosphere parameters using the IRI90 model. (step 4). The steps 1 to 4 are then repeated starting with this new location. An additional subroutine is used to correct the initial parameters if it is found that an expected mode cannot be reflected by the ionosphere. Finally the IRI90 model at the final estimate of 1 hop reflection point is plotted (see figure 2). If real time ionospheric profiles are available from a sounder, this data can be used to further improve the estimate of the electron density profile at the estimated reflection point(s). An analytical model is then fitted to the best estimate of the electron density profile at the reflection point(s). A pull down menu (labeled 'model') can then be used to plot any or all of the ionospheric profiles based on the three models: a single quasi-parabolic layer [12], a modified Bradley-Dudeney (CCIR) ionospheric profile [13] or an MQP (multi-quasi-parabolic). Greater accuracy in fitting the MQP model to the data can be achieved by using a greater number of quasi-parabolic layers but this may requires a careful choice of ionospheric parameters to include in the optimisation since it is not realistic to optimise for all the parameters of all the layers. The fitted profiles are plotted on the same axes so that their correspondence with the estimated electron density profile at the reflection point model can be determined (see figure 2). For each of these 3 analytical models (QP, CCIR, MQP), , analytical formulae exist whereby the propagation distance can be determined. Another pull-down menu (labeled 'formula') can then be used to select which of the three models is to be used in the optimisation procedure. In all these analytical models, the effect of the magnetic field is neglected so that individual magneto-ionic modes cannot be treated. The optimisation is performed to determine the "best-fit" ionospheric parameters and hence propagation distance. Only the most important parameters of the model are treated as variables. These can be determined by varying each parameter in turn over a given range e.g. $\pm$ 20% corresponding to the likely variation of the parameter in the real ionosphere and then examining the effect on the determined propagation distance. For example, in the case of the modified Bradley-Dudeney model, the ionosphere is represented using two quasi-parabolic layers, representing the E and F2 layers and a quasi-linear layer in between. The height of the maximum and semi-thickness of the E-region were chosen as fixed parameters with values of 110 and 20 km respectively. Then, there are 4 variable parameters, foE, foF2, hmF2, ymF2; the optimisation is performed for corrected values to these 4. The fact that the total range of the 2-hop path must equal that of the 1-hop distance is used as a constraint. The algorithm used is the Nelder Mead simplex search [1] which is a direct search method not requiring gradients or derivatives. If n is the number of

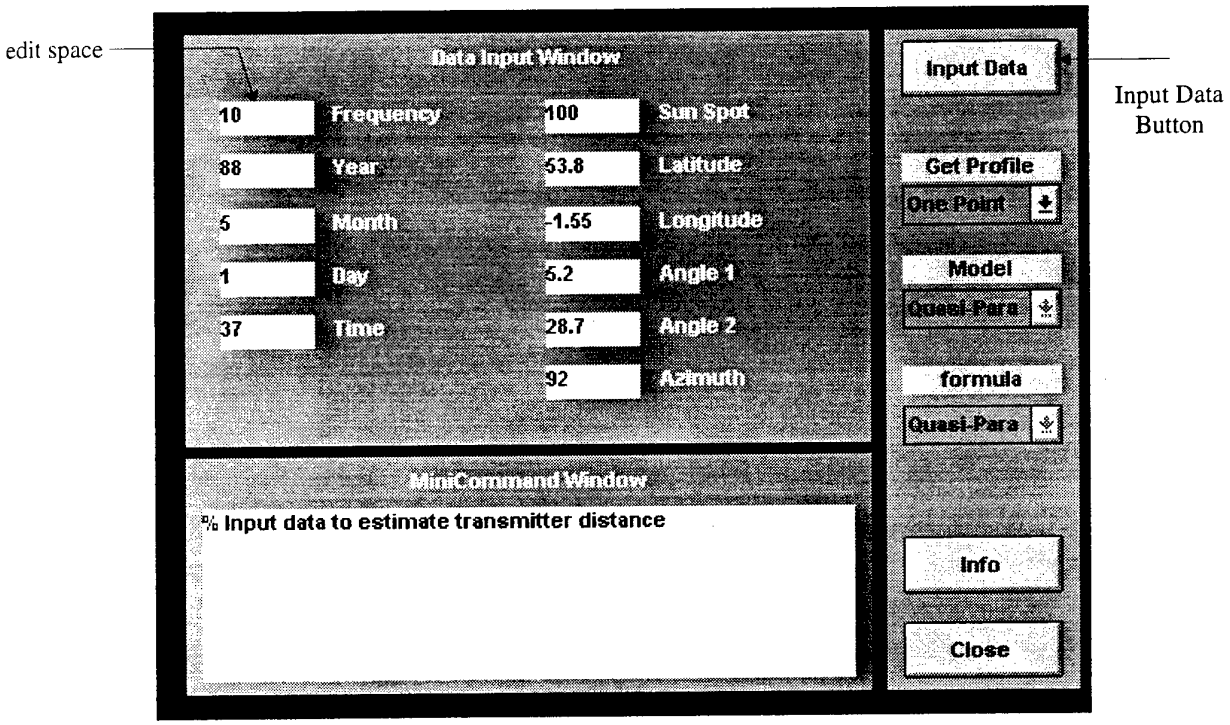


Figure 1 Input Data Screen

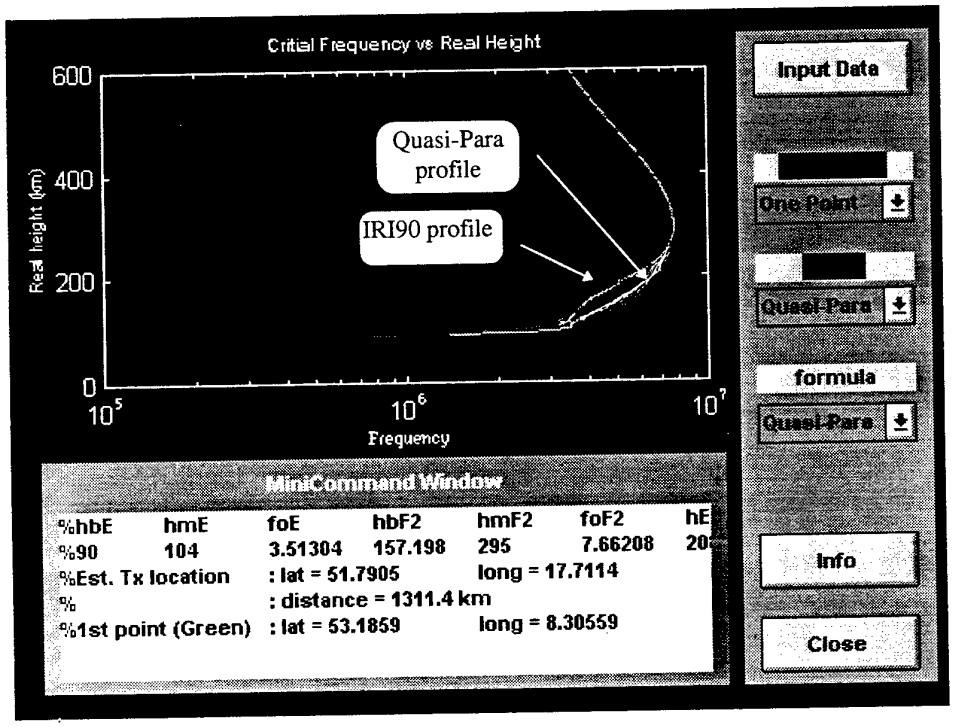


Figure 2 A QP profile plotted together with the IRI90 profile

variables, a simplex in n-dimensional space is characterised by $n+1$ distinct vectors which are its vertices e.g. in two-space, a simplex is a triangle and in three-space a pyramid. Using the appropriate analytical equations, an analytical function can be determined which will give the propagation distance in terms of the chosen variable ionospheric parameters of the chosen ionospheric model, the elevation angle α and transmission frequency f in the form, say,

D=F(foE, foF2, hmF2, ymF2, f,α).

Then the following function:-

| F(foE, foF2, hmF2, ymF2, f, α_1) - 2xF(foE, foF2, hmF2, ymF2, f, α_2) |

will have a minimum for the distance to the unknown transmitter in the case of the ionosphere parameters assumed the same at both the 1 and 2 hop reflection points. Thus minimising this function will determine an updated "best fit" ionosphere model. If the ionosphere is considered to vary between the one and two hop reflection points, then the following function is minimised:-

$| F(x_1, x_2, ... x_n, f, \alpha_1, \beta) - F_2(x_1, x_2, ... x_n, f, \alpha_2, \beta) |$

where F_1 ( )and F_2 ( ) are functions giving the 1 and 2 hop ranges respectively in terms of the ionosphere parameters. to be optimised $x_1, x_2, ... x_n$ etc which may include ionisation gradients, the transmission frequency f, elevation angles α_i. and azimuth angle β.

There is also a choice of including one or more reference paths to known transmitters (see section 5). In this case, the frequency, propagation distance and measured elevation at the receiver of each are also input using an additional input data box (see figure 3). Optimisation is then performed to find the same parameters of the ionospheric model but using the angle(s) of incidence of the unknown target transmitter together with the angles of incidence of the known transmitters for which the known transmitter-receiver distance acts as an additional constraint

The propagation distance can finally be determined from the optimised ionospheric parameters and this, together with the azimuthal bearing and known receiver location, used to find the unknown transmitter location. These are then printed out in the MiniCommand Window and the ionospheric profile, based on the chosen analytical model and optimised ionospheric parameters, also plotted (see figure 4).

5. USING ADDITIONAL INFORMATION

It is clear that for the basic method the number of variables optimised exceeds the number of constraints in the system so that where all four initial parameters differ substantially from their real values, good convergence cannot always be expected. Additional constraints can be included in the system by (i) including information on paths to known transmitters adjacent to that to the unknown transmitter (mentioned above) and/or (ii) using information on the comparative time delay of the different propagational modes. For (i), the optimisation is performed to find the same parameters of the ionospheric model but using the angle(s) of incidence of the unknown transmitter together with the angles of incidence $\alpha_{k,i}$ of the known transmitters of frequency $f_{k,i}$ for which the known transmitter-receiver distance D_i acts as an additional constraint. Then, for example, the function to be minimised could be of the form:-

| F(foE, foF2, hmF2, ymF2, f, α_1) - 2xF(foE, foF2, hmF2, ymF2, f, α_2) | +| Σ F(foE, foF2, hmF2, ymF2, $f_{k,i}$,$\alpha_{k,i}$) - D_i |

where the summation is taken over all i known transmitters. Different weightings could be applied to the first term and each of the i terms corresponding to known transmitters if some measurements were considered more accurate than others. The inclusion of this extra information can significantly improve the distance estimate. More successful optimisation can be achieved even when only one additional reference path is used. This method relies, however, on the ionosphere being correlated over the area of reflection of the different transmissions. Rush and Edwards [14] have determined ionospheric correlation distances in both North-South and East-West directions for different latitude ranges and seasons. For mid-latitudes, correlation is generally very good (coefficient > 0.8) for points separated up to about 1000 km There are likely to be many known transmitters within this range, suggesting the practicality of this method. The correlation measured is, however, between hourly deviations from the monthly mean value at the respective points. Thus a more accurate method would be to include small offsets to the ionospheric parameters in the equations for the known transmitters so that for example, foF2′ is taken as foF2+ε where fof2 is the variable being optimised and ε is the difference in the ionospheric model between foF2 at the midpoint of the path to the known transmitter and foF2 at the estimated midpoint of the path to the unknown transmitter.

6. ACCURACY

Although ideally the method should be tested using real data consisting of experimentally measured angles of arrival, tests were first made using modeled data. There were three reasons for this. The first was that errors in the range-finding method itself would be unambiguously determined. For measured angles of arrival, errors in determined distance would be due to both error in the measured angles of elevation as well as any errors inherent in the method. It is clearly important when optimising the method to achieve greatest accuracy that there are no errors in input angles. Secondly, using modeled data, the effect of random errors in all the ionospheric parameters e.g. foE, foF2, hmF2 could be easily determined. Thirdly, it was considered that for an operational system, angles of arrival could be determined more accurately than using the

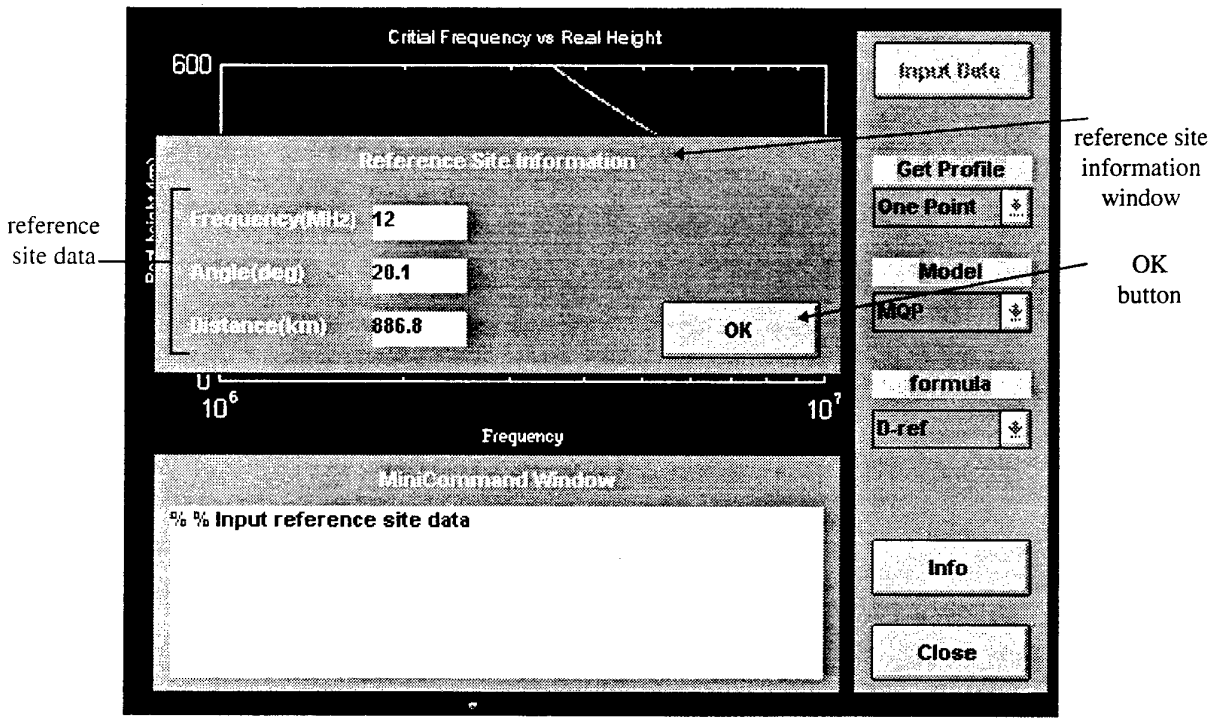


Figure 3 Reference Site Information Windows

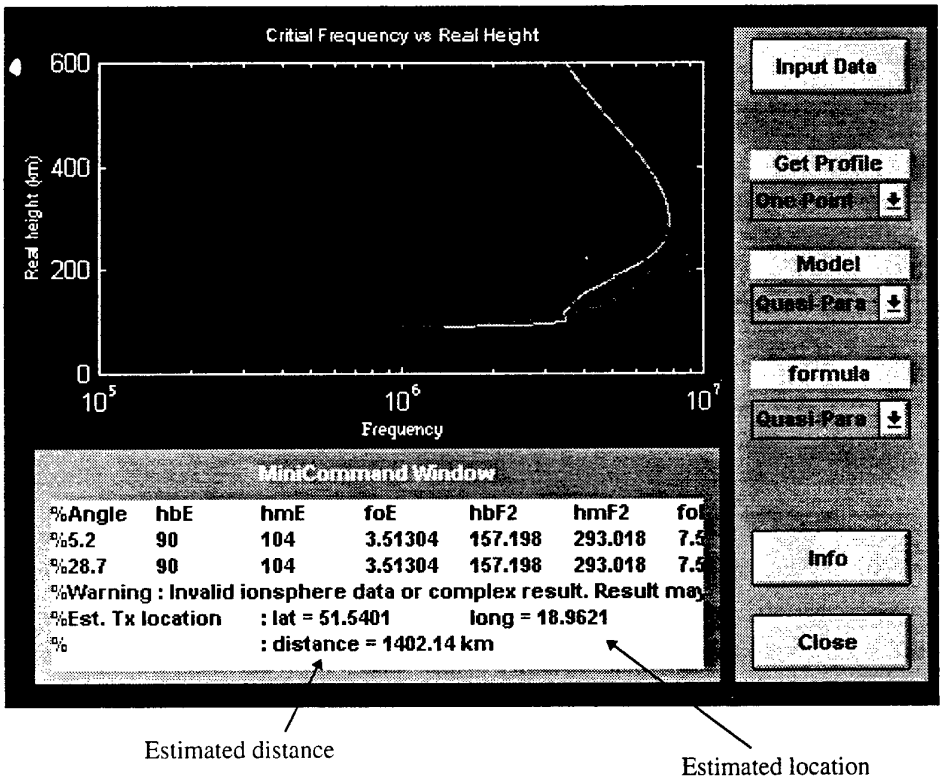


Figure 4 Final Screen with Results

very limited budget equipment built by ourselves and employed in our trials.

Tests against modeled data were accomplished by using a random number generator to add a random amount (in the range ± 15%) to each of the initial values of the ionospheric parameters e.g. foF2, hmF2 that were optimised in the model. The amount of variation was considered commensurate with day to day variation in ionospheric parameters determined by [13,14]. Tests were performed for different path combinations such as 1F2 and 2F2. It was found, for example, for a 560.5 km propagation path (see fig.5) that the error in the determination of the distance to the target transmitter using the MQP model was less than 2% when there was one added reference path (which was 1410 km).

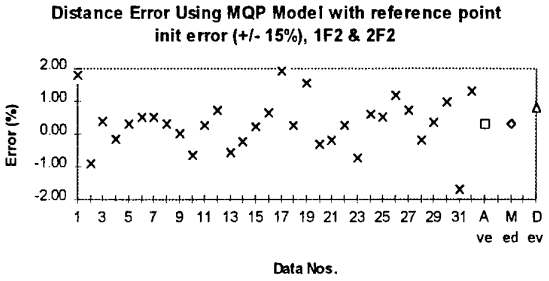


Figure 5 Distance error for 1F2 and 2F2 560.5 km path with an additional reference transmitter added.

The small errors shows the effectiveness of the method although more trials on real data are necessary. The problem with trials using real data is that it impossible to separate errors in determining the elevation angles using the super-resolution D/F technique with errors in distance introduced by the SSL optimisation process. It is considered that the method would work optimally when elevation angles are determined to about 0.1°accuracy. Best mode separation and bearing accuracy can be achieved in the presence of multipath when employing (i) an array long compared with the wavelength of the received transmission (ii) a 3D array using linear active antenna well above the ground and (iii) MLE based super-resolution methods [4--8] which can best cope with the coherent sources that arise in multipath..

7. MAGNETO-IONIC MODES AND TILTS

The optimisation can be performed for a more realistic 3D ionosphere model and including the effect of the geomagnetic field if numerical ray-tracing rather than analytical formulae are used in the optimisation process. This would enable the two magneto-ionic modes to be treated. For a 3D ionosphere model, optimisation could also be performed for additional parameters such as a horizontal gradient of electron density. The use of a numerical ray-tracing method does, however, greatly increase the time required for the solution.

Chirp soundings can also be made use of if angles of elevation at the receiver are measured. The chirp sounder can be considered as a number of co-located reference transmitters of differing frequencies of which each one would provide an additional term in the function that is minimised.

Since analytical formula exist for the time of propagation as well as the distance, the measured difference in time of propagation between two modes can be used as an additional constraint. The absolute time of propagation is not required. The time difference can be measured for a modulated transmitter by performing an autocorrelation of the received baseband signal. The accuracy of the measured time delay will of course depend on the bandwidth of the modulation but can be improved by employing superresolution techniques in the time domain. If $T = F(foE, foF2, hmF2, ymF2, f, \alpha)$ gives the time of propagation and Δt is the measured difference in propagation time between say 1 and 2 hop modes, then the following function should be minimised:-

$$|F(foE, foF2, hmF2, ymF2, f, \alpha_1) - 2xF(foE, foF2, hmF2, ymF2, f, \alpha_2)| + |T(foE, foF2, hmF2, ymF2, f, \alpha_2) - T(foE, foF2, hmF2, ymF2, f, \alpha_1) - \Delta t|$$

Different weightings could also be applied to the two terms reflecting different accuracies in angle and time measurement.

8. CONCLUSIONS

A new SSL method for HF transmitters has been described which employs a Nelder Mead minimisation of a multi-variable function which can include data from a variety of sources. Thus, the final determination of the location of the target transmitter can be based on data from the following sensors:-

1. Multi-element receiving antenna array
2. Vertical and/or oblique ionospheric sounders
3. Passive reception of known HF transmitters
4. Chirp sounders
5. Relative time of arrival of multimode components

An optimum solution for both an updated real-time ionosphere model as well as target transmitter location is obtained which is a single solution employing all the measured data.

The SSL determination is performed under MATLAB via a menu-driven graphical user interface. This employs pull-down menus for data input, program choice, ionospheric profile plotting and presentation of results.

9. REFERENCES

1. Nelder J.A. and Mead R., "A simplex method for function minimisation", Computer Journal, vol. 7, pp.308-313, 1965.

2. McNamara L.F., "Ionospheric modeling in support of single station location of long range transmitters", J. atmos. terr. Phys., 50, 781-795, 1988.

3. Bertel L., Brousseau C, Ferreol A., Rogier J.L and Demeure C., "Electron density profile updating from arrival angle measurements on HF links", Proceedings and program for the 7th international ionospheric effects symposium, 4-6 May, 1993.

4. Zatman, M.A., Strangeways H.J. and Warrington E.M., "Resolution of multimoded HF transmissions using the DOSE superresolution direction finding algorithm", Proc. ICAP '93, Conf. Publ. No. 370, part 1, 414-418. 1993.

5. M.A.Zatman and H.J. Strangeways, Bearing measurements on multi-modeded signals using single snapshot data, Proc. of IEEE AP-S International Symposium, Washington, Seattle, Vol.4, pp.1934-1937, 1994.

6. Zatman M.A. and Strangeways H.J., "The effect of the covariance matrix of ionospherically reflected signals on the choice of direction-finding algorithm", Proc. of the 6th International Symposium on HF Radio Systems and Techniques, York, U.K., pp.267-272, 1994.

7. Zatman M.A. and Strangeways H.J, "Super-resolution DF Algorithms for the Characterisation of Multi-moded HF signals", in " Multiple Mechanism Propagation Paths (MMPPs), their Characterisation and Influence on System Design" pp. 41/1- 41/10, AGARD conference proceedings 543, AGARD,Neuilly sur Seine, France, 1994.

8 . Strangeways H.J., and Zatman M.A., "Experimental observations using superresolution df and propagation path determination of additional off great circle paths due to the terminator.", Proc. ICAP '95, Conf. Publ. No. 407, vol.2, 69-72, 1995.

9. W. Featherstone, H.J. Strangeways, M.Darnell and H. Mewes, Application of superresolution direction-finding to a variety of channel conditions in the HF band, IEE 7th International Conference on HF Radio Systems and Techniques, pp.306-310, 1997.

10. Norman R.J. and Cannon P.S., "A two-dimensional analytic ray-tracing technique accommodating horizontal gradients", Radio Science, Vol..32, 387-396, 1997.

11. Baker, D. C., and Lambert, S., "Range estimation for SSL HFDF systems by means of a multiquasiparabolic ionosphere model", IEE Proc. Pt.H., Vol. 136 pp. 120-125, 1988.

12. Croft, T. A., and Hoogasian, H., "Exact ray calculations in a quasi-parabolic ionosphere with no magnetic field", Radio Science, Vol. 3, pp. 69-74.1968.

12. Milson, J. D., "Exact ray path calculations in a modified Bradley/Dudeney model ionosphere", IEE Proc. Pt. H, vol.132, pp.33-38, 1985.

13. Strangeways H.J. and Sow S.M. , " HF Transmitter SSL (Single Site Location) with Combined Ionospheric Update", Proc. ICAP '97, 1997.

13. Rush C.M., Miller D. and Gibbs J., " The relative daily variability of foF2 and hmF2 and their implications for HF radio propagation. ", Radio Science, vol.9, pp.749-756, 1974.

14. Rush C.M. and Edwards W.R. Jr, "An automated technique for representing the hourly behaviour of the ionosphere", Radio Science, 11, pp.931-937, 1976.

PAPER No: 22

DISCUSSOR'S NAME: P. Cannon

COMMENT/QUESTION:

You have used MQP and have shown that your errors are small. MQP does not allow for spatial variations along the path. Could you please comment?

AUTHOR/PRESENTER'S REPLY:

The figures presented were tto show errors due to departure of real-time ionosphere parameters from parameters determined from an ionosphere model with up to 15% difference determined using a random number generator independently for each parameter. These calculations are for an ionosphere with no horizontal gradients. In the written paper I mention how horizontal gradients could also be included but requires either numerical ray-tracing using e.g. Jones 30 program or the segmented MQP mdoel which my questioner has devised with R.J. Norman (my reference 10).

DISCUSSOR'S NAME: C. Goutelard

COMMENT/QUESTION:

When you introduce, as you indicate, the horizontal components of the ionisation gradients, are you sure that you will always obtain a unique solution for your optimisation.

AUTHOR/PRESENTER'S REPLY:

The simulations presented in my talk were for an ionosphere with no horizontal gradients. However, the method can be applied for such an ionosphere (as detailed in the written paper) by employing, either a ray-tracing program utilising numerical integration (e.g. Jones 30) or a spatially segmented MQP (see my reference 10), in each determination of the function that is minimised. I plan to do further trials to investigate this. In general the problem of a non-unique solution can be circumvented by making sure that the initial values chosen for parameters in the optimisation are close to their real values.

PAPER No. 22

DISCUSSOR'S NAME: G. S. Brown

COMMENT/QUESTION:

The results of your simulation seem to contain a bias in them; have you run enough simulations to see what the histogram of the error looks like?

AUTHOR/PRESENTER'S REPLY:

The number of sets of random errors presented in each figure was 32. I think that this is too small a number to be sure that the overall percentage error in distance would be expected to be zero. However, I intend to investigate the possible existence of any systematic errors by conducting similar trials with larger numbers of data sets including random errors.

SMART ANTENNA ARRAY PROCESSING FOR SPREAD SPECTRUM WIRELESS COMMUNICATION SYSTEMS

Y. F. Huang, S. Kapoor, S. Gollamudi and S. Nagaraj

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556 *

ABSTRACT

This paper examines the problem of incorporating smart antennas in wireless DS-CDMA systems. The use of array signal processing in DS-CDMA systems offers the possibility of jointly utilizing code and spatial resources to achieve superior performance compared to conventional multiple-access methods. The development of interference suppression techniques is essential to take full advantage of the available radio resources and requires data fusion in the spatial and code domains. Spatial signal processing can also lead to faster and robust synchronization and provide the capability to rapidly adapt to channel fading, co-channel and adjacent channel interferers. The overall method is suitable for real-time implementation and can lead to a direct increase in the overall capacity, quality and coverage of these systems.

1. INTRODUCTION

The use of spread spectrum multiple access techniques is well established in the wireless communications arena. In particular, *Direct Sequence Code Division Multiple Access* (DS-CDMA) has been widely studied in the literature and has been implemented in several commercial systems as well. Adaptive interference suppression techniques based on multiuser detection and antenna array processing have recently been considered as powerful methods for increasing the quality, capacity and coverage of these systems. They provide a superior, though computationally more expensive, alternative to conventional single sensor (matched filtered) detection which is severely limited by multi-access interference.

This paper focuses on co-channel interference (CCI) mitigation techniques for the uplink channel of DS-CDMA systems over frequency selective slowly fading channels. Multiuser detection [11, 16, 24] and adaptive array processing [2, 18, 22, 25] have been shown to be promising solutions to this problem. By carrying out multi-sensor data fusion using an antenna array, significant CCI reduction can be obtained. Furthermore, this approach is particularly effective when data fusion is done adaptively to combat time-varying channels and interference. Towards this end, we present here a novel adaptive array receiver structure which adopts a two-pronged approach to CCI suppression using multiuser detection and beamforming. Furthermore, a RAKE structure is retained for overcoming frequency selective fading. Thus, the overall receiver makes a combined use of frequency diversity (through the use of RAKE combining), code (or time) diversity and angle diversity (through the use of beamforming).

The principles of multiuser detection are by now well established [11, 24]. The optimal detector [24] is a major theoretical milestone but has limited practical utility due to its exponential complexity in the number of users. Decorrelating detectors have enjoyed much popularity due to their near-far resistant performance and much reduced linear complexity. A minimum energy formulation for blind adaptive near-far resistant multiuser detection was introduced in [11] and also used in [22]. While the MMSE (and the equivalent minimum energy) solution is known to be near-far resistant, a major obstacle lies in adaptively seeking the optimum solutions. Stochastic gradient algorithms such as the Least Mean Squares (LMS) method are very attractive because of their simplicity but suffer from poor convergence due to the large dimensionality in typical CDMA systems and the fluctuation in received signal powers [16]. Least squares (LS) algorithms on the other hand, provide adequate convergence speed under these circumstances, but can be prohibitively complex to implement. Note that most low complexity fast-RLS type algorithms are not applicable due to the lack of time-shift structure in the input vectors [10]. Thus, there is a clear imperative to develop reduced complexity multiuser detection schemes without having to sacrifice performance. The major difference in the adaptive multiuser detector presented in this paper is that it utilizes a novel recursive update algorithm based on set-membership parameter estimation theory which provides significant benefits in terms of tracking, convergence, and complexity.

Adaptive arrays are typically designed for either utilizing spatial diversity for mitigating the effect of fading or for using the inherent angle diversity in the received signals for CCI reduction. In [15], an adaptive array receiver configured as the cascade of a beamformer and matched filter detector is presented. The update of the beamformer weights in this structure is rather cumbersome, especially over fading channels. A receiver structure with a matched

*This work was supported in part by the National Science Foundation under Grant MIP-9705173, the Center for Applied Mathematics, University of Notre Dame, and Tellabs Research Center, Mishawaka, IN.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

filter detector followed by a beamformer is presented in [18]. However, the DOA estimation algorithm used for beamforming is computationally complex (it is based on eigen-decompositions) and is sensitive to angle spread. An extension of this receiver structure, called a 2D-RAKE is also presented. The 2D-RAKE is designed to operate over multipath channels and consists of a front-end beamformer which feeds the beam-steered signals into a standard RAKE receiver structure. Joint spatio-temporal reception, akin to broadband beamforming [23], has also been considered [2, 16].

2. CDMA SIGNAL MODEL AND PROBLEM FORMULATION

Consider an asynchronous DS-CDMA system with K users. Such a system may be used for a cellular wireless network with the K users distributed in a cell of interest. The centralized receiver at a cellular base station receives asynchronous transmissions from all active users on the uplink (mobile to base) multiple access channel and interfering signals from adjacent cells as well. Let the information symbol sequence from the kth user be denoted by $A_k(\cdot)$, chosen from a possibly complex alphabet. Assuming the symbol and chip duration to be T_s and T_c, respectively, resulting in a nominal processing gain $N = T_s/T_c$, the spread spectrum signal emanating from the kth user is given by

$$v_k(t) = \sum_n \sqrt{P_k} A_k(\lfloor \frac{n}{N} \rfloor) c_k(n\%N) \delta(t - nT_c) \quad (1)$$

where n is the chip index; % denotes the modulus operation and $\lfloor\rfloor$ denotes the floor operation; P_k is the transmitted power of the kth user; $c_k(n\%N)$ denotes the nth chip of the kth users' periodic spread spectrum sequence with period N. After performing baseband pulse shaping with a filter $\psi(t)$, the transmitted waveform is given by

$$x_k(t) = \sum_n \sqrt{P_k} A_k(\lfloor \frac{n}{N} \rfloor) c_k(n\%N) \psi(t - nT_c) \quad (2)$$

The pulse $\psi(t)$ is assumed to have unit energy and duration T_c. Assume that the transmitted signal from the kth user is received at an antenna array receiver with M elements. A L-path frequency selective slowly fading model is assumed in order to formulate the spatio-temporal impulse response of the mth sensor to the signals from the kth user (see Appendix A) according to

$$h_k^{(m)}(t) = \sum_{l=0}^{L-1} a_{kl}^{(m)} \delta(t - \tau_{kl} T_c) \quad (3)$$

where $a_{kl}^{(m)}$ denotes the composite response of the mth sensor to the lth multipath component from the kth user; $\tau_{kl} T_c$ is the time delay of the lth multipath component of the kth user's received signal and that the maximum delay is $\tau_{k(L-1)} T_c < T_s$. Thus, the received complex baseband signal from the kth user at the mth element is given by the convolution

$$\begin{aligned} r_k^{(m)}(t) &= x_k(t) * h_k^{(m)}(t) = \sum_{l=0}^{L-1} a_{kl}^{(m)} x_k(t - \tau_{kl} T_c) \\ &= \sum_n \sqrt{P_k} A_k(\lfloor \frac{n}{N} \rfloor) c_k(n\%N) \\ &\quad \sum_{l=0}^{L-1} a_{kl}^{(m)} \psi(t - nT_c - \tau_{kl} T_c) \end{aligned} \quad (4)$$

The received signal from K asynchronous users at the centralized receiver is obtained as the superposition of each user's signals according to

$$r^{(m)}(t) = \sum_{k=0}^{K-1} r_k^{(m)}(t - \nu_k) + n^{(m)}(t) \quad (5)$$

where ν_k is the flat propagation delay of the kth user and $n^{(m)}(t)$ denotes the mth sensor's front-end additive noise. The receiver structure used in Figure 8 is used to recover transmitted symbols from all desired users. For a particular desired user, the composite signal is chip match filtered and fed into multiple RAKE arms, each delayed by a chip time or more. It is assumed that the receiver uses conventional techniques for determining which RAKE arms contain delayed copies of the signal [18]. Also, RAKE receiver arms for user k are synchronized to the path delays $\tau_{kl} T_c$. Without loss of generality, chip synchronous reception is assumed. Thus, the chip rate samples for the kth user with delay $\tau_{kl} T_c$ are obtained at the mth sensor after demodulation as

$$\hat{v}_{kl}^{(m)}(n) = r^{(m)}(t) * \psi(-(t - \tau_{kl} T_c - \nu_k)) \mid_{t=nT_c} \quad (6)$$

Denoting the sequence of received samples of dimension N spanning one symbol as $\hat{\mathbf{v}}_{kl}^{(m)}$, the chip sample vector for the pth symbol of the desired user, *i.e.*, for $n \in [pN, pN+N-1]$, is given by

$$\begin{aligned} \hat{\mathbf{v}}_{kl}^{(m)}(p) &= \sqrt{P_k} a_{kl}^{(m)} A_k(p) \mathbf{c}_k \\ &\quad + \sum_{r \neq l} \sqrt{P_r} a_{kr}^{(m)} [A_{k,0} \mathbf{c}_{kr,0} + A_{k,-1} \mathbf{c}_{kr,-1}] \\ &\quad + \sum_{h \neq k} \sqrt{P_h} \sum_{l=0}^{L-1} a_{hl}^{(m)} \\ &\quad [A_{h,0} \mathbf{c}_{hl,0} + A_{h,-1} \mathbf{c}_{hl,-1}] + \tilde{\mathbf{n}}^{(m)}(p) \end{aligned} \quad (7)$$

where

$$\begin{aligned} \mathbf{c}_k &= [c_k(0), c_k(1), \ldots, c_k(N-1)]^T \\ \mathbf{c}_{kl,0} &= [0, \ldots, 0, c_k(0), \ldots, c_k(N - \tau_{kl} T_c - 1)]^T \\ \mathbf{c}_{kl,-1} &= [c_k(N - \tau_{kl} T_c), \ldots, c_k(N-1), 0, \ldots, 0]^T \end{aligned} \quad (8)$$

and $A_{k,0}$ and $A_{k,-1}$ denote the two overlapping symbols; $\tilde{\mathbf{n}}^{(m)}(p)$ is the vector of filtered noise samples during the pth symbol. Thus, in addition to the filtered background noise, each sample of the received chip vector for the pth symbol has two interfering components. These arise from the multipath components of the same user with different time delays and from all other users. In addition, each interferer (self-multipath as well as from other users) contributes two independent interference vectors to the received sample vector in each symbol time. The kth users' symbols are

extracted using a linear detector $\mathbf{w}_k$, characterized by the discrete time inner product with the sampled chip sequence as,

$$e_{kl}^{(m)}(p) = \mathbf{w}_k^H \hat{\mathbf{v}}_{kl}^{(m)}(p) \tag{9}$$

As can be seen in Figure 8, a linear multiuser detector is used for each RAKE branch at all array elements. The computation and adaptive update of these linear detectors is now addressed.

3. BLIND MULTIUSER DETECTION

The goal here is to blindly compute linear decorrelation weight vectors to preserve the desired signal and mitigate interference, see *e.g.*, [11, 22] for prior work. For the special case that the receiver treats multiple access interference as additive white noise (or when detector adaptivity is not feasible) the linear detector is merely the signature sequence of the desired user. The following algorithm is presented following the framework of the well known *Generalized Sidelobe Canceler* (GSC) [23]. Constraining the detector to present a unit response to the desired user's signature sequence, and decomposing $\mathbf{w}_k$ into corresponding constrained and unconstrained components denoted by $\mathbf{w}_{k,q}$ and $\mathbf{w}_{k,a}$ respectively, we have

$$\mathbf{w}_k = \mathbf{w}_{k,q} - \mathbf{C}_{k,n}\mathbf{w}_{k,a} \tag{10}$$

where,

$$\mathbf{w}_{k,q} = \mathbf{C}_k(\mathbf{C}_k^H\mathbf{C}_k)^{-1}g \tag{11}$$

with the constraint matrix $\mathbf{C}_k = \mathbf{c}_k$ and the output constraint $g = 1$. Thus, in this case, $\mathbf{w}_{k,q} = \mathbf{c}_k/N$. The columns of the $N \times (N-1)$ matrix $\mathbf{C}_{k,n}$ span the null space of $\mathbf{c}_k$. In general, of course, multiple constraints can be imposed on $\mathbf{w}_{k,q}$, for instance, to exploit the knowledge of signature sequences and timing of other interfering users. Such constraints appear as columns of $\mathbf{C}_k$ and result in a corresponding decrease in the dimensionality of $\mathbf{w}_{k,a}$. In the extreme case when all the columns of $\mathbf{C}_k$ are constrained, the detector ceases to be adaptive. $\mathbf{C}_{k,n}$ is readily obtained via one of many orthogonalizing procedures [23]. In fact, for the special case above, $\mathbf{C}_{k,n}$ can be precomputed offline for each desired signature sequence. The unconstrained component $\mathbf{w}_{k,a}$ represents the $N-1$ dimensional adaptive portion of $\mathbf{w}_k$. Thus, the output of the mth array element for the lth symbol is given by

$$e_{kl}^{(m)}(p) = \mathbf{w}_{k,q}^H\hat{\mathbf{v}}_{kl}^{(m)}(p) - \mathbf{w}_{k,a}^H\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(m)}(p) \tag{12}$$

Equation (12) can be viewed as a standard adaptive filtering problem with $\mathbf{w}_{k,q}^H\hat{\mathbf{v}}_{kl}^{(m)}(p)$ serving as the desired signal; $\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(m)}(p)$ serving as the input vector; $\mathbf{w}_{k,a}$ denoting the adaptive weight vector and $e_{kl}^{(m)}(p)$ denoting the estimation error.

3.1. Adaptive Solution Strategies

The conventional solution to the above is to perform an unconstrained optimization involving $\mathbf{w}_{k,a}$ which can be obtained via adaptive algorithms such as stochastic gradient descent or least squares [10]. A compelling solution to this problem is to consider a new approach based on *set-membership parameter estimation* theory. For excellent tutorial overviews of set-membership theory, see [3] and [6]. Other key papers detailing the structure, features, convergence and tracking properties, and signal processing applications of set-membership algorithms include [4, 5, 7] and the references therein.

Specifically, the focus of this paper is on a subset of set-membership techniques, namely the class of *optimal bounding ellipsoids* (OBE) algorithms. There are several features of OBE algorithms which render them attractive for the problem at hand. Experience has shown that OBE algorithms perform better than weighted recursive LS algorithms in tracking time varying parameters and in low signal-to-noise ratio (SNR) situations [4, 8, 13]. Furthermore, OBE algorithms are computationally efficient due to their *discerning update* property. They can also provide an explicit indication of any loss in tracking - a feature not possessed by point estimation algorithms. Although LS algorithms can also be equipped with such indicator functions [10], such a feature is an integral part of parameter estimation using OBE algorithms. As a point of common ground, the geometric centers of the bounding ellipsoids in OBE algorithms (which are usually taken as point estimates at any given time) are known to be weighted recursive LS estimates [6]. Simply stated, the *optimization* of the weighting (update) factors of data sets according to set-membership principles essentially leads to the discerning update property and superior convergence and tracking properties. The OBE algorithms may appear to have $\mathcal{O}(N^2)$ complexity from an inspection of the recursive update equations. However, their discerning update (or data selective) feature can be fruitfully exploited for significant reduction in complexity [5, 8]. In particular, $\mathcal{O}(N)$ implementations are described in [5, 9] and an efficient time-shared implementation between multiple users is described in [8]. Thus, the structure and properties of OBE algorithms can narrow the gap between performance and complexity which is encountered by several conventional adaptive filtering algorithms.

3.2. An OBE Algorithm

A recursive algorithm is now derived to estimate $\mathbf{w}_{k,a}$ based on set-membership principles using OBE. The idea here is to update the estimator such that the estimation error is constrained to lie within a specified performance bounds. The goal is to construct an OBE algorithm which attempts to seek solution vectors $\mathbf{w}_{k,a}$ that meet the following *specification* for all p,

$$\sum_{m=0}^{M-1} |\mathbf{w}_{k,q}^H\hat{\mathbf{v}}_{kl}^{(m)}(p) - \mathbf{w}_{k,a}^H\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(m)}(p)|^2 \leq \gamma_k^2 \tag{13}$$

where γ_k^2 is a specified constant corresponding to a desired performance level and may also be viewed as a design parameter. In this paper, a suitably selected constant γ_k^2 is assumed (see [6, 7] for general guidelines on selection of γ_k^2). Weight vectors which achieve (13) for all possible input sequences constitute the so called *feasible set*. The objective of the OBE methodology is to seek this feasible set by successively refining *membership sets* $\mathcal{W}_k(p)$ defined at time p

as

$$\mathcal{W}_k(p) = \bigcap_{r=0}^{p}\{\mathbf{w}\in\mathcal{C}^{N-1}:\sum_{m=0}^{M-1}|\mathbf{w}_{k,q}^H\hat{\mathbf{v}}_{kl}^{(m)}(p) - \mathbf{w}^H\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(m)}(p)|^2\le\gamma_k^2\} \quad (14)$$

However, the complexity in exactly computing $\mathcal{W}_k(p)$ is overwhelming even for small N and p. OBE algorithms circumvent this problem by recursively updating hyper-ellipsoids $\mathcal{E}_k(p)$ which tightly outerbound the membership sets $\mathcal{W}_k(p)$ for all p. It follows that $\mathcal{E}_k(p)$ also outerbounds the feasible set at all times, since the feasible set is a subset of $\mathcal{W}_k(p)$ for all p. Thus, a recursive formulation can be used to update $\mathcal{E}_k(p)$ with each incoming chip vector at the symbol rate. Equation (12) can be rewritten in matrix form for the array as,

$$\mathbf{e}_{kl}^*(p) = \mathbf{U}_{k,q}(p) - \mathbf{V}_{k,n}^H(p)\mathbf{w}_{k,a} \quad (15)$$

where

$$\begin{aligned}\mathbf{V}_{k,q}(p) &\triangleq [\hat{\mathbf{v}}_{kl}^{(0)}(p),\hat{\mathbf{v}}_{kl}^{(1)}(p),\dots\hat{\mathbf{v}}_{kl}^{(M-1)}(p)]\\ \mathbf{U}_{k,q}(p) &\triangleq \mathbf{V}_{k,q}^H(p)\mathbf{w}_{k,q} \quad \text{and}\\ \mathbf{V}_{k,n}(p) &\triangleq [\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(0)}(p),\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(1)}(p),\\ &\quad\dots,\mathbf{C}_{k,n}^H\hat{\mathbf{v}}_{kl}^{(M-1)}(p)] \quad (16)\end{aligned}$$

Then as per the set-membership framework, let the error specification for the kth user be,

$$||\mathbf{e}_{kl}(p)||^2 \le \gamma_k^2 \quad (17)$$

where $||\cdot||$ denotes the vector l_2 norm and γ_k^2 is an appropriately chosen constant. The selection of γ_k^2 is addressed later in this section. Define $\mathcal{S}_k(p)$ as a degenerate ellipsoid as

$$\mathcal{S}_k(p) = \{\mathbf{w}\in\mathcal{C}^{N-1}:||\mathbf{U}_{k,q}(p) - \mathbf{V}_{k,n}^H(p)\mathbf{w}||^2\le\gamma_k^2\} \quad (18)$$

Let the membership set at time $(p-1)$ be given by

$$\begin{aligned}\mathcal{E}_k(p-1) &= \{\mathbf{w}\in\mathcal{C}^{N-1}:[\mathbf{w}-\mathbf{w}_k(p-1)]^H\mathbf{P}_k^{-1}(p-1)\\ &\quad [\mathbf{w}-\mathbf{w}_k(p-1)]\le\sigma_k^2(p-1)\} \quad (19)\end{aligned}$$

where $\mathbf{P}_k(p-1)$ is a symmetric positive-definite matrix and $\mathbf{w}_k(p-1)$ is the center of the ellipsoid. An ellipsoid that contains $\mathcal{E}_k(p-1)\cap\mathcal{S}_k(p)$ is given by

$$\begin{aligned}\mathcal{E}_k(p) &= \{\mathbf{w}\in\mathcal{C}^{N-1}:[1-\lambda_k(p)]\\ &\quad [\mathbf{w}-\mathbf{w}_k(p-1)]^H\mathbf{P}_k^{-1}(p-1)\\ &\quad [\mathbf{w}-\mathbf{w}_k(p-1)]+\lambda_k(p)||\mathbf{U}_{k,q}(p)-\mathbf{V}_{k,n}^H\mathbf{w}||^2\\ &\quad \le[1-\lambda_k(p)]\sigma_k^2(p-1)+\lambda_k(p)\gamma_k^2\} \quad (20)\end{aligned}$$

where $\lambda_k(p)$ is a real number in $[0,1]$. It can now be shown that there exists a symmetric positive definite $\mathbf{P}_k(p)$ and a positive scalar $\sigma_k^2(p)$ such that

$$\begin{aligned}\mathcal{E}_k(p) &= \{\mathbf{w}\in\mathcal{C}^{N-1}:[\mathbf{w}-\mathbf{w}_k(p)]^H\mathbf{P}_k^{-1}(p)\\ &\quad [\mathbf{w}-\mathbf{w}_k(p)]\le\sigma_k^2(p)\} \quad (21)\end{aligned}$$

is a well defined ellipsoid.

Proposition 3.1. Consider the inequalities (18) and (19) above. Define

$$\begin{aligned}\mathbf{d}_k(p) &\triangleq \mathbf{U}_{k,q}(p) - \mathbf{V}_{k,n}^H(p)\mathbf{w}_k(p-1)\\ \mathbf{G}_k(p) &\triangleq \mathbf{V}_{k,n}^H(p)\mathbf{P}_k(p-1)\mathbf{V}_{k,n}(p)\\ \mathbf{Q}_k(p) &\triangleq [1-\lambda_k(p)]\mathbf{I}_M+\lambda_k(p)\mathbf{G}_k(p)\end{aligned}$$

The following recursive update equations may be obtained,

$$\begin{aligned}\mathbf{P}_k(p) &= \frac{1}{1-\lambda_k(p)}[\mathbf{P}_k(p-1)-\lambda_k(p)\\ &\quad \mathbf{P}_k(p-1)\mathbf{V}_{k,n}(p)\mathbf{Q}_k^{-1}(p)\mathbf{V}_{k,n}^H(p)\mathbf{P}_k(p-1)],\\ \mathbf{w}_k(p) &= \mathbf{w}_k(p-1)+\lambda_k(p)\mathbf{P}_k(p)\mathbf{V}_{k,n}(p)\mathbf{d}_k(p) \quad \text{and}\\ \sigma_k^2(p) &= [1-\lambda_k(p)]\sigma_k^2(p-1)+\lambda_k(p)\gamma_k^2\\ &\quad -[1-\lambda_k(p)]\lambda_k(p)\mathbf{d}_k^H(p)\mathbf{Q}_k^{-1}(p)\mathbf{d}_k(p) \quad \blacksquare\end{aligned}$$

The last three equations in Proposition 3.1 constitute the recursions of the OBE algorithm. In order to compute the optimal update factor the parameter $\sigma_k^2(p)$ is minimized. The parameter $\sigma_k^2(p)$ can be considered as a bound on the estimation error at the pth step and is closely related to other popular measures of optimization such as volume and trace of the bounding ellipsoid [4, 13]. Unlike these measures, minimization of $\sigma_k^2(p)$ lends itself to a very efficient test for innovation. A tight upper bound on $\sigma_k^2(p)$, denoted by $\sigma'^2(p)$, is given by

$$\begin{aligned}\sigma'^2(p) &= [1-\lambda_k(p)]\sigma_k^2(p-1)+\lambda_k(p)\gamma_k^2-[1-\lambda_k(p)]\\ &\quad \lambda_k(p)\frac{\mathbf{d}_k^H(p)\mathbf{d}_k(p)}{[1-\lambda_k(p)]+\lambda_k(p)g_k(p)} \quad (22)\end{aligned}$$

where $g_k(p)=||\mathbf{G}_k(p)||$. Denote the optimal $\lambda_k(p)$ by $\lambda_k^o(p)$ (which lies in $[0,\alpha_k]$ for some $\alpha_k<1$) and define the quantity

$$\beta_k(p)\triangleq\frac{\gamma_k^2-\sigma_k^2(p-1)}{\mathbf{d}_k^H(p)\mathbf{d}_k(p)} \quad (23)$$

Proposition 3.2. Minimization of $\sigma'^2(p)$ with respect to $\lambda_k(p)$, leads to the following update condition:

(1) if $\beta_k(p)\ge 1$, then $\lambda_k^o(p)=0$

(2) otherwise $\lambda_k^o(p)=\min(\alpha_k,\nu_k(p))$ where

$$\begin{aligned}\nu_k(p) &= \alpha_k \quad \text{if } \mathbf{d}_k^H(p)\mathbf{d}_k(p)=0;\\ &\quad \frac{(1-\beta_k(p))}{2} \quad \text{if } g_k(p)=1;\\ &\quad \frac{1}{1-g_k(p)}[1-\sqrt{\frac{g_k(p)}{1+\beta_k(p)[g_k(p)-1]}}];\\ &\quad \text{if } 1+\beta_k(p)[g_k(p)-1]>0;\\ &\quad \alpha_k \quad \text{if } 1+\beta_k(p)[g_k(p)-1]\le 0 \quad \blacksquare \quad (24)\end{aligned}$$

See [13] for proof of Propositions 3.1 and 3.2. This result is used for computing the optimal update parameter $\lambda_k^o(p)$ at each step. Equation (23) and Proposition 3.2 constitute the condition for data selectivity or no-update. At any time p, the center of the ellipsoid, $\mathbf{w}_k(p)$ is taken as an estimate of

the adaptive component of the detector in (10). To initialize the algorithm, $\mathcal{E}_k(0)$ is chosen such that

$$\mathcal{E}_k(0) = \{\mathbf{w} \in \mathcal{C}^{N-1} : ||\mathbf{w}||^2 \leq 1/\mu\}$$

where $\mu \ll 1$ resulting in $\mathcal{E}_k(0)$ being a suitably large initial hyper-sphere. In other words, the following initial values may be chosen

$$\mathbf{P}_k(0) = \mathbf{I}_{N-1}, \quad \mathbf{w}_k(0) = \mathbf{0} \quad \text{and} \quad \sigma_k^2(0) = 1/\mu \quad (25)$$

The next stage of processing in the receiver of Figure 8 entails the combining of the detector outputs corresponding to each RAKE branch from all array elements via a bank of beamformers.

4. DOA ESTIMATION AND BEAMFORMING

Carrying out DOA estimation prior to despreading can be a formidable task. Most subspace based methods are not applicable due to the large number of independent signals and typically small number of array elements. Post-detection DOA estimation, on the other hand, is a viable option. Several methods have been proposed in the literature, including subspace based methods such as Weighted Subspace Fitting (WSF), ESPRIT [20] and an eigendecomposition based method [18]. The structure of a DS-CDMA system, however, allows for a much simpler and robust approach. This section describes an approximate *Maximum Likelihood DOA Estimator* (DOA-MLE) used for beamforming. In addition to inheriting the desirable properties of MLEs, it also turns out to be simple and intuitive. The basic idea is to partition the array into groups of two consecutive sensors or doublets. The algorithm then exploits the fading correlation between closely spaced doublet elements to extract the spatially induced phase differences between the post-correlation complex baseband outputs. Each doublet operates independently to compute the DOA-MLE and the estimates from multiple doublets are then suitably combined.

4.1. Maximum Likelihood DOA Estimation

The relationship between angle of arrival, beamwidth of arriving signals and antenna spacing has been explored in [21] and the references therein. For an inter-element spacing of ρ and narrowband signal wavelength v, the fading experienced at adjacent sensors is almost perfectly correlated for small values of ρ/v (such as as 0.5 or less) and angle spread Δ (such as $\Delta < 10^o$). Thus, inter-element spacing ρ for each doublet is assumed to be such that the two sensors experience near identical fading. Consider sensor m and $(m+1)$ constituting a doublet. Using (9) and (10),

$$\begin{aligned} e_{kl}^{(m)}(p) &= \mathbf{w}_{k,q}^H \hat{\mathbf{v}}_{kl}^{(m)}(p) - \mathbf{w}_k^H(p)\mathbf{C}_{k,n}^H \hat{\mathbf{v}}_{kl}^{(m)}(p) \\ &= \sqrt{P_k} a_{kl}^{(m)} A_k(p) + \tilde{I}_{kl}^{(m)}(p) \end{aligned} \quad (26)$$

where $\mathbf{w}_k(p)$ denotes the updated weight vector from the recursions of Proposition 3.1; $I_{kl}^{(m)}(p)$ denotes the collective interference terms in (7) and $\tilde{I}_{kl}^{(m)}(p)$ is appropriately defined. Given P observations from time instants $(p-P+1)$ to p, the MLE of the spatial phase difference can be obtained straightforwardly (see Appendix B) as

$$\hat{\phi}_{kl}^{(m)}(p) = \tan^{-1}\left[\frac{\mathrm{Im}\left[\sum_{r=p-P+1}^{p} e_{kl}^{(m+1)}(r)(e_{kl}^{(m)}(r))^*\right]}{\mathrm{Re}\left[\sum_{r=p-P+1}^{p} e_{kl}^{(m+1)}(r)(e_{kl}^{(m)}(r))^*\right]}\right] \quad (27)$$

where $(e_{kl}^{(m)}(r))^*$ denotes the complex conjugate of $e_{kl}^{(m)}(r)$. At every time instant, each doublet contributes a DOA-MLE for each user being tracked. Since all array elements are used for beamforming, the DOA-MLEs from each doublet are combined to form a single estimate. The simplest method is to average the outputs from each doublet and other schemes may be readily conceived. For instance, a suitable selection criterion can be adopted for selecting the "best" DOA-MLE from among the doublets. In any case, denote the final DOA estimate at time p by $\hat{\phi}_{kl}(p)$. Let the lth beamformer weight vector be denoted by $\mathbf{b}_{kl}$ (see Figure 8); the data covariance matrix at the output of the lth RAKE branch by $\mathbf{R}_l$ and the steering vector for the kth users' lth path be given by

$$\mathbf{s}_{kl}(p) = [1, e^{j\hat{\phi}_{kl}^{(1)}(p)}, \ldots, e^{j\hat{\phi}_{kl}^{(M-1)}(p)}]^H \quad (28)$$

Adopting the classical *minimum variance distortionless response* (MVDR) criterion for computing the beamformer weights [23]

$$\mathbf{b}_{kl}(p) = \frac{\mathbf{R}_l^{-1}\mathbf{s}_{kl}(p)}{\mathbf{s}_{kl}^H(p)\mathbf{R}_l^{-1}\mathbf{s}_{kl}(p)} \quad (29)$$

The output of the lth beamformer is then given by

$$y_{kl}(p) = \mathbf{b}_{kl}^H(p)\mathbf{e}_{kl}(p) \quad (30)$$

$\mathbf{b}_{kl}(p)$ is typically block updated using (29) above at a rate commensurate with the available processing power while $\mathbf{R}_l$ is approximated by a suitably windowed a temporal average. Other beamforming methods such as the maximum SINR method [23] may also be used.

4.2. RAKE Reception

Consider now the final stage of the adaptive array receiver, just prior to the slicer. The $L \times 1$ vector output $\mathbf{y}_k(p) = [y_{k0}(p), y_{k1}(p), \ldots, y_{k(L-1)}(p)]^T$ of the bank of beamformers is fed into the RAKE combiner to obtain symbol estimates for each desired user. Denote the kth users' RAKE combining coefficients by $\mathbf{r}_k(p) = [r_{k0}(p), r_{k1}(p), \ldots, r_{k(L-1)}(p)]^T$. The output of the kth users' combiner is then given by

$$\hat{A}_k(p) = \mathbf{r}_k^H(p)\mathbf{y}_k(p) \quad (31)$$

There are several RAKE combining algorithms, the classical ones being maximal ratio, equal gain and selection combining [12]. Any one of these methods or other variations in the literature [25] may be used. Such methods typically hinge on the slowly fading assumption and make use of special pilot symbols or training sequences to update the RAKE combining coefficients. In a flat fading environment, the RAKE combining reduces to a single 1-tap equalizer yielding

$$\hat{A}_k(p) = r_{k0}(p)y_k(p) \quad (32)$$

The equalizer operates on the received complex baseband symbol before slicing by correcting the magnitude and phase of the single beamformer's symbol rate output.

5. SIMULATION RESULTS

Consider first the performance of the proposed receiver in a flat Rayleigh fading environment. Fading coefficients are generated using the standard Jakes model [12]. Figure 3 depicts the ensemble averaged signal-to-interference ratio (SIR) obtained using matched filtering and multiuser detection using OBE, standard LMS and RLS algorithms. The latter three algorithms are used for updating the GSC's $(N-1)\times 1$ adaptive weight vector component with a single signal preserving constraint. Coherent quadrature modulation is used with a processing gain of $N = 16$ with a single antenna element ($M = 1$) and a background SNR of 20 dB due to Additive White Gaussian Noise (AWGN). Similar curves are obtained for larger processing gains. There are a total of 11 active users including the desired user, each transmitting with equal power. Taking the processing gain into account, this results in an initial SIR value of about 2 dB representing a severe interference environment. The LMS update gain is fixed at 0.001 while the RLS forgetting factor is chosen to be 0.99. The error specification for OBE is set at $M\hat{P}_0 + \xi$ where $\hat{P}_0$ is an estimate of the desired user's power and ξ is design parameter chosen to be unity. The SIR at the pth symbol is calculated according to

$$SIR(p) = \frac{\sum_r |\mathbf{w}_r^H(p)\mathbf{v}_{sig,r}(p)|^2}{\sum_r |\mathbf{w}_r^H(p)\mathbf{v}_{int,r}(p)|^2} \tag{33}$$

where the ensemble average is carried out over 200 independent trials in each of which the chip sequences are generated randomly; $\mathbf{w}_r(p)$ denotes the detector weight vector in the nth trial at the pth symbol; $\mathbf{v}_{sig,r}(p)$ and $\mathbf{v}_{int,r}(p)$ denote the $N \times 1$ received signal and interference chip vectors in the nth trial at the pth symbol respectively. $\mathbf{w}$ is adapted using OBE, LMS and RLS or is fixed when the matched filter is used.

An antenna array with $M = 4$ elements is introduced in Figure 4 and the ensemble averaged mean squared error (MSE) at the receiver output is shown. It is assumed that channel estimates are available at the receiver to equalize the pre-slicer symbol rate samples. Some degradation in the achievable MSE is to be expected when the estimates are not perfect. MSE with a strong interferer introduced at the 450th symbol is depicted in Figure 5 using the RLS and OBE algorithms. The DOA's of all 11 users are randomly distributed in the range $[-60^o, +60^o]$, each with an anglespread of 5^o. An inter-sensor spacing of $\rho = v/8$ is used. Perfect correlation is assumed between the fading experienced at adjacent sensors. This is known to be a good approximation for the sensor spacing and angle spreads under consideration [21]. The DOA-MLE algorithm, (27) is used for DOA estimation and the beamformer weights are obtained using the MVDR criterion, (29).

Figures 6 and 7 show the performance of the DOA-MLE algorithm alone. The magnitude and root-mean-square (RMS) value of the estimation error under different post-detection SNR ratios are shown. Each data point is obtained by ensemble averaging over 1000 independent trials using a single sensor doublet with spacing $\rho = v/8$. In Figure 6, the angle spread Δ is held fixed at 0^0 while the number of symbols P used for forming the estimate is varied. In Figure 7, $P = 500$ while Δ is varied. The angle spread is assumed to arise from multipath sub-components uniformly distributed in the interval $[-\Delta/2, \Delta/2]$.

In the simulations results described above, it is observed that the OBE algorithm outperforms the RLS algorithm in terms of convergence and complexity. Most of the OBE updates are confined to first few hundred symbols and the update rate drops significantly after that. Overall, the percentage of updates is 20% or less in the simulations shown and far lower when the percentage of updates is computed over a larger number of symbols. This burst nature of updates has been fruitfully exploited for significant computational savings [8]. Relative to OBE and RLS, the LMS algorithm does not perform adequately - a finding consistent with those in the literature [16]. In situations where use of the LMS algorithm is not subject to choice, the above results are suggestive of using an variable step size LMS algorithm which can adapt to the time-varying interference conditions. The use an antenna array at the base station with even 2 or 4 elements is seen to be very beneficial for DS-CDMA reception, especially when coupled with multiuser detection.

6. CONCLUSIONS

This paper presented a new receiver structure for CCI suppression and fading compensation for CDMA signaling over frequency-flat or frequency-selective slowly fading channels. The CCI suppression and fading compensation are accomplished by combining multiuser detection, beamforming and RAKE reception in a single integrated receiver. Conventional RAKE reception is used to combat multipath fading while CCI suppression is carried out by multiuser detector and beamforming. $(N-1+M)$ adaptive weights are used for the multiuser detector and beamformer, where N is the spread-spectrum processing gain and M is the number of antenna array elements. The blind adaptive multiuser detector is formulated using a constrained energy minimization criterion and adaptation is carried out using a novel OBE algorithm. The OBE multiuser detector provides fast convergence and superior tracking relative to conventional adaptive algorithms such as LMS and RLS. Also, a simple and robust approximate Maximum Likelihood DOA estimator is presented for beamforming.

APPENDIX

Spatio-Temporal Impulse Response

The composite spatio-temporal impulse response of the channel and the mth sensor to the kth user's signals is given by

$$h_k^{(m)}(t) = \sum_{l=0}^{L-1} a_{kl}^{(m)} \delta(t - \tau_{kl} T_c) \tag{34}$$

where $a_{kl}^{(m)}$ is the response of the mth antenna element to the lth multipath component from the kth user. Each multipath component is received with an angle spread of Δ and is assumed to be distinct from all other paths of the same user. The angle spread arises due to a large number of rays

emanating from local scatterers in the vicinity of the transmitting source. Each RAKE branch sees a distinct fading signal with a certain angle spread. Thus,

$$a_{kl}^{(m)} = \sum_i p_{kli} e^{j(q_{kli}+\phi_{kli}^{(m)})} \tag{35}$$

where the summation is taken over all the sub-components; $\phi_{kli}^{(m)}$ denotes the angular deviation of the spatial angle of the ith sub-component of the lth multipath component and p_{kli} and q_{kli} denote the corresponding channel magnitude and phase response respectively. The spatial angle $\phi_{kli}^{(m)}$ is given by

$$\phi_{kli}^{(m)} = 2\pi(m\rho/v)\sin(\theta_{kl} + \delta\theta_{kli}) \tag{36}$$

where θ_{kl} denotes the kth users' nominal angle-of-arrival of lth multipath component and $\delta\theta_{kli}$ is the angular deviation of the ith sub-component's DOA. For small angle spread Δ,

$$\phi_{kli}^{(m)} \approx \phi_{kl}^{(m)} + \delta\phi_{kli}^{(m)} \tag{37}$$

where

$$\begin{aligned} \phi_{kl}^{(m)} &= 2\pi(m\rho/v)\sin(\theta_{kl}) \quad \text{and} \\ \delta\phi_{kli}^{(m)} &= 2\pi(m\rho/v)[\cos(\theta_{kl})\delta\theta_{kli} \\ &\quad -(1/2)\sin(\theta_{kl})\delta\theta_{kli}^2] \end{aligned} \tag{38}$$

Note that for point sources, $\delta\phi_{kli}^{(m)} = 0$. Using (35)-(38),

$$a_{kl}^{(m)} = e^{j\phi_{kl}^{(m)}} \sum_i p_{kli} e^{jq_{kli}} e^{j\delta\phi_{kli}^{(m)}} \tag{39}$$

$(\sum_i p_{kli} e^{jq_{kli}} e^{j\delta\phi_{kli}^{(m)}})$ is usually modeled as a complex Gaussian random variable since it is the summation of a large number of i.i.d. random variables constituting the channel attenuation for each multipath component.

DOA-MLE Derivation

Rewriting (26) in vector form using P samples for sensor m and $(m+1)$,

$$\mathbf{e}_{kl}^{(m)} = \sqrt{P_k}\mathbf{a}_{kl}^{(m)}\mathbf{A}_k + \tilde{\mathbf{I}}_{kl}^{(m)} \tag{40}$$

$$\mathbf{e}_{kl}^{(m+1)} = \sqrt{P_k}\mathbf{a}_{kl}^{(m+1)}\mathbf{A}_k + \tilde{\mathbf{I}}_{kl}^{(m+1)} \tag{41}$$

where $\mathbf{e}_{kl}^{(m)}$, $\mathbf{a}_{kl}^{(m)}$, $\mathbf{A}_k$ and $\tilde{\mathbf{I}}_{kl}^{(m)}$ denote vectors of length P of the respective temporal samples. Using (39),

$$a_{kl}^{(m+1)} = e^{j\phi_{kl}^{(1)}} a_{kl}^{(m)} + \epsilon_{kl}^{(m+1)} \tag{42}$$

where $\epsilon_{kl}^{(m)}$ denotes the residual difference between the channel attenuation at the mth sensor due to angle spreading. To obtain $\epsilon_{kl}^{(m)}$, a series expansion of (39) may be carried out for small $\delta\phi_{kli}^{(m)}$ according to

$$e^{j\delta\phi_{kli}^{(m)}} \approx 1 + j\delta\phi_{kli}^{(m)} \tag{43}$$

Thus, yielding

$$\epsilon_{kl}^{(m)} = e^{j\phi_{kl}^{(m)}} \sum_i (j\delta\phi_{kli}^{(1)}) p_{kli} e^{jq_{kli}} \tag{44}$$

Under isotropic scattering, $\epsilon_{kl}^{(m)}$ can be regarded as complex valued zero mean and Gaussian distributed. Denoting $\phi_{kl}^{(1)}$ by Φ for notational simplicity and the conditional probability density function of $\mathbf{e}_{kl}^{(m)}$ by $f_{\mathbf{e}}$, $\hat{\Phi}_{ML}$ is given by

$$\begin{aligned} \hat{\Phi}_{ML} &= \arg\max_{\Phi} f_{\mathbf{e}}(\mathbf{e}_{kl}^{(m+1)}, \mathbf{e}_{kl}^{(m)}/\Phi) \\ &= \arg\max_{\Phi} f_{\mathbf{e}}(\mathbf{e}_{kl}^{(m+1)}/\mathbf{e}_{kl}^{(m)}, \Phi) \end{aligned} \tag{45}$$

since $\mathbf{e}_{kl}^{(m)}$ is independent of Φ. Now, using (41) and (42),

$$\mathbf{e}_{kl}^{(m+1)} = \mathbf{e}_{kl}^{(m)} e^{j\Phi} - \tilde{\mathbf{I}}_{kl}^{(m)} e^{j\Phi} + \tilde{\mathbf{I}}_{kl}^{(m+1)} + \epsilon_{kl}^{(m+1)}\sqrt{P_k}\mathbf{A}_k \tag{46}$$

Conventional assumptions are now made on the post-detection interference vectors to enable use of a ML approach. Namely, they are assumed to be instances of a stationary, ergodic zero-mean complex valued Gaussian process. Thus, the mean of $\mathbf{e}_{kl}^{(m+1)}$ is $(\mathbf{e}_{kl}^{(m)} e^{j\Phi})$ and $f_{\mathbf{e}}$ can be factored into a product of partial densities. Thus, $\hat{\Phi}_{ML}$ can be obtained by maximizing the log likelihood function according to

$$\hat{\Phi}_{ML} = \arg\max_{\Phi} \left(\mathbf{e}_{kl}^{(m+1)} - \mathbf{e}_{kl}^{(m)} e^{j\Phi}\right)^H \left(\mathbf{e}_{kl}^{(m+1)} - \mathbf{e}_{kl}^{(m)} e^{j\Phi}\right) \tag{47}$$

Differentiating the right-hand side of (47) with respect to Φ and setting to zero yields

$$[\mathbf{e}_{kl}^{(m)}]^H \mathbf{e}_{kl}^{(m+1)} e^{-j\Phi} = [\mathbf{e}_{kl}^{(m+1)}]^H \mathbf{e}_{kl}^{(m)} e^{j\Phi} \tag{48}$$

Noting that the right-hand side of (48) is merely the complex conjugate of the left-hand side, and setting the imaginary part to zero yields the desired result in (27).

REFERENCES

[1] S. Anderson, M. Millnert, M. Viberg and B. Wahlberg, "An Adaptive Array for Mobile Communication Systems," *IEEE Trans. on Veh. Tech.*, Vol. 40, No. 1, pp. 230-236, February 1991.

[2] T. A. Brown, *The Use of Antenna Arrays in the Detection of Code Division Multiple Access Signals*, Ph.D. Thesis, University of Minnesota, 1995.

[3] P. L. Combettes, "The Foundations of Set-Theoretic Estimation," *Proc. IEEE*, Vol. 81, No. 2, pp. 182-208, 1993.

[4] S. Dasgupta and Y. F. Huang, "Asymptotically Convergent Modified Recursive Least Squares with Data Dependent Updating and forgetting factor for Systems with Bounded Noise," *IEEE Trans. on Info. Theory*, Vol. 33, No. 3, pp. 383-392, May 1987.

[5] J. R. Deller and S. F. Odeh, "Adaptive Set-Membership Identification in $\mathcal{O}(m)$ Time for Linear-in-Parameter Models," *IEEE Trans. on Signal Proc.*, Vol. 41, No. 5, pp. 1906-1924, May 1993.

[6] J. R. Deller, M. Nayeri and S. F. Odeh, "Least Square Identification with Error Bounds for Real-time Signal Processing and Control," *Proc. IEEE*, Vol. 81, pp. 813-849, June 1993.

[7] E. Fogel and Y. F. Huang, "On the value of Information in System Identification - Bounded Noise Case," *Automatica*, Vol. 18, No. 2, pp. 229-238, March 1982.

[8] S. Gollamudi, S. Kapoor, S. Nagaraj, and Y.F. Huang, "Set-Membership Adaptive Equalization and an Updator-Shared Implementation for Multiple Channel Communication Systems," Submitted to *IEEE Trans. on Signal Processing*, 1996.

[9] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.F. Huang, "Set-Membership Filtering and a Set-Membership Normalized LMS Algorithm with an Adaptive Step Size", submitted to *IEEE Signal Processing Letters*, April, 1997.

[10] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.

[11] M. L. Honig, U. Madhow, and S. Verdu, "Blind Adaptive Multiuser Detection," *IEEE Trans. on Info. Theory*, Vol. 41, No. 4, pp. 944-960, July 1995.

[12] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1994.

[13] S. Kapoor, S. Gollamudi, S. Nagaraj and Y. F. Huang, "Tracking of Time-Varying Parameters using Optimal Bounding Ellipsoid Algorithms," *Proc. 34th Ann. Allerton Conf. on Comm., Control and Comp.*, Monticello, Illinois, October 1996.

[14] S. Kapoor and Y. F. Huang, "Blind Multiuser Detection and Interference Cancellation in DS-CDMA Mobile Radio Systems," *Proc. IEEE Workshop on Signal Proc. Advances in Wireless Comm. (SPAWC'97)*, Paris, France, April 1997.

[15] R. Kohno et al., "Combination of an adaptive array antenna and a canceler of interference of direct-sequence spread-spectrum multiple-access system," *IEEE J. Sel. Areas Comm.*, Vol. 8, pp. 675-681, May 1990.

[16] U. Madhow, "Signal Processing for Interference Suppression in DS-CDMA Systems," *Proc. ICASSP'96*, Atlanta, GA, May 1996.

[17] R. L. Moses, T. Soderstrom and J. Sorelius, "Effects of Multipath Induced Angular Spread on DOA Estimators in Array Signal Processing," *Proc. IEEE/IEE Wksp. on Signal Proc. Methods in Multipath Environments*, Glasgow, Scotland, April 1995.

[18] A. F. Naguib and A. Paulraj, "Performance of CDMA Cellular Networks with Base Station Antenna Arrays," *Proc. Int. Zurich Seminar on Digital Comm.*, pp. 87-100, Zurich, Switzerland, March 1994.

[19] K. Pahlavan and A. Levesque, *Wireless Information Networks*, John Wiley, 1995.

[20] R. Roy and T. Kailath, "ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, Vol. 37, No. 7, pp. 984-995, July 1989.

[21] J. Salz and J. H. Winters, "Effect of Fading Correlation on Adaptive Arrays in Digital Mobile Radio," *IEEE Trans. on Veh. Tech.*, Vol. 43, No. 4, pp. 1049-1057, November 1994.

[22] J. B. Schodorf and D. B. Williams, "A Constrained Adaptive Diversity Combiner for Interference Suppression in CDMA systems," *Proc. ICASSP'96*, Atlanta, GA, May 1996.

[23] B. D. Van Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, Vol. 5, No. 2, pp. 4-24, April 1988.

[24] S. Verdu, "Multiuser Detection," In *Advances in Statistical Signal Processing*, Vol. 2, pp. 369-409, JAI Press, 1993.

[25] Z. Zvonar, "Combined Multiuser Detection and Diversity Reception for Wireless CDMA Systems, Proc. *IEEE Trans. on Veh. Tech.*, Vol. 45, No. 1, pp. 205-211, February 1996.

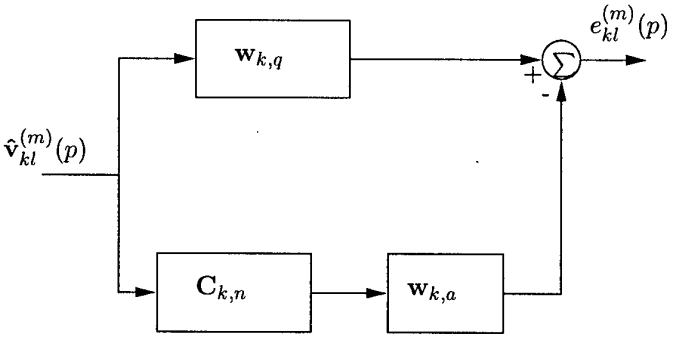


Figure 1. Adaptive Multiuser Detector structure in a Generalized Sidelobe Canceler (GSC) framework.

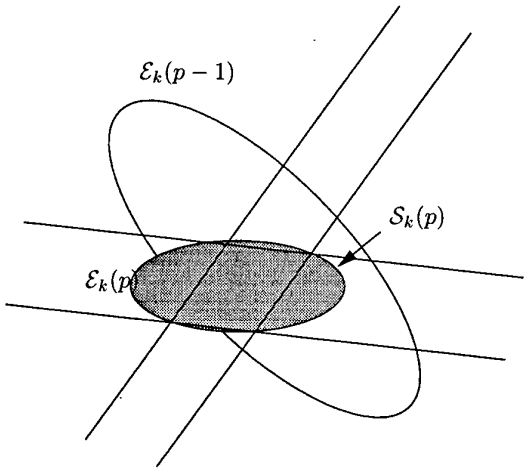


Figure 2. Schematic depicting operation of OBE recursions in two dimensions.

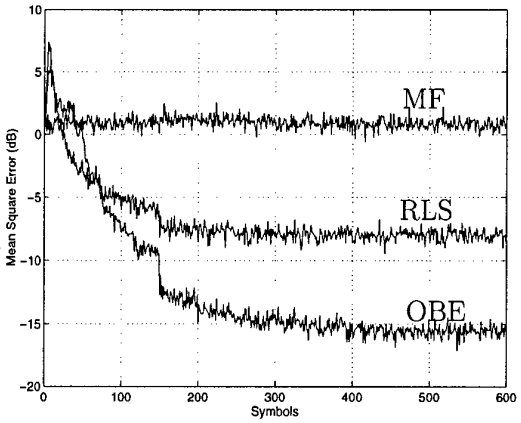


Figure 4. MSE using MF, OBE and RLS algorithms. Processing gain $N = 16$, $M = 4$.

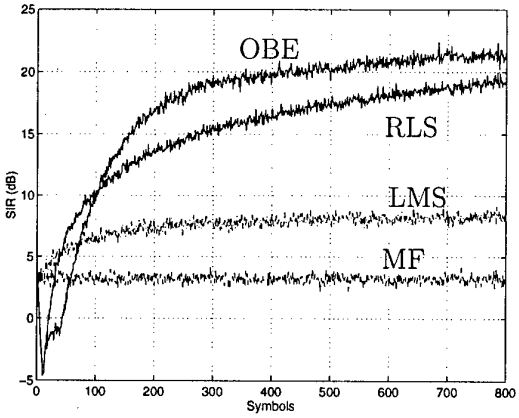


Figure 3. *SIR* using MF, OBE and RLS algorithms. Processing gain $N = 16$, $M = 1$.

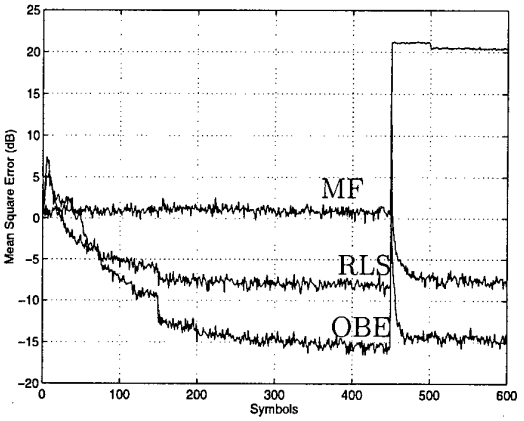


Figure 5. MSE using MF, OBE and RLS algorithms with strong interferer. Processing gain $N = 16$, $M = 4$.

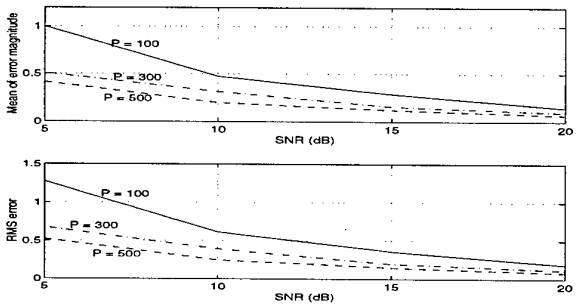


Figure 6. Magnitude and RMS value of DOA estimation error using DOA-MLE algorithm (in degrees): Number of symbols P versus SNR for angle spread $\Delta = 0$.

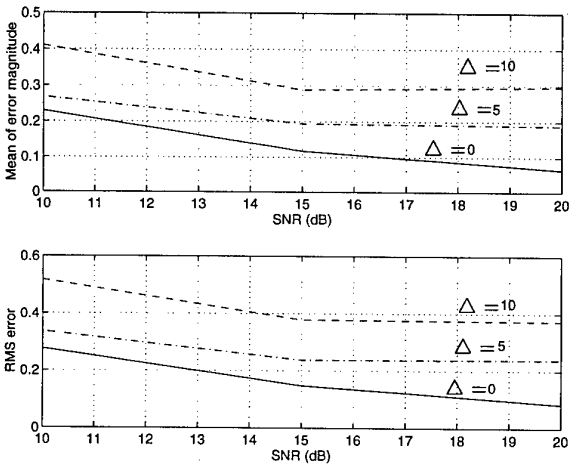


Figure 7. Magnitude and RMS value of DOA estimation error using DOA-MLE algorithm (in degrees): Angle spread Δ versus SNR for number of symbols $P = 500$.

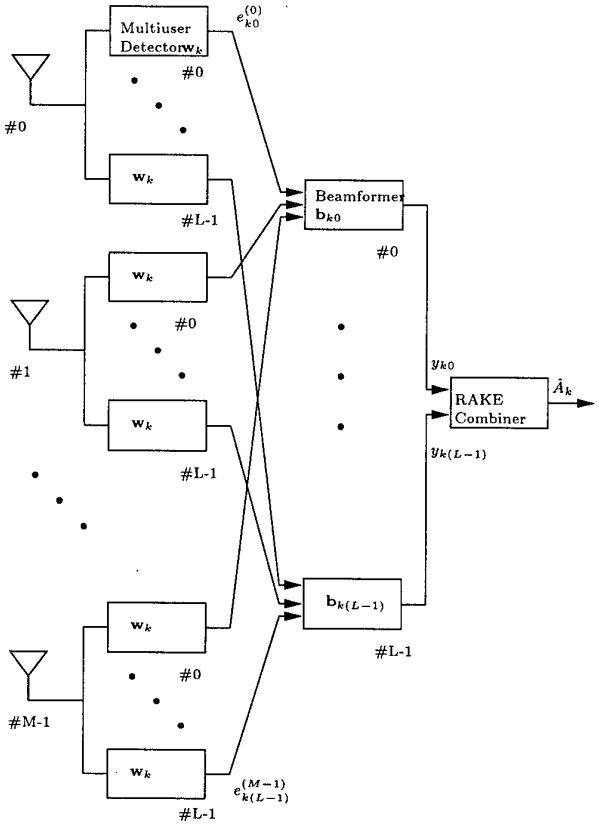


Figure 8. Receiver structure for L-path frequency selective fading channel for kth user.

Propagation Diversity Constraints on Reception of GPS Signals at High and Equatorial Latitudes

Jules Aarons and Bosheng Lin

Center for Space Physics
Boston University
725 Commonwealth Avenue
Boston, MA 02215 USA

1. SUMMARY

The coming of years of high solar flux, starting in 1999, will bring propagation problems to some users of GPS; we are concerned with ionospheric irregularities which produce phase and amplitude scintillation. The knowledge of when deteriorated signals will occur and their characteristics can help minimize their impact or at least allow users to understand that the fading phenomena are of natural origin rather than equipment related. At equatorial anomaly latitudes, fades of 20 dB at GPS frequencies have been observed in years of high solar flux; according to receiver experts these are enough to disturb any equipment. The sophistication of the equipment in moving through these signal problems is a function of hardware and complexity of software. Knowledge of diversity can help to minimize the effect of temporary loss of signal.

The equatorial region irregularity structures are dominated by patches of irregularities which are frequently of the order of 100-200 km in their East West extent. At times these extensions may be exceeded. In addition their effective magnetic North South dimensions may be effective to a few thousand kilometers along the earth's magnetic field. If the altitude that the plume of irregularities reached at the magnetic equator is high, the effects reach the anomaly region with its high electron density and show strong scintillation intensities. New data on the variety of altitudes of irregularities at the equator puts limits on forecasting of some structures but hope for forecasting some major plumes.

In the polar region severe problems are not expected to surface until 1999 since polar irregularities have been found to be a function of solar flux. In this area diversity of propagation paths can play an important role.

The irregularity region at auroral latitudes is primarily affected by individual magnetic storms which can occur during any phase of the sunspot cycle. Irregularity intensity is a function of various parameters of the magnetic storm relative to that station. Illustrations of GPS phase fluctuations will be shown as individual storms develop. New data clearly show the forecasting possibilities when using timing of the auroral oval occurrence at a site as a means of prediction.

L Band is to be used in the downlink of new systems proposed and being initiated for personal communications. Low link margins are forecast for several systems to be in the range 6 to 10 dB. These levels indicate that fading due to the ionospheric irregularities such as discussed in these GPS studies will cause problems.

2. PREFACE

In previous reports to AGARD on propagation problems with the reception of the 1.2 and 1.6 GHz signals from the Global Positioning System's 24 satellites, we outlined several areas where difficulties could be encountered. The observational problems outlined were in the area of ionospheric irregularities and variations of total electron content.

How the receiver deals with propagation problems is a function of the hardware-software combination of particular designs. When only simple single

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

frequency receivers are used, there could be problems ranging from a long period being necessary to acquire the satellite signal from a cold start to a loss of lock due to deep amplitude or phase scintillations. Erratic readings could result during large swings of fluctuations. The complexity of the receiver can be such that flywheel aspects can essentially reduce or eliminate certain groups of fading problems; equipment parameters have to be folded into the problem (Clynch and Aarons, 1996). However in differential GPS, phase fluctuations have to be evaluated.

The total electron content variations of importance in accurate measurements include horizontal gradients both east-west and north-south as well as travelling ionospheric disturbances of various types.

3. SCINTILLATION OCCURRENCE

While the general patterns of scintillation have been given in the literature and at AGARD Symposia, recent studies have advanced the ability to forecast high latitude occurrence. In the field of forecasting equatorial scintillation there can be methods of forecasting some major disturbances as well as localized turbulence. In any event the understanding of the problems and the propagation will allow users to correctly identify the behavior of the system as due to trans-ionospheric propagation.

4. DATA SET

The data set for these studies is obtained through the International GPS Service for Geodynamics (IGS). Using these data we can know the patterns of auroral fluctuations and equatorial effects in a more precise manner than earlier studies. The many stations reporting daily data (over a hundred) has allowed for the first time a comparison of phase fluctuation over large areas. Thus the development of irregularities whether proceeding equatorward in the high latitude region or moving westward with sunset in the equatorial region is now better understood.

Phase fluctuations reported in the present paper were obtained by examining 30 second values of total electron content (TEC). Each 30-s data point is obtained from the phase differences between the 1.2 GHz and the 1.6 GHz signals of each GPS satellite. The rate of change of these values is the source of the phase fluctuations reported.

5. HIGH LATITUDE REGION

It has been known that during magnetic storms, irregularities develop in the auroral oval of an intensity greater than during magnetically quiet periods. During magnetic storms the irregularity oval expands both equatorward and poleward - and the fluctuations intensify. A large amount of data from 1993 to 1997 has been studied. The data are processed in the manner illustrated in Figures 1a and 1b. For a high latitude station (Tromso) we show an example of the data input and a first look at the observations. Figure 1a is a map of the region in Corrected Geomagnetic Latitude and Longitude traversed by the satellite propagation path at 400 km. The total electron content is plotted in the middle panel. The rate of change of total electron content taken from the phase differences between 1.2 GHz and 1.6 GHz is shown on the right. This is not a very intense magnetic storm. For the station and the equipment taking the data the fluctuations were not enough to cause loss of lock. In Figure 1b we show two satellite propagation paths and the resulting phase fluctuations as noted from Santiago, Chile, a station in the equatorial anomaly region.

When a storm develops, irregularities are produced in the auroral region. With increasing magnetic indices the auroral region expands equatorward and poleward and the irregularities are both more frequent and of higher intensity. This can be seen in the illustration of Figure 2 where both days of low and high magnetic indices are shown. Tromso and Kiruna show high intensities for the magnetic storm of September 27-28. Ny Alesund (Spitzbergen) above the auroral oval shows some increase but does not intensify the way phase scintillations do over Tromso and Kiruna; Kiruna is approximately 1.5° lower in corrected geomagnetic latitude than Tromso.

6. EQUATORIAL REGION

The map of Figure 3 contains the geographic positions of many of the stations used in this study. At the anomaly latitude of Santiago we have extremely high total electron content and F2 electron density during certain times of the day when compared to observations at the magnetic equator. In a campaign in October 1996, one magnetic storm produced, at the South American longitudes, irregularities across the continent. GPS stations in Fortaleza, Brazil, La Plata, Argentina, Kourou, French Guiana, Bogota, Columbia, and Arequipa, Peru all showed several hours of phase scintillations in the post sunset to midnight time period, similar to that shown in Figure 1b. This was not a frequent occurrence in this solar minimum year. During the month of October 1996, only two nights showed scintillation activity at Santiago in this sunspot minimum year.

Figure 1b showed data from two satellites observed from Santiago when a high altitude plume developed at the magnetic equator. Each of the two satellites showed high levels of phase scintillations but there were periods of low level fluctuations indicating diversity path possibilities. The data were taken in the magnetic storm period of October 22-23, 1996. For the anomaly region there was a bifurcated structure to the irregularity region as viewed by Satellite 15 (02-04 UT). This allows periods of time when no scintillations would be observed on an individual path. In Figure 1b for example at 0230 UT on the path to Satellite 15, no scintillation activity is observed. However phase fluctuations are noted on the path to Satellite 14 at that time. It can be seen that the level of fluctuations did not produce loss of lock on either path for the receiver used in Santiago measurements.

Total electron content in the post sunset period in the anomaly region increases considerably during years of high solar flux. The ratio of total electron content (solar maximum compared to solar minimum) using model calculations has been made for the sunset time period by M. Fox (personal communication). For TEC the model indicates well over a factor of 5 during the sunset time period when maximum solar flux periods are compared to minimum solar flux periods. At L Band, Basu, et al., 1988 has found fades reaching peak to peak values of over 20 dB in the anomaly region in years of high solar flux; fading of this magnitude can last for several hours as has been shown even in 4 GHz results (Fang and Liu, 1984). In years of low solar flux only short lived periods of 10 dB peak to peak values have been reported. The question remains whether simple receivers will lose lock when solar flux increases in the coming years.

It is of interest that during years of low solar flux, several magnetic storms have produced disturbances at both equatorial and auroral latitudes. This may or may not be the case during high solar flux years. An example of one storm's effects at equatorial and auroral latitudes is shown in Figure 4 where both Yellowknife at high latitudes shows intense scintillations as does Santiago in the anomaly region. Auroral latitudes lower than the position of Yellowknife show considerably lower occurrence of intense phase scintillations. This storm with its high intensity lasting over a period of hours at anomaly latitudes is not the only storm of its kind found in recent studies. A series of magnetic storms is shown with similar effects in Aarons et al., 1996. In particular the storm of October 22-23, 1996 was similar to the storm of October 3, 1994; radar backscatter from the October 3, 1994 storm is shown in Figure 5 from Aarons et al., 1996. In this case irregularities developed over many hours.

It should be noted that systems being proposed for personal communications use L Band (Evans, 1997). Fading problems will certainly occur and for the anomaly region (which includes cities such as Hong Kong, Santiago, Bogota) there will be fading problems in the hours between sunset and midnight. Margins for these systems only range from 6 to 16 dB.

7. CONCLUSIONS

The spreading of irregularity development during magnetic storms at high latitudes is ubiquitous in the magnetic midnight sector. Areas equatorward and poleward of the auroral oval are affected. However in the equatorial anomaly regions the

high altitudes of the plumes show splitting or bifurcation and irregularities can move past a disturbed propagation path; there are paths which are relatively free of scintillations.

The ability of a particular GPS receiver to glide through phase and amplitude scintillations is a function of the hardware (signal to noise levels) and the software dealing with problems of loss of signal i.e. the flywheel effect. In the coming solar maximum, high latitude paths which show intense phase and amplitude scintillation will encompass the polar region; data are not available to determine if the irregularities which develop in the auroral region will reach greater intensity during years of solar maximum than during years of solar minimum. The knowledge of the effects of the ionosphere on transmissions can be of importance to users in the field. Knowledge of the phenomena will help to minimize effects and to understand that observational problems may be due to ionospheric effects rather than to equipment failures.

8. ACKNOWLEDGEMENTS

These studies were supported by the Office of Naval Research. Data from the International Service for Geodesy were supplied through the Jet Propulsion Laboratory.

9. REFERENCES

Aarons, J., M. Mendillo, R. Yantosca, E. Kudeki, GPS phase fluctuations in the equatorial region during the MISETA campaign. J. Geophys. Res., 101, 26,851-26,862, 1996

Basu, S., E. MacKenzie and Su. Basu, Ionospheric constraints on VHF/UHF communication links during solar maximum and minimum periods, Radio Science, 23, 363-378, 1988.

Bishop, G.J. and E.A. Holland, Morphology of solar maximum total electron content and L-Band scintillation in the Northern Polar Cap Ionosphere: A first look, Proc. of the Satellite Beacon Symposium, University of Wales, Aberystwyth, Wales, July 1994.

Clynch, J. and C. Henry, Ionospheric effects on GPS and DGPS in polar regions Proceedings of ION GPS-94, 2, 1579-1587, 1994

Clynch, J. and J. Aarons, High latitude GPS observations and receiver constraints, Proc. of the 1996 Ionospheric Effects Symposium, J.M. Goodman, Editor, 432-440, 1996

Evans, J.V., Satellite systems for personal communications, IEEE Antennas and Propagation Magazine, 39, June 1997

Fang, D.J. and C.H. Liu, Statistical characterizations of equatorial scintillation in the Asian region, Radio Science, 19, 345-358, 1984

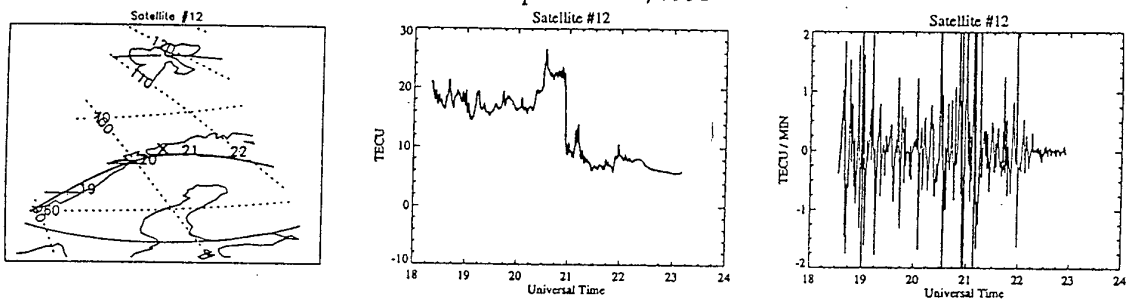


Figure 1a. Tromso data for one satellite includes a map in Corrected Geomagnetic Latitudes with the Tromso site shown by an X. Total electron content and rate of change of total electron content are shown in the next two panels.

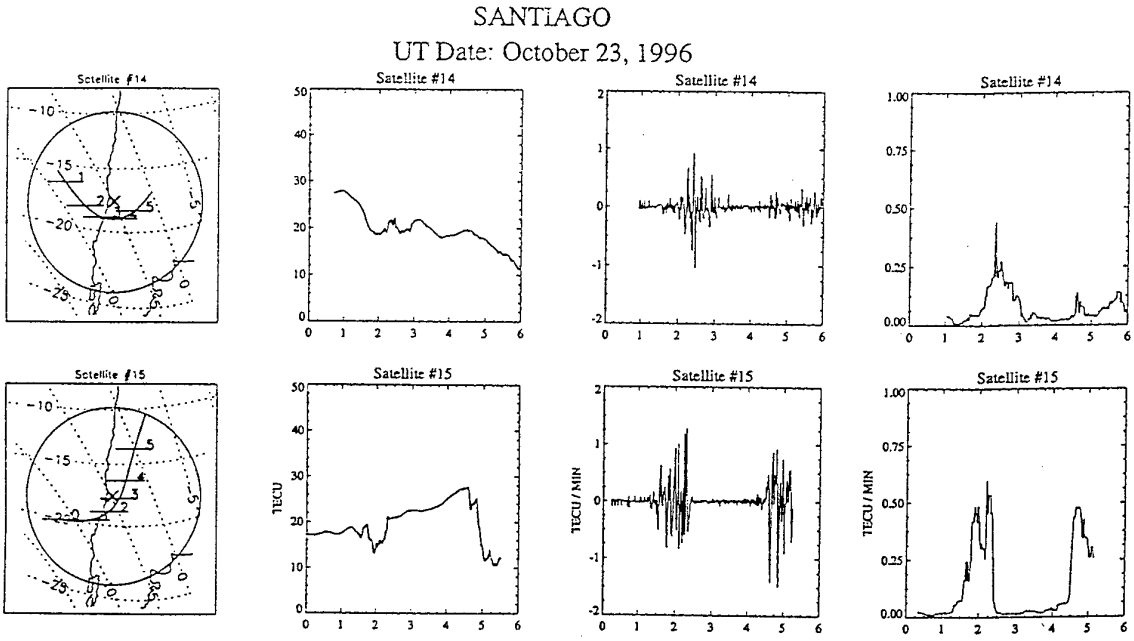


Figure 1b. Data similar to that of Figure 1a are shown for Santiago, Chile. There is an added fourth panel of relative intensity of phase fluctuations.

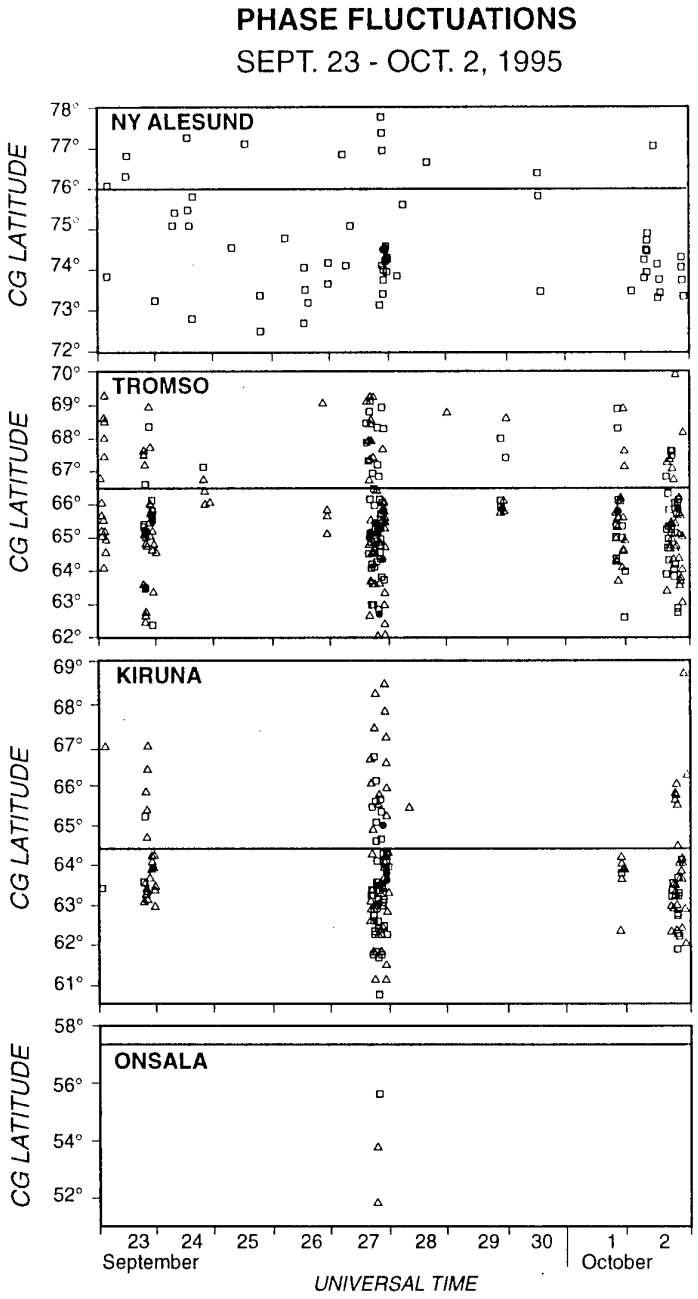


Figure 2. Scintillation activity of various levels are shown for the quiet period before the magnetic storm of September 27-28, 1995 and during the magnetic storm. Various levels are indicated where the rate of change of Total Electron Content Units exceeds plus and minus .5 TECU/minute. A Total Electron Content Unit is 1 x 10^{16} electrons per square meter.

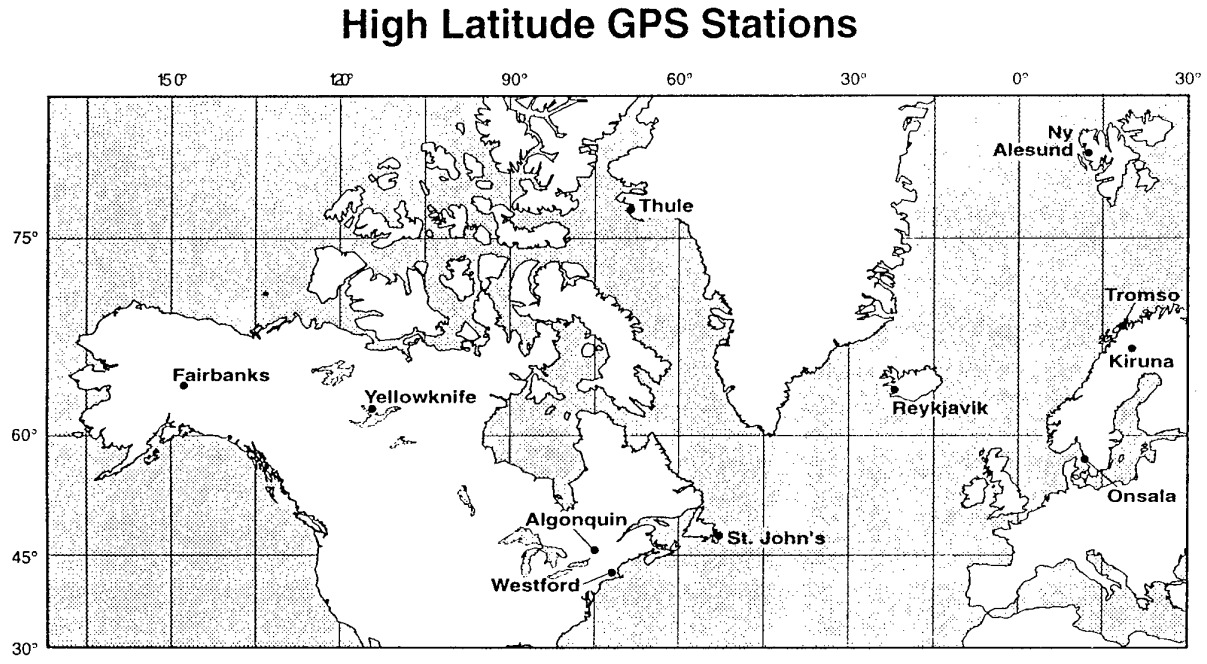


Figure 3. A map for several of the stations used for this study.

The Magnetic Storm Period of November 2-6, 1993

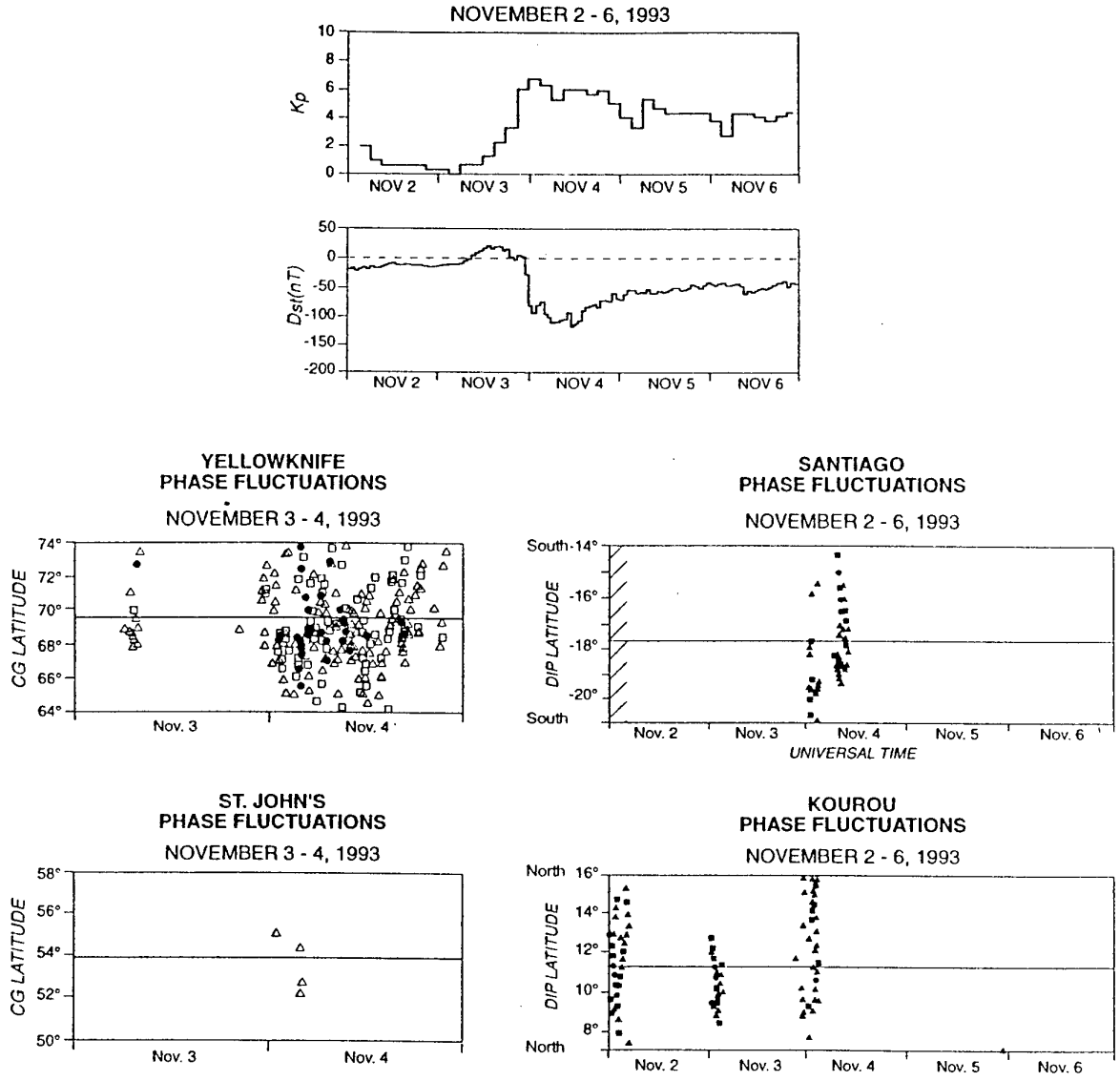


Figure 4. Magnetic storm effects increased occurrence and intensity of scintillation activity at high latitudes (Yellowknife) but less at St. John's at a lower latitude. At the equatorial anomaly latitudes of Santiago and Kourou high scintillation activity is shown for the magnetic storm day of November 3-4, 1993. For the quiet day of November 2-3, Kourou at a latitude closer to the magnetic equator than Santiago shows phase fluctuations but Santiago at 18° dip latitude does not. It should be noted that equatorial irregularities occur during both magnetically quiet and magnetically disturbed periods.

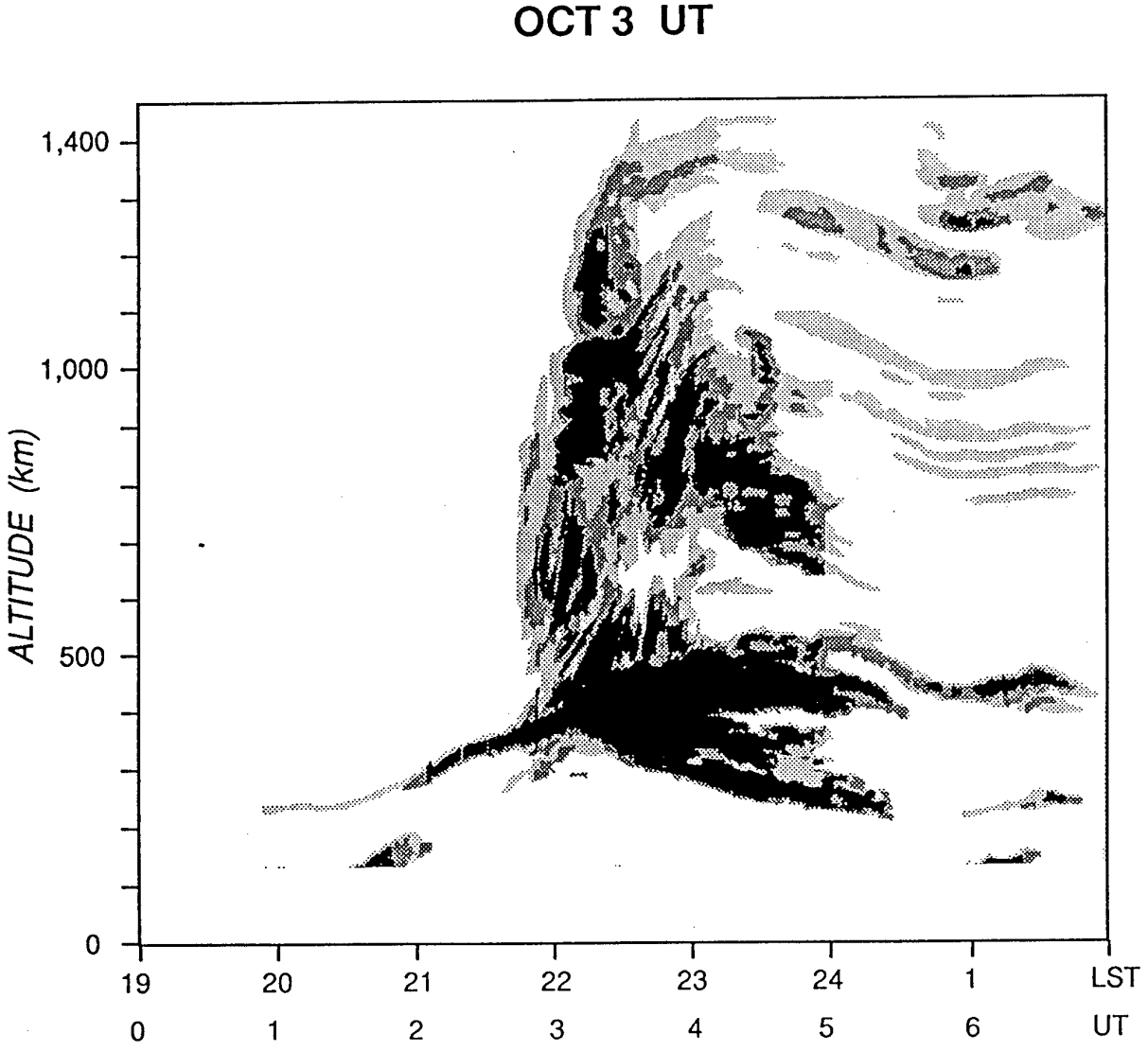


Figure 5. Backscatter returns on the magnetic storm day of October 3, 1994 as observed at Jicamarca, Peru (Aarons et al., 1996). A plume similar to this in duration and altitude also occurred on October 22-23, 1996; data shown in Figure 1b are from that storm period.

PAPER No: 24

DISCUSSOR'S NAME: R. Sabatini

COMMENT/QUESTION:

1. You mentioned that TEL fluctuations have a very important effect on GPS performance. Probably this is not true for P-code double frequency receivers. Is that correct?

2. Are there any substantial differences between the performance of DGPS using "carrier-phase" and DGPS using "pseudo range" with regard to atmospheric propagation effects?

AUTHOR/PRESENTER'S REPLY:

1. Yes, military P-code receivers are not affected by TEL fluctuations.

2. We did not investigate this particular aspect.

METEOROLOGICAL DATA FUSION TO ASSESS EM/EO PROPAGATION CONDITIONS IN COASTAL ENVIRONMENTS

J. Rosenthal
R. Helvey
P. Greiman
M. McGovern
Geophysics Branch (521400E)
Naval Air Warfare Center Weapons Division
Point Mugu, California 93042-5001 USA

SUMMARY

EM/EO propagation conditions can be estimated along specific routes or regions of interest by fusing on one display, the visible satellite imagery showing cloud cover, the operational geometry, and derived quantities such as duct height estimated from the IR-duct technique. In addition, by compositing conditions over a set of similar synoptic events, and when also fused with averaged synoptic field gridded data, an average depiction of duct height and associated variability can be obtained. To the extent the cases selected are typical of that type of event, the averaged picture can be used to forecast or estimate the distribution of EM/EO parameters in the near future if confidence exists in the forecast synoptic situation. Additional compositing is underway to broaden the range of synoptic weather types and regions examined.

BACKGROUND

Efforts have continued at the NAWCWPNS Geophysics Branch at Point Mugu to improve the automation and accuracy of satellite-based efforts to diagnose the character, height, strength and horizontal variability of EM/EO propagation conditions. Rosenthal et al (Ref. 1, 2) laid the groundwork for some of these efforts using descriptive and semi-objective techniques of cloud pattern recognition of stratus/stratocumulus clouds to infer information on the distribution of refractive properties. Lyons (Ref. 3, 4) and Helvey (Ref. 5, 6, 7) developed the more automated "Infrared (IR) Duct" technique to convert satellite-derived cloud-top temperatures to an estimate of duct height, and Greiman and Rosenthal et al (Ref. 8) have attempted to use the approach to measure the amount of horizontal variability in the marine boundary layer. Utility of these procedures for discerning some of the atmospheric features important to electro-optical (E-O) systems has also been evident.

Both the subjective and objective satellite techniques, together with synoptic and climatological conditions are being integrated or fused together in an expert system (EXPERDUCT) to allow different levels of user expertise to synthesize various factors in making refractive assessments (Ref. 7)

This paper summarizes some of the more recent efforts to determine variability by employing the IR-duct and statistical techniques.

HORIZONTAL VARIABILITY ASSESSMENTS

Several papers (Rosenthal, Helvey, Greiman (Ref. 7, 8)) have documented use of the IR-Duct approach to measure the horizontal variability of duct heights across specific synoptic/mesoscale features. By using this approach for the same region and time on successive days, a measure of the temporal changes associated with these case studies was observed as well. Though of significant value as a diagnostic tool for EM/EO assessment, (see fig. 1) the unique features of each case study make the approach of using IR-duct heights for predictive purposes dependent on the extent to which the case study approximates the typical dimensions and intensities for that type of feature.

In order to generalize the distribution of duct heights that can be expected for different types of synoptic regimes, an averaging technique has been employed which permits the characteristics of several similar synoptic cases to be lumped or composited together.

COMPOSITING TECHNIQUE

The methodology for this involves initial selection of case studies for a group of similar meteorological patterns observed over the western U. S. and Eastern Pacific Ocean, e.g. Santa Ana (Great Basin High Pressure/offshore wind flow); cut-off low pressure systems; stable ridge, or other common synoptic situation. U. S. Department of Commerce's "The "Daily Weather Map" weekly series (Ref. 9) was used to select common episodes. Using a commodore

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

Amiga 3000, gridded U. S. National Center for Environmental Prediction (NCEP) Medium Range Forecast (MRF) 1000 MB, 700 MB and 500 MB height data and concurrent GOES-9 satellite imagery was down loaded from
1 GB 'JAZ' diskettes onto floppy disks for display on an IBM PS/2 Model 70 computer as part of the Meteorological Interactive Data Display System (MIDDS).

Duct heights were calculated over areas covered by low stratus/stratocumulus clouds using the NAWCWPNS IR-duct technique along selected radials (or arcs) spanning the domain of analysis. The domains for both satellite and field data were synchronized. Generally, four to six 'radials' were sufficient to cover the domain, depending on cloud cover, with three to six points along each radial (again dependent on cloud cover at designated pixel locations).

After all duct heights were calculated, they were averaged for each point on the radial. This generated average duct heights for each latitude-longitude point on each radial for the selected cases.

Next, MRF gridded field data were also averaged and displayed on the MIDDS system, together with the calculated duct heights.

RESULTS

A composite of gridded field data at two levels and satellite-derived IR-Duct heights is shown in figure 2 for what has been termed a "San Diego Ridge" - a situation whereby strong ridging aloft prevails over the U. S. Southwest with a high centered just to the west of San Diego. The averaged height contours at 500 mb are shown in figure 2 at 20 meter intervals. Also shown are the averaged 1000 mb heights at 10 meter intervals. The bold numbers are averaged duct heights in feet (taken as the optimum coupling height (OCH)) derived for the GOES-9 IR data. The duct heights have been manually analyzed, - shown as dashed lines or contours.

The period of record used in this composite covers 3 episode periods, May 10-12, 1996; June 1 1996; and August 28-30, 1996.

Inspection of this composite reveals a subsidence-induced minimum in duct height off the Central and Southern California Coast with minimum of about 600 feet west of the Monterey Bay area.

Duct heights for this type of synoptic occurrence increased slightly to the south and southeast, but increase sharply to the north and west where average 500 mb heights are much lower, and where subsidence is presumably much reduced. The full range of duct height increase from the minimum to a region about 300 to 400 miles away was about 2,600 feet.

Figure 3 shows a composite for a different synoptic situation; - one dominated by a cut-off low over the U. S. Southwest with cyclonic curvature, instability and a minimum of subsidence over the Southern California region. This figure was obtained by averaging similar weather patterns from March 12, 1996; March 14-15 1996; May 24-25 1996; Oct 26, 28-29 & 31, 1996; and Nov 1-3, 1996.

In this kind of synoptic situation, highest duct heights (approximately 3,300 feet) occur in a zone from the coast south of Monterey Bay, extending seaward to the southwest for approximately 300 miles. To the northwest from this maximum, duct heights decrease by approximately 1400 feet in about 300 miles to a duct height minimum located off the Northern California coast. This marks the region where offshore flow predominates north of the low aloft, with subsidence of air from inland areas to the coast.

It is interesting to note that whereas the synoptic conditions are dramatically different in the two examples shown, both reveal a sharp slope in duct height along the coast from southeast to northwest, but of opposite sign.

Additional synoptic regimes are also being examined to characterize the variability of EM/EO parameters associated with each type. The additional compositing should broaden the range of synoptic weather types and geographical regions examined.

FUSION OF SATELLITE IMAGERY AND OPERATIONAL GEOMETRY

Another useful tool that is permitted by the Navy's current meteorological and oceanographic (METOC) architecture is the ability to combine satellite imagery with defined areas or lines of interest in such a way that the user's attention can be focused on cloud and other conditions in his or her immediate purview. Figure 4 shows an example demonstrating this simple technique. When combined with other information such as estimates of duct height, a real-time assessment can be made of expected propagation conditions along specific routes that apply to both

EM (duct height) and EO (clouds and fog).

REFERENCES

1. Rosenthal, J. S., Westerman, S. and Helvey, R. A., "Inferring Refractivity Conditions from Satellite Imagery", Pacific Missile Test Center, Geophysical Sciences Technical Note No. 96, 1985.

2. Rosenthal, J. S., Helvey, R. A. Lyons, S. W., Fox, A. D., Szymber, R. and Eddington, L., "Weather Satellite and Computer Modeling Approaches to Assessing Propagation over Marine Environments", AGARD CP No. 453, 1989

3. Lyons, S. W., "Satellite-Derived Refractive Duct Height Estimates", Pacific Missile Test Center, Geophysical Sciences Technical Note No. 98, 1985.

4. Lyons, S. W., "SPADS Automated Duct Height Statistics", Pacific Missile Test Center, Geophysical Sciences Technical Note 100, 1985.

5. Helvey, R. A., and Rosenthal, J. S., "Guide for Inferring Refractive Conditions from Synoptic Parameters", Pacific Missile Test Center, Technical Publication TP-5, 1983.

6. Rosenthal, J. S. and Helvey, R. A., "Refractive Assessments from Satellite Observations", AGARD CP 502, 1992.

7. Helvey, R. A., Rosenthal, J. S., Eddington, L., Greiman, P., and Fisk, C., "Use of Satellite Imagery and Other Indicators to Assess Variability and Climatology of Oceanic Elevated Ducts", AGARD CP 507, 1995.

8. Greiman, P., Rosenthal, J., and Helvey, R., "Synoptic Variability Revealed by Satellite and Equivalent Altitudes", AGARD CP 582, 1996.

9. U. S. Department of Commerce, National Oceanic and Atmospheric Administration, "Daily Weather Map" series, 1996.

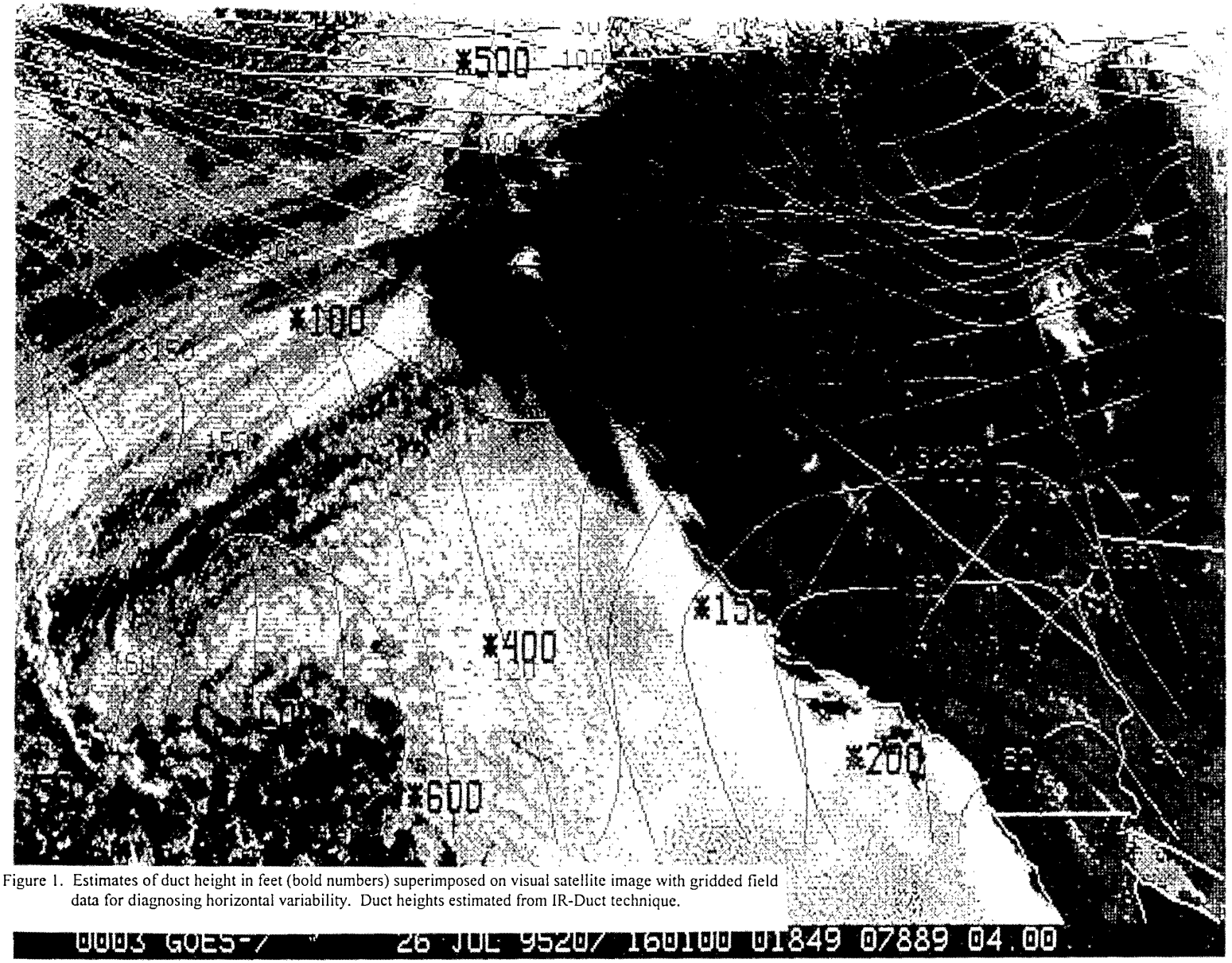


Figure 1. Estimates of duct height in feet (bold numbers) superimposed on visual satellite image with gridded field data for diagnosing horizontal variability. Duct heights estimated from IR-Duct technique.

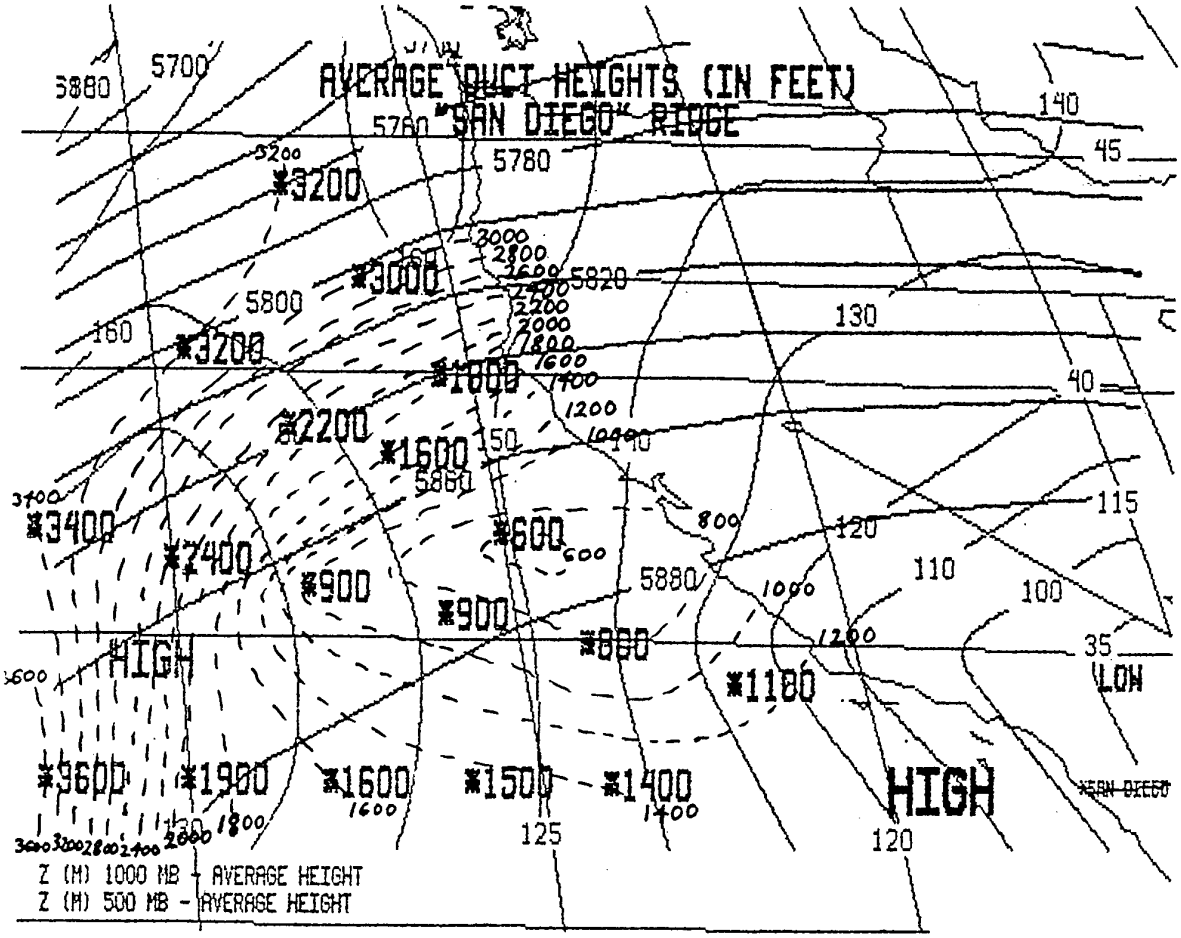


Figure 2. Composite of averaged IR-Duct heights in feet (bold numbers) along selected radials superimposed on averaged height fields at 500 mb and 1000 mb for a "San Diego Ridge", a feature that typically results in subsidence and low ducts. Analysis of estimated duct heights shown as broken lines.

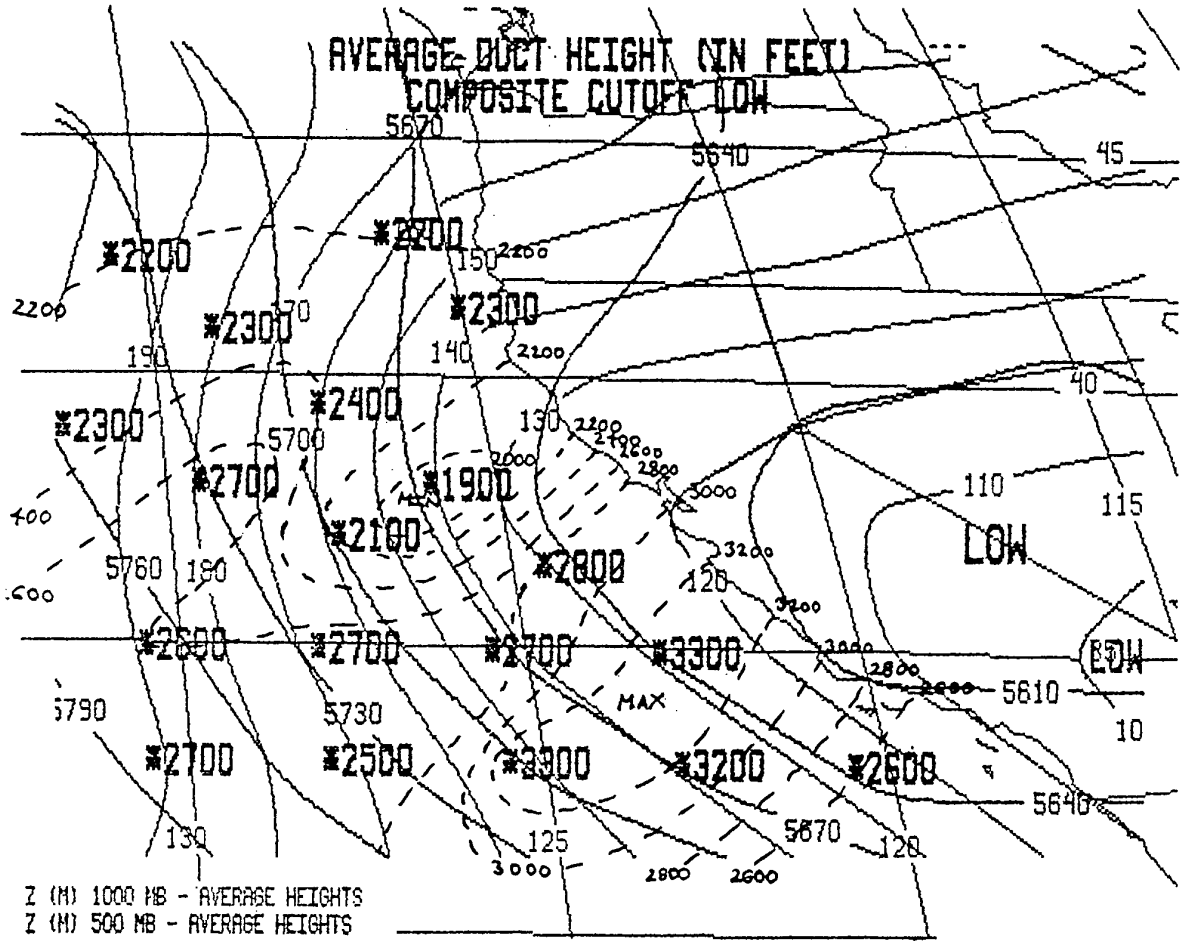


Figure 3. Composite of averaged conditions (same as above) except for synoptic events marked by cutoff low pressure system aloft.

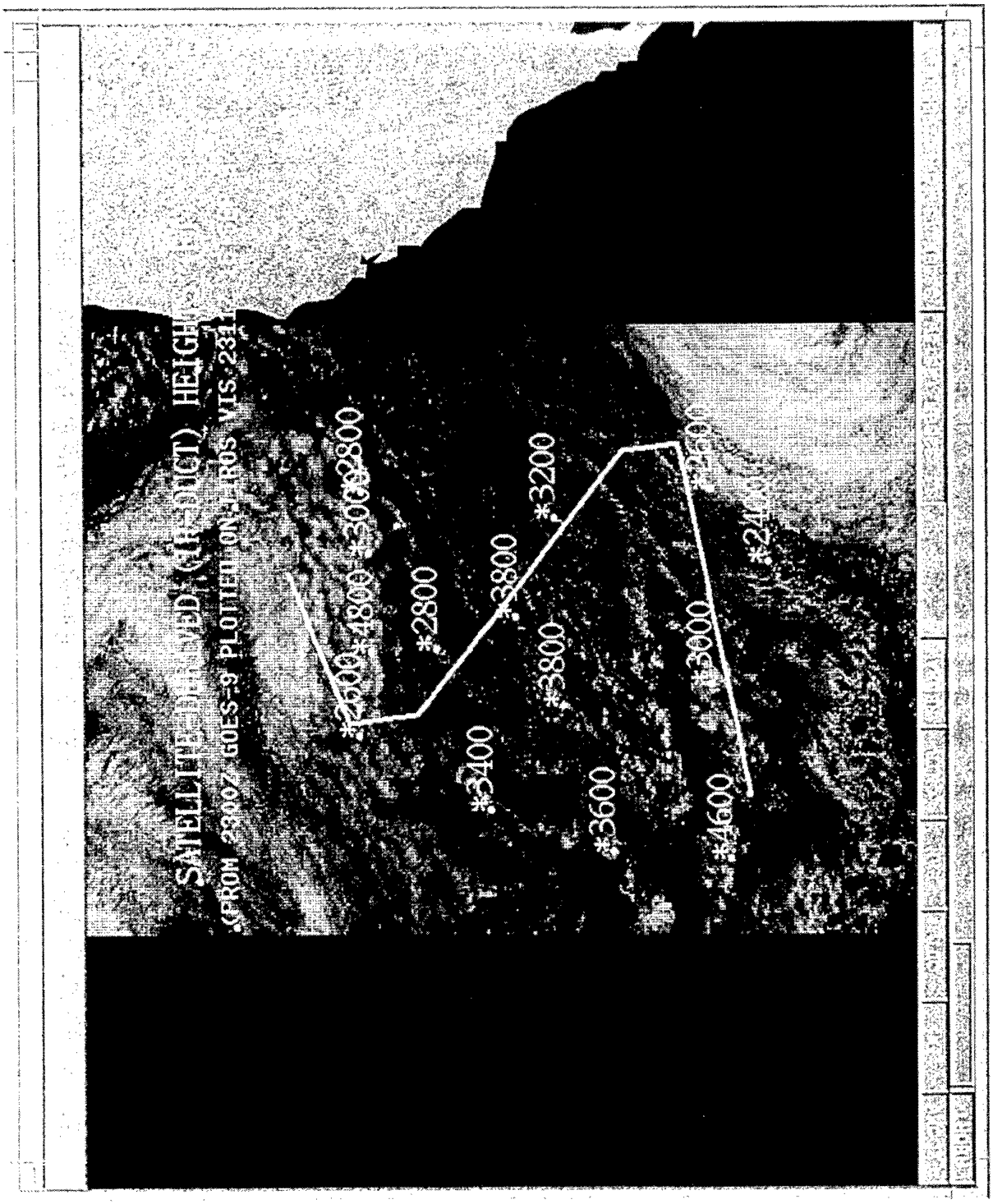


Figure 4. Satellite image superimposed by estimates of duct height and by routes of operational interest fused into one display.

UTILISATION D'INFORMATIONS CONTEXTUELLES DANS DES ALGORITHMES DE PISTAGE MULTICAPTEUR

Vincent NIMIER

ONERA BP72 92322 CHATILLON CEDEX
FRANCE

Résumé

Nous proposons, dans cet article, une méthode permettant de combiner des informations symboliques avec des informations numériques. Pour ce faire, nous proposons un processus d'estimation supervisée par un niveau d'analyse du contexte. L'application visée est la fusion de données et les algorithmes qui sont implantés dans les systèmes multicapteurs. Ces algorithmes doivent être conçus pour que le système fonctionne d'une façon nominale dans toutes les conditions. Pour cela le système devra s'adapter, de sorte qu'à tout instant l'estimation tienne compte du contexte considéré. Le résultat étant de privilégier, dans certains contextes, les mesures issues des capteurs en état nominal de fonctionnement et de réduire l'importance de celles qui sont aberrantes au vu de certains critères établis au préalable par un expert.

1. Introduction

Les développements récents des systèmes de perception convergent actuellement vers l'utilisation conjointe de capteurs multiples [1]. En effet, les bénéfices attendus sont prometteurs : une capacité plus importante d'analyse des situations complexes, une robustesse accrue à l'environnement. Les domaines concernés touchent aussi bien le milieu industriel pour les tâches d'assemblage, la robotique mobile, que le milieu militaire dans le domaine du commandement et du contrôle de champs de bataille, de la poursuite de cibles, ou de la navigation d'engins aériens. L'intégration et la fusion d'informations multiples sont devenues dès lors une voie d'investigation et de recherches très actives.

Si les techniques classiques d'estimation et de classification basées sur la théorie des probabilités ont vu leur champ d'application s'élargir à la fusion de données, l'émergence de problèmes nouveaux et propres aux systèmes multicapteurs a été à l'origine de recherches de nouvelles modélisations pour le traitement de l'information. Parmi celles-ci la logique floue [6] et la théorie de Dempster-Shafer [9] sont apparues comme des alternatives aux probabilités sans pour autant les remplacer. Elles permettent de manipuler et de traiter des informations souvent hétérogènes et d'origines incertaines. De façon générique, si la théorie des probabilités est notamment utilisée pour la modélisation de phénomènes aléatoires, la logique floue elle, a un champ d'applications tourné vers la représentation de la connaissance humaine. Une dualité apparaît donc entre ces deux théories qui peut être mise à profit, notamment dans un système multicapteur. En effet, les informations manipulées obéissent à cette même dualité, les unes, aléatoires, sont les mesures issues de chaque capteur, les autres relèvent d'une connaissance plus symbolique, et se révèlent utiles, de la conception du système jusqu'à son utilisation.
Dans le domaine de la poursuite de cibles [2], [3], [6], [7], [10], les algorithmes proposés pour la fusion de données sont basés sur une approche exclusivement probabiliste. Les filtres de Kalman développés et leurs extensions IMM (Interacting Multiple Models), PDAF (Probability Data Association Filter), JPDAF(Joint Probability Data Association Filter), ... [2], supposent que les capteurs qui composent le système ont des caractéristiques connues. En outre, le contexte n'est jamais pris en compte, ce qui suppose, implicitement, que celui-ci est favorable à l'utilisation simultanée de l'ensemble des capteurs. Cette hypothèse est, à l'évidence, souvent très loin d'être vérifiée. Pour qu'un système multicapteur puisse fonctionner d'une façon nominale dans toutes les conditions pour lesquelles il a été conçu, il faut analyser le contexte et s'y adapter. Le résultat est simplement de privilégier, dans les algorithmes, les mesures issues des capteurs en état nominal de fonctionnement et de réduire l'importance de celles qui sont aberrantes au vu de certains critères établis au préalable.

Nous proposons dans cet article une méthode, et les algorithmes associés, permettant de prendre en compte le contexte pour un système multicapteur. Cette méthode élargit les travaux initialement menés dans [8], [9]. L'organisation de cet article est la suivante : le principe général de la méthode est présenté dans la deuxième partie ainsi que la logique de fonctionnement qui en résulte. La troisième partie décrit les équations d'estimation qui prennent en compte le contexte. Nous distinguerons l'estimation statique, faite à partir d'un ensemble de mesures acquis à un instant donné, de l'estimation dynamique, calculée à partir de toutes les mesures passées, et qui conduit à des équations de filtrage spécifiques. Le dernier chapitre est consacré à une simulation.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

2. Logique de fonctionnement du système

2.1 PRINCIPE GÉNÉRAL

On considère qu'un contexte particulier peut être identifié par des variables dites "contextuelles". La nature et l'origine de ces variables sont très diverses. On peut considérer des mesures faites par des capteurs annexes tels que ceux mesurant la pluviométrie ou la température extérieure, etc... De même un opérateur, par l'intermédiaire d'une interface homme/machine, peut donner des indications précieuses sur les conditions opérationnelles du moment. Des traitements supplémentaires mesurant un rapport signal sur bruit, la largeur d'un pic de corrélation ou tout autre paramètre ou indicateur permettant, dans certain cas, d'évaluer la qualité du signal ou d'indiquer l'état de fonctionnement de chaque capteur, peuvent être pris en compte.

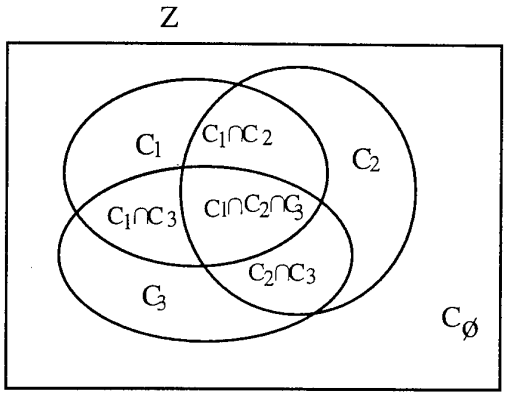
La représentation des connaissances qui décrivent l'état de fonctionnement de chaque capteur, et donc la qualité des mesures qu'il est susceptible de fournir, est basée sur une description en termes de sous-ensembles flous. Cette connaissance est établie par un expert capable d'évaluer les performances et les limites de chaque capteur, au moyen de fonctions d'appartenance définies sur les variables contextuelles. Dès lors, en situation réelle, et connaissant les valeurs que prennent chaque variable contextuelle, la conjonction des fonctions d'appartenance établira la validité des mesures issues de l'un ou l'autre des capteurs. Ainsi, il est possible de définir l'association de capteurs la mieux adaptée à un contexte particulier, et de ne prendre en compte, dans le processus d'estimation, que les mesures issues de cette association.

2.2 PARTITIONNEMENT DE L'ESPACE CONTEXTUEL

La prise en compte du contexte apparaît comme une idée assez naturelle pour une personne en charge de la réalisation d'un système. Cependant sa mise en œuvre effective n'est pas immédiate et aboutit souvent à l'élaboration de quelques heuristiques et à l'évaluation de coefficients dit " de confiance" ; l'ensemble fournissant un résultat satisfaisant quoique dépendant de l'application concernée. Il n'existe, à notre connaissance, pas de méthodologie générale permettant de formaliser le problème. Nous proposons dans cet article un formalisme. Celui-ci permet de définir un espace contextuel et d'établir, sur cet espace, une logique de fonctionnement du système. Cette logique sera ensuite utilisée pour superviser le processus d'estimation.

2.2.1 Définition

Nous considérerons dans la suite un système S constitué de n capteurs. Les variables contextuelles seront désignées par les lettres z_j, avec $j \in \{1, ..., p\}$, p étant le nombre de variables contextuelles considérées. Un contexte particulier z est donc défini par p mesures ou valeurs de chaque variable contextuelle, si bien que l'on peut noter $z = \{z_1, ... z_p\}$. Les contextes appartiennent à un espace à p dimensions noté Z. Un capteur est valide pour un ensemble donné de contextes qui est représenté par un sous-ensemble noté C_i de Z, avec $i \in \{1, ..., n\}$. La figure 1 illustre cet aspect pour un système composé de trois capteurs. Chaque sous-ensemble C_i, avec $i \in \{1, 2, 3\}$, est représenté ainsi que toutes les intersections entre les sous-ensembles.



Partitionnement de l'espace contextuel

Figure 1

Nous distinguerons ici deux domaines de validité, pour chaque capteur, représentés par les deux notations C_1 et c_1. Le domaine de validité inclusif C_1 représente le sous-ensemble de contextes pour lequel le capteur n°1 est valide, sans préjuger toutefois de la validité des autres capteurs. Le domaine de validité exclusif c_1 représente le sous-ensemble de contextes où seul le capteur n°1 est valide à l'exclusion de tous les autres. La relation logique qui lie les deux est de la forme: $c_1 = C_1 \cap \bar{C}_2 \cap \bar{C}_3$. Cette distinction peut être faite pour toute combinaison de capteurs, ainsi $c_{\{1,2\}} = C_1 \cap C_2 \cap \bar{C}_3$ est le sous-ensemble de contextes dans lequel les capteurs numéro 1 et 2 sont valides mais pas le capteur numéro 3. Plus généralement, pour un système comprenant n capteurs, on peut constituer l'ensemble $A = \{ c_\emptyset, c_1, c_2, ..., c_{\{1,2\}}, ..., ... c_{\{1,2, ..., n\}}\}$ des domaines de validité exclusif de toutes les combinaisons de capteurs, avec $c_J = \bigcap_{j \in J} C_j \bigcap_{i \in \bar{J}} \bar{C}_i$ et $J \in \{1, ..., n\}$ L'ensemble A, constitué d'éléments exclusifs, forme une partition de Z. Notons que $c_\emptyset$ représente une absence de capteur valide $c_\emptyset = \bar{C}_1 \cap \bar{C}_2 \cap ... \cap \bar{C}_n$. Dans la suite, nous noterons c_J un élément de A, où J désigne un sous-ensemble de

l'ensemble d'indices {1,...,n} représentant les capteurs valides.

2.2.2 Remarque

La représentation précédente appelle la remarque suivante. L'ensemble des contextes pour lesquels les trois capteurs sont simultanément valides $C_1 \cap C_2 \cap C_3$ contient, à l'évidence, un nombre de contextes moins important que chaque sous-ensemble C_i pris individuellement. De ce fait, une stricte utilisation d'un système multicapteur aux conditions pour lesquelles l'ensemble des capteurs est valide limite drastiquement le champ d'utilisation de ce système. D'où la nécessité d'identifier les contextes et de considérer alors l'association de capteurs adaptée à chacun d'eux.

2.3 PROBABILITÉ DE FONCTIONNEMENT D'UN CAPTEUR

2.3.1 Logique binaire

En logique binaire, le sous-ensemble C_i de Z peut être représenté par sa fonction indicatrice d'ensemble $I_i(z)$. Cette fonction prend alors les deux valeurs 1 ou 0 suivant que le capteur i est valide ou ne l'est pas pour le contexte z. Dès lors que z est un vecteur composé de p variables élémentaires z_j, la fonction indicatrice $I_i(z)$ est obtenue par l'opération logique de conjonction $\wedge$ des fonctions indicatrices élémentaires $I_{ij}(z_j)$.

$$I_i(z) = I_{i1}(z_1) \wedge I_{i2}(z_2) \ldots \wedge I_{ip}(z_p)$$

Chaque fonction élémentaire $I_{ij}(z_j)$ définit la plage de validité du capteur i dans le contexte identifié par la variable z_j. Par convention, on prendra $I_{ij}(z_j)=1$ si la variable z_j ne renseigne en rien sur la validité du capteur i.
L'origine des variables z_j conduit à considérer z comme un vecteur aléatoire de densité de probabilité $p(z / z^m)$ où z^m est le vecteur représentant les valeurs mesurées de la variable z. Cette densité modélise l'incertitude liée à la mesure. La prise en compte de cette incertitude ne permet plus d'établir d'une façon binaire la validité d'un capteur mais d'évaluer la probabilité pour que le capteur soit valide. Celle-ci est donnée par la formule :

$$P(C_i / z^m) = \int I_i(z)\, p(z / z^m)\, dz$$

Connaissant la valeur des paramètres mesurés z^m, $P(C_i/z^m)$ représente la probabilité pour que la valeur de z appartienne au sous-ensemble C_i .

2.3.2 Logique floue

La logique floue apporte une nuance supplémentaire dans la représentation des plages de validité de chaque capteur. En effet, dans une logique binaire, les bornes définissant ces plages sont souvent arbitraires, elles peuvent être avantageusement remplacées par des intervalles sur lesquels la validité de chaque capteur est mesurée par une variable prenant ses valeurs dans l'intervalle [0,1]. Les fonctions indicatrices sont alors les fonctions d'appartenance $\mu_i(z)$ du sous-ensemble flou C_i, la notation restant ici identique à celle adoptée en logique binaire puisque aucune confusion ne peut être faite. Comme précédemment $\mu_i(z)$ est obtenue par conjonction des fonctions d'appartenance élémentaires $\mu_{ij}(z_j)$.

$$\mu_i(z) = \mu_{i1}(z_1) \wedge \mu_{i2}(z_2) \ldots \wedge \mu_{ip}(z_p)$$

où $\wedge$ est un opérateur de conjonction de logique floue. Comme précédemment, dès lors qu'un capteur i n'a pas de lien direct avec un contexte identifié par la variable z_j, la convention est de prendre $\mu_i(z_j) = 1$ pour toutes les valeurs que prend z_j.

L'utilisation de fonction d'appartenance de sous-ensemble flou permet de s'affranchir d'un arbitraire portant sur la définition des bornes de validité de chaque capteur. Ainsi elle donne la possibilité de modéliser l'incertitude inhérente à cette définition. Une seconde source d'incertitude réside toujours dans la composante aléatoire des variables z_j. La combinaison des deux incertitudes est à l'origine de la définition de la probabilité d'un événement flou proposée par Zadeh [14] suivant la relation :

$$P(C_i / z^m) = \int \mu_i(z)\, p(z / z^m)\, dz \qquad (1)$$

Lorsque l'incertitude liée à la mesure est négligeable, ou encore, si la valeur de la variable est certaine, la densité de probabilité $p(z / z^m)$ est remplacée par un dirac $\delta(z-z^m)$ permettant l'identification de la probabilité de l'événement flou à la valeur que prend la fonction d'appartenance au point considéré. $P(C_i/z^m)$ représente la probabilité pour que la valeur de z appartienne au sous-ensemble flou C_i sachant que le contexte mesuré est z^m.

Lorsque les variables z_j sont indépendantes, et en prenant l'opérateur Min comme opérateur de conjonction, la formule (1) se développe suivant :

$$P(C_i / z^m) = \int \text{Min}\,(\mu_{i1}(z_1), \ldots, \mu_{ip}(z_p))\, p(z_1/z_1^m) \ldots p(z_p/z_p^m) dz_1 \ldots dz_p$$

2.4 PROBABILITÉS D'UN GROUPEMENT DE CAPTEURS

2.4.1 Probabilité d'association

Plusieurs capteurs peuvent être associés, la probabilité du groupement qui en résulte est égale à la probabilité de la conjonction des événements flous associés à chaque capteur. Pour deux événements cette probabilité est définie par la relation suivante :

$$P(C_i \cap C_j / z^m) = \int \text{Min}\,(\mu_i(z), \mu_j(z))\, p(z/z^m)\, dz,$$

lorsque Min est pris comme opérateur de conjonction. La généralisation à l'intersection de plusieurs événements est immédiate.

2.4.2 Probabilité du domaine de validité exclusif

La probabilité d'un domaine de validité exclusif en fonction des domaines de validité inclusifs est définie par la formule suivante, dont la démonstration est donnée en annexe :

$$\beta_J = P(c_J) = \sum_{\{I \subseteq N / J \subseteq I\}} (-1)^{|I-J|} P(\cap_{i \in I} C_i) \qquad (2)$$

et $\beta_\emptyset = P(c_\emptyset) = P(\cap_{j \in N} \bar{C}_j)$

où I et J sont deux ensembles d'indices correspondant chacun à une partie de l'ensemble N={1, .,n}. On note |I-J| le cardinal de l'ensemble I-J. Pour des raisons de simplicité d'écriture, nous avons omis d'écrire le conditionnement par la variable mesurée z^m dans les probabilités. Ainsi, $P(c_J) = P(c_J/z^m)$ représente la probabilité pour que le contexte mesuré z^m appartienne au domaine de validité exclusif c_J. Il existe alors autant de probabilité $P(c_J)$ qu'il y a d'éléments dans A c'est-à-dire 2^n. La condition de normalisation étant vérifiée :

$$\sum_{J \subseteq N \cup \emptyset} \beta_J = 1$$

3. Estimation avec prise en compte du contexte.

Les probabilités qui viennent d'être définies ci-dessus permettent d'établir la validité de chaque groupement de capteurs. Ainsi, elles permettent de valider les mesures issues des différents capteurs qui composent le groupement et, de ce fait, l'estimée fournie par la fusion partielle de celles-ci. Lorsque plusieurs groupements sont simultanément valides, avec des probabilités différentes, la fusion globale est le résultat de la moyenne des fusions partielles pondérées par les probabilités qui leur sont associées.

Nous distinguerons alors deux cas suivant que l'état ne dépend que de l'instant présent (estimation statique) ou bien des états passés (estimation dynamique). Les applications visées diffèrent d'un cas à l'autre.

3.1 ESTIMATION STATIQUE

3.1.1 Formulation du problème

Le système est composé de n capteurs chacun délivrant une mesure y^i, avec $i \in N$, permettant l'observation d'un état x à travers n équations d'observation :

$$y^i = H_i\,(x, b^i)$$

où H_i, $i \in N$, sont les n systèmes d'observation, b^i les bruits d'observation. Par souci de simplicité nous considérerons que le bruit est additif, gaussien, de moyenne nulle et de variance $E(b^i b^j) = \sigma_i^2\, \delta(i,j)$ où δ est le symbole de Kronecker. L'ensemble des observations disponibles est regroupé dans le vecteur d'observation $Y^T = \{ y^1, ..., y^n \}$.

3.1.2 Équations d'estimation

L'estimation de l'état x, optimal en moyenne quadratique, est obtenue par la moyenne conditionnée aux observations de la variable x :

$$\hat{x} = E(x/Y) = \int x\, p(x / Y)\, dx \qquad (3)$$

On peut développer la probabilité p(x / Y) sous la forme suivante :

$$p(x / Y) = \sum_{J \subseteq N \cup \emptyset} p(x / Y, c_J)\, P(c_J) \qquad (4)$$

Il existe donc 2^n probabilités élémentaires $p(x / Y, c_J)$ affectées à x et correspondant aux parties de N.

L'estimation de x avec prise en compte du contexte s'obtient en remplaçant la probabilité dans (3) par l'expression donnée en (4) :

$$\hat{x} = P(c_\emptyset) \int x p(x/Y, c_\emptyset) dx + \sum_{J \subseteq N} P(c_J) \int x p(x/Y, c_J) dx \qquad (5)$$

Le conditionnement par c_J signifie que seuls les capteurs dont les indices sont contenus dans J sont valides. De ce fait, seules les observations correspondant à ces capteurs doivent être considérées. Ces observations sont alors

regroupées dans le vecteur $Y_J = \{y_k\}_{k\in J}$ ce qui permet d'écrire la relation (5) sous la forme :

$$\hat{x} = x_0\beta_\emptyset + \sum_{J\subseteq N} \beta_J \, E(x / Y_J) \qquad (6)$$

Les coefficients β_J sont donnés par la relation (2).

La variable x_0 qui apparaît dans (6) correspond à la quantité :

$$x_0 = \int x \, p(x / Y, c_\emptyset) \, dx$$

Comme $c_\emptyset$ représente l'absence de capteur valide, x_0 est donc fixée a priori et doit être considérée comme la valeur limite, par défaut, que doit prendre l'état si aucun capteur n'est en mesure de délivrer une observation cohérente.

Dans certains cas cette situation peut paraître absurde. On suppose alors, qu'à tout instant, un des capteurs au moins est en mesure de délivrer une donnée acceptable, l'équation (6) peut être mise sous la forme :

$$\hat{x} = [\sum_{J\subseteq N} \beta_J \, E(x / Y_J)] / [1-\beta_\emptyset] \qquad (7)$$

Chaque probabilité β_J est divisée par $1-\beta_\emptyset$ de sorte que la condition de normalisation soit respectée.

3.1.3 Exemple

L'exemple suivant permet d'illustrer les équations proposées. Prenons un système composé de deux capteurs ; par exemple : un capteur radar C1 et une caméra visible C2. Deux variables contextuelles permettent d'identifier le contexte : z_1 est un indicateur de fonctionnement du radar, z_2 est un paramètre qui mesure la luminosité. Les deux fonctions d'appartenances qui permettent d'établir la validité de chaque capteur seront notées $\mu_1(z_1)$ et $\mu_2(z_2)$.

Pour ce système la formule (2) fournit les probabilités suivantes :

$$\begin{aligned}
\beta_{\{\emptyset\}} &= P(\bar{C}_1 \cap \bar{C}_2) = P(c_\emptyset) \\
\beta_{\{1\}} &= P(C_1) - P(C_1 \cap C_2) \\
\beta_{\{2\}} &= P(C_2) - P(C_1 \cap C_2) \\
\beta_{\{1,2\}} &= P(C_1 \cap C_2)
\end{aligned} \qquad (8)$$

Si les deux observations y_1 et y_2 sont indépendantes, gaussiennes, de moyenne x, et de variance respectivement σ_1^2 et σ_2^2 , pour chaque association de capteurs, les différentes estimées sont :

$$E(x / \{y_1\}) = y_1$$

$$E(x / \{y_2\}) = y_2$$

$$E(x / \{y_1,y_2\}) = y_1 \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} + y_2 \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} \qquad (9)$$

Lorsqu'aucun capteur n'est valide, on pose :

$$E(x / \{\emptyset\}) = x_0$$

où x_0 est une donnée a priori.

L'estimateur globale est fourni par la formule (6) ou l'on a remplacé les probabilités β_J par leurs valeurs données en (8), et les différentes estimées par leurs valeurs données en (9). Le résultat donne alors :

$$\hat{x} = P(c_\emptyset)x_0 + y_1[P(C_1) - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}P(C_1\cap C_2)] + y_2[P(C_2) - \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}P(C_1\cap C_2)]$$

Pour un opérateur de conjonction Min, et en supposant que les erreurs sur les mesures de z_1 et de z_2 sont négligeables, les probabilités sont alors :

$$\begin{aligned}
P(c_\emptyset) &= \text{Min}(1-\mu_1(z_1) \, ; \, 1-\mu_2(z_2)) \\
P(C_1) &= \mu_1(z_1) \\
P(C_2) &= \mu_2(z_2) \\
P(C_1\cap C_2) &= \text{Min}(\mu_1(z_1) \, ; \, \mu_2(z))
\end{aligned}$$

On peut ainsi étudier quelques situations particulières :

Cas 1 - Pour $\mu_1(z_1)=\mu_2(z_2)=1$, le contexte est favorable à l'utilisation simultanée des deux capteurs, l'estimée est alors :

$$\hat{x} = y_1 \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} + y_2 \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$$

Cette estimée est aussi obtenue par la moyenne conditionnée aux deux observations y_1 et y_2 et correspond à une approche probabiliste de la fusion sans que le contexte soit pris en considération.

Cas 2 - Pour $\mu_1(z_1)=0$ et $\mu_2(z_2)=1$, ces conditions invalident le capteur 1 alors que le capteur 2 est totalement valide, on obtient :

$$\hat{x} = y_2$$

Seule l'observation du second capteur est prise en compte. Ce résultat est conforme au souhait de voir le système ne prendre en compte que les mesures issues de capteurs valides dans le contexte considéré.

Cas 3 - Une situation intermédiaire est donnée par $\mu_1(z_1)=0.5$ et $\mu_2(z_2)=1$, qui invalident partiellement le capteur 1. L'estimée est alors :

$$\hat{x} = y_1 \frac{0.5\sigma_2^2}{\sigma_1^2+\sigma_2^2} + y_2 \frac{\sigma_1^2 + 0.5\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

On constate une pondération plus importante sur l'observation fournie par le capteur 2 par rapport à celle obtenue dans le cas 1 qui supposait alors les deux capteurs valides.

Cas 4 - Pour $\mu_1(z_1) = 0$ et $\mu_2(z_2)=0$. Ce cas limite correspond au cas où aucun des capteurs n'est valide. La seule estimée possible est celle fournie par la valeur a priori x_0.

3.1.4 Remarque

L'estimée $E(x / \{y_1,y_2\})$ de la formule (9) que l'on retrouve aussi dans le cas 1, est l'estimée fusionnée au sens probabiliste. On remarque que l'importance de chaque capteur dans le processus de fusion, qui n'est en faite qu'une simple moyenne pondérée, dépend de la variance du bruit d'observation. Plus la variance d'un capteur est forte, ce qui dénote une qualité moindre de la mesure, plus la valeur du coefficient qui pondère le second capteur est importante. Ce processus permet de privilégier dans la moyenne pondérée la mesure issue du capteur de plus faible variance.

3.2 ESTIMATION DYNAMIQUE

3.2.1 Formulation du problème

On note x_k le vecteur d'état à l'instant k. Le modèle représentant la dynamique du système est supposé ici linéaire, invariant, et d'équation :

$$x_k = F\, x_{k-1} + v_k$$

F est la matrice de transition du système. On suppose que v_k est un processus aléatoire gaussien de moyenne nulle et de matrice de covariance : $E(v_k v_j^T) = Q\, \delta(k,j)$. On suppose que les n équations d'observation sont linéaires, de la forme :

$$y_k^i = H_i\, x_k + b_k^i$$

avec $i \in N$. $H_1, \ldots, H_n$ sont les n matrices lignes d'observation. Lorsque les équations d'observation sont non linéaires ces matrices seront remplacées par les Hessiens ; les équations d'estimation résultantes seront alors celles du filtre de Kalman étendu. Les bruits d'observations $b_k^1, \ldots, b_k^n$ sont gaussiens de moyenne nulle et variance donnée par $E(b_k^i b_l^j) = R_i\, \delta(k,l,j)$.

L'ensemble des mesures fournies par le capteur i jusqu'à l'instant k est noté $Y_k^i = \{ y_l^i \}_{l=1}^{l=k}$ et l'ensemble de toutes les mesures pour tous les capteurs à l'instant k sera noté $Y_k = \{ Y_k^i \}_{i=1}^{i=n}$. De plus, pour toute partie $J \subseteq N$ on notera $Y_k^J = \{ Y_k^i \}_{i \in J}$ l'ensemble des mesures à l'instant k fournies par l'association de capteurs identifiés par J.

3.2.2 Équations de filtrage pour une association de capteurs donnée

Pour une association de capteurs dont les indices sont éléments de J, l'estimée au sens de l'erreur quadratique moyenne est donnée par la moyenne conditionnée :

$$\hat{x}_{k/k}^J = E(x_k / Y_k^J)$$

Les équations qui résultent de cette estimation sont les équations de Kalman adaptées à la fusion des données [4], [5] qui prend en compte un nombre d'équations d'observation supérieur à 1. L'estimée optimale à l'instant k est donc fournie par la relation :

$$\hat{x}_{k/k}^J = \hat{x}_{k/k-1}^J + \sum_{j \in J} K_J^j(k)\, (y_k^j - H_j\, \hat{x}_{k/k1}^J) \qquad (10)$$

où $K_J^j(k)$ est le gain de Kalman associé au capteur j appartenant au groupement J. L'écriture des équations de Kalman pour un système multicapteur est plus simple sous la forme information, et c'est cette forme que nous reprenons ici, sachant que la mise en œuvre effective du filtre se prête mieux à une programmation séquentielle que nous présenterons dans la partie algorithmie. Le gain de Kalman prend donc la forme :

$$K_J^j(k) = P_J(k/k) H_j^T R_j^{-1} \qquad (11)$$

La matrice des erreurs de prédictions a posteriori $P_J(k/k)$ est obtenue par la forme récurrente sur les inverses suivante :

$$P_J^{-1}(k/k) = P_J^{-1}(k/k-1) + \sum_{j \in J} H_j^T R_j^{-1} H_j \qquad (12)$$

Pour un système composé de n capteur le filtre de Kalman classique pour la fusion est basé sur les équations précédentes en remplaçant J par N.

3.2.3 Équations du filtre avec prise en compte du contexte

Nous allons considérer maintenant toutes les associations de capteurs J possibles avec $J \subseteq N$ et composer à partir de celles-ci une estimée globale. D'une façon générale, celle-ci s'écrit :

$$\hat{x}_{k/k} = E(x_k / Y_k)$$

La prise en compte du contexte s'effectue, comme dans le cas statique, par le développement de l'estimateur $\hat{x}_{k/k}$ suivant la formule :

$$\hat{x}_{k/k} = x_0 \beta_{\emptyset}(k) + \sum_{J \subseteq N} \beta_J(k) E(x_k / Y_k^J) \qquad (13)$$

L'estimée correspondant à la fusion globale est décomposée en fonctions des fusions partielles. En l'absence de mesure valide, x0 peut être avantageusement remplacé par la prédiction de l'état $\hat{x}_{k/k-1}$. Les fusions partielles $E(x_k / Y_k^J)$ sont remplacées par leurs valeurs données en (10), où l'on a substitué chaque prédiction $\hat{x}^J_{k/k-1}$ au sens de l'association J par la prédiction de l'état $\hat{x}_{k/k-1}$ fournie par la fusion globale à l'instant précédent. La relation (13) devient :

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + \sum_{J \subseteq N} \sum_{j \in J} \beta_J(k) \; K_J^j(k)(y_k^j - H_j \hat{x}_{k/k-1})] \qquad (14)$$

Cette dernière relation peut prendre une forme plus simple en adoptant la notation :

$$K_j(k) = \sum_{\{J / j \in J\}} \beta_J(k) K_J^j(k) \qquad (15)$$

où $\{J / j \in J\}$ est l'ensemble de tous les groupements de capteurs qui contiennent le capteur j, ce qui conduit à :

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + \sum_{j \in N} K_j(k) \; (y_k^j - H_j \hat{x}_{k/k-1})] \qquad (16)$$

La matrice de covariance a posteriori est obtenue grâce à la relation suivante [2] :

$$\hat{P}_{k/k} = \sum_{J \subseteq N \cup \emptyset} \beta_J(k)[\hat{P}_J(k/k) + (\hat{x}_{k/k} - \hat{x}^J_{k/k})(\hat{x}_{k/k} - \hat{x}^J_{k/k})^T] \qquad (17)$$

qui correspond à la matrice de covariance d'un mélange de lois gaussiennes. $\hat{P}_J(k/k)$ est la matrice de covariance a posteriori correspondant à chaque groupement de capteurs J.

Enfin, à partir de l'état et de la matrice de covariance, les équations de prédiction sont :

$$\hat{x}_{k+1/k} = F \hat{x}_{k/k} \quad \text{et} \quad \hat{P}_{k+1/k} = F \hat{P}_{k/k} F^T + Q \qquad (18)$$

3.2.4 Algorithmes

Bien que les équations précédentes puissent être directement utilisées, on préfère souvent la forme séquentielle [5] pour l'implantation d'un filtre de Kalman destiné à la fusion de données. D'une part, cette forme est plus facile à mettre en œuvre, et d'autre part, elle permet une extension immédiate aux traitement de capteurs asynchrones. L'algorithme est présenté sous les deux formes : sans prise en compte du contexte et avec prise en compte du contexte. Nous considérerons, pour les besoins de la présentation, un système composé de trois capteurs, mais la généralisation à un système composé de n capteurs ne pose aucune difficulté.

3.2.4.1 Algorithmes de fusion sans prise en compte du contexte.

Nous rappelons ici l'algorithme séquentiel du filtre de Kalman pour un système multicapteur synchrone [5]. Les équations d'état pour un système S composé de trois capteurs sont les suivantes :

$$S \begin{cases} x_k = F x_{k-1} + v_k & \text{Dynamique} \\ y_k^1 = H_1 x_k + b_k^1 & \text{Capteur 1} \\ y_k^2 = H_2 x_k + b_k^2 & \text{Capteur 2} \\ y_k^3 = H_3 x_k + b_k^3 & \text{Capteur 3} \end{cases}$$

Une forme équivalante de ce système d'équation est la suivante :

$$S1 \begin{cases} x_k^1 = F x_{k-1} + v_k & \text{Dynamique} \\ y_k^1 = H_1 x_k^1 + b_k^1 & \text{Capteur 1} \end{cases}$$

$$S2 \begin{cases} x_k^2 = x_k^1 & \text{Dynamique} \\ y_k^2 = H_2 x_k^2 + b_k^2 & \text{Capteur 2} \end{cases}$$

$$S3 \begin{cases} x_k = x_k^2 & \text{Dynamique} \\ y_k^3 = H_2 x_k + b_k^3 & \text{Capteur 3} \end{cases}$$

Cette décomposition suggère donc d'effectuer le traitement en quatre étapes successives : une étape est consacrée à la prédiction les trois autres sont des processus d'estimation. Le premier système d'équations noté S1 est à l'origine des deux premières étapes : la prédiction suivi du renouvellement de l'état par la donnée issue du capteur 1. Les deux étapes suivantes sont consacrées au renouvellement de l'état respectivement par la donnée du capteur 2 et celle du capteur 3.

Étape 1

À partir d'un état estimé à l'itération k-1, la prédiction s'effectue de façon classique par l'application des formules (18) ci-dessus.

Étape 2

La prédiction est ensuite mise à jour par les formules classiques suivantes :

$$\hat{x}^1_{k/k} = \hat{x}_{k/k-1} + K_1(k)\,(\,y^1_k - H_1\ \hat{x}^1_{k/k-1})$$

et

$$P^1(k/k) = (I - K_1(k)\,H_1)P_{k/k-1} \qquad (19)$$

avec :

$$K_1(k) = P_{k/k-1}\,H_1^T\,(\,H_1\,P_{k/k-1}H_1^T + R_1)^{-1}$$

Étape 3

Au vu du système S2, le processus de prédiction est simplement obtenu en identifiant l'état prédit par l'état estimé à l'étape précédente, et la matrice de covariance des erreurs a priori par la matrice des erreurs a posteriori obtenue à l'étape précédente. L'étape d'estimation est obtenue en substituant dans les équations (19) $P_{k/k-1}$ par $P^1(k/k)$, $\hat{x}_{k/k-1}$ par $\hat{x}^1_{k/k}$ et en remplaçant l'indice 1 par l'indice 2 sur les autres quantités.

Étape 4

Cette quatrième étape est rigoureusement identique à l'étape précédente. A la suite de cette étape il suffit se changer $P^1(k/k)$ par $P^2(k/k)$ et $\hat{x}^2_{k/k}$ par $\hat{x}^1_{k/k}$ et de remplacer l'indice 2 par l'indice 3 sur les autres quantités. L'estimée ainsi que la matrice de covariance des erreurs a posteriori issues de cette étape sont maintenant l'estimée et la matrice de covariance des erreurs a posteriori du système global S et sont notées $\hat{x}_{k/k}$ et $P_{k/k-1}$.

Dans la simulation qui va suivre, les résultats obtenus par l'algorithme qui vient d'être présenté seront notés xf.

Remarque

L'algorithme précédent donne des résultats rigoureusement identique à ceux résultant de la programmation des équations du filtre de Kalman sous la forme information présentées section 3.2.2. Par ailleurs, et en faisant le parallèle à la remarque 3.1.4, on remarque qu'à chaque étape la variance du bruit d'observation du capteur considéré est prise en compte dans le gain de Kalman. Cette procédure permet de tenir compte de l'importance des capteurs dans le processus de fusion. En effet, la matrice de covariance du bruit d'observation est différente d'un capteur à l'autre, celle-ci dépend notamment de la résolution du capteur et du rapport signal sur bruit. Plus cette matrice est importante, au sens d'un critère tel que la trace, moins les mesures du capteur correspondant auront de l'importance dans le processus global de fusion. Cependant, la hiérarchisation des capteurs en fonction d'un seul critère statistique n'est pas suffisante pour appréhender le processus de fusion dans sa globalité.

3.2.4.2 Algorithmes de fusion avec prise en compte du contexte.

L'algorithme proposé avec prise en compte du contexte est basé sur l'algorithme précédent, avec cette différence que l'on doit, pour un système multicapteur comportant n capteurs, déterminer 2^n-1 estimées, au lieu des n estimées dans la version précédente. Chaque estimée correspond à une association de capteurs. La combinatoire n'est pas ici un obstacle car, dans les systèmes actuels, n est de l'ordre de deux ou trois pour chaque coordonnée (site, azimut, gisement).

Nous distinguerons ici 5 étapes dans le traitement. La première étape est toujours dédiée à la prédiction. Les trois suivantes se distinguent par le nombre de capteurs inclus dans les différentes associations. La cinquième étape est celle qui permet de prendre en compte le contexte en attribuant à chaque association une probabilité, et en effectuant la moyenne pondérée des différentes estimées.

Étape 1

Comme précédemment cette étape effectue la prédiction. À partir des estimées acquises à l'itération précédente $\hat{x}_{k-1/k-1}$ et $P_{k-1/k-1}$ elle permet d'établir les prédictions $\hat{x}_{k/k-1}$ et $P_{k/k-1}$ à partir des équations (18).

Étape 2

Cette deuxième étape consiste à estimer, pour chaque capteur i, l'état $\hat{x}^i_{k/k}$ et la matrice de covariance $P^i_{k/k}$ à partir de chaque donnée y^i_k avec $i \in \{1,2,3\}$. Chaque vecteur d'état et sa matrice de covariance seront considérés ensuite comme des donnés a priori dans la seconde phase d'estimation.

Étape 3

On réalise ensuite la fusion partielle de toutes les paires de capteurs qu'il est possible de former dans le système. Dans le cas présent on en dénombre trois qui sont : {1,2}, {2,3}, {3,1}. L'estimée, que nous noterons $\hat{x}^{12}_{k/k}$, et sa matrice de covariance $P^{12}_{k/k}$, seront alors calculées en prenant comme état a priori $\hat{x}^1_{k/k}$ et $P^1_{k/k}$ et en

mettant à jour cet état grâce à la donnée y_k^2. Cette mise à jour tient compte des caractéristiques du capteur 2 notamment dans le calcul du gain et de la matrice de covariance. Ce processus est réitéré pour toutes les paires de capteurs considérés. La convention étant que le premier indice désigne l'état a priori et le second désigne le capteur effectuant la mise à jour.

Étape 4

Cette troisième étape est celle qui effectue la fusion globale des trois capteurs. L'état a priori est alors l'un des états résultant de l'une des fusions partielles des paires précédentes. L'état est ensuite mis à jour par la mesure du capteur dont l'indice n'est pas contenu dans la paire a priori. Par exemple la paire a priori peut être la paire {1,2} donnant l'état $\hat{x}^{12}_{k/k}$ et sa matrice de covariance $P^{12}_{k/k}$, la mise à jour est alors effectuée par la mesure issue du capteur 3 pour obtenir $\hat{x}^{123}_{k/k}$ et sa matrice de covariance $\hat{P}^{123}_{k/k}$. Il apparaît évident que l'ordre des indice n'a pas d'importance et que l'état désigné par $\hat{x}^{123}_{k/k}$ est identique à celui désigné par $\hat{x}^{321}_{k/k}$. L'état estimé $\hat{x}^{123}_{k/k}$ étant celui qui serait obtenu si le contexte n'était pas pris en compte, il est donc identique à celui obtenu par l'algorithme précédent si l'on considère que les états à priori sont les mêmes.

Étape 5

Cette quatrième phase permet de prendre en compte le contexte. A l'instant de l'estimation chaque variable contextuelle a une valeur déterminée qui permet d'évaluer la valeur de la fonction d'appartenance correspondante et ensuite d'en déduire l'ensemble des coefficients $\beta_J(k)$. Ces coefficients vont permettre d'évaluer l'état global $\hat{x}_{k/k}$ à partir de l'équation (10) ainsi que la matrice de covariance $P_{k/k}$ à partir de l'équation (17).

Dans la simulation qui va suivre, les résultats obtenus par l'algorithme qui vient d'être présenté seront notés xc.

Remarque

En plus du critère statistique, un second critère intervient dans l'attribution de l'importance de chaque capteur dans le processus de fusion. Ce second critère résulte d'une analyse contextuelle et mesure l'adéquation du capteur au contexte.

4. Simulation

4.1 Conditions de simulation

La simulation présentée ici a pour objet la fusion de mesures issues d'un système composé de trois capteurs : un radar de veille représenté par la donnée y_3, un radar de poursuite représenté par la donnée y_2, et une caméra infrarouge représenté par la donnée y_1. Deux des capteurs subissent des perturbations. Les mesures y_1 et y_2 issues des capteurs 1 et 2 ont, dans les conditions normales, les mêmes caractéristiques de bruit : celui-ci est gaussien, de moyenne nulle, et d'écart type $\sigma_1 = \sigma_2 = 1$. La mesure issue du capteur 3 a un bruit gaussien, de moyenne nulle, et d'écart type $\sigma_3 = 5$. Pour les capteurs 1 et 2 des perturbations accidentelles se rajoutent au bruit de mesure. Pour le capteur 2 le phénomène accidentel est un bruit blanc gaussien, se rajoutant au premier, de moyenne nulle et d'écart type 10. Par ailleurs on suppose connu, par la variable contextuelle $z_2(t)$, l'évolution de cette perturbation en fonction du temps ; celle-ci est donnée figure 2. Le second phénomène, attaché au capteur 1, est de type impulsionnel. La probabilité d'apparition d'une impulsion à l'instant t : I(t) est donnée par une loi binomiale de probabilité 0.15 et l'amplitude de la perturbation en fonction du temps est donnée par la formule :

$$pt(t)=z_1(t)*\sqrt{x^2}$$

où x est une variable aléatoire gaussienne de moyenne nulle et d'écart type égal à 10. La variable contextuelle $z_1(t)$ indiquant la présence d'une perturbation est donnée figure 2.

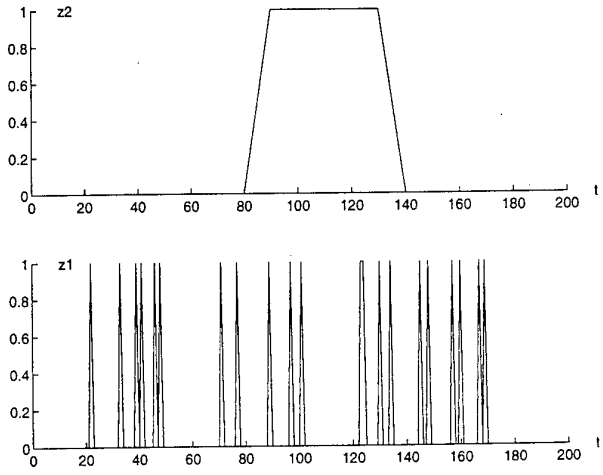


Figure 2

Variables contextuelles

z_2 est la variable contextuelle indiquant une perturbation sur le capteur 2. z_1 est la variable contextuelle indiquant une perturbation sur le capteur 1. Ces variables représentent la présomption qu'une perturbation affecte l'un des capteurs mais dont les caractéristiques sont inconnues.

Les signaux résultant de cette simulation sont présentés figure 3.

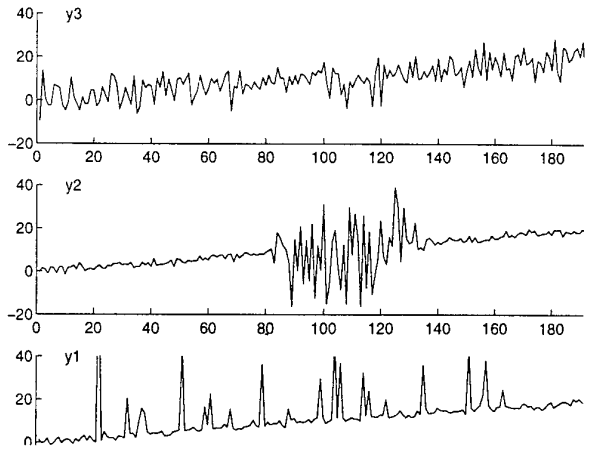


Figure 3

Signaux simulés.
y_3 : Radar de veille
y_2 : Radar de poursuite
y_1 : Caméra I.R.

4.2 Résultats

Les résultats des estimations sont présentés, figure 4, dans le cas d'une fusion sans prise en compte du contexte, xf, et dans le cas d'une fusion avec prise en compte du contexte, notée xc. L'estimée xf est calculée par la méthode séquentielle telle qu'elle est présentée dans [5].

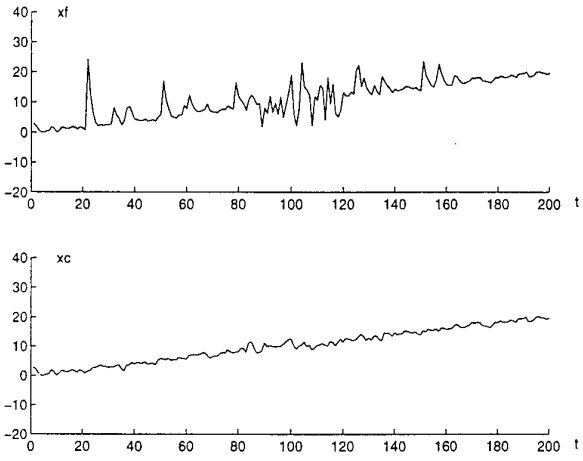


Figure 4
Résultats

xf est le résultat du filtrage par un filtre de Kalman adapté à la fusion de données mais sans exploiter l'information apportée par les variables contextuelles de la figure 2.
xc est le résultat obtenu par le filtre proposé qui tient compte des variables contextuelles de la figure 2.

On remarque que les perturbations affectent assez peu xc comparativement à xf. Ce constat est confirmé au vu des courbes d'erreur présentées figure 5. Aux instants où il n'existe aucune perturbation les deux courbes sont identiques comme c'est le cas au début et à la fin de la simulation ; ailleur l'erreur attachée à xc est presque toujours inférieure à celle associée à xf.

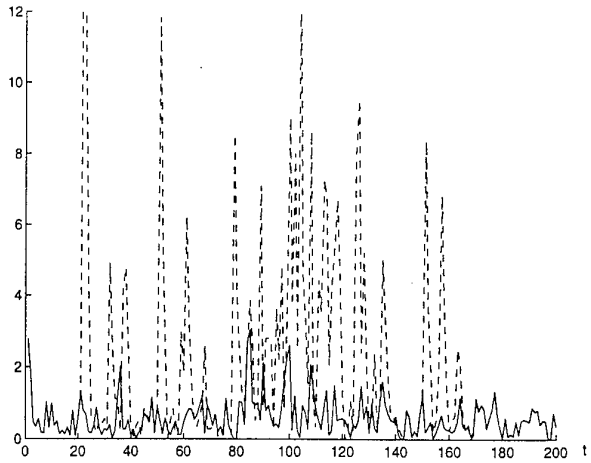


Figure 5

Erreur quadratique moyenne.
___ : erreur de xc
-. : erreur de xf

Globalement, l'erreur quadratique moyenne est de erf = 3,413 pour la fusion qui ne prend pas en compte le contexte et seulement de erc = 0,806 pour la fusion qui prend en compte le contexte. On constate que le gain est très net en faveur de l'approche proposée.

4.3 Remarques

Bien évidemment les deux algorithmes utilisés dans la simulation n'exploitent pas les mêmes informations et la comparaison des performances est de ce fait biaisée. Une comparaison avec des algorithmes calculant automatiquement les variances des signaux observés et qui les intègrent dans les gains de Kalman serait plus judicieuse et permettrait de comparer des algorithmes dont les fonctions sont plus proches l'une de l'autre. Toutefois, là encore, cette comparaison ne peut être que partielle puisque les sources d'informations ne sont toujours pas rigoureusement les mêmes. Dans un premier cas l'information ne provient exclusivement que des signaux issus des différents capteurs, dans le deuxième cas l'information est toujours composée des signaux des capteurs mais aussi de paramètres extérieurs ou bien de résultats d'heuristiques visant à évaluer la qualité de la mesure.

Comme nous l'avons remarqué toute la difficulté dans la définition d'un algorithme pour un système multicapteur réside dans la démarche qui permet d'accorder une importance à chaque capteur dans le processus de fusion.

Dans l'algorithme présenté cette importance est accordé au vu de deux critères. Le premier, d'ordre statistique, vise à évaluer la qualité du signal en terme de variance. Ce critère est déjà naturellement pris en compte dans le filtre de Kalman par le biais des matrices de covariance du bruit d'observation. Le second critère est l'adéquation du capteur au contexte et l'évaluation de son bon fonctionnement. Il peut subsister une certaine confusion entre ces deux critères puisqu'une non-adéquation du capteur au contexte peut entraîner une modification statistique du signal. Cela pourrait laisser à penser que les seuls paramètres statistiques sont suffisant pour évaluer l'importance du capteur dans le processus de fusion. Mais la relation, reliant la présence d'une perturbation à une modification des paramètres statistiques du bruit n'est pas une relation biunivoque, et un capteur peut délivrer des mesures d'une qualité statistiquement satisfaisante mais totalement aberrante au vu des objectifs suivis.

L'apport de la méthode proposée repose donc bien sur l'identification et l'utilisation de toutes les informations qui peuvent être disponibles au niveau du système pour évaluer l'adéquation du capteur au contexte, sachant que les caractéristiques du bruit sont déjà naturellement prises en compte dans le filtre de Kalman.

5. Conclusion

Nous avons proposé un algorithme de fusion de données permettant de tenir compte du contexte. Cette prise en compte est essentielle pour un système multicapteur car elle permet de ne sélectionner à tout instant que les mesures pertinentes et de réduire l'importance ou simplement d'exclure les mesures qui pourraient perturber le signal utile. Le contexte est donc analysé par un niveau de traitement que l'on pourrait qualifier de symbolique puisqu'il résulte d'une analyse préalable, par un expert, des différentes situations pouvant survenir dans l'utilisation du système et des moyens à mettre en œuvre face à celles-ci. Ce niveau de traitement permet de superviser des traitements classiques de traitement du signal qualifié de numérique. Cette approche permet de montrer la dualité qui existe entre les deux traitements et illustre la synergie qui peut en résulter. Outre la poursuite de cibles aériennes, nous pensons que de nombreuses applications peuvent bénéficier du formalisme développé dans cet article.

6. Annexe

Dans le cas général, la démonstration de la formule (2), issue de la formule de Poincaré, se fait par induction sur

la variable $k=|\bar{J}|$ et sur la variable $n = |N|$ avec $k \leq n$. Le principe est de démontrer que si la formule est vérifiée à l'ordre (n-1,k-1) et à l'ordre (n,k-1) alors elle est vérifiée à l'ordre (n,k). Sachant que la formule est vérifiée quelque soit n pour tous les couples (n,0), l'induction est complète pour tout couple (n,k) tel que $k \leq n$.

Par définition on a la relation :

$$P(c_J) = P(\cap_{i\in J} C_i \cap_{j\in \bar{J}} \bar{C}_j)$$

A l'évidence, la formule est vérifiée pour tout couple (n,0) en effet celle-ci s'écrie.

$$P(c_N) = P(\cap_{i\in N} C_i)$$

$$= \sum_{N\subseteq I} (-1)^{|I-N|} P(\cap_{i\in I} C_i)$$

Supposons que la relation soit vraie pour le couple d'indice (n-1,k-1) et pour le couple d'indice (n,k-1). Nous allons montrer qu'elle est vraie pour le couple d'indice (n,k).

Soit J tel que $|J| = n-k$ et $|\bar{J}| = k$, alors :

$$P(c_J)=P(\cap_{i\in J} C_i \cap \bar{C}_l \cap_{j\in \bar{J}-\{l\}} \bar{C}_j) \qquad (A1)$$

où l est un indice tel que $l \in \bar{J}$.

En utilisant la relation :

$$P(A\cap \bar{B}) = P(A) - P(A\cap B)$$

où A et B sont deux sous-ensembles quelconques, (A1) peut s'écrire :

$$P(c_J)=P(\cap_{i\in J} C_i \cap_{j\in \bar{J}-\{l\}} \bar{C}_j) - P(\cap_{i\in J\cup\{l\}} C_i \cap_{j\in \bar{J}-\{l\}} \bar{C}_j).$$

Cette relation pouvant aussi prendre la forme :

$$P(c_J)=\sum_{\substack{J\subseteq I \\ l\notin I}} (-1)^{|I-J|} P(\cap_{i\in I} C_i) - \sum_{J\cup\{l\}\subseteq I} (-1)^{|I-J|+1} P(\cap_{i\in I} C_i)$$

$$P(c_J)=\sum_{\substack{J\subseteq I \\ l\notin I}} (-1)^{|I-J|} P(\cap_{i\in I} C_i) + \sum_{\substack{J\subseteq I \\ l\in I}} (-1)^{|I-J|} P(\cap_{i\in I} C_i)$$

$$= \sum_{J\subseteq I} (-1)^{|I-J|} P(\cap_{i\in I} C_i),$$

l'induction est ici complète.

7. Références

[1] Appriou A. "Perspectives liées à la fusion de données" Science et défense 90, Dunod, Paris, Mai 1990.

[2] Bar-Shalom, Y., Fortman, T.E., Tracking and data association, New York: Academic Press, 1988.

[3] Bar-Shalom Y. "Multitarget - Multisensor Tracking: Application and Advances", Artech House, 1992.
[4] Chong C.Y. "Hierarchical Estimation", Proc. 2nd MIT/ONR Workshop on Distributed Communication and Decision Problems, Monterey, CA, June 1979.
[5] Chong C.Y., Mori S.,Tse E, Whisner R.P. "Distributed Estimation in Distributed Sensor Networks", Proc IEEE American Control Conference, Arlington, CA, June 1979.
[6] Haimovich A.M. "Fusion of Sensors with dissimilar Measurement/tracking accuracies", IEEE trans. on AES, Vol 29, Jan 1993.
[7] Houles A., Bar-Shalom Y. "Multisensor Tracking of a Maneuvring Target in Clutter", IEEE trans. on AES, Vol 25, March 1989.
[8] Nimier V. "Introduction d'informations contextuelles dans les algorithmes de poursuites multi-capteurs", IPMU 94, Paris 1994.
[9] Nimier V. "Introducing Contextual Information in Multisensor Tracking Algorithms", Advances in Intelligent Computing, Lecture Note in Computer Science, Springer, 1994.
[10] Roecker J.A., and McGillem C.D. "Comparaison of two-sensors tracking methods based on state vector fusion", IEEE trans. on AES, vol 24, July 1988.
[11] Shafer G. " A Mathematical Theory of Evidence ", Princeton University Press, 1976.
[12] Smets P. "probability of fuzzy event: an axiomatique approach", Fuzzy Sets Systems 7, 1982, pp 153-164.
[13] Zadeh L.A., "Fuzzy set", Information and control, vol 8, 1965.
[14] Zadeh L.A., "Probability Measures of Fuzzy Event", JMAA, vol 23, 1968.

PAPER No. 26

DISCUSSOR'S NAME: V. Nimier

COMMENT/QUESTION:

Les bruits de mesure de différents capteurs sont ils réajustés dynamiquement en fonction soit de l'environnement, soit des modes d'émission mis en oeuvre par les capteurs (formes d'ondes spécifiques).

(Is the measurement noise from different sensors readjusted dynamically either with respect to the environment, or to the transmission modes used by the sensors (specific waveforms).

AUTHOR/PRESENTER'S REPLY:

Les matrices des bruits d'observation des capteurs peuvent être modifiés automatiquement en fonction des modes d'émission. L'environnement est pris en compte par le haut niveau.

(The observation noise matrices of the sensors can be modified automatically to suit the transmission mode. The environment is allowed for by the high level).

SAGESSE : un modèle de représentation de donnéespour la Fusion de Données Symboliques

Yann Pollet [1], Sébastien Robidou [1] [2]

(1) MATRA SYSTEMES & INFORMATION
Direction des Filières Technologiques
Parc d'Affaires des Portes
BP n°613
27106 Val de Reuil Cedex

(2) Laboratoire PSI/LIR
Institut National des Sciences Appliquées
Place Emile Blondel
76130 Mont Saint Aignan Cedex

Tél : 02 32 63 40 45 Email : pollet@mcs-vdr.fr
Tél : 02 32 63 40 00 Email : robidou@mcs-vdr.fr

Abstract : un système de fusion de données symboliques a pour but d'élaborer et d'entretenir à tout instant une image la plus complète, exacte et cohérente possible d'un monde extérieur en évolution, appelée *situation*, à partir d'informations d'entrée incertaines, incomplètes et imprécises. La modélisation d'une *situation* nécessite la représentation de la notion d'imprécision sur les caractéristiques des objets perçus, la notion de doute quand à l'existence d'un objet, mais aussi la représentation de l'information hypothétique utilisée au cours du raisonnement. Nous présentons dans cette communication une approche formelle permettant d'intégrer de manière unifiée ces différentes notions dans un même modèle de représentation, en les dotant d'une sémantique précise. Nous décrivons ensuite l'application faite de cette approche et l'implémentation pilote réalisée dans le cadre du programme international DFD (Data Fusion Demonstrator).

Mots clés : Systèmes d'Information et de Communication, Bases de Données, Théorie des Possibilités, Fusion de données symboliques, représentation des connaissances, approche objet.

1. La fusion de données symboliques.

La fusion de données est de manière générale le processus visant à élaborer et à entretenir l'"image" la plus fidèle, complète et précise possible d'un univers extérieur évolutif à partir d'un flot d'informations partielles, incertaines et bruitées. Nous considérons ici la fusion de données dans le domaine du champ de bataille (détermination et évaluation des forces ennemies en présence), les problématiques soulevées pouvant s'appliquer plus généralement à une large classe de contextes tels que sécurité civile, gestion de crise, maîtrise des risques industriels, etc.

Les multiples informations en entrée peuvent être de nature ***symbolique*** (observations humaines, informations déjà partiellement fusionnées issues d'autres systèmes), ou ***numérique*** (mesure effectuée par un dispositif physique *capteur*, tels que radar, imagerie visible, écoute radioélectrique, etc.).

De manière générale, les sources d'information sont **hétérogènes** (elles reflètent différents phénomènes physiques partiellement indépendants) et **asynchrones** (elles délivrent des informations à des instants non nécessairement prédéterminés). Elles ont par ailleurs des **champs de vision limités** (elles ne perçoivent qu'une certaine partie du monde extérieur à un instant donné), et délivrent des **informations incertaines et imprécises** (ce sont des dispositifs physiques imparfaits par nature). L'image générée (en constante évolution) du monde extérieur sera dite *situation perçue*.

Dans le fusion de données dite *symbolique* à laquelle nous nous intéressons ici, on suppose que tous les signaux en entrée ont été préalablement transformés par le biais de traitements convenables en données symboliques.

Un processus de fusion peut être mené de manière manuelle, semi-automatique (avec l'assistance de traitements informatiques), voire complètement **automatique**[1]. Ce dernier cas fait l'objet du programme international de recherche DFD (Data Fusion Demonstrator) dans lequel se sont inscrits les travaux ici présentés. Dans ce cadre, un problème essentiel a alors trait à la modélisation de la *situation perçue*, et de sa dynamique au cours du temps. Ce problème fait l'objet de la présente communication.

Nous précisons tout d'abord les besoins de représentation de données incertaines, imprécises et hypothétiques dans le cadre de la fusion de données symboliques. Nous présentons ensuite une approche possibiliste proposée pour la représentation de la

[1] Les traitements de fusion s'appuient sur un ensemble de *connaissances a priori* reflétant les invariants et la dynamique des entités du domaine considéré. Dans le domaine ici considéré, on trouvera la cartographie numérique, les performances d'équipements ennemis, les ordres de batailles, les doctrines ennemies, etc.

Mentionnons enfin que le processus de fusion peut être distribué sur plusieurs sous-systèmes indépendants (travaillant en cascade et/ou en parallèle), appelés ici *noeuds de fusion*, au travers desquels transite une information de plus en plus en plus enrichie. On considérera ici non pas le processus complet mais ce qui se passe au sein d'un *noeud*.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

connaissance imparfaite, les notions de *mesures* et de *quantités fonctionnelles*, proposée pour lever les limitations de celle-ci, et enfin le modèle de données SAGESSE construit en réponse à notre problématique[2]. Enfin, nous décrivons l'application faite du modèle dans un noeud de fusion et l'implémentation réalisée.

2. Problématique.

Dans le domaine du champ de bataille, une *situation perçue* sera la description d'un ensemble d'entités évoluant au cours du temps. Dans ce cadre, il est nécessaire de disposer d'un modèle permettant la représentation de ***collections d'objets***, rendant compte simultanément de différents aspects [Bou 92], [Pol 93, 96] :

- **imprécision** sur une valeur **numérique**. Ex :
  - *"l'unité X est à 20km à 10%près. Son effectif est de 18 à 20 hommes."* *(cas 1)*

  ou symbolique. *Ex :*
  - *"l'unité Y est un Bataillon ou bien un Régiment de blindés."* *(cas 2)*
- **incertitude** sur l'existence d'une entité. Ex :
  - *"L'observation du capteur C laisse présumer l'existence d'un véhicule en (x, y)."* *(cas 3)*

De plus, la réalité opérationnelle nécessite d'opérer un certain type de fusion dite ***au fil de l'eau***, où tout nouvel *indice* est intégré immédiatement et "au mieux" dans la situation courante, c'est a dire dans l'image que se fait le système du monde extérieur à cet instant[3]

Dans cette hypothèse, que nous adopterons dans toute la suite, l'intégration d'un *indice* consiste idéalement à rechercher dans la situation courante l'entité de la *situation perçue* avec laquelle l'indice entrant doit être *corrélé*, puis à agréger la nouvelle description partielle d'une entité avec un état plus ancien et déjà connu de celle-ci.

Le caractère à la fois **incertain** (une information peut se révéler fausse) et **imprécis** (cas par exemple de positions et vitesses observées) de la donnée d'entrée, jointe aux déplacements inévitables de l'ennemi entre deux observations entraînent de nombreuses ambiguïtés dans l'intégration d'un indice. Ceci nécessite d'effectuer des **hypothèses** successives au cours des traitements, qu'il s'agit alors de gérer dans la représentation de la situation. Une hypothèse peut bien sur dépendre d'une hypothèse précédente en liaison avec une démarche de type hypothético-déductif.

Dans la fusion dite "**temps critique**" ici considérée, les retours arrières dans le raisonnement sont exclus, ce qui impose de représenter **et gérer en parallèle plusieurs description alternatives de la réalité**.[Pol 96]. Un besoin supplémentaire consiste donc à rendre compte du caractère **hypothétique** d'une description (avec expression d'hypothèses concurrentes ou dépendantes). Ex :

> *"Il est possible de corréler l'observation << unité vue en (x,y)>> avec l'objet de situation o_1 ou encore avec l'objet o_2.*
>
> - *dans le premier cas, l'état V_1 de O_1 devient V'_1 (et l'état V_2 de O_2 reste inchangé).*
> - *dans le second cas, c'est V_2 qui est changé en V'_2."* *(cas 4)*

La pertinence d'une approche possibiliste plutôt que probabiliste pour représenter de telles descriptions imparfaites est aujourd'hui largement admise [Bou 92]. Dans ce cadre, la modélisation de l'incertitude et de l'imprécision a été largement étudiée à travers la notion de sous-ensemble flou [Zad 65], et la théorie des possibilités [Zad 78], [Dub 88]. On trouve de nombreuses approches pour la mise en oeuvre de ces concepts au sein de bases de données relationnelles [Tes 84], [Pra 84, 86], [Pet 96] et dans une représentation objet [Zic 90], [VGy 93], [Sad 94]. Nous avons pour notre part déjà proposé de tels modèles dans le cadre particulier de la fusion de données symboliques [Pol 93a, 94a, b, 95].

Nous nous proposons ici d'étendre les modèles déjà envisagés en introduisant tout d'abord une modélisation possibiliste de situation rendant compte de l'existence incertaine d'une objet. Nous présenterons ensuite une seconde extension permettant la représentation d'hypothèses concurrentes au sein d'une approche formelle, basée sur la notion de *mesure fonctionnelle*, ce qui permettra de manipuler (créer, repondérer, détruire) les descriptions hypothétiques.

3. Un modèle de représentation possibiliste de situation.

3.1. Les *extensions* "nettes".

On considère ici les *situations* dont la structure "idéale" (celle du monde réel que l'on veut approcher) peut être représentée par une base de données modélisée par un ensemble :

$$B = \{ (S_i, E_i) ; i = 1, ...n\}$$

[2] Les traitements opérés sur les données n'étant pas considérés ici.

[3] Par opposition à une fusion a posteriori où l'on attendrait d'avoir collecté l'ensemble des indices afférents à une certaine tranche de temps dans une optique d'optimisation globale).

où S_i et E_i définissent respectivement l'*intention*[45](structure attributaire) et l'*extension* (ensemble des instances) d'une classe C_i. Les types de structures ici considérés sont les agrégations (produits cartésiens) de types élémentaires pris parmi entier, réel, types énumérés, *référence* à un autre objet[6].

Ex : Unité = (type_unité : énumération (Div, Rég, Esc), x : reel, y : reel, Vitesse : reel)

Une *extension* E_i sera une *collection* d'instances $\{<o_j, V_j>\}$, où :

- o_j est un identifiant d'objet (*oid*) issu d'un ensemble O;
- V_j est un vecteur de valeurs du domaine défini par E_i(espace V).

Ex : E = {<"1", (Régiment, 50km 60km, 40km/h>, <"2", (Division, 20km,30km, 20km/h>}

La seule contrainte d'intégrité considérée ici est celle d'*identité* d'objet, soit :

$$V_1 \neq V_2 \Rightarrow o_1 \neq o_2$$

Un telle extension idéale est donc ici une application de O dans un ensemble V[7]. L'ensemble considéré est ici V exp O.(notation pour V^O). Une telle extension sera appelée *extension nette.*

Il s'agit maintenant d'introduire l'imperfection de la connaissance dans une telle description. Celle-ci peut porter sur les **valeurs** associées à un objet, ou encore se traduire par un doute sur l'**existence** de celui-ci, ou plus généralement sur l'existence d'un sous-ensemble de la description.

Nous rappelons dans la suite comment peuvent être introduits dans un tel modèle l'imprécision (*valeurs floues*) d'une part et le doute existentiel (*extensions floues*) d'autre part. Nous présentons ensuite la généralisation de ce modèle à la représentation combinée de l'imprécision et du doute existentiel (*extensions floues*).

Ce modèle sera ensuite étendu à la représentation hypothétique (*extensions hypothétiques*).

[4]On peut construire $\mathbb{J}$ en considérant l'ensemble $[-1, 1]^n$ muni de la distribution de possibilités produit résultant des distributions partielles $\mu_i(\omega_i) = 1 - |\omega_i|$ (dimention finie), ou encore l'espace des suites infinies à valeur dans [-1, 1] (dimention dénombrable).

[5] Nous emploieront aussi dans la suite le terme de *domaine de valeurs*, plus explicite que celui d'*intention*.

[6] Cette vision ignore pour l'instant les notions propres aux bases de données telles que contraintes d'intégrité, définition de méthodes d'accès, de triggers, etc. qui sont ici hors de notre propos.

[7] convenablement étendu avec un extra-élément représentant l'absence d'objet pour un oid donné.

3.2. Les valeurs *floues.*

La notion d'imprécision dont est entachée la perception de la réalité peut être introduite au niveau des valeurs avec la notion d'*ensemble flou* sur les *domaines* de celles-ci.

On considère alors des extensions de la forme $\{<o_j, V_j^*>\}$, où V_j^* note un ensemble flou sur V [Pol 93]. L'espace ici considéré est alors classiquement ([0, 1]) exp V) exp O.

On pourra ainsi représenter des état d'objets dont un exemple est :

<o, (type_unite={Div, Reg}, x=[49, 51], y=[59, 61], vitesse=T(18, 19, 21, 22)>

où T(18, 19, 21, 22) note la distribution de possibilités trapézoïdale de support [18, 22] et de noyau [19, 21]).

3.3. L'existence incertaine.

Nos besoins imposent de pouvoir représenter l'incertitude au niveau de l'existence même d'un objet ou d'un ensemble d'objets dans une *extension.* Pour représenter un tel *doute existentiel*, il faut maintenant considérer des ensembles flous dont les éléments sont non plus des valeurs, mais des *collections d'objets* (soit des éléments de [0,1]exp({0,1}expO)). On peut doter cet ensemble d'une **structure algébrique** basée sur deux lois de composition internes "+" (alternative) et "$_*$" (union généralisée), et sur un produit externe "/", permettant la construction et la mise à jour de telles collections [Pol 94b, 95a]

On pourra par exemple représenter des composantes de *situation perçue* de la forme :

$$S = 1/(\{o_1\}*\{o_2\}) + 0.8/\{o_1\}$$

qui exprime l'alternative entre deux situations réelles possibles : l'une où o_1 et o_2 existent (possibilité 1), l'autre où seul o_1 est présent dans la réalité (possibilité 0.8).

Une expression de la forme

$a/\{o\} + b/\varnothing$, où ($a<1$ et $b = 0$) ou ($a=1$)

qui représente un objet d'existence incertaine, sera dite binôme *canonique* (un tel binôme est idempotent). Une forme particulière d'extension, dite *totalement factorisée*, sera un produit de la forme:

$$S = P_i\ (a_i/\{o_i\} + b_i/\varnothing)$$

où tous les o_i sont distincts.

Dans une telle représentation, des valeurs V peuvent bien sûr être associées aux objets. Toutefois, cette approche ne permet d'associer qu'un vecteur de valeur unique, ou un vecteur flou, à un objet donné. Ceci n'est pas suffisant dans la mesure où la description prêtée à un même objet peut dépendre de l'hypothèse considérée (cas 4 ci-dessus).

Nous proposons ci-dessous une modélisation plus riche intégrant la représentation simultanée de l'imprécision sur les valeurs et du doute existentiel.

3.4. Les *extensions* floues.

Une *extension floue* sera un ensemble flou sur l'ensemble des *extensions nettes*, soit un élément de l'ensemble B* = [0, 1] exp (V exp O).

A l'espace V des valeurs, on peut associer l'espace V* = [0, 1] exp V des ensembles flous sur V. V* peut être muni des lois de composition internes + et $*$, et d'un opérateur externe "/", définis par[8] :

- $V^* = V_1^* + V_2^*$: $m_{V^*}(v) = \text{Max}(m_{V1}(v), m_{V2}(v))$ (union)
- $V^* = V_1^* * V_2^*$: $m_{V^*}(v) = \text{Min}(m_{V1}(v), m_{V2}(v))$ (intersection)
- $V^* = \lambda / V_1^*, \lambda \in [0, 1]$: $m_{V^*}(v) = \text{Min}(\lambda, m_{V1^*}(v))$

Ces lois permettent de munir B* de lois de composition internes +, $*$ et d'un produit externe /, définis, si $B_1^* \in B^*$, $B_2 \in B^*$, $\lambda \in [0, 1]$, $B \in B$. par :

- $B^* = B_1^* + B_2^*$: $m_{B^*}(B) = \text{Max}(m_{B1}(B), m_{B2}(B))$ (opérateur alternative généralisé)
- $B^* = B_1^* * B_2^*$: $m_{B^*}(b) = \text{Sup}_{b \in D\,(bi,\,bj)} \text{Min}(m_{B1^*}(b_i), m_{B2^*}(b_j))$

  où D est la fonction associant à tout couple $(b_1, b_2) \in B^2$ d'*extensions nettes* l'ensemble d'*extensions nettes* ($\in \{0, 1\}$ exp B) défini par D $(b_1, b_2) = \{EN(b_1\ D\ b_2)\}$ où EN (b) représente toute extension nette dont le graphe fonctionnel est inclus dans b. En d'autres termes, un élément apparaît dans $B_1^* * B_2^*$ s'il apparaît dans B_1^* et pas dans B_2^*, ou vice-versa. Sinon, il apparaît dans $B_1^* * B_2^*$ avec l'ensemble des valeurs portées dans B_1^* ou dans B_2^*, excluant les valeurs communes (ou exclusif).
- $B^* = \lambda / B_1^*$: $m_{B^*}(b) = \text{Min}(\lambda, m_{B1^*}(b))$

On vérifie un certain nombre de propriétés généralisant celles mises en évidence dans les approches précédentes :

- + et $*$ sont commutatives, associatives et possèdent chacune un élément neutre (distribution nulle et $\varnothing$);
- * est distributive par rapport à +;
- on peut immerger l'ensemble des *extensions nettes* muni de la loi $\cup$ dans (B*, $*$). Ce résultat permettra en particulier de définir des opérateurs généralisant sur les extensions floues les opérateurs *insert*, *update*, *delete* et *select* des bases de données.

On vérifie également les identités remarquables :

- $\lambda_1/\langle o_1, V_1\rangle + \lambda_2 / \langle o_1, V_2\rangle = (\lambda_1+\lambda_2) / \langle o_1, V_1+V_2\rangle$
- $\lambda_1/\langle o_1, V_1\rangle * \lambda_2/\langle o_1, V_2\rangle = (\lambda_1.\lambda_2) / \langle O_1, V_1 * V_2\rangle$

  où $\lambda_1+\lambda_2$ et $\lambda_1.\lambda_2$ notent respectivement $\text{Max}(\lambda_1, \lambda_2)$ et $\text{Min}(\lambda_1, \lambda_2)$
- plus généralement

  $\Sigma_i\ \lambda_i/\langle o, V_i\rangle = \langle o, \Sigma\ \lambda_i/V_i\rangle$

  et :

  $(\Sigma_i\ \lambda_i/b_i) * (\Sigma_j\ \mu_j/b_j\rangle = \Sigma_i\ \Sigma_j\ (\lambda_i\ .\ \mu_j)\ b_1 * b_2$

Si une valeur floue V* n'est pas *normalisée*, et si elle apparaît dans un objet e sous la forme $e = \lambda/\langle O, V^*\rangle$, alors on peut la remplacer (en notant $||V|| = \text{Sup}\{m_{V^*}(v)\}$) par :

- $e = \lambda/\langle O, V'^*\rangle$, si $1 < ||V||$, où $V'^* = \lambda \hat{} V^*$ est défini par :
  - $m_{V'^*}(v) = m_{V^*}(v)$ si $m_{V^*}(v) \leq 1$
  - $m_{V'^*}(v) = 1$ sinon
- $e = (\lambda.||V||)/\langle O, V'^*\rangle$, si $1 > ||V||$, ce qui ramène au cas précédent.

ce qui donne dans tous les cas l'identité :

$\lambda/\langle O, V^*\rangle \equiv (1.||V||)/\langle O, \lambda \hat{} V^*\rangle$

qui permet de ne considérer que des valeurs *normalisées*.

L'exemple d'*extension floue* ci-dessous :

$S = 1/(\langle o_1, V_1\rangle * \langle o_2, V_2\rangle) + 0.8/(\langle o_1, V'_1\rangle)$

permet d'exprimer l'alternative entre deux situations réelles possibles : l'une où o_1 a l'état V_1 et o_2 a l'état V_2 (possibilité 1), l'autre où l'état V_1 de o_1 a été modifié en V'_1 et où o_2 n'existe pas (possibilité 0.8). Nos résultats ci-dessus permettent d'opérer des **mises en facteur** de S telles que :

$S = \langle o_1,\ V_1.V'_1\rangle\ *\ (1/(\langle o_1,\ V_1\rangle * \langle o_2,\ V_2\rangle) + 0.8/(\langle o_1, V'_1\rangle)$

qui mettent en évidence la partie "certaine" de la connaissance. Si $V'_1 \subseteq V_1$ (V'_1 "précise" V_1), S se réduit à

$S = \langle o_1,\ V'_1\rangle\ *\ (1/(\langle o_1,\ V_1\text{-}V'_1\rangle * \langle o_2,\ V_2\rangle) + 0.8/\varnothing)$

Une forme particulière simple d'*extension*, dite *totalement factorisée*, sera un produit :

$S = P_i\ (a_i/\langle o_i, V_i\rangle + b_i/\varnothing)$

où tous les o_i sont distincts.

Un binôme de la forme $a/\langle o, V\rangle + b/\varnothing$, où ($a<1$ et $b = 0$) ou ($a=1$) décrit un objet d'existence

[8] Les opérateurs + et $*$ ont avant tout un rôle formel et ne doivent pas être vus comme des opérateurs de fusion (par contre ces derniers pourront s'exprimer comme des combinaisons des premiers jugées pertinentes dans un contexte particulier).

incertaine. Le cas particulier de $a_i = 1$, $b_i = 0, \forall i$ est celui d'une *extension nette*.

On démontre par ailleurs qu'une bijection canonique permet d'opérer l'identification :

$$[0, 1] \text{ exp } (V \text{ exp } O) = ([0, 1] \text{ exp } V) \text{ exp } O) = [0, 1] \text{ exp } (V \times O)$$

qui autorise à considérer indifféremment une *extension floue* comme :

- un ensemble flou sur l'ensemble des *extensions nettes*;,
- une fonction associant à tout identifiant une valeur floue V*,
- un ensemble flou de couples (identifiant, valeur>.

Notre formalisme nous permet **généraliser les opérateurs classiques des bases de données**. Ainsi, les opérateurs classiques d'*insertion*, *mise à jour*, et *destruction* sont-ils des cas particuliers de l'affectation $S' = S * <o, V^*>$:

- ***insertion*** : $S' = S * <o, V^*>$, où o n'est pas dans B;
- ***mise à jour*** (à prendre au sens d'une réduction d'états[9]) : $S' = S * <o, V^*>$, où o est déjà dans S avec une "valeur" $V'^* \subseteq V^*$;
- ***destruction*** : $S' = S * <o, V^*>$, où o est déjà dans S avec la "valeur" V*.

La mise à jour avec changement de valeur (***correction***)[10] peut être décrite par :

$$S' = S * <o, V^*> * <o, V'^*> = S * <o, V^* \Delta V'^*>$$

La ***sélection*** pour sa part est exprimée par l'opérateur $s_{V'}$ généralisant la restriction ($V'^* = (..., v'^*_i, ..)$ est un vecteur de valeurs floues définissant un critère de restriction), défini par :

- $s_{V'}(P_i < o_i, V^*_i >) = P_i s_{V'}(< o_i, V^*_i >)$
- $s_{V'}(S_i < o_i, V^*_i >) = S_i s_{V'}(< o_i, V^*_i >)$
- $s_{V'}(< o, V^* >) = (V^*.V'^*)/< o, V^* > + (V^* \perp V'^*)/\varnothing$

où $V^*.V'^* = \text{Sup}_i (v^*_i \text{ min } v'^*_i)$,
et $V^* \perp V'^* = \text{Inf}_i (v^*_i \text{ max } (1 - v'^*_i))$

4. Un modèle de représentation d'objets hypothétiques.

4.1. Les limites de l'approche possibiliste.

La modélisation précédente permet de représenter des descriptions alternatives d'une situation réelle imparfaitement connue. Dans la fusion au fil de l'eau, une *situation perçue* est construite par raffinement successifs à partir des indices reçus qui viennent conforter ou infirmer telle ou telle éventualité formulée. Il est donc nécessaire de faire évoluer les confiances accordées aux hypothèses. De même, il faut rendre compte des liens de dépendance éventuels existant entre celles-ci, de manière à pouvoir retirer des éléments de connaissance insérés à tort. Dans ce cadre, l'approche proposée, qui ne représente pas la notion de dépendance/indépendance entre hypothèses n'est pas suffisante.

En effet, considérons par exemple une première *situation perçue* $S = S = <o_1, V^*_1>$ où l'on tient pour certaine l'existence d'une entité o_1 à l'état (flou) V^*_1[11]. Supposons que l'arrivée d'un nouvel indice laisse présumer l'existence (incertaine) d'une unité (nécessité 0.2) qui peut être (après calcul de corrélation) l'unité déjà perçue (possibilité 1), ou une nouvelle (possibilité 0.5). Un calcul simple donne :

$$S' = 1/<o_1, V'_1> + 0.5/(<o_1, V_1> * <o_2, V_2>) + 0.2/<o_1, V_1> \quad (1)$$

que l'on peut écrire aussi : $1/<o_1, 1/V'_1+0.2/V_1> + 0.5/(<o_1, V_1> * <o_2, V_2>)$

Si aucune nouvelle observation ne vient conforter la nouvelle "piste" o_2 a l'issue d'un temps déterminé, alors le processus de fusion peut décider de diminuer la confiance en son existence (possibilité 0.8 au lieu de 1). L'état de la connaissance souhaité est alors :

$$S'' = 0.5/(<o_1, V_1> * <o_2, V_2>) + 0.8/<o_1, V'_1> + 1/<o_1, V_1> \quad (2)$$
$$= 0.5/(<o_1, V_1> * <o_2, V_2>) + 1/<o_1, 0.8/V'_1+1/V_1>$$

Le passage de la forme (1) à la forme (2) requiert la résolution d'un système d'équations en min/max qui n'admet en général pas de solution unique. On intègre en effet dans un réel unique des éléments de provenances différentes et dont la mémoire n'est pas conservée. Il n'est donc pas possible d'effectuer à partir de (2) la transformation souhaitée.

Nous présentons dans la suite la notion de *mesure fonctionnelle* permettant de lever ces limitations, puis les *extensions hypothétiques* construites sur cette base.

4.2. Mesures fonctionnelles, quantités hypothétiques.

Nous présentons brièvement ci-dessous la notion de *quantité hypothétique* [Pol 95b], [Rob 96]. Nous utilisons pour cela une notion de *mesure fonctionnelle* apparentée à celle de *L-fuzzy set* de [Gog 67], et s'inscrivant dans une **modélisation possibiliste**.

---

[9] Consécutive à un gain d'information pour mieux décrire l'état au même instant du monde réel.

[10] La nouvelle valeur est indépendante de l'ancienne. Par exemple pour décrire l'état à un instant ultérieur.

[11] L'état d'un objet contient ici en particulier sa position géographique.

On appellera *mesure fonctionnelle* F sur un domaine de valeurs D une application de P(D), non plus dans [0, 1], mais plus généralement dans un treillis (T, +, *), et vérifiant :

$$F(\varnothing) = 0 ; F(D) = 1$$
$$\forall P_1 \in P(D), \forall P_2 \in P(D), F(P_1 \cup P_2) = F(P_1) + F(P_2)$$

On vérifie dans ce cas que $P_1 \subseteq P_2 \Rightarrow F(P_1) \leq F(P_2)$. Une *mesure fonctionnelle* F peut être complètement déterminée par l'application f de D dans T, dite *distribution fonctionnelle*, définie par f(a) = F({a}) (moyennant une hypothèse de continuité pour les domaines de cardinal infinis, non détaillée ici).

On considère ici le cas particulier où T est un sous-ensemble donné de $P(\Omega)$[12], stable pour l'union et l'intersection. Les opérateurs +, * et . ≤ sont alors respectivement ∪ , ∩ et Í (0 et 1 sont respectivement ∅ et Ω). T ne sera alors en général **pas totalement ordonné**.(contrairement à [0,1]).

Si Ω est un singleton, P(Ω) se réduit à {∅,Ω} (isomorphe à {0,1} muni des opérateurs *ou* et *et*), et F est alors une *valeur de vérité* classique. Si T est l'ensemble des sous-intervalles de [0, 1] de la forme [0, l], muni des lois ∪ et ∩, alors F est une *mesure de possibilités*. (∪ et ∩ sont alors le max et le min). Nous supposerons dans la suite que T est stable pour la complémentation (algèbre de Boole). F(A) sera appelé *support de vérité* de A Les éléments de T, qui généralisent la notion de valeur entre 0 et 1, seront appelés *pseudo-scalaires*.

Si on considère par ailleurs une *mesure de possibilités* P sur Ω, on vérifie que P' = P ∘ F est une *mesure de possibilités* sur P. Si H Î B, on notera dans la suite P(H) = |H|. Cette quantité sera appelée **possibilité de l'hypothèse H.**

Une *mesure fonctionnelle* sur D permettra de représenter **un doute sur la vérité d'une information** tout en conservant **la notion de *source* dont celle-ci est issue**, reflétant donc à tout instant un état **potentiellement affinable** de la connaissance acquise. Il est alors possible en particulier de repondérer les confiances associées aux sources de données, la *mémoire* du calcul d'une quantité étant conservée dans la quantité elle-même.

Deux *pseudo-scalaires* H_1 et H_2 seront dits *indépendants* ssi $|H_1 . H_2| . = |H_1| \min |H_2|$. (on écrira $Ind(H_1, H_2)$). Cette notion d'indépendance rend ici compte de l'idée de **sources de données** issues de **phénomènes physiques non interactifs**. Une famille de p *pseudo-scalaires* $H_1, ..., H_p$ sera dite *indépendante* ssi tous les H_i sont *indépendants* deux à deux. Pour T donné, le nombre maximum n (s'il existe) d'éléments *indépendants* sera dit *dimension hypothétique* de T.(dim(T)) Une famille *indépendante* de T de cardinal n sera dite *base hypothétique* de T. Le cas de [0, 1] et d'une *mesure de possibilité* définit la dimension 1.

On définit par ailleurs les notions suivantes :

- une *mesure* F sera dite *propre* ssi $A \cap B = \varnothing \Rightarrow F(A) \cap F(B) = \varnothing$. Elle sera dite sera *totale* ssi $\cup_{Ai \in P} F(A_i) = \Omega$ Une mesure à la fois *propre* et *totale* peut être définie par une application inverse $f : \Omega \dashrightarrow \{0, 1\}$ associant à tout $\omega \in \Omega$ une valeur de D.
- Ω sera dit *dense* ssi $\forall H \in T$ tel que $|H| = \lambda \in [0, 1]$, $\forall \mu < \lambda$, $\exists H'$ tel que $|H'| = \mu$, avec Ind(H', H'') $\forall H'' \in T$ tel que Ind(H, H''). Le plus grand élément (au sens de ⊆) vérifiant cette propriété pour H donné sera noté λ/H.(on montre que ce plus grand élément existe)
- Ω sera dit *normalisé* ssi p ≤dim(T), il existe une famille *indépendante* $\{H_i\}$ de cardinal p telle que $|H_i| = 1 \; \forall i$.

On supposera dans la suite que ces propriétés sont vérifiées. Dans ce cas, on remarquera qu'à partir d'un *support* donné H extrait d'une famille *indépendante*, **il est toujours possible de former un nouveau *support de vérité* de possibilité arbitraire** l (l/H) (si m=0) ou **de nécessité arbitraire** m (H + m/¬H) (si l=1).

On appellera *quantité hypothétique* sur un domaine D une telle *mesure fonctionnelle* sur un D. L'ensemble de telles quantités sera noté $H_\Omega(D)$. De manière générale, si D est muni d'une structure algébrique, on retrouve dans $H_\Omega(D)$.un certain nombre des propriétés algébriques de D [Pol 95b]. Par exemple, si on note IR le **corps des réels**, (H_Ω(D IR), +, .) est un **anneau unitaire** dans lequel on peut **immerger** (IR, +, .). Cette propriété est remarquable puisqu'elle permet d'opérer des calculs sur des "nombres flous" en conservant des propriétés des opérations sur les réels (distributivité, existence d'inverses) invalides sur les réels flous classiques.

4.3. Les *extensions* hypothétiques.

Nous généralisons donc ici l'approche précédente en introduisant la notion d'***extension hypothétique***. On constate que l'ensemble de la construction déjà opérée, et définissant les *extension floues*, peut être menée en remplaçant les valeurs de [0, 1] par les valeurs d'une algèbre de Boole T telle que définie en 3. Une *extension*

[12] ensemble des parties d'un ensemble donné Ω.

hypothétique. sera alors un élément de B* = T exp (V exp O). La bijection canonique analogue à celle déjà énoncée tient, et une valeur d'un objet pourra donc être considérée comme une *quantité fonctionnelle* sur l'espace des valeurs V., ce qui permettra de mener des calculs algébriques sur ceux-ci respectant les dépendances [Pol 93]. La structure d'algèbre de Boole rend compte par ailleurs du caractère d'indépendance ou de dépendance entre hypothèses, ce qui lève les limitations mentionnées ci-dessus.

Un *binôme canonique* sera de la forme $H/<o, V> + \neg H/\varnothing$ (l'une des quantités $|H|$ et $|\neg H|$ vaut 1), et une *extension totalement factorisée* sera de la forme :

$$S = P_i (H_i/<o_i, V_i> + \neg H_i/\varnothing)$$

où les H_i constituent une famille *indépendante* au sens donné en 3. Dans la pratique, on les construira à partir d'une base hypothétique $\{H_i^0\}$ unique prédéfinie, et l'on utilisera comme déjà énoncé les (l/H_i^0) (si m=0) et $(H_i^0 + m/\neg H_i^0)$ (si l=1) pour former un couple $(H_i, \neg H_i)$ de possibilités arbitraires (l, m).

La présente approche n'est autre qu'un calcul formel sur les H_i, $\neg H_i$,menant à des calculs de possibilités *in fine*, les constructions mathématiques proposées en 3 donnant la validité nécessaire aux opérations utilisés en leur conférant un sens rigoureux.

5. Application à un noeud de fusion.

On considère ici un *noeud de fusion* en montrant l'utilisation qui a été faite de l'approche précédente dans le cadre d'une fusion *au fil de l'eau* et *temps critique* qui est l'objet du programme DFD.

Un tel *noeud* peut être vu comme un ensemble ordonné d'opérateurs appliqués successivement à un indice entrant[13], visant en final à intégrer celui-ci dans la situation perçue locale du *noeud*. Ces opérateurs sont [Pol 96] :

- **alignement** : mise dans un référentiel d'espace et de temps homogène;
- **corrélation** : recherche des éléments de la *situation perçue* pouvant correspondre à l'entité du monde réel désignée par l'indice entrant;
- **association** : choix du ou des éléments les plus vraisemblables, avec création ou destruction éventuelle d'hypothèses;
- **identification** : enrichissement maximum de l'état en utilisant les connaissances a priori sur les caractéristiques des objets observés (la connaissance de certains d'attributs d'état permet par exemple de raffiner la catégorie de l'entité).
- **fusion** : mise à jour éventuelle de la situation locale en fonction de seuils prédéterminés, et génération en sortie du système d'un indice reflétant l'enrichissement de la situation opéré par le *noeud*.

Ces opérateurs sont complétés par un **gestionnaire d'hypothèses** ayant en charge d'une part la mise à jour des valeurs de confiance accordées aux hypothèses en fonction des délais écoulés, et d'autre part la suppression des hypothèses descendant en deçà d'un seuil de confiance déterminé.

La situation locale peut être représentée par une *extension hypothétique* d'objets d'une classe nommée ici *SituationElement*, définie [Pol 96] par les vecteurs de valeurs :

$$V = (c, t_v, X), \text{ où :}$$

- c est la *catégorie conceptuelle* d'appartenance (élément d'une énumération ou d'un treillis de symboles),
- t_v est le *temps valide*, c'est à dire le temps du monde réel auquel la description de l'objet s'applique (par opposition au *temps transactionnel* de la machine auquel cette description est considérée),
- X est l'état de l'objet proprement dit comprenant en particulier la position (x, y),

On adjoint d'autre part à un élément de situation la liste des indices justifiant l'existence supposée de celui-ci. Les indices sont éléments d'une *extension nette*. De plus, un objet de situation fait référence aux état anciens connus de celui-ci pour servir de base à des raisonnements temporels.

Une situation perçue revêtira donc la forme :

$$S = Expr (\varnothing , <o_1, V_1>, ..., <o_n, V_n>, ..., H_i, ..., \neg H_i, ...)$$

où Expr désigne une expression algébrique d'opérateurs +, * et /.

Une situation locale est dite *hétérochrone*, celle-ci se composant d'états d'entités ne faisant pas référence à un même temps unique du monde réel. La *corrélation* d'un indice avec une entité de la *situation perçue* fera donc appel à une opération F_C dite de *dérivation*, permettant d'associer à un état imprécis V* et à une durée donnée Dt un nouvel état V'* (en général plus imprécis) contenant l'ensemble des états possibles atteignables à partir de V* dans le temps Dt. Une telle fonction est spécifique de la *catégorie conceptuelle* C de l'entité et reflète ses capacités dynamiques (en particulier celles de déplacement dans l'espace). Une telle fonction peut

[13] Les problèmes de concurrence d'accès à la situation locale sont levés en excluant le traitement concurrent d'indices par le système à un instant donné.

également faire appel à la description du *milieu* (terrain, météo) si celle-ci existe.

On montre que chaque étape du processus de fusion peut être simplement modélisée à l'aide des opérateurs de l'algèbre des extensions hypothétiques. Ainsi, par exemple, si un indice entrant délivré par une source S est représenté par un quadruplet (E^*, t_{obs}, S, l), où E^* est l'état observé (imprécis), t_{obs} le temps (monde réel) de l'observation, et l une mesure représentative de la confiance accordée, celui-ci sera projeté en un élément de situation :

$$M = H_S/<o, c^*, t_v^*, X^*> + \neg H_S/\varnothing$$

- H_S est une nouvelle hypothèse *indépendante* telle que $P(H_S) = 1$; $P(\neg H_S) = 1 - l$
  (la confiance étant ici interprétée comme une *nécessité* d'existence de l'entité décrite dans la situation[14]).
- $c^* = j(E^*, S)$; $V^* = y(E^*, S)$ sont donnée par les connaissance a priori.

L'étape de *corrélation* nécessite pour sa part l'opérateur de *sélection* appliquée à une *extension* déduite de la *situation perçue* par application de l'opérateur de *dérivation*.
Toutes ces opérations son "linéaires" vis à vis de la somme et du produit externe /, sauf l'étape de fusion qui est non linéaire du fait de l'utilisation d'une seuil.

6. Implémentation.

Le logiciel **SAGESSE** (Système d'Acquisition et de Gestion d'Eléments de Situation pour les SIC Evolués) est une implémentation pilote de notre modèle en C++ sur station UNIX. Ce logiciel se compose du **Système de Gestion d'Objets** (SGO) proprement dit, auquel nous avons adjoint un algorithme de corrélation simple, une visualisation sur fond cartographique, et enfin une interface usager de manipulation d'objets permettant l'expression de requêtes sur la situation perçue. Un générateur de scénario permet d'alimenter ce noeud de fusion en indices.

Le SGO est basé sur une extension des types offerts par le modèle ODMG [Cat 93] par des types tels que fuzzy_int, fuzzy_float, etc., H_int, H_float, etc. permettant la représentation d'attributs dont la valeur est imprécise ou peut être conditionnée par une hypothèse. Les définitions de classes sont alors exprimées dans le langage ODL (Objet Definition Langage) de l'ODMG, convenablement étendu.

Par exemple :

```
interface Unite {
    H_extent Unites ;
    H_string nom;
    H_enum {Escadron, Regiment, Division} categorie;
    H_int effectif;
    H_location position;
    H_relationship Set<Unite> subordonnes;
    H_relationship Set<Indice> observations;
}
```

L'exemple montre un nouveaux types d'*extension* de classe représentant une *collection fonctionnelle* (implémentées comme un graphe dont les feuilles sont des références à des objets) est offert. Une extension spécifique des collections floues existe de même.

Des méthodes d'accès et de comparaison sont accessibles depuis l'interface C++ d'une part, et depuis le langage d'interrogation basé sur OQL (Object Query Langage) de l'ODMG d'autre part. La création d'une nouvelle hypothèse (indépendante de celles déjà créées) s'effectue en invoquant le "new" (méthode de création) de la classe "Hypothesis" fournie.
La fonctionnalité de *persistance* des objets, actuellement assurée par une bibliothèque logicielle basée sur les *Memory Mapped File* d'Unix, (les références à des objets étant classiquement "traduits" par des couples (oid, pointeur C)), est actuellement en cours de portage sur le SGBDOO[15] Objet Store. Les nouveaux *littéraux* tels que f_int, etc. sont alors simulés comme de nouvelles classes[16],
Les recherches en base nécessitant ordinairement l'accès à des descriptions alternatives pouvant être nombreuses (leur nombre croit avec le nombre d'hypothèses/sous hypothèses), des algorithmes particuliers utilisables lors de l'accès à l'information sont actuellement étudiés avec une estimation de leur complexité. Le but recherché est de conserver les temps d'accès du SGBDOO lorsqu'aucune hypothèse n'est active, ou encore lorsque la situation perçue est un produit de facteurs premiers indépendants.

7. Conclusion

Nous avons ici défini un modèle de données adapté aux futurs systèmes de fusion de données symboliques, supportant conjointement dans le cadre d'un modèle orienté objet les concepts d'imprécision sur les attributs, de doute existentiel, et enfin de représentation d'hypothèses concurrentes. Ce modèle est doté de facilités de représentation des données et de

[14] L'élément de situation $H_j/<o, V> + \neg H/\varnothing$ avec $\Pi(H) = 1$; $\Pi(\neg H) = 1$ n'apportant aucune connaissance.

[15] Système de Gestion de Bases de Données Orienté Objet.

[16] La liste des types de littéraux n'étant pas extensible en ODMG.

manipulation de celles-ci au cours d'un raisonnement. Notre approche permet donc de doter un système de fusion de la capacité de **fonctionner au fil de l'eau** et en **temps critique**.

D'autre part, une approche formelle reposant sur une construction mathématique rigoureuse dote les différentes facilités proposées d'une sémantique précise. Ceci nous a permis d'étendre les opérateurs usuels d'insertion, de modification et d'interrogation d'une base de données.

L'intérêt de ces concepts a pu être démontré dans le cadre d'un Système de Gestion d'Objets pilote, permettant le prototypage d'un noeud de fusion.

Toutefois, la mise en oeuvre du logiciel de gestion proposé dans le démonstrateur du programme DFD nécessite encore l'étude d'algorithmes d'accès optimaux d'une part, et le portage sur un véritable SGBDOO d'autre part, ce qui est le sens des recherches actuellement menées.

Un autre point d'intérêt concerne bien sur l'étude d'applicabilité des concepts proposés à d'autres domaines, en particulier issus du secteur civil (gestion de crise, maîtrise des risques industriels, etc.).

Remerciements. : nous remercions tout particulièrement Monsieur Michel Barès de la Direction des Recherches, Etudes et Techniques (DRET) grâce auquel ont pu être menés les présents travaux.

Bibliographie.

[And 89] Andres V., "Filtrage sémantique dans une base de données imprécises et incertaines: un système souple autorisant la formulation de requêtes composites pondérées", Thèse, Toulouse, 1989.

[Bos 92] Bosc P., Pivert O. "Sur l'utilisation de requêtes booléennes pour l'évaluation de requêtes relationnelles floues". Proc. VIIIèmes Journées Bases de Données Avancées 1992.

[Bou 92] B. Bouchon-Meunier, "Représentation et traitement de l'incertitude". DGA, Sciences et Défense 92.

[Cat 93] R. Catell, T. Atwood, J. Duhl, G. Ferran, M. Loomis, D. Wade. "The Object Database Standard ODMG-93". Morgan Kaufmann Publishers. Juillet 1993.

[Dub 87] Dubois D., Prade H., "On incomplete conjunctive information". Proc. of the North American Fuzzy Information Processing Society Workshop, Purdue Univ. W. Lafayette, 1987.

[Dub 88] Dubois D., Prade H., "Théorie des possibilités. Application à la représentation des connaissances en informatique", Editions Masson, 1988.

[Dub 92] Dubois D., Prade H., "Combination of fuzzy information in the Framework of Possibility Theory", Data Fusion in Robotics and Machine Intelligence. 1992.

[Gal 88] Galibourg M., "Mises en oeuvre d'opérateurs flous dans une base de données", Thèse, Rennes, 1988.

[Gog 67] Gogen J.A. "L-Fuzzy Sets". Journal of Mathematical Analysis and Applications, volume 18, Number 1. 1967.

[Mou 92] N. Mouaddib, O. Foucaut, P. Subtil, D. Vilmin, "Firms : un système orienté objet pour la représentation des information nuancées. Application aux sciences économiques". 2èmes journées nationales "Les applications des ensembles flous". Novembre 1992.

[Pet 96] Petry F.E. "Fuzzy Databases, Principles and Aplications", International series in Intelligent Technologies, Hardbound, 1996.

[Piv 91] Pivert O., "Contribution à l'interrogation flexible de bases de données: expression et évaluation de requètes floues", Thèse, Rennes, 1991.

[Pol 93a] Pollet Y., Bregeault L., Bridon Ph., "Bases de Données pour les systèmes de fusion de données". Troisièmes journées nationales "Applications des Ensembles Flous", Nïmes1993.

[Pol 93b] Pollet Y. "Une approche formelle des nombres mal connus ". Troisièmes journées nationales "Applications des Ensembles Flous", Nïmes1993.

[Pol 94a] Pollet Y"Un modèle de données spatio-temporel flou pour les SIC". Journées internationales "Avignon 94", conclave IA, Défense et sécurité civile, Paris, 1993.

[Pol 94b] Pollet Y. "Une approche pour la représentation des situations opérationnelles dans les systèmes de fusion ". 4èmeJournées sur les applications des ensembles flous Lille 1994.

[Pol 95a] Pollet Y., Robidou S. "An approach for the Management of Multivalued Attributes in Fuzzy Databases". Proc. of Fuzzy IEEE/IFES'95 Workshop on Fuzzy Databases Systems and Information Retrieval. Yokohama, 1995.

[Pol 95b] Pollet Y., Robidou S. "Une approche pour la simulation des systèmes en univers

incertain". Rencontres francophones sur la logique floue. Paris, 26, 27 novembre 1995.

[Pol 96] Pollet Y. "Représentation des connaissances et fusion de données symboliques".Journées Régionales Science et Défense. Nancy, 1996.

[Pra 84] Prade H., Testemale C., "Generalizing database relational algebra for the treatment of incomplete/uncertain information and vague queries. Inf. Sci., 34, 1984.

[Pra 86] Prade H., Testemale C., " Representation of soft constraints and fuzzy attribute values by means of possibility distributions in databases", The Analysis of Fuzzy Information, (J. Bezdzk, Eds), CRC Press, 1986.

[Rob 96] Robidou S., Pollet Y. "Représentation des données imparfaites dans les Systèmes d'Information et de Défense". Article soumis à la revue TSI. 1996.

[Sad 93] Sadri F., "Modeling Uncertainty in Object Oriented Databases". Proc. Workshop Incompleteness and Uncertainty in Information Systems, Springer-Verlag, 1994.

[Tes 84] Testemale C. "Un système de traitement d'informations incomplètes ou incertaines dans une base de données relationnelles". Thèse de doctorat, Université de Toulouse, 1984..

[Uma 83] Umano M. "Retrieval from fuzzy database by fuzzy relational algebra". Proc IFAC Symposium on fuzzy Information, Knowledge Representation and Decision Processes, Marseille, 1983..

[VGy 93] Van Gyseghem N., De Caluwe R., Vandergerghe R. "UFO Uncertainty and Fuzziness in an Object Oriented Model". 2nd IEEE Int Conference on Fuzzy Systems, 1993.

[Zad 65] L.A. Zadeh. "Fuzzy sets", Information and Control, Vol 8, New York : Academic Press, 1965.

[Zad 78] L.A. Zadeh. "Fuzzy Sets as a basis for a Theory of Possibility", Fuzzy Sets and Systems, 1, 1978.

[Zic 90] Zicari R. "Incomplete Information in Object-Oriented Databases", Sigmod Record 19, 1990.

PAPER No. 27

DISCUSSOR'S NAME: B. Dacre-Wright

COMMENT/QUESTION:

Avez-vous une idée de la robustesse de votre méthode à une erreur de registration entre les images?

(Do you have an idea of the ruggedness of your method when there is a registration error between images?)

AUTHOR/PRESENTER'S REPLY:

A ce jour, nous ne nous sommes pas intéressés à la robustesse de la méthode à une erreur de registration entre les images. Mais c'est quelque chose que nous comptons analyser.

(To date, we have not looked at this aspect. It is, however, something we intend to analyse.)

DISCUSSOR'S NAME: P. Verlinde

COMMENT/QUESTION:

Est-ce que vous avez verifié votre hypothèse de départ sur la "normalité" des distributions de probabilité?

(Did you verify your initial hypothesis concerning the "normality" of the probability distribution?)

AUTHOR/PRESENTER'S REPLY:

Les distributions de probabilité ne vérifient pas l'hypothèse de normalité. Nous avons fait une hypothèse simplificatrice pour utiliser la règle de décision de Bayes.

(The probability distributions do not check out the normality hypothesis. We made a simplified assumption in order to be able to use the Bayes decision rule).

ADAPTIVE INTENSITY MATCHING FILTERS : A NEW TOOL FOR MULTI-RESOLUTION DATA FUSION.

S. de Béthune
F. Muller
M. Binard
Laboratory SURFACES University of Liège
7, place du 20 août
B 4000 Liège, BE

1. SUMMARY

Many different multiresolution fusion methods have been proposed in the literature. An important actual aim in this field of research is to produce colour composites of multiresolution data preserving both the essential spatial information of the high resolution image and the spectral information content of the low resolution channels, so as to produce pseudo high resolution spectral channels which can be further processed for improved classification or other information extraction purposes.

The best integration results in this regard have been obtained by the HPF algorithm and by a new fusion method based on multiresolution analysis of the images using the wavelet transform.

A new methodology based on adaptive intensity matching filters using local image statistics to spectrally adjust high resolution images to the radiometry of low resolution channels is described in this paper. The algorithm tends to equalise the mean (LMM algorithm) or the mean and the variance (LMVM algorithm) of the high resolution image with those of the low resolution channels, on a pixel by pixel basis, from the values measured within a local window around each pixel position.

The INR (Intensity Normalised Ratio) transform, as defined in this paper, is a fast alternative to the RGB-IHS-RGB transform, and allows an efficient implementation of the intensity matching fusion method, generalised to images with more than three channels.

These algorithms are applied to a 1024 x 1024 window extract of a high resolution (5m) KOSMOS KVR 1000 panchromatic image to be fused with a registered low resolution (20 m) SPOT XS image, obtained over the city of Liège (Belgium).

The results obtained for varying filtering window sizes are compared with the HPF filter and the wavelet transform applied to the same set of images.

2. INTRODUCTION

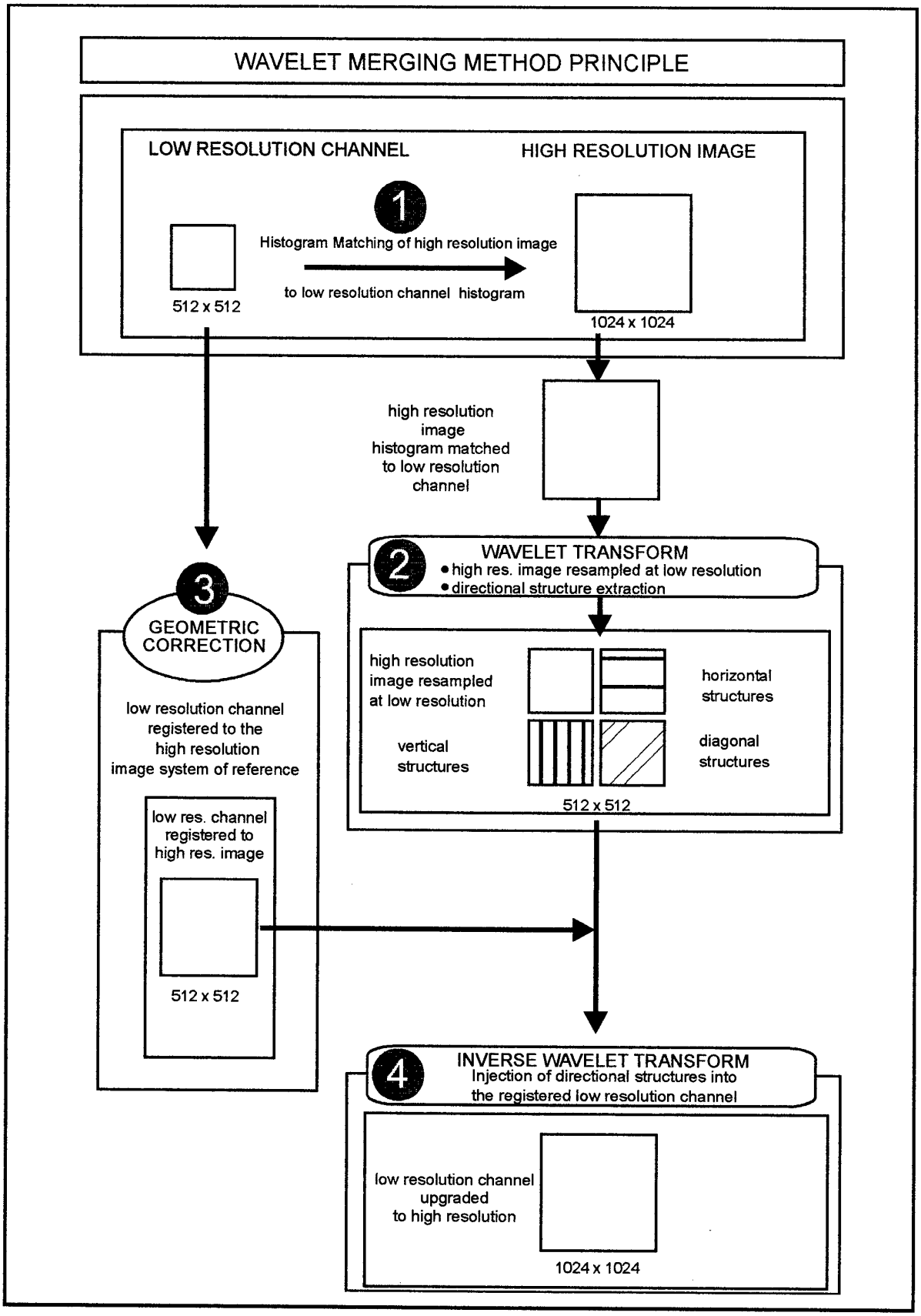
The terms «fusion», «integration», «merging», «bandsharpening» have been used interchangeably in the scientific remote sensing literature to describe processes aiming at condensing information coming from multisource and multitemporal satellite image data. More specifically, these processes aim at producing colour composites preserving in the best possible way most of the significant or desired information contained in the separate images or image channels. Clearly, a choice as to which information will be relevant of the final product has to be made.

In this regard, an important effort was realised in recent years in order to preserve the highest possible spatial resolution of the available data sets without destroying spectral information content. The most usual situation in this field is the integration of three spectral channels with an high spatial resolution image.

The onset of this type of research started as soon as multisource data with different spatial resolutions became available. The best known examples concern the merging of SPOT Panchromatic images with LANDSAT TM or SPOT XS channels. These first methods applied simple arithmetic operations between channels or used principal component analysis (PCA) to condense information ([1], [2], [3], [4]). The final products correspond in general to high spatial resolution images, but with important alteration of the spectral information content of the original channels [2].

A first step toward effective spectral information conservation was introduced with the RGB-IHS-RGB transform ([5], [6], [7]). This transform is performed in three steps : (1) transform three low resolution multispectral channels from RGB to IHS space ; (2) replace the intensity channel with the high resolution channel ; and (3) perform a backward projection from IHS to RGB space. The method preserves two of the essential characteristics of the spectral colour, namely the original hue and saturation of the three spectral channels. The quality of the final product however depends on the actual differences between the original intensity channel and the intensity of the high spatial resolution image. Similar hues with identical saturation values are still strongly affected by important differences in intensity values, and are therefore not always similar looking on the final colour composite. The method performs best if there is an evident correlation between the IHS intensity component and the high resolution channel, ensuring that bright objects on the original image remain bright in the merged product [7]. Unfortunately, high correlation has no reason to be expected in most situations, where the images to be fused stem from different sensors and have been obtained at different times.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.


WAVELET MERGING METHOD PRINCIPLE
LOW RESOLUTION CHANNEL
HIGH RESOLUTION IMAGE
1
Histogram Matching of high resolution image
to low resolution channel histogram
512 x 512
1024 x 1024
high resolution
image
histogram matched
to low resolution
channel
2
WAVELET TRANSFORM
• high res. image resampled at low resolution
• directional structure extraction
high resolution
image resampled
at low resolution
horizontal
structures
vertical
structures
diagonal
structures
512 x 512
3
GEOMETRIC
CORRECTION
low resolution channel
registered to the
high resolution
image system of reference
low res. channel
registered to
high res. image
512 x 512
4
INVERSE WAVELET TRANSFORM
Injection of directional structures into
the registered low resolution channel
low resolution channel
upgraded
to high resolution
1024 x 1024

To overcome this lack of correlation between high and low resolution channels, a simple integration method consists in adding the result of a high pass filter (HPF) of the high resolution channel to the spectral channels, as was already proposed by Showengerdt [8]. This high pass filter is directly related to the structural information content of the high resolution image and removes most of its spectral information. It has a zero mean value and can be integrated in the low resolution spectral channels by calculating appropriate weighted sums, so as to ensure minimum modification of the low resolution spectral channels ([8], [9]). Chavez *et al.* [2] showed that the results generated with the HPF method are less distorted than those obtained by the traditional IHS or PCA methods.

The same approach - injecting high spatial information into the low resolution spectral channels - has also been proposed by Ranchin [10] and others ([11], [12]), and has been developed by means of multiresolution analysis of remotely sensed images using the wavelet transform [13]. Undoubtedly this approach produces the best results in regard of preservation of spectral information content of the low resolution multispectral channels [14].

This paper proposes a different approach, pursuing the same goals and achieving similar results as the multiresolution wavelet transform, and is based on adaptive intensity matching techniques using local image statistics.
In the following we shall briefly describe the wavelet merging method principle before introducing the adaptive intensity matching approach. These two methods and the HPF method will next be compared as applied to the same set of multitemporal and multiresolution data.

3. DATA FUSION BY MULTIRESOLUTION ANALYSIS AND WAVELET TRANSFORM.

This first method, called the ARSIS method (Augmentation de la Résolution Spatiale par Injection des Structures) as described by Ranchin *et al.* [14], aims at injecting high spatial structural information from the high resolution image into the low resolution channel.

The multiresolution analysis by means of the wavelet transform used to perform this task is essentially a four stage process (see figure 1, adapted from Garguet-Duport *et al.* [11] : (1) histogram matching of the high resolution image to the corresponding low resolution channel in order to adjust radiometry and improve the initial correlation between the two images. (2) perform a forward wavelet transform of the high resolution channel so as to produce four new images at the same resolution as the low resolution image. Three of these images contain the structural information of the high resolution image to be injected into the low resolution channel. They are coded by the wavelet coefficients and correspond to directional high pass filters. The fourth image, a low pass filter, corresponds to the resampled high resolution image at the resolution of the low resolution channel. (3) registering the low resolution image to the same system of reference as the high resolution image. And (4), apply the inverse wavelet transform to the registered low resolution channel, to produce a pseudo high resolution spectral channel with the added structural information content.

In order to produce a spatially enhanced colour composite, this process is successively applied to the three separate spectral channels to be combined.

4. DATA FUSION BY ADAPTIVE INTENSITY MATCHING FILTERS

As mentioned above, the quality of spectral colour preservation with the IHS transform depends on the difference between the IHS intensity channel and the high resolution channel. High correlation between these two channels will usually produce good results [7]. If, as in most instances, correlation is poor, then the only remaining alternative is to minimise the difference between the high resolution channel and the IHS intensity channel. More precisely, the original spectral information of the low resolution channels will be better preserved if the high resolution channel is made to resemble the IHS intensity channel.

Image resemblance can be achieved by matching the histograms of the two images, and is used for instance, to minimise radiometric contrast between neighbouring images in a mosaic [15]. Another approach, also producing similar histograms, is to use normalisation functions [16], which tend to equalise the means or the means and variances of the two images. Experience shows however that these approaches are strongly case dependant and that they provide the best results when the initial image correlation is high. This is because histogram matching algorithms only remodel the image frequency distribution without any control on the actual spatial distribution of the intensity values within the images.

In order to ensure true intensity matching of the images spatially, one has to be able to control and remodel the frequency distributions at a local scale. This can be performed by using adaptive filtering techniques, which have also widely been used in other domains, such as speckle removal from radar images ([17], [18]) or for image reconstruction purposes of lost data [19].

Following this approach, two local intensity matching filters, have been devised, based on the normalisation functions of Joly [16], one adjusting the local means of the images (LMM) and one adjusting both local means and variances (LMVM). The general Local Mean Matching and Local Mean Variance Matching algorithms to integrate two images, a high resolution image (H) into a low

resolution channel (L) resampled to the same size as H, are given by :

- **the LMM algorithm**

$$F_{H>L(l,c)} = \frac{DN_{H(l,c)} * M_{L(l,c)}}{M_{H(l,c)}} \quad \textbf{(1)}$$

- **the LMVM algorithm**

$$F_{H>L(l,c)} = \frac{(DN_{H(l,c)} - M_{H(l,c)})(\sigma_{L(l,c)})}{\sigma_{H(l,c)}} + M_{L(l,c)} \quad \textbf{(2)}$$

where M corresponds to the local means of the images :

$$M_{\mathrm{Im}(l,c)} = \frac{\sum_{k=c-w/2}^{c+w/2} \sum_{j=l-h/2}^{l+h/2} DN_{\mathrm{Im}(k,j)}}{hw} \quad \textbf{(3)}$$

and σ to their local standard deviations :

$$\sigma_{\mathrm{Im}(l,c)} = \frac{\sqrt{hw\left(\sum_{k=c-w/2}^{c+w/2} \sum_{j=l-h/2}^{l+h/2} DN^2_{\mathrm{Im}(k,j)}\right) - \left(\sum_{k=c-w/2}^{c+w/2} \sum_{j=l-h/2}^{l+h/2} DN_{\mathrm{Im}(k,j)}\right)^2}}{hw} \quad \textbf{(4)}$$

measured within a window of size h*w centred on the image pixel at co-ordinates (l,c), DN corresponding to the pixel values of the images.

These algorithms will produce a simulated high spatial resolution image (F) pertaining the spectral characteristics of the low resolution channel (L). How well the spectral values are preserved will depend on the size of the filtering window. Small window sizes produce the least distortion. Larger filtering windows incorporate more structural information from the high resolution image, but with more distortion of the spectral values.

If the spatial resolution ratio between the two images is important, then the pixel values of the resampled low resolution image will already correspond to a local average value of the image. Hence, in this case, the LMM algorithm (1) can be rewritten as :

$$F_{H>L(l,c)} \approx \frac{DN_{H(l,c)} * DN_{L(l,c)}}{M_{H(l,c)}} \quad \textbf{(5)}$$

which shows that the LMM filtered image corresponds to the low resolution channel multiplied by the image_to_local_mean ratio of the high resolution image.

The LMVM algorithm becomes :

$$F_{H>L(l,c)} \approx \frac{(DN_{H(l,c)} - M_{H(l,c)})\sigma_{L(l,c)}}{\sigma_{H(l,c)}} + DN_{L(l,c)} \quad \textbf{(6)}$$

The resampling process of the low resolution image to the same scale as the high resolution image drastically reduces its local standard deviations ($\sigma_{L(l,c)} \approx 0$), especially if the resolution ratio between the images is high and for small filtering window sizes. In this instance, equation (6) simply reduces to the limiting case :

$$F_{H>L(l,c)} \approx DN_{L(l,c)} \quad \textbf{(7)}$$

The fused image will not significantly differ from the low resolution image. In this case, larger local environments should be taken to perform the filtering in order to incorporate more information from the high resolution image.

Without the constraint of variance equalisation, equation **(6)** becomes :

$$F_{H>L(l,c)} \approx DN_{H(l,c)} - M_{H(l,c)} + DN_{L(l,c)} \quad \textbf{(8)}$$

and corresponds to the HPF integration filter defined by Showengerdt [8]. This formulation also corresponds to the additive form of equation **(5)**.

4.1. Procedural steps for implementing image fusion by adaptive intensity matching.

As described earlier, the LMM and LMVM filters perform an intensity matching of two channels which have previously been registered. Hence, this process can be applied between the high resolution channel and three individual low resolution spectral channels producing three spatially enhanced images which can then be combined to form the desired colour composite.

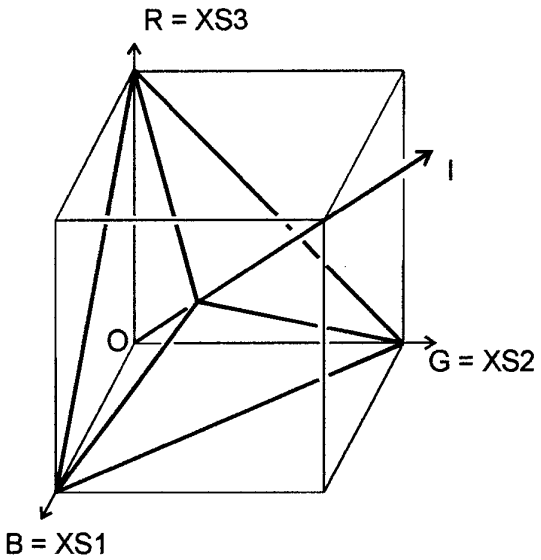
Instead of applying the filtering algorithms to each of the separate channels, a simplified method can be adopted based on the IHS conversion. If three channels have to be merged with a high spatial resolution image, these channels can first be converted from RGB to IHS space. The high resolution image will then be intensity matched to the IHS intensity channel and substituted to this channel before a reverse projection back into RGB space. The resulting RGB channels correspond to the spatially enhanced spectral channels, and are obtained after only one filtering process.

4.2. Generalising the data fusion implementation with the INR transform

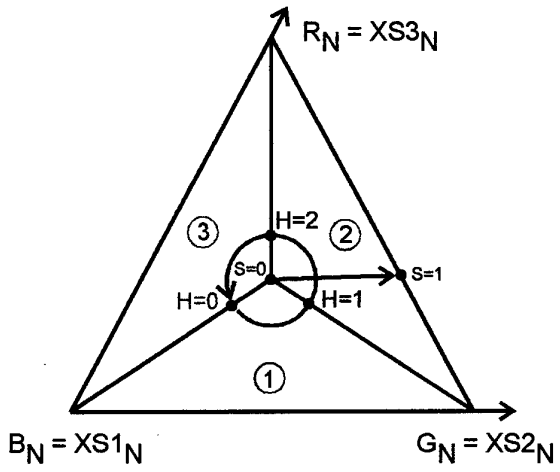
An improvement of this implementation method, leading to a further simplification and a generalisation of the procedure to more than three channels, has been devised. It is based on a new analytic formulation of IHS space.

GEOMETRIC RELATIONS IN RGB-IHS SPACE

RGB - IHS SPACE



NORMALISED RATIO PLANE
HUE-SATURATION PLANE



RGB coordinates	IHS coordinates	INR coordinates
R = XS3 G = XS2 B = XS1	I = XS1 + XS2 + XS3 ① H = (XS2-XS3)/(I-3XS3) C ② S = (I-3XS3)/(I-2XS3) C ③	I = XS1 + XS2 + XS3 $XS3_N$ = XS3/I $XS2_N$ = XS2/I

When an IHS-RGB transform is performed, only the intensity channel is taken into account and replaced, while the hue and the saturation channels are left unchanged. It is however necessary to compute both these channels as they are needed in the reverse conversion from IHS to RGB space.

The relationships between RGB and IHS space are presented on figure 2.

The RGB axes of the RGB cube correspond to an hypothetical SPOT multispectral image (XS3, XS2 and XS1 channels). Crosscutting the RGB cube, the IHS space is schematically represented by the RGB triangle, perpendicular to the Intensity (I) axis, which can be defined as the sum of the three XS channels ([5], [20], [21]) :

$$I = XS1 + XS2 + XS3 \quad (9)$$

The triangle itself corresponds to the two dimensional Hue-Saturation plane. The expressions of the Hue and Saturation co-ordinates are cyclic, and depend on the position of the individual colours inside the three subtriangles making up the RGB triangle [5]. We can express Hue on a scale ranging from 0 (= blue) to 3 (= 0 = blue) where the particular values 1 and 2 correspond respectively to green and red hues. Saturation can be expressed on a scale of 0 (= no saturation) to 1 (= total saturation). For each subtriangle we thus have :

$$\begin{aligned} H_1 &= (XS2 - XS3)/(I - 3XS3) \\ H_2 &= 1 + (XS3 - XS1)/(I - 3XS1) \\ H_3 &= 2 + (XS1 - XS2)/(I - 3XS2) \end{aligned} \quad (10)$$

$$\begin{aligned} S_1 &= (I - 3XS3)/(I - 2XS3) \\ S_2 &= (I - 3XS1)/(I - 2XS1) \\ S_3 &= (I - 3XS2)/(I - 2XS2) \end{aligned} \quad (11)$$

where the subscripts refer to the particular subtriangles illustrated on figure 2.

Individual colours plotted onto the RGB plane by projection from the origin (O) of the axes can also be described in terms of the normalised ratios of the three colour components (barycentric co-ordinates) as shown on figure 2. Hence, description of IHS space can advantageously be replaced by what we shall define as INR space, standing for Intensity-Normalised Ratio space. Three co-ordinates are necessary to define this space, the usual intensity co-ordinate and two normalised ratios :

$$\begin{aligned} I &= XS1 + XS2 + XS3 \\ XS1_N &= XS1 / I \\ XS2_N &= XS2 / I \end{aligned} \quad (12)$$

The third normalised ratio is redundant as it can be deduced from the first two:

$$XS3_N = XS3 / I = 1 - XS1_N - XS2_N \quad (13)$$

Performing a backward transformation of IHS to RGB space with preservation of hue and saturation, amounts to say that the normalised ratios of the three channels are also preserved. Hence, if a panchromatic SPOT image (PAN) has been substituted to the IHS intensity channel, then the backward transform should produce three new channels for which the following relations hold :

$$I_{new} = PAN = XS1_{new} + XS2_{new} + XS3_{new} \quad (14)$$

and

$$\begin{aligned} XS1_{new} / PAN &= XS1_N = XS1 / I \\ XS2_{new} / PAN &= XS2_N = XS2 / I \\ XS3_{new} / PAN &= XS3_N = XS3 / I \end{aligned} \quad (15)$$

Rewriting these equations yields :

$$\begin{aligned} XS1_{new} &= XS1 * PAN / I \\ XS2_{new} &= XS2 * PAN / I \\ XS3_{new} &= XS3 * PAN / I \end{aligned} \quad (16)$$

showing that a data integration procedure by INR transform only requires computation of the global intensity channel of the three low resolution multispectral channels even without having to compute any normalised ratio beforehand. These relations also clearly stress the fact that the original channel ratios are preserved by the new multispectral channels. The new channels are of course identical to those obtained by a classic RGB-IHS-RGB transform.

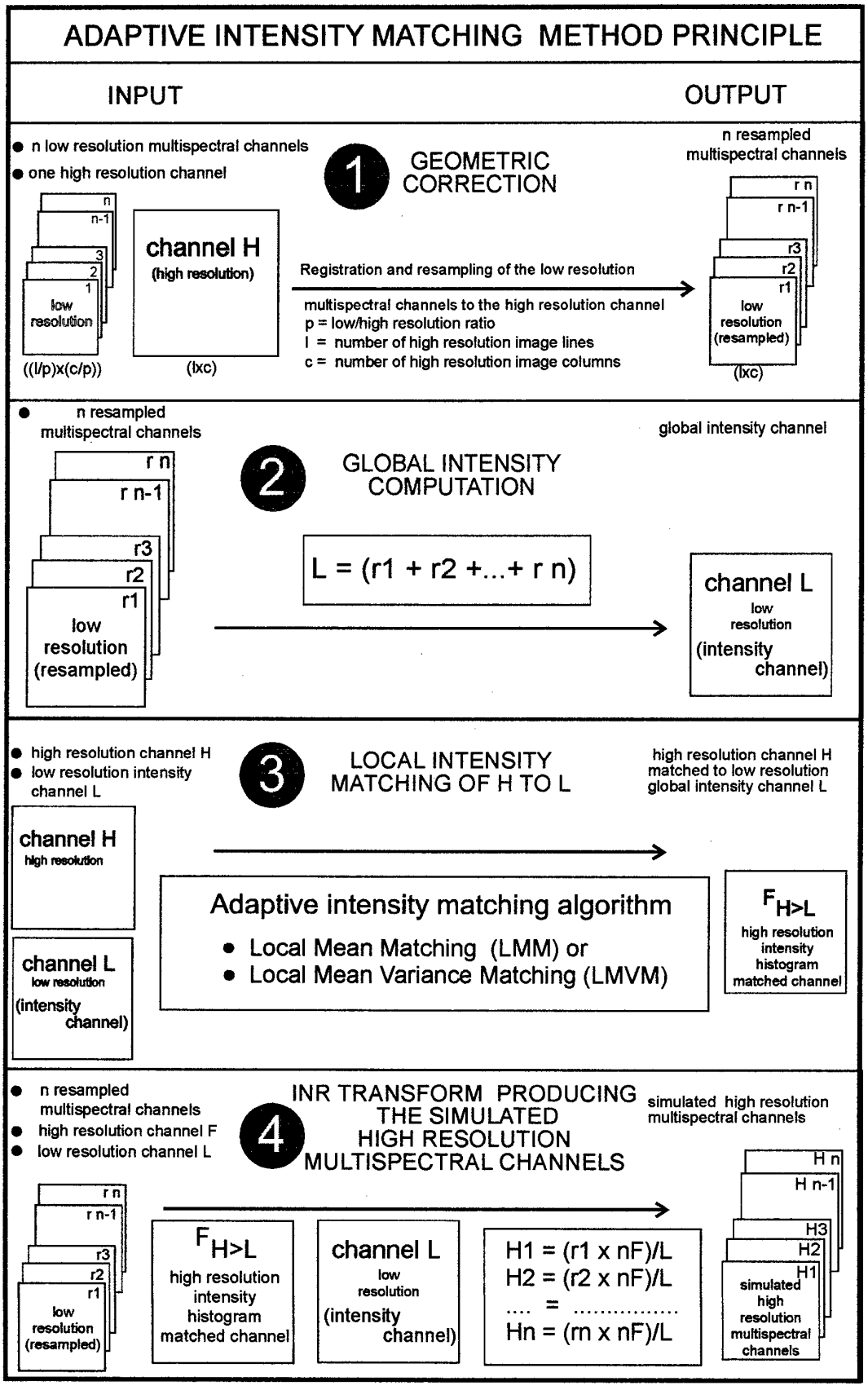
Furthermore the data fusion by INR transform is easily generalised for multispectral data of more than three channels, such as for instance LANDSAT TM multispectral images. For n channels (ch) to be upgraded with a high resolution image (H) we thus have :

$$I = \sum_{i=1}^{n} ch_i \quad (17)$$

$$new_ch_i = ch_i * H / I \quad (18)$$

By this method, only one intensity matching filtering of the high resolution image to the global intensity channel is needed to produce all the integrated upgraded spectral channels.

Based on this approach, the general merging method can be described in four steps as illustrated on figure 3.
The first step is to register the low resolution multispectral channels to the high resolution image (H), so as to obtain resampled multispectral channels (r) overlaying exactly the high resolution image. This increases artificially the spatial resolution of the channels, as their pixel size diminishes, but without increasing their spatial information content, so that they still have to be considered as low resolution channels. These channels are then summed up (step 2) to produce a low resolution global intensity channel (L) to which the high resolution image (H) will be intensity-matched (step 3), resulting in a high resolution global intensity matched image (F). The


ADAPTIVE INTENSITY MATCHING METHOD PRINCIPLE
INPUT
OUTPUT
n low resolution multispectral channels
one high resolution channel
1
GEOMETRIC CORRECTION
n resampled multispectral channels
n
n-1
3
2
1
low resolution
((l/p)x(c/p))
channel H
(high resolution)
(lxc)
Registration and resampling of the low resolution multispectral channels to the high resolution channel
p = low/high resolution ratio
l = number of high resolution image lines
c = number of high resolution image columns
r n
r n-1
r3
r2
r1
low resolution (resampled)
(lxc)
n resampled multispectral channels
global intensity channel
2
GLOBAL INTENSITY COMPUTATION
L = (r1 + r2 +...+ r n)
channel L
low resolution
(intensity channel)
high resolution channel H
low resolution intensity channel L
3
LOCAL INTENSITY MATCHING OF H TO L
high resolution channel H matched to low resolution global intensity channel L
channel H
high resolution
channel L
low resolution
(intensity channel)
Adaptive intensity matching algorithm
Local Mean Matching (LMM) or
Local Mean Variance Matching (LMVM)
F H>L
high resolution intensity histogram matched channel
n resampled multispectral channels
high resolution channel F
low resolution channel L
4
INR TRANSFORM PRODUCING THE SIMULATED HIGH RESOLUTION MULTISPECTRAL CHANNELS
simulated high resolution multispectral channels
F H>L
high resolution intensity histogram matched channel
channel L
low resolution
(intensity channel)
H1 = (r1 x nF)/L
H2 = (r2 x nF)/L
.... = ...
Hn = (rn x nF)/L
H n
H n-1
H3
H2
H1
simulated high resolution multispectral channels

KOSMOS KVR 1000 IMAGE OVER S-E OF LIEGE (BE)

resampled to 5 m resolution

1024 x 1024



final step applies the INR transform to the low resolution resampled channels (ri) using the F and L images, yielding the simulated high resolution multispectral channels.

5. DATA DESCRIPTION AND APPLICATION

The three fusion algorithms (LMVM, LMM and HPF) were applied with different window sizes to a SPOT XS scene to be integrated with a panchromatic KOSMOS KVR 1000 image, and the results are compared with those obtained by the ARSIS method computed by Dr. T. Ranchin at the University of Sophia-Antipolis (Nice, France).

The two images were taken above the city of Liège (Belgium), respectively in May and in June of 1992. The KOSMOS image, with a nominal resolution of two meters was registered to the 1:25000 topographic basemap and resampled by cubic convolution to 5 meters, in order to reduce the original resolution ratio between the two sets of data from 10 to 4. The three SPOT XS channels were then registered to the KOSMOS image, and also resampled by cubic convolution to 5 meter pixels. A square window extract of 1024 x 1024 pixels over the south-eastern part of the city was selected to perform the integration analysis (figure 4.).

This study area along the Meuse river is characterised by recent industrial development and suburban growth. Figure 4 shows the river harbour bordered by industrial zonings to the south and to the east, urban and suburban areas to the north-west and mixed forest and agricultural areas to the south-east.

Table 1 summarises the main statistical characteristics of the three SPOT XS channels and of the KOSMOS image. The XSI channel corresponds to the intensity channel of the IHS transform. The entropy value is a measure of the average amount of information of the channel, as described by different authors ([22],[23]) :

$$Entropy = \sum_{i=0}^{255} - p_i * \log_2(p_i) \quad \textbf{(19)}$$

where p_i is the probability of level i in the image.

As can be seen in table 1, the KOSMOS image is characterised by much higher standard deviation and entropy values than the corresponding SPOT XS channels.

Table 1 : Image statistics and spatial resolution

	Mean	Median	St. Dev.	Entropy	Range
KVR	137.72	139.5	56.29	7.6400	236
XS1	53.89	53.5	6.58	4.5555	100
XS2	43.14	42.5	8.76	4.9804	120
XS3	63.12	60.5	20.36	6.2546	157
XSI	53.05	52.5	8.32	4.9076	123

Table 2 presents the correlation between the different channels. The correlation between the two channels in the visible part of the spectrum (XS1, XS2) is elevated, but is low between the visible and infrared part (XS3) of the spectrum. Although the two images were taken at only two months interval, the KOSMOS image does not really correlate well with any of the XS channels. The quite low KVR-XSI correlation value precludes a satisfactory fusion between these images by the traditional IHS method, as it will significantly distort the spectral values of the original low resolution channels [7].

Table 2 : Correlation between the different image channels

	KVR	XS1	XS2	XS3	XSI
KVR	1.000				
XS1	0.6240	1.000			
XS2	0.6210	0.9611	1.000		
XS3	-0.0428	-0.0306	-0.0398	1.000	
XSI	0.3474	0.5756	0.5714	0.7934	1.000

The three LMVM, LMM and HPF merging algorithms were applied to this set of images with 6 different window sizes, ranging from 3 x 3 to 49 x 49 pixels in size. In order to assess the quality of the different fusion algorithms applied to the images, a set of statistical parameters based on the criteria proposed by Ranchin *et al.* [14] were computed (tables 3 to 5) and are illustrated on figure 5., together with the same data corresponding to the original images and those obtained by the ARSIS method. Ideally, the mean and the standard deviation of the integrated and of the original channels should remain as close as possible. Similarly, correlation between these channels shoud be as near as possible to one [14].

These tables also lists the entropies of the different results, allowing to analyse the added information content of the integrated channel as compared to the original channel. The last parameter listed in these tables is a deviation index (d_index) as defined by Costantini *et al.* [24] :

$$d_index = \frac{1}{lc} \sum_{i=1}^{l} \sum_{j=1}^{c} \frac{\left| F_{i,j} - L_{i,j} \right|}{L_{i,j}} \quad \textbf{(20)}$$

which measures de relative deviation of the fused image (F) of that of the original low resolution reference channel (L). The smaller this value, the better the image quality.

5.1 The LMVM filter (table 3 ; figure 5)

By construction, the mean and standard deviation of the fused images remain constant for all filtering window sizes. Entropy remains constant for small window sizes (3 to 7). Correlation values also remain practically constant, very near the maximum possible value. At these local scales the differences between the reference image and the fused images are so small that they do not bear any real significance. This is due to the fact that, at

Table 3 : Local Mean Variance Matching Statistics

window	KVR	XSI	LMVM 3	LMVM 5	LMVM 7	LMVM 15	LMVM 25	LMVM 49	ARSIS
Mean	137.72	53.39	53.39	53.40	53.40	53.40	53.41	53.40	53.05
St. Dev.	56.29	8.32	8.32	8.30	8.29	8.28	8.30	8.27	9.10
Entropy	7.6400	4.9062	4.9049	4.9033	4.9042	4.9169	4.9356	4.9371	5.1067
Correlat.	0.3474	1.0000	0.9973	0.9929	0.9858	0.9431	0.8878	0.8065	0.9253
d_index	1.5985	0.0000	0.0090	0.0137	0.0182	0.0347	0.0498	0.0685	0.0488

Table 4 : Local Mean Matching Statistics

window	KVR	XSI	LMM 3	LMM 5	LMM 7	LMM 15	LMM 25	LMM 49	ARSIS
Mean	137.72	53.39	53.12	52.98	52.96	52.95	52.94	52.91	53.05
St. Dev.	56.29	8.32	8.67	9.70	10.81	13.38	14.97	17.04	9.10
Entropy	7.6400	4.9062	5.0397	5.2555	5.4320	5.7528	5.9133	6.0825	5.1067
Correlat.	0.3474	1.0000	0.9541	0.8484	0.7572	0.6004	0.5298	0.4787	0.9253
d_index	1.5985	0.0000	0.0358	0.0712	0.0986	0.1529	0.1848	0.2233	0.0488

Table 5 : Local High Pass Filter Statistics

window	KVR	XSI	HPF 3	HPF 5	HPF 7	HPF 15	HPF 25	HPF 49	ARSIS
Mean	137.72	53.39	53.39	53.39	53.41	53.72	54.02	54.55	53.05
St. Dev.	56.29	8.32	10.24	15.06	19.20	26.43	30.38	35.22	9.10
Entropy	7.6400	4.9062	5.3488	5.9466	6.3097	6.7182	6.8319	6.8848	5.1067
Correlat.	0.3474	1.0000	0.8128	0.5540	0.4305	0.3152	0.2295	0.3271	0.9253
d_index	1.5985	0.0000	0.0838	0.1760	0.2483	0.3863	0.4647	0.5613	0.0488

these scales, the local variance values of the resampled reference image (L) are nearly zero because of the quite high resolution ratio between the two images, so that equation (6) simply reduces to the limiting case of equation (7).

With increasing window size (from 15 to 49), entropy, correlation and deviation index start changing significantly. Entropy and deviation index gradually increase with window size, whereas correlation values decrease. Referring to the numbers, the closest match to the ARSIS filtered image is the Local Mean Variance Matched image with a window size of 15, although bigger window sizes also produce good results with still quite high correlation values (0.8 for window size 49). Visually, the fused images obtained with large window sizes (from 15 to 49) remain very close to the ARSIS filtered image.

5.2. The LMM filter (table 4, figure 5)

Here too, by construction, the mean of the fused images remains practically constant for all filtering window sizes. Standard deviation, entropy and deviation index increase
steadily with filter window size, while correlation decreases continuously. Comparing the numbers, the ARSIS filtered image lies somewhere in between the LMM3 and LMM5 filtered images. Visually, for these small window sizes, the LMM algorithm produces images the most similar to the ARSIS image.

5.3. The HPF filter (table 5, figure 5)

The HPF filter, introduced by Showngerdt [9], simply adds a high pass filter of the high resolution image to the resampled low resolution channel. The high pass filter has a mean value of zero and therefore preserves the mean of the filtered images for all filtering window sizes. Standard deviation, entropy and deviation index increase rapidly with filtering window size, while correlation diminishes very fast. Although considered one of the best integration methods so far [7], it does not compare favourably with the three other integration methods described, producing images with more distorted spectral values.

6. CONCLUSIONS

This study has shown that data integration of a high spatial resolution image into low spatial resolution channels of a multispectral image can conveniently be performed by mean of intensity matching the high resolution image with the low resolution channels. This intensity matching is performed by locally adjusting the mean (LMM) or the mean and variance (LMVM) of the high resolution image to that of the low resolution channels.

The spectral values of the original low resolution channels are the best preserved when filtering the images with the smallest possible window sizes. In the case of the LMM algorithm, window sizes of 3 to 5 produce the results which compare best with

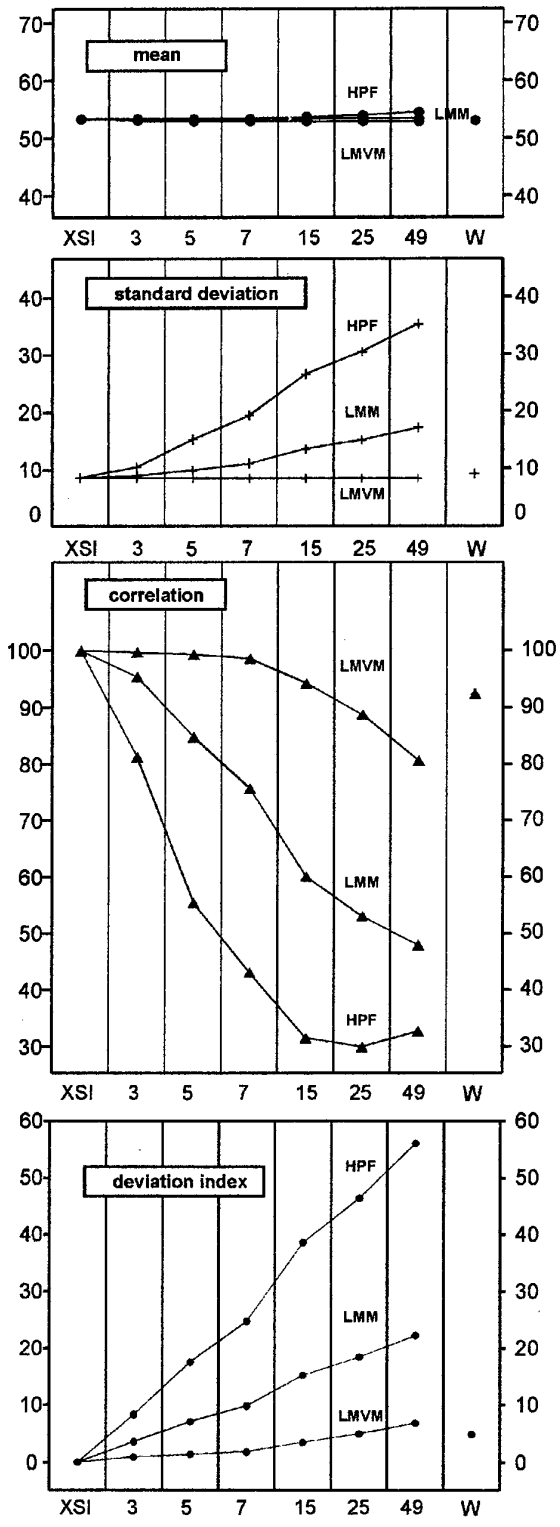


Figure 5
LMVM, LMM and HPF filtered image statistics in function of filtering window size. XSI is the reference image. W is the image filtered by the ARSIS method.

the image produced by the ARSIS method. In the case of the LMVM algorithm, the smallest filtering window sizes produce no significant difference between the original channel and the fused image. Only with bigger window sizes does the information content of the fused image start to increase significantly while still preserving the spectral characteristics of the original channels.

AKNOWLEDGEMENTS

This research was partially conducted within the framework of a cooperative pilot research project on feature extraction from high spatial resolution satellite imagery and a joint Sino-Belgian research project on spatiomapping funded by the Belgian OSTC (contracts n° T3/12/47 and IN-CH 005). The authors wish to thank Dr. T. Ranchin for providing the integrated image with the ARSIS method as also for his helpful suggestions and comments.

REFERENCES

[1] Chavez, P.S. Jr., (1986), Digital merging of Landsat TM and digitised NHAP data for 1: 24,000 scale image mapping, Photogrammetric Engineering and Remote Sensing, 52, 10, pp.1637-1646.

[2] Chavez, P.S. Jr., Sides, S.C. and Anderson, J.A., (1991), Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT Panchromatic, Photogrammetric Engineering & Remote Sensing, 57, 3, pp. 295-303.

[3] Cliche, G., Bonn, F. , Teillet, P., (1985), Integration of the SPOT panchromatic channel into its multispectral mode for image sharpness enhancement, Photogrammetric Engineering and Remote Sensing, 51, 3, pp. 311-316.

[4] Welch, R. and Ehlers, M., (1987), Merging Multiresolution SPOT HRV and Landsat TM data, Photogrammetric Engineering and Remote Sensing, 53, 3, pp. 301-303.

[5] Sabins, F.F. Jr., (1987), Remote Sensing. Principles and Interpretation. 2e ed.,Freeman, New York, 449 p.

[6] Lillesand, T.M. and Kiefer, R.W., (1987), Remote Sensing and image interpretation, 2e ed., John Wiley & Sons, 721 p.

[7] Carper W.J., Lillesand T.M., Kiefer, R.W, (1990), The use of Intensity-Hue-Saturation transformations for merging SPOT Panchromatic and Multispectral Image Data, Photogrammetric Engineering and Remote Sensing, 56, 4, pp. 459-467.

[8] Showengerdt, R.A., (1980), Reconstruction of multispatial, multispectral image data using spatial frequency contents, Photogrammetric Engineering & Remote Sensing, 46, 10, pp. 1325-1334.

[9] Vrabel J., (1996), Multispectral imagery band sharpening study, Photogrammetric Engineering & Remote Sensing, 62, 9, pp. 1075-1083.

[10] Ranchin, T., (1993), Applications de la transformée en ondelettes et de l'analyse multirésolution au traitement des images de télédétection. Unpublished PhD Thesis, University of Nice-Sophia Antipolis, 146 p.

[11] Garguet-Duport, B., Girel, J., Chassery, J-M ., and Pautou, G., (1996), The use of multiresolution analysis and wavelets transform for merging SPOT panchromatic and multispectral image data, Photogrammetric Engineering & Remote Sensing, 62, 9, pp. 1057-1066.

[12] Yocky, D.A., (1996), Multiresolution wavelet decomposition image merger of Landsat thematic mapper and SPOT panchromatic data, Photogrammetric Engineering & Remote Sensing, 62, 9, pp. 1067-1074.

[13] Ranchin, T., Wald, L, (1993), The wavelet transform for the analysis of remotely sensed images. International Journal of Remote Sensing , 14, 3, pp 615-619.

[14] Ranchin, T., Wald, L., Mangolini, M. and Penicand, C., (1996), On the assessment of merging processes for the improvement of the spatial resolution of multispectral SPOT XS images. In Proceedings of the conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Cannes, France, February 6-8, 1996, Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, pp. 59-67.

[15] Richards, A.J.,(1986), Remote Sensing Digital Image Analysis. An Introduction., Spriger-Verlag, 281 p.

[16] Joly, G., (1986), Traitements des fichiers-images, Télédétection Satellitaire 3, Ed. Paradigme, 137 p.

[17] Frost, V.S., Styles, J.A., Shanmugan, K.S., and Holzman, J.C. , (1982), A model for radar images and its application to adaptive digital filtering of multiplicative noise, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAM 1-4, pp. 157-166.

[18] Lee, J.S., 1981, Refined filtering of image noise using local statistics, Computer Graphics and Image Processing, 15, pp. 380-389.

[19] Fusco L. , Trevese D., (1985), On the reconstruction of lost data in images of more than one band., *International Journal of Remote Sensing,* 6, pp.1535-1544.

[20] Liu, J.G. and McM. Moore, J., (1990), Hue image RGB colour composition. A simple technique to suppress shadow and enhance spectral signature, International Journal of Remote Sensing, 11, 8, pp. 1521-1530.

[21] Edwards K., Davis, P.A., (1994), The use of Intensity-Hue-Saturation transformation for producing colour shaded-relief images, Photogrammetric Engineering & Remote Sensing, 60, 11, pp. 1369-1374.

[22] Moik, J., (1980), Digital processing of remotely sensed images. NASA Special Publication 431, Washington DC.

[23] Mather, P.M., (1987), Computer processing of remotely sensed images. An Introduction. John Wiley & Sons, 352 p..

[24] Costantini, M., Farina, A., Zirilli, F., (1997), The fusion of different resolution SAR images, Proceedings of the IEEE, 85, 1, pp. 139-146.

PAPER No. 28

DISCUSSOR'S NAME: B. Dacre-Wright

COMMENT/QUESTION:

Quel traitement d'interpolation avez-vous appliqué, lors du rééchantillonnage, pour éviter les distorsions spatiales dans l'image obtenue.

(What interpolation processing did you apply during re-sampling to avoid spatial distortions in the image obtained?)

AUTHOR/PRESENTER'S REPLY:

Tous les exemples traités et illustrés ont été rééchantillonnés par une convolution bicubique, qui est la moins mauvaise méthode de rééchantillonnage utilisable.

(All the examples presented were re-sampled by means of bi-cubic convolution, which is the least imperfect re-sampling method available).

PAPER No. 28

DISCUSSOR'S NAME: E. Schweicher

COMMENT/QUESTION:

1. Pourriez-vous définir ce que vous entendez par capteur à faible résolution spatiale et à haute résolution spectrale.

2. Par rapport aux différents types de capteurs que vous avez cités, dans quelle classe de capteurs mettriez-vous les radars?

(1. Could you define what you mean by low resolution sensor with high spectral resolution?

2. Of the different types of sensors you have mentioned, which category of sensor would you put radars in?)

AUTHOR/PRESENTER'S REPLY:

1. Dans le domaine de la télédétection spatiale on désigne généralement par haute résolution spectrale les canaux des images XS (SPOT) TM (Landsat) LISS (IRS), par opposition aux images panchromatiques qui ont une très large bande passante et qui ont de ce fait en général une meilleure résolution spatiale.

2. En théorie la méthode de fusion proposée permet l'application à tout type d'images. Un des premiers buts de cette méthode est d'essayer de préserver le mieux possible l'information spectrale d'origine afin de produire des images qui seront plus facilement interpretables afin de produire des images qui seront plus facilement interprètables ou qui seront plus facile à classer. L'information spectrale d'une image radar est "très différent" de celle du domaine visible or infra-rouge, et ne donne pas lieu à une signature spectrale du même type, mais en principe on peut également préserver le mieux possible cette information spectrale radar dans le cadre de l'application souhaitée.

(1. In the field of space remote sensing we generally reserve the term high spatial resolution for image channels XS(SPOT), TM(LANDSAT) and LISS(IRS), as opposed to panchromatic images which have a very broad bandwidth and which as a result have a better spatial resolution.

2. In theory, the fusion method proposed can be applied to all types of images. One of the main aims of this method is to try to preserve the original spectral data in the best possible way in order to be able to produce images which are easier to interpret or to classify. The spectral data obtained from a radar image is "very different" from that in the visible or infrared domain and does not produce a spectral signature of the same type, but in principle this radar spectral data can also be preserved in optimum fashion by the application in question.)

PAPER No. 28

DISCUSSOR'S NAME: C. Remus

COMMENT/QUESTION:

Existe-t-il une différence en grain de traitement entre la méthode proposée et la méthode ARSIS basée sur des calculs d'ondelettes.

(Is there a difference in grain of processing between the method proposed and the ARSIS method based on wavelet calculations?)

AUTHOR/PRESENTER'S REPLY:

La rapidité de la méthode ARSIS ne nous est pas connu. La méthode de fusion par ajustement des moyennes et variances locales est basée sur des paramètres statistiques simples à calculer, et la vitesse de calcul peut être rendue indépendante de la taille de la fenêtre de filtrage. On n'a pas fait de test précis sur l'optimisation, mais c'est très rapide.

(We have no information on the speeed of the ARSIS method. The method of fusion by adjustment of averages and local variances is based on statistical parameters which are simple to calculate, and the computation speed can be made independent of the size of the filtering window. We haven't run a specific test on optimisation, but it's very fast.)

The Fusion of Organic and Non-Organic Identity Information Sources Using Evidential Reasoning

Éloi Bossé, Jean Roy, Stéphane Paradis

Defence Research Establishment Valcartier
Decision Support Technology Section
2459 Pie XI Blvd. North
P.O.8800, Val-Bélair, G3J 1X5 CANADA

1. SUMMARY

The aim of this paper is to explore the problem of identity data fusion in the context of naval warfare where commanders and their staff require access to a wide range of information to carry out their duties. With the increasing use of the remote sources and the increasing use of information that has already been fused, there is a requirement to define an identity data fusion function that can handle organic and non-organic, local and remote types of information characterized by different accuracy and timeliness. This paper is a step towards the definition of an architecture that can perform the fusion of organic and non-organic identity data through the use of statistical analysis rooted in the Dempster-Shafer theory of evidence. The goal is to offer the decision maker a quantitative analysis based on statistical methodology that can enhance his/her decision making process regarding the identity of detected objects.

2. INTRODUCTION

Research and Development (R&D) activities by the Data Fusion and Resource Management (DFRM) Group at the Defence Research Establishment Valcartier (DREV) are aimed specifically at investigating system integration concepts for the automation of Command and Control (C2) processes dealing with data and information management and tactical decision making and action implementation issues at the shipboard level. The group has developed an approach to help counter the anticipated threat to our surface ships by increasing the Above Water Warfare (AWW) defence capability of Canadian Patrol Frigates (CPFs) through the development of a real-time advisory Decision Support System (DSS) (Ref. 1) that supports the tactical decision making and action execution processes in a ship's Operations Room.

The increasing tempo and diversity of open-ocean and littoral scenarios, and the volume and imperfect nature of data to be processed under time-critical conditions, threats characterized by high speeds, low approach altitudes or steep dive trajectories, maneuverability, and the ability to deceive defensive systems using countermeasures pose significant challenges for future shipboard Command and Control Systems (CCSs) and the operators who must use these systems to defend their ship and fulfill their mission.

The littoral environment is especially stressful. The presence of dense, commercial air traffic and merchant shipping challenges the operators in distinguishing between hostile, neutral, and friendly tracks. The proximity to hostile shores decrease the available battlespace and warning time and enable the use of land-based Electronic Counter-Measures (ECM), increase clutter levels and offer more vulnerability against land-based missile attacks. Furthermore, weapons developers recognize the vulnerabilities of the defensive systems and develop or evolve weapons in an attempt to defeat them. As the threat becomes more sophisticated, additional aids and decision support to the operators will be increasingly required.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

The proposed DSS (Ref. 2) would: continuously take in data from the ship's sensors and other information sources; form and dynamically maintain an accurate AWW tactical picture, and thereby provide the most likely interpretation of the tactical situation; formulate strategies and plans for responding to anticipated or actual threats, including, as necessary, options to defend the ship using the best possible combination of hardkill and softkill weapons or other defensive means; and present fused information and decision support analysis results with the opportunity for the Commanding Officer and AWW team to accept or reject recommended actions and plans in a timely manner, and coordinate and direct the implementation of actions and plans that result from interaction with the human decision maker.

The DSS would be an embedded component of the ship's combat system, integrated within the CCS system, providing real-time implementations of functions for Multi-Sensor Data Fusion (MSDF), Situation and Threat Assessment (STA), and Resource Management (RM). In view of the functional integration involved, we call this subsystem of the ship's CCS an MSDF/STA/RM system.

This paper addresses specifically the identity data fusion aspects of the MSDF/STA/RM system that may encompass the MSDF and STA domains such as defined in Refs. (1-3). It discusses the differences between the problem of fusing attributes (often from local and organic sources) and identity declarations (most of the time from non-organic and remote sources). The paper also addresses the refinement of the identity declarations in the situation assessment process by the use of other a priori information as well as the information inferred by the behavior and cluster analysis functions. The use of statistical analysis rooted in the Dempster-Shafer theory of evidence is proposed. The paper is organized as follows. Section 3 presents two generic models: one for MSDF and the other for STA. Section 4 the kinds of potential identity information that are going to be fused. In Section 5 we discuss the various architectures for identity fusion. Section 6 describes the identity information fusion process for organic and non-organic information.

3. GENERIC MODELS FOR MSDF AND STA FUNCTIONS

Throughout the 1980s, the three U.S. military services pursued the development of tactical and strategic surveillance systems employing data fusion and supported extensive research in the areas of target tracking, target identification, algorithm development for correlation (association) and classification, and the application of intelligent systems to situation assessment (Ref. 4). The large amount of fusion-related work in this period raised some concern over possible duplication of effort. As a result, the Joint Directors of U.S. Department of Defense (DoD) Laboratories (JDL) convened a Data Fusion Subpanel to (1) survey the activities across all services, (2) establish a forum for the exchange of research and technology, and (3) develop models, terminology and a taxonomy of the areas of research, development and operational systems. As a result of many years of effort to establish standardization and stability in the lexicon of data fusion, the definition of many terms is slowly achieving consensus across the diversified application community. Problem-specific nuances and shading in these definitions remain but agreement on a meaningful subset of terms does seem to exist.

Data Fusion (DF) is an adaptive information process that continuously transforms the available data and information into richer information, through continuous refinement of hypotheses or inferences about real-world events, to achieve refined (and potentially optimal) kinematic and identity estimates of individual objects, and complete and timely assessments of current and potential future situations and threats (i.e., contextual reasoning), and their significance in the context of operational settings. The process is also characterized by continuous refinements of its estimates and assessments, and by evaluation of the need for additional data and information sources, or modification of the process itself, to achieve improved results. Given these considerations, a complete data fusion system can typically be decomposed into four levels (Refs. 2-3):

Level 1 - Multi-Source Data Fusion (MSDF);
Level 2 - Situation Assessment (SA);

Level 3 - Threat Assessment (TA); and,
Level 4 - Process Refinement Through
Resource Management (RM).

Each succeeding level of data fusion processing deals with a higher level of abstraction. Level 1 data fusion uses mostly numerical, statistical analysis methods, while levels 2, 3 and 4 data fusion use mostly symbolic, Artificial Intelligence (AI) methods.

Multi-Source Data Fusion (Level 1)

Multi-source data fusion (MSDF) is concerned solely with individual objects, first in associating the sensor outputs with specific known objects or using them to initiate new objects. Level 1 processing uses sensor data to correctly and quickly derive the best estimates of current and future positions for each hypothesized object. In addition, inferences as to the identity of the objects and key attributes of the objects are developed.

Situation Assessment (Level 2)

Based on incomplete and inaccurate sets of data and information, situation assessment (SA) is devoted to the continuous inference of statements about the hypothesized objects provided by the lower level data fusion function in order to derive a coherent, composite tactical picture of the situation. This picture must be described in terms of groups or organizations of objects so that enemy intent can be estimated in the next level and decisions can be made by decision makers about how to use war fighting assets.

Threat Assessment (Level 3)

Threat assessment (TA) is focused at the details necessary for decision makers to reach conclusions about how to position and commit the friendly forces. By coupling the products of situation assessment with the information provided by a variety of technical and doctrinal databases, TA develops and interprets a threat oriented perspective of the data to estimate the enemy capabilities and lethality, identify threat opportunities in terms of the ability of own force to engage the enemy effectively, estimate enemy intent (i.e., provide indications and warnings of enemy intentions), and determine levels of risk and danger.

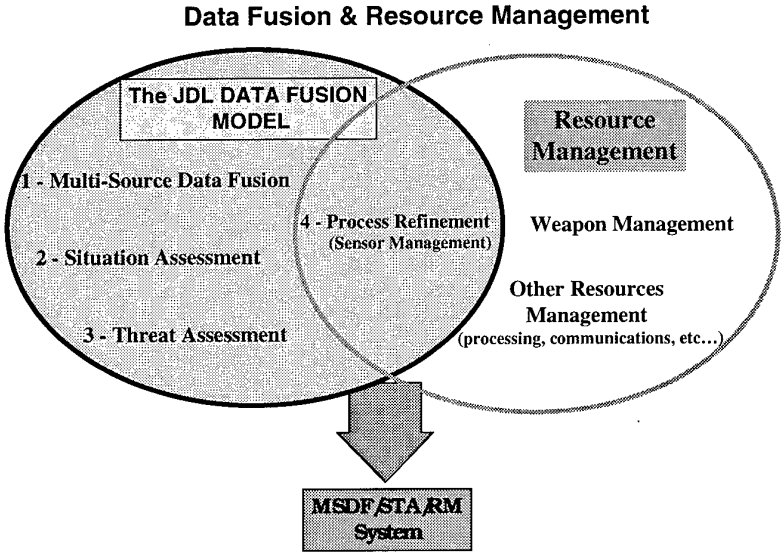


FIG. 1 - Overlap between the data fusion and resource management domains

Resource Management (level 4 - Process Refinement through Sensor Management)

Information resource management, level 4 processing, closes the loop by first examining and prioritizing what is unknown in the context of the situation and threat and then developing options for collecting this information by cueing the appropriate sensors and collection sources. Resource Management refers to the continuous process of planning, coordinating and directing the use of the ship or force resources to counter the threat. It is concerned with issues of both command and control. Note that resource management in the context of level 4 fusion is mainly concerned with the information gathering process refinement (i.e., sensor management). However, the overall domain of resource management also encompasses the management of weapon systems and other resources. Fig.1 illustrates the overlap between the data fusion and resource management domains.

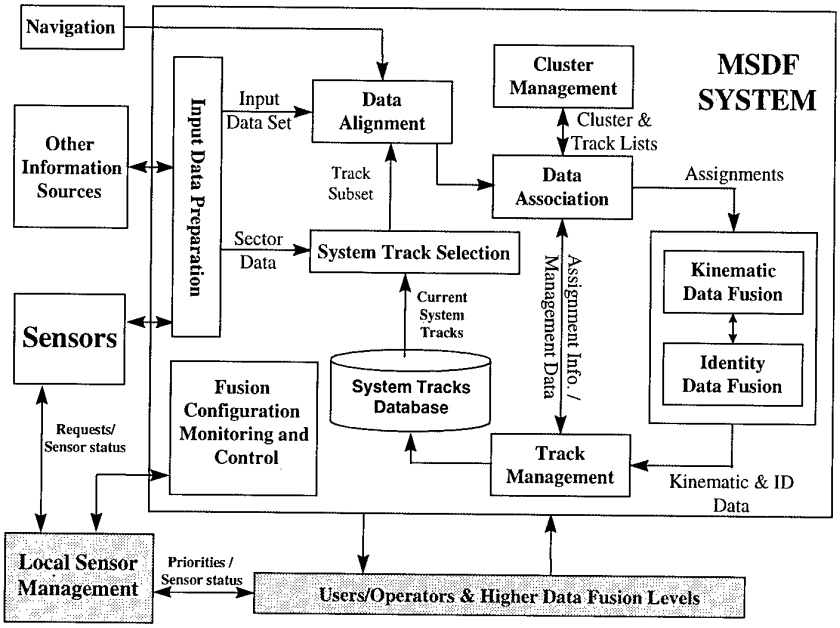


FIG. 2 - A Generic MSDF Model

3.1 A generic MSDF model

Figure 2 shows a generic MSDF model where the key functions are identified. The processing can be divided into blocks such as:

- data alignment (spatial, time),
- data association,
- target kinematic data fusion,
- target identity data fusion,
- track management process,
- cluster management process,
- input data preparation,
- track database,
- fusion configuration monitoring and control.

A detailed description of all these processes is provided in Ref. 5.

In any MSDF system, sensor data alignment in time and space must take place before any fusion can be performed. Navigation data is used to estimate and remove the effects of sensor motion from the received data. The functions of data association (labeling measurements from different origins and/or sensors, at different times, that correspond to the same object or feature) and data fusion (combining measurements from different times and/or different sensors) are also required in one form or another in essentially all multiple sensor fusion applications: data association determines what information should be fused, the fusion function performs the fusion.

In addition to the MSDF system which detects, localizes, and identifies targets, the local sensor management, on the basis of an evolving picture, and under the command of the overall Command and Control (C2) resource management, manages the information that the MSDF might receive by pointing, focusing, maneuvering, and adaptively selecting the modalities of its sensors and sensor platforms. The overall C2 resource management has the responsibility of first examining and prioritizing what is unknown in the context of the situation and threat and then developing options for collecting this information by sending priorities to the MSDF local sensor management function. The interaction of the local sensor management and the MSDF function is done through the fusion configuration, monitoring and control which is responsible for the initialization of the MSDF system, the setup and the adjustment of the various MSDF algorithm parameters to control the quality of the MSDF product.

3.2 A generic STA model

Ref. 3 presents a generic model for situation and threat assessment influenced by the human's mental processing. The model evolved from a three level descriptive model of situation awareness, where the first level is concerned with the perception of the elements in the environment, the second level is about the comprehension of the current situation and the last level deals with the projection of the future states of the situation.

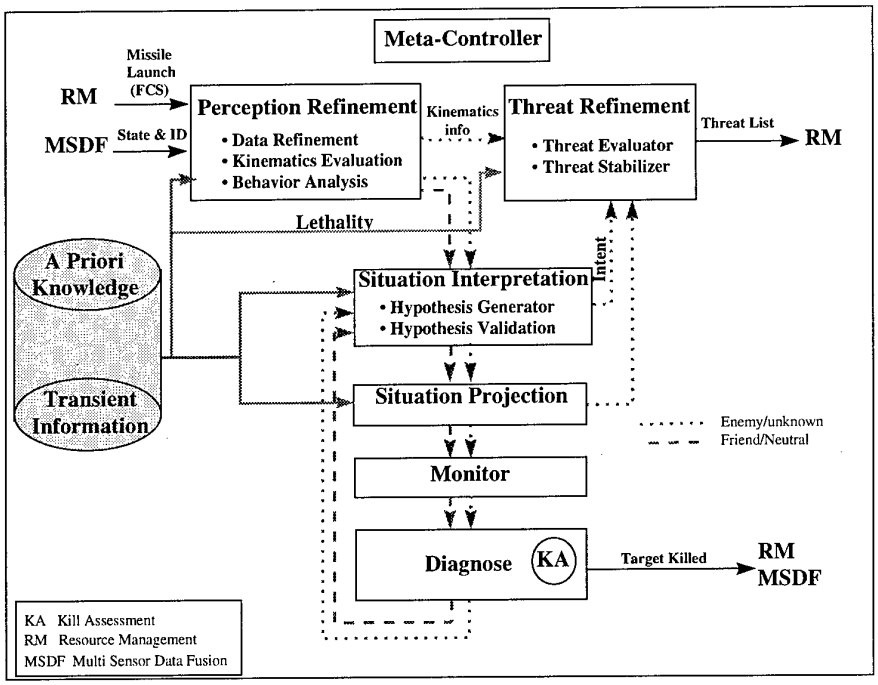


Fig. 3 - Generic Model for STA

The proposed generic model for STA is illustrated in Fig. 3 and consists of a Perception Refinement module, a Threat Refinement module, a Situation Interpretation module, a Situation Projection module, a Monitoring module and a Diagnosis module. The generic model has access to a priori knowledge and transient information and its behavior is modulated by a Meta Controller taking account of processing priorities, processing time and quality of the information. The idea here is not to describe all these entities but only those which can be used to complete the identity data fusion process i.e the perception refinement module. We refer the reader to Ref. 3 for a detailed description of the STA model of Fig.3.

3.2.1 The Perception Refinement module

The Perception Refinement module is composed of three sub-processes : the Data Refinement (DR) sub-process, the Kinematic Estimation (KE) sub-process and the Behavior Analysis (BA) sub-process. The idea of DR is to refine the track data (position, identity) already generated by MSDF to examine the data set for incompleteness and contradictions and to establish relationships among the entities (in terms of proximity, functionality and dependency) with the help of external data sources if necessary. At this point, no inferences about the situation is done. The only results obtained from this sub-process is perceptual refinements.

The Kinematic Estimation sub-process is used for the computations of kinematic information (mean line of advance, closest point of approach (CPA), time of flight (TOF)...) of a track for Threat Refinement processing, and for engageability calculations done in Resource Management. An history function, which records the positional tracking and identification information in time, is needed to accomplish these kinematic calculations. By means of this history function, the recording of the track information will allow us to address enemy information countermeasures (i.e. information warfare).

Behavior Analysis (BA) is the last sub-process under Situation Refinement and is used to analyze the behavior analysis of entities and/or cluster in order to help to the refinement of the data set. Also, BA becomes a preliminary step for the second level of the situation awareness model by providing processing cues or evidence about the behavior of the track or cluster to the Situation Interpretation module for the interpretation and understanding of the current tactical situation. The cues are obtained from the analysis of the refined information (track and cluster's kinematic data and identification) done in DR, the a priori knowledge and from transient information (i.e. electronic emissions from ESM, the datalink, participating units...). BA includes functions, such as corridor correlation, maneuver/pattern identification that generate the required cues to infer about the tactical picture.

4. KINDS OF IDENTITY INFORMATION

In a maritime environment, various surveillance systems, electronic intelligence and human observations are examples of information sources available to the commander. Output from these sources is partitioned according to the type of information they provide; output data may be characterized as either positional or identity information. *Positional information* represents the dynamic parameters describing the movement associated with an object (contact). This generally includes position, velocity and acceleration. *Identity information* can be defined as declarations, propositions or statements that contribute to establish the identity of an object. Equivalently, identity information may be seen as information from various sources that helps in distinguishing one object from another. Possible values for identity information can span the range from *sensor signals*, to *attributes*, to *identity declarations*, as depicted in Fig. 4.

The sensor signals represent some characteristics of the energy sensed. Attributes such as size, shape, degree of symmetry, emitter type, etc. are inferred from these characteristics. Identity declarations specify the detected object; in the Canadian Navy, for example, they can consist of a general classification of which the observed object is a member (surface combatant), a specific type of ship (frigate), a specific class (City Class) or a unique identity (Ville de Québec). Therefore surface combatant, frigate, City

Class and Ville de Québec are all examples of identity declarations. Identity declarations can also include information concerning the threat designation of an object: pending, unknown, assumed friend, suspect, friend, neutral or hostile.

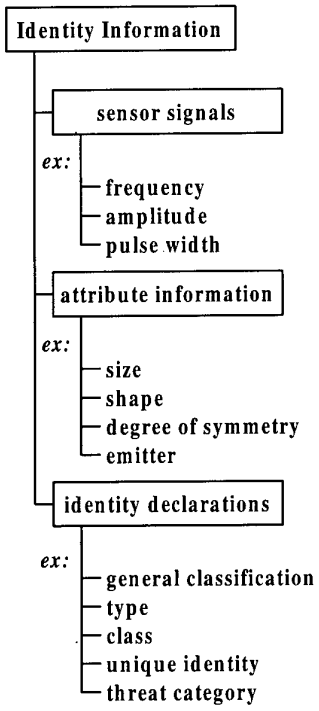


Fig. 4 - Identity Information

The quantity of identity declarations available may become quite imposing and diverse. A potential way of organizing identity declarations might be to use the NATO standards for representing maritime tactical information (STANAG 4420, Ref. 6). These standards are in the form of charts that define the full range of tactical information required by the operational user at the command level. The charts were created to establish the basis for developing a standardized representation of spatially displayed tactical data using symbology and colour for NATO maritime units. Ref. 7 presents an adapted version of a tactical information hierarchy for surface and air objects respectively; certain levels and entries have been omitted for simplicity.

The names in the various boxes, which are indicative of the taxonomy used in STANAG 4420, represent identity declaration entities. These entities illustrate the sort of identity declarations provided by various sources which must be combined in order to obtain an estimate of an object's identity.

In addition, the identity information can be characterized by being local, remote, organic and non-organic. The notion of local and remote are self-explaining but the organic and non-organic require to be defined.

Organic Information

Data collected and managed by assets under Commanding Officer's direct control can be defined or characterized as organic information. Organic information must be sufficiently timely and accurate to be used in real-time, responsive systems. Consequently, it can be used to produce a local tactical picture related to the situation. Note that remote sources can be organic and local sources can be non-organic. Here are examples of organic information: radars, datalink (e.g. Link-11), ESM, weather, ...etc.

Non-Organic Information

Data collected by agents not under CO direct control, is referred as non-organic information. Non-organic information is less timely, reduced in accuracy, differently structured and has differing identification confidence levels. Transient information such as intelligence reports or PU's information is non-organic as means to support the various processes providing the commander at sea to gain a level of situational awareness.

Other Type of Information

Another source of data is the human contribution to the identification process through a Human Computer Interface (HCI). Human can participate actively by providing inputs, based on his experience, his training and his awareness of the situation within the identification process. The precise nature of this human-computer interaction remains to be determined.

A Priori Knowledge

A Priori Knowledge contains static information as means to support the various processes providing the decider to gain a level of situational awareness. That knowledge is a component within the STA model as opposed to inputs derived from external sources. A Priori Knowledge is used by the Perception Refinement module to refine identity information. Here are some examples of a priori knowledge source: social and political, geographical, platforms characteristics, mission guidelines, weapons characteristics, corridor and flight paths, EM characteristics, lethality, emitter characteristics, doctrines, ...etc.

5. ARCHITECTURES FOR IDENTITY FUSION

One of the key issues in developing a MultiSource Data Fusion (MSDF) system is the question of where in the data flow to actually combine or fuse the data. The MSDF architecture is an important issue since the benefits are different depending on the way the sensor or other source data are combined. There are four broad alternatives to fusing positional information: at the signal level, at the contact (plot) level, at the track level and finally a hybrid approach which allows fusion of either contact or track data. The benefits of theses architectures for positional fusion have been extensively analyzed in (Ref. 8). For identity fusion, there are several types of architectures which can be used: signal level fusion, feature level fusion and decision level fusion. Fig. 5 displays the three fusion architectures providing the structural basis for fusing identity information in the case of two sensors. It should be noted that the output from each level of processing represents the three categories of identity information depicted in Fig. 4.

There is no one universal architecture which is applicable to all situations or applications. The selection of the MSDF architecture type should be aimed at optimizing the target detection, tracking and identification performance required for a specific platform given its missions. However, the selection is also constrained by the technological capabilities (both hardware and software). It depends on the quality of the sensors being fused, the availability of computer processing power, the bandwidth of the available data transmission paths, and the degree to which operator intervention is required or desired.

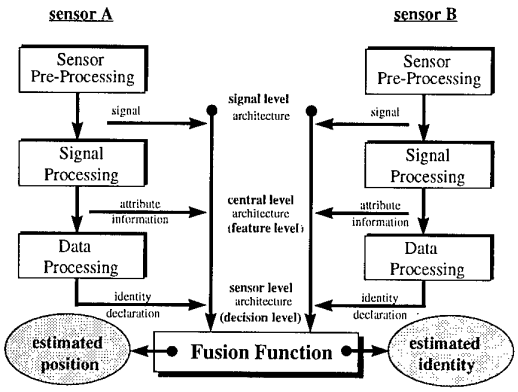


FIG. 5 - Identity Information Fusion Architectures

6. THE IDENTITY DATA FUSION PROCESS

In Section 4 the sources of information are classified as organic and non-organic with the possibility that remote sources can be organic and local sources can be non-organic. With the increasing use of the remote sources (e.g. wide area sensors) and the increasing use of information that has already been fused, there is a requirement to define an identity information fusion that can handle both organic and non-organic information characterized by different accuracy and timeliness (real-time, near real-time and non real-time). Fig. 6 shows an example of an identity information process for two platforms fusing identity information from organic and non-organic sources. When remote data is to be combined with organic fused data, the relationship between the two sets of data should be indicated to avoid the problem of data duplication or data looping. In Fig. 6, we propose a distributed approach where two data fusion centers are maintained at the platform level: an organic data fusion center and a global data fusion center.

The organic data fusion center produces the identity declarations from the local and remote organic sources and the result is broadcast to all other platforms. In Fig. 6, the organic data fusion results of platform #1

are made available to the global data fusion center of platform #2 and the same way, the organic data fusion results of platform #2 are made available to platform #1. Two track databases need to be maintained: an organic and a global. Fig. 6 also shows the remote organic sources. For example in an air engagement, some sensors may be slaved directly to some platforms for Cooperative Engagement Capability (CEC) in which selected data is transmitted from one unit to another through dedicated communications.

The global data fusion takes place at the platform as well but the results are made available only under dedicated broadcast to avoid data looping. Ideally, if each fusion center utilizes the same algorithms i.e. to produce the same database result when the same data are provided as inputs, each platform will share the same global picture. However, this is not likely to happen. Similar algorithms, procedures and format are most likely foreseeable. The goal here is for each platform to share a similar tactical picture where the differences are within the range of the uncertainty of the overall fusion process. To achieve this interoperability, some international standards need to be established and agreed upon.

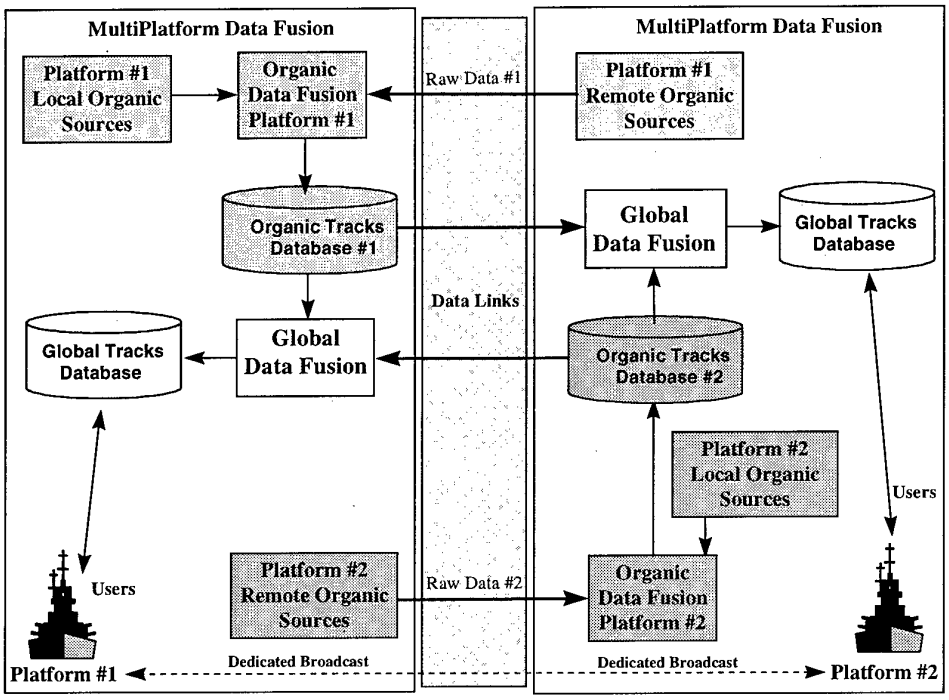


FIG. 6 - Multiplatform Organic and Non-Organic Data Fusion

6.1 The Organic Identity Information Process

In the organic identity information fusion process of Fig. 7, the attribute information and identity declarations obtained from various sources are compared with a Platform Data Base (PDB) containing all the possible identity values that the potential target may take. Each record of this database contains information related to the measured sensor attributes and identity declarations. Therefore, each sensor's attribute information is translated into a subset of the PDB and a confidence level for each subset is then computed. If the available identity information is only provided by self-contained sensors producing identity declarations in an autonomous fashion or by remote sources offering identity

might be quite different from those used in the organic identification process. The proposition construction subprocess is dealing only with identity declarations and then does not require the use of fuzzy logic. Since the identity declarations are hierarchically structured, fast algorithms such as the Shafer-Logan (Ref. 7) can be used. Ref. 7 also discusses how the hierarchy affects the decision rules. Unlike the organic identification process, where only organic information is used in the refinement of the identity propositions, the global identification process makes use of all the available information to refine the identity propositions.

7. CONCLUSION

This paper proposed an architecture that can perform the fusion of organic and non-organic identity data through the use of statistical analysis rooted in the Dempster-Shafer theory of evidence. The architecture has the benefit to provide a solution to the problem of data duplication or data looping when remote data is to be combined with organic fused data. The distributed approach consists of two data fusion centers maintained at the platform level: an organic data fusion center and a global data fusion center.

8. REFERENCES

1. Chalmers, B.A., " Design Issues for a Decision Support System for a Modern Frigate ", Proc. Of the 2nd Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment, Naval Air Warfare Centre, Maryland, USA, June 1997, pp.127-149.

2. Roy, J., Chalmers, B.A., Carling, R., Bossé, É., " Overview of Shipboard Data Fusion and Resource Management R&D Results and Rationale for its Real-time Implementation in the ASCACT Testbed", DREV-R-9518, April 1996.

3. Paradis, S., Chalmers, B.A., Carling, R., Roy, J., Bossé, É., "Towards A Generic Model For Situation And Threat Assessment"", DREV-R-9622, December 1996.

4. Llinas, J., Waltz, E.L., "Multi-Sensor Data Fusion", Artech House, Norwood, MA, 1990.

5. Roy J., Bossé, É., "Generic Model for a Multi-Source Data Fusion System", DREV-R-97xx, (to be published in 1997)

6. "STANAG 4420. Display Symbology and Colours for NATO Maritime Units". Edition 1. NATO UNCLASSIFIED.

7. Bossé, É., Roy, J., "Fusion of Identity Declarations from Dissimilar Sources Using the Dempster-Shafer Theory", *Optical Engineering, special section on Sensor Fusion*, March 1997.

8. Roy, J., Bossé, É., Duclos-Hindié, N., " Quantitative Comparison of Sensor Fusion Architectural Approaches in an Algorithm-Level Testbed", *Signal and Data Processing of Small Targets 1996, SPIE Proceedings Vol. 2759*, Orlando, 8-10 April 1996.

9. Simard, M.A., Jouan, A., " Platform Database and Evidential Combination Method to Fuse ESM Detections with the Attribute Data From Other Sensors", AGARDS Symposium on " EW Integration for Ships, Aircraft and Land Vehicles, Ottawa, May 1997.

10. Simard, M.A., Bossé, É., " Identity and Attribute Information Fusion Using Evidential Reasoning ", SPIE Proceedings, Vol. 3067, April 1997.

PAPER No: 29

DISCUSSOR'S NAME: M. Balci

COMMENT/QUESTION:

1. Is there any difference between remote-organic and non-organic information?

2. What do you mean by one single identity information from each sensor?

3. Do you give any reference for defining m (θ) values?

AUTHOR/PRESENTER'S REPLY:

1. Yes, remote organic information means that the decider or the commander has control over units that are remote, while non-organic means that the decider has no control over it.

2. If there is more than one attribute given by a particular sensor, we propose to integrate those attributes in a single proposition.

3. Up until now, the m(θ) values have been determined by common sense and are rather ad hoc. We plan to look more carefully at this problem in our future work.

DISCUSSOR'S NAME: M. Desbois

COMMENT/QUESTION:

I fully agree on the need to separate organic and non-organic data to reject "incestual fusion" (reinforcing the ID by using your own information). However, among the ID sources you have IDBO (Identification by Origin) and the Track Behaviour (altitude, aspect, heading, ...).

1. Do you process these sources.

2. If yes, would you classify these sources as organic data?

AUTHOR/PRESENTER'S REPLY:

1. Yes, we could.

2. It depends on the application, but generally speaking if the source delivering that information is under the control of the decider I would classify it as organic data.

PAPER No. 29

DISCUSSOR'S NAME: E. Schweicher

COMMENT/QUESTION:

Dans les menaces requérant une réaction rapide incluez-vous les mines dérivantes et les missiles à vol rasant tels que le missile EXOCET?

(Do you include drifting mines and sea-skimming missiles such as EXOCET in the category of threats requiring rapid reaction?)

AUTHOR/PRESENTER'S REPLY:

Absolûment. Pour ce qui a trait aux missiles à vol rasant, le système de fusion devrait être capable d'identifier cette menace le plus tôt possible et de déléguer à un sous-système automatique le soin de la pister avec grande précision.

(Absolutely. In so far as concerns sea-skimming missiles, the fusion system should be capable of identifying this threat as soon as possible and assigning the task of high precision tracking to an automatic sub-system.)

PIXEL FUSION FOR ROADS DETECTION AND CLASSIFICATION

S. Fabre *, ONERA
X. Briottet, ONERA
P. Marthon, ENSEEIHT
A. Appriou, ONERA
* ONERA DOTA, Complexe Scientifique de Rangueil,
BP 4025
31055 TOULOUSE Cedex, FRANCE
Email : sfabre@onecert.fr

SUMMARY

The opportunity to process several images being acquired simultaneously over the same landscape in different spectral bands provides a better perception of the scene using data fusion methods. This new method, classifying the image in two types (target and background) using the radiometric properties of the surface, is particulary powerful for the detection of non resolved objects. The fusion method is defined by using the normal probability density hypothesis to estimate the ponderation of each initial image. The classification is done by using a threshold where analytic expression depends on the type of the "a priori" knowledge. The method is then applied on an image acquired with the hyperspectral scanner sensor : AVIRIS. The chosen target is a non resolved road in one dimension. Different criteria are used to evaluate the performance of the method : statistical properties, good classification percentage, false alarm and non detection errors. We analyze the "a priori" probability uncertainty on the image obtained after fusion.

LIST OF SYMBOLS

i : the index of the considered band (i = 1, ..., n),
j, j': the current index of the landscape; j or j' = t for the target and j or j' = b for the background,
n : number of channels to be combined,
$\vec{x}$: pixel spectral signature of component x^i (dim $\vec{x}$ = n),
$D(d_t, d_b)$: decisions made about the pixel nature,
d_t shows that pixel is belonging to the target,
d_b shows that pixel is belonging to the background,
$W(t,b)$: the two possible classes,
$c_{jj'} = c(d_j / j')$: the cost combined with the decision d_j when the nature's statement is j',
R : the conditional risk,
P(error) : the mean error probability,
$p(\vec{x})$: the density probability of the $\vec{x}$ vector,
P(j) : the "a priori" probability that a pixel belongs to the database j,
$P(j/\vec{x})$: the "a posteriori" probability that the nature's statement is j with $\vec{x}$ well-known,
$p(\vec{x}/j)$: the density probability of $\vec{x}$ according to the j nature's statement,
Σ_j : the covariance matrix of the database j,
$\vec{m}_j$: the mean vector of the database j of component m_j^i,
σ_j^i : the standard deviation of database j from the band i,
$g(\vec{x})$: the combination rule,
$\vec{a}$: the ponderation vector,
$\tilde{m}_j$: the database mean of the fusionned image,
$\tilde{\sigma}_j$: the database standard deviation of the fusionned image,
s : the threshold,
R_j : the area associated to the j database,
n_j : the pixel number of the database j whose level values belong to the R_j area,
N_j : the total pixel number on the j database,
XS_i : the i spectral band of the SPOT instrument,
L_P : the equivalent radiance at a P point,
S_{XSi} : the spectral sensibility of the XSi sensor.

1. INTRODUCTION

The large quantities of data provided by satellite and airborne sensors lead to the consideration of efficient methods as fusion that combine several information from multiple sensors and significantly increase scene perception. The data diversity and the different processing level imply the development of several fusion techniques.
In the image domain, Waltz [1] distinguishes three levels according to the information abstraction degree :

- The fusion pixel, that is close to raw observations, handles with pixel level information that are directly provided by sensors.
- The numerical fusion integrates attribute data obtained after a segmentation stage.
- The symbolic fusion combines information made up by hypothesis or suggestions.

Our study is in the context of pixel fusion. Now, the various fusion methods can gather in two classes according to the used combination rule: weighted means or weighted means ratio. So, Gonzales [2] defines the multispectral image ponderation from the principal components analysis (PCA). Likewise, Arigides [3] and Stocker [4, 5] use the bi-spectral difference in infrared domain in order to remove the artificial targets background. According to the available "a priori" knowledge nature, various criteria are used to elaborate the multispectral observations ponderation. The correlation matrix allows to merge information by minimizing the background noise power. The further knowledge of pixel signature tends towards the maximum contrast between target and background and so provides the best results in signal to noise ratio [4]. Methods based on weighted means ratio generally use the physical properties of the studied target. An application is the vegetation detection by the vegetation index [6] defined as a normalized spectral contrast. However, these performance are sensitive to external parameters : precision of the sensor radiometric calibration, atmospheric and geometric viewing conditions. Various authors have tried to minimize their impact either by introducing adjustment parameters ([7]) or by developing a non linear expression of this index ([8]). In the infrared domain, another application [9,10] is the detection of buried minefields whose principle consists in dissociating the reflexion and emission phenomenon from images acquired in the 3-5 µm and 10-12 µm bands by exploiting the radiometric behaviour difference between natural and artificial surfaces.
Therefore, fusion allows to improve classification either by directly analyzing the fusionned image or by exploiting the fusion results at the beginning of the classification algorithm in order to increase performance. Most of these techniques are

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

based on a knowledge base whose performance is directly dependent of its quality. Whatever the fusion method used, except the principal components analysis, the effort is made to optimize the classification algorithm rather than analyze the spectral bands choice or the relative importance given for each band. Moreover, the performance obtained by the fusionned image is not directly evaluated on the image but is determinated only on the classification results.
We propose a new data fusion method at pixel level applied to estimate multispectral images ponderation coefficients for supervised classification of non-resolved objects and optimize target detection on background. The bayesian approach, described in § 2, is used to select the channel combination rule, that minimizes the global risk. Therefore, a channel combination rule is proposed by estimating the probability density functions of the target and background. Then thresholds are defined and applied to the initial and fusionned images. The originality of our work lies in the results assessment that is directly obtained on the image. Moreover, different criteria are defined (§ 2) to evaluate the classification performance. Our method is applied to multispectral images taken by the AVIRIS hyperspectral imaging spectrometer for road detection. This last part allows the comparison of results obtained according to the "a priori" knowledge.

2. METHOD DESCRIPTION

The pixel fusion method is based on a probabilistic approach that supposes the probability laws of each class known.

2.1. Pixel fusion

2.1.1. Principle

The target detection is obtained by minimizing the global risk. The conditional risk is the mean cost combined with the d_j decision when the multispectral observation is $\vec{x}$:

$$R(d_j / \vec{x}) = \sum_{j=b,t} c(d_j / j) P(j / \vec{x}) \qquad (1)$$

Let $D: \vec{x} \rightarrow D(\vec{x}) = d_t$ or d_b be the decision rule that points out the decision to make when the spectral signature is $\vec{x}$. Thus, the global risk R designates the mean cost combined with a decision rule : $R = \int R(D(\vec{x}) / \vec{x})\, p(\vec{x}) d\vec{x}$

The bayesian detection requires that the applied decision rule minimizes the global risk. Therefore, a pixel $\vec{x}$ belongs to a target if and only if the conditional risk is minimum, then :

$$R(d_t / \vec{x}) < R(d_b / \vec{x})$$

Using **(1)**, we obtain :

$$(c_{tb} - c_{bb}) P(b / \vec{x}) < (c_{bt} - c_{tt}) P(t / \vec{x})$$

We require that the error cost is higher than the cost in the absence of error: $(c_{tb} - c_{bb}) > 0$ and $(c_{bt} - c_{tt}) > 0$.

The Bayes' rule allows us to express the "a posteriori" probabilities as a function of the "a priori" probabilities and the conditional densities:

$$P(j / \vec{x}) = \frac{p(\vec{x} / j).P(j)}{p(\vec{x})}$$

Consequently, the decision rule becomes :

$$\textbf{Pixel belongs to target} \Leftrightarrow \frac{p(\vec{x} / t)}{p(\vec{x} / b)} > \frac{c_{tb} - c_{bb}}{c_{bt} - c_{tt}} \frac{P(b)}{P(t)}$$

Note: If the cost functions are unknown, the expression $0 \le c_{jj} \le 1$ *is used to determine the associated values :*

- $c_{tt} = c_{bb} = 0$ *: the cost to make the decision* d_j *when the nature statement j is minimum. So the cost to make a right decision is non-existent.*
- $c_{tb} = c_{bt} = 1$*: the cost to make the wrong decision to allocate the pixel to a class is maximum.*

The decision rule that makes the global risk be minimum value is then : $\frac{p(\vec{x} / t)}{p(\vec{x} / b)} > \frac{P(b)}{P(t)}$

DUDA [11] shows that this rule makes the mean error probability be the minimum value :

$$P(\text{error}) = \int_{-\infty}^{+\infty} P(\text{error}, \vec{x}) d\vec{x}$$

Then, we try to minimize the mean error probability.

2.1.2. Bayesian method

When the probability densities $p(\vec{x}, j)$ of target and background follow a gaussian law $N(\vec{m}, \Sigma)$ and the "a priori" probabilities are determinated, a function g of the $\vec{x}$ vector, that separates target from background, can be calculated. Let $p(\vec{x} / j) \equiv N(\vec{m}_j, \Sigma_j)$, the decision rule becomes :

$$d_{target} \Leftrightarrow p(\vec{x} / t).P(t) > p(\vec{x} / b).P(b)$$

Using the normal density probability expression, we obtain :

$$d_{target} \Leftrightarrow -\tfrac{1}{2}[\vec{x} - \vec{m}_t]^t \Sigma_t^{-1} [\vec{x} - \vec{m}_t] + \text{Log}(P(t)) - \tfrac{1}{2}\text{Log}(\text{dét}(\Sigma_t)) > -\tfrac{1}{2}[\vec{x} - \vec{m}_b]^t \Sigma_b^{-1} [\vec{x} - \vec{m}_b] + \text{Log}(P(b)) - \tfrac{1}{2}\text{Log}(\text{dét}(\Sigma_b))$$

that we rewrite as :

$$d_{target} \Leftrightarrow g(\vec{x}) > a'_t - a'_b \qquad (2)$$

with :

$$g(\vec{x}) = \vec{x}^t (A_t - A_b)\vec{x} + (\vec{a}_t - \vec{a}_b)^t \vec{x}$$

$$A_j = -\frac{1}{2}\Sigma_j^{-1}$$

$$\vec{a}_j = \Sigma_j^{-1} \vec{m}_j$$

$$a'_j = \frac{1}{2}\vec{m}_j^t \Sigma_j^{-1} \vec{m}_j + \frac{1}{2}\text{Log}(\text{dét}(\Sigma_j)) - \text{Log}(P(j))$$

According to equation **(2)**, $g(\vec{x})$ designates a quadratic function of $\vec{x}$. In other words, the spectral signatures combination of every pixel must be quadratic so that the error mean probability corresponds to the minimum value.

Note : When the normal density probability hypothesis is taken into account and the target and background covariance matrix are supposed equal $(\Sigma_t = \Sigma_b = \Sigma)$, we obtain $A_t = A_b$.
The decision rule that minimizes the error mean probability is given by :

$$\vec{a}^t \vec{x} > a'_t - a'_b \text{ with } \vec{a} = \Sigma^{-1}(\vec{m}_t - \vec{m}_b)$$

Thus, the fusionned image is now obtained according to the linear combination rule $g(\vec{x}) = \vec{a}^t \vec{x}$.

2.2. Pixel classification for bayesian method

The image classification in two objects (target, background) is obtained by applying a threshold on this image. The classification is carried out at 2 levels. The first level

corresponds to the initial image thresholding and the second level is the fusionned image thresholding. These two classification levels results allow us to assess the fusion performance in relation to the initial images classification. Threshold is chosen according to the classification level.

2.2.1. Initial images

We consider that images are independent in the sense that a threshold is computed for every initial image. Thus for every initial image, the databases are modelled by a one dimension gaussian density probability.
The band i hypothesis is the following :

$$p_i(x^i / j) \equiv N(m_j^i, \sigma_j^i) .$$

The decision rule can be written as :

$$d_{target} \Leftrightarrow p_i(x^i / t).P(t) \rangle p_i(x^i / b).P(b) .$$

To simplify our notations, the sign i identifying to the spectral band will be deleted. The above calculations have to be made for every spectral band.
The density formula allows us to simplify the decision rule expression :

$$d_{target} \Leftrightarrow -\frac{1}{2\sigma_t^2}\left[x - m_t\right]^2 + \mathrm{Log}(P(t)) - \mathrm{Log}(\sigma_t) \rangle$$

$$-\frac{1}{2\sigma_b^2}\left[x - m_b\right]^2 + \mathrm{Log}(P(b)) - \mathrm{Log}(\sigma_b)$$

Equality is obtained when the x value of the pixel P is equal to the classification *s* threshold which allows to separe the target from the background databases.
Equality leads to this quadratic equation :

$$\frac{1}{2}x^2\left(\frac{1}{\sigma_b^2} - \frac{1}{\sigma_t^2}\right) + x\left(\frac{m_t}{\sigma_t^2} - \frac{m_b}{\sigma_b^2}\right) - \frac{m_t^2}{2\cdot\sigma_t^2} + \frac{m_b^2}{2\cdot\sigma_b^2} - \mathrm{Log}\left(\frac{\sigma_t \cdot P(b)}{\sigma_b \cdot P(t)}\right) = 0 \qquad (3)$$

The equation **(3)** can be reduced to a a first degree equation if the coefficient of x^2 is non-existent :

$$a' = \frac{1}{2}\left(\frac{1}{\sigma_b^2} - \frac{1}{\sigma_t^2}\right) = 0 \Leftrightarrow \sigma_t = \sigma_b = \sigma$$

So, for equal target and background databases standard deviations, there are various cases :

- If $m_t = m_b$ and $P(t) = P(b)$, the two density probabilities are superimposed along with equal "a priori" probabilities, then there is an infinite number of the equation **(3)** solutions nevertheless none could separate the target from the background.

If $P(t) \neq P(b)$, the **(3)** equation doesn't have any solution and another threshold type, called barycentric threshold, is introduced :

$$s = \frac{\sigma_b m_t P(b) + \sigma_t m_b P(t)}{\sigma_t P(t) + \sigma_b P(b)}$$

- If $m_t \neq m_b$, the equation **(3)** solution leads to one threshold value, called gaussian threshold :

$$s = \frac{1}{2}(m_t + m_b) + \frac{\sigma^2}{m_t - m_b}\mathrm{Log}\left(\frac{P(b)}{P(t)}\right)$$

When $\sigma_t \neq \sigma_t$, the **(3)** equation roots number depends on the discriminant value :

$$\Delta = \frac{1}{\sigma_{tc}^2\sigma_b^2}(m_t - m_b)^2 + 2\left(\frac{1}{\sigma_b^2} - \frac{1}{\sigma_t^2}\right)\log\left(\frac{\sigma_t P(b)}{\sigma_b P(t)}\right).$$

We analyse the strictly necessary conditions to have :

- $\Delta = 0 \Leftrightarrow m_b - m_t = \pm\sqrt{2\left(\sigma_b^2 - \sigma_t^2\right)\cdot\mathrm{Log}\left(\frac{\sigma_t P(b)}{\sigma_b P(t)}\right)}$

The discriminant cancels out if and only if one of the two following conditions is achieved :

$\sigma_b \rangle \sigma_t$ and $\frac{\sigma_t}{P(t)} \rangle \frac{\sigma_b}{P(b)}$ **(a)** or $\sigma_t \rangle \sigma_b$ and $\frac{\sigma_b}{P(b)} \rangle \frac{\sigma_t}{P(t)}$ **(b)**

So, the threshold value obtained with the equation **(3)** is :

$$s = \frac{m_t\sigma_b^2 - m_b\sigma_t^2}{\sigma_b^2 - \sigma_t^2}$$

This threshold is called gaussian threshold.

- $\Delta \rangle 0 \Leftrightarrow (m_b - m_t)^2 \rangle 2\left(\sigma_b^2 - \sigma_t^2\right)\cdot\mathrm{Log}\left(\frac{\sigma_t P(b)}{\sigma_b P(t)}\right) \Leftrightarrow$

$(\sigma_b \rangle \sigma_t$ and $\frac{\sigma_t}{P(t)} \rangle \frac{\sigma_b}{P(b)})$ (a)

or

$(\sigma_t \rangle \sigma_b$ and $\frac{\sigma_b}{P(b)} \rangle \frac{\sigma_t}{P(t)})$ (b)

In that instance, the solution **(3)** leads to two threshold values *smin* and *smax* called gaussian thresholds :

$$s\max = \frac{-b + \sqrt{\Delta}}{2a'} \quad \text{and} \quad s\min = \frac{-b - \sqrt{\Delta}}{2a'} \quad \text{with}$$

$$b = \frac{m_t}{\sigma_t^2} - \frac{m_b}{\sigma_b^2} .$$

Therefore, the decision rule is : $d_{target} \Leftrightarrow s\min \langle x \langle s\max$

Two further conditions which depend on **(a)** or **(b)** conditions are obtained :

(c) $\sigma_b \rangle \sigma_t$ and $\sigma_t^2 m_b \langle \sigma_b^2 m_t$ if condition **(a)** is respected.

(d) $\sigma_t \rangle \sigma_b$ and $\sigma_t^2 m_b \rangle \sigma_b^2 m_t$ if condition **(b)** is respected.

- $\Delta \langle 0 \Leftrightarrow (m_b - m_t)^2 \langle 2\left(\sigma_b^2 - \sigma_t^2\right)\cdot\mathrm{Log}\left(\frac{\sigma_t P(b)}{\sigma_b P(t)}\right)$

The equation **(3)** doesn't supplied threshold value. Then the barycentric threshold is chosen in order to classify image.

2.2.2. Fusionned image

For normal density probability hypothesis, a threshold can directly be determinated by the bayesian inference (cf. § 2.1.2) and the density probability formula.
In fact, the decision rule leads to :

$$d_{target} \Leftrightarrow g(\vec{x}) \rangle a'_t - a'_b$$

with $a'_j = \frac{1}{2}\vec{m}_j\Sigma_j^{-1}\vec{m}_j + \frac{1}{2}\mathrm{Log}\left(\text{dé t}(\Sigma_j)\right) - \mathrm{Log}(P(j))$

The threshold, called bayesian threshold, is due to the decision rule : $s = a'_t - a'_b$.

2.3. Evaluation criteria

The images fusion use is justified if the result image performance is better than in the best initial images. This contribution is evaluated with several criteria depending on the selected processings. These criteria are all defined from the two databases.
The first and second order statistical characteristics analysis of the initial and fusionned images leads to make up two performance evaluation criteria.
The normalized deviation (in %) between the target and background mean levels, noted Δm , is :

$$\Delta m = \frac{|m_t - m_b|}{(m_t + m_b)/2} \cdot 100$$

with m_j mean equals to the mean $\tilde{m}_j$ in the fusionned image or to the mean m_j^i in the initial image i.

This criterion is comparable with a contrast measurement between target and background for a given initial band or a given fusionned image.

The class overlap in the fusionned or initial image, written Δr, is defined with the maximum and minimum levels of the target and background databases.

For instance, if $\min(x_b) < \min(x_t) < \mathrm{Max}(x_b) < \mathrm{Max}(x_t)$ then the class overlap percentage is written :

$$\Delta r = \frac{\mathrm{Max}(x_b) - \min(x_t)}{\mathrm{Max}(x_t) - \min(x_b)} \cdot 100$$

This criterion gives information about the position of the target database level variation area compared to the background one.

Three further criteria are drawn up from classified initial and fusionned images in addition to a classified images visual analysis. These criteria are defined from target and background databases analysis.

Let us consider $m_t \langle m_b$. The area R_t corresponds to the part situated before the classification threshold s and the area R_b is the one situated after the threshold.

The good classification rate, BC, is obtained by (in %) :

$$BC = \frac{n_t + n_b}{N_t + N_b} \cdot 100$$

The non-detection error, noted E_{nd}, corresponds to the case where the pixel is in the R_b area but belongs to the target class :

$$E_{nd} = \left(1 - \frac{n_t}{N_t}\right) \cdot 100$$

So, higher is this rate, more important is the number of pixels classed as "background" while there are coming from target object.

The false alarm error, noted E_{fa}, is obtained by computing the pixels number in the R_t area but belonging to the background :

$$E_{fa} = \left(1 - \frac{n_b}{N_b}\right) \cdot 100 .$$

These criteria will be used afterwards to evaluate the fusion and classification pixel performance.

3. DATA DESCRIPTION

3.1. Images

Data are provided by the AVIRIS airborne spectro-imager, covering the visible and the near infra-red spectral domain. AVIRIS is an instrument characterized by a 17 m by 17 m ground spatial resolution and an image area of around 10 km by up to 21.5 km. This instrument simultaneously acquires images in 224 narrow spectral bands which are adjacent 10 nm wide (+/- 2 nm) and covers the whole solar spectral domain from 0.4 µm to 2.5 µm (Table 1) [12,13].

Spectral characteristics	Values (nm)
Spectral domain	400 à 2450
Sample interval	10
Channels width	10
Calibration accuracy	≤ 2

Table 1 : AVIRIS spectral characteristics.

All the spectral manipulated pixel values are expressed in spectral radiance ($W.m^{-2}.sr^{-1}.\mu m^{-1}$).

The images (512 lines by 614 samples) used in this study were acquired above the La Crau plain on the 18th of June 1997 at 11h57 a.m. This site is a plain which is situated at 50 km in the north-west of Marseille and diagonally crossed by a highway divided with a grassy central reservation. On both sides on this road, there are some orchards in the North and marshes in the South. The rest of the plain is composed of stones and thick bushes. Some scattered clouds (along the image right border) as well as their shade appear (Figure2).

The images used to our processing simulate the SPOT instrument's three spectral bands (Band XS_1: 500 - 590 nm; XS_2: 610 - 680 nm; XS_3: 790 - 890 nm) and are reconstructed with AVIRIS hyperspectral acquisitions.

For the selected spectral band XS_i, the equivalent radiance [14] at a P point is :

$$L_p(XS_i) = \frac{\int_{XS_i} L_p(\lambda) S_{XSi}(\lambda) d\lambda}{\int_{XS_i} S_{XSi}(\lambda) d\lambda} \quad \text{in } W.m^{-2}.sr^{-1}.\mu m^{-1}$$

Fusion and pixel classification will be carried out on area that are extracted from these different images XS (Figure 2).

3.2. Databases

A minimum "a priori" knowledge of the target and background databases is required in order to have a reliable and effective pixel fusion. Our method is applied to realize supervised road detection. Then, the two databases are built by selecting target and background pixels in the SPOT images.

In order to construct the target database, pixels, the most representative of the radiances variations area, are taken from different parts of the highway. The background database (Figure 1) is taken from an area composed of various vegetation types (growed fields, orchards) and bare soils (bushes, stones).



Figure 1 : Background database extracted from XS2 image.

The characteristics of the databases are given in Table 2.

We compare the ratio σ/m between databases for a given band and notice that the background database dispersion is higher than the one from target for XS1 and XS2 bands (increasing approximately of 76 % from target to background). The σ/m ratio is high for the background database because the radiances are scattered owing to the variety of the surfaces which make up this database. The road radiances are fairly uniform because road is tared.

The Δm deviation between the target and background mean radiances increases with the wavelength (increasing of 60 % from XS1 to XS2 and of 70 % from XS2 to XS3). The XS1 and XS2 bands are characterized by a high Δr overlap. On the other hand, this overlap decreases of 50 % from XS2 band to XS3 but remains quite considerable.

		XS1	XS2	XS3
σ/m	Target	9	12	10
	Background	16	21	10
Δm		12	19	32
Δr		75	67	38

Table 2 : Target and background database statistics determinated on initial images (%).

4. PIXEL CLASSIFICATION RESULTS

We consider two cases : the "a priori" probabilities of target and background are well-known or not.

4.1. "A priori" probabilities well-known

The target and background "a priori" probabilities are directly computed from the initial images: $P_t = 0.01$ and $P_b = 0.99$.

4.1.1. Initial images pixel classification

Theoretical and visual thresholds that are applied on initial images lead to a image classification in two objects : target and background. The visual threshold is the optimal one in the sense of quality image criterion.
The best results are obtained in band XS3 and are display in Figure 2.

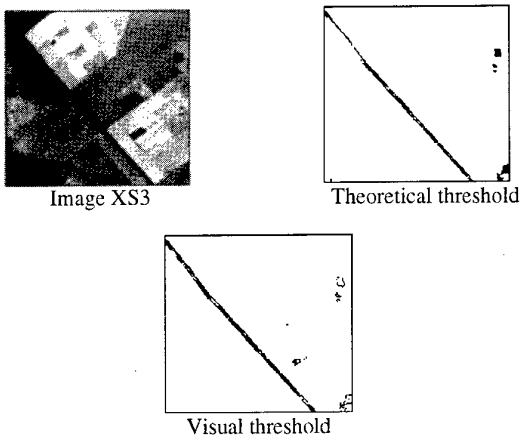


Figure 2 : Image XS3 classified by bayesian method with the known "a priori" probabilities.

The class recovery rates and the mean radiances deviation obtained in the 3.2 paragraph are in harmony with this remark. In fact, the best classification is realized when threshold is applied on the XS3 band because for this band the rate of radiances which are identical to target and background is lower and the target and background mean radiances deviation is higher. Nevertheless, scattered cloud shades (on right border image) and pixels stemming from the swampy area (situated at the left bottom of image) remain classified in target in the XS3 band.
The search for visual classification threshold leads to a different value from the one provided by theory (§ 2.2). For normal density probability hypothesis, only the XS3 image can be classified by gaussian thresholds (there isn't solution for the others bands cf. § 2.2.1.1.). For the XS3 band classified by visual threshold, only some pixels delimiting clouds shades belong to the target class while for theoretical threshold the whole shades were detected as target object. However, pixels belonging to the road beginning are obscured by trees which are located on the road border or their shades while the two road ways were detected on the XS3 image classified by visual threshold; what does not perfectly point out the real landscape.
The XS1 and XS2 bands are highly correlated, this can be shown in Table 3.
Afterwards the evaluation criteria results that are obtained by classification after fusion will be compared with results determinated by initial XS3 image and gaussian thresholds.

Criterion	E_{nd}	E_{fa}	BC
XS1 Image	43.0	17.6	81.7
XS2 Image	41.6	17.0	82.5
XS3 Image	43.7	0	98.9

Table 3 : Pixel classification performance evaluation criteria on initial images with the well-known "a priori" probabilities (%).

4.1.2. Pixel classification after fusion by bayesian method

The Δm deviation strongly increases and the Δr recovery rate is equal to zero after fusion (Table 4).

Criterion	Δm	Δr	E_{nd}	E_{fa}	BC
Well-known "a priori" probabilities	>100	1	9.9	1	99.1

Table 4 : Performance evaluation criteria of after fusion classification with the well-known "a priori" probabilities (%).

The fusion leads to a target from background radiances partition.
The fusionned image classification is obtained by the bayesian and visual thresholds (Figure 3).

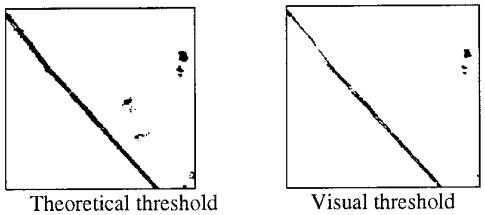


Figure 3 : Bayesian method and the known "a priori" probabilities.

In the case of classification by theoretical threshold, the classified image visual analysis shows that field pixels corresponding to the darkest levels, clouds, pixels close to road are detected as target after fusion.
In the case of classification by visual threshold, the road is perfectly detected. While some cloud shades belong to the

target class, the number of pixels belonging to clouds and classified as target (on the right border at the top) has strongly decreased and even some shades have disappeared. Moreover, the central platform that divides the highway clearly appears. The visual threshold improves the classification in relation to the theoretical one. In that case a low threshold value deviation creates considerable change on the classified image.
The visual analysis of the image fusionned and the reference image classified by visual threshold (Figure 2) shows that fusion improves classification: a part of the cloud shade is no more detected as target but belongs to background, the other part classified as target has decreased. This classification improvement provided by fusion is low because the XS3 band is high-performance and the associated others bands are mediocre.
At the database level, the non-detection error computed with bayesian threshold strongly decreases (77 %) according to the reference image result. On the other hand, the false alarm error and the good classification rate are approximately equal before and after fusion.
Threshold value is dependent on the database choice. The bayesian and visual thresholds lead to different results: the bayesian threshold allows rough road detection while the visual threshold realize a more efficient detection.
This kind of fusion can lead to small size target detection. On the fusionned image classified by visual threshold, the grassy central reservation is clearly detected in the middle of the road. Every way including central reservation width is equal to the spatial resolution instrument. The small target detection is thus conceivable for fusion with normal probability hypothesis and visual threshold.

4.1.3. Note

A high non-detection error E_{nd} rate is obtained on the initial and fusionned classified images.
The E_{nd} deviation is very sensitive owing to the low number of pixels forming the target database. Moreover this high E_{nd} value is due to the target database construction. In fact, the target pixels are of two types: those consisting only of road objet called "pure pixels" and those resulting from the combination of road object with either the grassy central reservation or the road borders stemming from the background class called "mixed pixels". At the time of the classification, the "mixed pixels" separated by the central reservation are allocated to the background class while they were chosen at the beginning for target database construction.

4.2. "A priori" probabilities unknown

When the target and background "a priori" probabilities are not known, they are chosen equal to 0.5 (equiprobability hypothesis).

4.2.1. Initial images pixel classification

The visual analysis proves that the best results are obtained for the XS3 image as in the case where the "a priori" probabilities are well-known. On the XS3 image classified by gaussian thresholds (Figure 4), some background areas characterized by the darkest levels, the cloud shades and some marsh pixels belong to the target class.

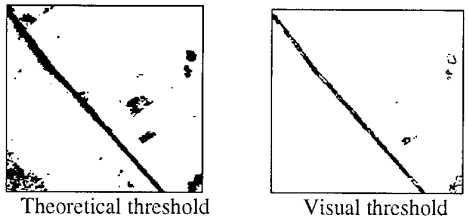


Figure 4 : Image XS3 classified by bayesian method with the unknown "a priori" probabilities.

The visual thresholds are independent of the target and background "a priori" probabilities values. The XS3 image visual classification leads to better result than these obtained by gaussian thresholds. In fact, the road separation is perfectly defined and some pixels of the cloud shades border belong to the target class.
By comparing results obtained for every bands (Table 5), we notice that the non-detection error increases of 75 % from XS1 to XS2 while the false alarm error decreases of 20 % due to some vegetation pixel and clouds shades detected as road class. The XS1 and XS2 good classification rates are approximately equal. The XS3 non-detection error is the same as the one obtained for XS1. The false alarm error strongly decreases from XS2 to XS3 (approximately 70 %) and the good classification rate increases of 8 %. These results prove also that the best classification is obtained by XS3 band.

Method bayesian classification			E_{nd}	E_{fa}	BC
Initial images	XS1		4.2	13.0	87.2
	XS2		7.0	10.2	89.9
	XS3	Theoretical threshold	4.9	3.3	96.7
		Visual threshold	26	0	99.3

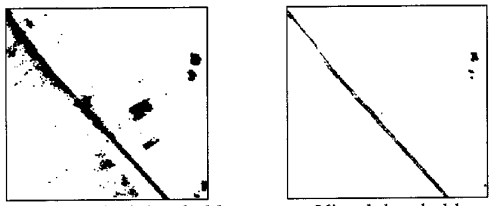
Table 5 : Performance evaluation criteria of the initial images with the unknown "a priori" probabilities.

The visual comparison of the XS3 image classified using equiprobability hypothesis (Figure 4) with the XS3 image classified by gaussian thresholds and well-known "a priori" probabilities (Figure 3) shows that the classified initial images are very dependent on the "a priori" probabilities choice. When the probabilities are known, the classification is better in relation to the case where they are unknown and arbitrarily fixed. In fact, the knowledge accuracy increasing allows to decrease the number of pixels located near the road, marsh pixels and pixels from darkest areas that are misclassified.
At the databases level, the non-detection error decreases while the false alarm error increases for the XS3 image and arbitrarily fixed "a priori" probabilities. This non-detection error decreasing is linked to the 4.1.3 note. In fact, for equiprobability hypothesis, the classification by gaussian thresholds leads to a rough road detection. However, the road detection is realised at the pixel level when the visual threshold is used or when the "a priori" probabilities are known.

Afterwards, the XS3 classified image will be considered as reference image in order to analyse the bayesian fusion performance.

4.2.2. Pixel classification after fusion

The fusionned image classification is obtained by the bayesian threshold and the visual one associated.
The overall visual analysis of the classified image by bayesian threshold (Figure 5) shows that pixels located near the road, some field pixels and clouds are allocated to the road class. However pixels stemming from swamp that belonged to target on the XS3 classified image, are classified as background on the fusionned image.
In the case of classification by visual threshold, the road is perfectly detected and the central reservation dividing the road clearly appears. However, some cloud shades are allocated to the target class. In that case, the visual threshold improves results in relation to bayesian threshold.



Theoretical threshold Visual threshold

Figure 5 : Bayesian classification and the unknown "a priori" probabilities.

The visual analysis of the fusionned image and the XS3 reference image that are classified by visual threshold (Figure 4) shows that fusion improves classification: none dark background area is classified as target, and the cloud shades (located on the right border image at the bottom) disappear. Moreover, pixels located on the road beginning are obscured by trees or their shades while the two roads ways appear on the XS3 classified image before fusion. So, the classified fusionned image perfectly represents the real landscape.
At the databases level (Table 6), the non-detection error obtained by theoretical threshold and by XS3 reference image is the same. However, the false alarm error increases as the good classification rate decreases after fusion. Fusion does not improve classification by bayesian threshold results.
In the visual threshold case, fusion increases non-detection error of 25 % in relation to the XS3 image value. This increasing proves that road detection is achieved at the pixel level (cf. § 4.1.3) by visual threshold. The false alarm error and good classification rate don't change.

Criterion	Δm	Δr	E_{nd}	E_{fa}	BC
Theoretical threshold	>100	1	4.9	4.4	95.6
Visual threshold	>100	1	43.7	0	99

Table 6 : Performance evaluation criteria of after fusion classification with the unknown "a priori" probabilities (%).

The visual classification after fusion is better than the XS3 band one.
The visual analysis of the images having been classified by visual threshold (Figure 3 and Figure 5) shows that results are similar when the "a priori" probabilities are known or not known. In fact, the target and background "a priori" probabilities don't take place in the bayesian fusion channels combination rule (cf. § 2.1.2). Therefore, the "a priori" probabilities choice has no effect on the classification results obtained by visual threshold.
The "a priori" probabilities are taken into account in order to determine the theoretical threshold. By comparing Figure 3 with Figure 5, we notice that a better classification is obtained by bayesian threshold when the "a priori" probabilities are known. In fact, for the equiprobability hypothesis, lots of pixels located near the road and some pixels belonging to dark vegetation area are classified as target.
As a conclusion, the classification results are closely dependent on the threshold choice and the "a priori" probabilities knowledge.

5. CONCLUSIONS AND PERSPECTIVES

The optimal low level fusion leads to a channel combination rule that minimizes the global risk and consequently the mean probability error. The bayesian approach is used to realize this optimization. It allows to assess hypothesis from independent sensor measurements when the target and background conditional distribution and the "a priori" probabilities are known.
The pixel fusion aims at classifying the image pixels as different objects (in that instance two types). The classification is obtained by applying on the fusionned or initial image theoretical threshold and visual threshold. The visual threshold is the optimal one in the sense of quality image criterion. The theoretical threshold depends on the classification level and is the result of databases measurements. The classification obtained by threshold depends on the database choice. This classification method is fast and efficient at the repartition of pixels to classes level.
The result performance evaluation is complete. Several criteria take into account databases: statistical errors, and two criteria providing information on the distributions position independently of threshold. A further qualitative analysis in image term completes these criteria.
The "a priori" probabilities knowledge effect on these classification methods is analysed by considering the case where they are unknown and arbitrarily fixed.
When the "a priori" probabilities are known, results show that the initial image classification applied to road detection depends on wavelength and threshold choice. In fact, the best classified image before fusion is obtained by the XS3 band with gaussian thresholds which means normal density probability hypothesis. The fusionned image classification depends on the threshold choice.In fact, a low threshold deviation creates important changes on the classified image before and after fusion. Moreover, visual and theoretical thresholds are different.
The pixel fusion that takes into account the gaussian databases probability laws and the visual threshold allows to improve classification in relation to XS3 image results. The fusion efficiency for pixel classification is here weakly highlighted because the XS3 band that provides a good classification is combined with the other bands that provide a bad classification in order to obtain the fusionned image.
The effect of target and background "a priori" probabilities values is analysed by considering the case where these probabilities are unknown. When the classification is obtained by visual threshold, classification before and after fusion are independent on the "a priori" probabilities choice. However classification by theoretical threshold is better when the "a priori" probabilities are known.
The results lead to introduce further spectral bands in order to improve the fusion performance.

These classification methods will be used to detect n objects ($n > 2$) on an image thanks to their simplicity, quickness and efficiency. Consequently, the probability densities of the main kind of surfaces withdraw from SPOT images are at the present time analysed and modelised.
From the fusionned image by bayesian method, that allows the more efficient classification according to our results, classification can be realized by two kind of thresholds: the theoretical threshold following from the analytical resolution and the visual threshold that provides the best visual result. It is difficult to evaluate the visual threshold because this threshold is obtained from a subjective criterion but it is independent of the target and background "a priori" probabilities values. On the other hand, the theoretical threshold depends on the "a priori" probabilities knowledge. But it is very difficult to obtain an "a priori" probability set that perfectly represents reality. A more accurate analysis of the "a priori" probabilities and visual threshold choices will be later developed by introducing the uncertainty theory. In fact, this theory allows to exploit the physical knowledge and take into account the more accurate as possible information.
The probability theory associated to the bayesian decision theory is a fusion method widely used also as comparison base to the other models [15]. However, this theory doesn't take into account the more or less partial physical knowledge. This is a further reason to deal the bayesian inference to the processings of uncertain values that are provided by the non exclusive hypothesis [16].

6. ACKNOWLEDGEMENTS

We sincerely thank Dr. R.O. GREEN from J.P.L. (U.S.A.) for having kindly provide us with AVIRIS images.
We sincerely thank E. CUBERRO from C.N.E.S. (France) for having contribute towards improve this document with her judicious remarks.
We thank the National Superior School of Aeronautics and Space Tchnologies (France).

7. REFERENCES

[1] Waltz E., Llinas J., "Multisensor Data Fusion", Artech House, Boston, M.A., 1990

[2] Gonzales R.C., Wintz P., "Digital Image Processing", Addison-Wesley Publishing Company, Advanced Book Program/World, Science Division, 1977

[3] Arigides A., Fernandez M., "Adaptative 4-D clutter suppression filtering technique", SPIE No 1481 on Signal and Data Processing of Small Targets, pp 110-116, 1991

[4] Stocker A.D., Reed I.S., Yu X., "Multi-dimensionnal signal processing for electro-optical target detection", SPIE, Vol. 1305, pp 218-231, april 1990

[5] Stocker A.D., Yu X., "Adaptative detection of sub-pixel targets using multi-band frame sequences", SPIE. Vol. 1481: "Signal and Data processing of Small Targets", pp 156-169, 1991

[6] Rouse J.W., Haas R.H., Schell J.A., Deering D.W., Harlan J.C., "Monitoring the vernal advancement of natural vegetation", NASA/GSFC Final Rep., Greenbelt, MD, 371 pp., 1974

[7] Huete A.R., "A soil adjusted vegetation index (SAVI)", Remote Sensing of Environment, Vol. 25, pp 295-309, 1988

[8] Pinty B., Verstraete M.M., "GEMI: a non-linear index to monitor global vegetation from satellites", Vegetation, vol. 101, pp 15-20, 1992

[9] Clark G.A., Sengrupa S.K., Schaich P.C., Sherwood R.J., "Data fusion for the detection of buried land mines", Lawrence Livermore National Laboratory: "Substance Identification Technologies", october 1993

[10] DelGrande N.K., "Airborne Detection of buried minefields", Energy and Technology Review, pp 9-21, Dec. 1991

[11] Duda R.O., Hart P.E., "Pattern Classification and Scene Analysis", Wiley-Interscience Publication, John Wiley & Sons, 1973

[12] Green R.O., Conel J.E., Carrere V., Bruegge C.J., Margolis J.S., Rast M., Hoover G., "Determination of the in-flight spectral and radiometric characteristics of the Airbone Visible/Infrared Imaging Spectrometer (AVIRIS)", ESA, SP-319, 5th International Colloquium on "Physical Measurements and Signatures in Remote Sensing", Vol. 1, pp 19-28, may 1991

[13] Vane, G.; Green, R. O.; Chrien, T. G.; Enmark, H. T.; Hanse, E. G.; Porter, W. M., "The Airbone Visible/Infrared Imaging Spectrometer (AVIRIS)", Remote Sensing of Environment, vol. 44, no. 2-3, May-June 1993, p. 127-143

[14] Begni G., Dinguirard M., Jackson R., Slater P., "Absolute calibration of the SPOT 1 HRV cameras", SPIE 660, Earth Remote Sensing Using the Landsat Thematic Mapper and SPOT Sensor System, 1986, pp 66-76

[15] Bloch I., Maitre H., "Fusion de données en traitement d'images: modèles d'information et décisions", Traitement du Signal 1994, Vol. 11, n°6, pp 435-446

[16] Appriou A., "Probabilités et incertitude en fusion de données multi-senseurs", Revue Scientifique et Technique de la Défense, 11, 1991, pp 27-40

PAPER No. 30

DISCUSSOR'S NAME: C. Remus

COMMENT/QUESTION:

Comment se passe l'entrée des données issues des senseurs dans la base lorsque ces données ne sont pas indepéndentes entre elles?

(How is sensor data input to the database when this data is not differentiated?)

AUTHOR/PRESENTER'S REPLY:

Le modéle proposé permet de rendre compte aussi bien d'indices indépendants les uns et autres, que d'indices partiellement ou totalement dépendants.

(The model proposed allows for input of both independent indices and partially or totally dependent indices.)

Sensor synergetics: The design philosophy of the sensor fusion demonstrator and testbed (SFD) at TNO-FEL

A.J. van der Wal
TNO-FEL Physics and Electronics Laboratory
P.O. Box 96864
NL-2509 JG The Hague
The Netherlands
e-mail: vanderwal@fel.tno.nl

1. SUMMARY

Although the potential of sensor and data fusion as a way to improve situation awareness nowadays is widely regarded as a valuable tool by the military R&D community, there is still little quantitative evidence available that supports the acclaimed benefits of generic fusion methodologies and thus justifies their application. It is therefore legitimate to ask how much value sensor fusion adds to the observation process. In case studies this question is seldom addressed and it is therefore tempting to ask, whether the acclaimed advantages of sensorfusion are "hype" or real. In this paper we will try to identify in what situation the application of sensor fusion techniques may be beneficial and which sensor fusion techniques seem to be among the most promising. The starting point of our discussion is the observation that the added value of fusion of similar sensors must originate from a *nonlinear* combination of sensor data streams. This observation naturally gives rise to the application of novel nonlinear models, *e.g.* from the area of softcomputing, *viz.* fuzzy logic, neural networks and evolutionary programming. In addition a trend in research can be observed to apply fusion techniques almost exclusively to "high-level" fusion (*i.e.* levels 2 and 3 of the JDL data fusion process model), implicitly considering fusion on the sensor level to be much less of a challenge. Motivated by the fact that high quality primary data are of paramount importance to the success of *all* subsequent signal processing stages, we argue against this view in this paper. In contrast to the classical data fusion approach, at TNO-FEL it was decided to start up a *sensor* fusion approach, in which an integrated approach towards sensor optimization, sensor management, and *early* fusion is pursued. Only in this way one may hope to attain the goal of sensor fusion, *viz.* an improved situation assessment. Concentrating on the primary information sources, *i.e.* the sensors, and using their synergy at the earliest stages of signal processing, inevitably forces one to invest in state-of-the art data-acquisition and real-time (pre)processing facilities. This will be illustrated by discussing some of the design issues of the FEL sensor fusion demonstrator (SFD) testbed. The SFD project has been started up by the end of 1996 and initial concepts have meanwhile been defined. The hardware and software concepts of this testbed will be described as well as the functionality that we hope to implement.

2. INTRODUCTION

In recent years, numerous research papers have been published dealing with the application of multisensor data fusion, also referred to as distributed sensing high-level fusion, especially in the domain of military observations [1-6]. Although intuitively appealing, one may conclude that data fusion did not yet bring about the expected breakthrough. Several explanations for this can be given, such as the particularity of the application domain, the limited availability of general nonlinear methods for fusion, and finally the quality of the primary 'raw' sensor data. Another problem may be the unrealistic expectations of the virtues of the synergy of multiple sensors.

The absence of a general way to approach the subject and the many ad hoc experiments and simulations that have been published, have motivated the present work at TNO-FEL. In the following we will shortly review the history of fusion, define sensor fusion as a field of research in its own right, and in Sec. 3 motivate the need for a sensor fusion demonstrator (SFD) testbed at TNO-FEL, sketch the problem of how to model sensor fusion and suggest some directions for answering some of the pertinent questions in this field using concepts from soft computing. Especially the use of fuzzy measures looks

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

promising as a way to model the sensor fusion process quantitatively. Next in Sec. 4 we describe some of the implementation issues and the plans for the first test experiment.

2.1 Historical overview

Historically the idea of sensor fusion is not new: As early as the sixties multi-radar trackers have been in use by the military for air traffic control and air defense. Multisensor data fusion seeks to combine information generated by multiple sensors to achieve goals that would be very hard or impossible to achieve with single sensors. From the point of view of efficiency, scheduling, accuracy, and redundancy it seems intuitively obvious that several sensors are 'better' than a single sensor.

Nowadays data fusion is a well-accepted method for making superior inferences in the field of industrial automation (*e.g.* for controlling a power plant, an oil refinery, a cement kiln (for a review on industrial applications, see *e.g.* [7,8]), or even a nuclear reactor [9,10], and for carrying out real-time pattern recognition in industry using a variety of sensors. Especially since the advent of softcomputing methods, such as fuzzy logic, data fusion has become a widely accepted successful fusion technology in industry. We note however that the success of such methods is primarily due to their ability to model human behavior or expertise in supervisory control. Sensor fusion also endeavors to mimic cognitive processes in humans by absorbing the signals of the human observation system, our five senses, from the real world and integrate, or 'fuse', these signal streams to arrive at a coherent picture of our environment. As such, sensor fusion is concerned with lower abstraction levels, much higher information rates, and generally requiring faster response than the data fusion used in supervisory control systems. This forms also the key problem in applying soft computing methods to this field: in controlling complex industrial or organizational processes at relatively long timescales human operators have accumulated over the years ample experience. In contrast, there is only limited insight in the way a human being builds up an environmental picture, his awareness, from continuous multisensate observations. It can therefore be a useful approach when setting up a sensor fusion testbed to have a close look at how the human cognitive system works. Cognition is still far from understood in detail, but a few global characteristics are apparent: the human recognition system consists of a massively parallel processor that merges vague, qualitative inputs and a priori knowledge, acquired by learning from experience into a more or less consistent picture of the environment. It consists of a large number of hierarchically ordered decision processes running concurrently, simultaneously inferencing on the same set of input data at different levels of granularity. We will not discuss these points in detail here, since they are outside the scope of this article. With the properties mentioned above in mind we will attempt to emulate in the sensor fusion testbed some of the functionality that we observe in human perception.

Although sensor fusion is important to virtually all phenomenological sciences and engineering disciplines, most work until now has been done in the field of *defense* research. One reason for this can be understood as follows. In *analytical* approaches, *e.g.* in a physics experiment, the measured quantities or interactions are often so small that the experimental setup has to be designed in such a way as to make sure that the desired quantity or effect is optimally measurable. If the measured quantities are small, the experiment is repeated many times and ergodicity (see Sec. 3.4) and statistics are used to arrive at average values with low relative standard deviation. Especially in case one tries to prove or disprove the correctness of a theoretical model, this often is a good approach. A final point to note here is that - apart from intrinsic physical real-time aspects - such experiments very often can be repeated many times and real time constraints are not a bottleneck.

In engineering approaches the use of sensors is more *synthetic*, as illustrated *e.g.* in the field of factory automation. Here one deals with a well-defined problem such as the quality control of products on a manufacturing line, e.g. checking the soldering joints on a PCB with an automated vision system. This problem certainly has real-time aspects, but the optimization can be done offline and the observation circumstances, like in the physics experiment, can be optimized offline, *e.g.* by testing the best combination of sensors, the proper cameras and illumination, parallel operation with more than one quality control station if the speed of production requires so.

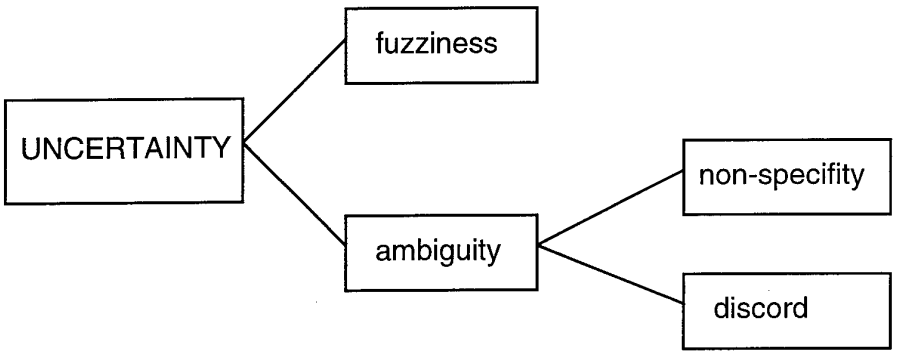


Figure 1 *A taxonomy of the different types of uncertainty.*

In military observations we deal with a situation that is far less comfortable than the situations described above: generally speaking it is necessary to assess in real time an often complex situation, that almost certainly is outside one's complete control. Handling such observations requires the modeling of *uncertainty*. Apart from the ordinary problems such as noise and clutter, radar and electro-optical sensors operate also under adverse weather and atmospheric conditions, without any possibility to improve the circumstances of the experiment, or to repeat the experiment, under strict real time constraints, with sometimes enormous consequences of false classification and even more serious penalties for non-detection. In addition, by the nature of the military métier, most interesting targets move at high speeds, try to avoid actively or passively or mislead sensors by jamming or using decoys and they are designed in such a way as to present a minimal scattering cross section to commonly used sensors and thus to be virtually invisible ('stealth').

Under these circumstances it is clear that doing military observations invariably implies the modeling of uncertainty. Classically this is often done by applying statistical methods, notably Bayes' theorem to formulate a (multi-) hypothesis testing problem [10]. It is however clear that statistical uncertainty can only model part of the uncertainty. The different types of uncertainty, whose measures are now well established in classical set theory, fuzzy set theory, probability theory, possibility theory and evidence theory [11] are schematically summarized in Fig.1 The breakdown distinguishes *fuzziness*, or vagueness due to a lack of definite and sharp conceptual distinctions and *ambiguity,* the situation where we are dealing with one-to-many relationships in the information obtained from sensors, yielding *non-specifity* in the case that the data leaves two or more alternatives unspecified, or even *discord, i.e.* disagreement in choosing from among several alternatives.

Recently methods that explicitly deal with ambiguity and partially overlapping hypotheses such as Dempster Shafer theory [12,13] and the application of belief functions instead of probability densities have become popular. Of even more recent date is the application of general fuzzy measures [14]. The difficulty inherent to making accurate observations in military applications and the lack of measurement statistics are the prime motivations to improve single sensor observations by merging (partial) inferences/conclusions from one sensor with inferences from the another one.

In addition recent history shows that the nature of military operations changes rapidly: Although sensors are vital to the success of any military mission, it becomes at the same time much more difficult to interpret these observations. This can be explained by the introduction of stealth technologies (radar), by which planes become much more difficult to detect by radar, the subtleties of 'peacekeeping' missions compared to classical, full scale warfare scenarios, and finally the complexities and greater vulnerability of navy vessels operating close to shore ('littoral warfare'). Finally it should be noted that there is a genuine need to fuse sensor generated information, at least at the higher levels of command and control: the

man-machine interface being the limiting factor. Although new sensors have been developed (e.g. GPS) and accuracy and resolution in space and time of most existing sensors have greatly increased in time, the bandwidth of the man-machine interface has not. The situation of having to deal with more information than one can process in a certain time is not unsimilar to the situation where a *lack* of information exists. Both situations involve taking decisions in the presence of uncertainty and would benefit from intelligent data reduction techniques, such as sensor fusion.

2.2 Definition of sensor fusion *vs.* data fusion

Following the definition of the functional model of the data fusion process that is widely accepted in the military research community, as *e.g.* presented by Hall [15], and using the terminology as agreed by Joint Directors of Laboratories (JDL) of the Data Fusion Subpanel. Multisensor data fusion is by the JDL defined as

"A continuous process of dealing with association, correlation and combination of data and information from multiple sources to achieve refined entity position and identity estimates, and complete and timely assessments of resulting situations and threats, and their significance."

We will in this paper define multisensor fusion as level 1 processing, with basic processes: data alignment, association and correlation, positional and identity fusion, complemented by the real-time part of level 4 processing ("maintenance"). The reason for including part of level 4 may perhaps seem strange at first sight, but is immediately apparent when maintenance is interpreted as the assessment of the status of each sensor in order to keep it optimally tuned. Apart from optimizing individual sensors monitoring sensor performance makes it possible to perform *'sensor management'*, *i.e.* to optimize groups of sensors, which is *e.g.* important for military observation systems where a large number of sensors co-operate in a coherent way and in which part of the sensors may be damaged during operation (*c.f.* a phased array radar system). The sensor management system contributes therefore directly to the *robustness* of a system.

After having clarified the scope of sensor fusion, we will next discuss the motivation to develop a sensor fusion testbed, some of the problems inherent to sensor fusion and the type of approach that has been selected to study the added value of sensor fusion. In Fig.2 a schematic representation is given of the different levels of fusion that can be distinguished in data fusion. For clarity the feedback paths originating from maintenance and on-line sensor optimization (fast level 4

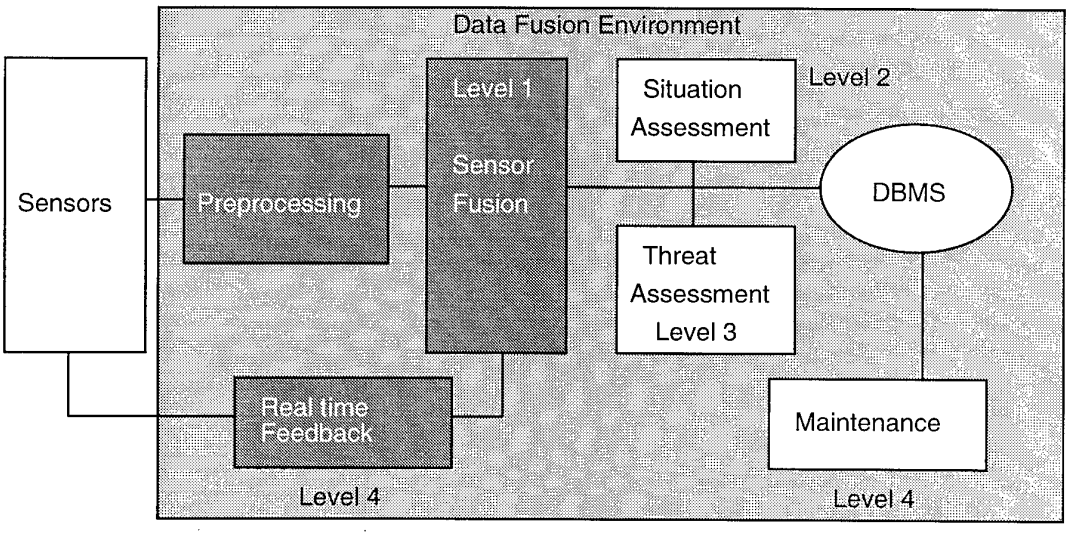


Figure 2 *The JDL process model for data fusion. The darker areas indicate the scope of the SFD.*

processing) have been included. Early sensor fusion can be viewed as a two-step process, a direct fusion step followed by a complimentary fusion step (Fig.3). The direct fusion process acts immediately on *raw* sensor data, after a possible preprocessing stage for alignment. This type of fusion is in practice limited to combining signals from similar sensors. In the next stage, in the complementary fusion process, very different types of sensors can be fused. In this stage it is possible that a considerable data reduction is achieved and that the information can be represented as a vector in feature space. Features such as range, position, orientation, effective cross section, shape, color, etc. are extracted from the different sensors and combined in qualitative or quantitative ways. Combining of information from complimentary sensors can thus be seen as augmenting the dimensionality of the feature space. After this fusion step all sensor information has been fused and one needs now to combine feature vectors with existing, a priori information about the environment, collected from previously measured data or intelligence. This more abstract fusion stage is typical for levels 2 and 3 of the JDL model and will not be considered in this article.

3 SENSOR FUSION

3.1 Motivation for designing the SFD testbed

The general motivation why multisensor fusion is pursued is generally answered in terms of "to improve the combined observations of different sensors and thus create a better situational awareness". Basically this boils down to a purely economic factor. From the operational point of view possible benefits include a greater user friendliness because of data reduction, a greater robustness of the resulting observation system, higher reliability by a higher plot rate and therefore a better performance of *e.g.* a tracking algorithm, and thus a better observation, also under adverse circumstances. Scientifically however one immediately faces difficulties with this formulation of the motivation, because it introduces subjectivity: How can one measure "situational awareness"? An even more complicated question to be answered is *what* exactly can be improved in conventional, single sensor observation and *how* different sensors can benefit from each others measurements. Answering these two questions is by no means trivial and they can in fact only be answered by considering a specific application. However before

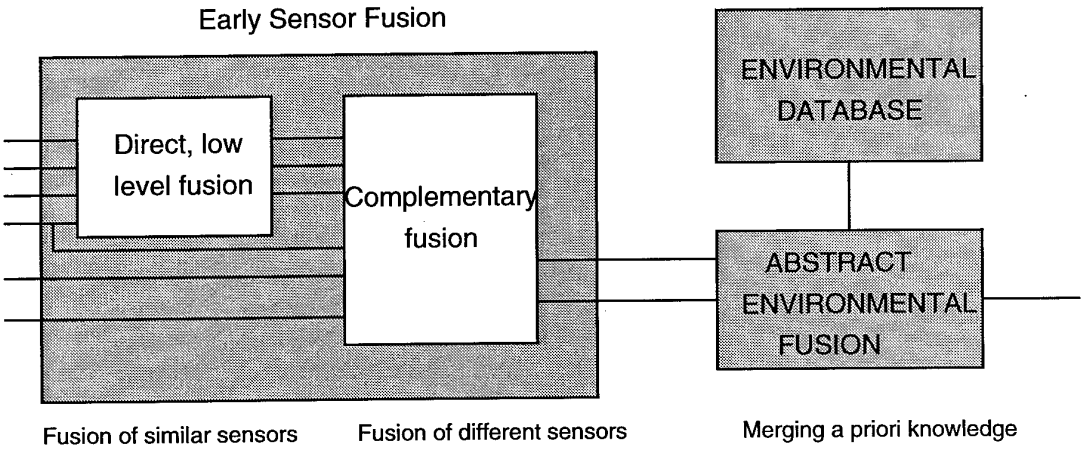


Figure 3 *Global information flow in sensor fusion; the level of abstraction in the fusion process increases to the right; to the left the data rate increases.*

doing so, it's worthwhile to take a moment and review the various abstraction levels, methods, hardware and software implementation methods of observation systems and the different ways that they can interact and cooperate. The basic difficulty is in the opinion of the author the absence of an unique theoretical framework for objectively combining the information streams generated by the various sensors. The difficulty lies essentially in adequately describing the information content in each sensor stream. The amount of *useful* information in a data stream created by a sensor is of course dependent on the ultimate goal of the complex of observations. Whenever this goal cannot be formulated in a clear, transparent and unambiguous way, it will be extremely difficult to develop synergy between the sensors and to compare the performance of the various fusion algorithms. We may therefore conclude that the order to study sensor fusion and designing a sensor fusion demonstrator (SFD) due attention must be paid to the following points:

- selection of suitable fusion experiments
- the goals to be improved by fusion
- selection of the quantities to be observed
- selection of number and type of sensors
- optimization of the individual sensors
- the handling of uncertainty
- architecture of the multi-sensor fusion setup
- selection of pre-processing algorithms
- clear distinction of objective and subjective modeling issues

In this context it is worth noting that optimization of individual sensors is a first step for understanding the benefits, and the limitations of sensor fusion, because if individual sensor observations have a marginal accuracy or do not pertain to the fusion goals, it is not realistic to expect improvements by combining these streams. If on the other hand the observations of the individual sensors are excellent, their is very little room for improvement. The ideal sensor fusion experiment shall therefore make use of well-defined, accurate and reproducible sensors that, when taking individual measurements are by themselves not effective to achieve the fusion goal. Note that this formulation also includes the situation that an individual sensor may ultimately attain the fusion goal by itself, but that the recognition or detection process can be substantially sped up by fusion.

3.2 Benefits and limitations of sensor fusion.

A number of potential benefits of fusion have been compiled by Waltz [16] and include

1. Robust operational performance
2. Extended spatial coverage
3. Extended temporal coverage
4. Increased confidence
5. reduced ambiguity
6. Improved detection
7. Enhanced spatial resolution
8. Improved system reliability
9. Increased dimensionality

We first note that most of the benefits quoted in literature are benefits that are exclusively associated with the presence of *multiple* sensors; they are *not* the benefits of sensor fusion. Most of these benefits are qualitatively and intuitively immediately clear. Globally we can distinguish three types of benefits:

- In the first place the effect of multiple sensors is an extended spatial, temporal, or spectral coverage of the associated phenomenon (benefits 2,3).
- A second type of benefit follows from statistical arguments: multiple sensors increase the measurement accuracy (7) and from this an increased confidence (4) may be derived, or at least a reduction in the number of hypotheses about the targets (5) and thus an improved detection, *c.q.* a shorter detection time (6). Only in the cases 5 and 6 a sensor *fusion* step is needed.
- Finally multiple sensors create overlap in observations and thus *redundancy*. If this redundancy is properly exploited in the system design, the maintenance (level 4) module will optimize the sensor scheduling and will result in the graceful degradation of system performance if sensors breakdown (1,8,9).

A quantitative aim of sensor fusion is to improve the accuracy of the observation, *e.g.* the position of a target. In Fig. 4 this situation is illustrated by combining a forward-looking IR sensor (FLIR) and a radar. The position determination of the radar and the FLIR observations is given together with the

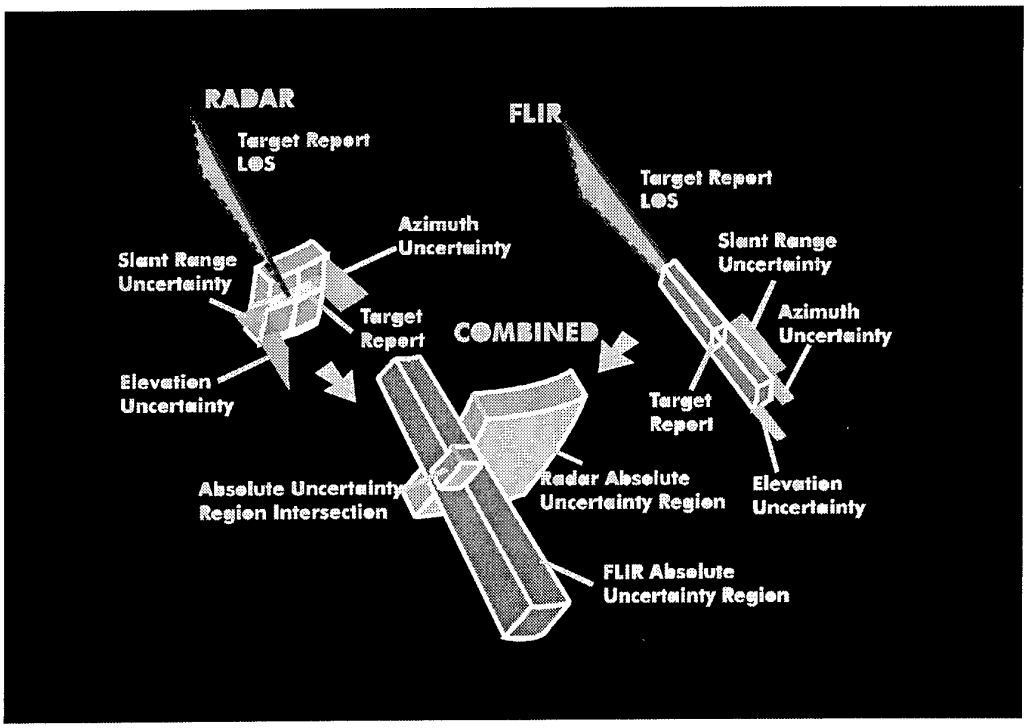


Figure 4 *Reduction of positional uncertainty by combining two complementary sensors: radar and FLIR.*

result of sensor fusion. The inaccuracy in azimuth and elevation of the radar sensor is compensated by the more accurate measurements of the FLIR sensor, while the pulsed radar accurately determines the range. This example illustrates how a radar can initially detect a target, because its wider field of view. Subsequently the FLIR can be cued using the
inaccurate coordinates of the radar to initiate the FLIR measurement. Together they determine a small region of interest around the target, so that the benefit of fusion is an improved estimate (or reduced uncertainty) of the position of the target.

An interesting attempt to illustrate in a quantitative way the virtues of sensor fusion is described in [17]. In the article an odd number N of identical sensors are fused with the aim to classify an observed phenomenon following a majority vote rule. The sensors are assumed to be statistically independent and the a priori probabilities are taken equal to 1/N, corresponding to the principle of maximum entropy, equivalent with a minimum of a priori knowledge. Although the example is a very idealized model case of identical, independent, unbiased sensors, all following the same statistics, and using binary classification and a majority vote scheme as fusion aggregator, a number of qualitative results are worth mentioning here:

- Fusing data from multiple sensors (each with an individual probability of correct detection *c.q.* classification of less than 0.5) results in a *decrease* in performance in going from a single sensor classification to the multiple sensor fused result.
- If the individual sensors are very accurate (probability of correct detection larger than 0.95) sensor fusion cannot significantly improve the results of the inference process.
- The relative improvement in performance of an N sensor fusion process over single sensor performance increases as a function of N leveling off at about N=10. Adding more, identical sensors does not pay off beyond this number. (see Fig.5)
- The maximum relative improvement of N sensor performance for $N \rightarrow \infty$ compared to a single sensor is of the order of 15-25 %,

depending on the fusion scheme. The maximum is assumed if the single-sensor probability of correct detection is in the range between 0.60 and 0.75.

Of course the numbers mentioned above should be treated with care because they depend on the type of aggregation operator chosen to represent the fusion process. Moreover these conditions refer to the simplified case of *identical* sensors, *i.e.* same positioning, calibration statistics, biases, sampling rates, bandwidths, sensitivities, dynamic behavior and the same measured quantities. If a new type of sensor is added to the sensor suite, the dimensionality of the observation is increased and a substantial increase in information content may be expected, depending on fusion objective.

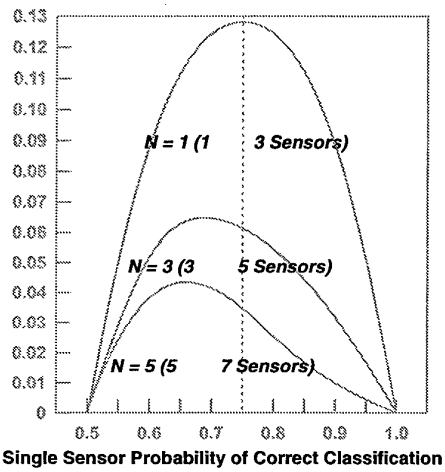


Figure 5 *Marginal gain in the probability of correct classification by adding sensors in going from 1 to 3, 3 to 5, and 5 to 7 sensors as a function of the single sensor probability P, as calculated in [17].*

In another recent study [18] the dependence of decision reliability on different fusion strategies for the case of 2 different sensors (i.e. sensors that are not identical) has been investigated based upon simple Boolean-type operators, along the lines of a probabilistic detection system (two hypotheses that are mutually exclusive and span up the whole universe of discourse). In a way this study allows to extend the validity of the rules formulated above to more subtle fusion strategies than the 'majority vote'. Also from these results we may conclude that for fusion to be successful the single sensor detection probability must be higher than 0.50 and that, in the case of two-sensor fusion, increases in detection performance of maximum 11% (from a single-sensor probability of detection P_1 of 0.8 to a two-sensor fused probability of detection P_2 of 0.899), or 17% (an increase from $P_1 = 0.80$ to $P_2 = 0.941$) have been calculated, depending on the fusion method.

Although these results do in no way preclude a substantial increase in the combined sensor performance brought about by a suitably chosen sensor fusion strategy, these numbers indicate that if measurements from identical, hard decision making sensors are combined under the assumption of statistical independence, the marginal gain in performance will be limited to a percentage of a few tens, say 10-20% .

3.3 The sensor fusion problem

The process of sensor fusion can be viewed in a number of different ways, depending on the goal that one wants to achieve: One may view the synergy between sensors as a scheduling problem: How can one schedule the available resources (an ensemble of sensors) in such a way that the observation process is optimal (*e.g.* that a typical category of targets can be detected)? This clearly depends on the availability and performance of the sensors as given by the physical properties and capabilities of each sensor, but in a military situation the availability may also depends on such factors as physical damage or jamming of sensors and environmental conditions (*e.g.* fog, sunset, night, and "clutter", *i.e.* (back)scattering from land, see and atmosphere). Basically the application of sensor fusion therefore involves the solution of an multivariate, generally nonlinear, time-dependent optimization problem.

It is worth noting in this context is the importance of *optimization* as a keyword in sensor fusion at various levels. Above we have already mentioned 'optimal' in the definition of fusion, as well as in the remark on sensor management. Combining these two observations naturally gives rise to the study of *nonlinear optimization*, a field that has since long attracted the attention of researchers from very different fields, *e.g.* mathematics, game theory, AI, systems theory, information science, physics,

biology, operations research and various other areas, and that despite all these efforts still holds many secrets. The general approach in arriving at (approximate) solutions is by *heuristic* methods. Exhaustive search is generally precluded because even small practical problems already suffer from the combinatorial explosion.

One of the problems involves the formulation of the function to be optimized, because of the difficulties to define a suitable measure for such notions as: better usage of resources (economy), robustness *vs.* redundancy, increasing the statistical accuracy of an observation, earlier, timelier achievement of an acceptable level of observational accuracy. For all these notions it is possible to come up with ad hoc definitions, but often it is felt that including too much heuristics makes the solution too subjective. A final point to note is that it is often desirable to have the possibility to delay decisions (classifications) a while, thereby increasing the confidence level of the final decision .

The nonlinear combination of sensor information holds possibly a key to gaining insights in the exploitation of sensor fusion. Nonlinearity can be introduced in signal analysis in a number of ways:

- Decision making is a nonlinear process: hard decision making consists of making a choice out of a (finite) set of alternatives, in its simplest form choosing between two or more mutual exclusive hypotheses H of the form: "IF (signal > threshold) THEN (H is true) ELSE ($\neg$H is true)", where H $\cup$ $\neg$H = U, using binary logic. The detection of a target basically follows this simple scheme. Statistics is subsequently applied in order to take into account the stochastic nature of sensor performance, fluctuations in the environment and in the target parameters, and notions as "false alarm rate", "detection probability", "misclassification rate" are introduced. So it appears relatively straightforward to view the fusion process as a synergy between taking decisions based upon two or more sensors. the difficulty associated with crisp, binary type decisions is that (too) much of the intrinsic sensor information is discarded, thus making it virtually impossible to use the decision in a subsequent decision stage. In analogy with the paradigm of human decision making one needs to introduce two additional properties, not present in the classical approach:
  1. A complex decision consists of a hierarchy of simpler, partial decisions.
  2. In order to create a continuous mapping between inputs and the final outcome, it is desirable that decisions are *soft* decisions, *e.g.* represented by a fuzzy number of by a degree of membership associated with a certain hypothesis (see *e.g.* [19]). In this way a smooth propagation from sensor input to final decision is warranted. The use of soft decisions has the added advantage of increased robustness and ability to handle conflicting information.

- An alternative to soft decision taking for constructing an accurate, fused decision is by *delaying* the final decision in time and by recursively reconsidering the decision in those cases where sensor inputs would yield ambiguous or, at least partially, conflicting decisions. In its simplest form this means that a three-valued logic has to be selected in the case of having to make a binary decision (*i.e.* having to choose among two mutually exclusive hypotheses H_0 and H_1 that cover together all cases: $H_0 \cap H_1 = \varnothing$). Thus a third hypothesis H_2, corresponding to "cannot decide yet between the alternatives", is introduced. If the initial decision is H_2, sensor inputs are scanned again in the next data acquisition cycle to resolve the outstanding decision. In this way one may construct a successive refinement of the decision by which a better support for the final decision is created at the cost of response time. In this way has shown that it is possible to obtain positive synergy for suitable parameter regimes and after enough recursions [18].

- Once we allow soft decisions to exist for some time, we have to answer the question what type of 'fusion operator', or "fusor", should be used when aggregating soft, partial decisions, and if possible select the 'best' aggregation scheme. One can easily imagine that this is not an easy question, since a multitude of schemes are possible and the answer is clearly context dependent. In [18] several schemes have been investigated. Although the detailed final decision is dependent on the actual fusion operator, some general features prevail. All

fusors have in common that they are *nonlinear*, ranging from fuzzy generalizations of the Boolean "AND" and "OR" operator, to a "majority decision" fusion scheme.

These observations guide us in the search for an effective sensor fusion algorithm. The relevance of fuzzy logic, fuzzy measures and fuzzy aggregation operators for the development of fundamental sensor fusion concepts in detection and classification is striking. the optimum way how to combine information from different sensors depends on the particular gals of the mission for which the observation is made, as well as on the environmental circumstances and the properties, observables and relative level of confidence of the individual sensors. This will be the subject of studies with the SFD. Note that fuzzy logic is *not* introduced here to allow for human vagueness and subjectivity, but rather to get rid of the limitations of Boolean logic operations and the inherent crisp decisions at early stages of a complex decision tree.

3.4 The key concept of sensor fusion: Synergy

A general concept that is intimately connected to the idea of fusion is that of *ergodicity*, *i.e.* the concept that the outcome of an observation is unique, independent of the fact that one makes a series of consecutive measurements with one system, or that one makes N parallel setups and combines the N different outputs at one time. In the macroscopic physical world the concept of ergodicity generally is assumed to hold without exceptions, although in microscopic physics some examples have been found in spinglasses, where ergodicity does not hold. For a review the reader is referred to [20]. Throughout this paper we will assume ergodicity. However it should be pointed out that this statistical principle is sometimes difficult to apply in the real world, such as in military observations, because of the fact that we are mostly dealing with *single,* isolated events.

The fusion of sensor observations, *i.e.* the combination of observations obtained from the same sensor at different times (temporal fusion), or the combination of simultaneous observations take by a number of equivalent sensors (repetition) or the combination of sensor observations with a priori information obtained from previous measured data, is the focus of attention in the present work. The key concept in fusion is how to take advantage of *nonlinearity.*

We are not so much concerned with increasing the accuracy of an observation by repeating a measurement a number of times and thus reduce the statistical variation of the average. Rather it is our objective to extract additional information out of this data set (reduction of information) by correlating (*not* superimposing) the measurement with observations from other sources. Moreover in the case of fast moving targets it is generally impossible to obtain a sufficient number of samples to apply statistics.

Two different types of measurements are of interest in military observation systems:

1. determination of presence, position, orientation, and speed of a target and
2. identification and recognition of a target (type, affiliation etc.)

Although identification clearly is an entirely different characteristic of a target compared to its position and speed and although the latter can generally be determined at much larger distances than those at which identification can the accomplished with reasonable confidence, identification can help improve the accuracy with which speed can be determined and vice versa. In particular the identification of a target may be helped through a wealth of observations, whereas establishing the position and speed of a target can only be accomplished by the few sensors. The identification of a target will be accomplished more easily, because of the higher dimensionality of the 'feature vector', provided that a good database of properties is available. The basic problem in recognition is to exploit the high dimensionality and representing data in such a way that differentiation between various possibilities becomes easier. Therefore the task of sensor fusion is the combination of, possibly incompatible, measurements and to try and construct from these an improved environmental picture. In this modeling one needs to include the confidence level of the new measurement, as well as *how* to combine this information with the already existing picture. Various schemes have be proposed in the past: *e.g.* Bayes' rule of combination from statistics, belief measures, and

Dempster Shafer theory. Although most of these methods have a sound theoretical basis, their application sometimes lacks theoretical justification or simply yield non-intuitive results in specific situations. This makes it difficult to compare results that are obtained with different methods.

3.5 A systematic approach to Level 1 processing

In this section we will outline a practical approach to sensor fusion in military observation, following the theoretical framework referred to as Level 1 data fusion in the JDL model. Two generic tasks are of importance in almost all observation processes, *viz.* detection and recognition. Despite the fact that different sensors and methods commonly used to accomplish these tasks, it is obvious that the successful completion of one task will almost certainly have a positive effect on the other one. If however the identification and recognition tasks are considered to be "hard" decisions that result from the independent processing of separate sensors, interaction of the two processes can only take place *after* the first process has reached a decision. In executing more complex tasks we have already indicated in Sec 3.3 that it may be more advantageous to allow for partial, delayed, or "soft" decisions which may

Table 1. Typical operational characteristics of military observation sensors.

SENSOR CHARACTERISTIC	DESCRIPTION
Detection performance	Detection characteristics, *e.g.* false alarm rate, detection probability, detection range) for a calibrated target in a given noise background.
Spatial and temporal resolution	Ability to distinguish between two targets that are moving close together
Spatial coverage	Spatial volume covered by a sensor. For staring sensors this would be the field of view, for scanning sensors also the scan pattern must be taken into account
Detection/Tracking modes	Search and tracking modes performed: 1. Staring and scanning 2. Single or multiple target tracking 3. Single or multimode (track-while scan/stare)
Target revisit rates	Rate at which a given target is revisited in case of a scanning sensor
Measurement accuracy	Statistical accuracy of sensor measurements
Measurement dimensionality	Number of independent physical observables that can be measured with the sensor
Hard vs. soft reporting	Sensor outputs can be Boolean (signal above or below a threshold) in the case of hard reporting, or alternatively they can make a soft decision, i.e. output quantitatively a measure to what extent a hypothesis is supported by the measurement.
Detection/ track reporting	Sensor reports each individual target detection or maintains a time sequence representation (track) of the target's trajectory.

offer a way to separate the various goals and thus allow us to break down a complex task into a hierarchy of relatively simple decisions. Partial or "soft" decisions can be combined at earlier moments in the processing chain without discarding too much information and thus offer a method for applying early fusion of sensor streams. Before discussing in more detail how soft computing methods can be used to achieve sensor fusion, we will first review the signal processing steps that are necessary to benefit from sensor fusion. An overview of the physical characteristics of observation sensors that are relevant for defense applications is given in Table 1 (from Hall [15]). In addition sensors can be free running, synchronized if they have commensurate sample rates, or can be operated in a 'master-slave' mode.

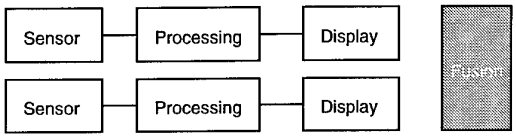
From Table 1 it can be seen that most sensors are decision-oriented towards achieving the detection, positioning and tracking or recognition of a target. These tasks are generally considered sequentially and independently. The classical way in which fusion is applied is by transforming a physical measurement into some hard decision (*e.g.* a 'plot', 'track', 'identity', etc.) that is communicated to the user via the man-machine interface, generally an optical display. The fusion process of the information shown on a number of different displays then takes place *in* the mind of the operator, who assesses the situation, or makes a threat analysis. All these fusion processes take place in the human mind, after the sensor signals have been processed completely (Fig 6a).

A first step towards true multi-sensor fusion is 'late' fusion (Fig. 6b): the construction of a special, goal-oriented architecture that fuses on the level of images, with the goal to enhance the image (*e.g.* combining IR and visible light images using some false color scheme), or to fuse the consecutive plots of moving targets into a single track by taking into account some type of assumed target dynamics, or the fusion of tracks generated by different sensors (*e.g.* two radars or a radar and an electro-optical sensor). The results are qualitatively better, especially if the sensors are one different platforms, or, alternatively, under difficult circumstances (weather, jamming) when one sensor can be supplemented by the other. In the case of *early* fusion with many same sensors, one can process directly on the raw data without much preprocessing, possibly except for bias removal (Sec. 3.5.1). In this sense one may view SAR (Synthetic aperture radar) also as a 'temporal' fusion process, combining different samples of single sensor to simulate a kind of multi-sensor phased array with a much larger aperture than that of the actual sensor.

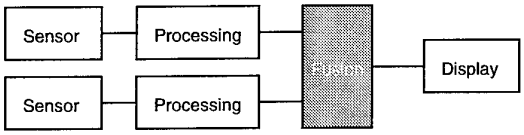
The early fusion process that we would like to investigate with the SFD is schematically given in Fig. 6d: a collection of interacting sensors, each contributing to one or more decision trees of positional *c.q.* feature declarations. Each tree represents a hierarchical decision process, building up from fast, low level, noisy decisions, up to well-founded, abstract decisions that require some time to take. The sensor fusion architecture outlined in Fig. 6e illustrates the ultimate goal of sensor fusion: mixed early and late fusion processes on different levels of abstraction, each subgoal benefiting from its own support set of sensors and the results presented on one display to the human observer in a representation that is highly informative, indicating at the same time alternatives, as well as the associated confidence levels.

We have already noted in Sec. 3.2 that application of multiple sensors in general can be beneficial to the accuracy of an observation, even though one low resolution sensor directs a sensor with a higher resolution ("cueing") without actually sensor *fusion* occurring. For making hard detection statements explicit knowledge on the clutter is required, and in order to improve single-sensor detection capabilities, one needs to carefully analyze the model assumptions that have gone into the sensor design. New, more detailed clutter models corresponding to the state of the art of sensor technology, may be needed for the detection of harder targets. On top of that sensor fusion may be applied. But it should be noted that sensor fusion can never make up for poor clutter modeling.

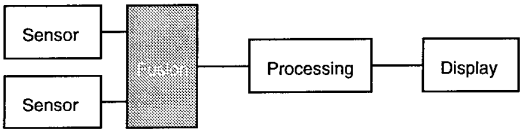
For any sensor fusion process to be successful, one has to properly prepare the raw signals originating from the single sensors. From a system point of view one has to determine the stage at which fusion has to take place (ranging from 'early' to 'late') in relation to the goal that must be achieved by the fusion process. Once this has been decided the first step in preparing the sensor signals for fusion is *alignment*, to guarantee that the fields of view (FOV) have maximum overlap. Early fusion



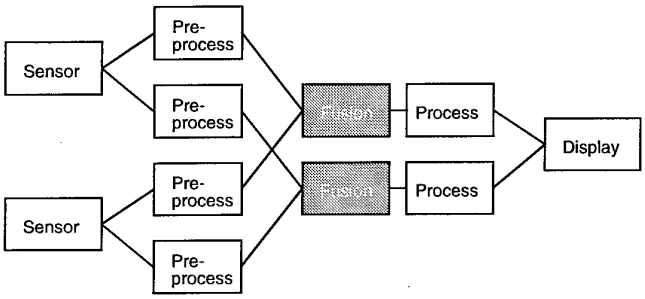
Fusion of information in human mind



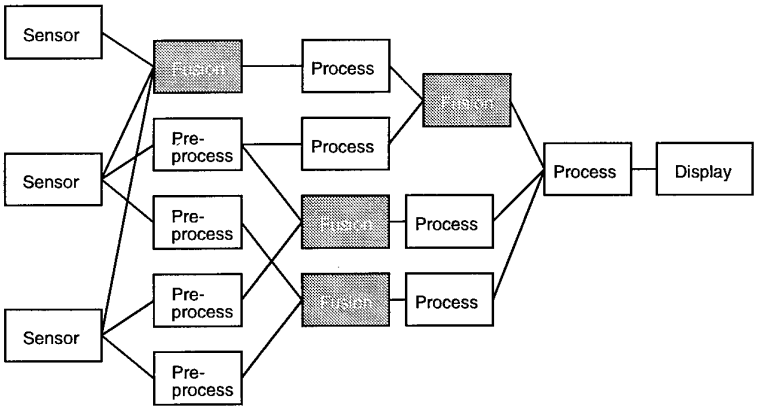
Late fusion: e.g. Image fusion, Track fusion



Early fusion with similar sensors, e.g. SAR



Early fusion with different sensors and with different fusion objectives



Distributed multisensor fusion: concurrent early and late fusion on different levels

Figure 6 *Different fusion architectures starting from late fusion (i.e. in the human mind), via early, multisensor fusion with the same and with different sensors to the general case of multisensor multiobjective fusion on different levels concurrently.*

can only be successful if there is an *overlap* between the FOVs. The second step is the proper correlation of the same objects in the FOV of the sensors. This task can in practice be quite laborious when many targets are observed simultaneously (large FOV; scanning sensors). Before any of these preprocessing transforms and associations can be carried out however, attention has to be paid to the optimization of performance of the isolated sensor by *removal of the biases* in the observed quantities.

3.5.1 *Bias removal*

An aspect of particular importance to sensor fusion is that of on-line *bias removal* [21]. As discussed in the previous sections, sensor fusion is expected to improve the observation process, by extracting more information from the sensors than could otherwise be done by processing the different sensor streams independently and therefore to be beneficial by enhancing the performance of observation and detection tasks, especially under difficult circumstances. However, it should be noted that combining information from different sensors at an early stage initially *complicates* the analysis compared to *e.g.* combining subsequent images from the same sensor, or to the separate processing of the single sensor streams and combining the results at a later stage, *i.e.* at a more abstract level. The reason for this is that the relative importance of systematic measurement errors compared to accidental errors *grows* when increasing the number of sensors. Accidental measurement errors follow some type of stochastic and will average out to zero by repeating the measurement. In contrast, *systematic* errors will persist and produce a *bias* in the measured quantity that can only be removed by careful theoretical analysis and proper design of the experimental setup. In a single sensor observation, *e.g.* a video sequence from a CCD visible light camera, bias errors in position are not important as long as one is only interested in recognition or in the relative positions of objects in the scene. If the *absolute* position of a target should be estimated from the scene, the correct estimation of bias errors becomes essential and many more data, such as the positioning of the tracking platform of the camera, the specification of the optics and the observed spectral range become important. Typical quantities that are sensitive to bias errors are position coordinates. As soon as the observations of more than one sensor have to be combined, we are confronted with differences in bias. These bias errors originate from the relative positions of the sensors, their orientations, different fields of view, different temporal and spatial resolutions, as well as from differences due to the spectral properties of the medium. In fusing long range optical and radar observations one has to take into account that, depending on the specific circumstances, the optical path and the rf path to the same target are not necessarily straight nor equal and have to be accounted for ('alignment'). The number of bias parameters to be estimated for a typical sensor, such as a radar antenna, could be as high as ten [21]. This number illustrates that the task of estimating the various systematic errors is by no means a trivial one. All these effects have to be corrected for *before* sensor fusion can be applied. If the preparatory corrections are not precise enough they could easily deteriorate the performance of the sensor fusion process, or even destroy it completely. In a practical situation the removal of the many systematic errors contributing to the outputs of the sensors is a matter of compromise: because the high data rates at the sensor level, it will be impossible to let a bias removal process run at the same pace in parallel with the data acquisition process. Some time averaging will be necessary, filtering out the highest order fluctuations. On the other hand, if the estimates for the bias levels are incorrect or too scarce, the effectiveness of the sensor fusion process is negatively affected, sometimes even reducing functionality below the performance of a single sensor system. An example of this is the tracker function of a radar sensor: the existence of a bias error in the position vector is of only minor importance in a single tracker system, but may result in creating multiple tracks that are originating from the same target in a multi-radar tracking system, if no compensation is provided for the sensor biases. The possible benefits of sensor fusion, *e.g.* constructing one track of high accuracy using twice the number of plots, are thus precluded in the preprocessing stage, even before a sensor fusion algorithm could be applied. Therefore raw single sensor data should be compensated for bias errors, before they are associated with tracks. In [21] a system has been described that constructs at regular intervals a set of linear equations describing to first order how the various bias errors in position range and

direction of the sensor affect the output signal. At regular intervals the set of linear equations is solved to obtain the best estimate in a least squares sense of the single sensor bias signals, which are then used to filter the new samples obtained in the next data acquisition cycle before applying to the sensor fusion process. The bias update rate is much slower than the sampling rate, partly because of the real-time processing constraints, partly because adopting a coarser granularity in time effectively limits the bandwidth of bias fluctuations. This is motivated by the assumption in the analysis that the bias is constant and therefore its fluctuations are by definition slow on the time scale of the measurement process, including the sensor fusion processing. This batchlike bias calculation is called 'quasi-recursive since it combines the structure of batch processing with an infinite batch size, because of the recursion.

3.5.2 *Fusion*

As we have seen in the previous section, it is worth to select the proper sensors and it is also necessary to spend sufficient effort in the preparation of the signals before they can be fused. It should be noted that fusion is *not* a magic way to reduce the quality or the price of a sensor and still get the same observation performance. Making accurate observations requires a great deal of study, modeling and experimentation; the successful application of sensor fusion initially means *more* work than the application of isolated sensors.

Depending on what the goal of the fusion process is, there are different time scales to consider:

- The maximum sampling rate, limiting the highest instantaneous bandwidth of a signal, which is essentially a measure for the sensor resolution: in a pulsed radar it is the range resolution and in a camera system it is the transverse 2D spatial resolution.
- The frame rate. This rate is important in extracting information from a time sequence of camera images, *e.g.* with optic flow analysis. By using the temporal correlation of objects in the pictures when the platform is moving, it is possible to make estimates of the distance of each of these objects.
- The batch processing rate: If signal processing is not performed on a continuous basis, as is *e.g.* the case of Kalman filtering, one is not forced to make *a priori* assumptions and may therefore be more accurate than Kalman filtering especially in early stages of the signal processing chain. A disadvantage is that batch processing is considerable costlier than Kalman filtering in terms of processor (CPU) time.

Although we will here focus only on sensor fusion, many more aspects in modeling need to be considered, before an attempt to apply sensor fusion should be made. We mention here only a few:

- clutter modeling: the type of statistics, correlation times and correlation lengths.
- construction of an 'a priori' environmental data base, necessary to make (partial) decisions on identity and position.
- the modeling of the target dynamics if the target is moving, and its significance for improving classification.

We have already seen in Sec. 3.2 that fusion of multiple same sensors mainly improves the *statistics* of the observed phenomenon compared to the single sensor statistics. In the case a sensor suite exclusively consists of identical sensors, one may wonder how by fusion one can extract additional qualitative information from the single sensors. The following scenarios can be thought of to illustrate the point:

1. Increase in spatial resolution via synergy between the sensors: By suitable sensor setup and associated processing a synthetic aperture can be formed that is much larger than the single-sensor aperture.
2. Certain NCTR observations (*e.g.* jet engine modulation **[REF]**) are confined to relatively narrow solid angles. In those cases it can be very effective to have multiple sensors located relatively widely apart to increase the probability of detection.
3. During tracking multiple same sensors may help resolve details of targets flying in formation, or identify targets that vary strongly in aspect dependent on their orientation (*e.g.* a stealth bomber or a helicopter). Multiple sensors of the same type make it easier to do the recognition by correlating the different aspects of the target.

If the observables of the fused sensor suite are mutually 'orthogonal', complementary fusion will invariably yield more information than each of the separate sensors can provide. It is therefore conceptually the simplest way to demonstrate the practical benefits of fusion. In this context one could make an analogy between a single sensor observation of the real world as a (stochastic) *projection* of the real world into a sensor observation space. In this analogy, fusion can be seen as (partially) reconstructing the real world, representing it as the direct product space of all observation spaces of sensors that participate in the fusion suite. Effectively the dimensionality of the observation space increases by adding up the dimensions of complementary sensor spaces. Adding sensors of the same type through the Ergoden hypothesis basically improves the *statistics* of the observation in the particular sensor space. However the *dimensionality* of the single sensor space does *not* increase by adding more sensors of the same kind.

In the case that the observables of the sensor suite are *not* 'orthogonal', the fusion process can increase the speed with which a predefined accuracy or resolution of observation is achieved by acting as a smart scheduler, *e.g.* via cueing. The shorter response time is realized by first determining areas of interest by the sensor with the lower resolution or accuracy and the largest FOV, and then focusing attention on these areas using the high accuracy sensor, instead of scanning the entire area with a high accuracy sensor with a small FOV. In addition this type of sensor fusion increases the robustness: if the cued sensor fails or is jammed, the other one can take over, although with lower resolution or accuracy *(graceful degradation).*

An example of multisensor fusion with different sensors is the combination of a radar measurement and an optical image: if an airplane is observed by radar, the range from observer to the airplane is accurately measured, while azimuth and elevation are only coarsely determined. In contrast, an optical measurement provides azimuth and elevation with relatively high accuracy, whereas the uncertainty in range is high. Combination of the two sensor types can considerably diminish the absolute uncertainty in the position of the airplane in 3D space, which is a natural consequence from the two complementary measurement principles (see Fig.4).

Finally we note that from a system theoretical standpoint the observation process of a sensor may be viewed as a projection operator. In this context sensor fusion holds the more potential the earlier it is applied (no significant reduction of information). Intuitively we can express the expected effect of the fusion process symbolically as:

$$S_1 \oplus S_2 \geq S_1 + S_2$$

where $\oplus$ represents the fusion operator and S_i is a quality measure associated with sensor i.

3.5.3 *Fusion with fuzzy aggregation operators as a way to reduce complexity.*

In our study we will concentrate on the synergy of sensors at the signal level (*"early fusion"*). Although this does not preclude the use of *a priori* information, or taking into account any human generated inputs and feedback, we focus specifically on the sensor area, because of its fundamental nature, because the signals are not yet distorted or corrupted by incoherent, independent, or *ad hoc* signal analysis operations, and because there is relatively little room for subjectivity. The attractivity of this approach is of course that by operating close to the primary sources of information, one expects to be able to significantly enhance the detection and recognition processes by applying sensor fusion.

In terms of the functional model for data fusion as proposed by the Joint Directors of Laboratories, we will be exclusively concerned with level 1 processing. It should however be pointed out that the JDL model is somewhat artificial and in our view does not take into account the wealth of possibilities that sensor fusion allows. For this reason we will also include into level 1 processing part of level 4, dealing with real-time sensor maintenance or sensor scheduling. It is absolutely necessary to include real-time feedback aspects *e.g.* the on-line calibration, cueing, and sensor suite optimization in case of damage.

In reviewing data fusion literature it is apparent that the vast majority of the research is concerned

with level 2 and level 3 processing, dealing with respectively situation and threat assessment. This can be understood in the sense that there exists a need to automate the decision process at the higher layers and that it is necessary to frequently interact with humans and a priori knowledge. However from the point of view of information theory it is surprising, because it is obvious that the information content in the *early* stages of data processing is higher than in the more abstract, condensed streams of inference results processed in levels 2 and 3. Even in level 1 processing there are more papers dealing with *track* level fusion compared to *plot* level fusion, although processing of plots would be preferred being earlier than tracks. Of course there is a price to be paid for using "raw" sensor data: one needs considerable processing power, because of the much higher bandwidths. However it seems that this is not the main reason for the interest in levels 2 and 3; more researchers seem to be attracted by the neatness of well-structured information-theoretical data structures instead of focusing on real-time processing of vast amounts of data. As is well-known from *e.g.* high-energy physics experiments, such situations are best handled by constructing a hierarchy of decision levels ('trigger hierarchy'), starting from relatively crude binary decisions taken in very short time, to more elaborate, subtle inferences on the resulting data stream.

A sensor fusion system that receives raw signal inputs from all sensors, retains control over all primary sensors and has, at least theoretically, a number of advantages over secondary (level 2 and higher) fusion. Apart from the larger information content of raw information, it should however be noted that each fusion step requires a certain processing time and that in early fusion it is effectively the slowest sensor in the fusion suite that determines this latency, even if we neglect the execution time for the fusion process itself. In addition the latency is increased because fusing information from autonomous, asynchronous, and dissimilar sensors requires synchronization. A designer of a sensor fusion testbed should be aware of this and take precautions to ensure that the pileup of latencies does not degrade the real-time performance of the fusion system as a whole, or jeopardize the quality of the fusion process, *e.g.* by constructing a deficient situation awareness picture.

It is relatively straightforward to illustrate the idea of sensor fusion by the improvement of operation of a target tracker during the observation of a maneuvering target in cluttered areas. A variety of different sensors can be used to generate plot reports and these can be combined on the basis of a simple confidence criterion, in turn based on the presence of clutter for a particular sensor. In [21] this has been illustrated. However, although useful as an idea this example basically supports the idea of increasing robustness by increasing the number of different sensors. Our goal is more ambitious: we would like to improve the *quality* of the single sensor conclusions in such a way that

$$P(S_1 \cup S_2) \geq P(S_1) + P(S_2),$$

or if this condition is too strong, at least

$$P(S_1 \cup S_2) \geq MAX(P(S_1), P(S_2)),$$

where $P(S_i)$ indicates the performance or 'added value' (the effectively information accumulated over time) of stream S_i, measured by sensor i. From this formulation it is clear that in order to model sensor fusion, we will need *nonlinear* operators.

The earliest attempts to combine measurements from multiple sources is by Bayes [22].

There are a number of difficulties connected with the application of the Bayesian sensor fusion formula:

- difficulty in assigning a priori probabilities
- complexity when there are multiple hypotheses and/ or multiple conditional events
- requirement that hypotheses have to be mutually exclusive and exhaustive
- absence of uncertainty modeling

In trying to find an appropriate way to model fusion in the presence of uncertainty and take advantage of the nonlinearity of the process, Dempster created a generalization of Bayesian theory that allows the incorporation of uncertainty by using (overlapping) probability intervals and uncertainty modeling to determine the likelihood of hypotheses based on multiple evidence. The essential generalization of Dempster-Shafer (DS)

theory is that not all hypotheses need to be mutually exclusive as in the Bayesian theory. In DS fusion an evidence interval is assigned both to single and to more general propositions, instead of assigning directly a probability to hypotheses as in Bayesian theory.

Evidence in DS theory is modeled based on the concept of probability assignment m(A), a function m: $\wp(X) \rightarrow [0,1]$, where $\wp(X)$ is the power set of X, $m(\varnothing)=0$ and

$$\sum_{A\in\wp(X)} m(A) = 1$$

For each set $A \in \wp(X)$, m(A) expresses the proportion to which all available and relevant evidence supports that a particular element of X belongs to the set A. Although this normalization condition resembles a similar equation for probability density functions, there is a fundamental difference: probability density functions are defined on X, where as the basic probability function is defined on the *power set* $\wp(X)$ of X.

DS evidence theory [23] has a clear advantage over the Bayesian formulation that it allows for uncertainty in the form of conflicting evidence by adopting belief and plausibility measures that unlike probability measures are no longer additive. Instead, weaker conditions hold for these measures as expressed by:

$$Bel(A_1 \cup A_2 \cup \ldots\ldots \cup A_n) \geq \sum_j Bel(A_j) - \sum_{j<k} Bel(A_j \cap A_k) + \ldots\ldots + (-1)^{n+1} Bel(A_1 \cap A_2 \cap \ldots\ldots \cap A_n)$$

and

$$Pl(A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_n) \leq \sum_j Pl(A_j) - \sum_{j<k} Pl(A_j \cup A_k) + \ldots\ldots + (-1)^{n+1} Pl(A_1 \cup A_2 \cup \ldots\ldots \cup A_n)$$

where the belief and plausibility measures, *Bel(A)* and *Pl(A)*, respectively, are defined by the basic probability assignment function *m* by:

$$Bel(A) = \sum_{B\subseteq A} m(B)$$

$$Pl(A) = \sum_{A\cap B\neq\varnothing} m(B)$$

From the definition it is obvious that $Pl(A) \geq Bel(A)$. the duality of the two measures is given by:

$$Pl(A) = 1 - Bel(\overline{A})$$

Although the probability measure Prob(*A*) is *not* a fuzzy measure, it may be seen as a limiting case of belief and probability measures:

$$Bel(A) \leq \text{Prob}(A) \leq Pl(A)$$

In the DS approach Bel (sometimes called 'support') and Pl are interpreted as *lower* and *upper* probabilities, respectively, or alternatively as the boundaries of the evidencial interval. From the definition of the belief and probability measures the following identities hold:

$$Bel(X) = 1 \text{ and } Pl(\varnothing) = 0$$

These identities reflect that total ignorance may be expressed by $Bel(X) = 1$ and $Bel(A\neq X) = 0$ for all $A\in\wp(X)$, or, alternatively, by $Pl(\varnothing)=0$ and $Pl(A\neq\varnothing) = 1$ for all $A\in\wp(X)$. In DS theory the combined evidence $m_{1,2}$ from two sensors (1 and 2) is expressed as:

$$m_{1,2}(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K} \text{ for } A \neq \varnothing \text{ and } m_{1,2}(\varnothing) = 0$$

where $K = \sum_{B\cap C=\varnothing} m_1(B)m_2(C)$.

The renormalization term (1-K) follows from the condition that $m_{1,2}$ must again be a basic probability assignment, *i.e.* $\sum_{A\in\wp(X)} m_{1,2}(A) = 1$, and reflects the handling of conflicting evidence. K may be interpreted as a measure of the resulting uncertainty after fusion.

Noting that belief and plausibility measures are both examples of Sugeno's λ-fuzzy measure g_λ [24], the question arises whether it is possible to combine the intuitive ideas on sensor fusion and

the properties of g_λ, we willl argue that in contrast the basic probability assignment in DS theory, fuzzy g_λ measures can be tuned to the problem consideration.

We will take a closer look at this in the following and propose to view the multisensor fusion process in terms of a synergy between (sets of) sensors that are grouped in such a way as to support a certain decision or hypothesis. Instead of attempting to make a decision (detection or classification) in one step, either by a single sensor, or by a linear combination of a group of sensors, it is proposed to combine supporting evidence for a hypothesis in a hierarchical way by building a tree structure that combines at the lowest level clusters and in the next levels combines the outputs of several initial cluster in superclusters and so on.

At each level in the tree decisions need to be made from different sources with different importance. This is conveniently modeled by the Fuzzy λ-measure g_λ ($0 \leq g_\lambda \leq 1$). In particular we have in the absence of relevant information towards the classification/detection goal: $g(\varnothing) = 0$ and $g(A) \leq g(B)$ if $A \subset B$. This coincides with the intuitive feeling that if the evidence support is larger (*i.e.* if we observe the same scene with more sensors.), that then the information content should also increase. In addition the following property holds for all $A, B \subset X$ with $A \cap B = \varnothing$:

$$\exists \lambda > -1 \quad g(A \cup B) = g(A) + g(B) + \lambda\, g(A) g(B)$$

This again supports the intuition that adding more independent data ($A \cap B = \varnothing$) cooperates towards an increase in confidence about the final decision. In addition both intuitive features about the fusion of two independent sensors are reproduced, *viz.*

$\lambda \geq 0 \qquad g(A \cup B) \geq g(A) + g(B)$, *i.e.* fusion is more than superposition and

$-1 < \lambda \leq 0 \quad g(A \cup B) \geq MAX(g(A), g(B))$, implying that even when the level of confidence is larger than 1 (as reflected by the negative λ) then still it may be fruituitous to apply sensor fusion. In the event that λ=0, *i.e.* the case where all sensors have the same importance and complexity covers the whole universe of discourse, the degree of importance g towards the final decision become additive and coincides with the definition of a probability measure.

Following ideas put forward in [25], sensor fusion may be modelled using the concept of fuzzy integration. For a review on the role fuzzy integrals in the framework of multiple criteria decision making see [26]. A fuzzy integral may be interpreted as an aggregation functional of subjective evidence, where the subjectivity is expressed in the fuzzy measure, and integation is defined over measurable sets [27]. In contrast to normal (Lebesque) integrals, fuzzy integrals are *non-linear* functionals. It is exactly this nonlinearity and the possibility to include a fuzzy measure g_λ that is attractive in the context of fusion.

Formally Sugeno's fuzzy integral is defined as [24]:

$$\int_{fuzzy,A} h(x) \circ g(*) = \sup_{E \subseteq X} \left[\min \left(\min_{x \in E} (h(x), g(A \cap E)) \right) \right]$$

Hence the evaluation of the fuzzy integral may be interpreted as evaluating the degree of agreement between objective evidence h(x) and the expectation.
We will not discuss the properties of this fuzzy fusion operator here, but note that it is ideally suited to combine information of different sources without the risk of the combinatorial explosion.

A similarity between fuzzy fusion (FF) aggregation and the way DS theory fuses data from different sources is that both make use of fuzzy measures: DS uses the belief measure exclusively, whereas the FF operator uses the g_λ measure. For $\lambda \geq 0$ this measure is equivalent with the DS belief measure). The conceptual difference between both methods is twofold:

1. the frame of discernment (or universe of discourse) is different
2. clear separation of objective and subjective uncertainty in the case of FF

We will illustrate these points in the following: For the FF scheme the frame of discernment contains the information sources (the sensors) related to the hypothesis under consideration, whereas in DS theory the universe of discourse contains *all* possible hypotheses. In combining the

different information streams, the fuzzy fusion aggregator fuses all sources according to their relative a priori importance as well as to the degree to which each sensor supports the hypothesis under consideration. In contrast, the fusion process in DS theory associates with each knowledge source a belief function that is defined over the power set of the set of hypotheses and combines these in the fusion process. The evaluation of Dempster's rule of combination therefore has exponential complexity $O(2^N)$, where N equals the number of hypotheses under consideration. In contrast, in FF one fuzzy integration has to be calculated, which implies that g_λ has to be calculated nN times, where n is the number of sensors. The evaluation of the fuzzy integral can then be carried out in O(n) steps. The second advantage of fuzzy aggregation is that both the weighting with the degree of support by which a sensor supports a certain hypothesis, but also the weight of importance of a certain sensor, reflecting a subjectivity or a *a priori* confidence in the particular sensor.

We therefore conclude that the formalism of fuzzy measure theory may offer an opportunity to model the process of sensor fusion in a natural, intuitive and adequate way, allowing arbitrary sensors to be fused and allowing different ways to weigh different combinations of observations (see Table 2). The example reflects the property that if sensor 2 ("the most decisive sensor in support of the hypothesis under consideration of the three sensors") is combined with one of the other two sensors (1 or 3), the combined evidence as reflected by g_λ must be larger than the sum of g_2+g_3. This is indeed the case, as we see from the example: $g_2 \oplus g_3 \equiv g_{23} = 0.687 > 0.5 = g_2 + g_3$ and also $g_{23} > g_{21} > g_{13}$, which we would also expect intuitively on the basis of fusion as a nonlinear operation with positive synergy.

Table 2. Example of a fuzzy g_λ measure with $\lambda > 0$ (belief measure).

Member of Powerset $\wp(A)$	Belief measure $g_\lambda(A)$ $\lambda = 3.109$
$\varnothing$	0
$\{x_1\}$	0.1
$\{x_2\}$	0.3
$\{x_3\}$	0.2
$\{x_1, x_2\}$	0.493
$\{x_1, x_3\}$	0.362
$\{x_2, x_3\}$	0.687
$\{x_1, x_2, x_3\}=X$	1.0

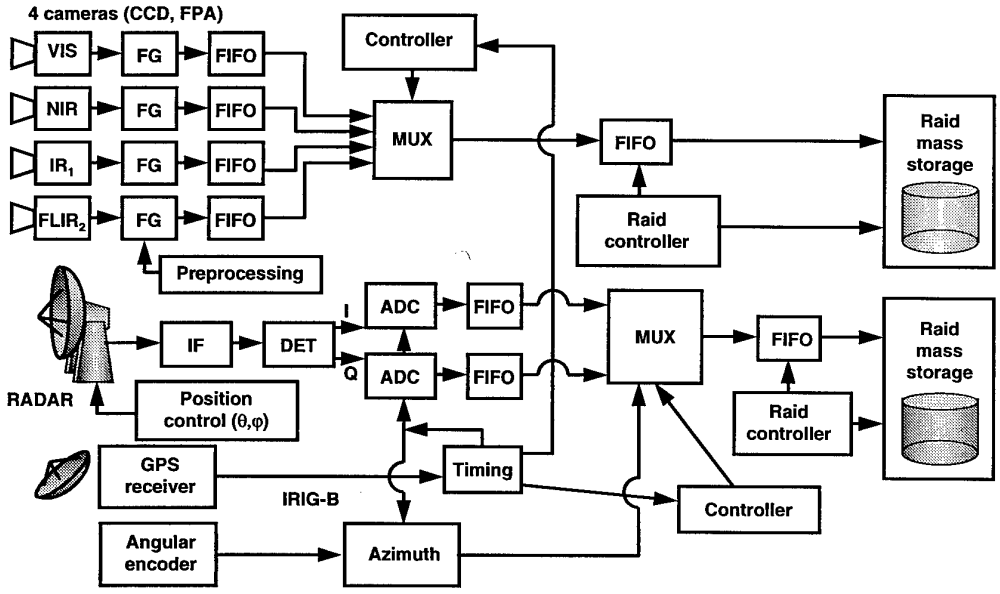


Figure 7 *The architecture of the SFD.*

4 IMPLEMENTATION OF THE SFD

4.1 System architecture

Multisensor fusion can be defined as the study of concepts and techniques developed for optimal information processing in a generally distributed multisensor environment through a clever integration of sensor data streams.

We will limit us here and consider only fusion at the lowest level, *i.e.* fusion of *raw* sensor data, possibly after some preprocessing, but excluding any combination with other sensor streams or large database management systems and other abstract (non-physical) information streams. Although the final goal is to do sensor fusion in real time, the initial goal is to select and develop, test and evaluate suitable sensor fusion algorithms. The minimum real-time demand on the sensor fusion testbench therefore is that as much as possible raw data can be stored on disk in real time. Even this rather obvious demand is in practice not so trivial.

The option of data reduction via filtering, *e.g* by MPEG or JPEG encoding schemes can be ruled out. The fact that video signals generated by commercial CCD cameras are interleaved for the comfortable perception of the observer, is in terms of data acquisition an undesirable feature that either halves the vertical image resolution or displays artifacts, corresponding to the superposition of two half-resolution frames captured at different times. This becomes *e.g.* visible in the case of fast moving targets close to the camera. Compression techniques are only allowed if they can be applied in real time and if they can guarantee to be invertable (lossless compression).

Because raw sensor data still have *maximum* information content, they are best suited for studying data fusion. At this level one also has the maximum data rates. During the conceptual design phase it was thought useful to design a data-acquisition system that should be capable of digitizing the inputs from four optical cameras, and a radar system. The radar signals should be stored with an effective bandwidth of 20 MHz, corresponding to an instantaneous data rate of 40 Msamples/s @ 12 bit and a sustained rate of 20 MB/s. A short survey of available commercial-of-the-shelf hardware indicated that AD conversion for radar can be done and that also frame grabbers for the CCD cameras are readily available. It seemed initially also possible to meet the required sustained data rate of approximately 48 MB/s for the datastream of all cameras and a similar data rate for the radarvideo signals. Apart from coping with these massive data streams, the signals must also be labeled in real time: the frames from the different cameras and each radar echo need to be timestamped before writing the signal to disk. A typical scene lasts, say, 5 minutes, so that if we want to record 3 sequences we would need in total about 50 GByte storage, requiring a redundant array of "inexpensive" disks (RAID). In the next step the measurements are stored on magnetic tape.

Both for practical as well as for storage reasons the data-acquisition system will be built around two powerful PCI-bus systems, each equipped with its own RAID storage system. The PCI-bus is specified up to a sustained transfer rate of 70 MB/s, so, apart from some operating-system overhead, no throughput difficulties are to be expected. For reasons of stability the Windows NT vs. 4.0 operating system was selected, which allows interference of the user in defining task priorities so that the full PCI-bus capacity can be dedicated to the data-acquisition process.

The architecture of the SFD is shown in Fig. 7 for the case of 4 cameras (2 CCD's, 2 IR FPA's) and a pulsed radar. Camera outputs may be either analog or digital, may be free-running or locked to each other and the radar sensor may be either staring or rotating. After AD conversion signals are time-stamped and stored on two RAID mass storage devices, together with the azimuth angle of the radar antenna. Care is taken to maintain the full dynamic range of the signals as much as possible, without reducing the original bandwidth. Absolute time is stored with each sample, *i.e.* for the cameras with each frame and for the radar per pulse return (time histogram). In this way the relative order of the various streams is ensured. The time stamp is derived from an IRIG-B coded timing signal produced by a GPS receiver and has a resolution of 1 ms. In order to guarantee also a unique ordering of radar histograms at the maximum PRF of 4 kHz, we would need a resolution of 250 μs. Special provisions have been devised to ensure that even at the highest PRF

unique ordering is achieved despite the lower time resolution of IRIG-B. Although IRIG-B synchronizes with absolute time (UTC), only relative timing is important, because the SFD is intended for use on the same platform, with the sensors positioned close together.

In case of a scanning radar antenna we store in addition to the proper sensor signals also the azimuth and elevation angles with their own time-stamp. Due to external influences, the scanspeed will in general not be constant, so that the sampling of the azimuth angle is not equidistant in time.

The data-acquisition system is currently under development in close cooperation with a system integrator. It has been shown that despite the advertised specifications given in the data sheets for *e.g.* commercial RAID systems, the actual system performance in terms of *sustained* data rates as guaranteed by harddisk manufacturers is down by at least a factor of 3, transferring at maximum at a data rate of about 12 MB/s. Further investigation has shown that the solution of the storage problem needs the investment of a considerable amount of time into this problem in absence of expert support by the equipment manufacturers. Real-time standards are absent, data sheets vague or simply wrong and in one case it was actually found that the manufacturer published transfer rate performance data measured with a tool of unknown origin. We are by now convinced that the required data transfer rates can indeed be realized. Attaining the necessary performance however, may in the end come down to developing a new hardware interface board, including the development of a new PCI chipset to obtain the maximum PCI-bus performance, as well as the development of new drivers for industrial standards under Windows NT. This experience illustrates that there is still a long way to go in developing real-time data-acquisition systems with commercial-of-the-shelf products.

In conclusion it is clear that although ideal standards for data acquisition are not yet commercially available and universal libraries for hardware and software for data acquisition do not yet exist, it is still possible to build a cost-effective system out of commercial available components that does the job, provided that one keeps in mind what the consequences of this suboptimality are for the final system performance. Some of the functional requirements that have been outlined above may be realized with VMEbus systems, but at a much higher price. At the system software level the situation is more alarming: effective real-time operating systems, with a good GUI, are hard to find and real-time drivers for the boards that one wishes to combine are seldom available. Alternatively, the driver that one buys with a board appears to be 8-bit, whereas the ADC on the board delivers 12 bits. If maximum performance is required one should be prepared to invest time in the problem oneself or work together with an expert system integrator. At the moment it is the only way that will solve hard real-time data-acquisition problems.

In Fig. 7 an optional preprocessing process is shown. Although initially data will be analyzed and algorithms developed off-line, on a different computer platform, this option makes it possible to test fusion algorithms in real time by inserting extra processing capability. To this end SHARC DSP boards can be installed in the SFD that add scaleable processing power to the SFD.

Via the system software it is possible to get a first impression of the data during storage by selecting regions of interest in the radar stream by shifting a delayed time window relative to the transmitter pulse.

A standard magtape drive is used for backup of data from the RAID sytem. After completing the measurement the user writes selections of data to tape as *metafiles*, *i.e.* in an intermediate representation, which enables him to store these metafiles from the RAID system to tape using the Windows NT operating system and to read these metafiles back to a SUN workstation under Solaris for analysis and *vice versa.*

4.2 Model experiment

In defining a model experiment to test the functionality of fusion algorithms a number of possible relevant model experiments have been considered. Suitable experiments are such experiments that try to observe or detect targets that are hard to detect by individual single sensors and that one can hope for that sensor fusion could offer a solution. It should be noted that there should, at least theoretically, be a reasonable

probability of detection for the sensors that participate in the fusion process, because otherwise there is not much point in attempting the fusion, as we have seen in Sec. 3.2. The dissatisfactory performance of single sensors may be due to our inability to attach a (preferably scene-based) correct confidence level to the observation under a vast range of changing environmental conditions, or, alternatively, it may be due to resolution and visiting rate of a sensor in combination with an insufficient modeling of clutter, even under well-defined circumstances. In the first case we have essentially a scheduling problem: Which sensor in the suite is optimal given the circumstances? or Which sensor can cue the other sensor(s) given a particular observation task? In the second example effort must be invested in analyzing the detection steps and the validity of the clutter models used therein. Two experiments have been considered in some detail: Detection of buried anti-personnel landmines (APM) and the detection and identification of small sea targets in a wide variety of circumstances. Small targets means small patrol boats, dinghies, sea mines, the head of a frogman, buoys, debris, *etc.* Because of the still ongoing developments of constructing suitable sensors for mine detection, it has been decided to attack first the problem of detecting small targets at sea with the SFD.

It is expected that a combination of visible light and IR sensors together with radar must be capable of detecting small, floating objects. Sensor fusion is expected to play an important part in the detection and recognition of small sea targets because small sea objects

- have both a small optical cross section and a small RCS (radar cross section)
- are not permanently observable (this is especially relevant for scanning sensors), depending on their height and the sea state
- have almost the same temperature as the surrounding sea water
- have a low velocity relative to the water, so that effective clutter suppression becomes difficult
- must be detected in a sea clutter spectrum for radar and electro optics that is complicated because of its time variance and its spatial and temporal coherence.

Parallel to the construction of the SFD a research effort has been started to make an inventory on radar sea clutter models and their validity.

5 CONCLUSIONS

In this article we presented early sensor fusion as a possible way to demonstrate the advantages of sensor synergy. Early sensor fusion enables us to maximally benefit from sensor observations, but requires at the same time extensive data acquisition efforts. Sensor fusion is motivated by the expected qualitative and quantitative improvement of observations and thus of situation awareness. In order to test and verify these benefits experimentally in a quantitative way, a sensor fusion demonstrator testbed (SFD) has been designed and is presently under construction at TNO-FEL. The system architecture and its design philosophy have been outlined. The real-time constraints and the need to (re)use intermediate fusion results in different decision processes suggest that in early fusion soft decisions are more effective than hard decisions. Finally it can be concluded that from the theoretical point of view key concepts of fuzzy logic, such as fuzzy measures and Sugeno's fuzzy integral, provide us with suitable tools to combinevery different sensor streams with soft decisions that show the type of synergy behavior expected of a fusion process.

6 REFERENCES

[1] J. Llinas and E. Waltz, Multisensor Data Fusion, ArtechHouse, Norwood, MA. 1990.

[2] D. Hall and R. Linn, "A taxonomy of multisensor data fusion techniques", Proc. 1990 Joint data fusion symposium, vol 1, pp 593-610, 1990.

[3] C.B. Weaver, Ed., "Sensor Fusion", Proc. of SPIE, vol 931, Orlando, FL, 1988.

[4] P.S.Schenker, Ed., "Sensor Fusion, Spatial reasoning and scene interpretation", Proc. of SPIE, vol 1003, 1988.

[5] C.B. Weaver, "Sensor Fusion II", Proc. of SPIE, vol 1100, Orlando, FL, 1989.

[6] P.S.Schenker, "Sensor Fusion II, human and machine strategies", Proc. of SPIE, vol 1198, Philadelphia, PA, 1989.

[7] A.J. van der Wal, "Application of fuzzy logic control in industry", Fuzzy Sets Syst 74, pp 33-41, 1995.

[8] A.J. van der Wal, "Fuzzy Logic: Foundations and Industrial Applications", ed. D. Ruan, Kluwer, Chapter 14, 275-311 (1996).

[9] D.Ruan, Z. Liu, L. Van den Durpel, P.D'hondt, and A.J. van der Wal, "Progress of Fuzzy Logic control applications for the Belgian Nuclear reactor BR1", Proceedings EUFIT '96, Aachen vol2, 1237-1241 (1996).

[10]D. Ruan and A.J. van der Wal, "Controlling the output power of a nuclear reactor with fuzzy logic", proceedings JCIS 97, vol1, 136, 1997.

[11] G. J. Klir and B. Yuan, Fuzzy sets and fuzzy logic, Prentice Hall, New Jersey, 1995.

[12]A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", Ann. Math. Statistics, 38, pp 325-339,1967.

[13]G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, 1976.

[14]K. Leszczynski, P. Penczek, and W. Grochulskki, "Sugeno's fuzzy measures and fuzzy clustering", Fuzzy Sets Syst., vol 15 pp 147-158, 1985.

[15]D.H. Hall, Mathematical techniques in multisensor data fusion, Artech House, Boston, 1992.

[16]E.L. Waltz, "Data fusion for C3I: A tutorial", Command, control, communications and Intelligence handbook, EW Communications Inc, Palo Alto, CA , pp 217-226, 1986.

[17]P.J. Nahin and J.L. Pokoski, "NCTR plus sensor fusion equals IFFN", Proc. IEEE Trans. Aerospace Electronic Systems, Vol AES-16, pp 320-327, 1980.

[18]B.V. Dasarathy, "Fusion strategies for enhancing decision reliability in multisensor environments", Opt. Eng. 35 (3), pp 603-616, 1996.

[19]L.A. Zadeh ,"The linguistic approach and its application to decision analysis", in Y. C. Ho and S. K. Mitter, eds., Directions in large scale systems, Plenum Press, New York, pp 339-370, 1976.

[20]A.S. Mikhailov and A.Yu. Loskutov, Foundations of Synergetics II; Chaos and Noise, Springer, Berlin, 1996.

[21]E. Sviestins, multi-sensor tracking for air traffic control and air defense, ATC Systems, Jan, pp 10-16, 1995.

[22]T. Bayes, "Essay towards solving a problem in the doctrine of chances", Philos. Trans. Royal Soc. London, vol 53, pp 370-418, 1763.

[23]S.C.A. Thomopoulos, "Sensor integration and data fusion", J. Robotic, Syst. Vol 7(3), pp 337-372,1990.

[24]M. Sugeno, "Fuzzy measures and fuzzy integrals: A survey", in Fuzzy Automata and Decision Processes, North Holland, Amsterdam, pp 89-102, 1977.

[25]J. M. Keller, H. Qiu, and H. Tahani, "The fuzzy integral and image segmentation", Proc. NAFIPS June 1986, pp 324-338, 1986.

[26]M. Grabisch, "On the equivalence classes of fuzzy connectives-The case of fuzzy integrals", IEEE Trans Fuzzy Syst.3(1), pp 96-109, 1995.

[27]W.F. Pfeffer, Integrals and measures, Marcel Dekker, New York, 1977.

PAPER No: 31

DISCUSSOR'S NAME: E. Bossé

COMMENT/QUESTION:

- First observation: By trying to fuse data at the lowest possible level you are imposing a constraint of similarity among the sensors.

- Question: In the voting algorithm used for the fusion of n identical sensors, do you consider any weighting function to reject bad quality sensor information? If not, it may be dangerous to conclude that sensor fusion can deteriorate the end product. In any sensor fusion implementation, sensor information should be weighed according to its quality.

AUTHOR/PRESENTER'S REPLY:

- Yes, this is absolutely correct: we intend to demonstrate that sensor fusion is not always advantageous. For this reason, a simplified example of N (N odd) same copies of sensors is taken to make sure that the results are only dependent on the single-sensor possibility of detection (and of course on the decision method, in this case majority voting.

- Answer to the Question

No, we did not apply any weighing in the process, for the same sensors as formulated above. For the sake of simplicity we assume that the sensors are perfectly the same and no sensor is malfunctioning. Of course, it is always possible to take this type of behaviour into account, but it only complicates the argument. However, it can easily be seen that including mulfunctioning sensors will at best exclude these sensors from the fusion process and therefore the fusion process reduces again to an effectively smaller N - fusion, i.e. no improvement by adding more sensors. In case they are not rejected, the fusion process will only deteriorate even further. All the above conclusions are valid under the assumption of using positive weights only.

PAPER No: 31

DISCUSSOR'S NAME: P.M. Zanker

COMMENT/QUESTION:

What plans do you have to construct a test bed, and what form will this take?

AUTHOR/PRESENTER'S REPLY:

The testbed will primarily be concerned with designing the hardware on the level of sensor fusion. The initial stage will be concerned with storage of new data to facilitate design and testing of fusion algorithms (on different platforms .... SUN, PC etc.). Finally, the most suitable algorithms will be implemented in hardware preprocessing units that add scaleable (DSP) processing power to the testbed.

THE NEMESIS IDENTIFICATION DATA FUSION DEMONSTRATOR

P N. Griffith, L A Hooper, W M Everitt
The Defence Evaluation and Research Agency
St Andrews Rd, Malvern Worcestershire
WR14 3PS, England

(c) British Crown Copyright 1997 / DERA / MOD

1. SUMMARY

The NEMESIS programme at DERA Malvern was instituted to investigate aspects of multi-sensor, single node data fusion applied to the task of air target identification (ID) in the military environment. In recent years, a reduction of force numbers combined with the introduction of more capable weapon systems, such as precision guided weapons and beyond visual range missiles has made the requirement for accurate, timely, reliable and robust target identification a high priority for air defence forces. After much research during the last 25 years into ID technologies and techniques, it is generally conceded that no single identification technology or system can now nor will, in the foreseeable future, satisfy all the operational capabilities and performance requirements needed from air target identification systems. The enforced use of several identification sensors and sources implies the need for the fusion of the ID data derived from these different sensors and sources. Studies have indicated that automated ID fusion is likely to provide accuracy, consistency, timeliness and manpower benefits. The NEMESIS programme, sponsored by the UK MoD, has investigated the application of a dedicated NATO air target ID data fusion algorithm, NATO STANAG 4162 (see reference [1]), to a multi-sensor, single node SHOrt Range Air Defence (SHORAD) demonstrator system. This paper describes the demonstrator package, the sensors, the fusion process and the results.

2. INTRODUCTION

Although, by 1993, much theoretical development and simulation of STANAG 4162 effectiveness had occurred, the UK MoD required a greater insight into the practical difficulties of applying the Identification Data Combining Process (IDCP), as described in STANAG 4162, to a real air target tracking and ID system. The solution chosen used several sensor and source packages which reflected current practices in regard to aircraft ID, which were suitable (and available) for aggregation in a practical demonstrator, and which encompassed the wide range of different sensors and sources detailed in STANAG 4162 Annex D. The demonstrator brought together many techniques ranging from track correlation, ID association, fusion, Human Computer Interface displays (HCIs), to the handling of large quantities of air target library data and innovative visual identification processes. NEMESIS, once developed, was used during a later phase to gather target data for the extraction of performance metrics to gain insight to the perceived benefits from this form of ID fusion.

3. STANAG 4162

STANAG 4162 describes in detail the IDCP and the recommended manner in which it should be implemented. It has been in development for some years and the concept is now accepted by the major NATO nations who continue to support its further development. This specific method of Bayesian fusion is currently being employed for the ID function in at least one major NATO programme and one UK significant procurement programme. Other NATO nations are also known to be carrying out demonstrations with the concept. It is important that nations continue to employ the STANAG due to the obvious inter-operability benefits that will accrue from ID data exchange in the very difficult ID task.

4. DESCRIPTION OF THE NEMESIS DEMONSTRATOR

NEMESIS consists of a radar system supported by a range of sensors designed to use specific aircraft characteristics such as shape, emissions, position etc to identify air targets. The track forming systems for NEMESIS are an I band primary radar and a secondary surveillance radar deriving target tracks from aircraft transponders, where they are fitted. Fundamental to STANAG 4162 fusion principles is the requirement for a target track to exist prior to the attachment of any ID data. The ID sensor package was ambitious (in order to test as many as possible of the STANAG Annex D options) and incorporates a wide range of disparate sensors ranging from acoustic and IFF to thermal and visual identification. The conversion of source and sensor declarations to a common format suitable for algorithmic fusion was an involved and complex process. The radar, sensors and sources are distributed

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

between three cabins and two vehicles all linked by fibre optic data and communications' networks. The central data fusion cabin also provides the control functions for directed sensors and management of all library and analysis processes. The demonstrator handles a maximum of 50 air tracks and can provide real-time ID recommendations of target allegiance within a few seconds. To permit additional study to be performed in the absence of a live environment, a simulation capability is built into the demonstrator to generate the conditions that do not occur naturally in peace time. The HCI consists of two Sun workstations, configured for system management, display of the recognised air picture, sensor control and analysis. Access to all stages of the data fusion process are available through various displays. The NEMESIS system layout is shown in Figure 1.

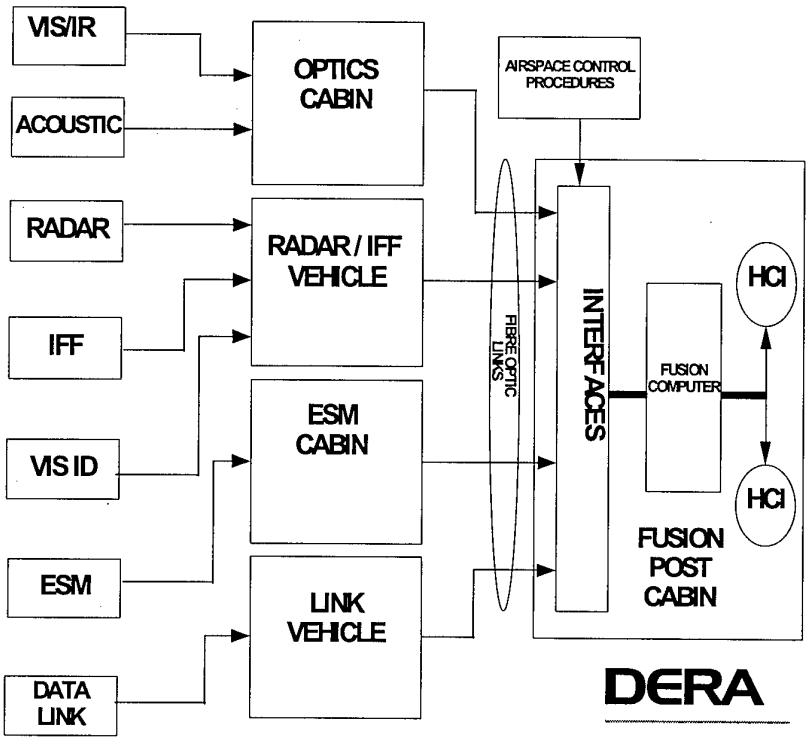


FIGURE 1: THE NEMESIS SYSTEM LAYOUT

5. THE SENSOR PACKAGE

The sensor and source package used for deriving aircraft allegiance consists of the following blend of co-operative and non co-operative identification systems:

- Identification Friend or Foe - this military standard system comprising Mark 10a and Mark 12 IFF equipments derives identification by a question and answer routine with co-operating aircraft, some aspects of which are secure.
- Electronic Support Measures (ESM) - this non co-operative identifier detects RF transmissions over a broad spectrum and classifies the target emitter(s) by reference to a look up table of library data. In some cases emitter ID can be mapped (see section 7) to platform type and allegiance.
- Visual identification - this auto-tracking dual (wide and narrow) field of view camera system, uses a man-in-the-loop to make judgements of target type based upon his training and experience of platform images. The allegiance of the target is then linked via library data to the platform type, and allowance is made in the mapping process for occasions where both own forces and enemy forces operate the same platform type.
- Thermal identification - This is a similar process to the visual function but operating in the 8-12 micron infra-red band with a Thermal Imager Common Module (TICM) system. This imager extends the platform identification function through the hours of darkness.
- Acoustic detection - This sensor, comprising a linear array of microphones, is dedicated to the

identification of helicopter platforms and provides platform type, range and bearing by comparison of a noise signature with a signature library.

- Airspace Control Procedures (ACPs) - With prior knowledge of flight corridors, safe lanes and mission plans, the tracks of aircraft crossing the area of interest can be compared with expected routes, positions and times and from their degree of compliance a judgement can be made regarding their allegiance.
- Data Link ID - with a secure data link established between friendly platforms and the NEMESIS demonstrator, a co-operative target's allegiance may be assumed with a high degree of confidence.

Four of the above systems (The IFF system, the visual and infra-red imaging devices and the ESM package) can all be tasked by the sensor management system to steer on to detected targets for narrow field of view active or passive interrogation, as appropriate. This requires sophisticated processing (initially a major hurdle), especially in the case of fast moving, close in targets, where a few seconds delay can result in the target never being in the field of view of the sensor. The data fusion block diagram is shown in Figure 2.

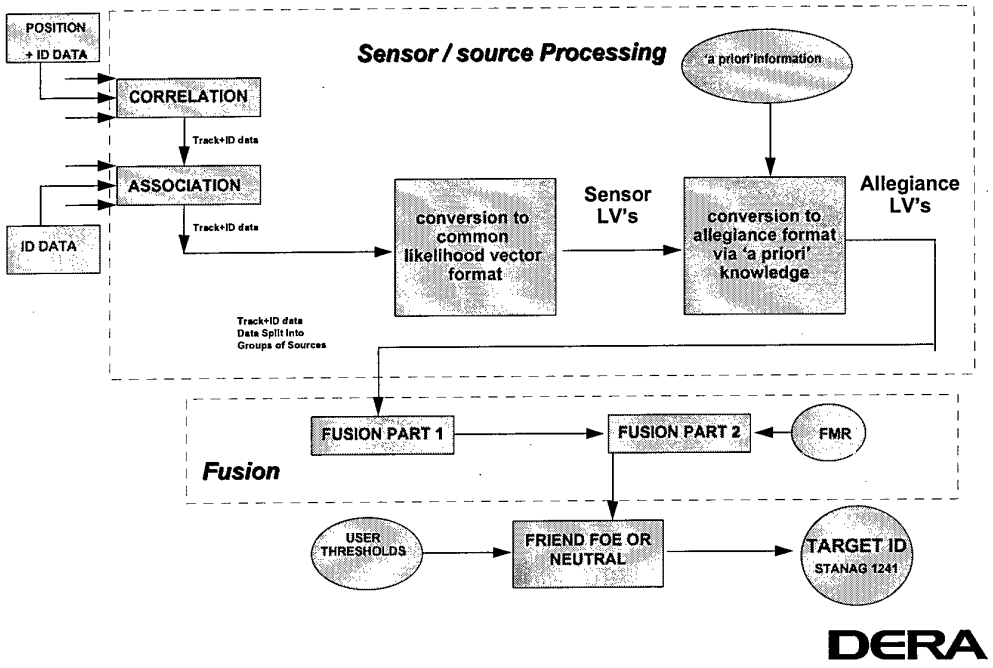


FIGURE 2: THE DATA FUSION BLOCK DIAGRAM

6. THE DATA FUSION PROCESS

The IDCP employs a Bayesian algorithm to fuse the ID data from the sensors and sources if they have been associated with a track. The algorithm, described later in this document, derives probabilities of hypotheses given an observation and an 'a priori' estimation of the observation being possible. The Identification Data Combining Process consists of two major sub-processes, sensor processing and fusion, both described in the paragraphs below and depicted in Figure 2.

7. SENSOR PROCESSING

Declarations on air targets from disparate sensors can take a number of very different forms. The declaration can range from an aircraft characteristic matching an entry in a database to a correct response from an interrogation from the ground. Before any identification data can be combined or fused the data has to take a common format. This is accomplished in two stages, by the use of a source probability matrix and a mapping matrix. The first stage in this process takes a sensor declaration and determines a level of confidence in its accuracy. An example of this stage is described as follows: Consider a Mode 3/A IFF sensor and its possible declarations of Evaluated Correct Response (ECR), Evaluated Incorrect Response (EIR), and Evaluated No Response (ENR). The sensor would normally interrogate a target and return either a no response or a four digit octal code. The code received is compared with a library and deemed to be a correct or an incorrect response. The causes for a particular response however are not conveyed (e.g. incorrect code entered in the transponder, transponder not working) and so the confidence in the declaration is lost. The source probability matrix contains information, derived from a system engineering

evaluation, that attaches a confidence level to a declaration and an example is shown in Figure 3.

	Fitted And Operating With Correct Code	Fitted And Operating With Incorrect Code	Fitted And Not Operating Or Not Fitted
ECR	0.90	0.01	0.01
EIR	0.01	0.70	0.01
ENR	0.09	0.29	0.98

FIGURE 3 : IFF MODE 3/A - TYPICAL SOURCE PROBABILITY MATRIX (SPM)

The system engineering evaluation needed for the population of these matrices is considered to be one of the more difficult tasks in implementing the IDCP and it should be noted that the values in this and the following figures are for descriptive purposes only. The matrix allows factors such as terrain screening, garbling and transponder reliability to be taken into account, building in a level of declaration probability. The matrix is used every time a declaration is made by the sensor. The likelihood vector is created by using the SPM row corresponding to the declaration made. A declaration of EIR would produce a likelihood vector as shown in Figure 4.

	Fitted And Operating With Correct Code	Fitted And Operating With Incorrect Code	Fitted And Not Operating Or Not Fitted
EIR	0.01	0.70	0.01

FIGURE 4 : IFF MODE 3/A - TYPICAL SOURCE DEPENDANT LIKELIHOOD VECTOR

This is the end of the first phase of data formatting but it should be noted that the likelihood vector is still sensor dependant. The second stage in this process converts the source dependant likelihood vector into a common format, which is an estimate of allegiance, (own forces (OF), non-aligned forces (NA) or enemy forces (EF)). This stage involves the combination of the source dependant likelihood vector and the mapping matrix by matrix multiplication. The mapping matrix contains more 'a priori' information expressing probabilities of an allegiance making a declaration, e.g. there is a high probability that own forces will make correct responses. This matrix also accommodates more unexpected events such as the probability of friendly aircraft having either no or defective IFF equipment. A typical mapping matrix is shown in Figure 5.

	OF	NA	EF
Fitted And Operating With Correct Code	0.70	0.20	0.10
Fitted And Operating With Incorrect Code	0.20	0.40	0.40
Fitted And Not Operating Or Not Fitted	0.25	0.15	0.60

FIGURE 5 : IFF MODE 3/A - TYPICAL MAPPING MATRIX (MM)

The combination of the source dependant likelihood vector and the mapping matrix results in a likelihood vector which is no longer source dependant and instead is a sources' estimate of allegiance. All source types, IFF, ESM, airspace control, for example, produce data in such a form and this data is eligible to be fused. An example of an allegiance likelihood vector is shown in Figure 6.

OF	NA	EF
0.70	0.20	0.10

FIGURE 6 : TYPICAL ALLEGIANCE LIKELIHOOD VECTOR

8. Data Fusion within the IDCP

The previous section describes the generation of a likelihood vector from a sensor declaration, and each source type produces likelihood vectors in this format ('Allegiance LVs' in Figure 2). This section describes the method of fusing data from the individual sensors.

'Fusion Part 1' in Figure 2 combines each likelihood vector from the different source types to produce a single likelihood vector for the target. This operation involves multiplication of similar matrices to produce the single vector. An example of such a target likelihood vector is shown in Figure 7.

OF	NA	EF
0.60	0.30	0.10

FIGURE 7 : TARGET LIKELIHOOD VECTOR

This single vector is passed to the second stage 'Fusion Part 2' in Figure 2, which incorporates an estimate of the force mix of aircraft in the local area, expressed as a normalised ratio of OF:NA:EF. An example of the format of this information is shown in Figure 8.

OF	NA	EF
0.80	0.10	0.10

FIGURE 8 : TYPICAL FORCE MIX RATIO

This information provides the prior probabilities required for Bayes theorem as shown below.

$$p(O_j|D_i) = \frac{p(D_i|O_j) \bullet p(O_j)}{\sum_{all\ j} p(D_i|O_j) \bullet p(O_j)}$$

← *Fusion of sensor and a priori data*

← *Normalising function*

where :

$p(Oj|Di)$ = *Final ID Likelihood Vector*

$p(Di|Oj)$ = *Target Likelihood Vector*

$p(Oj)$ = *Estimate of Force Mix*

Bayes' theorem fuses the target likelihood vector produced by stage one of fusion and the 'a priori' force mix ratio (FMR) to form a final normalised identification likelihood vector for the target being observed. An example of such a vector is shown in Figure 9.

OF	NA	EF
0.90	0.05	0.05

FIGURE 9 : FINAL ID LIKELIHOOD VECTOR

Throughout the above description of the IDCP process, the sensor declarations being used are assumed to indicate an allegiance. Sensors and sources such as IFF and ACP supply data from which the allegiance of a platform can be directly determined. Some sensors and sources however supply information from which allegiance cannot be directly determined. ESM and visual identification for example can supply platform information only. This data can not only be operationally valuable in its own right but can also be converted into an estimate of allegiance if it is known which nations operate certain platforms. Unfortunately with many nations often operating the same platform type, it can be difficult to obtain an estimate of allegiance with a high degree of confidence. In addition to the fusion of allegiance information, the IDCP accommodates this fusion of platform information. The same process is adopted for this

option with an output of a likelihood vector, estimating the platform type of the target being observed. Whether allegiance or platform information is fused the final result of the IDCP may not be in the correct form for display. This demonstrator uses the STANAG 1241 categories (friend, assumed friend, unknown, neutral, suspect, or hostile) for ID indication on the recognised air picture (RAP) and in order to arrive at these categories, the three state output (OF, EF, NA) from the IDCP must be thresholded.

9. CONFLICTS

Identification data from different individual sensors and sources on the same track can often be in contradiction. The IDCP has the ability to detect conflicting sensor information by testing the identification data at each combination stage. Conflicting identification data can arise for a number of reasons; mis-association of identification data to tracks being the most common. A malfunctioning sensor can also cause conflicting information and it is considered that if a conflict occurs the identification officer will resolve it manually by either breaking associations, inhibiting a sensor input or resetting the fusion process. There are currently no means by which identification data conflicts in the IDCP can be resolved automatically.

10. TRIALS PROGRAMME AND OPERATING PROCEDURES

NEMESIS was deployed to two operational UK airfields to gather and analyse live data. Targets of opportunity were designated friend, foe, and neutral to permit assessment of system performance. Air truth was derived from numerous sources but principally the air traffic control radar picture allied with IFF code and call sign data. The identification of targets with NEMESIS is a process initiated by the acquisition of radar tracks, at a range of 20-25 miles, and often confirmed (in the case of friends) with an IFF identification and an IFF derived track. Radar and IFF tracks are then correlated to provide the global track table to the fusion processor. The global track table prioritises these tracks in threat priority order, based upon range, speed and direction and commands the visual and infra-red sensors onto the correct azimuth for more detailed search. Lack of any target height information demands that the 'pointing' sensors then search in elevation upon that bearing. When a target comes within range of the optical and infra-red sensors, an auto-track routine takes over (with greater bearing accuracy and less latency than the radar) and the platform identification is performed by a man-in-the-loop who inputs data via a menu driven touch screen. In parallel with this reactive process, the ACP identification source compares the position, time and direction of the detected targets and makes a prediction of their intention to follow known pre-arranged procedural practices. During the whole phase the IFF continues to report on the track and the ESM and acoustic sensors continue to 'listen', and association routines attempt to tie any detected 'strobes' to known tracks. The link sensor also passes allegiance indications forward to the fusion processor. While these automatic processes continue, the fusion post operator monitors the evolving recognised air picture and assesses the allegiance recommendations made by the Bayesian fusion engine. Lack of ID information or conflicting data can sometimes be remedied by a command from the operator to the IFF, ESM, visual and thermal sensors to make repeated 'hits' upon the track to strengthen knowledge about the target. Graphical and tabular fusion displays meanwhile describe the evolving synergy of the ID picture. The output from the fusion process is an automatic classification of all targets in cover, updated every few seconds with a recommended STANAG 1241 target classification. These classifications derive from the friend, foe, neutral vectors output from the fusion process after thresholding. Thresholds are set reflecting any current state of tension.

11. DIFFICULTIES ENCOUNTERED

Two primary problems were encountered during the development and study phases of the demonstrator. The primary radar regularly formed false tracks and failed to maintain fast manoeuvring target tracks resulting in missed and erroneous ID associations. The second problem arose with high traffic densities when the processing times increased with the added workload causing excessive delays between the appearance of the target and the ID declaration stage. Additionally the characterisation of sensors, the development of matrices and establishing other 'a priori' data were non-trivial tasks and with hindsight, attendance at a full scale exercise would have provided more realistic 'enemy' targets, but logistics prevented this from happening.

12. CONCLUSIONS

The intention to practically implement the STANAG 4162 data fusion algorithm was satisfactorily met and the risks associated with the use of this form of identification data fusion are now better understood. Study of the quality of ID accuracy indicates that a few high quality sensors provide more accurate results than many low quality sensors. A sophisticated fusion algorithm obviously does little to improve the declarations made by poor quality sensors. Conflicts of identification that occur on any track when data from two sources differ are flagged to the operator as an automatic by-product of fusion. This is seen as an important benefit in alerting the identification officer who is able to intervene and correct any anomalies that occur. This function provides early alerting to an operator when a source or sensor becomes unreliable.

The major obvious benefit is that of the effect upon operator workload. An in-service system scaled to handle greater than 200 tracks and the equivalent of 5 sensor or source declarations every 10 seconds on each track for example, is a very demanding recognised air picture for any operator to assess in a timely fashion. This automatic classification technique enables the ID operator to manage his entire area of responsibility in a prompt and timely fashion and permits him more time to investigate thoroughly the tracks that are of special interest, (primarily the hostiles and unknowns).

13. FUTURE PROGRAMME

The NEMESIS programme to develop a STANAG 4162 real-time, single node, technical demonstrator is nominally complete and all the investigations are finished. The follow-on stage, which is in its infancy, will assess the implications and problems associated with networking up to four NEMESIS style nodes in a flexible architecture with simulated military data links. The programme will be laboratory based unlike the NEMESIS single node and will draw much of its data from the live environment. STANAG 4162 architecture studies are progressing and the performance of centralised, de-centralised and autonomous structures will be addressed during this programme.

14. REFERENCES

[1] NATO STANAG 4162 Draft Standardization Agreement, Technical Characteristics of the NIS Identification Data Combining Process, 1st September 1995, NATO Restricted.

PAPER NO: 32

DISCUSSOR'S NAME: M. Desbois

COMMENT/QUESTION:

The NC3A also conducted an IDCP trial with live data based on information recorded during Central Enterprise NATO exercise. The current STANAG 4162 states rules for war time, but it is very difficult to work with in peace/crisis time (typical exercise ....) where Modes 2,3 and 4 could be used simultaneously, and not all nations use them with the same rules.

Question: Did you find any problems of this kind during the NEMESIS experiment?

AUTHOR/PRESENTER'S REPLY:

This problem was not observed as, although NEMESIS was capable of Mode 1,2,3 and 4 interrogations, only Mode 3 replies were ever received.

NEMESIS implemented an early version of STANAG 4162 which did not appreciate peacetime IFF problems. This part of the STANAG is currently being reviewed and the new version will specifically address peacetime IFF use.

Communications Management In Battlespace Data Fusion

R.H. Deaves, D. Nicholson, P. Greenway & P. Vangasse

Advanced Information Processing Department
Sowerby Research Centre
British Aerospace
FPC 267, PO Box 5 Filton, Bristol
BS12 7QW
United Kingdom

Abstract

Algorithms for managing bandwidth-limited communications links in a multi-platform and multi-target battlespace data fusion scenario are investigated. The platform architecture is modelled on a fully connected decentralised system based on a simplified tactical air warfare scenario. Information filters on each platform maintain track estimates for each target. These are transmitted around the system in accord with a time-division-multiple-access communications protocol based on JTIDS. Here we address a practical limitation on the communications bandwidth: it is unable to support the full set of state information for all of the targets in each communications cycle. Strategies must therefore be found to manage the bandwith. In this paper we consider two different communications management algorithms: (i) a somewhat ad hoc cyclic, or round-robin, algorithm, and (ii) a more principled information theoretic algorithm. In general we find that the system has better tracking performance under the information theoretic algorithm than it does for the round-robin algorithm. This result appears to be fairly robust to uncertainties in a transmitting platform's prediction of the state estimates that are currently being maintained at the recipient platforms. In practice – since the information theoretic algorithm is more computationally costly to implement than the round-robin algorithm – the decision of whether or not to employ intelligent communications management is likely to depend upon the severity of the bandwidth constraint, the number and type of tracked targets, and the predictive power of a platform with regard to other platforms in the network.

1. Introduction

Multisensor data fusion is expected to play a crucial role in the digital battlespace of the 21st century [1]. Arguably one of the greatest challenges to be faced is the increasingly complex nature of battlespace scenarios, and the huge volumes of disparate data that these will generate. This may conceivably strain the sensing resources, computer processing speed and memory, and communications bandwidth to which a system has reasonable access. Algorithms for resource management are therefore likely to be an important element of future battlespace data fusion systems.

A number of authors have investigated system resource management in multisensor networks. However, this effort has almost entirely been focussed on sensor management [2, 3, 4]. It has led to a principled *information theoretic* approach to sensing resource allocation. In contrast, there has been relatively little published work that provides a systematic investigation into the issues that are involved in managing a limited communications bandwidth resource in multisensor systems. This is somewhat surprising given that communications issues may be critical in military applications. Aircraft on an intercept mission may, for example, wish to limit their usage of communications bandwidth in order to achieve some measure of stealth. Alternatively, reduced bandwidth can arise as a result of signal jamming by the enemy aircraft.

Our previous work on resource management in multisensor architectures has considered the relative performance of two different approaches to communications management in decentralised identity fusion [5] and tracking [6] systems. One approach was ad hoc but computationally cheap, the other approach was firmly rooted in information theory but involved more calculation. Process models were developed to indicate the conditions under which the information theoretic approach was likely to be more desirable than the simple alternative. This current work is intended to revisit the process model for the decentralised tracking system, in the light of more realistic simulations that permit us to explore the robustness of the information theoretic approach to modelling inaccuracies.

The paper is organised as follows. Section 2 covers some of the relevant background in multisensor systems architectures, decentralised filtering, and information theory. Section 3 outlines a computer simulation of a simple decentralised tracking system. Section 4 describes the communications management algorithms in some detail. The simulated battlespace scenario and numerical performance diagnostics are described in Section 5. Section 6 presents our main results and these are discussed in Section 7. Finally, Section 8 concludes and summarises the paper.

2. Background

This section presents some brief background material

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

on the key concepts – decentralised systems, state estimation, and information theory – that underpin the work described here. Further details can be found in the cited references.

2.1 Decentralised Systems

Centralised multisensor systems are characterised by having a central processing node through which all of the data from around the distributed sensing network is channeled. This architecture can result in catastrophic failure of an aircraft intercept mission, for example, if the inter-platform communications bandwidth is reduced or lost – a highly probable outcome in battlespace scenarios. Successive levels of decentralisation add some measure of fault tolerance to a multisensor system. The extreme case is a fully decentralised architecture with no central facility of any kind. This is the architecture that we have investigated in our previous work [5, 6], and is also the focus of this present investigation. In further references to decentralised systems in this paper, full decentralisation with full network connectivity should be assumed. Fig. 1 illustrates a range of different multisensor system architectures.

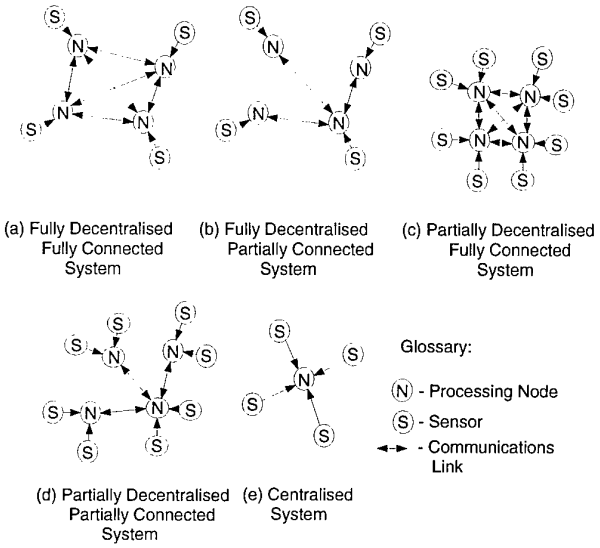


Figure 1: Multisensor System Architectures

In addition to being resilient, decentralised architectures offer a number of other features that are desirable for battlespace systems. These include modularity, scalability, and flexibility [7]. It can be shown that decentralised systems require less communication bandwidth than their centralised counterparts to obtain identical fused identity or track estimates. This advantage is offset by an increase in the number of communications links required.

2.2 State Estimation

The most common state estimator is the Kalman filter, which is proven to satisfy a number of optimality criteria under certain conditions [8]. A variant of the Kalman filter, known as the information filter, is utilised here. This is expressed in terms of the information state variables, $\mathbf{y}$ and $\mathbf{Y}$, that are related to the standard Kalman state-space variables, $\mathbf{x}$ and $\mathbf{P}$, as follows:

$$\mathbf{Y}(i|j) = \mathbf{P}^{-1}(i|j); \qquad \mathbf{y}(i|j) = \mathbf{P}^{-1}(i|j)\mathbf{x}(i|j) \quad (1)$$

where $(i|j)$ refers to an estimate at time step i based on all the observations up to time step j (where $j \leq i$). The information filtering equations are straightforward to decentralise, with local measurement update and global data assimilation terms being much less complex than those for the Kalman filter [7].

2.3 Information Theory

The information content, I, of a multivariate probability density function, $f(\mathbf{x})$, is defined as:

$$I = \langle \ln f(\mathbf{x}) \rangle \quad (2)$$

where $\langle \cdot \rangle$ denotes statistical expectation. In tracking applications, $\mathbf{x}$ refers to the state vector of a target and the uncertainty in the components of $\mathbf{x}$ is usually described by a multivariate Gaussian distribution. This can be expressed as follows:

$$f(\mathbf{x}; \bar{\mathbf{x}}, \mathbf{P}) = \frac{1}{(2\pi)^{l/2}|\mathbf{P}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})\mathbf{P}^{-1}(\mathbf{x}-\bar{\mathbf{x}})^T\right] \quad (3)$$

where $\bar{\mathbf{x}}$ is the mean value of the state vector, $\mathbf{P}$ is the state error covariance matrix ($|\mathbf{P}|$ is the determinant of $\mathbf{P}$), and l is the dimension of the state vector. The information relating to a target's track is then given by:

$$I_t = -\frac{1}{2}\ln\left[(2\pi e)^l |\mathbf{P}|\right] \quad (4)$$

or equivalently by:

$$I_t = \frac{1}{2}\ln\left[(2\pi e)^{-l} |\mathbf{Y}|\right] \quad (5)$$

where t is the identity index of the target. In fact later we shall only be interested in comparing relative information values between the targets state estimates and not in the absolute information values for individual targets track estimates. This means that the constant term in the expansion of the logarithm drops out. Also, since the logarithm of $|\mathbf{Y}|$ is monotonically increasing with $|\mathbf{Y}|$ we are able to replace $\ln|\mathbf{Y}|$ simply with $|\mathbf{Y}|$. Therefore let us define the *comparative information value* to be:

$$I'_t = |\mathbf{Y}| \quad (6)$$

Hence, the track information at a given node (i.e. a sensor and its processer) can by defined as:

$$I_n = \sum_{\forall t} I_t' \tag{7}$$

where n is the index of the node. Similarly, the system information, I^S, is given by:

$$I^S = \sum_{\forall n} I_n \tag{8}$$

where S indicates system information.

3. Simulation Details

A computer simulation was developed to investigate communications management in a simple decentralised tracking system. This paper only addresses the effect that communications management has upon one aspect of low-level data fusion, namely the tracking of multiple targets from multiple platforms. In this section the components of the simulator are described. Our generic model of a decentralised system – developed in conjunction with Hugh Durrant-Whyte of the University of Sydney – is illustrated in Fig. 2. Although situation and threat assessment are shown in this diagram, their inclusion in the simulator is deferred to a future investigation. A further sophistication that we have also chosen not to include in the simulator at this stage is a feedback loop that allows targets to respond *reactively* to the behaviour of the platforms, see A in Fig. 2. In addition, the communications model is ideal in the sense that all transmitted data are received successfully, see B in Fig. 2.

3.1 World Generator

The world is populated with N_T targets of three basic types. One type is meant to be representative of a bomber aircraft, another is representative of a fighter aircraft, and the third type is representative of a high performance (advanced) fighter aircraft. All of the targets execute a single turning manoeuvre during the simulation interval. The bombers turn with a maximum acceleration of Bg, the fighters at Fg, and the advanced fighter at Ag. Here, A > F > B which infers that the advanced fighter is the most and the bomber the least manoeuvrable targets.

3.2 Platforms

The world is observed from three platforms which may, in a tactical air warfare scenario, represent an Airborne Warning and Control System (AWACS) aircraft and a lead/wingman pair of fighter aircraft. The platforms form a fully connected system and it is assumed that the targets are visible from all three platforms at all times.

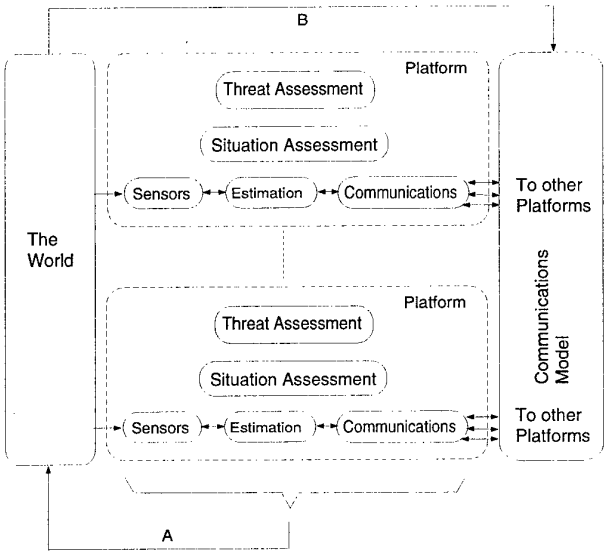


Figure 2: Components in a Decentralised Tracking System

3.3 Sensors

Each platform is equipped with a single sensor. The sensors measure the x and y range of the targets in 2d Cartesian space. These measurements are synchronised at the platforms and are updated at constant time intervals. All of the sensors have a fixed observation noise covariance $\mathbf{R}$. The noise in the individual sensors is independent in the x and y range measurements. The noise is also independent between the different sensors.

An important feature of these simulations is that we have fixed the probability of target detection, P_d, at the sensors to a value that is less than unity. This introduces some uncertainty into a transmitting platforms prediction about the track state error covariances that are being maintained at the recipient platforms. The reason is that at any given time a platform will not know whether the other platforms have based their current state estimate for a target's track on a new observation, or, in the absence of detecting a target, simply predicted forward the previous state estimate for the corresponding track. The implications of this for communications management are discussed later.

3.4 Estimation

The targets are tracked with information filters. These assume a constant velocity process model for the targets with a process noise variance tuned to their accelerations [9]. Local state estimation is performed separately at each of the platforms and the local estimates are fused into global estimates following a data assimilation step after each communication cycle. In order to avoid tricky data association problems, which

are not really the issue here, measurements are correctly assigned to tracks from within the simulation software. Local and global estimation does not therefore suffer from incorrect data association here, although this would be an issue in a real system.

3.5 Communications

The communications protocol that is followed in the simulator is a time-division-multiple-access system, based on JTIDS [10]. All the communication links are assumed to be maintained throughout the simulation and to have no latency. Each platform in the network is designated a time-slot in which it sends a burst of formatted data. Here we consider the data to comprise of state estimates and associated error covariances for up to N_T targets, depending on the severity of the bandwidth constraint. In decentralised systems one must be careful to handle the problem of multiple data counts. These can lead to over estimated filters, which are clearly undesirable in mission critical systems. In fully connected decentralised systems the problem is solved by maintaining an additional filter, known as the channel filter [11]. The communications module in our software is responsible for updating the channel filters and implementing the communications management algorithms. The latter component of the module is described in more detail in Section 4.

3.6 Implementation Details

The simulator has been coded in the 'C' programming language with results displayed from the `Tcl/Tk` scripting language. Simulations are carried-out by running the code on a Sun Ultra 140E networked computer under the Unix operating system. Results are stored to file for off-line analysis.

The two communications management algorithms are run concurrently during the simulation. This presents the same data set to each algorithm during a particular simulation. Hence, a reasonable comparison of performance can be carried-out on a single simulation. This is useful for demonstration purposes.

4. Communications Management

We now turn to the main feature of our investigation: the algorithms for managing limited communications resources in a decentralised tracking system. Here the following somewhat arbitrary restriction is placed on the communications bandwidth: it is sufficient to support the transmission of state information (estimation and covariance) for only N_T' out of N_T targets in each communications cycle. The management problem is basically to decide which are the most appropriate targets on which to communicate track data. We have taken two approaches to this problem. These are described below.

4.1 Round-Robin Algorithm

This is a computationally cheap and *ad hoc* algorithm. The transmitting platform essentially selects the targets on which to communicate state information in a wraparound cyclic, or round-robin, manner. The round-robin algorithm is initalised randomly on each platform by shuffling the N_T indices that label the targets and then transmitting data on the targets that correspond to the first N_T' indices in the list. Consider a scenario in which $N_T = 5$ and $N_T' = 3$. Suppose the initialisation process at one of the platforms selects targets T_1, T_3, and T_5 for the first communication cycle. The next time that this platform communicates it will select targets T_2, T_4, and T_1. This is followed by targets T_3, T_5, T_2, and so on until the simulation is complete.

4.2 The Information Theoretic Algorithm

This is a principled approach to communications management based on information theory. It is computationally more expensive than the round-robin algorithm to implement. The fundamental idea is that a transmitting node communicates data corresponding to the N_T' targets that will produce the maximal overall change in information, in a single communications cycle, at the receiving platforms.

The mathematical notation for this communications management algorithm is somewhat involved. Hence, for clarity and simplicity we provide a worded description of its behaviour. The algorithm begins with the transmitting node estimating the total system information before a communication has been made. This results in a single information value being generated for each iteration of the algorithm. This is referred to as the *prior system information (estimate)* given by:

$$\begin{array}{c} \textit{prior} \\ \textit{system} \\ \textit{information} \\ (\textit{estimate}) \end{array} = \begin{array}{c} \textit{transmitter} \\ \textit{information} \\ (\textit{real}) \end{array} + \begin{array}{c} \textit{prior} \\ \textit{receivers} \\ \textit{information} \\ (\textit{predicted}) \end{array} \tag{9}$$

In the second stage of the algorithm the transmitting node estimates the total system information after a communication has been made on each target, *i.e.* N_T values in all. This is referred to as the *posterior system information (estimate)* and is given by:

$$\begin{array}{c} \textit{posterior} \\ \textit{system} \\ \textit{information} \\ (\textit{estimate}) \end{array} = \begin{array}{c} \textit{transmitter} \\ \textit{information} \\ (\textit{real}) \end{array} + \begin{array}{c} \textit{posterior} \\ \textit{receivers} \\ \textit{information} \\ (\textit{predicted}) \end{array} \tag{10}$$

The third stage of the algorithm is to produce a list, comprising N_T elements, of the change in information

value that will occur on transmission of data relating to a particular target. These are referred to as *decision values* and are calculated as:

$$\begin{array}{c} \textit{decision} \\ \textit{value} \end{array} = \left| \begin{array}{c} \textit{posterior} \\ \textit{system} \\ \textit{information} \\ (\textit{estimate}) \end{array} - \begin{array}{c} \textit{prior} \\ \textit{system} \\ \textit{information} \\ (\textit{estimate}) \end{array} \right| \tag{11}$$

The final stage of this algorithm is to sort the *decision values* into descending order. Data on the first N'_T indexed targets in the list are then transmitted to the other nodes.

This information theoretic algorithm implies that the transmitting platform has knowledge of the information state at the receiving platforms. In our previous work [5] this was the case, effectively since the platforms and sensors were identical and $P_d = 1$. Clearly this was a somewhat unrealistic assumption and we have relaxed it artifically by introducing noise to the *decision values* before they are sorted into descending order. This permits us to investigate the robustness of the information theoretic communications management algorithm to inaccuracies in the decision process.

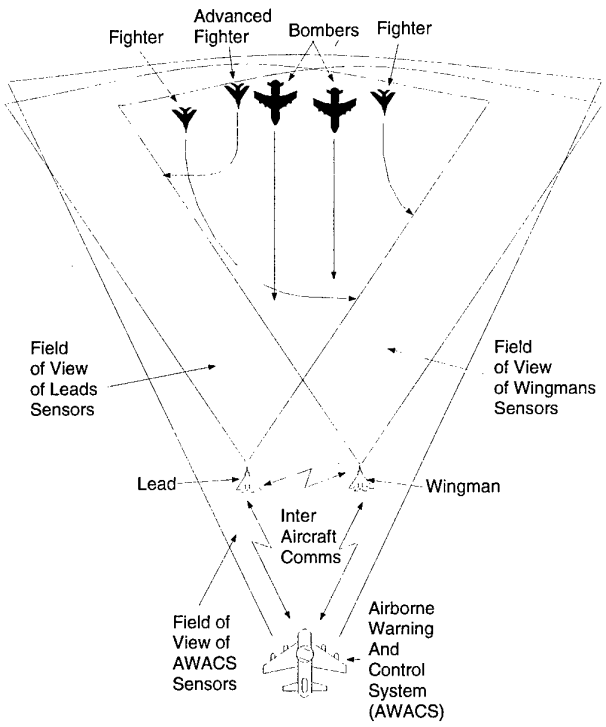


Figure 3: The Multi-Platform and Multi-Target Tracking Scenario

5. Investigation Process

This section of the paper describes the battlespace scenario, inter-platform communications scenario, and the statistical analysis methods that were applied to the simulation results.

5.1 Battlespace Scenario

The scenario on which the computer simulations were executed is illustrated in Fig. 3. We have somewhat arbitrarily chosen values of $N_T = 5$ targets (shown as black aircraft in the diagram) and a decentralised system comprising 3 platforms (white aircraft). This scenario is modelled on a simplistic air warfare scenario in which 'possible hostile' aircraft are intercepted by 'friendly' aircraft comprising of a lead/wingman pair in communication contact with an Airborne Warning And Control System (AWACS) aircraft. The scenario is not intended to reflect a true 'intercept' mission. However, a number of realistic elements are evident, these include: multi-aircraft friendly force, multi-aircraft hostile force, imperfect sensors, and bandwidth constrained data communication channels.

The 'possible hostile' aircraft comprise two large bomber aircraft performing turning manoeuvres with an acceleration of Bg. These aircraft are flanked by two fighter aircraft and one advanced fighter aircraft that are able to perform higher turning manoeuvres of Fg and Ag respectively. All the aircraft fly with slightly different speeds. Their mission begins with all the aircraft flying in parallel. However, on detecting that they are being observed, the fighter aircraft perform maximum acceleration manoevers for turning angles of approximately $+60°$, $-90°$ and $+45°$. During this time the bomber aircraft maintain their original paths. The trajectories for all the 'possible hostile' aircraft are represented (approximately) by the associated arrowed lines in Fig. 3.

The 'friendly' aircraft maintain parallel flight trajectories through the portion of the intecept mission that is simulated here. The lead/wingman pair fly ahead of the AWACS and towards the 'possible hostile' aircraft. During the mission it is assumed that *all* of the 'possible hostile' aircraft are in the field of views (FOVs) of *all* the 'friendly' aircraft *all* of the time. Further, it is assumed that the sensor information generated by all three 'friendly' aircraft are comparable. For simplicity of simulation the assumption is made that sensor clutter levels and distance between the 'possible hostile' aircraft are such that near perfect data association is maintained. This also implies that the number of spurious tracks that are generated is negligible.

5.2 Communications Scenario

For the work presented in this paper we are concerned with a performance comparison of the communications management algorithms as the precision of the *decision values* generated by the information theoretic approach is varied. This is achieved by introducing a

percentage error in the decision values before deciding which data to transmit. Arbitrary values for these errors of of 10, 50 and 90 % were selected. These errors are introduced as follows. First, a sum of the decision values is calculated. This allows the percent error in decision values to be represented as an *information error in decision values*, e.g. 10% of the sum of the decision values. An equal portion of this information error is then subtracted from each decision value. The information error in decision values is then randomly redistributed over the decision values. For each value of percent error in the decision values, simulations are performed with the bandwidth constrained to be a percentage of the full bandwidth that is required, *i.e.* $N_T'/N_T \times 100$, as N_T' is varied from 0 to 5. This gives values of 0, 20, 40, 60, 80 and 100%.

5.3 Analysis

Although the tracking and communications management algorithms are employed in information space, it is more meaningful to produce the results in standard space state. This does not affect the advantages of implementing the information form of the Kalman filter as discussed in Section 2.2, since the analysis is performed off-line. Hence, the performance of the two communications management algorithms are considered in the context of the determinant of the covariances produced by the tracking algorithm, $\mathbf{P}$. In general, this quantity provides a measure of the volume of the uncertainty ellipsoid associated with the track estimate.

The results are presented from the view point that the 'friendly' aircraft have a system requirement to track the targets with a covariance that should, on average, not exceed $\mathbf{P}_{req}$. This value is set during system specification. Hence, the results are presented using the following average (ave) and standard deviation (σ) values for each target:

$$\text{ave}\left(\frac{|\mathbf{P}_{xx}|}{|\mathbf{P}_{req}|}\right) = \frac{1}{N_S}\sum_{\forall m}\sum_{\forall k}\sum_{\forall n}\frac{|\mathbf{P}_{xx}^{n,k,m}|}{|\mathbf{P}_{req}|} \quad (12)$$

$$\sigma\left(\frac{|\mathbf{P}_{xx}|}{|\mathbf{P}_{req}|}\right) = \left[\frac{1}{N_S}\sum_{\forall m}\sum_{\forall k}\sum_{\forall n}\left(\frac{|\mathbf{P}_{xx}^{n,k,m}|}{|\mathbf{P}_{req}|}\right)^2 - \left(\frac{1}{N_S}\sum_{\forall m}\sum_{\forall k}\sum_{\forall n}\frac{|\mathbf{P}_{xx}^{n,k,m}|}{|\mathbf{P}_{req}|}\right)^2\right]^{\frac{1}{2}} \quad (13)$$

where:

- $|\mathbf{P}|$ - Is the determinant of the state error covariance on a track
- xx - Denotes the round-robin (RR) or information based (IN) communications management algorithm.
- N_M - The number of Monte Carlo simulations carried-out, i.e. 10 for the work presented in this paper.
- m - The index of a particular simulation i.e. $m \in \{1, 2, \ldots, N_M\}$.
- N_K - The number of discrete time intervals in a simulation (after the track has been initialised), i.e. 55 for the work presented here.
- k - The index of a particular discrete time interval, i.e. $k \in \{1, 2, \ldots, N_K\}$.
- N_N - The number of nodes or platforms used in the simulation, i.e. 3 for the work presented in this paper.
- n - The index of a particular node or platform, i.e. $n \in \{1, 2, \ldots, N_N\}$.
- N_S - Number of sample points, where $N_S = N_M + N_K + N$.

These average and standard deviation values are used to statistically summarise the results of this investigation.

6. Results

The main results are presented in this section. Fig. 4 provides results for three different percent error in decision values of 10, 50 and 90 % in plots (a), (b) and (c) respectively. Here the average fractional covariance (from Equation 12) is plotted (on a logarithmic scale) against the percentage of full bandwidth that the nodal communications resource can support. Coincidental plots for one of the fighters, one of the bombers, and the advanced fighter, for both the communications management algorithms, round robin (RR) and information based (IN), are provided on each of the figures, (a) to (c). The corresponding variance plots are provided in Fig. 5 with y axes ranging from 0.001 to 10. We refer to these plots as *process models* which relate the performance of the communications management algorithms to the percentage of full bandwidth employed for inter-platform communications.

7. Discussion

This section discusses the results in more detail. In addition, a qualitative comparison of the results are described with a specific application to battlespace scenarios.

Two process model metrics describe the performance of the communications management algorithms. They are the average fractional covariance volume, $|\mathbf{P}_{xx}|/|\mathbf{P}_{req}|$, and its associated standard deviation. The latter is comprised of a very small

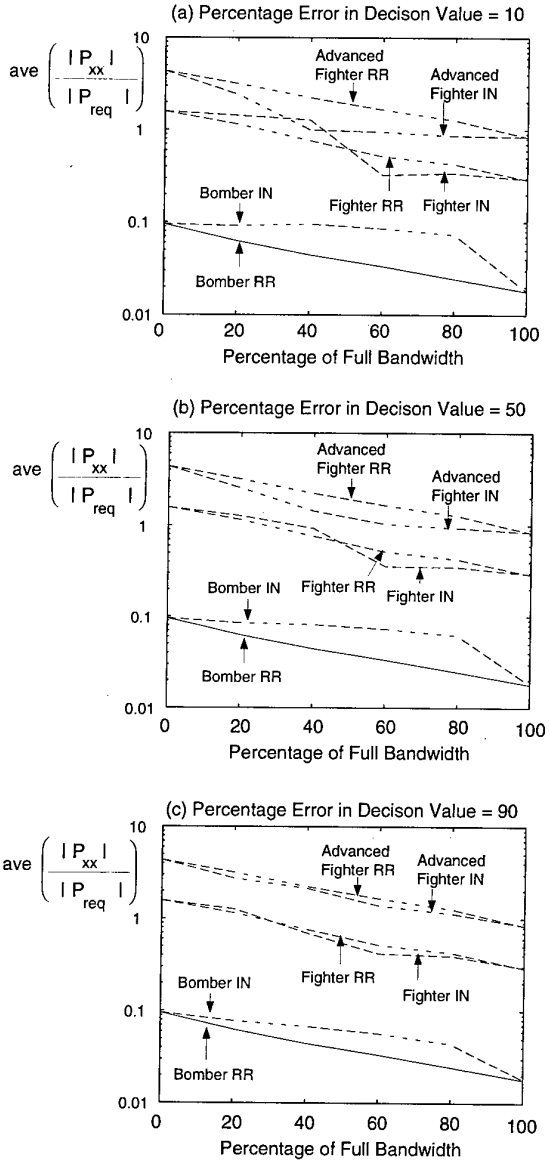


Figure 4: Average Results (10 Monte Carlo 'runs')

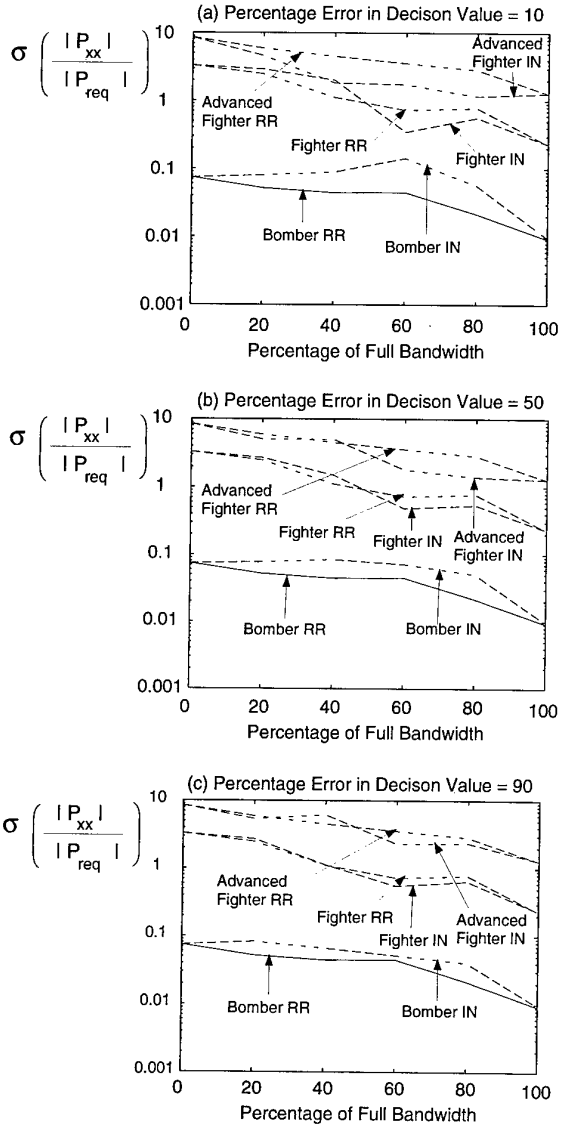


Figure 5: Standard Deviation Results (10 Monte Carlo 'runs')

random component (introduced by having the sensor probability of detection, P_d, in the simulations set to a value less than unity) and additionally, for the round-robin algorithm, by a random ordering of the targets. It is also comprised of a systematic component. This is a function of the time-division-multiple-access communications protocol in which each node follows a staggered sequence of . . .transmit-receive-receive. . ..

Note that for the round robin and information based communications management algorithms, the average and standard deviation metrics exhibit a greater difference margin for lower percent error in decision values. In fact we predict that as the percent error in decision value approaches 100%, the statistical performance of each algorithm would be similar. In addition, on an individual target basis, the information based algorithm performs better than the round robin algorithm on the advanced fighter target, while the situation is reversed for the bomber aircraft. A comparison of the curves for the fighter aircraft shows that the relative performance of the communications management algorithms is sensitive to the percentage of the full bandwidth that is being employed. This is a consequence of the information based communications management algorithm allocating more overall bandwidth to the targets that are difficult to track.

We now consider the application of these process models to a system design example. Let us assume that the scenario described in Section 5.2 has a specification requirement that all targets should have an average covariance volume of less than $|\mathbf{P}_{req}|$ with an inter-platform bandwidth of 60% that required to communicate all the available data. For this example, the system design decision is to determine what communications management algorithm to employ.

A sample of possible solutions have been generated from the results presented in Section 6 and illustrated in Fig. 4. Here the average performance metric (shown in Fig. 6) associated with the advanced fighter is considered since its value is larger than that for the bomber and fighter aircraft, in a range around 60% of full bandwidth. From the results generated in our investigation two possible communications management options are available: using the round robin algorithm, or using the information based algorithm. The latter algorithm was subject to 10, 50 or 90 percent errors in the decision values. Since, the system specification has to meet the required covariance, $\mathbf{P}_{req}$, a value of $|\mathbf{P}_{xx}|/|\mathbf{P}_{req}| > 1$ indicates when the specification is not met and a value of $|\mathbf{P}_{xx}|/|\mathbf{P}_{req}| \leq 1$ indicates when the specification has been met. Here, the round robin algorithm does not meet the specification producing a volume of uncertainty that is approximately twice the requirement, see Fig. 6. Likewise the information based algorithm with a 90 percent error in decision value also fails, although not so badly as the round robin algorithm. The information based algorithm with a 50 percent error in decision value achieves the specification. This would probably be chosen in preference to the 10 percent error in decision value implementation as it would be less computationally expensive, requiring a less rigorous algorithm for predicting the receiver's information (see equations 9 and 10). This would release valuable computation for other processes[1].

This is an example of applying the communications management process models developed during our investigation to meeting specification requirements while preserving a reasonable computational requirement.

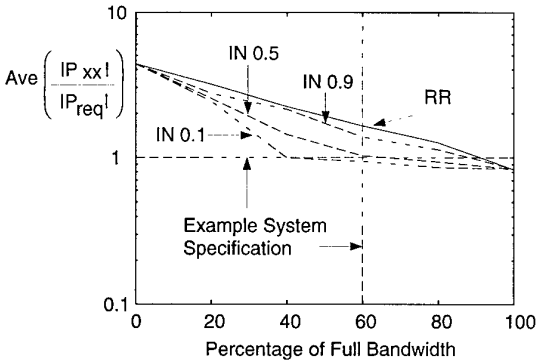


Figure 6: Application Example (Communications bandwidth = 60 % of Full Bandwidth, Performance requirement = $\mathbf{P}_{req}$).

8. Conclusions and Summary

Algorithms for managing a constrained communications resource in a decentralised target tracking system have been investigated. In a simulated multi-target and multi-platform battlespace scenario, the inter-platform communications bandwidth was varied between two extremes. At one end of the spectrum the bandwidth was able to support state information transmission on all the targets at every communication cycle. At the other end of the spectrum it was unable to support the transmission of any information about the states of the targets. Results were also obtained for a range of intermediate bandwidth values. We presented our results in a form that is useful in terms of gauging whether the management algorithms allow the system to meet some arbitrary system requirement. The results indicate that a system which implements the information theoretic communications management algorithm is better (in terms of producing smaller average state error ellipsoids around its tracks) than a system which employs the round-robin algorithm. The degree of improvement depends on the

[1] Alternatively, the *more cautious designer* may decide to implement the 10 percent error in decision value algorithm allowing a larger *margin of error* between the specification requirement and that given by the process model.

percentage of the full bandwidth that is being used, and is more evident for the highly manoeuvring targets. The decision on whether or not to employ the information theoretic communications management algorithm will depend on a number of factors. The critical factors are likely to be the amount of bandwidth that is available, the computing resources on the platforms, the nature of the battlespace scenario (particularly the number and the mix of target types), and the predicitve power of platforms in the network relative to other platforms.

Acknowledgements

We are grateful to Hugh Durrant-Whyte (University of Sydney), Colin Noonan, Mike Everett and Glenn Gapper (all British Aerospace) for their useful contributions to this work.

References

1. E. Waltz and J. Llinas. *Multisensor Data Fusion.* Artech House, 1990.

2. J.M. Nash. Optimal allocation of tracking resources. *IEEE Int. Conf. Decision and Control*, pages 1177–1180, 1977.

3. K.J. Hintz and E.S. McVey. Multi-process constrained estimation. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(1), January 1991.

4. K. Kastella. Discrimination gain to optimize detection and classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 27(1), January 1997.

5. R.H. Deaves, M. Bridges, P. Greenway, D.R. Bull, and D. Nicholson. Evaluation of communications management in a simulated decentralised identity fusion system. *SPIE 11th Annual International Symposium on Aerospace/Defence Sensing, Simulation, and Control*, 3068(08), April 1997.

6. R.H. Deaves, D. Nicholson, P. Greenway, and D.R. Bull. Evaluation of communications management in a simulated decentralised tracking system. *SPIE 11th Annual International Symposium on Aerospace/Defence Sensing, Simulation, and Control*, 3068(30), April 1997.

7. J. Manyika and H.F. Durrant-Whyte. *Decentralised Sensor Fusion: An Information Theoretic Approach.* Kluwer, 1994.

8. P.S. Maybeck. *Stochastic Models, Estimation, and Control.* Academic Press, Vol 1, 1979.

9. Y. Bar-Shalom and X. Li. *Estimation and Tracking: Principles, Techniques, and Software.* Artech House, 1993.

10. J. Toone. *Introduction to JTIDS.* Signal p. 55, 1978.

11. S. Harper-Grime. *Communications Within Decentralised Sensing Systems.* PhD thesis, Oxford University, 1993.

NATO Alliance Ground Surveillance Interoperability

Peter J. Lenk
Gert Retzer
NATO Consultation, Command and Control Agency, The Hague
PO Box 174
2501 CD The Hague
The Netherlands

1. SUMMARY

Through the NATO Alliance Ground Surveillance (AGS) Project, NATO is proposing to procure an airborne ground surveillance system based on a NATO-owned and operated core capability which would be supplemented by inter-operable national assets, as required. The NATO Consultation, Command and Control (C3) Agency (NC3A), together with support from six nations, France, Germany, Italy, Norway, The United Kingdom and The United States, has established a testbed capability which includes simulations of the candidate sensor systems and various exploitation capabilities. This testbed, coupled with various other command and control testbeds resident at NC3A, allows the study of many aspects of command and control associated with the AGS problem.

Essential to the selected ownership option is the concept of the interoperability of the NATO-owned assets with each other and with the nationally-owned assets. In order to address this problem, the NC3A, in conjunction with the six nations has developed an architecture and a standard format for the sensor data, so that any ground station could be made to receive and process data from any of the candidate platforms. Important to the solution is the ability to simultaneously receive and process data from different sensors and sensor platforms allowing for the fusion of data and potentially leading to improved surveillance coverage, tracking performance and accuracy. This approach has been prototyped and tested in the laboratory utilizing the AGS testbed. Recently, a large experiment was conducted during which several real AGS platforms and ground stations were utilized in order to prove the interoperability concepts that have been developed at NC3A in a more practical environment.

In order to set the stage, the paper begins with a brief overview of the NATO AGS project followed by a definition of the interoperability problem. Next, the hardware and data architecture developed at NC3A for achieving interoperability is described. Finally, the paper concludes with a description of a recent AGS interoperability experiment used to prove the concepts is presented.

2. NATO AGS PROJECT BACKGROUND

The idea of using airborne platforms to provide surveillance information of enemy troop positions and movements is not a new one. Since the dawn of the age of manned flight Armies have been exploiting the use of airborne observation to look deep beyond the front lines. In as early as the summer of 1911 the concept was being applied [Ref. 1] and by the end of the First World War it had been refined and well developed. However, until after the end of Second World War, the observation techniques were largely based on optical means such as observers or photographic cameras and were thus limited in their ability to "see" to times of fair weather and daylight. The sensors provided only limited range which meant that only limited observations were possible and that the platforms had to penetrate enemy lines in order to make their observations exposing them to enemy fire. It was only during the cold war that technology had advanced to the point where one could start to think of using alternate sensors. Recent advances in technology have made it possible to push the airborne ground surveillance concept in several important directions. Advances in sensor and computing technology allow observation to much greater ranges during all types of weather at any time of the day or night. As well, because of their increased range, the coverage can be over a much greater area providing a complete, theatre-wide, situational awareness as well as target selection capability. Because of the advances in computer processing power and communication systems the information gathered can be delivered in an immediate manner, giving commanders access to a continuous and current picture of the battlefield.

In April 1993 NATO formed an Ad Hoc Multi-Service Group (MSG) to investigate the requirements for a NATO Alliance Ground Surveillance (AGS) capability [Ref. 2]. Following this work, an Embryonic Project Structure (EPS) was established consisting of an Embryonic Project Office (EPO) and a Steering Committee. The EPS' work was divided into two phases. During the first phase the ownership options for a NATO AGS capability were to be investigated and during the second phase a NATO Staff Requirement (NSR) and other documents which will define the NATO AGS programme were to be drafted.

The first phase of the EPS' work concluded in the fall of 1995 and led to the endorsement by the North Atlantic Council (NAC), at the Ministerial level, of the concept of a "NATO-owned and operated core capability supplemented by interoperable national assets." The exact composition of this NATO owned core is not known at this time but could consist of both fixed wing and rotary wing surveillance platforms. The EPS has been reconstituted since November 1996 as the Provisional Project Structure and is at present in the midst of their second phase of work. A decision is expected in the autumn as to whether to proceed with immediate procurement due to the urgency of the need or to take a more deliberate approach and study the problem more fully.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

Four systems are being studied as primary candidates for the core capability, namely: the Italian Complesso Radar Eliportato Per La Sorveglianza (CRESO), the French Helicoptère d'Observation Radar et d'Investigation sur Zone (HORIZON), the United States' Joint Surveillance and Target Attack Radar System (JSTARS) and the United Kingdom's Airborne Stand-Off Radar (ASTOR) systems. All of these systems consist of an airborne sensor platform and ground-based exploitation stations. The airborne platforms communicate with the ground stations over system-specific medium/high-bandwidth data links. Complementary assets could range from current reconnaissance assets, UAVs, satellites, etc.

The fixed wing platforms both offer a *Synthetic Aperture Radar* (SAR) sensor which is capable of detailed imaging of the ground at considerable stand-off distances and the detection of fixed targets. These SAR sensors will have to operate in either a high resolution *Spot* mode, where a small area is imaged in detail, and a wider-area, lower detail, *Swath* mode. As well, these platforms offer a *Moving Target Indicator* (MTI) radar mode. This mode facilitates the detection of any targets which are in motion at or near the surface of the earth, again, at a considerable stand-off distance. The rotary-wing platforms each have MTI sensors, but no SAR capability. The complementary assets could provide a variety of other sensor data, including ESM, Electro-Optic, Infra-Red, etc.

3. THE INTEROPERABILITY PROBLEM

Although the interoperability problem could be considered in its entirety, it has generally been divided into two components: *Intra-AGS* Interoperability and *Extra-AGS* Interoperability. By Intra-AGS Interoperability is meant the interoperability between all of the various NATO owned and non-NATO owned AGS and other complementary surveillance assets. That is, Intra-AGS Interoperability deals with the problem of how to get all available sensor data to the ground stations in such a way that it can be exploited and possibly fused with data from other sensors and intelligence sources. By Extra-AGS Interoperability is meant the integration of the AGS capability and AGS data with the existing and planned NATO and national C^3I infrastructures. The Extra-AGS interoperability provides the channels and means to share the AGS data with the end users and to import data into AGS that can assist in the exploitation process. These two aspects of interoperability are illustrated in Figure 1. This report deals primarily with the first issue, that is, the interoperability between various AGS assets and does not address integration of the AGS data into the larger NATO C^3I infrastructure in great detail.

There have been many studies in the past which have addressed the issue of interoperability of the Alliance Ground Surveillance capability specifically or airborne reconnaissance systems in general. Some of these are listed in the references at the end of this report [Refs. 3-10]. NATO Air Group IV, through efforts such as STANAGs 7023 and 7085 have been working to achieve standards for data formats and data links in order to facilitate information exchange. The approach taken in this report is to construct an architecture which allows enough flexibility that it can capitalize on these standardisation efforts as they become available, but is not dependant on them for its success. Because of the existence of legacy national systems, the mandating of any one standard or STANAG with which the NATO owned capability must be compliant is not considered to be a complete solution to the problem.

4. PROPOSED ARCHITECTURE FOR INTRA-AGS INTEROPERABILITY

This section describes a proposed architecture for the solution of the Intra-AGS Interoperability problem. In the first sub-section, the proposed hardware architecture is described. Next the data that is to be interchanged is identified and finally the format for the data is discussed. These three components together constitute a complete architectural solution to the problem.

4.1 The Proposed Hardware Architecture

A generic AGS capability, as presently configured, is illustrated in Figure 2. It consists of an airborne sensor platform, a native data link specific to the particular platform, a *Ground Data Terminal* (GDT) for receiving the data on the ground and an *Ground Exploitation Capability* (GEC) for converting the data into information useful to the operational commander.

Although the configuration of the GEC for each candidate system is different, they all perform a number of generic functions, such as display and exploitation of MTI data, display of cartographic information, visualization of SAR information, etc. The GDT and GEC together form what is commonly called a Ground Station (GS) or Ground Station Module (GSM). Thus the ground station consists of two components: a system specific GDT, unique to each of the candidate systems; and a GEC, which is certainly different for each system but which performs many of the common, generic functions required for exploitation of AGS data.

Analysis has shown that the optimum solution [Ref. 15] to the interoperability problem, from a purely operational perspective, is to have all platforms utilize a common data link and common data formats. This allows the operational commander maximum flexibility in the employment of the surveillance assets. Unfortunately, when one includes technical and cost considerations and consideration of legacy systems, this solution does not appear to be practical in the near-term.

An alternate solution, which shares many of the operational and other benefits of the common data link solution, proposes that the two components to the GS be de-coupled so that one can think of the GDT and GEC as independent entities, connected together through a standardized common data format. This is the key concept of the proposed architecture and is illustrated in Figure 3. By adopting such an approach, the possibilities illustrated in Figure 4 become possible. One can think of combining any GDT with any available GEC. It is thus possible to deploy one GEC and several GDTs in order to receive data from any AGS or other surveillance asset operating in an area. The architecture also makes it possible to attach several GDTs to the same GEC and thus achieve reception of data from several airborne assets simultaneously. This capability allows for the possibility of real-time fusion of data from multiple sensors which could lead to advantages such as improved detection, increased surveillance coverage, improved track-continuity, etc. [Ref. 16]. The architecture can accommodate SATCOM data links or standardized common data links when they become readily available and commonly used. Finally, the proposed architecture provides a

clear and relatively inexpensive upgrade path for legacy systems. By providing an input for data in this common format, the legacy ground stations can be made to receive data from other sensors. Any NATO-owned ground stations could receive data from national assets utilizing the appropriate GDT and the common interface format. A very important point is that the architecture requires no expensive modifications to existing airframes - the changes are all implemented in the ground segment.

4.2 The Proposed Data Types

The hardware architecture for Intra-AGS Interoperability has been established in the previous section. In this section the data types that should be used within the architecture are established. This is done by first looking at data in a generic sensor system and then looking at the data existing in the airborne and ground segments of the candidate AGS platforms.

4.2.1 Generic Surveillance Sensor Data Flow

Data can exist in a number of different forms within AGS or other surveillance systems, ranging from raw sensor data through processed sensor data and finally to fully exploited data which results in useful operational information.

By *raw sensor data* is meant the data available directly at the output of the sensor sub-system. For radar systems this is usually in the form of a digitised video signal (time series). In the case of SAR radar data, this would equate to data prior to processing in the SAR processor where the SAR image is formed (so called I and Q channels). In the case of ESM, this again would be a sampled time series. Depending on the particular sensor, area of coverage, resolution, etc., the precise bandwidth requirements for raw sensor data can vary but are in the order of hundreds of megabits per second.

By processed sensor data is meant sensor data that has been manipulated in a signal processor in order to make it of use to an analyst. In the case of MTI data this would be the radar data, after it has undergone detection processing. In its simplest form, this process is accomplished by comparing the data to a threshold, and declaring a target present if the signal exceeds the threshold. Further information is then extracted from the data, such as, position, radial velocity, and perhaps some idea of target classification. In the case of SAR data, the intermediate step is the formed SAR image, usually consisting of an array of geo-located pixel values. At this stage the images would also be corrected for any airborne platform motions and focused in order to increase the sharpness of the images. Precise geo-location information is added to the image so that in subsequent stages accurate target location and mensuration information can be extracted. Data at this intermediate stage is referred to as *pre-exploitation data*. For pre-exploitation data, the bandwidth required is on the order of tens of megabits per second.

Finally, after the exploitation stage, the data is transformed into information on which commanders can base their decisions. For MTI this could take the form of intelligence reports or target tracks, giving position, true velocity, identification, etc. This is achieved by correlating consecutive "looks" with the MTI sensor to form tracks. MTI detections can also be grouped together to be identified as convoys. Targeting information such as predictions of future positions can be made and MTI data can be combined with other sources of information such as order of battle data bases, logistics movement plans, etc. in order to identify each target of interest, etc. For SAR data, the exploited version could be an annotated image or simply a list of fixed targets, including their location, identification, etc. These operations at present, unlike the operations in the first two stages of processing, rely heavily on a skilled and experienced operator and are accomplished at the exploitation stations, either aboard the airborne platform or within the ground-based exploitation station. It is the operationally significant information which results from the exploitation that we shall refer to as *post-exploitation data*. The data rates required can vary greatly depending on the exact nature of the information being transmitted and the timeliness requirements. In general this can be on the order of tens or hundreds of kilobits per second

These stages are illustrated in Figure 5.

4.2.2 Sharing Pre- vs Post Exploitation Data

At the highest level, interoperability can be achieved in one of three ways, namely: raw sensor data can be shared; data can be shared pre-exploitation; or data can be shared post-exploitation. First comparing the last two possibilities, the sharing of unexploited data results in a higher accuracy and more continuous tracking of targets [Ref. 16] than the sharing of exploited data. The data from several sensors may be complementary and gaps occuring in any single sensor coverage (due to terrain screening or doppler blind problems) may be filled in by the fusion of data from multiple sources so that a more continuous track is achievable. If tracks are established on the basis of data from individual sensors and then fused subsequently, due to gaps in sensor coverage it may not be possible to maintain a track in certain circumstances. However, sharing unexploited data requires higher bandwidth communication channels for transferring the data from all sensors to the exploitation stations, rather than just passing low bandwidth exploited data between exploitation capabilities.

4.2.3 Data Types Available in Candidate AGS Systems

In this section, the availability of each of the three data types is discussed for the four AGS candidate systems.

4.2.3.1 ASTOR

At this time it is unclear whether the ASTOR system will provide raw sensor data or pre-exploitation data to the ground station. In either case, some exploitation will be possible on board the aircraft. For the case of ASTOR then, in the worst case, raw sensor data, pre-exploitation data and post-exploitation data exist in the airborne platform and on the ground.

4.2.3.2 CRESO

The CRESO system transmits only pre-exploitation data to the ground. Therefore, for CRESO, raw sensor data and pre-exploitation data are available in the air and pre-exploitation and post-exploitation data are available on the ground.

4.2.3.3 HORIZON

HORIZON has essentially the same exploitation capability in the airborne platform as is available in the ground station. Therefore, for HORIZON, raw sensor data, pre-exploitation and post-exploitation data are available in the air. On the ground, as HORIZON links only pre-exploitation data to the

ground station, pre-exploitation and post exploitation data are available.

4.2.3.4 Joint STARS

Joint STARS has an extensive airborne exploitation capability so that raw sensor data, pre-exploitation data and post exploitation data are all available in the aircraft. On the ground, as the data link carries only pre-exploitation data, pre-exploitation and post-exploitation data are available.

4.2.3.5 Other, Complementary Sensors

A variety of other sensor types might be brought to bear on the ground surveillance problem. These include not only the ones discussed in some detail such as SAR and MTI, but also could include ESM, ECCM, Electro-Optic (EO), Long-Range Oblique Photography (LOROP), Infra-Red (IR), etc. Depending on the sensor and implementation, data sent to the ground can be either raw sensor data or pre-exploitation data.

4.2.4 Summary

For the proposed hardware architecture, it can be concluded that there is no point, at this time, in considering the sharing of raw sensor data as three of the four candidate AGS systems do not send this data to the ground. The sharing of post-exploitation data could be considered; however, this presupposes that the data from each sensor be individually exploited. This approach offers no savings in terms of sharing exploitation stations. As well, this solution suffers from the problems discussed above related to track accuracy and continuity. It is therefore concluded that the preferred data to be shared is pre-exploitation data.

4.3 Data Formats

At present, there exist a number of STANAGS and other standards which define data formats that could be used to represent the data within an AGS system [Refs. 6, 12, 13, 14, 15]. However, there exists no standard which can satisfactorily represent pre-exploitation data. Raw sensor data can be adequately represented in the format specified in STANAG 7023 [Ref. 6]. Post-exploitation data can be represented in the formats specified in STANAG 3377 [Ref. 12], STANAG 5516 [Ref. 13], ADatP-3, [Ref. 14], STANAG 4545 [Ref. 15] and others. Therefore a new set of standardized formats are required to represent pre-exploitation data. The work at NC3A over the last several years has focused on developing and implementing formats suitable for this purpose [Ref. 18]. Discussions with Air Group IV have commenced with a view to incorporating these formats into a future release of the STANAG 7023.

4.4 Proposed Architecture Summary

In summary, the proposed architecture consists of three parts: the hardware architecture based on the concept of de-coupling of the GDT and GEC and insertion of an intermediate standard common data format; the sharing of pre-exploitation data; and the use of the formats developed at NC3A for data representation. These concepts have been developed, prototyped and tested in the NC3A AGS Testbed over the past several years. The final proof of concept occurred in June 1997 when an interoperability experiment was held outside of Paris. This experiment is described in the following section.

5. PARIS INTEROPERABILITY EXPERIMENT

The Paris Interoperability Experiment (PIE) was a large experiment involving over 200 people from six nations and NC3A. It was held at the Centre d-Essais en Vol (CEV) Brétigny-sur-Orge, South of Paris, during the week following the 1997 Paris Air Show. The experiment is described in terms of its aims, conduct and achievements in the following paragraphs.

5.1 PIE Aims

There were two major aims of PIE. The first was to prove that the interoperability concepts described above could be made to work in a practical environment. They had been prototyped and tested in a laboratory environment and had reached a sufficient state of maturity that it was time to take them out in the field and prove them in practice. The second major aim was to gather a data set that could be used for analysis of the complementary nature of AGS assets. For the first time there would be the possibility of gathering data including simultaneous coverage of the same terrain by two or more AGS platforms. Specifically the aims were:

5.1.1 Prove Interoperability Concept

The specific test conducted were used to test the following:

a. the transfer of MTI in real-time from Joint STARS to the HORIZON and other ground station;
b. the transfer of MTI in real-time from HORIZON to the Joint STARS and other ground stations;
c. the transfer of SAR data from Joint STARS to HORIZON and other ground stations;
d. the simultaneous real-time transfer of SAR and MTI data from Joint STARS and MTI data from HORIZON to all ground stations;
e. the transfer of Radar Service Requests (RSR) from each ground station to each of the airborne platforms;
f. understand the Electro-Magnetic Compatibility (EMC) problems related to the simultaneous operation of multiple data links in close proximity; and
g. transfer of secondary (post-exploitation) data between ground stations and between ground stations and end-users. This included the transfer of Link-16 formatted messages, ADatP-3 messages and messages describing locations and other characteristics of detected convoys.

5.1.2 Gather Multi-Platform Data Set

The aircraft orbits and convoy routes were established in order to gather a data set which could be used to analyze and quantify the following measures of performance related to the complementary use of multiple AGS platforms:

a. the achievable improvements in the detection of moving targets;
b. the achievable improvements in target resolution;
c. the achievable improvements in target classification;
d. the achievable improvements in track continuity, compensating for gaps in coverage due to:
   i. zero target Doppler situations;
   ii. terrain masking; and
   ii. coverage area.

5.2 PIE Conduct

The PIE Experiment was conducted from 23 to 27 June at CEV. Prior to these dates all equipment was brought to

Brétigny. During the first day of flying, each platform was used individually for three hours in order to conduct tests related to the ability of single platforms to distribute the data in accordance with the prototype architecture. The remainder of the time involved simultaneous flights of the two platform types.

Three target convoys were employed for the first three days of PIE. The convoys were furnished by the French Army and consisted of seven vehicles each. The lead vehicle was soft-skinned while the remaining six were wheeled armoured vehicles. During the last day of flying, a single convoy of similar composition was used. The convoy routes involved a variety of terrain types, including: relatively flat terrain presenting few visibility problems; more challenging terrain presenting some screening problems from certain directions; and quite challenging terrain.

The convoys were equipped with Global Positioning System (GPS) receivers. The positions of the convoys were relayed via a commercial satellite system known as EuTelTracs, back to CEV. This provided real-time updates of the convoy's positions for monitoring and control purposes during the experiment and also truth data for post experiment analysis.

The equipment was set-up as indicated in Figure 6. Participating in the experiment were one United States Joint STARS E-8 aircraft and two Joint STARS GSMs, two French HORIZON helicopters and two HORIZON GSMs, software and hardware representative of the ground stations of the United Kingdom ASTOR, Italian CRESO and German Dornier system. Additionally, the Norwegian Mobile Tactical Operations Centre (MTOC), the NC3A Prototype Ace Intelligence System (PAIS) and Recognized Air and Surface Picture (RASP) display system were included as representative of users in the Army, Intelligence and Air Force communities.

5.3 PIE Achievements

In short, all of the technical aims of PIE were met. The data from both HORIZON and Joint STARS was exchanged and also distributed to the other participants in real-time and simultaneously. RSRs were sent to both platforms and serviced. There were no ECM problems experienced during the experiment. All three data links that were operating (two Joint STARS Surveillance and Control Data Link (SCDL) and one HORIZON Agatha data link) in the area functioned correctly and simultaneously.

In terms of the second aim, that is the gathering of a data set, 16 hours of Joint STARS MTI and SAR data and 17 hours of HORIZON data were gathered. Of this data approximately 10 hours includes overlapping sensor coverage, in the following combinations:

a. 8 hours with Joint STARS and one HORIZON;
b. 1 hour with Joint STARS and two HORIZON; and
c. 1 hour with two HORIZON.

This data represents a very unique data set which will be used as a basis for the investigation of many issues in the future.

6. CONCLUSIONS

This report has presented a proposed solution to the problem of Intra-AGS interoperability for the NATO AGS capability. The approach represents a complete architecture in terms of hardware, data types and data formats. This architecture has been extensively tested in the laboratory and through a significant practical experiment in the field.

It is considered that the best approach to the achievement of Intra-AGS interoperability, from an purely operational perspective, is the use of a common data link. However, when one also takes into account the technical and related cost and risk implications, the most viable approach, in the short to medium term, is considered to be to share the pre-exploitation data at the interface between the GDT and the GEC through the use of a standardized, common format. This is considered a low risk solution to interoperability which will allow tremendous flexibility in the employment of AGS assets, a clear path for nations to follow in order to achieve interoperability with the NATO core fleet and a growth path for the future.

7. REFERENCES

[1] Bregnet, M. Emmanuel, "Le Renseignement arien et la Conduite de la guerre, l'Example Francais de 1914-1918.", Service historique de l'armee de l'air, France, Colloque international: "L'aspect court terme dans la conduite de la bataille aerienne," Copenhague, 3-7 Octobre 1988.

[2] AC/259(SURV)D/6, "MSG Study Report to CNAD on an Alliance Ground Surveillance Capability", 1 October 1994, NATO CONFIDENTIAL

[3] AGARD-AR-334, Aerospace Applications Study 39, "Development of a Reconnaissance, Surveillance and Target Acquisition (RSTA) Architecture to meet NATO's Reaction Forces Requirements", February 1995, NATO RESTRICTED

[4] Alcatel, Alenia, Computing Devices Hastings, Dasa-Dornier, ESG, Fokker, Matra-Cap-Systemes, Motorola, Northrop-Grumman, "NATO Joint Stars Ground System Integration Preliminary Study", 1 December 1995

[5] STANAG 7023 Edition 2 - Study Draft 2, "Air Reconnaissance Imagery Data Architecture", 6 June 1997

[6] STANAG 7024 Edition 2 - Study Draft 1, "Imagery Air Reconnaissance Tape Recorder Standard", 6 June 1997

[7] STANAG 7085 (Proposal for draft Edition 2), "Interoperable Data Link for Imagery Systems", 6 October 1995

[8] Chairman Standing Interoperability and Applications Working Group For Reconnaissance, Surveillance, and Target Acquisition (SIAR WG), "Background Paper on Air Group IV's Imagery Interoperability Architecture Program For NATO Reconnaissance Systems", Draft, 9 October 1995

[9] AC/224(AG.IV)D/54, "Proposed Approach for Achieving Interoperability of NATO Stand-Off

Surveillance and Target Acquisition Systems", 16 December 1993

[10] AC/259 (SOSTAS)D/8, Multi-Service Group on Stand-Off Surveillance and Target Aquisition Systems (SOSTAS), "Final Report for the Airborne Radar Demonstrator System (ARDS), 12 July 1991

[11] STANAG 3377 Edition 4, "Air Reconnaissance Intelligence Report Forms", Military Agency For Standardization, 12 June 1978.

[12] STANAG 5516, "Tactical Data Exchange - Link 16", Military Agency for Standardization, Edition 1, March 1990

[13] ADatP-3, "NATO Message Text Formatting System (FORMETS)", Change 9, NATO IMS CIS Division, 6 February 1995

[14] STANAG 4545, "NATO Secondary Imagery Format (NSIF)", Edition 1 - Ratification Draft

[15] Lenk, P.J., "NATO AGS Interoperability Study", published in Proceedings of The NATO C3 Agency Alliance Ground Surveillance Interoperability Symposium, Held 26 and 27 February 1997, at The NC3A, The Hague.

[16] Waltz, Edward, and Llinas, James, "Multisensor Data Fusion", Artech House, Boston, 1990

[17] "Interface Control Document for Milestone I of The NC3A AGS Capability Testbed", NATO C3 Agency, Edition 1.09 (Draft).

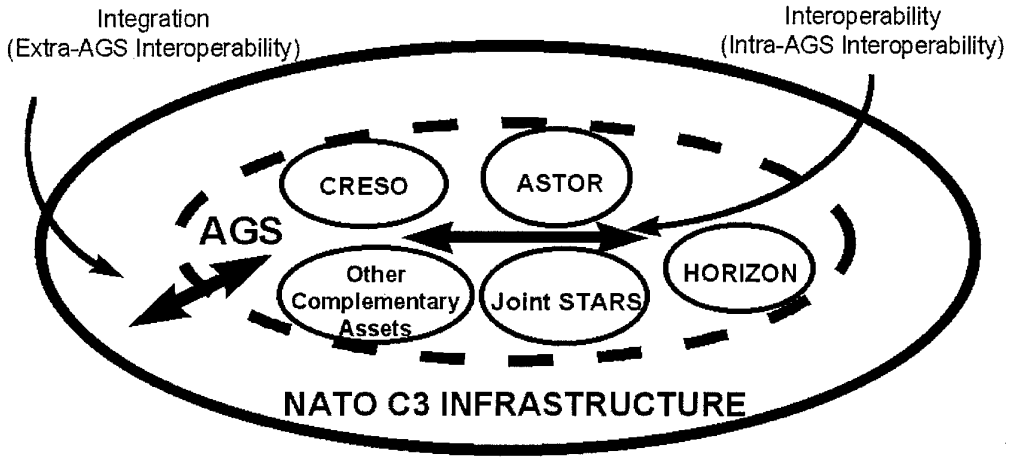


Figure 1: AGS Interoperability

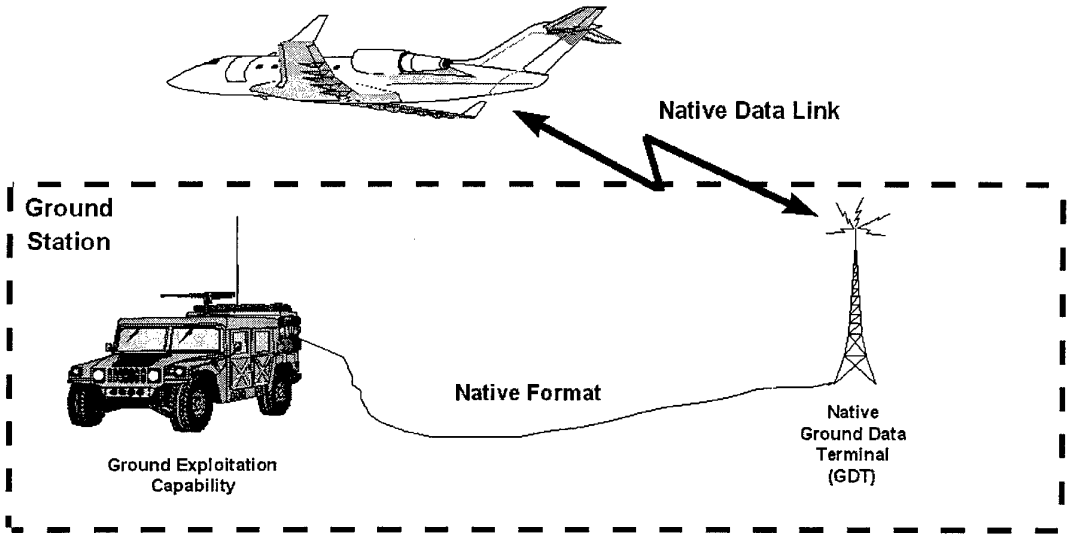


Figure 2: Present AGS System Configuration

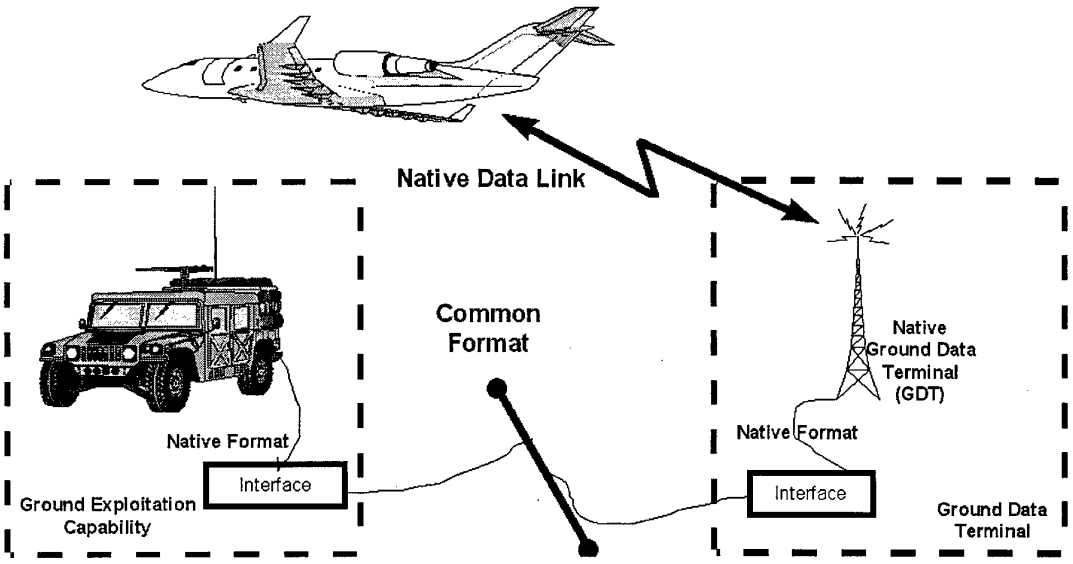


Figure 3: Proposed AGS System Configuration

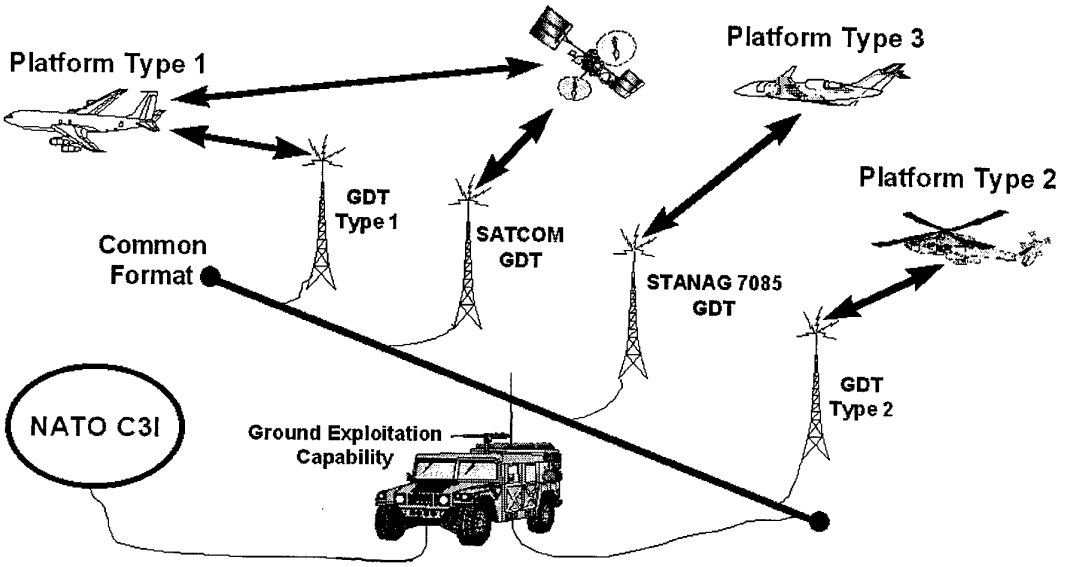


Figure 4: Possible Application of Proposed Architecture

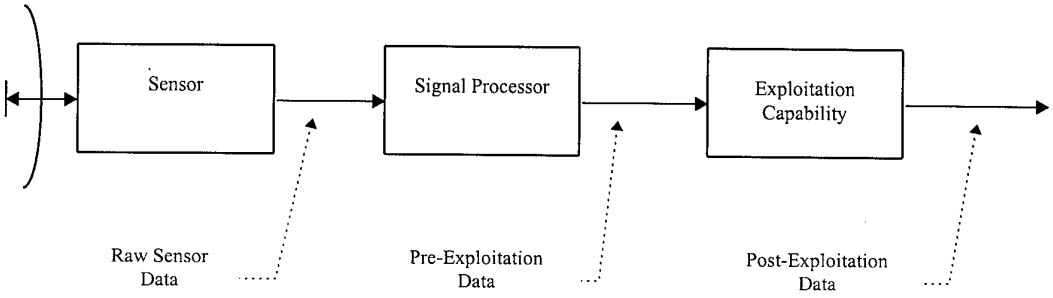


Figure 5: Data Flow For Generic AGS Sensor

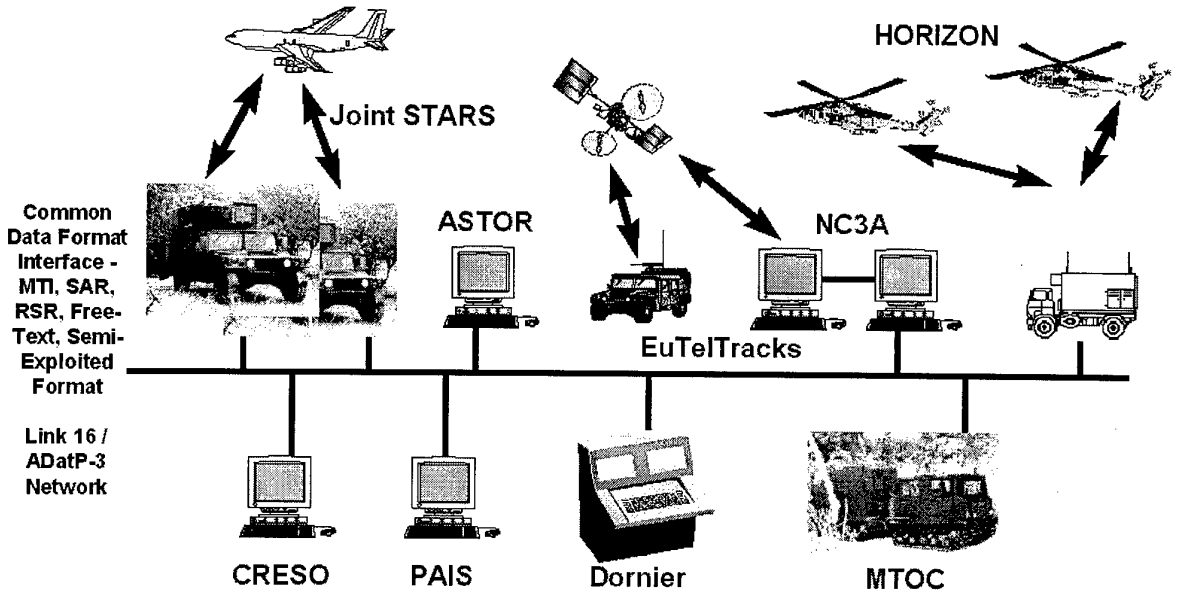


Figure 6: Paris Interoperability Experiment Equipment Configuration

Imagery and Signals Intelligence Data Fusion: Issues and Methods

Edward L. Waltz
ERIM
PO Box 134001
Ann Arbor, MI 48113-4001 USA
waltz@erim-int.com

Jerry Hart
BTG, Inc.
1945 Old Gallows Road, Suite 700
Vienna, VA 22182 USA
jhart@btg.com

SUMMARY

Imagery and Signals Intelligence (IMINT and SIGINT) data sources and processing have traditionally followed independent channels, with cross use of the information being achieved only by manual human interaction between users at the end of their respective channels. The characteristics of each of these sources are quite different, making automatic correlation of the data a challenging task, but the rewards of correlated data hold the promise for more accurate and robust surveillance than now available from either independent source. IMINT, for example, is characterized by a low revisit rate, high spatial fidelity and broad spatial context, while SIGINT provides a high revisit rate, low spatial fidelity and a broad temporal context.

This paper describes the specific issues that face designers who seek to automatically fuse imagery and signal data - and the methods used to overcome the obstacles posed by the fundamental differences in the data types. Four areas of IMINT-SIGINT fusion are addressed: 1) fusion for cross-cueing and sensor management, 2) fusion for improved target geolocation, 3) fusion for increased target detection and identification confidence, and, 4) fusion for improved visualization of the battlefield. The paper provides examples illustrating future solutions using terrain data, UAV imaging sensors, and real-time-in-cockpit (RTIC) techniques illustrated by ERIM-developed technologies. The integration of these capabilities into joint intelligence systems is illustrated by BTG-developed technologies.

The paper addresses two of the specific topics listed in the AGARD call for papers: 1) Applications of multiple sensors and data fusion in - target recognition and surveillance, 2) Sensor data networks across imaging and non-imaging sensors.

1 INTRODUCTION

Imagery and Signal intelligence disciplines have developed to high levels of maturity over the past 40 years, each developing mature technologies to collect, analyze and exploit data to deliver finished intelligence products to users. Within each of the disciplines, fusion of data is being performed to enhance the overall intelligence product, and to maximize the extraction of meaning (intelligence) from the raw data. It has been recognized that fusion *between* the intelligence disciplines offers the potential for further improvements in both performance and effectiveness.

1.1 Fusion *Within* IMINT and SIGINT

Representative SIGINT systems with highly automated collection and processing capabilities include the U.S. RC-12 Guardrail, RC-135 V/W Rivet Joint, or EP-3 Aries (see [1], [2]). These systems employ sophisticated broadband receivers to collect and sort signals, both communications and more general electronic emissions (radars, jammers, etc.) to:

- Provide efficient scanning of multiple bands to achieve high intercept probabilities for intermittent emitters, band-to-band handoff, instantaneous intercept, record and *"gisting"* (abstracting) of emitters.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

- Develop an electronic order of battle describing emitter types, quantities, functions, relationships, capabilities, and tactical behaviors.
- Locate individual emitters
- Monitor use, modes and capacity of emitters by *external* behavior, or by inference from *internal* data to extract intelligence data.

Because these systems search multiple bands for emitters, which are often related, fusion of data across bands must be performed to develop relationships between, for example, low-band communications and high-band radar activities. SIGINT testbeds, such as the Canadian Data Fusion and Correlation Techniques Testbed (DFACTT) [3] are developing and refining automated techniques to correlate the information gleaned from multiple SIGINT channels to support analysts operating in today's dense, complex electromagnetic environment.

Similarly, the IMINT community has developed automated means to register and combine multiple images and terrain data of the same ground areas to enhance the utility of imagery intelligence products. Much of the U.S. drive to digitize and visualize the battlefield is based upon technologies which register and combine imagery, maps, digital terrain and other spatial data sources to provide a comprehensive and accurate geographic information system (GIS) to:

- Provide efficient use of archived high-fidelity imagery and tasking of reconnaissance imaging sensors, sensor-to-sensor cueing and handoff, near-real time collection, exploitation and extraction of objects of tactical interest.
- Develop physical orders of battle describing identity and location of fixed, mobile and moving objects of tactical interest.
- Develop tracks of moving targets by revisits at sufficient rates to maintain tracks
- Precisely locate individual objects for surveillance, targeting, and change detection.
- Survey targets to monitor function and behavior, or for battle damage assessment (BDA).

Next generation IMINT systems, such as the U.S. Common Imagery Ground Surface System (CIGSS) have specified common standards for imagery formats and metadata[4] to provide the capability to permit spatial data fusion and visualization of data from multiple IMINT sources.[5] Numerous techniques exist to perform imagery and spatial data fusion to synthesize two and three-dimensional views of the battlefield which maximize the information content from all available imaging sensors. Imagery data of differing resolutions, geospatial accuracy and spectral content may be registered, orthorectified to remove distortion due to terrain relief, mosaicked and combined to form a common base for extracting features, generation of rapid map updates, and analysis for a variety of military applications. (See [6] for a survey of this technology.)

1.2 Issues Facing Intelligence Fusion *Between* the IMINT and SIGINT Disciplines

Although both of these disciplines have matured and are applying the principles of data fusion to correlate, merge and combine data, the challenge of fusion IMINT-to-SIGINT data remains a challenge due to the distinct differences in SIGNT and IMINT data sources. Figure 1 illustrates the major difference in the two sources: the fidelity of measurements differ in each of the two dimensions of observations. IMINT provides excellent spatial fidelity, but lacks the revisit rate and instantaneous coverage afforded by SIGINT. SIGINT lacks spatial fidelity, but provides high temporal fidelity. While these are general characteristics, they apply to most

cases. Consider the combination of UAV synthetic aperture radar (SAR) and airborne electronic support (ES). The SAR can provide spatial resolutions on the order of one meter or better (e.g. spatial accuracies on the order of meters or tens of meters) and revisits of a target on the order of minutes or tens of minutes at best. A tactical ES system can provide instantaneous revisits in less than a second, but with single-look spatial accuracies ranging from hundreds of meters to kilometers.

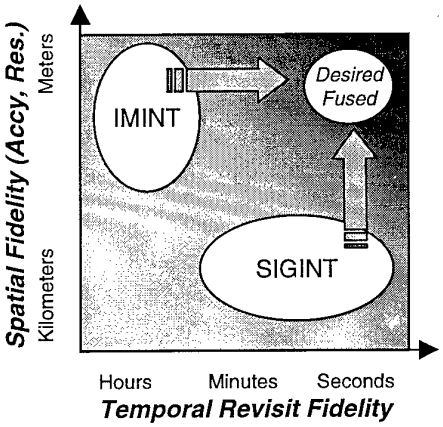


Figure 1- IMINT – SIGINT Contributions provide the potential for high-fidelity pace-time intelligence products.

The objective of fusing IMINT and SIGINT data is to provide combined intelligence which used the best of each source - a combined, high fidelity space-time product. The ability to achieve this fusion, however, requires a means to unambiguously correlate data across sources to, for example, locate a specific emitter within an image.

Figure 2 details additional contrasting characteristics of these two sources:

1) IMINT captures spatial objects (entities or phenomena from events – pixels), while,
2) SIGINT captures temporal events (emission events: pulses, modulations).
3) IMIT provides inherent spatial information in the data, in contrast,
4) SIGINT derives spatial information from measurements of the time and direction of arrival of signals.
5) IMINT delivers high-fidelity spatial information and context, while,
6) SIGINT delivers high-fidelity temporal content (and derived information, e.g. the identity of an emitter, or internal content) and context.

These contrasting characteristics make the two sources ideal candidates to deliver high fidelity spatial and temporal reports on battlefield targets and their behavior. These contrasting characteristics also pose major challenges to fusion:

1) The platforms and collection methods of each discipline are sufficiently different that coordination of sensors to achieve near-simultaneous acquisition of both

Characteristic	IMINT	SIGINT
	OBJECTS	EVENTS
Information Content	PRIMARY: Sensed data describes real-world entities and their spatial context SECONDARY: Time of observation	PRIMARY: Sensed data describes electronic events (emissions) and their temporal context (waveforms) SECONDARY: Location derived from direction and time of observation
Context of Data	SPATIAL Location of existence	TEMPORAL Time of occurrence
Subjects	Physical Objects	Emitting Objects
Focus of Analysis	Signature of spatial shape, size, texture and context	Signature of space-time signal behavior.
Resolution Dimensions *(on the order of)*	Space – meters Time – minutes, hours	Space- 100's meters to kilometers Time – seconds
Update Intervals	Hours - Minutes	Seconds- Milliseconds

Figure 2 – Contrasting Features of IMINT and SIGINT Sources

IMINT and SIGINT data is difficult.

2) Spatial correlation between IMINT objects (entities in a scene) and SIGINT objects (spatial ellipses) provides ambiguity and uncertainty due to the large difference in spatial resolutions of the data sets, and the time difference in collection.
3) Combination of the data is also difficult, because the sources each reports different characteristics about the entities and events.

1.3 Benefits of IMINT-SIGINT Fusion

The expected benefits of automated fusion of IMINT and SIGINT data include:

- Improved tasking and utilization efficiency of both SIGINT and IMINT sensors, via improved sensor-to-target pairing.
- More rapid warning, cueing and intelligence understanding in the presence of uncertainty.
- Improved detection, identification, characterization and geolocation of targets of interest. (The improvement may provide a reduction in uncertainty, resolution of conflicts, improvement in accuracy.)
- Rapid geolocation of threats for sensor-to-shooter targeting. [7] (This improvement may include a reduction in time-to-locate with sufficient precision to react.)

These contributions are important technical elements needed to achieve the information superiority objectives on the modern battlefield. The U.S. Advanced Battlespace Information System (ABIS) Task Force identified multisensor fusion and sensor cross-cueing as a "needed technology" to achieve precision information direction.[8] The task force identified the progressive incorporation and integration of IMINT and SIGINT sensors as an ultimate stage to achieve automatic sensor-to-target pairing capabilities.[9] One example of the move toward these capabilities is the consideration of developing Global Hawk UAV payloads that will autonomously cue imaging sensors by emissions detected by Electronic Support sensors.[10]

Traditionally, IMINT and SIGINT are combined by intelligence analysts who have oversight and access to both imagery and signals data. With broad intelligence oversight, their manual fusion functions include the following kinds of actions:

- Overlay SIGINT data (error ellipses, lines of bearing) on archive imagery or maps, and search recent imagery for signs of the physical emitter.
- Cue IMINT sensors to look in regions of emitter activities.
- Cue SIGINT sensors to search for emissions in areas where military activity is located in imagery.

Automation of these types of IMINT-SIGINT fusion functions requires the ability to automate the following activities:

Alignment (Registration) – Spatial and temporal alignment of the two sources must be performed to relate the data sets. Spatial alignment requires transformation of the signal location data onto imagery coordinates; this generally requires the mapping of an uncertainty volume (e.g. error ellipse) onto an image surface.

Correlation – Possible associations between SIGINT objects (e.g. an error ellipse) and IMINT object (e.g. all military-like entities within the ellipse) are enumerated, and an association metric is computed for each. The metric may account for spatial, temporal, behavioral and identifying feature similarities. A single, highest scoring candidate may be assigned ("correlated") or multiple hypotheses may be correlated and retained for further consideration and refinement as more data arrives ("deferred-decision" or "multiple hypothesis").

Combination – Correlated objects may be combined in various ways to refine or verify the identification of the target entity. If the emitter is a general fire control director, for example, and the correlated objects in imagery reveal a particular firing unit, the specific category of missile system may be identified.

Reasoning – Contextual data (e.g. terrain, road networks, lines of communications and line-of-sight considerations) may also be used to support the correlation and combination processes to restrict alternatives.

Resource Tasking – Tasking of collection and processing resources may be coordinated to provide simultaneous, near simultaneous or synchronized acquisition of data and efficient processing.

In the following sections, we discuss the functional opportunities to integrate and automate IMINT and SIGINT sources and the technologies that will permit this automation.

2. ALTERNATIVE METHODS FOR IMINT-SIGINT FUSION

A number of alternatives exist to improve the coordination and combination of IMINT and SIGINT Sources. Figure 3 illustrates the typical independent intelligence collection and processing "stovepipes" and the four major areas where improvements can be achieved. The following four subsections detail each of the four areas of opportunity and offer conceptual approaches to perform fusion functions.

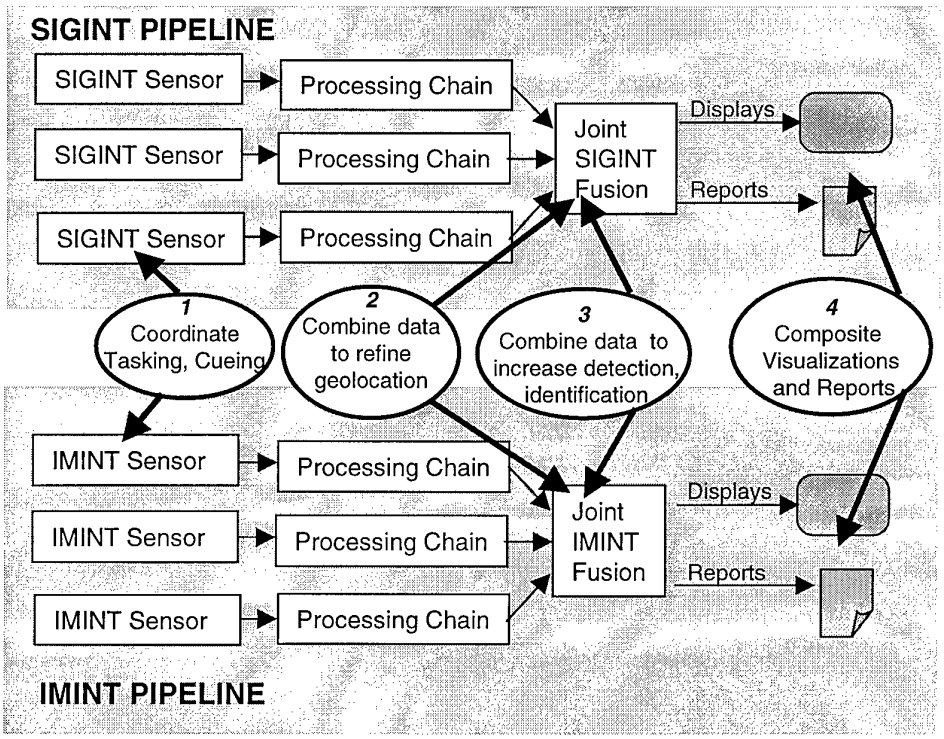


Figure 3- Four major areas of coordination and combination exist

2.1 Coordinate Tasking, Cueing & Control

The coordination of tasking, and real-time cross cueing of sensors hold the promise to improve the combined coverage volume, and speed to focus on targets of interest. Several tasking modes may be considered (Table 1) provide the ability to perform combined search and iterative focus to improve the composite target intercept performance (for SIGINT) and characterization (for IMINT).

Table 1 – Candidate Coordinated Tasking and Cross -Cueing Modes

Mode	Functions
SIGINT Area Search	• Search broad areas for emitters of interest • Cue IMINT to locate and verify detected targets
IMINT Area Search	• Search broad areas for non-emitting mobile and fixed targets • Cue SIGINT to focus on probable emitting targets
SIGINT Focus	• Task dwell and precise tune on designated area to verify identity and location
IMINT Focus	• Task precise imagery, or image sequence on designated area to verify identity and location

Consider the following representative coordination sequence between the two sources (as indicated by the numerical sequence in figure 4):

1. SIGINT are search locates a low probability detection, with very coarse error ellipse;
2. A cross cue to IMINT is generated, requesting a search of the area;
3. IMINT searches archive imagery for potential locations, and tasks the next available collector to collect imagery over the probable locations;
4. The probable locations are used to cue SIGINT;
5. SIGINT collects high dwell and precisely tuned data over the probable areas, then issues reports to IMINT;
6. IMINT used this data to search recent imagery (collected by the tasking in step, 3, earlier) to detect and locate the target.

The state diagram in Figure 4 illustrates the iterative and supportive sequence by which the two sources can be tasked and cued to refine low-confidence detections to either eliminate, or validate and locate targets.

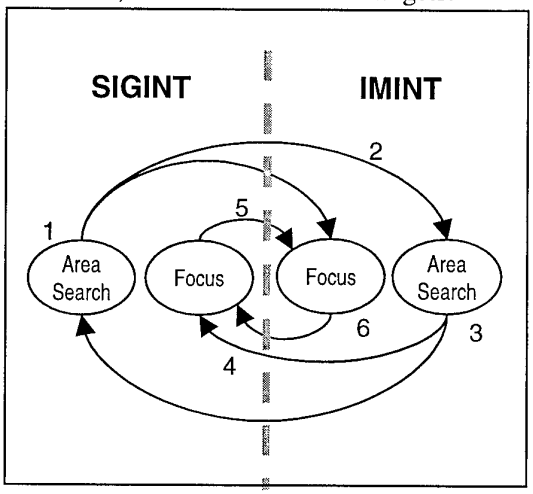


Figure 4- Transition of Search and Focus States for Coordinated cueing and tasking

This kind of cooperative and iterative search requires very close coordination of both sources, real-time interaction between processing of both data, and the ability to rapidly re-task both sensors. The U.S. Integrated Collection Management ACTD (Advanced Concept Technology Demo) is developing these kinds of capabilities, and has published a desired goal of 2-4 hours for response time between sources to achieve "tasking within friendly and enemy cycle times."[11]

2.2 Fusion for Improved Geolocation

The coordination process described above can yield improvements in the geo-location of critical targets that have been located within coarse SIGINT error ellipses. The use of imagery and terrain data to search probable locations (and to delimit unlikely locations) within the coarse SIGINT ellipse

may yield precise locations or precise probable locations.

2.3 Fusion for Increased Target Detection and Identification Confidence

In addition to geolocation improvement, the process may also provide for more accurate detection and identification of targets. Consider the following possible situations in which joint IMINT and SIGINT data may be combined to improve detection and/or identification:

1) SIGINT may locate an individual emitter, but an IMINT search refines the associated military components associated with the emitter.
2) Low-confidence SIGINT hits may be rejected or upgraded to higher confidence with collaborating IMINT or terrain contextual knowledge.
3) IMINT data may identify candidate activity areas, with no observable targets, and SIGINT may be tasked to search for relevant emissions to confirm or eliminate the areas.
4) SIGINT data may precisely locate an emitter, with ambiguous identity, and IMINT may provide contextual data to support high-confidence identification.

Programs such as the U.S. DARPA Dynamic Multi-user Information Fusion (DMIF) program are developing state-of-the-art methods to combine IMINT and SIGINT report objects. In that program, tactical decision reports (messages containing identity and location of targets) are parsed, common objects are correlated, and linkages between objects are discovered using force level structure and behavioral templates. [12]

2.4 Improved Visualization of the Battlefield

The effective coordinated tasking, correlation and combination of IMINT and SIGINT sources and data provide the potential to maintain comprehensive views of the battlefield – and to present these in dynamic visual presentations to optimize human comprehension of the battle situation. Visualizations of joint IMINT and SIGINT data may take on many forms, dependent upon system and user requirements, but several alternative candidates are presented in Figure 5. The figure illustrates three successive levels of sophistication in overlaying a coarse SIGINT error ellipse on imagery-derived products to aid an analyst in searching for potential sources. These visualizations provide both the *temporal* context of SIGINT and the *spatial* context of IMINT in a common form – and should allow the analyst to "drill-down " into the data bases from either source to improve the understanding of the battlefield.

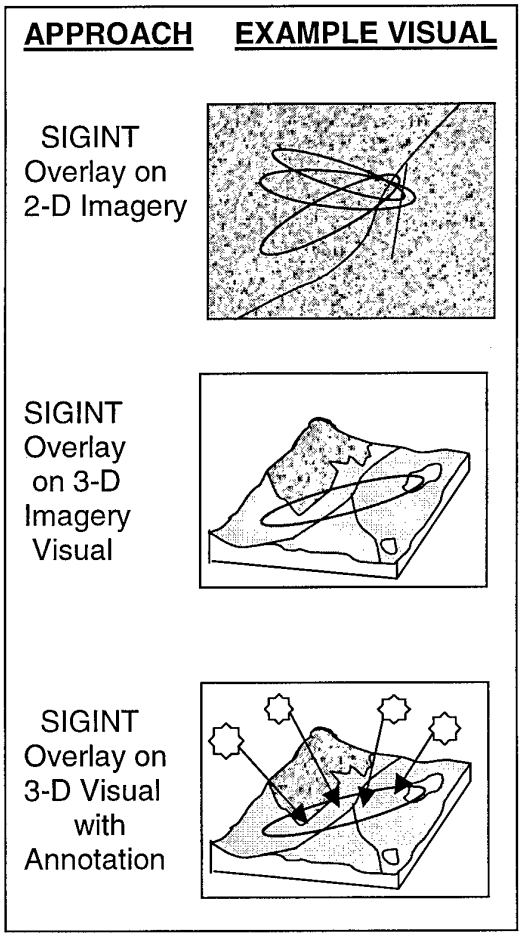


Figure 5- Alternative combined visualizations of IMINT-SIGINT data.

4. CONCEPTUAL FUTURE SOLUTIONS

The ability to achieve effective IMINT-SIGINT fusion requires the integration of several technologies and system commitments; the U.S. programs cited throughout this paper have illustrated some of the efforts to accomplish this. The technologies and system commitments to make this possible include:

- *Refined Collection Assets* – The introduction of medium and high altitude endurance unmanned air vehicle (UAV's), in concert with manned platforms, will provide more continuous, longer dwell views by IMINT to increase the potential for near-simultaneous collection with SIGINT. This will increase the effectiveness of cross-cueing for dynamic targets, while increasing the likelihood that IMINT can be tasked to geolocate SIGINT targets.
- *Coordinated Intelligence Collection Architectures*- The IMINT and SIGINT information chains cannot be operated as independent "stovepipes" to achieve timely coordination of tasking and exploitation. The architecture for planning, tasking and exploitation must provide for rapid interaction based on *overall* prioritization of joint resources. As an example, the U.S. Defense Airborne Reconnaissance Office (DARO), has defined a "Technical Architecture" which provides data paths between IMINT and SIGINT with the requirements for: 1)"reach-back" (an operation where sensor data is passed back to supporting site facilities for processing and/or exploitation/analysis) and, 2) reach-forward" (capabilities for data from supporting intelligence centers to be forwarded to the airborne reconnaissance system for relay, exploitation, cross-system cueing, data correlation, or other intelligence functions) capabilities.[13]
- *Source Data Reporting*- SIGINT and IMINT report formats (and data content) should provide metrics to support the correlation and combination processes. These metrics include: time of collection, common statistical uncertainty measures for location and identification, exploitation support data, etc.
- *Exploitation Architectures and Processing* – The exploitation architectures must also provide the capability to store, link and jointly visualize both sources of data. Fusion algorithms providing correlation and combination (managing uncertainty and able to provide multiple hypotheses or uncertainty metrics to users) is also required.
- *Precision Terrain and Feature Data* – Spatial context data consisting of precision terrain, feature and other data (See [14] for description of U.S. National Imagery and Mapping Agency's categories of topographic data, and [15] and [16] for a typical analyses of the applicability of this data for military application.)

In addition to technology changes, doctrinal changes will also be required to exploit the potential of IMINT-SIGINT fusion. In recent test exercises, such as the massive Force XXI Advanced Warfighting Experiment (AWE), the Army's 1st Brigade Experimental Force (EXFOR) reported significant benefits of fusing and cross-cueing UAV IMINT with ground and airborne SIGINT.[17] But the exercise also illustrated the need of rethinking operational procedures to complements the introduction of fusion technologies. The kinds of doctrinal changes that may be considered include:

- *Joint Sensor Utilization* – As the technology to allow coordinated tasking is developed, so must the doctrine to support this new use of the assets – as a joint sensor network, rather than two independent, but cooperating, sources.
- *Fusion Analyst Position*- Analysts with joint IMINT and SIGINT skills may be required to take maximum benefit of integrated visualizations. A new intelligence discipline, the *fusion analyst*, may be required, with an appropriate

training curriculum, to reap the full potential.

4. CONCLUSION

The automated fusion of imagery and signal intelligence data, now performed manually and infrequently, has the potential to provide significant timeliness, and information accuracy benefits to the warfighter. In order to achieve the maximum benefits of automation, 1) system architectures must be structured to permit collection interaction and coordination of the intelligence chains, 2) fusion processing and visualization technologies must be applied to optimize the extraction of joint intelligence, and, 3) supporting data, such as precision terrain data, must be available to permit contextual reasoning about the earth terrain.

5. REFERENCES

[1] LTC Bruce Jette, *Guardrail/Common Sensor Program*, SEMA Users's Conference, Ft. Huachuca, Jan 22-26, 1996.

[2] David A. Fulghum, "Navy EP-3s Gain Upgrades Plus Confidence", *Aviation Week and Space Technology*, May 5, 1997, pages 48-55.

[3] DFACTT: Data Fusion and Correlation Testbed, *Defence Research Establishment Ottowa* (DREO), URL: www.dreo.dnd.ca/ pages/electwf/ewd005.htm.

[4] *Metadata* refers to data that describes the context of the imagery data: time of collection, area covered, source, source parameters, annotation, etc.

[5] Common Imagery Ground/Surface System (CIGSS) Acquisition Handbook, Version 2.0,Volume I Standards, Defense Airborne Reconnaissance Office, 11 April 1997, p. 37.

[6] Waltz, Edward L. " The Principles and Practice of Image Data Fusion, "*Proc. of 8th National Symposium on Sensor Fusion*, 15-17 March 1995, pp. 257-278.

[7] Lum, Zachary, "KILLINT: EW on the Offensive", J. of Electronic Defense, July 1997, pages 37-42.

[8] Advanced Battlespace Information System (ABIS) Task Force, May 1996, Volume II Major Results, pages 3-21 to 3-28.

[9] Advanced Battlespace Information System (ABIS) Task Force, May 1996, Volume IV, Sensor-to-Shooter Working Group Results, pages 2-104 to 2-117.

[10] "JROC Beginning UAV Mission Ranking Process; SIGINT, Data Links High Priority", *Defense Information and Electronics Report*, February 7, 1997, pages 17-18.

[11] Joint Warfighter DTO Information Superiority, Integrated Collection Management ACTD, http://www.dtic. mil/dstp/ STP/97_dtos/jw_dtos /jw_dto_a.htm#A.09

[12] U.S. Army Corps of Engineers Topographic Engineering Center, *DMIF Bidders Brief Presentations Page*, DMIF Description, URL: www.tec.army. mil/dmif/dmif_bid.htm

[13] DARO Airborne Reconnaissance Information Technical Architecture (ARITA), Version 1.0, 30 September, 1996, http://www.acq.osd. mil/daro/homepage/arita/arita.html., section 3.7.2.

[14] Geospatial Information Infrastructure Master Plan, Version 1.0 ,October 2, 1997, National Imagery and Mapping Agency, Appendix B. http://164.214.2.57/ prg_docs/documents/ MasterPlan/vol2/ vol2app_b_e.htm#I4

[15] Louis A. Fatale,James R. Ackeret, Jeffrey A. Messmore, Impact of Digital Terrain Elevation Data (DTED) Resolution on Army Applications: Simulation Versus Reality, U.S. Army Topographic Engineering Center, 1994 ASPRS/ACSM, http://wwwsgi.ursus.maine. edu/gisweb/spatdb/acsm/ac94100.html.

[16] Ed Wright, Warren Olsen, Evaluation of The Military Operations In Built-up Areas (MOBA) - Terrain DataBases (TDB), TASC, http://www.tasc.com/simweb/papers/166/full. htm.

[17] Lum, Zachary, "SIGINT Scores with Army Task Force XXI", *J. Electronic Defense*, May 1997, pages 52-53.

PAPER No: 35

DISCUSSOR'S NAME: G. Wyman

COMMENT/QUESTION:

Can you identify a priori the domain in which you will undertake the correlation?

AUTHOR/PRESENTER'S REPLY:

The domain of the correlation operation can be described by the following example sequence:

1. SIGINT detects a new threat emitter, Type X.

2. The SIGINT uncertainty ellipse area is detrermined and the EMINT sensor (e.g. a SAR on an unmanned air vehicle) is tasked to collect imagery for the area).

3. The emitter type (X) is used to infer the types of equipment (vans, shelters, antennas) expected. These are used to form a matched filter.

4. The matched filter is scanned across the imagery (automatic target recognition) to locate the equipment associated with emitter X in the imagery.

The correlation, therefore, is performed in the image domain, using a reference template inferred from the SIGINT detection.

DISCUSSOR'S NAME: A. Hume

COMMENT/QUESTION:

How does the proposed system cope with multiple emitters in the imager's view?

AUTHOR/PRESENTER'S REPLY:

The concept is focused on locating individual, high-value emitters within imagery. While many emitters are detected across the battlefields, the SIGINT system will select high-value, time-critical targets (e.g. mobile SAM's) and will cue IMINT to locate those individual targets.

PAPER NO: 35

DISCUSSOR'S NAME: G.S. Brown

COMMENT/QUESTION:

Does the use of spaceborne imaging assets have any impact on your fusion which was primarily aimed at airborne platforms use?

AUTHOR/PRESENTER'S REPLY:

Airborne high-altitude endurance (HAE) surveillance platforms (.e.g. Global Hawk and Darkstar) provide a capability for rapid revisit and immediate tasking that is not feasible with spaceborne assets, such as SPOT or Radarsat.

NOSTRAMARINE : UN CONCEPT DE DETECTION MULTISTATIQUE ADAPTÉ À LA SURVEILLANCE DES CIBLES BASSE ALTITUDE.

M. Lesturgie, M. Flécheux
Office National d'Etude et de Recherches Aérospatiales - ONERA
Chemin de la Hunière et des Joncherettes
91120 PALAISEAU - France

RESUME

Ce papier décrit un nouveau concept de détection transhorizon adapté à la surveillance des cibles basse altitude. Le radar, de configuration bistatique, est constitué d'un système de réception implanté au voisinage de la zone à surveiller, et d'un émetteur éloigné, situé en retrait de la zone à surveiller. Le fonctionnement global du radar repose sur l'association d'un mode de propagation '*ionosphérique* ' et d'un mode de propagation par '*ondes de surface*'.

Ainsi, grâce aux propriétés de réflexion des ondes électromagnétiques sur l'ionosphère, les cibles basse altitude peuvent être illuminées par l'émetteur éloigné, supposé fonctionner en hautes fréquences. Les ondes électromagnétiques, après interaction avec les cibles se propagent en suivant la courbure terrestre (par '*ondes de surface*') jusqu'au réseau de réception. Ce concept constitue une alternative au radar *transhorizon par ondes de surface* traditionnel (OTH-S), en présentant au niveau de la zone géographique à surveiller, l'avantage d'un dispositif passif, discret, et doté de capacités d'intégration intéressantes.

La *propagation* joue un rôle déterminant dans l'optimisation des caractéristiques de ce dispositif; aussi doit on mener un choix judicieux de la fréquence, en fonction des prévisions ionosphériques d'une part, et des conditions d'excitation des ondes de surface d'autre part.

Du point de vue système, l'émission peut provenir d'un émetteur radar (par exemple OTH 'Nostradamus') ou d'un émetteur de radiodiffusion puissant. Le fonctionnement en réception repose sur une *antenne réseau* à formation de faisceaux par le calcul à laquelle sont appliquées des techniques de filtrage adaptatif. Enfin, la *fusion*, en réception, des plots de détection obtenus à partir d'émetteurs de nature et emplacement différents (*multistatisme*) confère au dispositif des propriétés intéressantes en matière de détection et de localisation. Des résultats de simulation, illustrant les capacités de détection et l'intérêt d'associer les plots issus de plusieurs émetteurs éloignés, sont décrits. Ce concept de détection a conduit à des campagnes de mesures sur cibles réelles; un résultat obtenu à partir d'un émetteur de radiodiffusion non coopératif est présenté.

INTRODUCTION

La détection des cibles rasantes, depuis un site côtier ou un bâtiment de surface pose un problème de préavis d'alerte pour les radars classiques dont la visibilité 'quasi optique' permet tout juste de détecter les cibles à leur lever sur l'horizon. A moins de disposer le radar sur un point haut - ce qui n'est pas toujours envisageable - la portée basse altitude se trouve limitée à quelques dizaines de kilomètres, ce qui est faible compte tenu des vitesses potentielles des cibles et du temps de réponse limité des procédures de riposte à déclencher en cas d'alerte.

Les concepts de radars transhorizon fonctionnant en bande décamétrique (radar OTH-B et radar OTH-S) sont déjà anciens. Solution au problème de la détection des cibles basse altitude (malgré une précision de mesure nettement en retrait par rapport ce que fournissent les dispositifs hyperfréquences traditionnels) ce type de radar semble toujours présenter un intérêt. Les progrès en matière de traitement numérique du signal ont permis d'accéder aux performances optimales de ces dispositifs dont le fonctionnement est largement basé sur la formation de faisceaux par le calcul (FFC) et le traitement Doppler cohérent. D'autre part, la difficulté, en gamme HF (mais également en V-UHF) d'utiliser des matériaux absorbants capables de diminuer la réflectivité des cibles renforce l'intérêt porté de manière générale aux dispositifs fonctionnant en basse fréquence, dispositifs que l'on peut qualifier « d'antifurtifs ».

Le dispositif radar décrit dans ce papier appartient à la classe des radars transhorizon basés au sol et opérant en gamme HF. De configuration bistatique, il exploite d'une part l'émission d'une onde de ciel dirigée vers les cibles à détecter (émission en provenance d'un radar OTH ou d'un émetteur HF de forte puissance), il s'attache d'autre part à traiter de manière optimale les ondes de surface induites par les cibles évoluant au voisinage de la surface. Le concept exploite les modes de propagation bien connus en HF et notamment utilisés dans le radar transhorizon par onde de ciel

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

(OTH-B) et le radar transhorizon par ondes de surface (OTH-S). L'appellation 'Nostramarine' résulte d'une contraction de "*Nostradamus*" (radar transhorizon développé par l'Onéra pour le compte du SPAé à Dreux Senonches) et du "*Radar HF marine* (projet de radar à ondes de surface étudié par l'Onéra pour de compte de la DCN en 1989)".

Le principe de fonctionnement du dispositif Nostramarine est décrit dans la première partie de ce papier. La seconde partie traite des performances théoriques; des résultats de modélisation de la couverture basse altitude sont notamment présentés. Le traitement du signal est ensuite décrit. Enfin, dans la dernière partie, un résultat expérimental, obtenu à partir d'un émetteur d'opportunité, est présenté.

1. PRINCIPE DE FONCTIONNEMENT

1.1. Les radars transhorizon existants

Les radars transhorizon fonctionnant en gamme décamétrique sont de deux types:

- les radars transhorizon par réflexion ionosphériques (OTH-B) dont les portées de détection peuvent atteindre plusieurs milliers de kilomètres. Ces radars possèdent une zone de silence ("zone aveugle") de l'ordre de 500 km à 750 km.

- les radars à ondes de surface (OTH-S), fonctionnant exclusivement au dessus de la mer avec des portées de détection pouvant atteindre plusieurs centaines de kilomètres.

Les radars OTH-B sont employés pour des missions de surveillance aérienne très longue portée, également pour des missions de surveillance antibalistique. Les radars OTH-S sont employés pour des missions de surveillance transhorizon au dessus de la mer: surveillance vis à vis des missiles en pénétration basse altitude, également surveillance des bateaux(lutte contre le passage de la drogue,..). L'utilisation de tels dispositifs à bord de bâtiments de surface est également envisagée pour contrer la menace des missiles mer-mer et pour détecter de manière précoce la présence d'autres bâtiments de surface. En revanche l'importance des niveaux de puissance à émettre ainsi que la dimension des aériens d'émission constituent un point dur à l'emploi d'un tel radar à bord de bâtiments de surface.

1.2. Le concept "passif" Nostramarine

Le concept "Nostramarine" repose sur la mise en oeuvre, en association avec un émetteur transhorizon, d'un réseau de réception au voisinage de la zone de menaces à surveiller. La figure 1 décrit schématiquement le principe de fonctionnement du radar. Dans la zone à surveiller, évoluent des cibles qui sont éclairées par les *ondes de ciel* provenant d'un émetteur de forte puissance (radar transhorizon ou émetteur de radiodiffusion puissant). Les cibles diffusent l'énergie électromagnétique dans toutes les directions; si les cibles évoluent à basse altitude cette diffusion génère par couplage avec la mer des *ondes de surfaces* qui se propagent en suivant la rotondité terrestre, conférant au dispositif une capacité de détection transhorizon. Les ondes de ciel provenant de l'émetteur parviennent au voisinage de la cible avec une polarisation généralement elliptique; l'interaction des ondes incidentes avec la cible est complexe; en revanche seules les ondes polarisées verticalement se propagent le long de la surface de la mer vers le récepteur. Le réseau de réception peut être implanté en bordure de côte (surveillance côtière) ou à bord d'un bâtiment de surface.

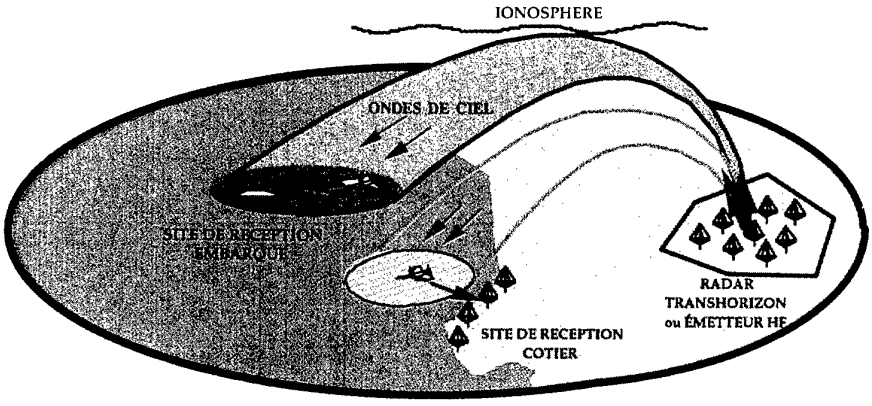


fig.1 : principe de fonctionnement du dispositif Nostramarine

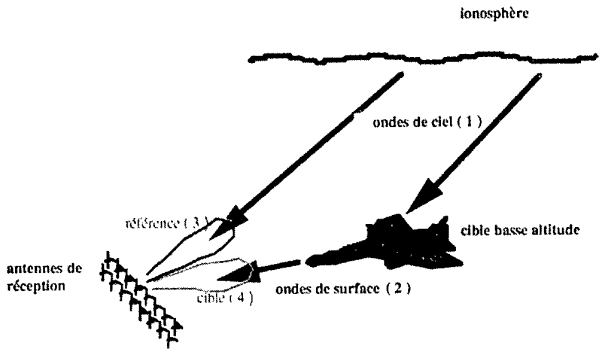


fig.2 : séparation des signaux en réception

Le rôle du réseau d'antennes est de séparer et sélectionner les ondes de ciel (trajet direct) provenant directement de l'émetteur (1) des ondes de surfaces provenant des cibles (2). La sélection des différentes ondes est faite par traitement adaptatif, les antennes du réseau étant couplées à une batterie de récepteurs et de codeurs permettant de procéder au traitement numérique des signaux. Le traitement, dont les bases seront détaillées par la suite consiste à former une voie de référence (3) dans la direction du trajet direct et une voie utile (4) dans la direction présumée des cibles.

Le fonctionnement du radar n'est effectif qu'après une phase de calibration à l'émission et à la réception. La calibration à l'émission comprend plusieurs opérations; dans le cas d'un fonctionnement couplé à un radar transhorizon, ces opérations sont, chronologiquement :

- le choix de la gamme de fréquence : le choix grossier de la fréquence (à 1 MHz près) est effectué à l'issue d'une prévision des conditions de propagation ionosphérique. Le choix de cette fréquence est ensuite affiné de manière à positionner précisément le faisceau d'émission sur la zone à surveiller (*pilotage du faisceau d'émission*); cette affinage conduit à déterminer la fréquence avec une marge de l'ordre de 100 à 200 KHz. A l'intérieur de cette bande de fréquence, on recherche alors une fréquence claire présentant le niveau de bruit le plus faible (*recherche de fréquence claire*).

- le pilotage du faisceau d'émission: dans le cas de l'utilisation d'un émetteur radar, au faisceau directif, une procédure de positionnement du faisceau est requise. Cette procédure peut être menée de façon entièrement adaptative, grâce à la conjugaison de phase. Dans ce cas, on dispose au niveau du réseau de réception (site local) un petit système d'émission de faible puissance (inférieure à 50 w), fonctionnant à une fréquence F_0. Le radar OTH reçoit sur chaque antenne, le signal à la fréquence F_0 et extrait après démodulation le vecteur directionnel des phases du signal reçu. Ce vecteur directionnel signe l'emplacement du réseau de réception Nostramarine et est alors utilisé en émission pour focaliser l'énergie du radar dans la direction du réseau local. La conjugaison de phase permet de compenser les erreurs de phases ionosphériques et garantit ainsi une focalisation optimale du faisceau d'émission. La liaison nécessitée par la calibration en émission doit se faire sur une fréquence F_o claire au niveau du radar OTH. L'émission du radar OTH vers le réseau local peut se faire à une fréquence F_1 différente. Cette solution est préférable: en effet, d'une part une recherche de fréquence claire unique et valable sur les deux sites est contraignante, d'autre part le changement de fréquence permet de déplacer la tâche focalisée au sol le long du segment OTH-Réseau local; ceci permet d'ajuster au mieux l'emplacement du faisceau d'émission sur la zone à surveiller. En augmentant légérement la frÈquence ($F_1 > F_0$), la tache au sol s'éloigne du site de réception. Cette situation schématisée par la figure 3 est intéressante dans un contexte d'alerte côtière.

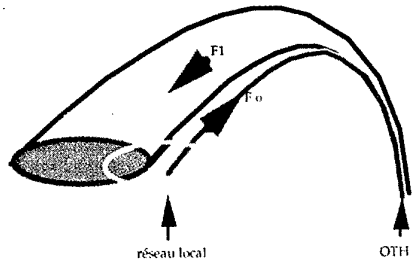


fig. 3 : pilotage du faisceau d'émission

- la recherche de fréquence claire : effectuée au niveau du site de réception, elle consiste à analyser des canaux de largeur comprise entre 5 et 30 KHz (selon la largeur de bande requise par l'émission radar) dans la gamme de fréquence déterminée après calibration du faisceau d'émission. La recherche de fréquence claire doit être fréquemment réactualisée car l'environnement radioélectrique fluctue de manière importante en gamme HF. Le

recours à une procédure de recherche de fréquence claire simultanée (ou entrelacée) avec le fonctionnement du radar constitue la solution idéale.

La calibration en réception est effectuée à partir du trajet direct en exploitant la connaissance a priori de la position de l'émetteur utilisé et les prévisions ionosphériques des conditions de propagation.

1.3. Avantages et inconvénients du dispositif Nostramarine

Le réseau déployé au voisinage du secteur à surveiller est entièrement passif. Ce dernier est par conséquent très peu vulnérable , notamment vis à vis des missiles antiradar. Les antennes peuvent être de dimensions réduites; l'utilisation de fouets actifs large bande (hauteur = 1 à 2 m) permet de concevoir un réseau facilement déployable ne nécessitant aucune infrastructure. Rappelons, qu'à l'inverse, un radar OTH monostatique utilisant des antennes d'émission accordées, présente des dimensions importantes (antenne log-périodique ou antennes biconiques de hauteur supérieure à 10 m). Enfin, l'implantation d'antennes de réception de petites dimensions est envisageable sur un bâtiment de surface (frégate..). Par ailleurs, le risque de perturbations radioélectriques causées par le radar sur les équipements du bâtiment est limité.

Enfin, contrairement aux radars OTH-S monostatiques, la configuration bistatique de Nostramarine lui permet de fonctionner avec des formes d'onde continues, et sans distance aveugle. (la distance aveugle, dans le cas de radar OTH-S monostatique est de l'ordre de 20 km.)

Le premier inconvénient du dispositif Nostramarine, si on le compare à un dispositif OTH-S (radar à ondes de surface) vient principalement des contraintes liées à la propagation ionosphérique. La distance entre l'émetteur et la zone à surveiller doit être comprise typiquement entre 800 et 2500 km. Par ailleurs, les fluctuations temporelles du milieu de propagation imposent une réactualisation fréquente des paramètres, notamment de la fréquence porteuse.

Le second inconvénient tient à la complexité du traitement de signal. Pour une exploitation optimale des capacités de détection et localisation du dispositif, des procédures de filtrage adaptatif angulaire doivent être mises en oeuvre. Toutefois, la charge de calcul supplémentaire, comparée à ce que requiert le traitement classique d'un radar OTH-S , ne doit pas être considérée comme un obstacle, eu égard aux performances sans cesse croissantes des processeurs de calcul disponibles actuellement sur le marché.

2. PERFORMANCES

2.1. Influence de la propagation et de l'environnement

La propagation et l'environnement jouent un rôle déterminant dans l'évaluation des performances du radar Nostramarine.

Les conditions de propagation par *onde de ciel* , ainsi que la contrainte de recherche d'une fréquence claire imposent dans la pratique le choix d'une largeur de bande de fonctionnement.

La propagation par *onde de mer*, le long du trajet cible-récepteur ne nécessite aucune prédiction particulière. Le fonctionnement en polarisation verticale est en revanche nécessaire. Les pertes de propagation présentent un comportement favorisant l'utilisation de basses fréquences. Au delà de 15 MHz, et pour des distances importantes (> 300 km) les pertes deviennent très importantes (> 30 à 50 dB). L'état de la mer, par la rugosité de surface qu'il induit, tend à augmenter encore les pertes de propagation, et ce, d'autant plus que la fréquence de fonctionnement est élevée (notamment au delà de 10 MHz).

2.2. Modélisation

Les modèles de propagation utilisés pour l'évaluation des performances théoriques sont d'une part le modèle de Bradley-Dudeney pour la propagation ionosphérique [1], d'autre part le modèle de Bremmer pour la propagation par onde de mer[2,3,4] . Le modèle Bradley-Dudeney consiste en un tracé de rayon à partir d'un modèle de profil d'indice de réfraction à 2 couches paraboliques. Le modèle de Bremmer procède au calcul de la série des résidus du potentiel de Hertz, au dessus d'un interface sphérique lisse. L'état de la mer et la rugosité des vagues sont prises en compte sous la forme d'une impédance de surface modifiée, d'après les travaux de Barrick[5]. A titre d'exemple, la prise en compte d'un état de mer 5 sur une propagation par ondes de surface opérant à 15 MHz conduit à des pertes supplémentaires de 6 dB à 70 km.

Le bruit radioélectrique, quantifié par la notion de facteur de bruit, est délicat à modéliser. Le bruit fluctue dans le temps, et ses propriétés dépendent de l'emplacement géographique; les zones rurales sont à privilégier devant les zones urbaines qui sont la source de bruits industriels contribuant à une augmentation significative du facteur de bruit ambiant. Des différences de facteur de bruit de 10 à 20 dB entre un site 'calme' et un site 'bruyant' sont courantes en gamme HF. En

environnement calme, le facteur de bruit moyen peut être modélisé simplement par l'expression suivante: F(dB) = 56 - 25 $\log_{10}(f_{MHz})$. Cette expression valable en dessous de 30 MHz pour une antenne courte non directive, est cohérente avec les informations connues du CCIR .

2.3. Bilan énergétique

En l'absence d'interférences, le bilan énergétique pour le radar Nostramarine s'écrit:

$$\left[\frac{S}{B}\right]_{Nostramarine} = \frac{P_e \cdot G_e \cdot G_r \cdot \lambda^2 \cdot \sigma \cdot T_{int}\,\alpha}{(4\pi)^3 \cdot D_{iono}^2 \cdot D_{os}^2 \cdot L_{iono} \cdot L_{os} \cdot FKT_0}$$

En préambule à un examen détaillé des performances, il est intéressant de présenter les ordres de grandeur des paramètres:

f = c/ l	Fréquence	12 MHz
F	Facteur de bruit extérieur	30 dB
$P_e.G_e$	PIRE OTH	10 MW
α	Facteur de forme	1
D_{iono}	Distance ionosphérique	1500 km
D_{os}	Distance cible - réception	100 km
L_{iono}	Pertes onde de ciel	13 dB
L_{os}	Pertes onde de sol	15 dB
G_r	Gain réception (Nostramarine) directivité	15 dB 20 °
σ	SER de la cible	10 m^2
T_{int}	Durée d'intégration	4 sec.

Le rapport signal à bruit est alors égal à :

S/B [Nostramarine] = 18 dB

Il est intéressant de comparer ce résultat au bilan énergétique que l'on obtiendrait à l'aide du radar OTH-B utilisé de manière conventionnelle, ou à l'aide d'un radar OTH-S monostatique. Les données complémentaires suivantes sont nécessaires (ces données sont toujours à considérer comme des ordres de grandeur):

$P_e^{'} \cdot G_e^{'}$	PIRE - OTH-S	100 kW
α	Facteur de forme	¼
$G_r^{'}$	Gain réception - OTH-B directivité	28 dB 3 °

$$\left[\frac{S}{B}\right]_{OTH-S} = \frac{P_e^{'} \cdot G_e^{'} \cdot G_r \cdot \lambda^2 \cdot \sigma \cdot T_{int}\,\alpha}{(4\pi)^3 D_{os}^4 \cdot L_{os}^2 \cdot FKT_0} = 14dB$$

$$\left[\frac{S}{B}\right]_{OTH-B} = \frac{P_e \cdot G_e \cdot G_r^{'} \cdot \lambda^2 \cdot \sigma \cdot T_{int}\,\alpha}{(4\pi)^3 \cdot D_{iono}^4 \cdot L_{iono}^2 \cdot FKT_0} = 4dB$$

Sur le plan énergétique, Nostramarine est équivalent à un radar monostatique dont la puissance moyenne rayonnée serait de l'ordre de la centaine de kW. Outre l'économie d'un système d'émission encombrant (antennes accordées nécessitant une infrastructure lourde), le dispositif Nostramarine peut être conçu à partir d'un système d'antennes de réception légères, facilement transportables et déplaçables. Par ailleurs les performances obtenues avec le radar OTH-B sont très en retrait des performances obtenues avec Nostramarine. En effet, la différence de distance, le long du trajet retour [cible-récepteur], soit 1500 km/ 100 km n'est que partiellement compensée par la différence de gain en réception (28 dB/15 dB). Dans le cas considéré, les pertes de propagation sont également plus élevées sur le parcours de l'onde de ciel que sur le parcours de l'onde de sol.

Pour ce qui concerne la localisation des cibles, les résolutions en distance sont comparables, qu'il s'gisse des radars OTH-S, OTH-B ou OTH-S/B (Nostramarine). La résolution transversale est liée à la résolution azimutale du radar et à la distance des cibles :

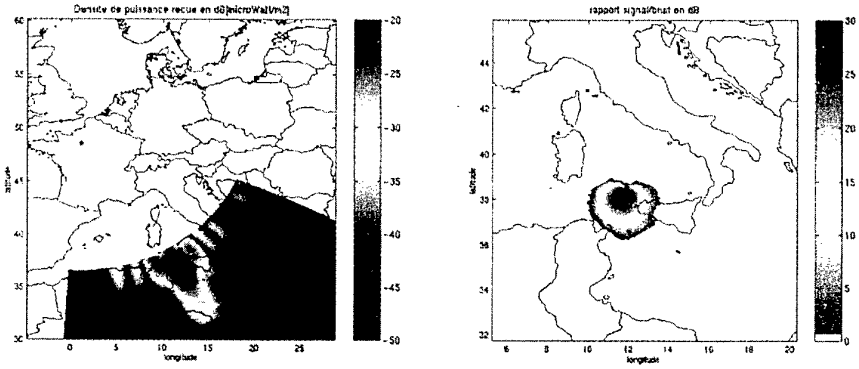
- de l'ordre de 20° à 100 km pour les OTH-S et Nostramarine, soit 35 km,
- de l'ordre de 3° à 1100 km (distance au sol) pour les OTH-B, soit 50 km.

2.4. Modélisation de la couverture

Un logiciel de calcul de la couverture radar selon le procédé Nostramarine a été mis au point à l'Onéra. Les figures 4 et 5 fournissent quelques exemples de couvertures simulées. Les couvertures sont calculées à partir d'une modélisation du rapport signal à bruit en environnement calme; la vitesse des cibles considérées est telle que les échos de fouillis n'en perturbent pas la détection. L'effet de secteur aveugle, qui peut apparaître du fait de l'élimination angulaire du trajet direct n'a pas, pour l'instant été modélisé.

La figure 4 est relative à l'utilisation d'un émetteur OTH de type Nostradamus implanté à l'ouest de Paris. Le dispositif de réception est implanté sur la côte méditerranéenne, à proximité de Toulon. Les caractéristiques de la simulation sont résumées dans le tableau ci dessous. La portée théorique sur cible aérienne est de l'ordre de 170 km.

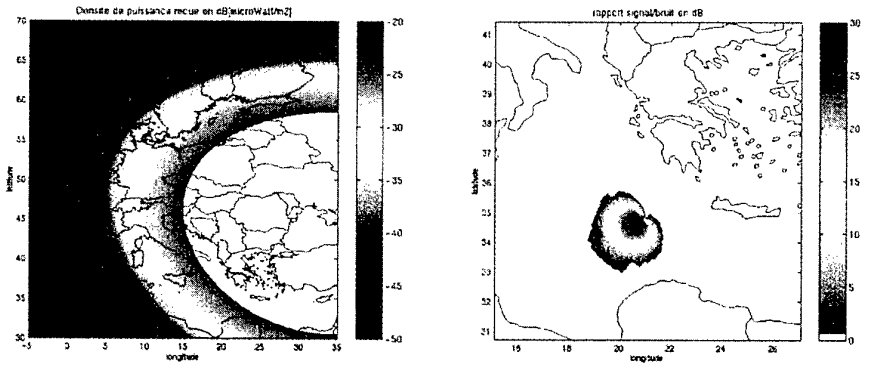
La figure 5 est relative à l'utilisation d'un émetteur HF non coopératif. Ici, l'émetteur de Simféropol permet d'obtenir des portées de l'ordre de 115 km sur cible aérienne.



(a) *(b)*

fig. 4 : exemple de couverture radar obtenue à l'aide d'un émetteur OTH

- (a) : densité de puissance incidente ; (b) : couverture radar (portée de l'ordre de 170 km)
- fréquence : 14.00 MHz, puissance 10 MW,
- réseau de réception à 12 antennes (secteur 80°), intégration cohérente sur 10 secondes
- cible de SER 2 m^2, altitude 30 m, mer calme
- caractéristiques de l'ionosphère : hauteur de la couche E : 100 km (F_c=3 MHz)
hauteur de la couche F : 320 km (F_c=10 MHz)



(a) *(b)*

fig. 5 : exemple de couverture radar obtenue avec l'émetteur de radiodiffusion de Simféropol

- (a) : densité de puissance incidente ; (b) : couverture radar (portée de l'ordre de 115 km)
- fréquence : 15.38 MHz, puissance 1MW, (heure TU : 15H00)
- réseau de réception à 12 antennes (secteur 80°), intégration cohérente sur 10 secondes
- cible de SER 5 m^2, altitude 15 m, mer calme
- caractéristiques de l'ionosphère : hauteur de la couche E : 100 km (F_c=3 MHz)
hauteur de la couche F : 320 km (F_c=10 MHz)

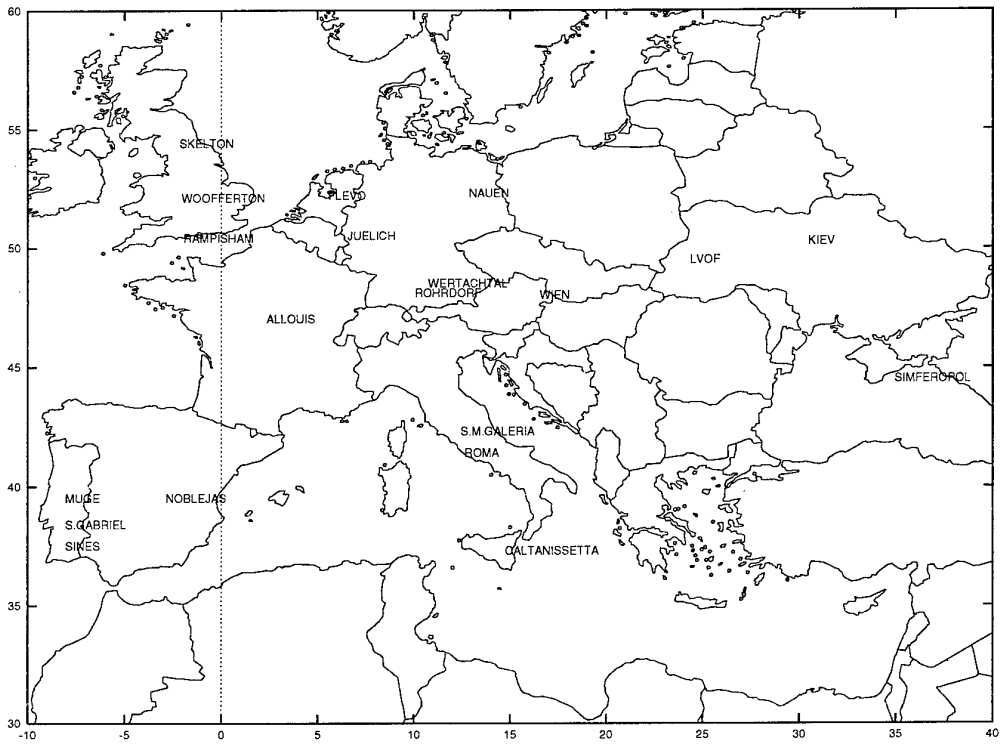


fig. 6 : Exemple de répartition d'émetteurs non coopératifs (PIRE > 100 Kw - octobre 96)

La figure 6 montre une carte des émetteurs de radiodiffusion puissants (PIRE> 100 kW) et disponibles à une période donnée. Le fonctionnement en réception sur plusieurs fréquences (fréquences d'émetteur OTH coopératifs et fréquences d'émetteur de radiodiffusion) confère au dispositif Nostramarine des propriétés multistatiques. La détection des cibles à partir de plusieurs émetteurs va permettre de confirmer les plots de détection et affiner la localisation des cibles.

3 . TRAITEMENT DU SIGNAL

3.1. Traitement de détection

Le traitement des signaux reçus consiste à séparer les contributions du trajet direct (ondes de ciel) et des trajets diffus présumés contenir les échos de cibles mobiles.

Du fait de la configuration bistatique, la forme d'onde radar peut être continue; la bande du signal émis est en revanche limitée à la largeur de bande disponible autour de la fréquence de fonctionnement du radar. Typiquement, des largeurs de bande comprises entre 10 et 30 kHz peuvent être obtenues, et ce, d'autant plus facilement que la fréquence porteuse se situe dans le haut de la bande décamétrique (> 15 MHz).Le traitement en réception comprend, après le codage et le filtrage de Hilbert des signaux, le traitement d'antenne, la compression d'impulsion et l'analyse Doppler. La figure 7 présente le synoptique général du traitement .

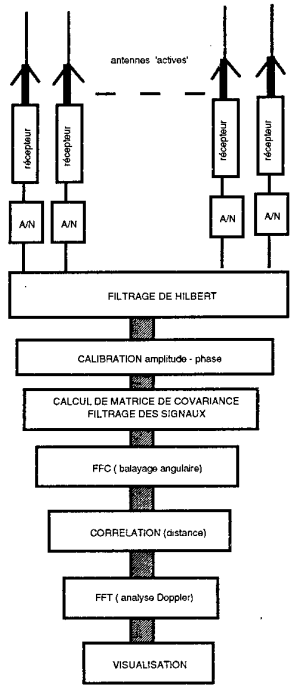


fig.7 : synoptique général du traitement de signal

Le traitement d'antenne est nécessaire pour détecter et localiser les cibles; son rôle est de diminuer la puissance du trajet direct incident afin de restituer au voisinage de la cible un contraste suffisant.
Le traitement angulaire est de type adaptatif, la mise en place d'un "zéro" dans la direction du trajet direct permettant d'améliorer de 30 à 40 décibels le taux de contraste de la cible. Le bruit de phase, qui accompagne inévitablement les signaux émis est également éliminé par le traitement angulaire. Si l'on note $S_n(t)$ le signal reçu sur le capteur n, alors le traitement adaptatif angulaire consiste à déterminer le vecteur des pondérations $W = \{w_n\}$ de façon à maximiser en sortie le rapport signal à [bruit + interférences]. Ce problème est un problème classique de minimisation sous contrainte que l'on peut exprimer sous la forme:

$$\min(W^H R W) \quad \text{avec} \quad R = \langle S_b^* S_b \rangle$$
$$W^t D = 1$$

où D est le vecteur directionnel de la cible et R la matrice de covariance des interférences (notées S_b). La solution à ce problème de minimisation est :

$$W = \frac{R^{-1} D^*}{D^t R^{-1} D^*}$$

Le vecteur D étant spécifique d'une direction particulière, le balayage angulaire en mode veille impose de calculer pour chaque valeur angulaire θ:

$$W(\theta) = \frac{R^{-1} D^*(\theta)}{D^t(\theta) R^{-1} D^*(\theta)}$$

La matrice de covariance R est estimée à partir des données elles mêmes, car les signaux de bruit, notés S_b ne sont pas disponibles séparément. La matrice est calculée en moyennant sur un nombre de récurrences radar N_r (N_r>2N dans la pratique) les produits de corrélation de signaux prélevés de préférence dans une case distance différente de la case distance présumée de la cible. Cette précaution permet d'éviter de dégrader le niveau de la cible lorsqu'elle rentre dans l'estimation de la matrice de covariance. En revanche, si la cible se trouve dans la direction du trajet direct, elle se trouve éliminée. L 'élimination angulaire du trajet direct ionosphérique peut ainsi conduire à un secteur aveugle en détection. Ce phénomène peut être contourné en utilisant un réseau surfacique (présentant de la directivité en site) ou, plus simplement, un réseau à 2 rideaux d'antennes correctement orientés. Dans le cas représenté sur la figure 8 (cas b), l'élimination du trajet direct à site et gisement non nuls ne conduit à aucun secteur aveugle à site nul.

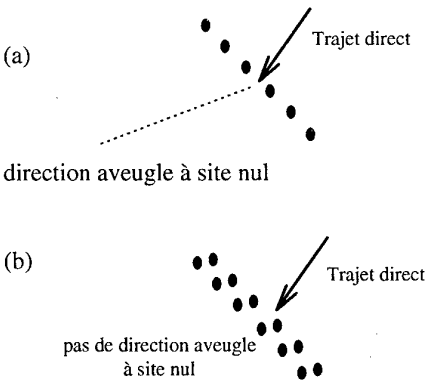


fig. 8 : choix d'une configuration permettant de limiter l 'effet de secteur aveugle

La compression d'impulsion, conventionnelle, repose sur la corrélation des signaux reçus avec une réplique du code émis. La réplique utilisée peut également provenir du trajet direct lui même, lorsque ce dernier présente un bon rapport signal à bruit. Cette technique permet de compenser une grande partie des dérives ionosphériques (retard de groupe et doppler), dont on peut faire l'hypothèse qu'elles sont semblables le long du trajet direct et des trajets émetteur-cibles dans la zone locale considérée. Cette compensation automatique évite la correction a posteriori des biais ionosphériques.

La compression d'impulsion est suivie d'une analyse Doppler par FFT, technique classique et peu coûteuse en temps de calcul. Les signaux, avant analyse Doppler, sont pondérés (fenêtre de Hanning ou de Blackman). L'analyse Doppler permet de séparer les échos de cibles mobiles du trajet direct résiduel et du fouillis. Ce dernier, après analyse Doppler fine (intégration cohérente sur plusieurs secondes) présente autour de zéro un étalement de l'ordre du Hertz. Le traitement Doppler améliore le contraste entre les cibles mobiles et le fouillis + trajet direct, d'environ 60 à 70 dB.

L'amélioration globale du contraste entre les échos de cibles et les signaux à Doppler nul est ainsi de l'ordre de 100 dB, ce qui compense parfaitement la dynamique initialement élevée entre les signaux utiles et les signaux parasites.

3.2. Association de plusieurs détections

L'utilisation simultanée de plusieurs émetteurs est envisageable. Les émetteurs de radiodiffusion puissants sont utilisables grâce à leur porteuse stable, propice à une analyse Doppler cohérente. En revanche, l'absence de 'code radar' (ou forme d'onde récurrente) compromet la localisation en distance des cibles. Les plots de mesure sont donc

constitués de l'azimut et du Doppler de la cible relatif à chacun des émetteurs envisagés. L'association des plots (et plus précisément des Doppler) en vue de la localisation plane des cibles a déjà été utilisée dans le cas de la détection à l'aide des émetteurs de télévision analogique [6] . Le principe repose sur la résolution d'un système d'équations non linéaires reliant les fréquences Doppler de la cible ses coordonnées planes et son vecteur vitesse:

$$-\lambda_1 f_{d_1} = V_x A^1_{x,y} + V_y B^1_{x,y}$$

$$-\lambda_2 f_{d_2} = V_x A^2_{x,y} + V_y B^2_{x,y}$$

$$-\lambda_n f_{d_n} = V_x A^n_{x,y} + V_y B^n_{x,y}$$

avec

$$A^n_{x,y} = \frac{x - x_{en}}{d(C,E_n)} + \frac{x - x_r}{d(C,R)}$$

$$B^n_{x,y} = \frac{y - y_{en}}{d(C,E_n)} + \frac{y - y_r}{d(C,R)}$$

avec :

(x,y)	: coordonnées de la cible C
(V_x,V_y)	: vecteur vitesse de la cible
(x_{en},y_{en})	: coordonnées de l'émetteur n
(x_r,y_r)	: coordonnées du site de réception R
$d(M_1,M_2)$	: distance de groupe entre M_1 et M_2

Dans le cas d'une localisation plane (x,y), l'utilisation de 4 émetteurs est nécessaire. Pour que la localisation soit correcte, il est indispensable de séparer au préalable les cibles en vue d'associer les Doppler correspondant à la même cible. Ceci est envisageable grâce aux capacités de séparation angulaire du réseau de réception

4. RESULTATS EXPERIMENTAUX

Des campagnes de mesures ont été menées dans le but de démontrer la faisabilité du procédé Nostramarine. Des essais ont d'abord été conduits à l'Onéra-Palaiseau à l'aide de 3 antennes, en présence d'avion civils (Orly) en utilisant un émetteur de radiodiffusion situé à Madrid; ces essais ont permis de valider les différentes étapes du traitement de signal. Des campagnes de mesure spécifiques ont eu lieu sur le site de Losquet (contrat DRET), en coopération avec le CNET, et au CEM (contrat SPAé). Au cours de ces deux campagnes de mesures, les capacités de détection au delà de l'horizon du dispositif Nostramarine ont pu être étudiées, les émetteurs utilisés étant Nostradamus ou des émetteurs de radiodiffusion puissant.

La figure 9 présente un résultat de détection (piste Doppler-temps) obtenu depuis le site de Losquet (CNET) à partir de l'émetteur de radiodiffusion de Kiev. La fréquence porteuse est de l'ordre de 15 Mhz; la cible évolue à une altitude de 300 pieds à une distance de 75 km. Le réseau de réception comprend 8 antennes isotropes en gisement à polarisation verticale. L'horizon radioélectrique (distance maximale de visibilité d'un radar conventionnel) est de l'ordre de 45 km.

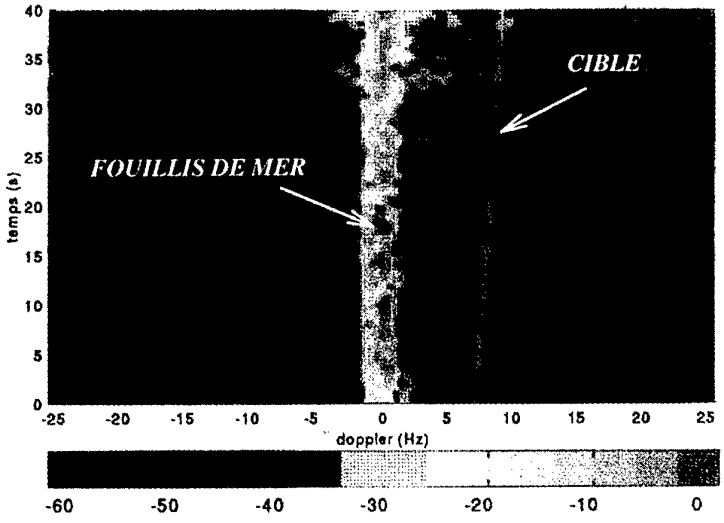


fig.9 : exemple de résultat de détection transhorizon (émetteur de radiodiffusion de Kiev)

CONCLUSION

Alternative au radar transhorizon à ondes de surface, le dispositif Nostramarine présente au niveau de la zone à surveiller l'attrait d'un système passif, peu vulnérable et facilement déployable. Le fonctionnement suppose de combiner un mode de propagation par onde de ciel (pour éclairer les cibles basse altitude) et un mode de propagation par onde de surface. Le fonctionnement en liaison avec un émetteur radar OTH ou un émetteur de radiodiffusion est donc nécessaire. L'opportunité fournie par la multiplicité des émetteurs de radiodiffusion puissants (puissance pouvant atteindre le mégawatt) a été introduite d'un point de vue multistatique. Le système de réception et de traitement peut être agencé de manière à traiter simultanément les signaux provenant de plusieurs émetteurs. Les émetteurs radar (OTH-B), par leur forme d'onde spécifique, permettent une localisation en distance; les émetteurs de radiodiffusion ne permettent pas de localiser en distance les cibles. Par contre, un traitement d'association des fréquences doppler, obtenues sur une même cible à partir de plusieurs émetteurs d'opportunité devrait permettre d'envisager une localisation plane de cette dernière.

Plusieurs campagnes de mesures ont déjà été menées dans le but de valider le fonctionnement du dispositif Nostramarine. Un résultat de détection obtenu à l'aide d'un émetteur de radiodiffusion a permis d'illustrer les capacités de détection transhorizon du radar.

Au delà de ses performances de détection transhorizon, Nostramarine est un radar passif reconfigurable, pouvant tirer partie du multistatisme, et dont l'intérêt repose sur une faible vulnérabilité, un déploiement rapide et discret d'antennes légères. Enfin, le coût du dispositif est également faible, comparé aux systèmes radar traditionnels: le coût est conditionné pour une grande partie par le calculateur de traitement de signal, ce dernier incluant des processeurs classiques, disponibles sur étagère avec un rapport coût/performances évoluant favorablement au cours du temps.

REMERCIEMENTS

Les travaux concernant l'étude du concept Nostramarine et les premières expérimentations associées ont été menés sur contrat DRET. Les expérimentations sur cibles réelles réalisées à l'Ile de Losquet, se sont déroulées avec le concours du CNET-Lannion.

REFERENCES

[1] J.D Milson, B. Sc - Exact ray path calculations in a modified Bradley/Dudeney model ionosphere - IEE proceedings, Vol 132, N° 1, feb. 85

[2] H. Bremmer - Terrestrial radio waves. Elsevier Pub. Co. (1979)

[3] M. Lesturgie - Contribution des gammes métrique et décammétrique au concept de radar de veille: performances à site bas et comparaison avec les radars classiques. AGARD - 47th EPP symposium - 1990

[4] P. Golé - Propagation des ondes électromagnétiques VHF. Recherche aérospatiale - Revue n° 1983-1

[5] D.E. Barrick - The interaction of HF/VHF radio waves with the sea surface ands its implications. Electromagnetics on the sea. AGARD conf. Proc. 77 - 1983-1

[6] M. Lesturgie, C. Delhote - Radar sans émetteur dédié
Entretiens "Sciences et Défense 1996" - Paris 24-25 janvier 1996

PAPER No. 36

DISCUSSOR'S NAME: C. Goutelard

COMMENT/QUESTION:

Vous utilisez la conjugaison de phase pour focaliser votre faisceau, émis d'un OTH, sur le site de réception. C'est, je pense, une méthode très bonne mais dans le cas présent, l'ionosphère, le milieu de propagation est non-réciproque et de plus la propagation est multimode. Avez-vous examiné l'influence de ces effets?

(You use phase conjugation to focus your beam, emitted by an OTH, on the receiving site. I think this is a very good method, but in the present case the ionosphere, the propagation medium, is non-reciprocal and what is more, the propagation is multimode. Have you looked at the influence of these effects?)

AUTHOR/PRESENTER'S REPLY:

La comparaison de phase n'a pour l'instant pas été testée expérimentalement dans le cadre de transmissions HF ionosphériques. En présence de multitrajets de niveaux différents, la technique de conjugaison de phase conduit à sélectionner le trajet dont le niveau et prépondérant. Dans la pratique, le faisceau est, pour le moment, postionné de manière manuelle, et la détermination d'un trajet prépondérant ne semble pas poser de problème. On peut ainsi espérer que la conjugaison de phase, qu'il reste à implementer, sera efficace dans la majorité des cas. Mais les problèmes que vous citez, spécifiques du milieu ionosphériques doivent être pris en compte.

(For the moment, phase comparison has not been experimentally tested in the framework of HF ionospheric transmissions. In the presence of multipaths of different levels, the phase conjugation technique involves selection of the path with the predominant level. In practice, the beam is for the moment positioned manually and the determination of the predominant path does not seem to pose any problem. Hopefully then, phase conjugation, which remains to be implemented, will be effective in most cases. However, the problems which you mention, which are specific to the ionospheric medium, should be taken into account.)

THE NATO DATA FUSION DEMONSTRATOR PROJECT

J M Skinner
Command Systems Department
Defence Evaluation and Research Agency
Malvern, WR14 3PS, UK

SUMMARY

This paper defines data fusion in the context of the land battle and the military intelligence cycle, it describes the NATO Data Fusion Demonstrator concept and the key aspects of the current design and concludes with a description and evaluation of the results achieved to-date.

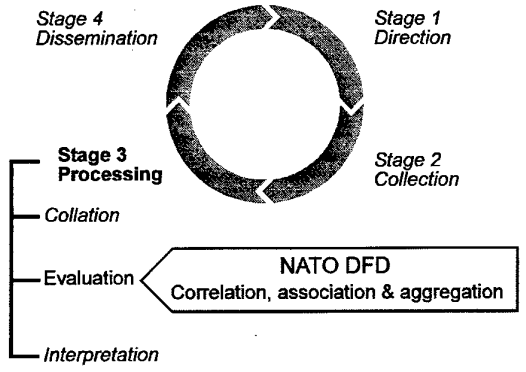
1 INTRODUCTION

1.1 The Problem

Land commanders at all levels need timely and accurate intelligence on the enemy's locations, strength, activities, subordination and intentions in order to manage their battle space effectively and get within the Opposing Forces (OPFOR) reaction time. The ever increasing amount of sensor data, Initial Processing Site (IPS) information together with information and intelligence from subordinate, peer and superior commands as well from other sources and agencies now makes it increasingly difficult for manual All Source Cells (ASC) to process this wealth of information sufficiently quickly and accurately to meet their commander's intelligence needs.

1.2 The Intelligence Cycle

Figure 1. The Intelligence Cycle



The processes carried out by an ASC are defined by the Intelligence Cycle shown in Figure 1. The first stage, Direction, consists of receiving the commander's intelligence requirements, determining the information needed to fulfil those requirements, preparing a collection plan, tasking sources and agencies and monitoring their activity. The second stage, Collection, consists of the exploitation of the sources of information by the collection agencies and the delivery of the collected information. The third stage, Processing, consists of collation, evaluation and interpretation. The DFD is concerned only with the evaluation phase which may be sub-divided into the processes of classification, correlation, association and aggregation. Interpretation, which includes predicting the OPFOR's objectives and future courses of action is outside the scope of the current project. Dissemination, the fourth stage of the cycle, consists of providing the commander and others with the intelligence they have requested and need to fulfil their roles effectively.

1.3 The Definition of Data Fusion

For the purposes of this paper, data is defined as the output of a transducer or sensor produced by a change in its surveillance environment. Information is the classification of that data by the application of collateral knowledge and inferencing in order to recognise or identify the detected entity, activity or occurrence. Finally, intelligence is produced by interpreting the significance of that information by the application of inferencing together with collateral knowledge of the enemy commander's Order of Battle (ORBAT), doctrine and potential deception tactics together with deductions of how his activities may be constrained by the terrain and weather conditions,.

As can be seen from the foregoing, the application of the term data fusion to this project could be regarded as a little misleading as the DFD is concerned primarily with the fusion of information from IPS and other sources rather than with the fusion of raw sensor data.

Nevertheless data fusion is the commonly accepted term for the processes of combining data and or information. Even so there are still many different definitions of data fusion. However, in the land military context, the product of data fusion should be regarded as an integral and vital component of the intelligence cycle and hence battle management and, for the purposes of the DFD project, (Ref.1) is defined as:-

> ***The process of creating a timely, consistent and effective, task orientated perception of the world from all source data and information which may be incomplete and/or inaccurate and/or misleading and/or untimely.***

Data fusion is achieved by some or all of a series of processes in which the input data or information is collated, classified, correlated, associated and aggregated to provide a reduced data set whose information content is greater than that of its constituent parts. Thus the simultaneous display of data is not strictly data fusion, nor is the process of limiting or discarding data in an arbitrary process. The essence of data fusion is data reduction with synergy.

The DFD definitions of the processes which go to make up data fusion are based on current British and NATO usage and are as follows.

Collation. The grouping together of related items of information to provide a record of events and facilitate further processing. Collation is the precursor to classification and only involves collecting similar sets of data and information from similar sources.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

Classification. The inferencing process which, using collateral information, classifies or recognises or indentifies the reported entity, activity or occurrence.

Correlation. The process which combines data and information from similar sources about the same entity, activity or occurrence.

Association. The logical blending of related but dissimilar sets of correlated information, from a variety of dissimilar sources, about the same entity, activity or occurrence. This is the process which some workers call data fusion. It may also be called picture fusion when the process combines the separate ELINT, COMINT, IMINT and HUMINT pictures.

Inferencing. The use of collateral knowledge to deduce or infer additional information about an entity or an activity or occurrence. Inferencing can occur in all the processes of data fusion and in conjunction with correlation and association can confirm or improve the recognition or identification and location of an entity or class of activity or occurrence.

Aggregation. The process which combines or aggregates those associated entities that are related to each other in order to form parent entities, similarly, the process will aggregate parents to form grandparents and so on to produce and identify the enemy's hierarchical structure or dynamic Order of Battle (ORBAT) within the fusion system's areas of responsibility and interest. Thus the presence of an HQ entity and three company sized entities with the appropriate characteristics and within sensible distances of each other would lead the aggregation process to deduce the existence of a particular battalion or type of battalion. This in turn, using the collateral Order of Battle and doctrine databases may give rise to hypotheses about the existence and role of the parent regiment.

1.4 The Objective of the NATO DFD Project

The prime purpose of the DFD is to alleviate the intelligence analysts' work loads by demonstrating the feasibility of providing machine assistance for the creation of the Opposing Forces (OPFOR) Perceived Ground Truth or Situation, that is the OPFOR dispositions, activities and subordination, by using Artificial Intelligence (AI), Knowledge Based Systems (KBS) and other techniques to fuse the messages from representative ELINT, COMINT, IMINT and HUMINT IPS and other sources.

Consequently, in the context of the DFD project, data fusion is confined to the creation of the Perceived Situation, that is the system's perception of the ground truth. Automatic or machine assisted analysis of the perceived situation in order to predict future activities and objectives is outside the scope of the current project.

1.5 The NATO DFD Environment

The simulation environment and role chosen for the demonstrator is that of a divisional ASC, driven by machine readable ADatP-3 messages (Ref.2) from representative COMINT, ELINT, IMINT and HUMINT Initial Processing Sites. Each simulated IPS has a sensor model or models deployed, to an approved collection plan, in a simulated land scenario that depicts an unopposed divisional advance accompanied by guerrilla deception activity. The scenario (Ref.3) covers a 24 hour period which commences with the movement of the Orange Forces out of their assembly area, their advance along two routes, an unopposed river crossing and concludes with their deployment into combat battle formation.

The initial phase of the advance is conducted in radio silence and the Blue Commander's information gathering capability is further hindered as he is not allowed to deploy any sensors or reconnaissance elements across the international boundary.

1.6 The Project Participants

The project is governed by a Memorandum of Understanding (MOU) (Ref.4) which was signed by Canada, Denmark, France, Germany, Italy, The Netherlands and the UK in September 1992. However, Italy were unable to commence work until early 1996 and France ceased work in April 1996. These two events have necessitated some re-structuring and down sizing of the project. The participants are currently negotiating an extension to the MOU to enable the final phase to be completed by Italy.

1.7 The DFD Programme

The project was conceived as a five year plan comprising of three evolutionary phases, Mockup, Prototype 1 and Prototype 2. The Mockup phase, now completed, used a preliminary scenario script and provided the knowledge and data bases, the Human Computer Interface (HCI) and the infra-structure.

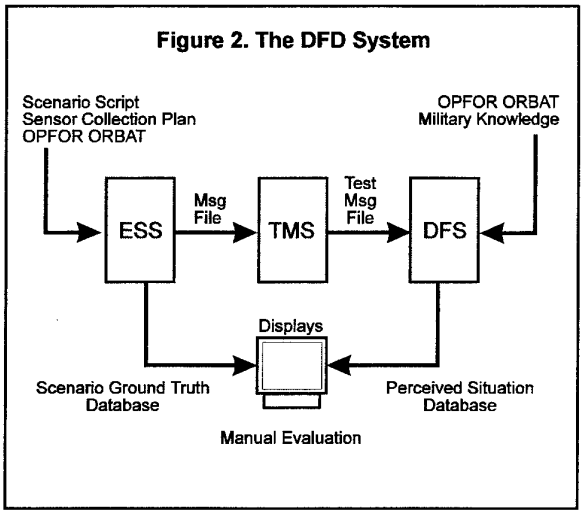
Prototype 1, due to be evaluated in October this year, provides the greater part of the functionality of the final system, including all the fusion algorithms, and uses the full International Scenario Script (ISS), the Simulated Ground Truth (SGT) data base and the ADatP-3 message file.

Prototype 2, now due in mid 1998, will provide the final system which will include updates to Prototype 1, an additional classification algorithm, based on Bayesian networks, by DDRE, Denmark, a German aggregation algorithm and the additional sets of fusion algorithms to be provided by the Italian participants.

1.8 The DFD System

To achieve the required functionality the DFD is divided into three systems or segments as shown in Figure 2.

Figure 2. The DFD System



The first segment, the Event Simulation System (ESS) (Ref.5) simulates the scenario, the sensor models and their deployment

and produces the ADatP-3 IPS message file and the SGT data base. The second segment, the Test and Monitor System (TMS), controls the tests and passes the message file to the DFS. It was also intended that the TMS would have had the facilities to compare the perceived situation produced by the DFS with the Simulated Ground Truth produced by the ESS. However, the withdrawal by France, who was responsible for the TMS, has resulted in these evaluation facilities being abandoned or at least, delayed until Prototype 2, consequently, the TMS will not be described any further in this paper. The third segment, the Data Fusion System (DFS) (Ref.6) processes and fuses the information contained in the message file and produces and displays the resultant Perceived Situation. In addition to these segments there is a fourth system, the Communications and Supervision System (CSS) which manages the TMS and DFS.

1.9 The DFD Development and Target Environment

The major components of the agreed development and target environment for the project are:

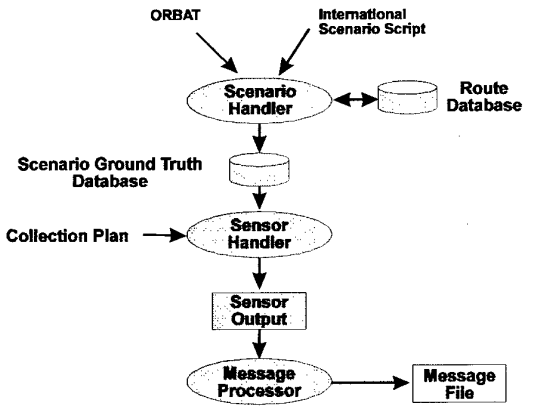
(a) Sun SPARCstation 10/41
(b) Solaris 2.5 operating system
(c) C++
(d) Sybase V4.9.2 and Objectstore 4.00c databases
(e) TerraLogic GIS
(f) IlogViews HCI support
(g) IlogRules

2 THE EVENT SIMULATION SYSTEM (ESS)

2.1 The ESS Architecture

The roles of the ESS are twofold, firstly, to provide the messages necessary to exercise the DFS and secondly, to provide the Scenario Ground Truth database, the reference against which the DFS's output, the Perceived Situation database is assessed. Figure 3 shows the three major components and processes of the ESS; the Scenario Handler, which creates the detailed scenario and populates the SGT database, the Sensor Handler which deploys the sensor models in the simulated scenario and creates sensor target data and the Message Processor which converts the target data into information formatted in ADatP-3 IPS messages.

Figure 3. THE ESS HIGH LEVEL ARCHITECTURE



The inputs to the ESS consist of a high level textual scenario script, the International Scenario Script produced by Dassault Electronique from an earlier Siemens AG scenario, the ORBAT, a modified version of the NATO Generic Enemy Forces Catalogue, the ADatP-3 message format database, the Sensor Collection plan and the digital map and route database for the scenario area.

The key outputs of the ESS are the SGT database and the Message File. The SGT database provides at minute intervals the position, identity, subordination, structure, activity, combat effectiveness, boundaries and radio and radar emissions of each independent unit, which are typically at either platoon, company or battalion level.

The message file consists of a total of 1053 ENEMY SITREP, RECCEXREP and METREP ADatP-3 messages containing 2028 reports about Orange units, activities and occurrences over the 24 hour scenario duration. The average report rate is over 80 per hour rising to a peak of over 150 per hour.

The subsidiary ESS deliverables are the Collection Plan, the digital route database and the scenario Trafficability Overlay. The latter defines the scenario terrain Go, Slow Go and No Go areas and potential Avenues of Approach.

2.2 The ESS Components

Unlike the remainder of the DFD, the ESS uses the DERA Carapace Human Computer Interface (HCI). Carapace (Ref.7) was specifically designed and developed for intelligence applications and its use in the ESS has proved to be very cost effective. The main Carapace components are the Graphics Information Viewer (GIV), the Scenario Editor (SE), the Message Editor (ME) and the Route Planner (RP). The Forces Editor (FE) is a PC based ORBAT editor. The GIV, SE, RP and FE were produced for DERA by Hi-Q Systems Ltd. The ME was the Canadian contribution to the ESS and was produced for the Defence Research Establishment Valcartier (DREV) by Le Groupe CGI.

The core component of the ESS is the GIV which provides central control of the required area and time frame and enables the SE, ME and RP client applications to have their data and information displayed either separately or simultaneously upon a map background. Each client has the ability to selectively filter its output, making it possible, for instance, to view ELINT messages produced by the ME superimposed upon the SGT ELINT emitters called up by the SE.

2.3 ESS Scenario Generation

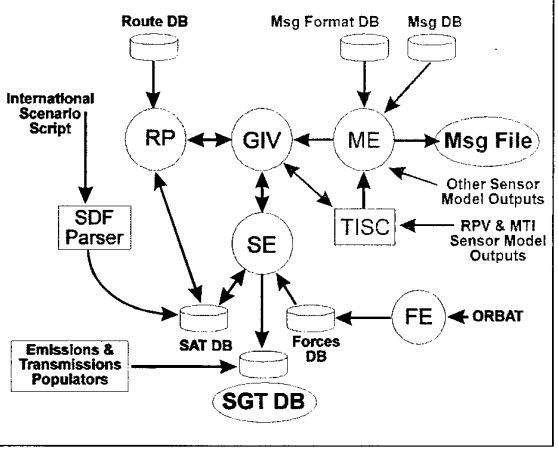
Figure 4 shows a high level diagram of the ESS, the shaded components are those required to enter and decompose the International Scenario Script and populate the Simulated Ground Truth database.

The scenario is entered, decomposed and the SGT created in five stages. Firstly, the locations and activities of each independent unit are manually transcribed into the Scenario Description text File (SDF). A unit's location is updated at each significant way point on its route, typically at 30-40 minute intervals. The SDF is then validated by the parser and entered into the Scenario Activity Tables of the Scenario Activity database.

Secondly, the scenario ORBAT is created by using the Forces Editor (FE) to edit the NATO GEFC to describe the composition and subordination of the scenario Orange forces.

The Orange ORBAT thus created is then entered into the Forces Tables of the Scenario Activity database.

Figure 4. ESS SCENARIO GENERATION



Thirdly, the Route Planner is used to define each unit's route between its way points; a route may be either along the road network or cross country or a mixture of the two.

Fourthly, the SE uses this data to decompose each unit down to its constituent vehicles in order to determine unit length, it then moves each unit from its assembly area onto its route in the correct sequence and updates the unit's position every minute as it moves between its way points. The SE warns the operator of conflicts such as the head of one unit overlapping the tail of another and provides an editing facility to enable the operator to resolve these and other conflicts.

Finally, when all conflicts have been resolved, the SGT database is populated with each unit's identity, subordination, location and activity and updated every minute. The SE also provides facilities to view the SGT against the map background and filter the data by area, time period, unit type and unit size. In addition the SE allows the unit history to be displayed as a track for any selected time period.

2.4 ESS Sensor Models and their Deployment

The ESS simulates six different types of sensor, a stand off, airborne Moving Target Indicator (MTI) radar, a ground based ELINT Electronic Support Measures (ESM) system, a COMINT ESM system, a number of Remotely Piloted Vehicles (RPV) with Thermal Imaging systems, a deployment of Forward Observers (FO) and an Air Reconnaissance system.

The models are deployed to the agreed collection plan which ensures that each unit and significant event is detected by at least one sensor system. This ensures that the DFS can be evaluated not only with message file that reports all units and activities but also with depleted message files in which messages or sensor missions have been removed.

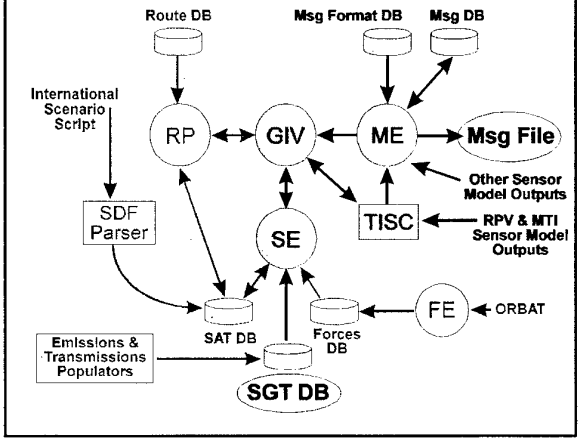
Each sensor model is a distinct Unix process that reads the SGT database and simulates the detection of scenario events according to its configured capabilities and detection and recognition performance. If necessary the models will decompose the scenario activity within the vicinity of their detection footprints into a second-by-second sequence in order to produce the required outputs.

The ELINT, COMINT, Forward Observer and Air Reconnaissance models directly generate their output in the form of ENEMY SITREP ADatP-3 messages whilst the radar and the RPV models produce binary output files which are converted into ADatP-3 messages by an operator using the Target Information Source Controller (TISC), the GIV and the ME.

2.5 ESS Message Generation

The ESS components used for message generation are shown shaded in Figure 5. The Message Format database contains the RECCEXREP format used for the Air Reconnaissance messages, the ENEMY SITREP format used for the RADAR, RPV, ELINT, COMINT and FO messages and the METREP format used for hand crafted weather messages.

Figure 5. ESS MESSAGE GENERATION



The function of the ESS message processing capability is firstly, to enable military analysts to view and replay the MTI RADAR and RPV data on the GIV via the TISC. And to create information reports, in the ADatP-3 ENEMY SITREP format, by selectively transferring data from the GIV to the ME and then editing it to add their classification of the detected targets and activities.

Secondly, to accept and validate the ADatP-3 messages created automatically by the other sensor models and add them to the Message database. Thirdly, to enable analysts to view the SGT and create specialised reports such as Intelligence Summaries from Corps and flanking divisions by transferring data to the ME and editing it. Fourthly, to enable hand crafted messages, such as the METREP weather reports to be validated and inserted into the Message database.

The messages from each sensor, including each FO, are stored in separate clusters in the Message database so that different Message Files can be created which include either all of the sensor missions or selected sub-sets. A typical FO message in the ENEMY SITREP format reporting the detection of two BTR armoured personnel carries (APC) heading north east at 19.40 Zulu on the 4th of July is shown below.

MSGID / ENEMY SITREP / FO / 65//
RPTTYPE / ENSIT //
EFDT / 041940Z / JUL //
EGROUP / U0067 /ORC //
LOCATION/REAL/AREA /-/-/-/ POINT / 32UPC8044178887//
SOURCE / C3 / OBSN //
TIME / AT / 041908ZJUL //
ACTIVITY / REAL / MOVING //
DIRCTN / NE //
EMAT / REAL / APC / BTR/ 2 //.

In addition to the message creation role the ME also provides the facility to selectively filter messages by source, area and time period and replay and display the individual reports in each message graphically upon the map background. As these reports can be superimposed upon the SGT this facility provides not only a valuable means of validating the Message File but also a very useful Intelligence tool in its own right.

2.6 ESS Message File Validation

In October 1996, at The Netherlands School of Military Intelligence, Ede, the DFD Advisory Group held a Manual Fusion Exercise (MFE) to validate the Prototype 1 ESS Message File. The exercise consisted of three All Source Cells, one manned by French Canadian and French military analysts, the second by Danish and UK military analysts and the third by Dutch and German military analysts.

All three cells were fed the ESS messages in the AdatP-3 formats via a printer in real time. The exercise was split into three eight hour sessions spread over three days, each cell was required to give the Exercise Control team a commander's briefing at the end of each period.

All three cells were successful in determining the number of enemy battalions facing them but were, in some cases, reluctant to believe their conclusions and consequently, failed to assess accurately the level of enemy forces. This was not due to a lack of information in the Message File but rather due to the analysts developing a range of hypotheses, sometimes based on little or no concrete evidence.

The Advisory Group concluded that the exercise successfully validated the Message File and that only some minor changes and additions were required. The players also agreed that the exercise provided them with a credible scenario.

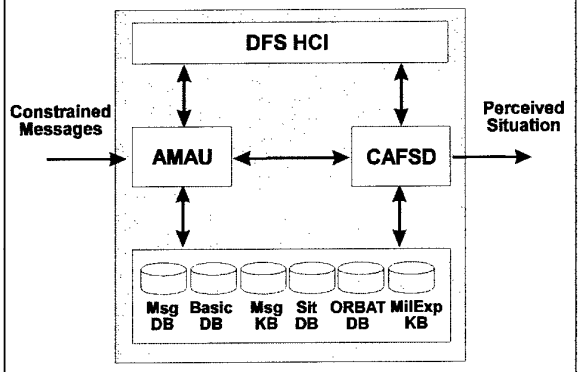
The Prototype 2 (P2) Message File has now been completed, delivered and validated by the UK Advisory Group representative. This is the same as the P1 version but with the addition of COMINT and RECCEXREP messages and a limited number of hand crafted INTSUMS from the adjoining divisions.

3 THE DATA FUSION SYSTEM

3.1 The DFS High Level Architecture

The DFS high level architecture is shown in Figure 6, the main components are the Asynchronous Message Assembly Unit (AMAU) produced by Le Groupe CGI, Quebec. The Correlation, Aggregation, Fusion and Situation Determination (CAFSD) system, the DFS Human Computer Interface (HCI) and the various data and knowledge bases. The DFS was or is being jointly produced by TNO and BSO in ORIGIN, in The Netherlands; Thomson CSF in France; Siemens AG, Dornier and Berner and Mattner Systemtechnik in Germany; DREV and Le Groupe CGI in Canada and Sirti, Italtel and Contraves in Italy. For historical reasons the designers of the CAFSD have used slightly different terminology to that defined in paragraph 1.3, here correlation refers to both the correlation and association processes and fusion to the algorithms used to combine the information about two or more associated entities. The purpose of the AMAU is to receive, process, validate and decompose the constrained messages into their individual reports and store the results in the Message Database.

Figure 6. THE DFS HIGH LEVEL ARCHITECTURE



The purpose of the CAFSD is to create the Perceived Situation, it does this by converting each individual report into a series of Situation Elements (SE) which it stores in the Situation Database. It then classifies, correlates (associates) and aggregates the SEs using a suite of classification, association and aggregation algorithms and stores and ranks all the resultant hypotheses, which represent all the likely consistent situations, in the Situation Database.

3.2 The AMAU

The AMAU receives the constrained messages from the TMS, it parses each message, rejects messages which are too stale or refer to activities outside the area of intelligence interest and diverts messages with errors to the military analyst for correction. It then decomposes each validated message into its individual reports in the DFS internal format and, in conjunction with the Level of Support (LOS) component of the CAFSD, assigns a credibility rating to each report. This credibility rating is based upon the existing number of Situation Elements which are related to the new report. The AMAU then assigns a global reliability rating to the whole message based on the type of source and its reliability history.

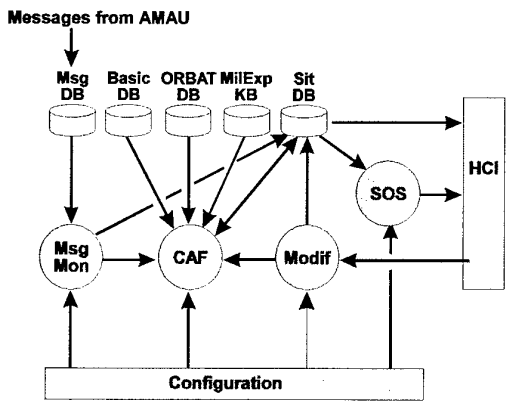
The individual reports are then given a priority rating based on whether or not they contain information which meets the commander's Priority Intelligence Requirements (PIR). Finally each processed report together with its assigned number, priority and reliability and credibility ratings is stored in the Message Database and the CAFSD is informed of its arrival.

3.3 The CAFSD

The main components of the CAFSD, apart from the LOS, are shown in Figure 7. When informed of the arrival of a new

message and its reports, the Message Monster (Msg Mon) takes each report in the message, converts it into an Observed Situation Element (SE), stores it in the Situation Database and informs the Correlation Association Fusion (CAF) system of its arrival.

Figure 7. THE DFS CAFSD



The CAF then retrieves the new Observed SE and integrates it into the perceived situation. Firstly, it uses the Basic and ORBAT data bases and the Military Expert Knowledge base to classify the observed SE and create a new Basic SE which is used in all further processing. The classification algorithm uses, for instance, the information about the number and type of vehicles detected together with the collateral information in the data and knowledge bases to derive the unit type, specialization and command level. If there is insufficient information to do this it creates an unidentified Basic SE. These processes are shown in Figure 8.

Secondly, it carries out a pre-association filtering to select only those SEs which might refer to the same unit on the battlefield. This is done by taking into account the age of the information, (stale information is not taken into account), unit identification information, unit classification information and the distance between the SEs together with their detection times. At the moment the location and time checks do not take account of the route network. By these processes it determines whether or not the new SE can be correlated (associated) with any of the existing SEs and produces a Derived SE for each potential correlation and computes the strength of the correlation links. The Derived SE data is produced by the fusion module which combines the information held by each of the correlated SEs.

Aggregation up to regiment level is performed by an aggregation algorithm which uses disposition templates stored in the Military Expert Knowledge base to find the best fit between the Hypotheses and a template. The introduction of a new SEs triggers the aggregation process. After aggregation the resulting Aggregation Hypotheses provide feedback to the correlation process so that other observations of the higher level unit can be identified.

The CAFSD Modification (MODIF) module allows the analyst to modify the decisions of the fusion system in the light of his expert knowledge and experience by creating, modifying or deleting SEs. These changes are passed to the CAF to enable it to take these changes into account.

The CAFSD HCI and the Situation Object Selector (SOS) enable the analyst to query the Perceived Situation database, which contains all the hypotheses, and display and filter the most likely Consistent Perceived Situation either for the current or any chosen historical time period. A Consistent Perceived Situation is one which incorporates all the SEs but only those hypotheses which have the highest probabilities and are consistent with each other. Once a current Consistent Situation has been selected the SOS maintains the display by sending updates as and when they occur. It is also possible to select and display lower probability Consistent Situations.

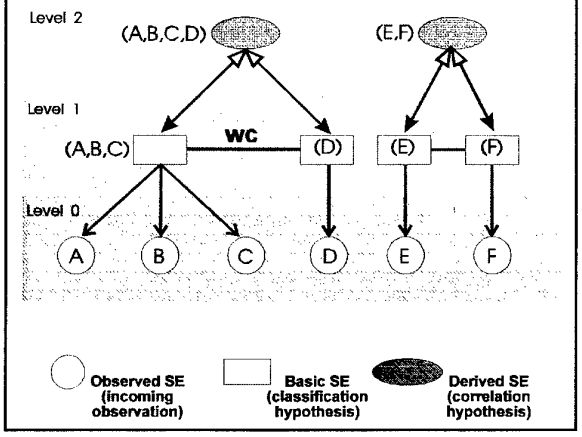
The Configuration module which will provide fine control of the CAFSD has not yet been integrated.

3.4 The Correlation Algorithm

The correlation algorithm developed by TNO and ORIGIN uses the Like Situation Element First principle together with Pairwise Correlation and is based on the work of Miles (Ref.8). In this approach SEs from similar sources are correlated first as their data sets are similar. Thereafter the new SE is correlated in turn with each of the existing filtered SEs and the strength of the correlation links determined.

Level 0 in Figure 8 shows the Observed SEs, Level 1 the Basic SEs and Level 2 the Derived SEs. In this example the initial correlation process has shown that the A, B and C Observed SEs refer to the same entity and consequently, only one Basic SE has been produced, whereas Basic SEs have produced for the D, E and F Observed SEs.

Figure 8. THE ASSOCIATION PROCESS



The correlation process determined a weak correlation link between (A,B,C) and (D) and produced and maintains two hypotheses (A,B,C,D) and (A,B,C) and (D). There is a strong correlation link between (E) and (F) and again two hypotheses are maintained, (E,F) and (E) and (F).

3.5 The Aggregation Algorithm

The aggregation engine developed by TNO and ORIGIN uses a library of disposition templates in the Military Expert Knowledge base to match the observed unit locations and movements with known types of disposition such as an

unopposed advance. The creation of a new hypothesis in the Situation database triggers the aggregation process. The templates are positioned to the latest timed event and all the possible matches are examined and the results stored in the Situation database. The selection of the candidate templates is based on the new unit hypothesis, type and command level and activity. The aggregation hypothesis strength is based upon the number of template slots filled and the hypothesis strengths of the units and their slots.

4 THE DFS PROTOTYPE 1 PERFORMANCE

The Prototype 1 DFD is currently being integrated and tested prior to being exercised in an Automated Fusion Exercise in October 1997. Consequently, the system has not yet been tested with the full ISS and a number of stability and integration problems are currently being resolved. Initial tests using the start up situation, a series of ADatP-3 reports derived from a simulated satellite mission over the Orange's forces initial assembly, have shown that the system can create and classify SEs at about 50 times the speed of an experienced analyst with a success rate of about 98%.

Initial tests of the correlation algorithms, using the last 4 hours of the scenario, showed a capability of handling 12 times the peak ESS message rate, again with a very high success rate.
More recent tests using the P2 ISS have shown that the DFS can classify and correlate SEs for the first 12 hours of the scenario in much faster than real time. However, some additional work is required to ensure that the system maintains this performance for the full 24 hours of the scenario.

Assessment of the aggregation algorithms by TNO and ORIGIN, using a Netherlands national scenario, have shown an excellent performance in aggregating up to regiment level. The aggregation algorithms have now been integrated in the DFS and initial tests with the International Scenario Script are providing very promising results.

5 THE FUTURE PROGRAMME

5.1 The Automated Fusion Exercise

After the success of the Advisory Group's Manual Fusion Exercise in October 1996 to validate the ESS Message File, it was decided to hold an Automated Fusion Exercise (AFE) in October 1997 to test and validate the complete DFD with military analysts.

The current plan is to have a number of Divisional All Source Intelligence Cells operating in parallel, each being fed with the ESS messages in real time. There will be four or five cells, each with two military analysts and two or more support scientists. The first cell will be the reference cell and be manual only, the next cell have the ESS graphical Message Viewer only and the third cell will have the DFS Prototype 1.21 with the classification, correlation and aggregation algorithms. The fourth cell will be similar to the third but with the addition of MODIF if the latter can be completed in time. In addition there will be fifth cell to demonstrate Prototype 1.5 which incorporates the new fusion algorithms currently being developed by the Italians.

Having a number of cells operating in parallel will not only provide the means to assess the effectiveness of the different levels of automation but also give a degree of redundancy.

This is intended to be a research exercise and its purpose is two fold, firstly, to provide a fitting end to the major part of the project and secondly, to provide the military and technical assessments necessary to enable the Italian consortium to complete an effective Prototype 2 in mid 1998.

6 CONCLUSIONS

The Prototype 2 Event Simulation System has been completed on schedule and has provided an effective means of producing and decomposing scenarios, deploying simulated sensors and generating realistic ADatP-3 message streams. In addition the key components of the ESS are currently being re-created on a PC NT platform in order to provide low cost training and operational aids.

The ESS Prototype 2 key deliverables, the Message File, the Scenario Ground Truth database, the Route database, the Avenues of Approach and the Collection Plan have also been completed, validated and delivered.

The Data Fusion System, the heart of the Data Fusion Demonstrator, is unfortunately behind schedule due primarily to contractual and political constraints. Nevertheless the completion and integration of the Prototype 1 DFS is proceeding apace and it is anticipated that there will be a stable and effective version available for the Automated Fusion Exercise in October 1997. The evidence todate is that within the next few months the DFD will provide the participating nations with not only a very valuable research tool but also with a number of modules with the potential for operational use.

7 ACKNOWLEDGEMENTS

The NATO DFD is a large, complex and ambitious project and represents a considerable amount of work by a large number of scientists and engineers in all the participating nations and their contractors. The success that we are now beginning to achieve is due to the very willing and effective co-operation of all the members of the DFD community and I would like to acknowledge the very friendly and effective co-operation that DERA and Hi-Q Systems Ltd. have received from all our DFD colleagues.

This paper is only an incomplete overview of the DFD Project and necessarily does not do full justice to many important areas of work and I would like to apologise to those friends and colleagues whose contributions have not been described as accurately and in as much detail as they justly deserve.

Finally, I would like to thank Ralph Wols and Marco Woestenburg of TNO, The Netherlands, for their invaluable contribution to the description of the DFS.

8 REFERENCES

1 "Programme Management Plan" (PMP) for the NATO AC243 (Panel 3) Data Fusion Demonstrator (DFD) Project. Appendix C Glossary of Terms. Version 4.0, January 1996.

2 "Software Requirements Specification for the ESS Prototype 2 Message File, (WP2501/SRS). Version 3.2 February 1997. Hi-Q Systems Ltd.

3 "The International Scenario Script Design Document (WP2010). V3.0." Dassault Electronique, March 1996.

4 "Memorandum of Understanding" (MOU) for the NATO AC243 (Panel 3) Data Fusion Demonstrator (DFD) Project. Less annexes A and B. Version 9.0, October 1991.

5 "Segment Design Document for the Event Simulation System (WP2000/SSDD). Version 3.2 February 1997. Hi-Q Systems Ltd.

6 "Correlation, Aggregation, Fusion and Situation Determination (WP1200 SRS/SDD) Version 3.0. April 1996. TNO-FEL, BSO in ORIGIN, Signaal and Berner and Mattner Systemtechnik. Unpublished.

7 "The Carapace User Guide Document" HCI/UG/004 Issue 2, 1993. Crown Copyright.

8 "MFE Evaluation Report, C34-EP-61, Version 1.0. February 1997. Hi-Q Systems Ltd.

9 Miles, J A H. "Artificial Intelligence Applied to Data Fusion and Situation Assessment for Command and Control." University of Southampton, 1988.

PAPER No: 37

DISCUSSOR'S NAME: A. Hume

COMMENT/QUESTION:

Can the system be used for fusion at a lower level for brigade or battle group?

AUTHOR/PRESENTER'S REPLY:

Yes, provided the ORBAT contains the necessary information on the structure of the lower level units. The aggregation algorithms have been designed for use at battleground level.

PIXELLESS INFRARED IMAGING USING QUANTUM WELLS

Gail J. Brown and Frank Szmulowicz*
Wright Laboratory, Materials Directorate
WL/MLPO, 3005 P St. Ste 6
Wright Patterson AFB, OH 45433-7707

1. SUMMARY

A new concept for designing infrared imaging systems is under development. In this concept, absorbed long wavelength infrared light is up-converted to near infrared emission through the vertical integration of a p-type multiple quantum well infrared photodetector (QWIP) heterostructure and a light emitting diode (LED). The infrared (IR) radiation is absorbed in the QWIP part of the heterostructure. The optically excited charge carriers from the QWIP are used to stimulate near IR emission from the LED layers. This emitted radiation can then be imaged using a commercial, high resolution silicon CCD camera. Because of the extremely short path that the photoexcited charge carrier traverses between the QWIP and LED layers, the emitted photon originates at the same spatial position as the absorbed photon, such that there is a one-to-one correspondence between the long wavelength infrared image and the near infrared image. With this device, the need for fabricating thousands of tiny pixels on the infrared array to generate a high resolution image is eliminated. With the elimination of pixels comes the elimination of the need for thousands of contacts to apply a bias voltage to each individual pixel. Only two wires are required to bias this QWIP and LED heterostructure. For a two color infrared system, this concept would require only three wires for the entire large area imaging array. Our preliminarary results for a vertically integrated p-type GaAs/AlGaAs QWIP and LED heterostructure will be discussed.

2. INTRODUCTION

For high resolution infrared imaging, focal plane arrays (FPA) composed of more than 256x256 individual detector elements (or pixels) are desired. For the mid- to long wavelength infrared region (3 to 12 μm), the current infrared FPA technology is based on HgCdTe or InSb detector arrays indium bump bonded to a silicon multiplexer for signal readout. This hybridization scheme has several drawbacks.

Large focal plane arrays require high uniformity of response over a large area. For InSb and HgCdTe arrays this large area material uniformity is difficult to obtain and drives up the cost of the imaging array. The recently developed quantum well infrared photodetector arrays, grown by molecular beam epitaxy , have the potential for lower materials cost and high uniformity across a three inch wafer.[1] However, even with well controlled, low defect density epitaxial materials, the fabrication of the thousands of pixels in the arrays can result in processing induced non-uniformities. Another drawback of hybridized arrays is the mismatch between the thermal coefficient of expansion between the silicon multiplexer and the IR detector array. This thermal mismatch can lead to debonding at the indium bumps after repeated thermal cycling.

To overcome these disadvantages of standard FPAs, Liu et al.[2] proposed a simple and innovative concept in which a quantum well infrared photodetector is in series with a light emitting diode to upconvert the long wavelength infrared photons absorbed by the QWIP into near IR photons emitted by the LED. The near IR emission serves as an optical readout of the longer wavelength IR image. The near IR optical output can then be captured with a standard silicon CCD camera to produce the desired visible image.

3. UPCONVERSION PROCESS

The fundamental processes of the vertically integrated QWIP and LED, mid to near infrared wavelength up-conversion device, are shown in Fig. 1. There are several different epitaxial layers in this device. These thin layers (40Å to 500Å) are grown by molecular beam epitaxy on a GaAs substrate. The GaAs and InGaAs layers of the LED are grown directly on top of the multiple GaAs and AlGaAs layers of the QWIP during the same growth run. Figure 1 shows the conduction and valence band edge profile of these layers using a p-type QWIP. The p-type QWIP is designed to absorb mid-infrared light which photo-excites a charge carrier (hole) from the confined ground state of the GaAs valence band quantum well to a state above the potential energy barrier of the wider bandgap AlGaAs layers.

Under an applied forward bias, the photo-excited hole moves to the LED layers and is collected in the bottom of the InGaAs valence band quantum well. Electrons from the top n^+ contact fall into the InGaAs conduction band quantum well and recombine with the confined holes, thereby emitting a near infrared photon. The intensity of the near IR emission is directly related to the intensity of infrared radiation absorbed. These devices have demonstrated 100% internal conversion efficiency of the injected hole to near infrared photon.[2]

* Also at the University of Dayton Research Institute, 300 College Park, Dayton , OH 45469-0178, USA

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

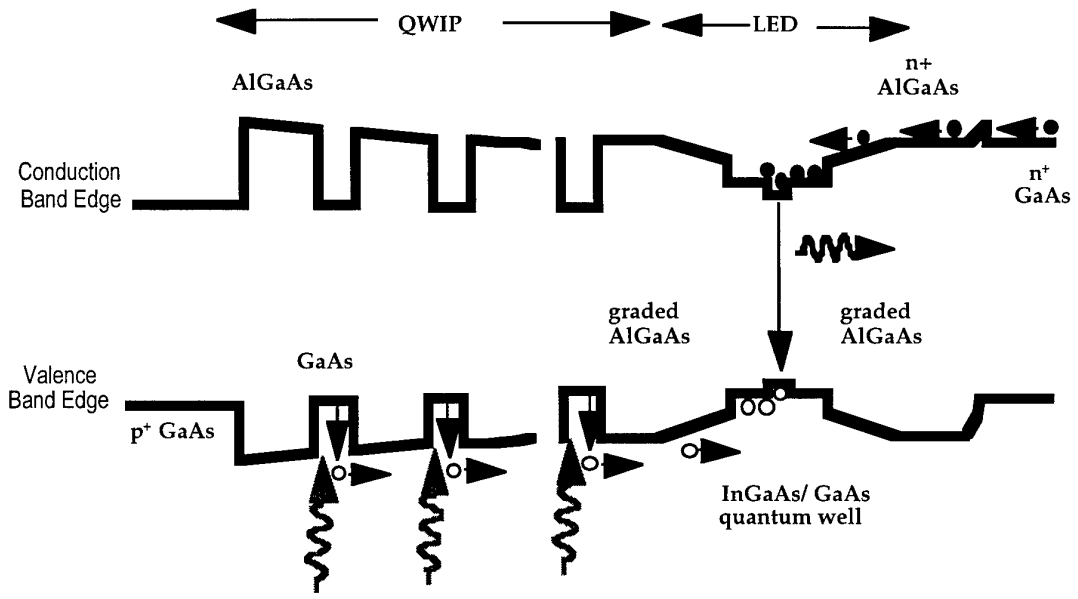


Fig.1 Energy band edge profile of a p-type QWIP-LED heterostructure for optical upconversion.

Only two contacts are needed for this device: one to the heavily n-doped LED contact layer and one to the heavily p-doped QWIP emitter. This is a remarkable change from the number of electrical contacts required on a hybridized imaging array.

4. NORMAL INCIDENCE ABSORPTION

The quantum well in the photodetector heterostructure could be doped with either donor atoms (n-type) or acceptors (p-type) and still be successfully integrated to the LED. In fact, the initial single element QWIP-LED device reported by Liu et al.[1] was based on a n-type QWIP. The advantage of a p-type GaAs/AlGaAs QWIP is that infrared radiation perpendicular to the epitaxial layers can be absorbed in the quantum wells formed in the GaAs valence band. This is not the case for n-type quantum wells in the GaAs conduction band. Because of the isotropic and parabolic nature of the GaAs conduction band, the confined electrons in the quantum wells can only couple to radiation with a polarization component in the confinement direction, i.e. normal incidence absorption is forbidden by the dipole selection rules.[3] Therefore, n-type QWIPs require diffraction gratings to be fabricated on the photodetector elements to bend the light so that it passes at an angle through the quantum well layers. Szmulowicz and Brown [4] have shown that the complex nature of the GaAs valence band allows a p-type quantum well to efficiently absorption of radiation perpendicular to the plane of the well.

5. P-TYPE QWIP DESIGN

For QWIPs, the preferred heterostructure design utilizes a bound-to-continuum optical transition. In a bound-to-continuum QWIP, the confined charge carrier is photo-excited from the quantum well ground state to an extended state above the potential energy barrier, i.e. in the continuum. In this case, the photo-excited charge carrier is free to move through the AlGaAs layers under an applied voltage. For a p-type GaAs/AlGaAs QWIP, there are several different final states that could be used. The calculated optical oscillator strengths for various possible final hole states are shown in figure 2. The optical transition with the largest oscillator strength for normal incidence absorption is the transition from the heavy-hole ground state (HH1) to the second light-hole (LH2) state. The HH1 to LH2 transition provides the largest optical absorption at the center of the Brillouin zone, where the majority of the confined holes reside.

The infrared absorption band of a QWIP is determined by the energy separation between the initial and final states of the quantum well. The two parameters which control this energy separation are the quantum well width, and depth (or $Al_xGa_{1-x}As$ composition). For a bound-to-continuum QWIP, care should be taken to design the final extended state just above the top of the potential energy barrier. The closer the extended state is to the top of the quantum well, the stronger the optical absorption will be. Figure 3 demonstrates the effect of changes in quantum well width on the calculated absorption spectrum intensity and cut-off wavelength.

With our advanced theoretical model, the p-type QWIP heterostructure can thus be designed to give maximum photoresponse in a specific wavelength band of interest. Utilizing the QWIP designs of Fig. 3, a series of p-type QWIPs were grown by molecular beam epitaxy for our studies.[5] The measured photoresponse spectrum, at normal incidence, for one of these samples is shown in figure 4. For a p-type QWIP with 30% aluminum in the barrier layers and a well width of 40Å, the peak photoresponse is at 8.5 μm and the cut-off wavelength is 9.5 μm. The measured photoresponse spectrum is in excellent agreement with the calculated absorption spectrum.

6. LED DESIGN

The active region of the light emitting diode, shown in Fig. 1, is very similar to that of a high performance quantum well laser. The advantage of using an InGaAs well in the active region is that the emitted photons will not be absorbed in the wider bandgap GaAs and AlGaAs layers. A low indium content $In_xGa_{1-x}As$ layer is used to avoid large strains caused by the lattice constant mismatch between GaAs and $In_xGa_{1-x}As$ with increasing indium. A typical LED active region would consist of a 90Å wide quantum well composed of $In_{0.1}Ga_{0.9}As$ and surrounded by GaAs. An LED of this design has a measured output wavelength of 875 nm.

7. PIXELLESS IMAGING ISSUES

In standard infrared imaging arrays, pixels are required to prevent optical and electrical cross-talk between adjacent areas on the array. However, the QWIP-LED device thickness is much smaller than typical alloyed infrared detector materials, i.e. <2 μm versus 6 μm. The combination of the thin device structure with the optical readout capability is what provides the opportunity for pixelless imaging with a QWIP-LED device.

There are mainly two key requirements for upconverting a long wavelength infrared image into a LED emission image with negligible distortion, smearing, and cross-talk. First, the thickness of the entire active region of the QWIP-LED structure must be less than the infrared wavelengths to be imaged. This requirement is based on the diffraction limit in optical elements.

Second, the charge carrier photo-excited in the detector layers must reach the LED active region with negligible motion perpendicular to the electric field of the applied bias. Again, considering the diffraction limit of the device, a perpendicular motion less than the infrared wavelength will have negligible effects. The photo-excited charge carrier from the QWIP only has to traverse a distance of about 0.07 μm before reaching the InGaAs well. The carrier diffusion lengths are very short in this device, on the order of 0.1 μm.

So, in principle the QWIP-LED architecture should lend itself to large area imaging without the need for the fabrication of individual detector elements. And without the need to form electrical interconnects to thousands of individual elements.

8. QWIP-LED DEVICE PERFORMANCE

To date, several tests of this pixelless imaging concept have been made. Allard et al.[6] performed these tests using a p-type QWIP-LED heterostructure grown by molecular beam epitaxy on a semi-insulating GaAs substrate. Details of the epitaxial layer thicknesses and compositions used in the device are given in Ref. 6. For testing the imaging characteristics, a large area device was fabricated using standard GaAs lithography and processing techniques. A 3mm by 4 mm mesa was etched to uncover a portion of the buried bottom contact. A metallized ring contact was deposited on the top of the mesa, as well as a metallized pad on the bottom contact layer. Therefore, only two wires were needed to operate this large area device. A schematic of the wavelength upconversion device geometry is shown in Fig. 5. [Please note that this drawing is not to scale. The device thickness (~2 μm) is actually much smaller than the device length (4 mm).]

Long wavelength QWIPs need to be cooled to cryogenic temperatures to operate efficiently. Therefore, the packaged QWIP-LED slab "array" was cooled to 80K using a liquid nitrogen optical cryostat. The reported applied bias used for testing was -3.0 volts between the top and bottom contacts.

Two imaging tests have been made with this pixelless QWIP-LED device in combination with a silicon CCD camera. The first test used a low intensity CO_2 laser beam focussed to one spot on the upconversion device. This test was to check for any optical cross-talk in the large area, pixelless device, i.e. will it really image without pixels?. A single bright spot, that was the size of the mid-infrared laser spot, was imaged off the rectangular surface of the QWIP-LED "array" by the Si CCD camera. No secondary emission of light from the areas surrounding the laser spot was observed. Liu et al [2] also showed that the LED output power increased as the input CO_2 laser power was increased.

The second test involved imaging the output of a 1000K blackbody-like source that had a 1 mm diameter aperature in front of it. Results from this test showed that the emitted image was a circular spot with about a 1mm diameter as expected. However, the spot imaged from the LED emission was ~150 μm smaller than the diameter of the infrared spot. This change in the emitted versus absorbed beam diameters is believed to be an effect of a reflection within the device's total thickness. For this test sample, the main contributor to the packaged device thickness was the 150 micrometers

of the thinned GaAs substrate. The recent theoretical work of Ryzhii et al.[7] found that the image transfer from the infrared generated current into the near infrared emission should be very close to one-to-one with no image smearing during the upconversion. Therefore, further thinning or removal of the GaAs substrate should result in improved image transfer from the mid-infrared scene to the Si CCD camera.

9. MULTI-SPECTRAL SENSING

The QWIP-LED concept can be applied to other quantum well material systems and other GaAs/AlGaAs quantum well designs, i.e. with different AlGaAs alloy compositions and well widths, in addition to the heterostructures discussed in this paper. This flexibility allows us to design QWIP and LED combinations for more than one spectral band. The tremendous advantage of needing only two contacts for the whole array, versus the thousands of electrical contacts on a hybridized array, becomes even more obvious for a multi-color sensor array. For a two-color QWIP-LED array, we need only design a second QWIP-LED combination that absorbs in a second infrared band and has a LED layer that emits a different near IR wavelength. Both sets of QWIP-LED layers would be sequentially grown during the same molecular beam epitaxy growth run. Three color silicon CCD cameras are already available, so the two separate near IR or visible emissions could be captured and fused into one visible image. Or, the two separate emissions could be used for other image processing needs. This double QWIP-LED structure would require at most three wires for the whole array.

10. CONCLUSIONS

A new concept for infrared imaging devices has been presented. By sequentially growing light emitting diode layers on top of quantum well infrared photodetector layers, a pixelless imaging device with an optical readout is created. Since the active region of the device is very thin, there is a direct mapping of the spatial and intensity information in the infrared scene into a near IR image. Without the need for fabricating and bump bonding tens of thousands of tiny pixels, a much simpler and lower cost infrared imaging array can be created.

These pixelless QWIP-LED imaging devices are only in the early stages of development. The full capabilities of these structures have not yet been explored. Further testing on improved p-type QWIP-LED structures is planned. The elegant simplicity of these integrated devices, especially for multi-spectral imaging, makes this a technology capable of creating a new generation of infrared imaging arrays.

10. ACKNOWLEDGEMENTS

We wish to acknowledge our collaboration with Dr. H. C. Liu, Institute for Microstructural Sciences, National Research Council of Canada, for research on developing p-type GaAs/AlGaAs quantum wells for QWIP-LED devices.

REFERENCES

1. B. F. Levine, "Quantum Well Infrared Detectors", J. Appl. Phys., **74**, R1 (1993).
2. H. C. Liu, J. Li, Z. R. Wasilewski and M. Buchanan, "Integrated Quantum Well Intersub-band Photodetector and Light Emitting Diode", Electron. Lett. **31**, 832 (1995).
3. L. C. West and S. J. Eglash, "First Observation of an Extremely Large-Dipole Infrared Transition within the Conduction Band of a GaAs Quantum Well" Appl. Phys. Lett., **46**, 1156 (1985).
4. F. Szmulowicz and G. J. Brown, "Infrared Absorption Calculation and Photoresponse Measurement of the Bound -to-Continuum Transitions in p-type GaAs/AlGaAs Quantum Wells", Phys. Rev. **B51**, 13203 (1995).
5. G. J. Brown, M. A. Capano, S. M. Hegde, K. Eyink and F. Szmulowicz, "Normal Incidence Photoresponse as a Function of Well Width in p-type GaAs/AlGaAs Multiquantum Wells", Proc. of MRS **450**, 231 (1997).
6. L. B. Allard, H. C. Liu, M. Buchanan and Z.R. Wasilewski, "Pixelless Infrared Imaging Utilizing a p-type Quantum Well Infrared Photodetector Integrated with a Light Emitting Diode", Appl. Phys. Lett. **70**, 2784 (1997).
7. Ryzhii, H. C. Liu, I. Khmyrova and M. Ryzhii, (in press), IEEE J. Quan. Elec. (1997).

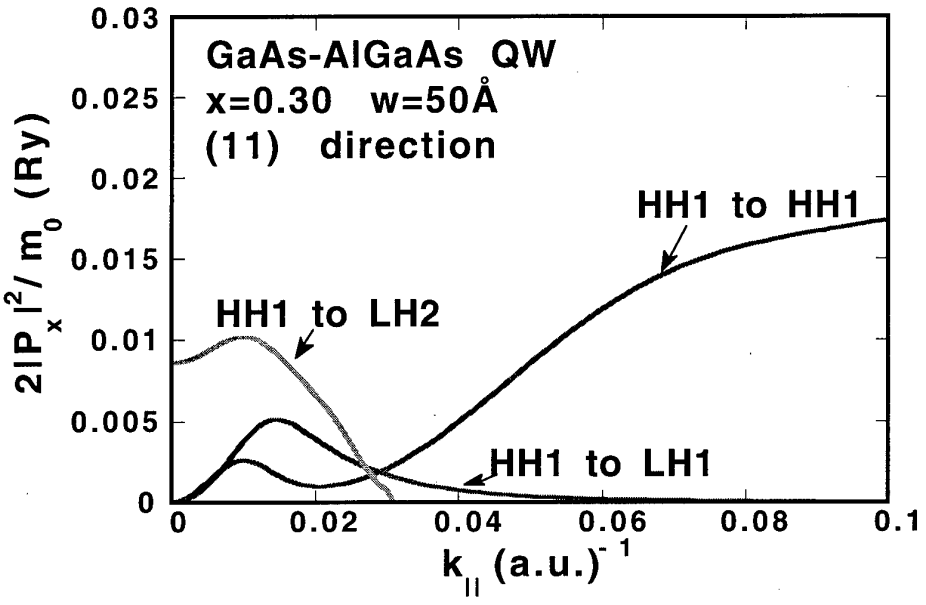


Fig. 2 Squared momentum matrix elements in the x-direction for intersubband transitions originating with the HH1 band as a function of $\mathbf{k}_{||}$ along the (11) direction for a 50Å $GaAs/Al_{0.30}Ga_{0.7}As$ quantum well (a.u. = 0.529177 Å).

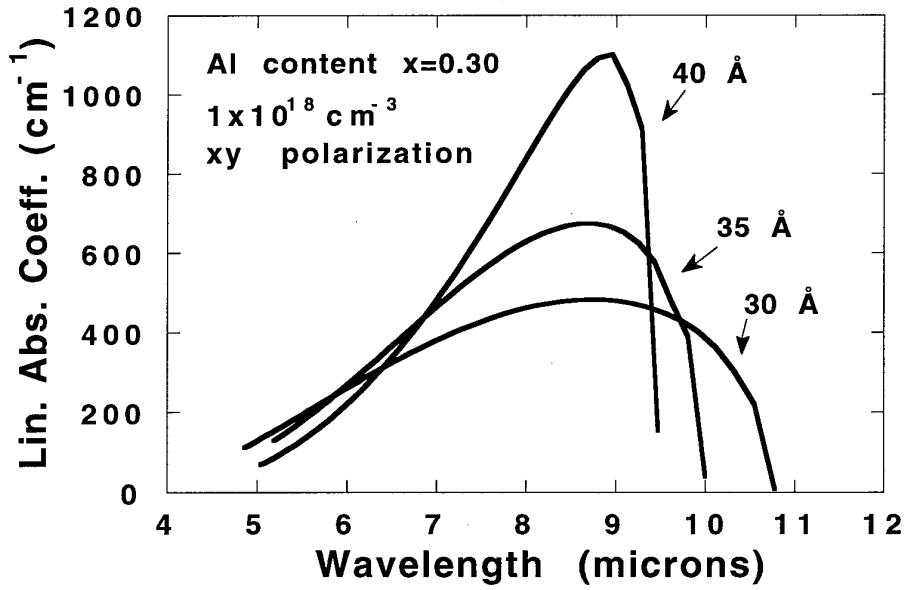


Fig. 3 Comparison of the calculated bound-to-continuum absorption spectra of three p-type quantum wells, for infrared radiation at normal incidence.

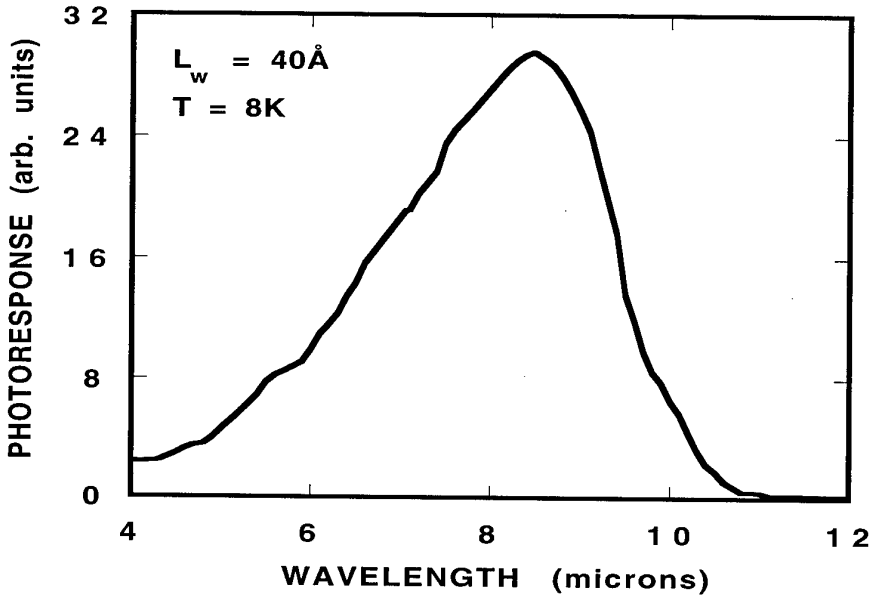


Fig. 4 The measured photoresponse spectrum at normal incidence of a p-type (beryllium doped) GaAs/$Al_{0.3}Ga_{0.7}As$ QWIP with a well width of 40Å.

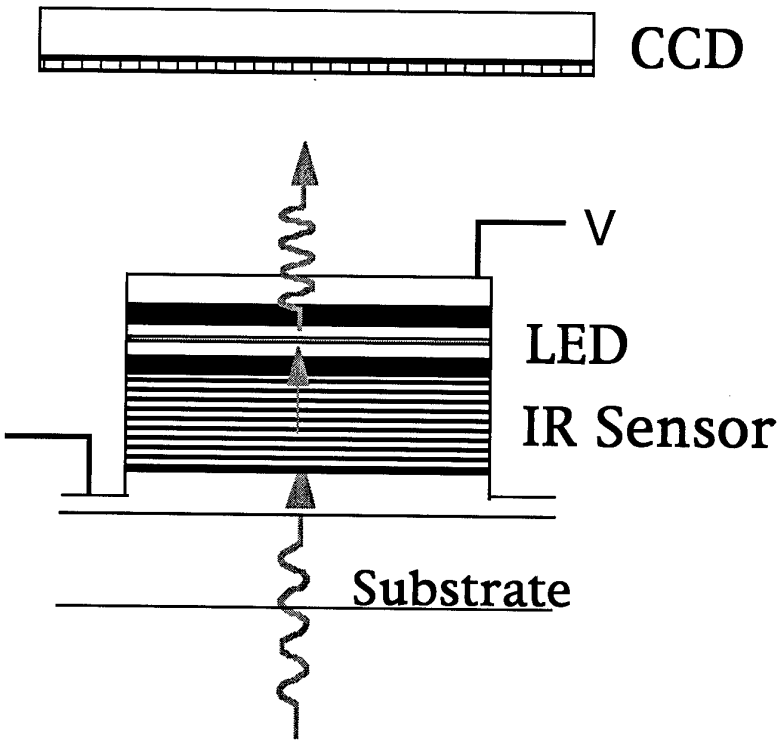


Fig. 5 Schematic of the wavelength up-conversion device geometry depicting the large area mesa of the QWIP-LED device in combination with a silicon CCD imager.

PAPER No: 38

DISCUSSOR'S NAME: P. Vangasse

COMMENT/QUESTION:

Is D* compromised by addition of LED interface and what D* performance is likely?

AUTHOR/PRESENTER'S REPLY:

LED interface doesn't compromise D*.
D* is about x2, x3, less (worse than) HgCd devices.

NONLINEAR OPTICAL FREQUENCY CONVERSION MATERIALS FOR MULTI-BAND REMOTE SENSING AND SURVEILLANCE

S. Caracci, N. Fernelius, M. Ohmer, & D. Zelmon
WL/MLPO Building 651
3005 P Street Suite 6
Wright-Patterson AFB, OH 45433-7707 USA

SUMMARY

There exists many nonlinear frequency conversion materials today which provide access to wavelength bands not previously possible by other means (solid state, gas, or semiconductor lasers). Of particular interest are optical sources for the 3-5 μm and the 8-12 μm wavelength bands. Operation in these optical transmission bands is very advantageous due to the large optical absorption of important chemical and biological materials and the low absorption of the atmosphere. A brief introduction of the uses of optical remote sensing and the unique properties of nonlinear materials for laser frequency conversion are described.

1 INTRODUCTION

Many future optical remote sensing and surveillance systems will benefit from the use of nonlinear frequency converters to provide high optical power at a diversity of wavelengths. These materials and laser sources are currently under development for applications such as pollution monitoring, urban and agricultural planning, weather forecasting, industrial process monitoring, biological pollution detection, wind shear detection, and chemical agent detection. Many current laser systems do not possess wavelength tunability and are often limited in their usefulness for many advanced remote sensing applications.

Accurate remote detection of biological and/or chemical materials requires the use of these frequency agile laser sources capable of high power optical output in multiple optical wavelength bands. This requirement stems from the diversity of absorption peaks in these materials and the existence of atmospheric absorption bands. Optical source requirements can be fulfilled with the use of nonlinear frequency conversion materials. Highly efficient frequency converters are now capable of reaching many of the important optical bands including ultraviolet, visible, near infrared, 3-5 μm, and 8-12 μm with sufficient output power.

2 OPTICAL REMOTE SENSING

Remote sensing has been a topic of interest for many years. Modern examples include the use of radars for image mapping from satellites and ground based systems.[1-3] These systems rely on active or passive detection of radiation from a target of interest and the interpretation of this detected signal. In the case of microwave remote sensing, information such as weather, foliage cover, and water temperature can be gathered.[2] For optical remote sensing (often called LIDAR for LIght Detection And Ranging) more detailed information of the molecular composition of a target can be determined from any one of many measurement techniques.

These techniques range from the use of light scattering from the medium which returns and is detected with no shift in wavelength (Rayleigh and Mie scattering) to techniques which rely on small wavelength shifts in the detected signal (Raman scattering). Other techniques are based on the optical absorption spectrum of a chemical or biological sample (Differential Absorption and Scattering (DAS), and DIfferential Absorption in LIDAR (DIAL)).[1]

The technique which has gained the most favor recently is differential optical absorption. DIAL makes use of two optical wavelengths. One of these wavelengths is centered at the chemical of interest's optical absorption peak. The second wavelength, the reference wavelength, is typical chosen to be off the peak in optical absorption and is used to help reduce noise and increase sensitivity. Due to the tunability needed to make such a technique possible, DIAL systems are ideal candidates for wavelength agile laser sources. Typically, a DIAL system consists of two major

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

subsystems; (1) the optical source system to generate a probe beam and, (2) the collection system (Figure 1).

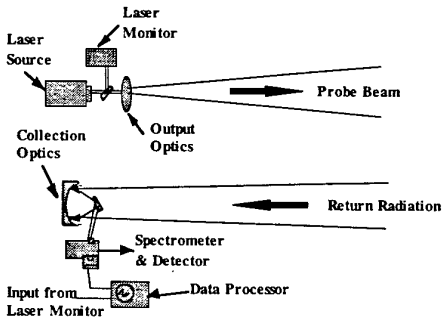


Figure 1. A typical LIDAR system consisting of a transmit and receive subsystems.

The probe beam is generated from a well controlled high power pulsed laser system. Diagnostic measurements are made of important pulse characteristics such as energy, timing and spectroscopic information. The optical beam is then sent out into the atmosphere though a lens system for beam shaping and collimation. When the reflected light returns from the target, a collection system consisting of lenses and mirrors, directs and focuses the optical radiation into a spectrometer system and detector. The detected signal is then sent to a signal processor for interpretation, data reduction, and readout. With this fairly simple setup, a mapping of the molecules between the LIDAR system and the target can be determined.

Methods such as DIAL are useful because they are capable of remotely determining properties of materials and aerosols. It can determine if a chemical is present above a certain concentration threshold or even determine its absolute concentration. Thermal structure of clouds or exhaust plumes can be monitored due to the temperature variation of absorption characteristics (both static and dynamic). Finally, it can determine a chemical species presence due to its unique spectral fingerprint.

High performance optical remote sensing systems also require high sensitivity and resolution, large standoff range, and wavelength tunability to look at many types of materials. These requirements are demanding on both the detection hardware and the laser probe source. For example, lasers are required to have large output powers to increase the range and sensitivity of the LIDAR system. They should also have a narrow optical bandwidth and short pulse duration for increased sensitivity and resolution. High pulse repetition rates are needed for noise reduction and averaging.

Such stringent requirements on the optical source also imposes severe requirements on the optical source materials. As next generation systems are designed and built, nonlinear materials are being increasingly sought to help provide such an ideal laser source. The advantages of designing a source based on nonlinear materials are; (1) solid state materials offer high rugginess, high efficiency and low part count, (2) a large operating wavelength spectrum covering visible, mid-infrared, and far-infrared spectral windows, (3) continous wavelength tunability over several micron, (4) high performance optical beam properties (high output power, large repetition rates, and short pulses), (5) decreased cost.

3 FREQUENCY CONVERSION MATERIALS

The use of nonlinear crystals for frequency conversion followed shortly after the discovery of the first laser. Whether the need is to gain access to wavelengths unobtainable from the current laser technology or to increase the ability of an optical system for frequency tunability, the need to shift optical power from one wavelength to another has always been important.

The earliest nonlinear materials were often inefficient, expensive, and riddled with defects. Over the last thirty years, many material improvements and newer materials with better properties have been realized. These include materials for frequency conversion in the ultraviolet, visible, and infrared. Because of the wide range of application and requirements, the material is usually optimized for a specific application and pump source.

Frequency conversion occurs in a dielectric medium when a high energy optical field is incident upon a material causing a nonlinear response of the material's refractive index. This change allows coupling of the electric field (pump beam) at one wavelength to electric fields at other wavelengths (signal and idler beams). This comes with the requirement that energy is conserved between the pump, signal and idler beams. Also, the process

will only occur with high efficiency if the relative phase of all the beams remains nearly zero. Therefore, two requirements must be satisfied, namely, conservation of energy, expressed as:

$$\omega_p = \omega_s + \omega_i \qquad (1)$$

where ω_p, ω_s, and ω_i are the wavelengths of the pump, signal and idler beams respectively, and conservation of phase (or momentum) which is expressed as:

$$k_p = k_s + k_i \qquad (2)$$

where k_p, k_s, and k_i are the wave vectors of the pump, signal, and idler beams respectively. In the case of nonlinear frequency conversion, the second condition (Equation 2) is the more restrictive. In order to have the phase of different wavelengths beams remain the same, the velocity of the waves must be identical. One method to accomplish this condition in a material is depicted in Figure 2. In this case, the material's natural birefringence or difference in refractive index for different crystalline directions is used to compensate for the different propagation speed of each wavelength. The directions are known as the ordinary and extraordinary direction. This method of optical phase compensation is known as birefringent phasematching. Unfortunately, this technique requires very strict control over the direction of propagation in the crystal, necessitating accurate polishing and cutting techniques.

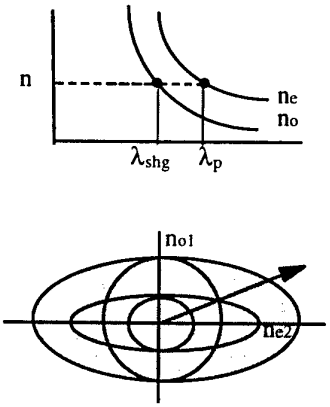


Figure 2. Birefringent phasematching requires a material to be anisotropic and limits the direction of propagation of electric fields for efficient frequency conversion.

This can force the use of a lower nonlinear coefficient since the largest nonlinear tensor does not always lie in the phase matching direction. Even with this restriction, many high power systems are based on these materials. Figure 3 shows a graph of many well developed nonlinear materials, for infrared applications and their transmission range and figure of merit. Two of the most important birefringent materials for this wavelength band, $ZnGeP_2$ and GaSe, are currently being developed for optical sources.

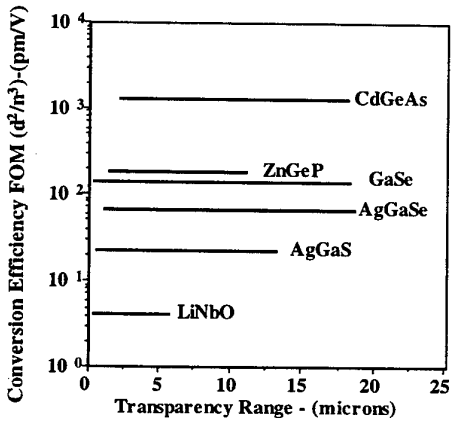


Figure 3. Figure of merit and transmission range for many common nonlinear frequency conversion materials.

$ZnGeP_2$ belongs to the chalcopyrite family of crystals. As it can be seen in figure 4, $ZnGeP_2$ has a very large optical gain making it very useful for high power optical sources. One problem with this material, however, is the observation of a crystal defect absorption band near 2 μm (see Figure 5).

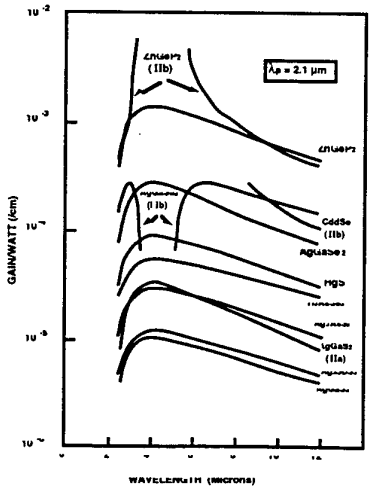


Figure 4. Optical Gain in $ZnGeP_2$ (From ref. 5)

This absorption band has recently been the topic of intense study. Reductions in the optical absorption through better crystal growth techniques and a better understanding of the defect structure have been achieved.

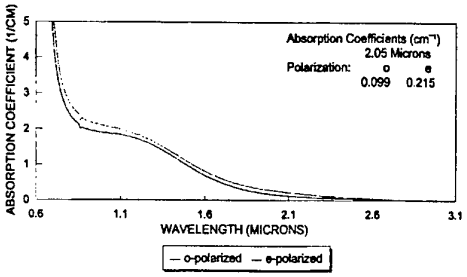


Figure 5. Optical absorption in $ZnGeP_2$ crystals grown by the horizontal Bridgeman method.

Gallium Selenide (GaSe) is another nonlinear material which has recently received attention. This material has a very large transparency range of approximately 1 μm to 18 μm (Figure 6), making it an extremely interesting material for applications in the mid- and far-infrared. [8] The measured nonlinear coefficient of ~55 pm/V for GaSe is adequate for most application. This material, because of its mica-like structure, is difficult to process. Cutting and polishing is of particular difficulty in this material because the crystal's tendency to cleave in thin sheets only along the [0001] crystalline direction. Recent attempts at doping and optimizing the growth parameters has reduced this problem. Even so, many optical systems have been designed using GaSe resulting in very functional sources with optical wavelength outputs over a 3 μm to 18 μm range.

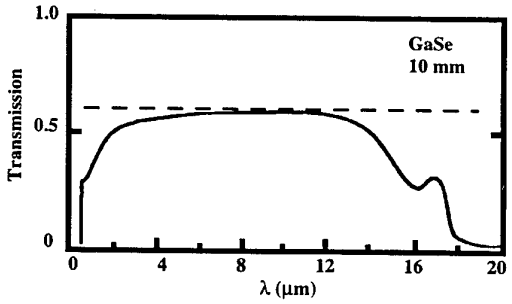


Figure 5. Optical transmission curve for GaSe.

Another methodology for achieving efficient frequency conversion is quasi-phase matching (QPM). Although this type of phase matching was first described in 1962, it wasn't until recently that materials and processes were developed to take full advantage of this technique. [6] In this method, a grating is used to compensate for the mismatch of phase in the propagating wavelengths instead of crystal birefringence (Figure 7). This is accomplished by periodically inverting the direction of the nonlinear tensor so that as the waves move out of phase the next periodic section brings them back into phase. This is equivalent to adding a grating wave vector of magnitude δK.

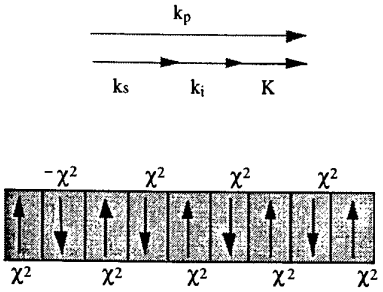


Figure 7. Quasi-phase matching uses a grating to conserve momentum during frequency conversion. The grating is formed by periodically inverting the nonlinear tensor.

Quasi-phase matching is a very powerful technique. The direction of the optical axis can now be in the direction of the highest nonlinearity . The crystal can also be customized for specific applications or wavelength conversion through external processing after crystal growth. This technique does however require a material whose nonlinear tensor can be altered in a periodic fashion. Materials like ferroelectrics have the property of being periodically poled under a large field. This field causes the ferroelectric domains to reverse and hence reverse the nonlinear tensor. For example, QPM has been successfully demonstrated in multi-grating periodically poled lithium niobate (PPLN). The periodicity of the ferroelectric domains can be easily controlled through photolithography to achieve any grating spacing to phase match the pump, signal and idler optical waves. Large tuning ranges (1.4 μm - 4.8 μm) and powers (>3 Watts) have already been demonstrated using this technique (Figure 8). [7]

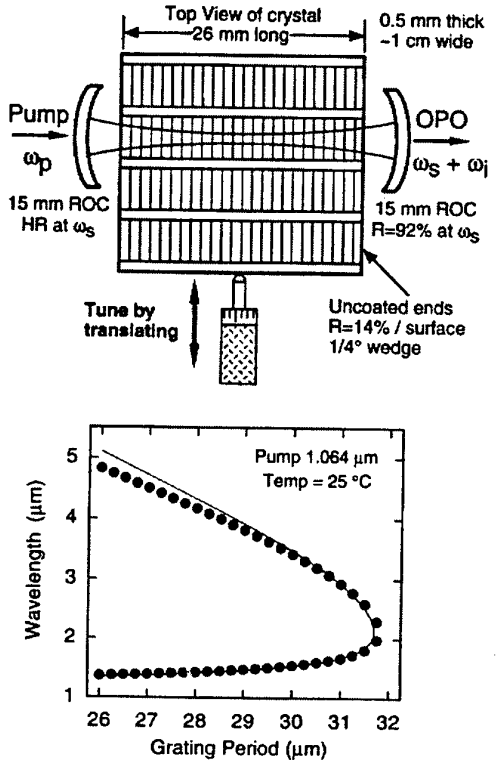


Figure 8. (a) Typical optical parametric oscillator configuration for a multi-grating Periodically Poled Lithium Niobate crystal and (b) corresponding tuning curve showing a large wavelength tuning range.

4 CONCLUSIONS

Nonlinear materials exist today, and others are being developed, capable of highly efficient laser frequency conversion. Optical sources based on these materials are ideal for insertion into remote sensing systems. By providing wavelength tunability over large optical spectral bands, improved sensing techniques are possible for systems with better sensitivity and operational performance. Materials like ZGP and GaSe have established themselves as high performance materials, what remains is the transition of this technology to new application like remote sensing. Other materials and technique, periodically poled lithium niobate for example, are still in their infancy but are rapidly making an impact on this technology.

5 REFERENCES

1. R.M. Measures, "Laser Remote Sensing: Fundamentals and Applications", 1st Ed., Krieger Publishing Co., Florida, 1984.

2. A. Schanpf, "Monitoring Earth's Ocean, Land, and Atmosphere from Space - Sensors, Systems, and Applications", American Institute of Aeronautics and Astronautics Inc., New York, 1985.

3. National Aeronautics and Space Administration, "Remote Measurement of Pollution", NASA Report #SP-285, National Aeronautics and Space Administration, Washington, D.C., 1971.

4. V.G. Dmitriev, G.G. Gurzadyan, D.N. Nikogosyan, "Handbook of Nonlinear Optical Crystals", 2nd Ed., Springer-Verlag, New York, 1997.

5. E.R. Nichols, J.C. Corbin, V.L.Donlan, "A Review of Parametric Oscillators and Mixers and an Evaluation of Materials for 2-6 μm Applications", Air Force Avionics Laboratory Technical Report #AFAL-TR-74-161, Air Force Systems Command, 1974.

6. J.A.Armstrong, N Bloembergen, J. Ducuing, P.S. Pershan, "Interaction Between Light Waves in a Nonlinear Dielectric", Phys. Rev., vol. 127, pp. 1918-1939, 1962.

7. L.E. Myers, R.C. Eckardt, M.M. Fejer, R.L. Byer, W.R. Bosenberg, "Multi-grating Quasi-phased-matched Optical Parametric Oscillators in Periodically Poled $LiNbO_3$", Optics Letters, vol. 21, pp. 591, 1996.

8. K.L. Vodopyanov, "Parametric Generation of Tunable Infrared Radiation in $ZnGeP_2$ and GaSe Pumped at 3μm", J. Opt Soc. Am. B, vol. 10, pp. 1723, 1993.

Nonlinear Optical Frequency Conversion Materials for Multi-Band Remote Sensing and Surveillance

S. Caracci
N. Fernelius
M. Ohmer
D. Zelmon

WL/MLPO Bldg. 651
3005 P Street Suite 6
Wright-Patterson AFB, OH 45433-7707 USA
(937) 255-4474 x3216

Outline

- Introduction to Remote Sensing
- Requirements for Optical Remote Sensing
- Laser / Optical Parametric Oscillator Fundamentals
- Nonlinear Crystals for Frequency Conversion
  - $ZnGeP_2$
  - Periodically Poled Lithium Niobate
  - GaSe
- Conclusions

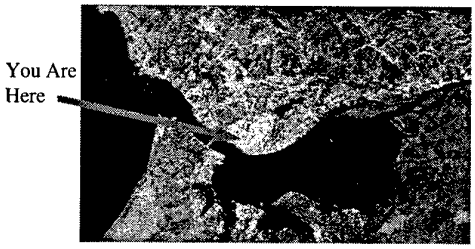
Why Do We Need Remote Sensing?

Environmental
- Urban Planning
- Weather Forecasting
- Agricultural Planning

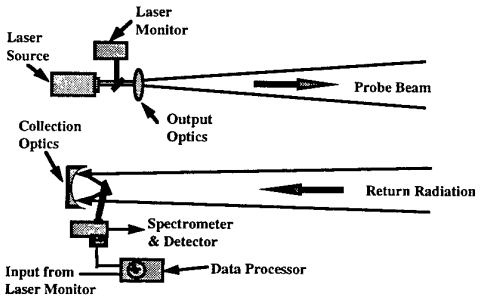
Chemical
- Pollution Monitoring
- Chemical Agent Detection
- Process Monitoring

Biological
- Biological Pollution (Ride Tide)
- Biological warfare agent detection

Example of RADAR Remote Sensing



Typical Optical Remote Sensing System (LIDAR)



Optical Remote Sensing Measurement Capabilities

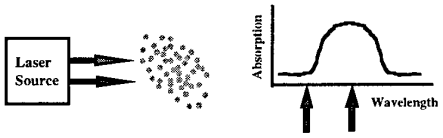
- Concentration measurements
- Thermal structural, and dynamic properties
- Threshold detection
- Plume dispersal
- Spectral fingerprinting

Laser Remote Sensing Techniques

- Rayliegh Scattering
- Mie Scattering
- Raman Scattering
- Resonance
- Flourescence
- Absorption
- Differential Absorption (DIAL)

DIAL for Remote Sensing of Chemical Species

- Very Sensitive detection of chemical species
- Good Spatial Resolution
- Requires two wavelength (On-peak, Off-Peak)



Laser Source Requirements for LIDAR

- High Peak Output Power (>1 MW Peak)
- Narrow Bandwidth (<2 nm)
- Frequency Agility (>50 nm)
- Output Power and Wavelength Stability
- Narrow Optical Pulses (<100 ns)
- High Repetition Rate (>1 kHz)

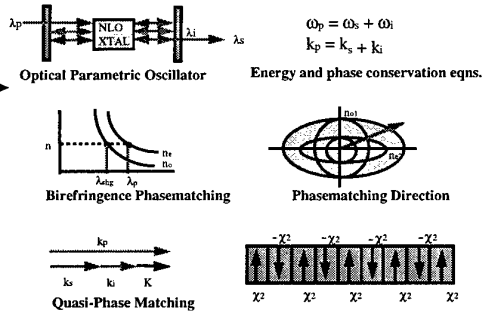
Current Laser Types for Remote Sensing

- Nitrogen Lasers (Raman Backscattering)
- Q-switched Ruby Laser
- Quadrupled Nd-YAG
- Carbon Dioxide (CO_2)
- Dye Lasers (Useful for DIAL Measurements)
- Diode Lasers (Short Distances)

NLO Crystals Advantages

- Solid State (Increased Reliability)
- Large Tuning Range (Many Microns)
- Availability of Many Wavelengths (Visible-IR)
- High Efficiency
- High Power, Repetition Rate, Narrow Pulses
- Cheap

Nonlinear Frequency Conversion



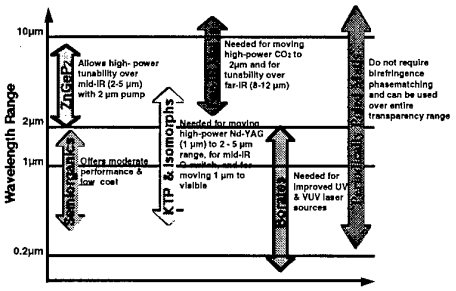
Optical Parametric Oscillators

- Similar to Laser
- Parametric Mechanism
- Satisfy Conservation of Energy and Momentum
- Angle or Temperature Tunable
- High Efficiency Possible

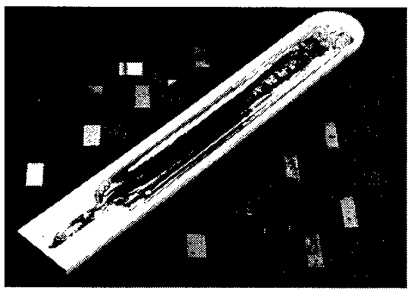


Optical Parametric Oscillator

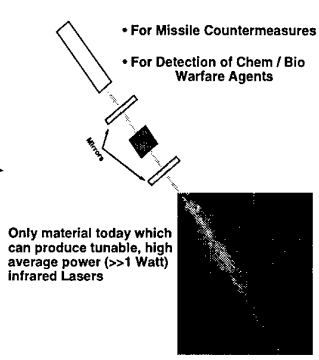
Broad Range of Bulk NLO Materials



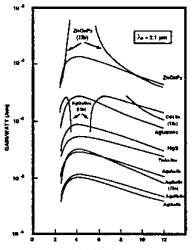
Zinc Germanium Phosphide ($ZnGeP_2$)



$ZnGeP_2$ - The Material for New Infrared Lasers



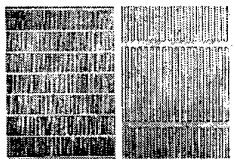
Only material today which can produce tunable, high average power (>>1 Watt) infrared Lasers

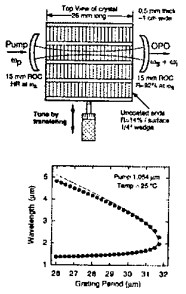


Dual Use:
- Environmental Monitoring
- Wind-Shear Detection
- Laser Radar
- Medical Lasers
- Materials Processing

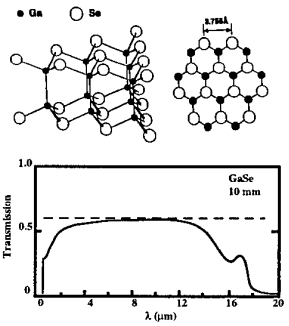
Mid IR Multigrating PPLN

Tuning from ~1.4 - 4.8 μm
Multi -Watt Output Power
Suffers from Photorefractive Effects
Low Cost Material
Uses 1.06 μm Laser Pump





Gallium Selenide (GaSe)



Other Applications for NLO Crystals

- Infrared Countermeasures
- Medical Lasers
- Optical Memories
- Guide Stars
- Spectroscopy

Conclusions

- Remote Sensing is on the Rise
- Current Laser Sources Not Optimal
- Nonlinear Bulk Crystals Can Add Increased Functionality to Remote Sensing Systems
- NLO Crystals on the Verge of Going Commercial

For further information

Dr. Stephen Caracci

WL/MLPO Bldg. 651
3005 P Street Suite 6
Wright-Patterson AFB, OH 45433-7707

(937) 255-4474 x3216
caraccsj@ml.wpafb.af.mil

Fusion d'informations et réseaux physiques pour les systèmes multi-senseurs sol-air

C. Nahum
Ingénieur de Recherche
O.N.E.R.A.
BP 72, 92332 Châtillon
France

1. SOMMAIRE

Les travaux entrepris pour le compte du Groupe de Coordination Défense Aérienne de la DRET, proposent une méthodologie d'aide pour la configuration, l'évaluation et l'amélioration des performances de systèmes multi-senseurs sol-air de type SACP. Ils ont été concrétisés par la réalisation du logiciel ROSACE (Recherche pour l'Optimisation de Systèmes Anti-aériens Conception Evaluation) décrit dans [5] et [10].
La première partie de ce document présente le logiciel ROSACE. La seconde partie est consacrée au problème de l'acheminement des informations au sein du système multi-senseurs. En particulier un algorithme est présenté, qui permet de gérer l'allocation des ressources et le transfert des informations pour que la logique algorithmique soit respectée.

2.INTRODUCTION

Les travaux entrepris pour le compte du Groupe de Coordination Défense Aérienne de la DRET proposent une méthodologie d'aide pour la configuration, l'évaluation et l'amélioration des performances de systèmes multi-senseurs sol-air de type SACP. Ils ont été concrétisés par la réalisation du logiciel ROSACE (Recherche pour l'Optimisation de Systèmes, Anti -aériens Conception Evaluation) décrit dans [5] et [10]. La première partie de ce document présente le logiciel ROSACE. La seconde partie est consacrée au problème de l'acheminement des informations au sein du système multi-senseurs. En particulier, les conditions de compatibilité entre logique algorithmique et réseau de communication physique, sont étudiées.

2.1 Pose du problème

Un système multi-senseurs sol-air est constitué d'un ensemble de capteurs, (radars, caméras optiques ou infrarouges) répartis sur une région dont la surveillance ou la défense doivent être assurées. Des moyens de communication (fibres optiques, radio, câbles coaxiaux) permettent l'acheminement des informations vers des centres de décision, l'exploitation de données d'origine différente par des processeurs de traitement. Un système de défense comporte en outre des moyens de défense (batteries de tir, missiles...).
Le problème de l'opérationnel militaire consiste dans un premier temps à sélectionner les capteurs qui répondent au mieux à un cahier des charges. Le coût, la fiabilité, les performances des composants sont alors étudiés en fonction des objectifs. Des outils d'aide au choix des composants ont été développés comme par exemple l'outil SADDAA (Voir [4]). Mais l'opérationnel est ensuite confronté au problème de la configuration du système sur le théâtre des opérations, c'est-à-dire au positionnement géographique des capteurs, des moyens de défense, à la définition du réseau de communication. En effet, son objectif est d'assurer une bonne visibilité compte tenue de l'environnement et en particulier des masquages par le terrain, une bonne détection, une localisation suffisamment précise des engins aériens. Tout ceci doit être pensé avant même que le système ne soit confronté à un scénario réel.

Le scientifique se pose quant à lui la question des traitements et de leur amélioration pour l'accomplissement des différentes tâches de détection, pistage, classification/dénombrement d'engins aériens. Notamment, l'apport, l'intérêt et le coût de la fusion d'informations doivent être étudiés. Mais qui dit fusion dit échange, transfert de données. Les algorithmes sont donc évidemment contraints par les possibilités de communication au sein du système.
L'objectif visé est l'élaboration d'un logiciel interactif, évolutif, et modulaire d'aide à la conception et à l'évaluation d'un système multi-senseurs pour la surveillance et la défense anti-aérienne. Il ne s'agit pas de mettre au point un simulateur qui reproduirait dans leurs moindres détails les fonctionnements des capteurs dans un environnement donné. Il s'agit plutôt de rechercher des critères qualifiants les performances opérationnelles globales d'un système puis d'améliorer ces performances par modification des paramètres du système ou de l'algorithmie.
Afin de démontrer la faisabilité d'un tel produit, certaines limitations sont imposées quant à la variété des capteurs, des menaces, des paramètres d'environnement. Certaines simplifications concernant les phénomènes physiques sont adoptées mais néanmoins, les situations envisagées conservent un caractère réaliste.
Les capteurs considérés sont des radars de veille, ou des caméras infrarouges. Dans un premier temps, il est proposé de synthétiser les informations pertinentes pour la spécification d'une situation et la simulation des missions d'engins aériens ennemis. Les menaces sont des hélicoptères, avions ou missiles. Ensuite, il s'agit de répertorier et d'évaluer les algorithmes de fusion qui accomplissent les tâches de détection, localisation, classification et pistage des engins aériens.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

2.2 L'outil ROSACE, outil méthodologique

L'outil ROSACE a été réalisé pour répondre à ces problèmes. C'est un logiciel interactif, ergonomique, portable sur station Unix supportant X-Windows et modulable. Ce n'est pas un logiciel écrit selon les normes industrielles mais une maquette. Dans son état actuel, il comporte quatre modules :

- Le module SITUATION offre la possibilité de définir et de configurer interactivement un système multi-senseurs sur une région d'intérêt. Une carte de la région peut être visualisée à l'écran. L'opérateur sélectionne les composants du système (capteurs, moyens de défense) dans une base de données préétablie. Ces composants sont représentés à l'écran par des icônes. Il les positionne et les oriente sur la carte. Il définit ensuite un réseau de communications (des moyens de communications type figurent dans la base de données).
- Le module ANALYSE propose une évaluation globale des performances du système (c'est-à-dire indépendamment des scénarios auxquels le système sera confronté). Ces critères comprennent le calcul de la visibilité et de l'intervisibilité des capteurs avec prise en compte du masquage par le relief du terrain, le taux de détection d'un aéronef supposé situé en un point de l'espace, ainsi que le taux de fausse alarme, la précision de la localisation, la létalité. Les résultats sont représentés graphiquement de manière synthétique. De plus un partitionnement de l'espace aérien surveillé est proposé. Ce partitionnement permet d'attribuer la responsabilité d'une région de l'espace à un groupe de capteurs.
Les régions insuffisamment surveillées ou protégées apparaissent alors et suggèrent à l'opérateur une modification de la situation donc un rebouclage sur le module SITUATION. Ce module est le coeur du logiciel et en fait son originalité. Il ne s'agit pas d'effectuer des traitements mais d'évaluer avant tout leurs performances.
- Le module SCÉNARIO permet de définir et de visualiser interactivement des raids (ensemble de trajectoires d'avions, d'hélicoptères ou de missiles) grâce à l'implémentation d'un éditeur de trajectoires adapté. En outre le scénario peut comporter brouilleurs, aérosols, et différentes conditions atmosphériques peuvent être précisées.
- Enfin le module SIMULATION prend en compte la problématique temporelle et donne une idée du comportement du système en état de fonctionnement. Les tâches de détection, pistage, classification sont donc simulées et accomplies par des algorithmes de fusion qui exploitent le réseau de communications pour améliorer le comportement du système.

ORGANISATION GÉNÉRALE DU LOGICIEL ROSACE

SITUATION

Théâtre
Capteurs
Postes de défense
Logique de fusion
Communications

SCÉNARIO

Edition de raids
Brouilleurs
Conditions météo
Aérosols

ANALYSE

Couverture par capteur
Détection/Fausses alarmes
Précision en localisation
Létalité

SIMULATION

Mesures
Détection
Pistage
Classification
Dénombrement

OPTIMISATION

Seuils capteurs
Position/orientation
Algorithmie

A titre d'illustration, la figure 1 propose un exemple de configuration d'un système multi-capteurs de veille. Le théâtre des opérations est représenté par une image SPOT.

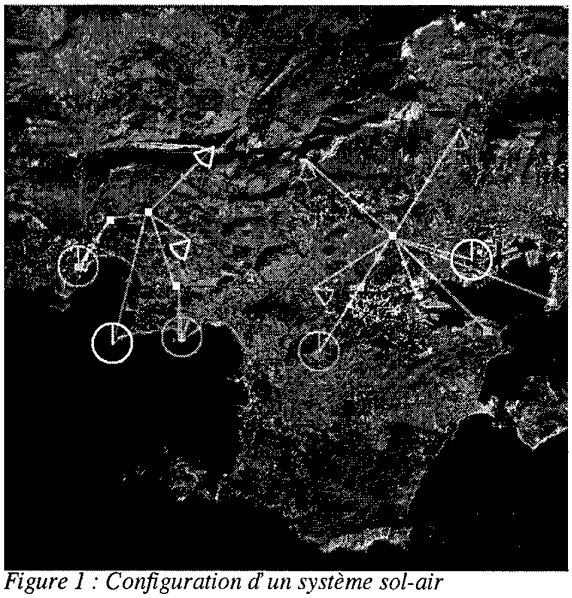


Figure 1 : Configuration d'un système sol-air

Chaque composant du système multi-senseurs est représenté à l'écran par une icône qui indique le type du composant, ainsi que son état d'activité. La symbologie adoptée est la suivante :

Radar : Le secteur angulaire observé en azimut est figuré en blanc.

Caméra : Le secteur angulaire observé en azimut est figuré en rouge.

Moyen de défense : Il est représenté par un carré rouge plein.

Position à protéger : Elle est représentée par un carré blanc en trait épais.

Logique de fusion : Symbolisé par des lignes bleues orientées par une flèche reliant les capteurs aux nœuds de fusion et aux moyens de défense.
Réseau physique : Matérialisé par des lignes vertes dont l'épaisseur caractérise le type de ligne.

Le réseau virtuel indique la logique d'association des informations. Les nœuds de fusion sont représentés par un carré bleu dans lequel le type de fusion est symbolisé par un signe en blanc. Les signes sont Σ pour un nœud de combinaison linéaire des observations, + pour un nœud à seuil haut (test ou) et Π pour un nœud à seuil bas (test et). Le choix de ces trois types de nœuds est justifié dans [5]. Il n'arrive qu'une seule ligne à un moyen de défense (une seule décision pour un point donné de l'espace : détection ou non détection d'une cible en ce point). Ceci permet de définir la notion de sous-système comme étant l'ensemble des capteurs qui contribuent à donner une décision à un même moyen de défense. Par contre d'un capteur peuvent partir plusieurs lignes vers des nœuds différents. Les nœuds indiquent le type de fusion logique des décisions (ou des observations) et symbolisent un terme ou un facteur dans une équation de fusion.
Le réseau virtuel décrit donc de façon graphique les opérations de fusion et de décision de réaction des défenses sans tenir compte explicitement du réseau de communications physique existant.

Dans [5], nous avons considéré trois groupes type de lignes de communication :
- Les lignes haut débit qui permettent la transmission de quelques séquences vidéo. On peut citer parmi ces lignes les fibres optiques, (30 Méga-octets par seconde) ou Ethernet (10 Méga-octets par seconde)
- Les lignes moyen débit, qui permettent la transmission d'une imagette au voisinage d'une menace potentielle ou d'une carte radar. Citons par exemple les liaisons RS232 dont le débit est de l'ordre de 50 Kilo-octets par seconde.
- Les lignes bas débit permettent de transmettre des valeurs numériques, position, variance, confiance de quelques plots, leur débit est de l'ordre de 500 octets par seconde.

En pratique, dans le module SITUATION du logiciel ROSACE, nous avons intégré la possibilité de tracer interactivement trois types de liaisons pour lesquelles on indique dans la base de données le débit et la portée.
- Type 1 : Câbles coaxiaux, fibres optiques.
Ces liaisons sont tracées au sol comme le chemin qu'elles empruntent. La longueur est limitée par la portée indiquée dans la base de données.
- Type 2 : Liens radio (par exemple le poste PR4G de Thomson, autorise des transmissions sécurisées à 4800 bits/s pour data, phonie etc... Les délais d'acquisition de cibles sont inférieurs à 1 seconde.
- Type 3 : Liens optiques pour lesquels on indique en plus la hauteur au sol des transducteurs. Si il y a un masquage entre les deux objets, alors l'opérateur est averti de l'impossibilité de la liaison optique.
Ces liaisons sont visualisées par des traits verts épais pour le type 1, des tirets pour le type 2 et des lignes en pointillés pour le type 3.
Notons que la désactivation d'une ligne du réseau physique peut se traduire par la suppression des éléments du réseau logique qui ne sont plus compatibles avec la topologie et les débits du réseau physique encore actif. D'autre part la cohérence entre les seuils de détection et donc les taux de fausse alarme des capteurs et le réseau physique devra être systématiquement vérifiée. Il faudra aussi respecter les capacités de transmission et de traitement en particulier dans le cas des attaques "saturantes".

Pour configurer un système de veille ou de défense, l'opérateur sélectionne donc un à un les composants, les positionne et les oriente à l'aide de la souris, sur l'écran représentant le théâtre. Ensuite, il spécifie les liens logiques entre les différents composants, et définit le réseau de communications.
Lorsque l'opérateur a configuré un système et que cette configuration lui paraît intéressante, il peut sauver la situation. Sinon, il peut la modifier par ajout d'un élément de la base de données, déplacement, suppression ou désactivation.

La visibilité des capteurs figure 2, la précision de localisation d'un hélicoptère figure 3, sa vulnérabilité figure 4 sont représentées avec une échelle de couleurs sur un plan d'altitude constante, ici 500m. Notons que par rapport à la situation de la figure 1, le radar nord-ouest a été déplacé.

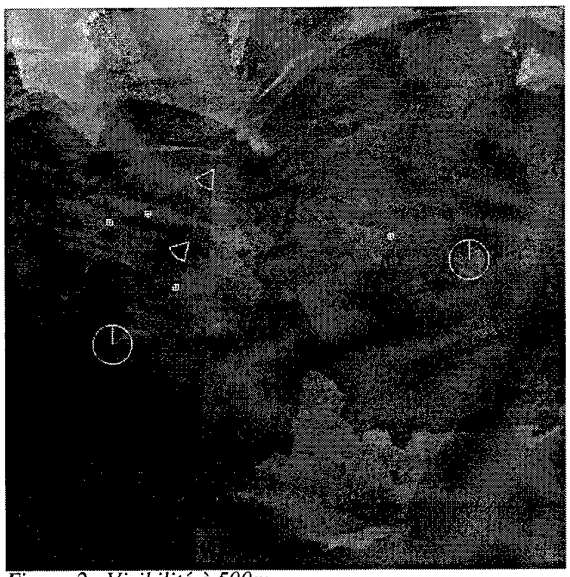


Figure 2 : Visibilité à 500m

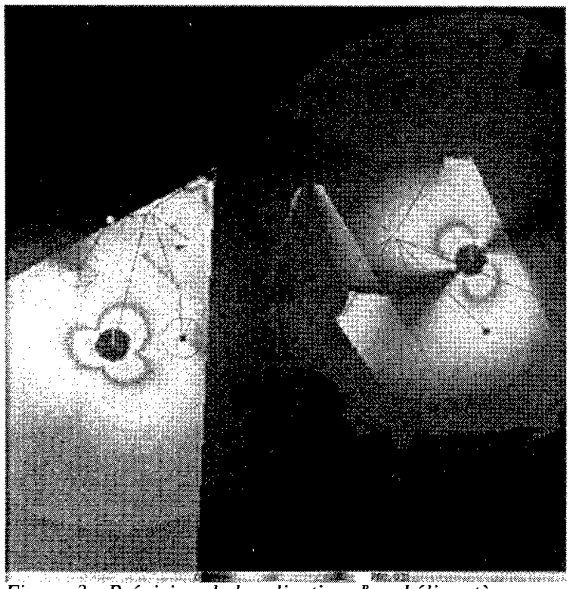


Figure 3 : Précision de localisation d'un hélicoptère



Figure 3 : Vulnérabilité d'un hélicoptère

2.3 Applications et perspectives

Initialement développé pour la conception de systèmes de surveillance sol–air, les applications du logiciel ROSACE peuvent trouver leurs sources dans d'autres contextes. Signalons que le CEM (Centre d'Essais de la Méditérannée) a montré un intérêt pour l'outil dans le cadre de la préparation d'essais. Une adaptation ergonomique du logiciel pour les besoins du CEM est en cours.
D'autre part, l'outil peut servir de plate–forme pour la mise au point et la comparaison d'algorithmes.

3. ACHEMINEMENT DES INFORMATIONS

Dans ce chapitre nous décrivons les moyens de communication et d'échange d'informations entre les différents composants du système multi–senseurs. Nous avons dans [10] distingué deux niveaux de communication :
- le réseau physique est constitué de lignes physiques avec leur débit défini comme le nombre de messages transmissibles par seconde ainsi que le délai de transmission que l'on supposera nul dans ce qui suit.
- le réseau virtuel décrit l'équation logique entre les composants pour la fusion des informations.

3.1 Le réseau physique

Rappelons que les composants d'un système multi–senseurs sont essentiellement de trois types : les capteurs, les moyens de défense et les processeurs de traitement. Les capteurs effectuent une observation du milieu et transmettent une information cartes d'énergie en distance–azimut par exemple pour les radars 2D, plots, etc... vers un processeur de traitement, via une ligne physique du réseau. Les processeurs de traitement effectuent une fusion des informations afin d'aboutir à une décision qu'ils communiquent aux moyens de défense. On supposera dans toute la suite que les capacités de traitement ne sont pas limitées. Autrement dit la fusion n'est contrainte que par le débit des lignes de transmission.
Les informations pourront être acheminées via des multiplexeurs/démultiplexeurs qui ne réduisent en aucun cas le débit nécessaire. La figure 5 montre un exemple de réseau physique pour un système comprenant huit capteurs et deux processeurs. Les flêches indiquent la direction de circulation des informations, les ronds verts sont des multiplexeurs/ démultiplexeurs. Les carrés verts représentent les processeurs de traitement. Nous n'avons pas représenté de moyen de défense.

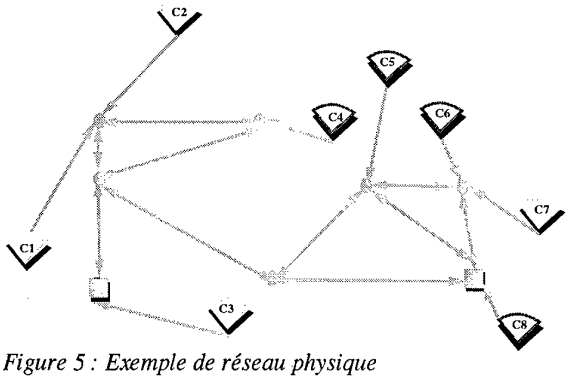


Figure 5 : Exemple de réseau physique

3.2 La logique de fusion

Rappelons que dans [5], [8] et [9] nous avons étudié l'apport de la fusion d'informations pour la tâche première du système multi–senseurs c'est–à–dire la détection. L'apport a été quantifié par l'amélioration de la probabilité de détection P_d fonction du rapport signal à bruit pour une probabilité de fausse alarme P_{fa} fixée. Par conséquent, dans [10], nous avons considéré qu'il pouvait y avoir en matière de détection, trois opérations de fusion d'un intérêt notable. La fusion optimale (celle qui fournit les meilleures performances ou encore celle qui donne une probabilité de détection P_d maximale pour une valeur de P_{fa} fixée consiste à calculer un rapport de vraisemblances multi–dimensionnelles et à le comparer à un seuil. Dans le cas où les capteurs peuvent être délocalisés, nous avons montré que ce type de fusion n'est pas praticable. D'autre part, même dans le cas où les capteurs sont colocalisés, la fusion linéaire optimale fournit quasiment les mêmes résultats.

La fusion linéaire consiste à effectuer une combinaison linéaire des observations et à comparer cette valeur avec un seuil. Ce type de fusion nécessite donc la transmission des observations (on ne considèrera que les énergies en pratique) et donc des lignes à haut débit. La fusion linéaire optimale entre deux capteurs sépare le plan des observations en deux régions Z_0 et Z_1 par une droite qui fournit parmi toutes les droites possibles la meilleure valeur de P_d, P_{fa} étant fixée. La conjonction pour laquelle un "et" logique est effectué entre les décisions émises par deux capteurs, et la disjonction pour laquelle un "ou" logique est alors effectué, sont deux opérations de fusion qui ne nécessitent pas un gros débit. Ces deux tests peuvent être optimisés. Nous avons montré dans [5] qu'il existait un choix de seuils optimal, ce choix fournit la valeur maximale de P_d, P_{fa} étant fixée.
Dans le cas de deux capteurs, la figure 6 montre les régions de décisions de ces différents tests.

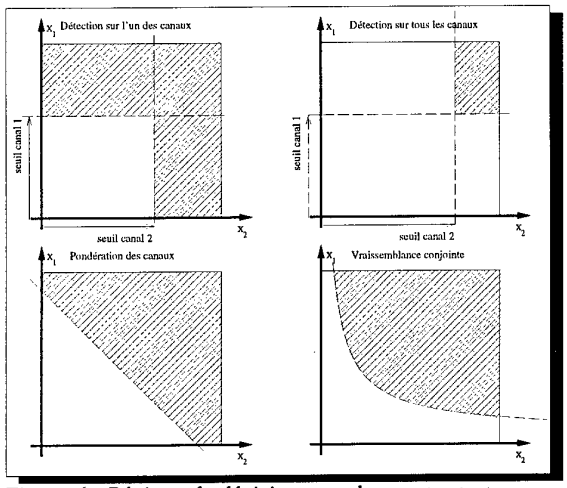


Figure 6 : Régions de décision pour les quatre tests

La notion d'optimalité des différents tests montre que dans le module OPTIMISATION du logiciel ROSACE, il est tout à fait envisageable d'optimiser les seuils de détection pour améliorer les performances du système multi-senseurs. Pour un sous-système du système multi-senseurs, la façon dont les observations ou les décisions doivent être fusionnées est donnée par une équation que nous appellerons "équation logique de fusion". Cette équation est représentée pour l'opérateur du logiciel ROSACE par des lignes (comme nous l'avons décrit dans [10]) et en machine par un arbre dont la figure 7 illustre un exemple.

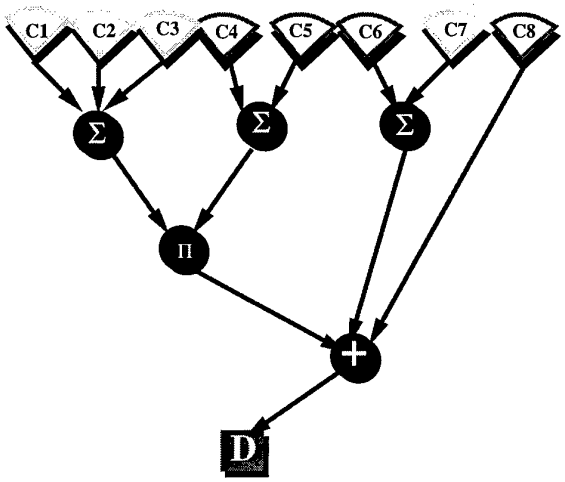


Figure 7 : Exemple de logique de fusion pour la détection

L'équation logique de fusion peut s'écrire de manière condensée

$$\left(\left(C_1 \Sigma C_2 \Sigma C_3\right) \Pi \left(C_4 \Sigma C_5\right)\right) + \left(C_6 \Sigma C_7\right) + C_8$$

L'opération Σ symbolise le test linéaire, Π la conjonction, + la disjonction. Il faut noter que le test linéaire est le plus riche en informations et s'effectue toujours en premier si il y a lieu.
Si u_1, $u_2 \ldots u_8$ désignent les observations délivrées par les capteurs alors on calcule les combinaisons linéaires L_1, L_2, L_3 telles que

$$\begin{cases} L_1 = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 \\ L_2 = \lambda_4 u_4 + \lambda_5 u_5 \\ L_3 = \lambda_6 u_6 + \lambda_7 u_7 \end{cases}$$

On les compare aux seuils γ_1, γ_2, γ_3 et u_8 est comparé à γ_4. Enfin, on effectue les opérations logiques

$$\left((L_1 \overset{?}{>} \gamma_1) \text{ et } (L_2 \overset{?}{>} \gamma_2)\right) \text{ou } (L_3 \overset{?}{>} \gamma_3) \text{ ou } (u_8 \overset{?}{>} \gamma_4)$$

On prend une décision de présence de cible si l'un des trois termes vaut 1, donc soit si u_8 dépasse le seuil, soit si L_3 dépasse le seuil ou encore si à la fois L_1 et L_2 dépassent leur seuil.

4. COMPATIBILITÉ DU RÉSEAU PHYSIQUE ET DE LA LOGIQUE DE FUSION

La compatibilité entre l'équation logique de fusion et le réseau physique doit être examinée. La question est de savoir si l'opération de fusion est réalisable.
Dans l'affirmative on peut se demander comment, à quel prix et quel processeur aura le résultat de l'opération. Ces problèmes s'apparentent aux problèmes de transbordements dans un graphe et pour cela nous reprenons des éléments de la théorie des graphes dont les détails figurent dans [2].

4.1 Éléments de la théorie des graphes

Un graphe G est déterminé par la donnée d'un ensemble X de sommets ou nœuds, et d'un ensemble U de couples ordonnés de sommets, appelés des arcs. On notera alors

G=[X,U]. On dira que j est un successeur de i si il existe un arc de la forme u=(i,j). Γ(i) désigne l'ensemble des successeurs du sommet i.
Étant donné un sous-ensemble de sommets S, on définit $\omega^+(S)$ ensemble des arcs ayant leur extrémité initiale dans S et leur extrémité finale dans X–S, et $\omega^-(S)$ ensemble des arcs ayant leur extrémité terminale dans S et leur extrémité initiale dans X–S.
La matrice d'incidence sommets–arcs associée à un graphe G=[X,U] est une matrice $A=(a_{iu})$ avec i=1...N, u=1...M à coefficients entiers 0, +1, −1 telle que chaque colonne correspond à un arc de G et chaque ligne à un sommet de G avec
Si u=(i,j) alors $a_{iu}=+1$ et $a_{ju}=-1$
Les autres termes de la colonne u étant nuls.
De manière équivalente, si on considère une ligne i quelconque (correspondant au sommet i) alors

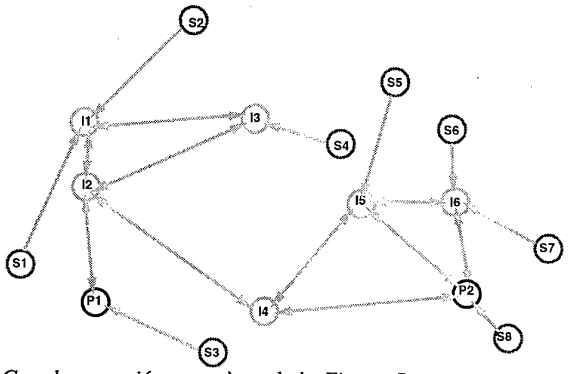
$$\begin{cases} \omega^+(i) = \{u\ ;\ a_{iu}= +1\} \\ \omega^-(i) = \{u\ ;\ a_{iu}= -1\} \end{cases}$$

4.2 Chemin dans un graphe

Un chemin de longueur q est une suite de q arcs $u_1,u_2,...u_q$ avec

$$\begin{cases} u_1=(i_0,i_1) \\ u_2=(i_1,i_2) \\ \dots \\ u_q=(i_{q-1},i_q) \end{cases}$$

Un graphe est dit connexe si pour tout couple de sommets i et j, il existe un chemin de i à j. Dans le cas d'un graphe non connexe, on associe à chaque sommet i sa composante connexe c'est-à-dire l'ensemble des sommets j tels qu'il existe une chaîne de i à j.
La vérification de la connexité ou la recherche des composantes connexes d'un graphe est l'un des premiers problèmes de la théorie des graphes. Il a une importance vitale pour l'élaboration des réseaux électriques, téléphoniques.
Dans le contexte des systèmes multi-senseurs, nous considèrerons les sommets particuliers que sont les processeurs et nous chercherons à connaître l'ensemble des capteurs susceptibles de transférer leurs informations vers un même processeur. Autrement dit, nous chercherons la composante connexe d'un poste de traitement. Notons toutefois que plusieurs postes de traitement peuvent appartenir à une même composante connexe. Dans ce cas il faudra déterminer celui qui effectuera le traitement (problème de l'allocation des ressources).
La figure 8 représente le graphe associé au système multi-senseurs de la figure 5. Les processeurs sont P1 et P2, les capteurs S1, S2,...S8 et enfin les concentrateurs I1,...I6. Ces notations seront explicitées ultérieurement.



Graphe associé au système de la Figure 5

Nous ne présentons pas l'algorithme classique qui permet de trouver la composante connexe associée à un sommet. Cet algorithme est dû à Trémaux 1882 et Tarjan 1972. Il explore les sommets en profondeur d'abord le plus loin possible dans le graphe sans former de cycle, puis il remonte jusqu'à la dernière bifurcation laissée de côté et ainsi de suite jusqu'au retour au sommet de départ. L'ensemble des sommets ainsi rencontrés forme la composante connexe. Si tous les sommets ont été rencontrés, le graphe est connexe. La complexité de l'algorithme est en O(M).
Il existe d'autre part des algorithmes qui permettent de trouver le chemin le plus court entre deux sommets. Dans le cas simple où les longueurs des arcs l(u) valent toutes 1, il existe un algorithme en O(M).

4.3 Flots sur un graphe

La notion de flot sur un graphe fournit un modèle général applicable à une grande variété de problèmes : circulations du courant électrique, réseau de transport, structuration des réseaux de communications etc...
Soit G= [X,U] un graphe connexe dont les arcs sont numérotés u = 1,2,...M. Un flot dans G est un vecteur à M composantes
$\phi = (\phi_1, \phi_2, \dots \phi_M)^T$
de $I\ R^M$ tel que en tout sommet $i \in X$ de G, la première loi de Kirchhoff soit vérifiée

$$\sum_{u \in \omega^+(i)} \phi_u = \sum_{u \in \omega^-(i)} \phi_u$$

La composante ϕ_u du vecteur ϕ est appelée flux sur l'arc u. La relation exprime la loi de conservation des flux aux nœuds (la somme des flux entrant en un sommet est égale à la somme des flux sortant).

On appelle réseau de transport un graphe G=[X,U] dans lequel chaque arc u de U est muni d'un nombre $c_u \geq 0$ appelé la capacité de l'arc. Lorsqu'on fait circuler un flot sur G, ce nombre indique la limite supérieure du flux admissible sur l'arc u. On supposera que les capacités sont des nombres entiers et que $c_u \geq 0$.
Étant donnés deux sommets particuliers de X dans le graphe G, s la source et p le puits, ($s \neq p$), on considère le graphe $G^0 = [X,U^0]$ déduit de G en ajoutant l'arc (p,s) de retour de flot. On convient de lui attribuer le numéro 0. Les arcs de G^0 sont donc numérotés 0,1,2,...M.

On dit que le vecteur $\phi=(\phi_1,...\phi_M)^T$ est un flot de s à p dans G si les lois de conservation aux nœuds sont vérifiées pour tous les sommets de G sauf aux sommets s et p où l'on a

$$\sum_{u\in \omega^+(s)} \phi_u = \sum_{u\in \omega^-(p)} \phi_u = \phi_0$$

ϕ_0 est la valeur du flot. Si ϕ est un flot de s à p dans G de valeur ϕ_0 alors $\phi'=(\phi_0,\phi_1,...\phi_M)^T$ est simplement un flot dans G^0.
Le problème de la recherche du flot maximum dans G revient à déterminer un flot ϕ' dans G^0 vérifiant les contraintes de capacités pour $u \in \{1,...M\}$ et tel que ϕ_0 soit maximum. L'algorithme de Ford & Fulkerson (1956) permet de rechercher un flot maximum

4.4 Problème de transbordement

Soit G=[X,U] un graphe orienté, supposé connexe. On suppose que G a N sommets. On distingue parmi les sommets de G des sommets *sources* et des sommets *puits*. Un sommet i est une source si il émet une quantité de données a_i et un sommet j est un puits si il requiert une quantité de données b_j pour effectuer une opération. Pour le problème qui nous concerne, les capteurs constituent les sources du graphe et les processeurs de traitement les puits.
Soit X_s l'ensemble des sommets source et X_p l'ensemble des sommets puits. On supposera dans la suite que : $X_s \cap X_p = \emptyset$
mais G peut posséder d'autres types de sommets ni source ni puits (c'est le cas par exemple les concentrateurs !)
Les quantités a_i et b_j sont supposées être des entiers et de plus dans un premier temps :

$$\sum_{i\in X_s} a_i = \sum_{j\in X_p} b_j$$

A chaque arc u=(i,j) du graphe, on associe un coût γ_u représentant le prix de passage d'une unité de flot sur l'arc. Les capacités des arcs sont supposées infinies. En pratique, cela signifie que les lignes de communications sont choisies de manière à ne pas être saturées.
Le problème que l'on se pose est de trouver un flot ϕ sur G vérifiant les contraintes aux sources et aux puits et de coût total minimum donc tel que :

$$\begin{cases} \sum_{u\in \omega^+(i)} \phi_u = a_i \;\forall\, i \in X_s \\ \sum_{u\in \omega^-(j)} \phi_j = b_j \;\forall\, j \in X_p \end{cases}$$

On peut transformer le problème en une recherche de flot à coût minimum sur un graphe ne possédant qu'une seule source et qu'un seul puits. Pour cela on ajoute à G un sommet s "super-source" relié à tous les sommets i de X_s par des arcs (s,i) de capacité a_i et de coût nul. On ajoute un sommet p "super-puits" relié à tous les puits j de X_p par un arc (j,p) de capacité b_j et de coût nul. On ajoute implicitement l'arc de retour (p,s).

On a ainsi un graphe G' et on dit que l'on a mis le problème sous forme canonique. Dans G' le problème revient à trouver un flot de valeur ϕ_0 entre s et p tel que

$$\phi_0 = \sum a_i = \sum b_j$$

de coût minimum et compatible avec les capacités a_i sur les arcs (s,i) et b_j sur les arcs (j,p).
L'algorithme de Ford & Fulkerson (1962) permet de trouver un flot optimum compatible.

4.5 Discussion

Les éléments de la théorie des graphes sont directement applicables pour l'étude de la compatibilité du réseau physique et de la logique de fusion.

Dans un premier temps, il faut trouver les composantes connexes du graphe associé au réseau de communication du système multi-senseurs. Seuls les capteurs qui font partie d'une même composante connexe pourront "fusionner" leurs informations.
Un graphe peut être connexe, mais il suffit qu'une liaison particulière soit coupée pour retirer cette propriété. De tels arêtes sont appelées des isthmes. Il existe un algorithme permettant de les trouver.
Si l'arc (I2,I4) est saturé ou détérioré, le graphe se retrouve avec deux composantes connexes et deux groupes de capteurs se distinguent alors : S1, S2, S3, S4 en liaison avec P1 d'une part, S5, S6, S7, S8 en liaison avec P2 d'autre part. De même, plus simplement, les capteurs S3 et S8 peuvent se retrouver isolés suite à une coupure de la liaison (S3,P1) ou (S8,P2).
Par contre les arcs (I1,I2) ou (I5,P2) n'ont aucune incidence sur la connexité du graphe.
Il apparaît donc nécessaire de favoriser les cheminements de données qui n'empruntent pas des isthmes.
En second lieu, dans le cas où les capacités des lignes sont supposées nettement supérieures aux besoins, un algorithme permet de trouver (si celui-ci existe) le flot solution de coût minimal au problème de transbordement. Le coût d'une liaison peut être une fonction de la fiabilité de la transmission, ou du temps de transmission.
Si la composante connexe considérée contient plusieurs processeurs, on pourra appliquer l'algorithme à ces différents puits et assigner finalement le traitement au processeur qui fournit le coût minimal. Mais cependant, il n'est pas nécessaire qu'un seul processeur effectue l'intégralité du traitement. Pour l'exemple considéré dans le chapitre précédent, on peut envisager d'appliquer l'algorithme pour chaque "fusion élémentaire". Elles seront implémentées alors sur les processeurs qui fourniront le coût minimal. Ainsi une première allocation sera effectuée. On procède ainsi en parcourant les différents niveaux de l'arbre logique. Si le nombre de processeurs est faible, cette euristique convient, sinon sa complexité croît vite.

5. CONCLUSION

Ce document a présenté l'outil ROSACE émanat de l'étude "Méthodologie multi-senseurs en défense anti-aérienne" menée pour le compte du GC/DA de la DRET.

En second lieu, le problème original de la compatibilité entre logique de fusion et réseau de communications physique, a été abordé. Même si il n'a été que partiellement traité, nous estimons qu'il est crucial et qu'il ouvre la voie vers un thème de recherche riche.
En effet, il est nécessaire de contraindre les algorithmes qui réalisent les tâches de détection, de pistage ou de classification, à la structure du réseau physique donc aux possibilités de transmission et d'échange des informations au sein du système multi-senseurs.
L'optimisation des algorithmes (afin d'obtenir les meilleures performances) ne peut se faire sans tenir compte de la fiabilité des lignes, des capacités et des délais de transmission, du coût.

BIBLIOGRAPHIE

[1] Darricau J.
Physique et théorie du radar
éditions SODIPE (1993)
[2] Gondran M. & Minoux M.
Graphes et algorithmes
Éditions Eyrolles (1979)
[3] Le Chevalier François
Principes de traitement des signaux radar et sonar
Collection technique et scientifique des télécommunications, éditions MASSON (1989)
[4] Le Tallec Claude
Recueil d'expertise SADAA
ONERA, Rapport technique RT 36/3522 SY (1995)
[5] Nahum Carole
Méthodologie multi-senseurs en défense anti-aérienne
ONERA RT 60/7244 SY (1993)
[6] Nahum Carole & Cantalloube Hubert
Conception de systèmes de surveillance et de défense anti-aérienne multi-senseurs
58th AGARD GCP Symposium Copenhague (1994)
[7] Nahum Carole & Cantalloube Hubert
Association de données 2D pour la localisation 3D
Colloque international sur le radar Paris (1994)
[8] Nahum Carole & Cantalloube Hubert
Évaluation du taux de détection par fusion d'informations
Colloque international sur le radar Paris (1994)
[9] Nahun Carole & Cantalloube Hubert
Evaluation of detection rates via data association
SPIE Symposium, Florida, U.S.A. (1994)
[10] Nahum Carole & Cantalloube Hubert
Méthodologie multi-senseurs en défense anti-aérienne II
ONERA RT 69/7244 SY (1995)

DEVELOPMENT OF A WIDEBAND AIRBORNE LASER DATA LINK

Robert A. Gill and Robert J. Feldmann
United States Air Force Research Laboratory
WL/AAZP
Wright-Patterson AFB OH 45433-7303

Summary: The United States Air Force Research Laboratory is conducting a program to develop a wideband airborne laser data link. The program successfully ground demonstrated a 1.1 gigabit/second (sec) full duplex data llink over a distance of 150 kilometers (km). This ground demonstration was accomplished in Hawaii in September 1995. The system used in the ground demonstration has been redesigned and will be installed in two jet aircraft for flight demonstration at distances up to 500 km. These demonstration flights will begin in September 1998.

Introduction: Modern battlefield strategy is predicated on knowing where the enemy's (or potential enemy's) assets are located and their operational capability. This vital information is constantly being gathered and updated by various ground, space, and airborne sensors. The requirement to send ever increasing amounts of tactical military information between sensor aircraft and information processing facilities has begun to press the limits of present airborne data links, even when data compression techniques are used. Utilization of optical data links is under consideration by the United States Air Force and development of a possible airborne laser data link is under way by the United States Air Force Research Laboratory Avionics Directorate (AFRL/WL/AA). The first part of this development process was a successful ground demonstration in September 1995 of a full duplex 1.1 gigabit/sec laser data link over a distance of 150 km between Mt. Haleakala and Mt. Mauna Loa in the state of Hawaii. This demonstration serves as the basis for a three year AFRL development program begun in August 1996, which will culminate in a series of air-to-air flight demonstrations beginning in September 1998.

Background: The feasibility of laser airborne data links was demonstrated in the mid-80's by the AFRL HAVE LACE (Laser Airborne Communications Experiment) Program. This program developed and tested two laser communications terminals that operated at 19.2 kilobits/sec. The terminals were tested using two KC-135 aircraft that nominally flew at 20,000 to 25,000 feet (ft) altitudes with separation distances out to 160 km. The most significant result of the HAVE LACE flights was the difficulty of initial signal acquisition between the two moving platforms, since it had to be performed manually. However, once signal acquisition was accomplished, tracking proved to be robust and communications performance was consistently measured at 10^{-6} bit error rate (BER) or better.

Since the HAVE LACE program, laser terminal development and data rates have improved dramatically. The Ballistic Missile Defense Organization (BMDO) Innovative Science and Technology (IST) office has, since 1988, invested in the development of laser communications technology primarily for space based laser crosslinks. Two laser communications terminals (LCT) built by ThermoTrex Corp. for BMDO under that development effort were the basis of the ground laser data link system demonstrated in September 1995 in Hawaii.

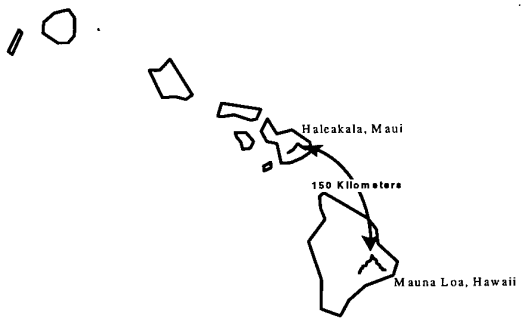


Figure 1. Hawaii Test Location

Ground Laser Data Link System Description: The laser data link demonstration equipment on Mt. Haleakala consisted of a ThermoTrex Corp. built

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

LCT along with its associated control electronics, a Lockheed Martin Corp. built Common Data Link (CDL) Airborne Modem Assembly (AMA), a Litton Corp. built P-IIIA Inertial Navigation Unit (INU), a McFadden Systems Inc. built Model 612A hydraulic motion base simulator, and electromagnetic shakers (along with their control electronics) which were provided by the AFRL. The laser data link equipment on Mt. Mauna Loa was similar to the Mt. Haleakala equipment setup but it was mounted on a fixed, stationary base. The equipment installation on Mt. Haleakala provided the motion and vibration spectrum to simulate operation from a high altitude aircraft, while the Mauna Loa installation served to simulate a fixed ground terminal. The LCT was the major developmental item of interest for the ground demonstration while the other equipment had been readily available for several years and were considered nondevelopmental items. The integration of this equipment, however, into a high data rate laser data link system was unique to this ground demonstration.

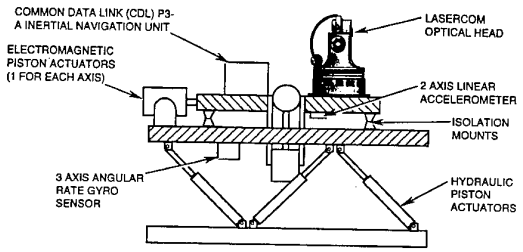


Figure 2. Motion Base Configuration

Ground Laser Communications Terminal Description: The ground demonstration LCT contained the various optical apertures for transmitting and receiving the optical signals and they were mounted in a single two axis gimbaled composite structural framework.

The receive telescope, a 14 centimeters (cm) diameter Schmidt-Cassegrain design with spherical primary and secondary mirrors, collected the incoming combined laser signals of 810 and 852 nanometers (nm) wavelength. A flat dichromatic mirror at the hole in the primary mirror split the high speed communication signal (810 nm) and the wavelength locked beacon signal (852 nm) and focused them onto their respective detectors. Light at 810 nm was reflected from the dichromatic mirror towards the front of the telescope where it was focused onto a 200 micron pinhole, which acted as a field stop. The light was relayed from the pinhole to optics located on the back of the secondary mirror that demultiplexed the signals and imaged them onto low noise EG&G SLIK avalanche photodiodes.

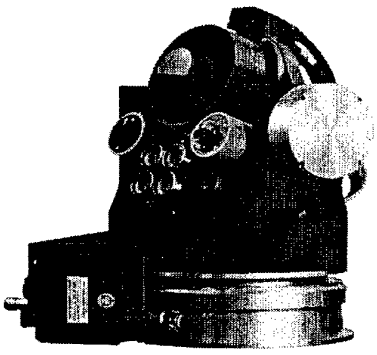


Figure 3. Ground Demo Laser Communication Terminal

Light at 852 nm was passed through the dichromatic mirror and directed to the rear of the telescope through an atomic line filter and was imaged on a charge coupled device (CCD) camera. The critical element in the operation of the LCT wide field of view (FOV), 1.25°, laser communication signal acquisition scheme was the use of an atomic line Faraday filter which was placed in front of a Dalsa 256 x 256 pixel, 225 Hertz (Hz) frame rate tracking CCD camera. The Faraday filter bandwidth is 0.019 nm, with an in band transmission of 70% (using polarized light) and an out of band rejection of better than 10^{-5}. Use of the Faraday filter required that the two LCT beacon lasers be locked to the transmission peak wavelength of the filter, 852 nm. This was accomplished by temperature and current stabilization and optical feedback through a reference atomic line filter. There were two 50 milliwatt beacon diode lasers and four 140 milliwatt peak communications diode lasers, two of which were left hand circularity polarized and two of which were right hand circularity polarized, for transmitting laser signals from the terminal. This provided two independent optical channels for high speed data, each channel capable of over 500 megabit/sec for a cumulative data link rate of over 1 gigabit/sec. The beacon lasers had a divergence of 2 milliradians while the communication lasers had a divergence of approximately 100 microradians. All of the

apertures were coaligned and moved together when the gimbals were moved.

The gimbaled framework was moved in azimuth and elevation by two direct drive pancake DC motors through two 23 cm angular contact bearings. A Canon angular shaft position optical encoder with a 1 arcsecond (4.484 microradians) resolution sampled at 350 millisecond intervals was used to derive gimbal position. A Type II control loop design was used to accomplish signal source tracking. There were two nested loops, an inner stability loop (which also provided inertial pointing correction) and an outer CCD tracking loop).

The inner stability loop incorporated a two axis angular rate gyro which stabilized open loop gimbal movements and provided a correction signal for maintaining a constant line of sight gimbal orientation while tracking motion that had a higher bandwidth than the CCD centroid data. The outer control loop was compromised of the tracking CCD camera, its associated thresholding electronics, and a software implemented centroid calculator and position algorithm. This digital signal processor error signal was low pass filtered, scaled, and summed into the inner loop to provide low bandwidth (less than 3 Hz) pointing correction for the tracking system.

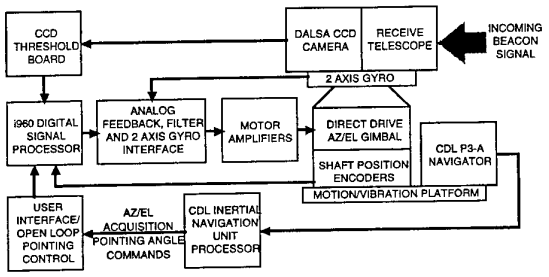


Figure 4. Tracking System Block Diagram

The electronics associated with the LCT consisted of a VME cage housing cards for beacon laser control, communications laser drive, and digitized video conversion. A second VME cage had the tracking and gimbal control cards, consisting of an i960 DSP, a Dalsa CCD camera drive, a CCD threshold and storage, a Cannon optical encoder interface, a motor amplifier, an analog filter and gyro interface, and a Radisys 486 imbedded PC for user and communications interface. There were two single rack mounted C-Cor units that performed the digitized video multiplexing to 560 megabits/sec and demultiplexing before digital to analog conversion for one optical channel. Another rack mounted unit provided a repeating bit pattern at 560 megabits/sec and had a bit pattern receiver for testing the second channel. An interface to the CDL system's INU processor provided azimuth and elevation pointing angles based on ephemeris data to the Mt. Haleakala LCT for initial pointing prior to optical tracking. The LCT disregarded the INU pointing information when the system was optically tracking the signal from the other terminal. The critical direction for the high bandwidth data transmission for this demonstration was from the simulated "aircraft" to the ground, which required high accuracy tracking since the transmitted communication lasers had less divergence than the communication receiver's field of view. The tracking system was required to remain pointed to the Mauna Loa terminal to within 40 microradians (±20 microradians peak from center).

Ground Test Description: The Mt. Haleakala LCT was mounted on a six degree of freedom motion base platform. The platform consisted of a table supported by six hydraulic legs oriented at various angles. The platform provided up to 20° of pitch, roll, and yaw motions and a limited amount (approximately 30 cm) of lateral motion. On top of this platform were three electromagnetic linear actuators oriented at right angles to each other, all attached to a single phenolic plate supported by isolation mounts. These linear actuators provided translational vibration motion with respect to the table of the hydraulic base. They also created angular vibration due to softness in the platform mounts and the slight off center of mass mounting of the actuator pistons to the plate. The hydraulic platform provided a vibration spectrum of DC to 20 Hz, while the electromagnetic actuators operated from 20 Hz to 200Hz. Motion sensors were mounted on the phenolic plate and the table to enable feedback control and measurement of the platform vibration spectrum to verify its matching with the actual high altitude aircraft electronics bay vibration spectrum specified for the ground demonstration.

The two Hawaii mountains were chosen as the demonstration location because of their clear atmospheric conditions, the long ground path (150 km), and the high elevation (10,000 and 11,200 ft) of the test sites. Due to the curvature of the earth

over this distance, the laser beams dipped to approximately 8,000 ft altitude mid way between the two mountains. The demonstration setup commenced on 24 August 1995 and because of low cloud cover, link closure did not occur until 3 September 1995 with all subsequent testing completed by 13 September 1995. At that time of year in Hawaii, a thick cloud layer extending from approximately 3,000 to 8,000 ft altitude is omnipresent between the islands of Hawaii and Maui. The actual link tests took place at various times of the day in an attempt to avoid these unfavorable weather conditions with operations between 4 am and 10 am appearing to be optimum. Weather conditions during the tests ranged from 5° Celsius and heavy winds at 4 am in the morning to almost 30° Celsius, calm and sunny at 12 noon. This morning atmospheric heating tended to raise the cloud layer to above 8,000 ft, thus blocking the beam path and ending the test for the day.

Testing was initiated by directing the Mauna Loa LCT to point to the Haleakula LCT site while simultaneously energizing the motion base and LCT on Haleakula. The Haleakula LCT would initiate with an arbitrary pointing direction which was corrected by the INU to point the beacon lasers to within the FOV of the Mauna Loa acquisition receiver which then responded with a beacon laser signal. Once the beacon laser signal was detected by the acquisition receiver, of either terminal, the terminal was steered to within the FOV of the communications receiver and tracking of the beacon signal was maintained within this FOV.

Ground Demonstration Results: The laser communication terminal pointing, acquisition, tracking, and reacquisition in a simulated high altitude aircraft vibration and turning maneuver environment were successfully demonstrated. Duplex operation of the terminals was tested while simulating a turn of 20° roll and 15° yaw. The maximum roll rate was approximately 8° per sec. Under the full dynamic motion of the moveable platform, acquisition of the laser signal occurred in less than 2 sec and reacquisitions were consistently performed in less than 0.25 sec. The tracking errors were estimated to be 110 microradians peak in azimuth and 60 microradians peak in elevation. These numbers were somewhat higher than the 80 microradians that was measured during previous ground acceptance tests in the integration laboratory primarily because of image blooming on the CCD due to varying intensity of the received beacon signal after propagation over the long atmospheric path between the islands. Even though there was software to compensate for image blooming, the algorithm proved to be too simple and produced a perceived error of only 80 microradians for large intensity blooms.

The simulated aircraft LCT on Mt. Haleakala sent two channels of high-speed data. One channel was CDL video data at 274 megabits/sec and the other channel was multiplexed uncompressed digitized video at 560 megabits/sec. The Mauna Loa LCT was simultaneously transmitting a 143 megabits/sec digitized video signal on one channel and a low rate CDL audio link on the other channel. The lasers were modulated using on/off keying for the digitized video channel and intensity modulation was used for the CDL channel. No forward error correction (FEC) coding was employed on either channel. There was no provision for measuring BER directly from the received communication signal which was primarily video. Estimating the link quality from the video display was also difficult because of the effects of resynchronization delays of the video demultiplexer, which was designed for a fiber optic channel and was not optimum for the frequent burst errors which were experienced. The average BER was greater than 10^{-6}. However, there were periods of exceptionally good signal quality. The LCT was designed to have 10 dB of signal margin to combat scintillation fading. Preliminary analysis of the scintillation fade statistics as measured over the propagation path indicates that an additional 15 dB of margin would have been necessary to achieve a BER of 10^{-6}, given that no FEC coding was used. A link margin of 25 dB is too large a factor to achieve by simply increasing the laser power (for cw diode lasers) and therefore the implementation of an error correction scheme is clearly indicated in order that acceptable BER performance be obtained in such a ground test environment.

Ground Demonstration Conclusions: A formal ground laser data link demonstration was successfully conducted for sponsoring governmental agencies and industry on 7 and 8 September 1995. The dual channel, full duplex capabilities of the LCTs were demonstrated during a full simulated aircraft vibration spectrum coupled to a simulated 20° roll (8°/sec rate) and 15° yaw maneuver of the motion platform. Initial signal acquisition, reacquisition (following both

short and long term outages), and tracking stabilization were also demonstrated.

The communications performance exhibited occurrences of large signal degradation due to atmospheric scintillation effects. While the signal fades due to the atmosphere at the Hawaii test sites were in excess of 25 dB, it is believed that they could be mitigated through a combination of increased signal power and FEC coding. The ground test results were sufficient proof of concept of laser data link acquisition and tracking. However, it was obvious that an LCT redesign was required to accomplish long range communications between two high altitude aerial platforms under operational scenarios. An analysis of high altitude long horizontal path fade statistics was also required to determine the best combination of signal and processing gain that would provide a robust communication link with acceptable BER performance.

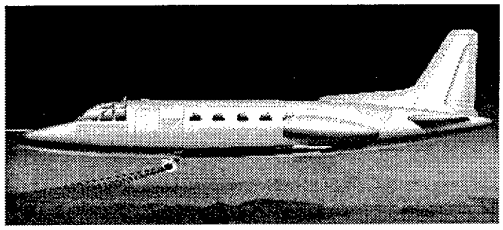


Figure 5. Airborne LCT Test Vehicle

Airborne Demonstration Requirements: The next phase in the development of the wideband airborne laser data link is the design, fabrication, and flight testing of an airborne LCT which will demonstrate 1.1 gigabit/sec communications between two aerial platforms at a range of 500 km. The LCT is to have at least a hemispherical field of regard (FOR) during laser data link operation from each platform. Two business size jet aircraft are to be the aerial platforms and are to operate at up to 40,000 ft altitude for the flight demonstration. An open and competitive AFRL contract solicitation resulted in the selection of TrexCommunications Corp. of San Diego, CA to design and fabricate the airborne LCT and Veda Inc. of Dayton, OH to provide the vehicle integration and flight operations. The two contracts were initiated in August 1996 and the LCT critical design review completed in June 1997.

Airborne Laser Data Link System Description: The requirements to have the LCT on a business size jet aircraft, a T-39A or Sabreliner 40 in this case, and the need for a hemispherical FOR drove the size, weight, and installation constraints of the design for the airborne LCT. The LCT is contained within a ball turret mounted on the underside of the aircraft fuselage.

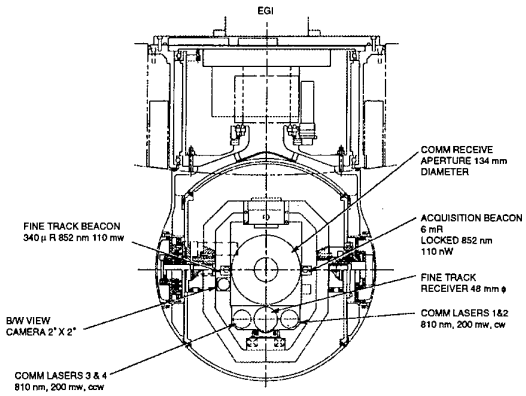


Figure 6. Ball Turret Assembly

The 48 cm diameter aluminum ball turret, containing the LCT, is structurally mounted in the lower escape hatch opening of the aircraft, thus requiring minimal aircraft modification. This installation provides a slightly greater than hemispherical FOR since the LCT within the turret has 360° azimuth coverage and 110° elevation coverage (20° above horizon to straight down or nadir). The turret follows the guidance commands to the LCT and provides the large angle, coarse steering function with a 7 Hz bandwidth. Within the turret is a tracking gimbal assembly for coarse tracking within its 12° range of motion and 50 Hz bandwidth. Mounted within the gimbal assembly is the actual LCT optical bench containing the beacon and communication lasers, fast steering mirror, the beacon and communications detectors, and a black and white viewing camera. The fast steering mirror has a 500 microradians range of motion with a 3 kHz bandwidth. This four-gimbal axis design (six axis when considering the fast steering mirror) allows the LCT to track through nadir smoothly and provides high bandwidth tracking to overcome aircraft kinematics and vibration. Atop the turret azimuth drive motor structure, mounted on a solid aluminum plate, is a Honeywell H-764G Embedded GPS Inertial (EGI) unit. The EGI, using GPS P-Code, provides the platform reference necessary to stabilize the LCT and the aircraft location information necessary for initial data link acquisition.

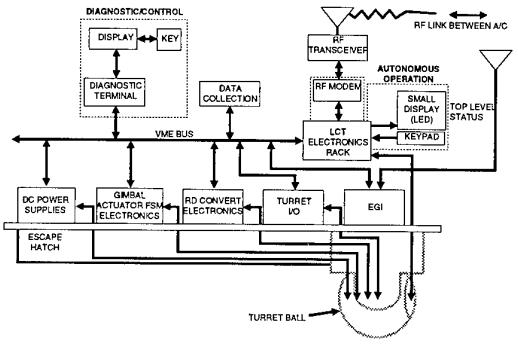


Figure 7. System Block Diagram

Airborne Laser Communications Terminal Description: The airborne LCT contains the various optical apertures for transmitting and receiving the optical signals. They are mounted on a composite optical bench in a single two axis gimbaled aluminum structural framework which is in turn mounted within a two axis gimbaled aluminum ball turret. The airborne LCT is very similar to the ground demonstration LCT in that the same beacon and communication lasers (852 and 810 nm) are used. Instead of the 14 cm diameter Schmidt-Cassegrain telescope, a 14 cm diameter Maksutov telescope is used. The optical bench is a new design and the layouts of the optical paths are different. Additional components are used to increase tracking resolution and received signal strength. An analysis of a high altitude (40,000 ft) long horizontal path (400 km) indicated that the airborne LCT using the four 810 nm communications lasers produces a link margin of 9 dB. Other wavelengths (e.g. one micron) were analyzed because of their higher laser output power and commercial availability. However, because of the need to mount more complex detectors and their associated expense, the ground demonstration LCT lasers and detectors were deemed preferable for the airborne LCT design.

The detection scheme for beacon tracking is different in the airborne LCT. Whereas the ground demonstration LCT used two 50 milliwatt 852 nm beacons with a divergence of 2 milliradians for both coarse and fine tracking; the airborne LCT uses one 110 milliwatt 852 nm beacon with a divergence of 6 milliradians for coarse tracking and one 110 milliwatt 852 nm beacon with a divergence of 340 microradians for fine tracking. The coarse beacon signal is still received through a 14 cm diameter telescope and passed through an atomic line filter onto a 288 X 384 pixel CCD detector running at 100 Hz. However, a separate 5 cm aperture telescope is used to receive the fine track beacon signal and pass it off a 700 Hz bandwidth piezoelectric fast steering mirror onto a separate 32 x 32 pixel CCD detector running at 10 kHz. The four 100 milliwatt 810 nm communication laser transmit signals are also aligned by the same fast steering mirror so that the communication signals are sent back along the same direction as the received fine track beacon signal. This coupling of the incoming beacon and communication transmit signals through the same high bandwidth fast steering mirror provides quick beam path correction and no reliance on an error feedback control loop with its associated time delay. The communication signal processing has been redesigned to overcome the long signal dropouts due to resynchronization delays of the video demultiplexer encountered during the Hawaii ground demonstration. The C-Cor unit used for multiplexing/demultiplexing the video signal in the ground demonstration LCT has been replaced in the airborne LCT by an IPiTEK IMTRAN CQ-508 unit. This unit provides resynchronization times in sub-milliseconds, thus improving the overall BER.

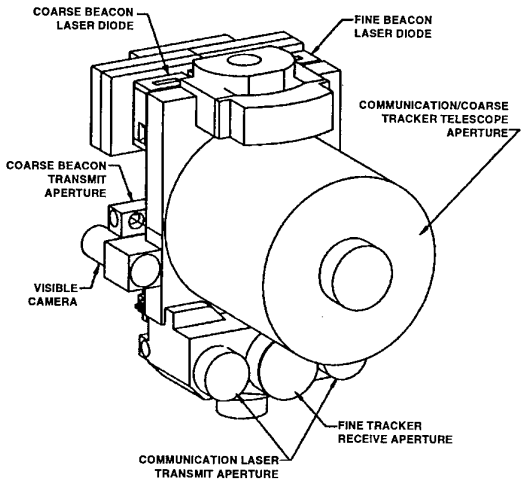


Figure 8. Optical System

The position resolvers and drive motors of the ground demonstration LCT have been upgraded for quicker response and higher torque. Flex pivots are used instead of bearings on the gimbal assembly to reduce static friction. The ball turret contains high torque azimuth and elevation motors to quickly respond to steering commands

under varying aerodynamic loads. The airborne LCT control loops are more complex than the ground demonstration LCT units because of the dynamic aircraft environment and redesign of the airborne LCT. These multiple new control algorithms include such functions as initial signal acquisition, platform stabilization, turret positioning, coarse track steering, fine track steering, short term search and reacquisition, and long term reacquisition. These algorithms are in the process of being written and programmed into software. The algorithms and their associated processing are accomplished in the dedicated System Processor mounted in the LCT Electronics Rack within the test aircraft. The System Processor also serves as the interface to the onboard flight test instrumentation system. A separate digital processor is also mounted in this rack to accomplish LCT diagnostics.

Air-To-Air Test Description: Two T-39A (or Sabreliner 40) aircraft will be modified during the summer of 1998 and a LCT integrated into each aircraft. Flight testing is scheduled to begin with aircraft flightworthiness tests in September 1998 at Crestview, FL. This will be followed by a series of flights at Dayton, OH and other locations to evaluate beacon signal acquisition and tracking, communication signal acquisition and quality, link atmospheric effects, aircraft kinematic and airflow effects, and operational mission requirements. The flight testing is an integral part of the LCT development process; therefore, the LCT is extensively instrumented and each test aircraft contains a complete and comprehensive onboard flight test instrumentation system to monitor in real-time LCT operation and the data link quality. Each test mission is expected to consist of approximately two hours of data gathering at various test conditions.

Initial air-to-air data link operations will be controlled by an onboard test engineer and monitored by an onboard systems engineer. A pilot and copilot/test director will also be part of the onboard crew. Later flights, such as the operational mission requirements test flights, will delete the onboard engineers as the LCT is operated in an autonomous mode. The saving in aircraft gross weight will allow higher altitude flight and subsequent longer link ranges.

The procedure for the conduct of each test flight is to first accomplish a ground checkout and alignment calibration prior to takeoff. Safety switches have been incorporated into the laser power circuits to prevent autonomous laser operation on the ground since the eye safe distance is approximately 1.5 km. The ball turret will be in a stowed position with the LCT inoperative during takeoff and landing to prevent turret window damage and to decrease system cooling requirements. When the two aircraft have reached their test altitude and range, the crew of the "master" test aircraft will request a position fix from the "slave" aircraft via an omnidirectional radio frequency (RF) data link. A one time message exchange of each aircraft's position and velocity vector then takes place. This information is used by each LCT to acquire the other aircraft's coarse track beacon signal. A transition is then made to the fine track beacon signal and link communications is initiated. This acquisition sequence is expected to be accomplished in less than 10 sec. If a test aircraft loses the track beacon signal, reacquisition is performed using a propagated vector of the other aircraft based on a track file. If reacquisition does not occur within one sec, a spiral search pattern is initiated by the LCT for 10 sec. If there is still no reacquisition after 10 sec, a new RF transmitted data point is obtained and the initial acquisition process is repeated. The same RF data link is used continuously by each aircraft's flight test instrumentation system to display the other aircraft's location to the flight crew. This RF data is not fed continuously to the LCT Systems Processor for a track file; the track file is based only upon information derived from the track beacon signals. Later operational mission requirements test flights, each aircraft will attempt to accomplish beacon signal acquisition based upon a priori knowledge of each aircraft's location at a designated time and will not utilize the RF data link in the acquisition process. The 1.1 gigabit/sec laser communication signals will consist of four video/audio channels, a serial digital data stream, and a BER generator data stream.

Expected Air-To-Air Demonstration Results: The data collected during the test flights should prove the ability of a wideband airborne laser data link to communicate in the upper atmosphere to 500 km with a BER of 10^{-6}. The tests should provide data on atmospheric attenuation and beam scintillation. Data on the effects of aircraft airflow upon beam steering should also be gathered. This information will be made readily available to other laser communication development efforts (e.g. an air-to-space capability).

Fusion of Visible and Thermal Imagery Improves Situational Awareness

A. Toet, J.K. IJspeert

TNO Human Factors Research Institute
Kampweg 5, 3769 DE Soesterberg, The Netherlands

A.M. Waxman, M. Aguilar

M.I.T. - Lincoln Laboratory
Lexington, MA 02173 USA

1. SUMMARY

A new colour image fusion scheme is applied to visible and thermal images of military relevant scenarios. An observer experiment is performed to test if the increased amount of detail in the fused images can improve the accuracy of observers performing a detection and localisation task. The results show that observers can localise a target in a scene (1) with a significantly higher accuracy, and (2) with a greater amount of confidence when they perform with fused images (either gray or colour fused), compared to the individual image modalities (visible and thermal).

2. INTRODUCTION

Scene analysis by a human operator may benefit from a combined or fused representation of images of the same scene taken in different spectral bands. For instance, after a period of extensive cooling (e.g. after a long period of rain or early in the morning) the visible bands may represent the background in great detail (vegetation or soil areas, texture), while the infrared bands are less detailed due to low thermal contrast in the scene. In this situation a target that is camouflaged for visual detection cannot be detected in the visible bands, but may be clearly represented in the infrared bands when it is warmer or cooler than its environment. The fusion of visible and thermal imagery on a single display may then allow both the detection and the unambiguous localisation of the target (provided by the thermal image) with respect to the context (provided by the visible image). The abovementioned line of reasoning is frequently adopted to promote image fusion, and has resulted in an increased interest in image fusion methods, as is reflected in a steadily growing number of publications on this topic [8–13]. A large effort has been spent on the development of these new image fusion methods. However, until now there are no validation studies that investigate the applicability domain and the practical use of these techniques. Ultimately the performance of a fusion process must be measured as the degree to which it enhances a viewers ability to perform certain practical tasks. Preliminary results on enhanced detection of targets embedded into real scenes [3] have shown potential benefits of fusion. The present study is performed (a) to investigate the conditions for which the fusion of visible and thermal images may result in a single composite image with extended information content, and (b) to test the capability of a recently developed colour image fusion scheme [2–7] to enhance the situational awareness of observers operating under these specific conditions.

3. METHODS

3.1. Image capture

3.1.1. Apparatus

The *visible-light* camera was a Siemens K235 Charge Coupled Device (CCD) video camera, with a 756 × 581 CCD chip, and equipped with a remotely controlled COSMICAR C10ZAME-2 (Asahi Precision Co. Ltd., Japan) zoom lens (f = 10.5–105 mm; 1:1,4). The *infrared* (IR) camera was an Amber Radiance 1 (Goleta, CA, U.S.) Focal Plane Array (FPA)-camera, with an array of 256 × 256 pixels, operating in the 3–5 μm (mid-range) band, and equipped with a 100 mm f/2.3 Si-Ge-lens. Each pixel corresponds to a square 1.3 min of arc wide instantaneous field of view. The entire array of 256 × 256 pixels therefore corresponds to a field of view of about 5.6 degrees wide.

The CCD and IR images must be aligned before they can be fused. The signals of the thermal and visual cameras are therefore spatially registered as closely as possible. A 2nd order affine warping transformation is applied to map corresponding points in the scene to corresponding pixel locations in the image plane.

Paper presented at the AGARD SPP Symposium on "Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar", held in Lisbon, Portugal, 29 September - 2 October 1997, and published in CP-595.

3.1.2. Conditions

The recording period is just before sunrise, and the atmosphere was slightly hazy. The visual contrast is therefore low. The thermal contrast is low because most of the objects in the scene have about the same temperature after having lost their excess heat by radiation during the night.

3.2. Image fusion

The computational image fusion methodology is developed at the MIT Lincoln Laboratory [2–7] and derives from biological models of colour vision and fusion of visible light and infrared (IR) radiation.

In the case of colour vision in monkeys and man, retinal cone sensitivities are broad and overlapping, but the images are contrast enhanced *within bands* by spatial opponent processing (via cone-horizontal-bipolar cell interactions) creating both ON and OFF center-surround response channels [14]. These signals are then contrast enhanced *between* bands via interactions among bipolar, sustained amacrine, and single-opponent colour ganglion cells [15,16], all within the retina.

Fusion of visible and thermal IR imagery has been observed in the optic tectum of rattlesnakes and pythons [17,18]. These neurons display interactions in which one modality (e.g. IR) can enhance or depress the response to the other sensing modality (e.g. visible) in a strongly nonlinear fashion. Such interactions resemble opponent-processing between bands as observed in primate retina.

For opaque surfaces in thermodynamic equilibrium, spectral reflectivity ρ and emissivity ϵ are linearly related at each wavelength λ: $\rho(\lambda) = 1 - \epsilon(\lambda)$. This provides a rationale for the use of both on-center and off-center channels when treating infrared imagery as characterized by thermal emissivity.

In the colour image fusion methodology the individual input images are first enhanced by filtering them with a feedforward center-surround shunting neural network [19].This operation serves to (i) enhance spatial contrast in the individual visible and IR bands, (ii) to create both positive and negative polarity IR contrast images, and (iii) to create two types of single-opponent colour contrast images. The resulting single-opponent colour contrast images represent grayscale fused images that are analogous to the IR-depressed visual and IR-enhanced visual cells of the rattlesnake [17,18].

To obtain a natural colour representation of these single-opponent images (each being an 8-bit grayscale image) we have developed two alternative methodologies to choose from, based on the relative resolution of the visible and IR images.

In the case where the IR camera is of significantly lower resolution (i.e. half that of the visible camera), the enhanced visible is assigned to the green channel, the difference signal of the enhanced visible and IR images is assigned to the blue channel, and the sum of the visible and IR images is assigned to the red channel of an RGB display [2–6]:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \mathrm{CCD}^+ & + & \mathrm{IR}^+ \\ \mathrm{CCD}^+ & & \\ \mathrm{CCD}^+ & - & \mathrm{IR}^+ \end{pmatrix} \qquad (1)$$

where the $(\ldots)^+$ indicates the center-surround operation. These channels correspond with our natural associations of warm (red) and cool (blue).

In the case where the IR and visible cameras are of comparable resolution, the enhanced sum of the enhanced visible and IR images is assigned to the green channel, the difference signal of the enhanced visible and infrared images is assigned to the blue channel, and the difference of the enhanced infrared and visible images is assigned to the red channel of an RGB display [7]:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \mathrm{IR}^+ & - & \mathrm{CCD}^+ \\ (\mathrm{CCD}^+ & + & \mathrm{IR}^+)^+ \\ \mathrm{CCD}^+ & - & \mathrm{IR}^+ \end{pmatrix} \qquad (2)$$

These transformations are followed by hue and saturation remapping. The result is a more natural appearing colour fused representation. Because the resolution of the cameras used for our experiments are comparable, this is the methodology chosen to process the imagery used here. At MIT-LL this approach to image fusion has been implemented on a dual-C80 hardware platform capable of processing 30 fps at 640 × 480 resolution.

The grayscale fused images are produced by taking the luminance component of the corresponding colour fused images.

3.3. Stimuli

The stimuli used in this experiment are 5 different types of images:

- graylevel images representing the signal of the video (CCD) camera,

- graylevel images representing the signal of the infrared (IR) camera,
- colour images representing the result of the fusion of corresponding CCD and IR image pairs (i.e. the combination of CCD and IR images of the same scene and registered at the same instant),
- graylevel images representing luminance component of the abovementioned colour fused images, and
- schematic graylevel images, representing segmented versions of the original visual (CCD) images.

The graylevel images are quantized to 8 bits. The colour images are quantized to 24 bits (8 bits for each of the RGB channels), and bitmapped to a 256 colour map, adding some colour dither.

The individual images correspond to successive frames of a time sequence. The time sequences represent 3 different scenarios. These scenarios were developed by the Royal Dutch Army [20]. They simulate typical surveillance tasks and were chosen because of their military relevance.

The corresponding schematic images are constructed from the visual images by

- applying standard image processing techniques like histogram equalization and contrast stretching to enhance the representation of the reference contours in the original *visual* images,
- drawing the contours of the reference features (judged by eye) on a graphical overlay on the contrast enhanced visual images, and
- filling the contours with a homogeneous graylevel value.

The images thus created represent segmented versions of the visual images.

Scenario I corresponds to the guarding of a UN camp [20], and involves monitoring a fence that encloses a military asset. To distinguish innocent bypassers from individuals planning to perform subversive actions the guard must be able to determine the exact position of a person in the scene at any time. During the image acquisition period the fence is clearly visible in the CCD image. In the IR image however, the fence is merely represented by a vague haze. A person (walking along the fence) is clearly visible in the IR image but can hardly be distinguished in the CCD image. In the fused images both the fence and the person are clearly visible. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the fence.

Scenario II corresponds to guarding a temporary base [20]. Only a small section of the dune like terrain is visible, the rest is occluded by trees. The assignment of the guard is to detect and counter infiltration attempts in a very early stage. During the registration period the trees appear larger in the IR image than they really are because they have nearly the same temperature as their local background. In the CCD image however, the contours of the trees are correctly represented. A person (crossing the interval between the trees) is clearly visible in the IR image but is represented with low contrast in the CCD image. In the fused images both the outlines of the trees and the person are clearly visible. As a result it is difficult to determine the position of the person relative to the trees using either the CCD or the IR images. The fused images correctly represent both the contours of the trees and the person. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the midpoint of the interval delineated by the contours of the trees that are positioned on both sides of the person.

Scenario III corresponds to the surveillance of a large area [20]. The scene represents a dune landscape, covered with semi-shrubs and sandy paths. The assignment of the guard is to detect any attempt to infiltrate a certain area. During the registration period the sandy paths in the dune area have nearly the same temperature as their local background, and are therefore represented with very low contrast in the IR image. In the CCD image however, the paths are depicted with high contrast. A person (walking along a trajectory that intersects the sandy path) is clearly visible in the IR image but is represented with less contrast in the CCD image. In the fused images both the outlines of the paths and the person are clearly visible. It is difficult (or even impossible) to determine the position of the person relative to the sandy path he is crossing from either the IR or the CCD images. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the sandy path.

3.4. Apparatus

A Pentium 100 MHz computer, equipped with a Diamond SVGA board, is used to present the stimuli, measure the response times and collect the observer responses. The stimuli are presented on a 17 inch

Vision Master (Iiyama Electric Co., Ltd) colour monitor, using the 640 × 480 pixels mode and a 100 Hz refresh rate.

3.5. Procedure

The subject's task is to assess from each presented image the position of the person in the scene relative to the reference features.

In Scenario I the reference features are the poles that support the fence. These poles are clearly visible in the CCD images (Fig. 1a), but not represented in the IR images (Fig. 1b) because they have almost the same temperature as the surrounding terrain. In the (graylevel and colour) fused images (Figs. 1c and d) the poles are again clearly visible.

In Scenario II the outlines of the trees serve to delineate the reference interval. The contours of the trees are correctly represented in the CCD images (Fig. 2a). However, in the IR images (Fig. 1b) the trees appear larger than their physical size because they almost have the same temperature as the surrounding soil. As a result, the scene is incorrectly segmented after quantization and it is not possible to perceive the correct borders of the area between the trees. In the (graylevel and colour) fused images (Figs. 2c and d) the outlines of the trees are again correctly represented and clearly visible.

In Scenario III the area of the small and winding sandy path provides a reference contour for the task at hand. This path is represented at high contrast in the CCD images (Fig. 3a), but it is not represented in the IR images (Fig. 3b) because it has the same temperature as the surrounding soil. In the (graylevel and colour) fused images (Figs. 3c and d) the path is again clearly visible.

For each scenario a total of 9 frames is used in the experiment. In each frame the person is at a different location relative to the reference features. These different locations are equally distributed relative to the reference features. Each stimulus is presented for 1 s, centered on the midpoint of the screen, preceded by a blank screen with an awareness message, which is presented for 1 s. A schematical representation of the reference features is shown immediately after each stimulus presentation (Figs. 1–3(f)). The position of the center of the reference image is randomly displaced around the center of the screen between presentations to ensure that subjects can not use prior presentations as a frame of reference for detection and localization. A complete run consists of 135 presentations (5 image modalities × 3 scenarios × 9 frames per scenario), and typically lasts about one hour.

The subject's task is to indicate the perceived location of the person in the scene by placing a mouse controlled cursor at the corresponding location in this schematical drawing. The subject has in principle unlimited time to reach a decision. When the left mouse button is pressed the computer registers the coordinates corresponding to the indicated image location (the mouse coordinates) and computes the distance in the image plane between the actual position of the person and the indicated location. The subject presses the right mouse button if the person in the displayed scene has not been detected. The subject can only perform the localization task by memorizing the perceived position of the person relative to the reference features.

The schematic reference images are also used to determine the optimal (baseline) localisation accuracy of the observers. For each of the three scenarios a total of 9 baseline test images (Figs. 1–3e) are created by placing a binary (dark) image of a walking person at different locations in the reference scene. The different locations of the person in these images are equally distributed over the entire reference interval. The image of the walking person was extracted from a thresholded and inverted thermal image. In the resulting set of schematic images both the reference features and the person are highly visible. Also, there are no distracting features in these images that may degrade localisation performance. Therefore, observer performance for these schematic test images should be optimal and may serve as a baseline to compare performance obtained with the other image modalities.

A complete run consists of 162 presentations (6 image modalities × 3 scenarios × 9 frames per scenario) in random order, and typically lasts about one hour.

3.6. Subjects

A total of 6 subjects, aged between 20 and 30 years, serve in the experiments reported below. All subjects have normal (or corrected to normal) vision, and no known colour deficiencies.

3.7. Viewing conditions

Viewing is binocular. The experiments are performed in a dimly lit room. The images are projected onto the screen of a CRT display. This screen subtends a

viewing angle of 25.5 × 19.5 degrees at a viewing distance of 0.60 m.

4. RESULTS

Figure 4 shows that subjects are uncertain about the location of the person in the scene for about 20% of the visual image presentations and 22% of the thermal image presentations. The (graylevel and colour) fused images result in a smaller fraction of about 13% "not sure" replies. The lowest number of "not sure" replies is obtained for the baseline reference images: only about 4%. This indicates that the increased amount of detail in fused imagery does indeed improve an observer's subjective situational awareness.

Figure 5 shows the mean weighted distance between the actual position of the person in each scene and the position indicated by the subjects (the perceived position), for the visual (CCD) and thermal (IR) images, and for the graylevel and colour fusion schemes. This Figure also shows the optimal (baseline) performance obtained for the schematic test images representing only the segmented reference features and the walking person. A low value of this mean weighted distance measure corresponds to a high observer accuracy and a correctly perceived position of the person in the displayed scenes relative to the main reference features. High values correspond to a large discrepancy between the perceived position and the actual position of the person. In all scenarios the person was at approximately 300 m distance from the viewing location. At this distance one pixel corresponds to 11.4 cm in the field.

Figure 5 shows that the localisation error obtained with the fused images is significantly lower than the error obtained with the individual thermal and visual image modalities (p=0.0021). The smallest errors in the relative spatial localisation task are obtained for the schematic images. This result represents the baseline performance, since the images are optimal in the sense that they do not contain any distracting details and all the features that are essential to perform the task (i.e. the outlines of the reference features) are represented at high visual contrast. The lowest overall accuracy is achieved for the thermal images. The visual images appear to yield a slightly higher accuracy. However, this accuracy is misleading since observers are not sure about the person in a large percentage of the visual images, as shown by Figure 4. The difference between the results for the graylevel fused and the colour fused images is not significant (p=0.134), suggesting that spatial localisation of targets (following detection) does not exploit colour contrast as long as there exists sufficient brightness contrast in the gray fused imagery.

5. DISCUSSION

This study investigates (a) for which conditions the fusion of visual and thermal images results in a single composite image with extended information content, and (b) whether a recently developed colour image fusion scheme [2–5,9] can enhance the situational awareness of observers operating under these specific conditions and using visual and thermal images.

Conditions in which fusion of visual and thermal imagery are most likely to result in images with increased information content occur around sunrise. At this time the contrast of both the visual and the thermal images is very low. One can construct other scenarios involving night operations in which both modalities are lacking in contrast. The visual contrast is low around sunrise because of the low luminance of the sky. However, contours of extended objects are still visible. After some image enhancement (like center-surround shunting, histogram equalization or contrast stretching) even an appreciable amount of detail can be perceived. Small objects with low reflectance, like a person wearing a dark suit or camouflage clothing, or objects that are partly obscured, are not represented in the visual image under these conditions, and can therefore not be detected. The thermal contrast is low around sunrise because most of the objects in the scene have about the same temperature after losing their excess heat by radiation during the night. As a result the contours of extended objects are not at all or incorrectly represented in the thermal image. The fusion of images registered around sunrise should therefore result in images that represent both the context (the outlines of extended objects) and the details with a large thermal contrast (like people) in a single composite image. To test this hypothesis a large set of image sequences is captured around sunrise on different days. The scenes used in this study represent 3 different scenarios that were developed by the Royal Dutch Army [20]. The images are fused using the recently developed MIT colour fusion scheme [2–5,9]. Graylevel fused images are also produced by taking the luminance component of the colour fused images. Visual inspection of the results shows that the fusion of thermal and visual images indeed results in composite images with an increased amount of information.

An observer experiment is performed to test if the increased amount of detail in the fused images can yield an improved observer performance in a task that requires a certain amount of situational awareness. The task that is devised involves the detection and localisation of a person in the displayed scene relative to some characteristic details that provide the spatial context. The person is optimally represented in the thermal imagery and the reference features are better represented in the visual imagery. The hypothesis is therefore that the fused images provide a better representation of the overall spatial structure of the depicted scene. To test this hypothesis subjects perform a relative spatial localisation task with a selection of thermal, visual, and (both graylevel- and colour-) fused images representing the above-mentioned military scenarios. The results show that observers can indeed determine the relative location of a person in a scene with a significantly higher accuracy when they perform with fused images, compared to the individual image modalities.

This study shows no significant difference between the localisation performance with colour fused images and with their luminance components (the derived graylevel fused images). However, in some conditions colour fused images are easier to visually segment than graylevel fused images. As a result, colour coding may greatly improve the speed and accuracy of information uptake [21], and fewer fixations may be required to locate colour coded targets [22]. Therefore, dynamic tasks like navigation and orienting, that probably depend on a quick and correct scene segmentation, may benefit from a colour fused image representation.

It is likely that the fusion of thermal and low-light level imagery may yield an even better observer performance over extended exposure times which often lead to exhaustion or distraction. Further research, preferably involving dynamic scenarios, is needed to test the hypothesis that colour image fusion schemes can boost observer performance in these tasks.

6. CONCLUSIONS

The fusion of thermal and visual images registered around sunrise results in composite images with an increased amount of detail that clearly represent all details in their correct spatial context.

Observers can localise a target in a scene (1) with a significantly higher accuracy, and (2) with a greater amount of confidence when they perform with fused images (either gray or colour fused), compared to the individual image modalities (visible and thermal).

Acknowledgement

TNO Human Factor Reseach Institute was sponsored in part by the Royal Netherlands Air Force and by the Royal Netherlands Army.
MIT Lincoln Laboratory was sponsored in part by the U.S. Office of Naval Research and the U.S. Air Force, under Air Force Contract F19628-95-C-0002.

REFERENCES

1. A. Toet and J. Walraven, New false colour mapping for image fusion, Optical Engineering, 35(3) (1996) 650–658.
2. A.M. Waxman, D.A. Fay, A.N. Gove, M. Seibert, J.P. Racamoto, J.E. Carrick and E.D. Savoye, Color night vision: fusion of intensified visible and thermal IR imagery, Proceedings of the SPIE Conference on Synthetic Vision for Vehicle Guidance and Control, vol. SPIE-2463 (1995) 58–68.
3. A.M. Waxman, A.N. Gove, M. Seibert, D.A. Fay, J.E. Carrick, J.P. Racamoto, E.D. Savoye, B.E. Burke, R.K. Reich, W.H. McGonagle, and D.M. Craig, Progress on color night vision: visible/IR fusion, perception & search, and low-light CCD imaging, Proceedings of the SPIE Conference on Enhanced and Synthetic Vision, vol. SPIE-2736 (1996) 96–107.
4. A.M. Waxman, A.N. Gove, D.A. Fay, J.P. Racamoto, J.E. Carrick, M. Seibert, E.D. Savoye, B.E. Burke, R.K. Reich, W.H. McGonagle, and D.M. Craig, Solid state color night vision: fusion of low-light visible and thermal IR imagery, Proceedings of the 1996 Meeting of the IRIS Specialty Group on Passive Sensors II, Ann Arbor: Infrared Information Analysis Center, ERIM, 1996, 263–280.
5. A.M. Waxman, J.E. Carrick, D.A. Fay, J.P. Racamato, M. Aguilar, and E.D. Savoye, Electronic imaging aids for night driving: low-light CCD, thermal IR, and color fused visible/IR, Proceedings of the SPIE Conference on Transportation Sensors and Controls, vol. SPIE-2902 (1996).
6. A.M. Waxman, A.N. Gove, D.A. Fay, J.P. Racamoto, J.E. Carrick, M. Seibert and E.D. Savoye, Color night vision: opponent processing in the fusion of visible and IR imagery, Neural Networks, 10(1) (1997) 1–6.
7. A.M. Waxman, J.E. Carrick, J.P. Racamato, D.A. Fay, M. Aguilar, and E.D. Savoye, Color night vision - 3rd update: Realtime fusion of low-light CCD visible and thermal IR imagery, Proceedings of the SPIE Conference on Enhanced and Synthetic Vision, vol. SPIE-3088 (1997).

8. P.J. Burt and R.J. Kolczynski, Enhanced image capture through fusion, Proceedings of the Fourth International Conference on Computer Vision, Washington, USA: IEEE Computer Society Press, 1993, 173–182.
9. A.N. Gove, Cunningham, R.K. and A.M. Waxman, Opponent-color visual processing applied to multispectral infrared imagery, Proceedings of 1996 Meeting of the IRIS Specialty Group on Passive Sensors II, Ann Arbor, US: Infrared Information Analysis Center, ERIM, 1996, 247–262.
10. H. Li, B.S. Manjunath and S.K. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models and Image Processing, 57 (1995) 235–245.
11. A. Toet, Hierarchical image fusion, Machine Vision and Applications, 3 (1990) 1–11.
12. A. Toet, L.J. van Ruyven and J.M. Valeton, Merging thermal and visual images by a contrast pyramid, Optical Engineering, 28 (1989) 789–792.
13. T.A. Wilson, S.K. Rogers, and L.R. Myers, Perceptual-based hyperspectral image fusion using multiresolution analysis, Optical Engineering, 34 (1995) 3154–3164.
14. P. Schiller, The ON and OFF channels of the visual system, Trends in Neuroscience, 15 (1992) 86–92.
15. P. Schiller and N.K. Logothesis, The color-opponent and broad-band channels of the primate visual system, Trends in Neuroscience, 13 (1990) 392–398.
16. P. Gouras, Color vision, in E.R. Kandel, J.H. Schwartz & T.M. Jessell (eds.), Principles of Neural Science, 3rd ed., Oxford, UK: Elsevier Science Publishers, 1991, 467–480.
17. E.A. Newman and P.H. Hartline, Integration of visual and infrared information in bimodal neurons of the rattlesnake optic tectum, Science, 213 (1981) 789–791.
18. E.A. Newman and P.H. Hartline, The infrared vision of snakes, Scientific American, 246 (1982) 116–127.
19. S. Grossberg, Neural networks and natural intelligence, Cambridge, MA: MIT Press, 1988.
20. A.R. Buimer, Scenarios for multi-spectral image fusion (in Dutch), Memo 21 December 1993, Section Planning, Training Center Infantry (OCI), Harderwijk, The Netherlands, 1993.
21. R.E. Christ, Review and analysis of colour coding research for visual displays, Human Factors, 17 (1975) 542–570.
22. P.K. Hughes, and D.J. Creed, Eye movement behaviour viewing colour-coded and monochrome avionic displays, Ergonomics, 37 (1994) 1871–1884.

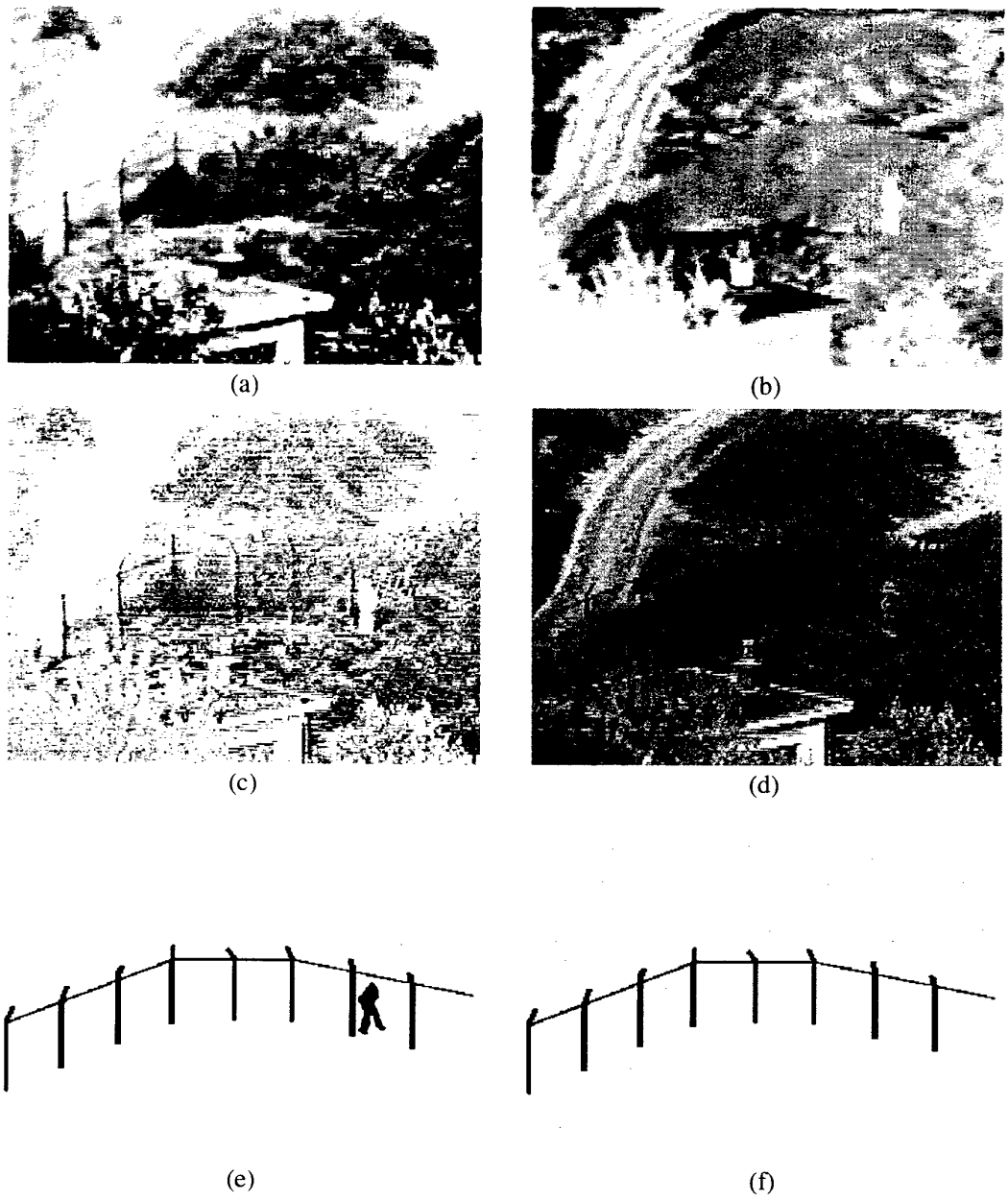


Figure 1. **(a)** Original CCD, **(b)** original IR, **(c)** graylevel fused, **(d)** colour fused, **(e)** baseline test, and **(f)** reference images of Scenario I.

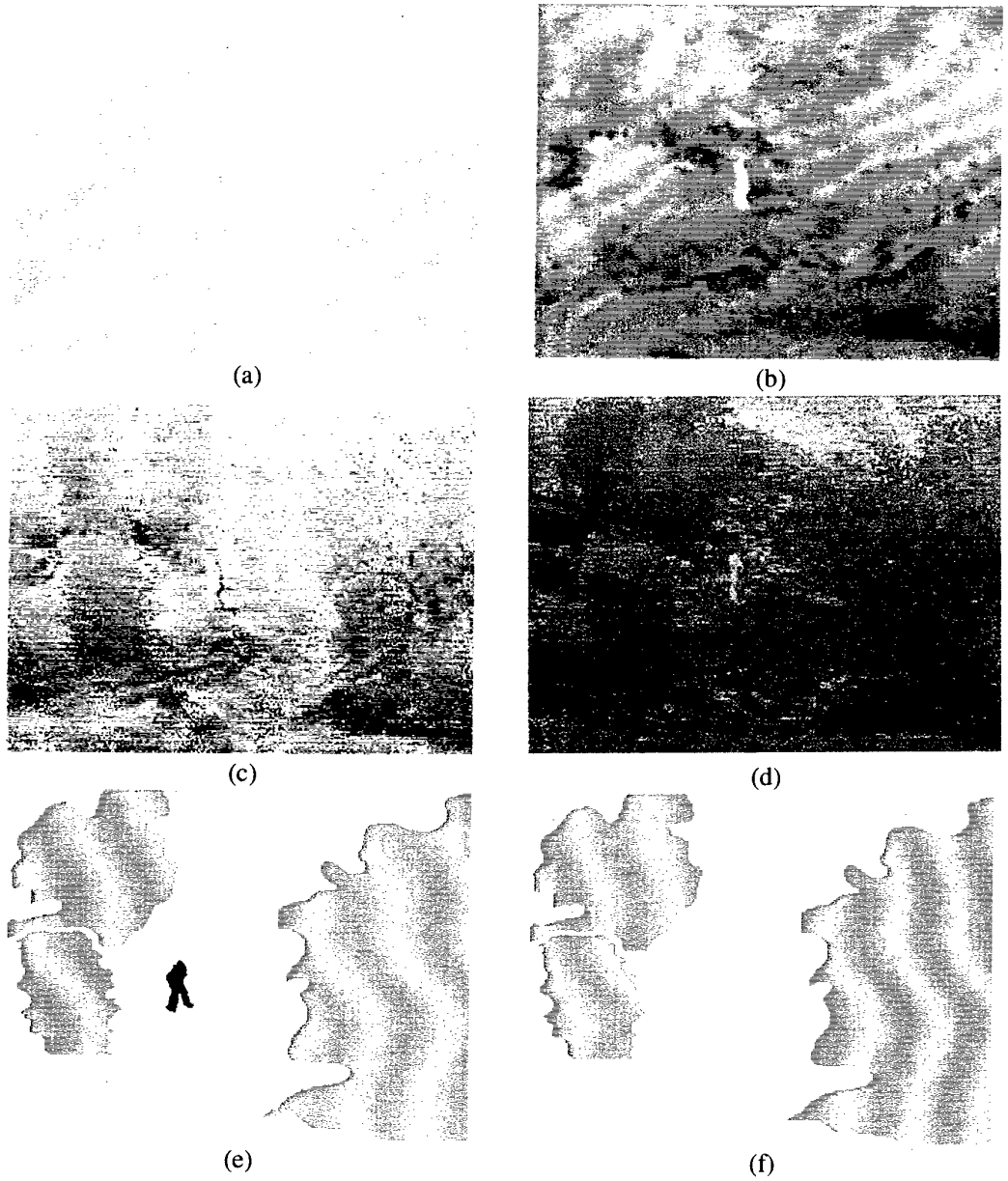


Figure 2. **(a)** Original CCD, **(b)** original IR, **(c)** graylevel fused, **(d)** colour fused, **(e)** baseline test, and **(f)** reference images of Scenario II.

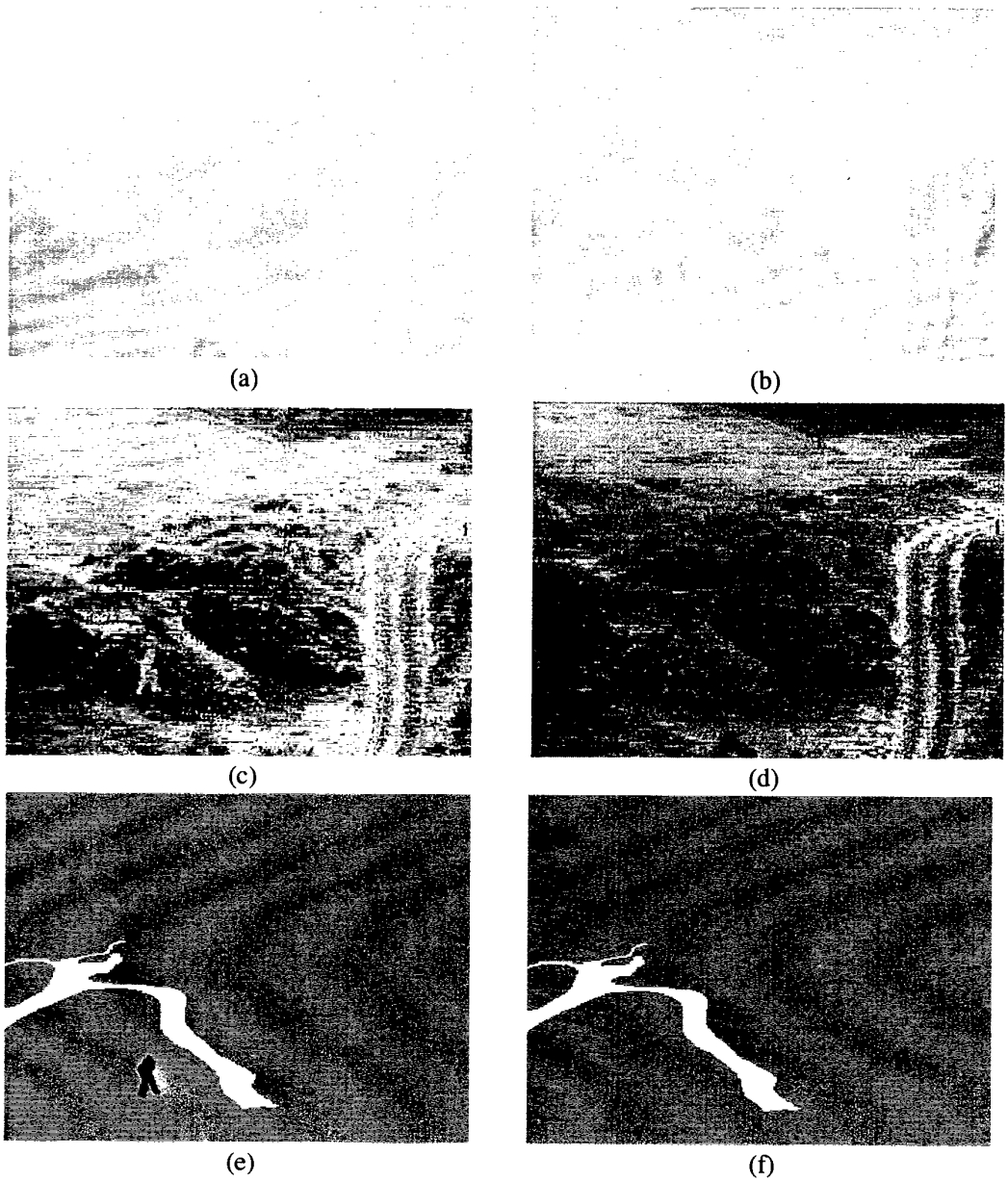


Figure 3. **(a)** Original CCD, **(b)** original IR, **(c)** graylevel fused, **(d)** colour fused, **(e)** baseline test, and **(f)** reference images of Scenario III.

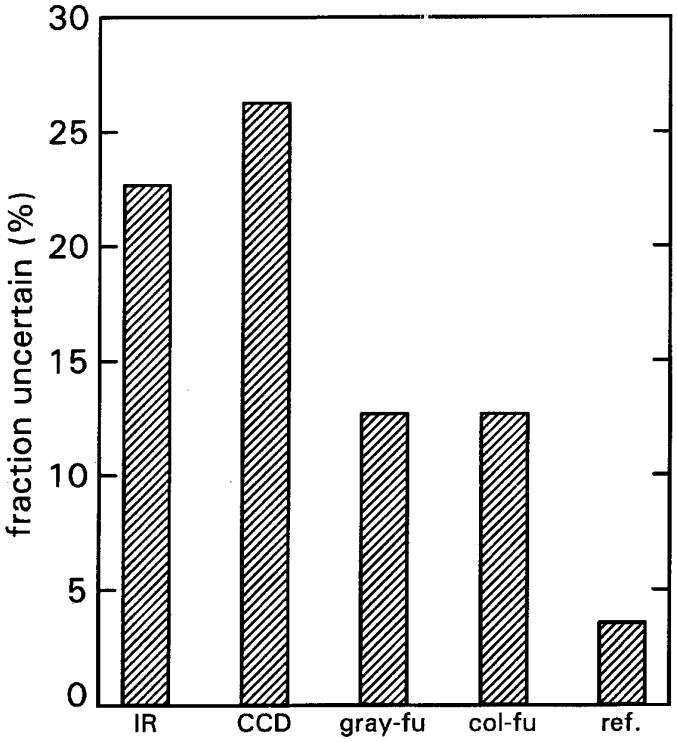


Figure 4. The percentage of image presentations in which observers are uncertain about the relative position of the person in the scene, for each of the 5 image modalities tested (IR, CCD, graylevel fused, colour fused, and schematical reference images).

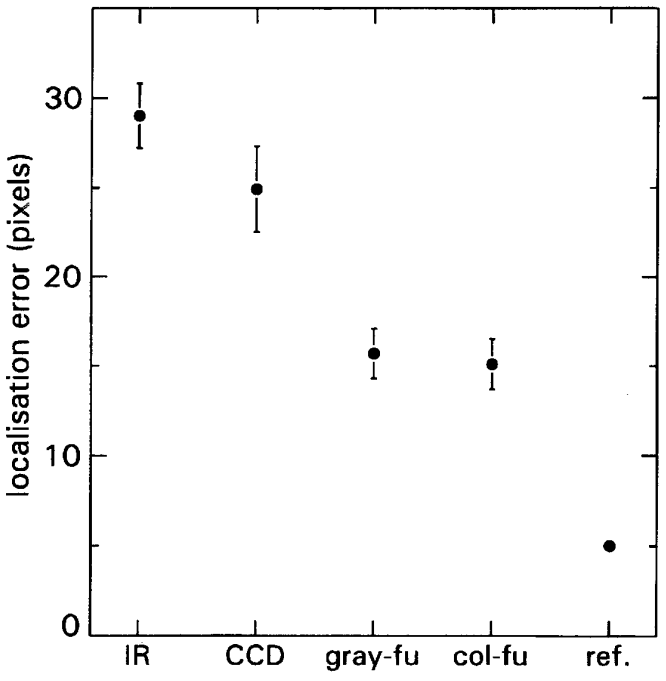


Figure 5. The mean weighted distance between the actual position of the person in each scene and the perceived position for each of the 5 image modalities tested (IR, CCD, graylevel fused, colour fused, and schematical reference images). The error bars indicate the size of the standard error in the perceived location.

PAPER No: 42

DISCUSSOR'S NAME: ?

COMMENT/QUESTION:

How are the images registered finely, and what is the robustness of the registration with disturbances?

If there is a decay of one or more pixels between images, does it produce completely wrong colours on the boundary of the objects?

AUTHOR/PRESENTER'S REPLY:

The CCD and IR images must of course be aligned before they can be fused. The signals of the thermal and visual cameras are therefore first spatially registered as closely as possible.

This is effected by (a) choosing a set of fiducial points that are visible both in the thermal and in the visual images and (b) ensuring that their positions coincide as much as possible at each location over the common field of view of both cameras.

Well-defined fiducial points are created by placing 3 large plastic jerrycans on the scene. The cans are bright (white) and filled with hot water so that they are clearly visible in both image modalities. They are evenly distributed over a horizontal row such that the outer two are just inside the field of view. The difference signal of the CCD and IR cameras is displayed on a CRT. Using both this difference image and the fiducial points both the zoom ratio (display size of the common field of view) and the optical axis of the CCD camera are adjusted to those of the IR camera. This tuning is iteratively performed with the fiducial points at different positions in the common field of view (for different orientations of the optical axis of the camera suite). When the optimal horizontal image registration has been achieved there is still a size difference of about 5% in the vertical direction. The reason for this is the fact that the pixels of the CCD camera represent rectangular image segments, whereas those of the IR camera correspond to a square field of view. A final correction for this misalignment is performed by geometrically transforming and resampling the digitized video frames. By registering the scene for 3 different vertical positions (top, middle, bottom) of the vitual line joining the 3 jerrycans (situated in the common field of view of both cameras) 9 matching points are obtained.

For the scenes used in this experiment an additional number of matching points is opportunistically available, because a row of street lamps is clearly visible from the viewing location. The field of view of the camera covers a highway section containing 6 lamps. By registering this highway scene for 3 different vertical positions of the row of lamps an additional number of 18 matching points is obtained. The scenes containing the matching sets of fiducial points are registered at the beginning of each recording session. Digitized frames from this recording are used to compute the coefficients of the affine warping transformation that maps corresponding points in the scene to corresponding pixel locations in the image plane. The digitized CCD images are then warped and resampled so that their pixels are in registration with the corresponding pixels of the IR image. The alignment of both cameras is not changed during a session. Therefore, a single set of match points can be used to register all scenes that are recorded during a session.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-595	ISBN 92-836-0051-7	UNCLASSIFIED/ UNLIMITED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title

Multi-Sensor Systems and Data Fusion for Telecommunications, Remote Sensing and Radar

7. Presented at/sponsored by

The Sensor and Propagation Panel Symposium, held in Lisbon, Portugal, 29 September - 2 October 1997.

8. Author(s)/Editor(s)	9. Date
Multiply	April 1998
10. Author's/Editor's Address	**11. Pages**
Multiple	416

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Multisensors
Weapon systems
Data fusion
Telecommunication
Remote Sensing
Radar
Electronic countermeasures
Deception
Target acquisition
Target recognition
Delivery
Accuracy
Command and control
Military intelligence
Surveillance
Battlefields
Signal processing

14. Abstract

This publication reports the unclassified papers presented at a specialists' meeting held by the Sensor and Propagation Panel at its Fall 1997 meeting.

The topics covered included:

— Applications of multiple sensors and data fusion

— Data fusion techniques and methods

— Sensor data networks and management techniques

— Validation studies, experiments, technologies.

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie 0(1)55.61.22.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

L'AGARD détient un stock limité de certaines de ses publications récentes. Celles-ci pourront éventuellement être obtenus sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous à l'AGARD par lettre ou par télécopie à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner.*

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres de diffusion nationaux indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordonnateur AGARD - VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Directeur - Gestion de l'information
(Recherche et développement) - DRDGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio Stralcio AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
Netherlands Delegation to AGARD
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076
Etats-Unis

The British Library Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax 0(1)55.61.22.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of some of its recent publications, and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. *Please do not telephone.*

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) in their distribution.

AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD - VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Director Research & Development
Information Management - DRDIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio Stralcio AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076
United States

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic Database or on CD-ROM)



Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-836-0051-7