NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(SET-118)TP/393

www.rto.nato.int

RTO TECHNICAL REPORT **TR-SET-118**

3D Modelling of Urban Terrain

(Modélisation 3D de milieu urbain)

Final Report of Task Group SET-118.

Published September 2011

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(SET-118)TP/393

www.rto.nato.int

RTO TECHNICAL REPORT **TR-SET-118**

3D Modelling of Urban Terrain

(Modélisation 3D de milieu urbain)

Final Report of Task Group SET-118.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

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

Published September 2011

Copyright © RTO/NATO 2011
All Rights Reserved

ISBN 978-92-837-0144-6

Single copies of this publication or of a part of it may be made for individual use only. The approval of the RTA Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

Table of Contents

	Page
List of Figures	**vi**
List of Tables	**ix**
Acknowledgements	**x**
Programme Committee	**xi**
Executive Summary and Synthèse	**ES-1**
Chapter 1 – Introduction	**1-1**
1.1 Background	1-1
1.2 Objectives	1-1
1.3 Summary of Activities	1-1
1.4 Common Misunderstandings	1-2
Chapter 2 – Military Applications	**2-1**
2.1 Mission Planning/Training	2-3
2.2 Change Detection	2-4
2.3 Sensor Simulation	2-6
2.4 Autonomous Navigation	2-7
Chapter 3 – Sensor Systems for Generating 3D Point Clouds	**3-1**
3.1 Passive Sensing	3-1
3.1.1 3D Point Clouds from 2D Images	3-1
3.2 Active Sensing	3-3
3.2.1 Time-of-Flight Systems	3-4
3.2.1.1 2-Dimensional (Flying Point) Scanning Systems (LIDAR)	3-5
3.2.1.2 1-Dimensional Scanning Systems	3-7
3.2.1.3 3D Staring Arrays (Flash LIDAR)	3-7
3.2.1.4 Through-Wall Synthetic Aperture RADAR	3-10
3.2.2 Estimating Distances through Waveform Analysis	3-11
3.2.2.1 Detecting a Pulse	3-11
3.2.2.2 Measuring Distance in Practice	3-13
3.3 Registration of 3D Point Clouds	3-14
3.4 Georeferencing of 3D Point Clouds or Models	3-15
Chapter 4 – From Point Clouds to 3D Models	**4-1**
4.1 Concepts and Requirements	4-2

4.2 Model Representations 4-7
4.2.1 Explicit Surface Representations 4-7
4.2.1.1 Octree Representation 4-8
4.2.1.2 Tetrahedral Occupancy Cells 4-9
4.2.1.3 Comparison of Oct-Cubes and Tetrahedral Bin-Trees 4-10
4.2.2 Implicit Surface Representations 4-11
4.2.3 Hybrid Methods 4-12
4.3 Classification of Terrain Objects 4-13
4.3.1 Bare-Earth Extraction 4-13
4.3.2 Classification of Above-Ground Objects 4-14
4.3.2.1 Buildings 4-14
4.3.2.2 Trees 4-15
4.3.3 Classifier Examples 4-16
4.3.4 Data and Information Fusion 4-16
4.3.5 Building Reconstruction from Airborne LIDAR Data 4-17
4.4 Sample Processed SET-118 Data Sets 4-18
4.4.1 WP-AFRL: PILAR (Polarimetric Imaging Laser RADAR) 4-18
4.4.2 WP-AFRL: Flash LIDAR Data 4-19
4.4.3 FOI's Norrköping Collection 4-21
4.4.4 UK – Defence Science and Technology Laboratory's Burst Illumination LIDAR 4-23
4.4.5 Processing of FOM – Urban SLAM Data Set 4-24
4.4.6 Defence Research and Development Canada, Valcartier LIDAR Data Collection 4-25
4.5 Model Instantiation 4-26
4.5.1 Semi-Automatic 3D Extraction of Roads 4-26
4.5.1.1 Road Boundary Detection 4-27
4.5.1.2 Centreline Detection 4-27
4.5.1.3 Road Modelling by the "Snake" Algorithm 4-29
4.5.1.4 Road Modelling by the "Ziplock Snake" Algorithm 4-30
4.5.1.5 Altitude Determination from Stereo-Couples 4-32
4.5.1.6 Results and Validation 4-33
4.5.2 Building Reconstruction from Multi-Aspect High-Resolution InSAR Data 4-36
4.5.2.1 State-of-the-Art 4-36
4.5.2.2 Signature of Buildings in High Resolution InSAR Data 4-37
4.5.2.3 Building Reconstruction Approach 4-39
4.5.3 Extraction of 3D Information of Individual Objects with Monoscopic Techniques 4-42
4.5.3.1 Estimation of Height of a Feature from the Shadow of the Feature 4-43
4.5.3.2 Features Height Estimation by Means of Photo-Interpretation 4-45
4.5.3.3 Pros and Cons 4-45

Chapter 5 – Assessment of 3D Models 5-1
5.1 Current Metrics for Quantitative Assessment 5-1
5.1.1 L_p Metrics 5-1
5.1.2 Hausdorff Metric 5-2
5.1.3 Quality Measures Based on Overlap of Volume 5-5
5.2 Future Human-Goal-Based Metrics for Quantitative Assessment 5-8

Chapter 6 – Conclusions and Recommendations **6-1**
6.1 Recommendations 6-1
6.2 A Path to the Future 6-1

Chapter 7 – RTG Participants and Meetings **7-1**
7.1 RTG Participants 7-1
7.2 Meetings 7-1

Chapter 8 – References **8-1**

List of Figures

Figure		Page
Figure 1-1	Continuous 2.5D, 2.75D and 3D Surfaces	1-2
Figure 2-1	Some Illustrations of Military Applications	2-3
Figure 2-2	A Soldier Training in a Semi-Urban Environment Using a CAVE System (DRDC)	2-4
Figure 2-3	Example of Change Detection in 3D LIDAR Data for IED Detection Purposes	2-5
Figure 2-4	Change Detection in Urban Setup	2-6
Figure 2-5	Simulated SAR Image and a View of the 3D Model Including Five Military Vehicles (FOI)	2-7
Figure 2-6	Snapshots from a Sequence of Simulated IR Images	2-7
Figure 3-1	Triangulation Principle Depicted Using Three Images Captured from Two Different Viewpoints	3-2
Figure 3-2	Three Representative Infrared Images Showing an Urban Area with Plotted Interest Points	3-3
Figure 3-3	Two Views of the Reconstructed Urban Scene	3-3
Figure 3-4	Farm House and Outbuildings at 18 km (Ground-Ground Demonstration)	3-6
Figure 3-5	Transmitted and Received Signals in an Amplitude Modulated LIDAR System	3-7
Figure 3-6	3D Flash Illumination Imaging System and Sample 3D Image of a Truck	3-8
Figure 3-7	UK 3D Laser Imaging Camera System and Sample Imagery of Land Rover	3-9
Figure 3-8	TWSAR Prototype	3-10
Figure 3-9	2D TWSAR Image of a Wood Building with Building Layout Outlined in Black	3-11
Figure 3-10	Histogram of the Width of Detected Pulses in an Area Containing Buildings and Vegetation	3-12
Figure 3-11	Deconvolution Example	3-13
Figure 3-12	A Number of FLASH LIDAR Data Sets Collected from a Car while Driving Down a Street in Bonnland, Registered with the ICP Algorithm	3-15
Figure 4-1	Examples of Urban Terrain	4-1
Figure 4-2	Requirement of Multi-Scale Localization	4-3
Figure 4-3	Domed Building Circled in Figure 4-2	4-3
Figure 4-4	Eglin LIDAR Flyover with Overlapping Swaths	4-4
Figure 4-5	Image Sequence of Adaptive Sub-Division Builds for On-Line Surface Reconstruction from Eglin LIDAR Overflight Data	4-4
Figure 4-6	Close Up of Parking Lot Light Pole from Eglin LIDAR Overflight Data; and Fine Scale Surface Reconstruction Using Level Sets is Used to Fit the Point Cloud Data and is Able to Resolve Additional Detail	4-5
Figure 4-7	Original LIDAR Data for Underpass; Close Up View from a Different Perspective than the Imaging Sensor, in particular, Looking Under the Underpass; Top View of Intersection; Reconstructed Surface of Underpass and Neighbouring Vegetation	4-6

	and Trees; and Alternate View Showing Overfitting of the Sparse Data Along the Roadway Borders	
Figure 4-8	Simulated Flash LIDAR Data of Patio Area – Ft. Benning; Coarse Learned Surface from Point Clouds Resulting from 10 Registered 128 x 128 Flash LIDAR Images; and Finer Resolution of Isosurface of Point Cloud of Table Constructed Using Unsigned Distance Field from Point Cloud in b)	4-6
Figure 4-9	Raw LIDAR Data Generated by Hokuyo Line Scanning Sensors Which Were Rotated at a Fixed Position about a Vertically Aligned Axis; and Surface Reconstruction Using a Level Set of an Unsigned Distance Captures Some Important Topology, Such as Building Faces and Columns, and Demonstrates Complexities Arising Both from the Sensed Surfaces as well as Processing Artefacts	4-7
Figure 4-10	Octree with the Corresponding Tree Representation	4-8
Figure 4-11	Cube Partitioned into Six Tetrahedral Simplices	4-9
Figure 4-12	Point Cloud of Wright-Patterson ASC Flash LIDAR Data of DRDC Valcartier Armoured Personnel Carrier Rendered by Range Colouring; Partitioned into Occupied and Force-Split Tetrahedra; and Rough Surface Representation Formed by Surfaces of Leaf Node Tetrahedra	4-10
Figure 4-13	Synthetic Point Cloud of a Simulated Urban Terrain with Moderately Complex Topology; and Level Set Reconstruction of Sensed Surface Using Variational Surface Fitting	4-12
Figure 4-14	Elevation Data (Digital Surface Model); and Ground Surface (Digital Terrain Model)	4-13
Figure 4-15	Elevation Image; and Black Pixels Indicate where Ground Hits have Occurred	4-15
Figure 4-16	Illustration of the Process of Creating a 3D Building Model from Airborne LIDAR Data	4-18
Figure 4-17	Gridding of PILAR Data Swath is Adaptively Refined in Order to Estimate Heights of Local Surface Patch to Fit the Point Cloud; and Surface Statistics are Progressively Updated On-Line as Data is Received	4-19
Figure 4-18	Rendering of Post-Assembled Bonnland Point Cloud, Time-Coded by Colour from Blue to Red, from a View Looking up the Roadway Haupstrasse	4-19
Figure 4-19	Referencing the LIDAR Point Cloud to Video from the Sensor Platform	4-20
Figure 4-20	Bonnland Data Features and Processing	4-21
Figure 4-21	Segmentation and Surface Reconstruction of Cropped Portions of Norrköping Data Set	4-22
Figure 4-22	Results from Automatic Object Classification for a Part of Norrköping	4-22
Figure 4-23	3D Building Reconstruction of a Cropped Portion of Norrköping Data	4-23
Figure 4-24	Iterated Mahalanobis Segmentation Applied to Burst Illumination LIDAR Data Provided by DSTL-UK NATO SET-118 Collaboration	4-24
Figure 4-25	Segmentation and Surface Reconstruction of FOM SLAM Data	4-25
Figure 4-26	Iterated Mahalanobis Segmentation of Scanning LIDAR Data Provided by DRDC Valcartier NATO SET-118 Collaboration	4-26
Figure 4-27	Boundary Detection – the Roads to Extract and the Detected Boundaries	4-27
Figure 4-28	Output of the “Steger Line Detector” Algorithm	4-28
Figure 4-29	Road Centreline Calculated by the Path Finding Algorithm	4-29
Figure 4-30	Snake Initialization by Centreline and Boundary Widths	4-30

Figure 4-31	Ziplock Snake Elements	4-31
Figure 4-32	Ziplock Snake	4-32
Figure 4-33	Cross-Correlation Example in a Stereo-Pair	4-33
Figure 4-34	Reference Model #1 and Automatically Extracted Road #1	4-34
Figure 4-35	Reference Model #2 and Automatically Extracted Road #2	4-34
Figure 4-36	Reference Model #3 and Automatically Extracted Road #3	4-35
Figure 4-37	Stability Analysis for the xy-Plane and the z-Axis	4-35
Figure 4-38	Appearance of Flat- and Gable-Roofed Buildings Under Orthogonal Illumination Conditions	4-38
Figure 4-39	Workflow of Algorithm	4-40
Figure 4-40	Results of Mainly Flat-Roofed Building Detection; Results of Gable-Roofed Building Detection; Oversized Building Hypothesis; Corrected Building Hypothesis; Optical Data of Hip-Roofed Building; Detection Result of AeS-1 Data; and TerraSAR-X Data	4-42
Figure 4-41	Height Calculation from the Shadow; and Exemplary Case of Application Under Ideal Conditions, i.e., Measurable Shadow, of an Airport Control Tower	4-43
Figure 4-42	Satellite Image with No Visible Shadows	4-44
Figure 4-43	Exemplary Situation Where the Shape from Shading Method Cannot be Applied	4-44
Figure 4-44	Building Example Where the Height is Extracted from Image Knowledge	4-45
Figure 5-1	Flat Surface	5-2
Figure 5-2	Flat Surface with Long Thin Ridge	5-2
Figure 5-3	ε Distance Measurements between Black and Green Surfaces in the L_∞, L_2, and Hausdorff Metrics, Respectively	5-3
Figure 5-4	Baltimore Height Data Used to Estimate Distortions of Viewshed Using Different Metrics	5-4
Figure 5-5	Correct and Erroneous Portions of Viewsheds for Convoy Route in Baltimore	5-5
Figure 5-6	Building Models Obtained by Two Different Acquisition Methods	5-7
Figure 5-7	Detail of the Screened Test Data Set T and Difference Set T\R	5-7
Figure 5-8	Diagnostics by Visualisation of the Local Quality Rates, Computed in 3D	5-8

List of Tables

Table		Page
Table 1-1	Discussion and Assessment of Various Scene Representations	1-3
Table 2-1	Military Applications of Urban 3D Models	2-1
Table 3-1	Provides an Overview of Commercial Airborne LIDAR Systems and Achievable System Performance	3-5
Table 3-2	Comparison of Performance Parameters for Amplitude-Modulated Camera Systems	3-10
Table 4-1	Detailed Analysis of Erroneously Classified Objects, as Compared to Ground Truth	4-16
Table 4-2	The Table Shows the Results of Information (Decision-Level) Fusion of Individual Results Through Majority Voting	4-17
Table 4-3	Horizontal and Vertical rms Differences for Three Test Cases	4-35
Table 5-1	Quality Measures (Globally Calculated) for the Two Models of Figure 5-6	5-7

Acknowledgements

The group would like to thank all those who contributed to the work documented in this report. Each country had a number of technical, scientific and other staff without whose effort this work could not have been accomplished. In particular, the group would like to thank Antje Thiele from Fraunhofer-IOSB, Germany, for her comprehensive contribution concerning the SAR technology, Matthew Hielsberg and Borislav Karaivanov from IMI, University of South Carolina, for their contributions in analyzing and processing various point cloud data collections provided by SET-118 members. Further, we would like to thank David Holland from Ordnance Survey, United Kingdom, and Andrew Kurdila from Virginia Tech, USA, for hosting meetings and making them successful. Finally, the authors would like to thank Richard Richmond, Jack Woods and Matthew Gebhardt from the Air Force Research Laboratory, Dayton, OH for providing us with the technical description and data from the flash LIDAR system.

Programme Committee

CANADA

Mr. Jonathan FOURNIER
DRDC Valcartier, Defence R&D Canada
Valcartier Tactical Surveillance and Reconnaissance Section
2459 Pie-XI Blvd North
Quebec City, QC G3J 1X5
Email: jonathan.fournier@drdc-rddc.gc.ca

Mr. Benoit RICARD
DRDC Valcartier, Defence R&D Canada
2459 Pie-XI Blvd North
Quebec City, QC G3J 1X5
Email: benoit.ricard@drdc-rddc.gc.ca

GERMANY

Dr. Jochen MEIDOW (Chair)
Fraunhofer IOSB SZA Scene Analysis
Gutleuthausstrasse 1
D-76275 Ettlingen
Email: jochen.meidow@iosb.fraunhofer.de

Mr. Ulrich THÖNNESSEN
Fraunhofer IOSB SZA Scene Analysis
Gutleuthausstrasse 1
D-76275 Ettlingen
Email: thoennessen@fom.fgan.de

ITALY

Ing. Stefano PASQUARIELLO
Strategy & Product Planning
Research & Development
Selex Sistemi Integrati S.p.A. (Finmeccanica Group)
Via Laurentina, 760
I-00143 Rome
Email: spasquariello@selex-si.com

SPAIN

Mr. Daniel Emilio CARRASCO DIAZ
Indra Espacio S.A.
Dpto. De Teledeteccion
C/Mar Egeo 4
Poligono Industrial No. 1
28830 San Fernando de Henares (Madrid)
Email: dcarrasco@indra.es

SWEDEN

Dr. Lars Gustav TOLT
FOI Swedish Defence Research Agency
Olaus Magnus Vag 42
P.O. Box 1165
SE-58111 Linköping
Email: gustav.tolt@foi.se

UNITED KINGDOM

Dr. Kenneth McEWAN
Defence Science and Technology Laboratory
DSTL Porton Down
Bldg 5, Rm 101, Desk 362
Sensors & Countermeasures Dept.
Salisbury, Wiltshire SP4 0JQ
Email: kjmcewan@dstl.gov.uk

UNITED STATES

Mr. John Edward LAVERY
US Army Research Office Mathematics Division
US Army Research Laboratory
P.O. Box 12211
Research Triangle Park
4300 South Miami Blvd
Durham, NC 27709-2211
Email: john.lavery2@us.army.mil

Professor Robert C. SHARPLEY
University of South Carolina USC
Department of Mathematics
Industrial Mathematics Institute
1523 Greene Street
Columbia, SC 29208
Email: sharpley@math.sc.edu

3D Modelling of Urban Terrain

(RTO-TR-SET-118)

Executive Summary

Modelling the geometric and physical properties of 3D urban terrain represents a significant opportunity to enhance the generation of the Common Relevant Operational Picture (CROP). Such modelling can support visualization, which in turn enhances the user's situational awareness in complex urban scenarios. Of equal and increasing significance are inputs to non-visualization tasks such as line-of-sight determination, mission planning, change detection, sensor network capability assessment, threat analysis and the calculation of acoustic, chemical and EM propagation.

The Research Task Group (RTG) "3D Modelling of Urban Terrain" was formed to study various representation forms for objects in urban terrain, to inspect procedures for automatic scene reconstructions by exploiting the data of modern sensors, and to discuss potential assessment metrics and criteria. By doing so, the group identified and clarified several conceptual misunderstandings which result from the colloquial use of scientific terms.

For the discussion of the relevance of 3D models, military applications have been itemized and discussed which benefit from the existence of 3D information. 3D point clouds have been identified as the most important intermediate result or representation. Therefore, a review of different sensors and sensing techniques that can be used to collect 3D point clouds is provided. The technologies under investigation cover both active and passive sensors, including flash laser, video, and Interferometric SAR (InSAR). Methods for geo-referencing and co-registration of different point clouds or models are presented.

For the modelling and the automatic 3D model instantiation, i.e. the scene reconstruction, the concepts and requirements are discussed along with the corresponding representation forms. For military applications this implies the capability of timely processing large data set, to update and augment existing models, and to provide multi-scale representations. During its three year activity, the group compiled and pooled data sets for benchmarking and investigations on different data sources. Workflows have been established applying state-of-the-art algorithms, e.g. starting with pre-processing by waveform analysis, surface reconstruction by the assimilation of sensed points, and scene interpretation by classification.

Last but not least, the group collected, developed and studied various metrics and evaluation criteria for the specification of the accuracy of scene representation at hand. Human-goal-based metrics were developed for potential quantitative assessment in the future.

Modélisation 3D de milieu urbain

(RTO-TR-SET-118)

Synthèse

La modélisation des propriétés géométriques et physiques de milieu urbain représente une opportunité significative d'améliorer la génération de la présentation de la situation opérationnelle valide (CROP). Les modèles générés peuvent permettre la visualisation et servir à améliorer la connaissance de la situation par l'utilisateur dans des scénarios urbains complexes. Ces modèles peuvent également être utilisés pour supporter des tâches n'impliquant pas de visualisation telles que le calcul de ligne de visée, la planification de mission, la détection de changements, l'évaluation des performances de réseaux de capteurs, l'analyse de la menace et le calcul de la propagation acoustique, chimique et électromagnétique.

Le groupe de travail de la RTO (RTG) « Modélisation 3D de milieu urbain » a été mis sur pied afin d'étudier différentes façons de représenter les objets en milieu urbain, d'évaluer les méthodes utilisant des données provenant de capteurs modernes pour la reconstruction automatique de scène et de discuter des critères et mesures d'évaluation potentiels. De ce fait, le groupe a identifié et clarifié plusieurs conceptions erronées qui sont dues à l'utilisation familière de termes scientifiques.

La nécessité des modèles 3D a été abordée en énumérant et en décrivant les applications militaires qui bénéficient de l'existence d'information 3D. Les nuages de points 3D ont été identifiés comme étant le résultat intermédiaire (représentation) le plus important. Ainsi, une revue des différents types de capteurs et techniques qui peuvent être utilisés pour faire l'acquisition de nuages de points est présentée. Les technologies à l'étude comprennent les capteurs actifs et passifs, incluant le laser flash, la vidéo et le radar interférométrique à ouverture synthétique (InSAR). Des méthodes permettant de géoréférencer et de mettre en registre différents nuages de points ou modèles sont aussi présentées.

Les concepts et conditions requises ainsi que les types de représentation inhérents à la modélisation et à la génération automatique de modèles 3D (reconstruction de scène) sont décrits. Pour les applications militaires, cela inclut la capacité de traiter rapidement des ensembles de données volumineux, la mise à jour et l'ajout d'informations aux modèles existants et la génération de représentations multi-échelle. Durant ses trois ans d'activité, le groupe a compilé et rassemblé des ensembles de données pour être en mesure de conduire des tests de performance et d'effectuer des expérimentations en utilisant différentes sources de données. Des procédures de traitement utilisant des algorithmes de pointe ont été établies, par exemple en commençant par le pré-traitement par analyse de forme d'onde, la reconstruction de surface par assimilation des points acquis et l'interprétation de scène par classification.

Enfin, le groupe a rassemblé, développé et étudié différents critères et mesures d'évaluation pour l'identification de la précision des représentations de scène. Des mesures basées sur les besoins-utilisateur ont été développées pour permettre d'effectuer une potentielle évaluation quantitative dans le futur.

Chapter 1 – INTRODUCTION

1.1 BACKGROUND

Modelling the geometric and physical properties of 3D urban terrain represents a significant opportunity to enhance the generation of the Common Relevant Operational Picture (CROP). Such modelling can support visualization, which in turn enhances the user's situational awareness in complex urban scenarios. Of equal and increasing significance are inputs to non-visualization tasks as line of sight, mission planning, change detection, sensor network capability assessment, threat analysis and the calculation of acoustic, chemical and EM propagation. In the frame of this effort, 3D models include geospatial, physical and contextual information in addition to geometric models.

An increasing need for 3D urban models for use in situational awareness, mission planning is suggested by military operations, peace-keeping, humanitarian operations and homeland security embedded in urban areas. The increasing prevalence and performance of ground-based, aerial and space-based sensor systems and networks used for tactical reconnaissance and surveillance of urban areas are now capable of providing data for generation of the 3D models.

1.2 OBJECTIVES

The group focused on the exploitation of data representing urban areas and the generation of the corresponding 3D model instances. In general the objectives of the SET-118 Task Group were the modelling of urban 3D structures, the automatic reconstruction of scenes, and the assessment of results. Research areas were:

- Data exploitation techniques;
- Data preparation and fusion;
- Extraction of 2D and 3D geospatial data (information) for urban terrain; and
- Automatic 3D model reconstruction.

The technologies under investigation covered both active and passive sensors, including LIDAR, video, and Interferometric Synthetic Aperture RADAR (InSAR).

1.3 SUMMARY OF ACTIVITIES

One of the main interests of the Task Group was the regular exchange of information on national research and development activities. In the following a brief overview about the most important activities of group during the period of 2007 through its end in 2010 is provided:

- In September 2007 German and U.S. members of the group participated in a field experiment in the training site "Bonnland" in Germany. The main purpose of the test was the acquisition and provision of point clouds captured by a modern flash laser system mounted on a moving vehicle. Great efforts were made to understand this new technology, to study the recorded waveforms, and to co-registered the individual points clouds by appropriate algorithms.
- Concerning the system technologies under consideration several workflows were established with data from passive and active sensors and the corresponding up-to-date evaluation algorithms. Data captured by modern flash laser (Section 3.2.1) was successfully co-registered by the ICP algorithm explained in Section 3.3. Results are shown in Section 3.3. Airborne captured infrared images were used to determine the sensor's path and to reconstruct the scene simultaneously

(Section 3.1). Based on the resulting point cloud a surface reconstruction was performed by computing the isosurface of a signed distance field (Section 4.2.4). The processing was done offline, buts most of the applied algorithms possess real-time capabilities.

- For the specification of the accuracy of scene reconstructions at hand the group collected, developed and studied various metrics and evaluation criteria (Chapter 5).
- For benchmarking and the investigation on different data sources a common data pool was built up. But because of the diversity of the sensors data and the need to implement corresponding specialized algorithms, this data pool has not been embraced.

1.4 COMMON MISUNDERSTANDINGS

The group identified several conceptual misunderstandings. Two of them are discussed and highlighted here:

1) Model vs. Model Instance

A model is a description of our environment or of observed phenomena as simple as possible containing representations appropriate for specific applications, parameters and rules how to interpret theses representations. Specifying the parameter values for given models leads to the task of model instantiation, which, in the context of urban terrain, is the reconstruction of the scene. Colloquially, these terms are used interchangeably and one has to infer from the context what is meant. In Chapter 5, model instances are evaluated but not the models that were used.

2) Dimensionality of Representation

Many approaches and products bear the label “3D” but are actually not three-dimensional from the scientific point of view. Terrain models which represent the surface by a single-valued height as a function of latitude and longitude have the dimension 2.5D and appear like rubber blankets (cf. Figure 4-14 in Section 4.3.1) – for each point in the XY-plane exactly one height value Z can be specified. Geo-information systems often store the third dimension, e.g., building heights, as attributes which lead to 2.75D representations permitting the representation of vertical walls. Real 3D representations permit representation of tunnels, passages and roof overhangs also. Figure 1-1 illustrates the situation in cross-sections and Table 1-1 compiles pros and cons for different representations. In this table, “+” and “++” denote positive and very positive, respectively, and “–” and “– –” denote negative and very negative.

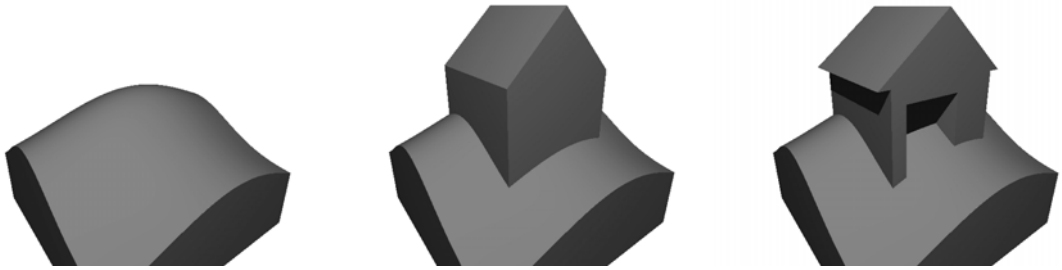


Figure 1-1: Continuous 2.5D, 2.75D and 3D Surfaces.

Table 1-1: Discussion and Assessment of Various Scene Representations.

	Point Cloud	**2.5D Surface Representation (e.g., a depth image)**	**3D Model**
Amount of Data	very large (– –)	large (o)	moderate (+)
Level of Abstraction	low (– –)	moderate (o)	high (+)
Procedure (Level of Readiness)	–	automatic (+)	automated (o)
Time Required for Instantiation	little (+)	moderate (o)	large (–)
Model Knowledge	no (–)	generic, often implicit	explicit
Semantics, Object Description	no (–)	no (–)	yes (++)

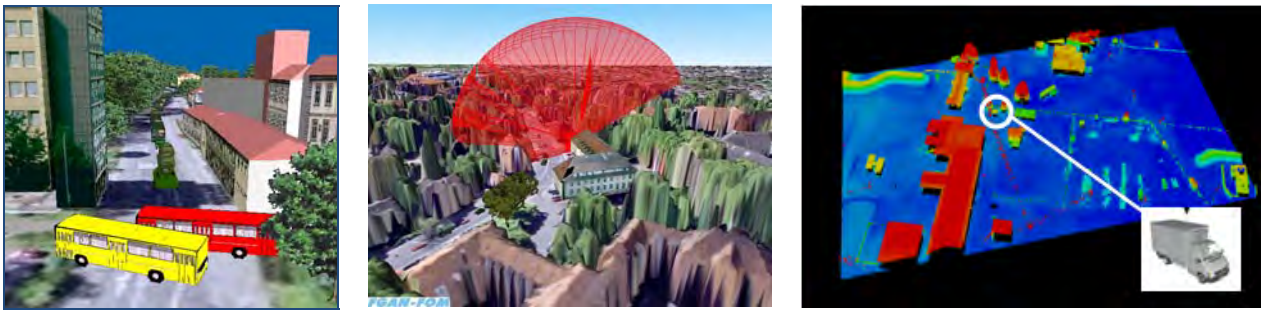
Chapter 2 – MILITARY APPLICATIONS

Urban 3D models are currently exploited to conduct a small number of military tasks such as autonomous vehicle navigation and mission training. By increasing the availability of urban 3D models, it is believed that a larger number of military tasks will benefit from the great range of information provided in this type of representation. In many situations, military personnel presently rely on two dimensional geographic maps and hand drawn sketches to perform their duty. Having access to accurate and complete urban 3D models would enhance their ability to conduct operations in a more efficient manner. The aim of this section is to present current and potential military applications of 3D models of urban terrain. Table 2-1 gives a list of military applications of urban 3D models and the following sub-sections describe some of these applications in greater details.

Table 2-1: Military Applications of Urban 3D Models.

Military Application	**Definition/Example**
Situational Awareness	A 3D model is generated and used to get a better knowledge of a given location.
Line of Sight (Visibility)	Model is used to evaluate the sight line of possible enemy posts.
IED Damage Simulation	3D model is exploited to simulate what would be the damage done by an IED explosion.
Change Detection	Compare models generated for a given location at two different times to highlight differences.
Battlefield Area Evaluation	Perform an evaluation of an area where military operations are to be conducted.
Battlespace Management	Use a model to manage ongoing military operations in urban terrain.
Briefing	Use model as a support to brief personnel for a given mission.
Collateral Damage Assessment	Generate a model of an area that has been under attack to assess the damage done to civilian and Army infrastructures.
Cooperative Reconnaissance	Use cooperative vehicles (e.g., UAVs) to perform reconnaissance of a given area and store the information acquired by each vehicle in a common 3D model.
Mission Rehearsal	Rehearse a mission using a 3D model of the area where the deployment will take place.
Mission Planning	Perform mission planning using a 3D model of the target location.
System Assessment	Use of accurate terrain models in assessment of optical and other communication systems.
Sensors Simulation	Simulate the behaviour of sensors for different types of targets and materials under varying conditions.

Mission Simulation	Use 3D models to perform simulation of different scenarios for a given mission.
Target Selection	Use a 3D model to precisely select the target of interest.
Battle Damage Assessment	Using a 3D model generated after an attack, assess the damages made to a specific target/area.
Terrain Analysis	Use 3D models to evaluate the density of vegetation of a specific area, determine the position of possible threats and/or determine line of sight regions for emplacement of assets.
Terrain Familiarization	Situational awareness / Mission rehearsal. Urban 3D models can be provided to soldiers to enable them to examine a given area from different points of view prior to actually being deployed in this area.
Threat Assessment	Capture in a 3D model possible threats (e.g., person, device, vehicles).
Wargaming	A 3D model is used to elaborate/test/refine strategies and theories of warfare.
Training	Urban 3D models can be used to create virtual environments where soldiers can train for different types of operations.
Automatic Target Detection (ATD)	Given a 3D model of an urban area, ATD algorithms aim at detecting the location of certain types of targets.
Automatic Target Recognition (ATR)	From a list of targets location output by ATD algorithms, ATR algorithms output possible target classification with corresponding probability.
System Assessment	Use of accurate terrain models in assessment of optical and other communication systems.
Calculation of Acoustic, Chemical and Electro-Magnetic Propagation	Simulations can be performed for various types of signals to assess how they would propagate in different types of built-up areas.
Terrain Reference for Autonomous Navigation and Path Planning	Autonomous vehicles have to be able to build a 3D representation of their environment in order to perform navigation and path planning tasks in an efficient manner.



a) b) c)

Figure 2-1: Some Illustrations of Military Applications – a) Mission Planning/Simulation; b) Threat Assessment; and c) Automatic Target Detection and Recognition.

2.1 MISSION PLANNING/TRAINING

Urban mission planning requires taking into consideration several factors that may have a positive or negative influence on the mission. It is believed that providing mission planners with 3D models of urban terrain could improve their understanding of these factors and facilitate their work. Urban 3D models may be exploited at the planning, briefing and training stages.

3D models used in the context of mission planning and training may include terrain surfaces, roads, vegetation as well as information on buildings such as doors, windows, building materials, roof types and interior layout. When additional intelligence is available, 3D models may also be augmented with potential enemy locations, barriers location, objectives, sight lines and intelligence estimates [1].

Depending on the type of mission, urban 3D models may be used at the planning stage to identify possible enemy observation posts and identify line of sight for these posts, establish a strategy to clear a set of buildings, select the best approach routes to reach a given objective, etc.

Once the mission planning phase is complete, 3D models can be used to brief personnel. At first, commanders may use the 3D models as a support for explaining the mission to their troop. Depending on the time available, the 3D models may then be made available to the soldiers so they can explore the area of interest from multiple angles and with different zoom levels. Being able to perform fly-through and walk-through in an urban environment prior to mission execution would result in greatly enhanced situational awareness.

Research is ongoing to provide soldiers with simulation tools to perform mission training and operation rehearsal using synthetic environments built from 3D models of urban terrains. These tools can be used to train soldiers for a specific task (e.g., clearing a building) or to simulate a whole mission. Simulations are usually performed on a desktop computer using a 2D or 3D display. Immersive virtual reality environments such as a Computer Assisted Virtual Environment (CAVE) can also be used for this purpose. These simulation environments enable soldiers to evolve in a simulated urban area with a real sense of 3D. Figure 2-2 shows a soldier during a training session in a semi-urban environment using a CAVE system. Mission training and operation rehearsal using 3D models of urban terrain enables soldiers to develop abilities and gain situational awareness in an innovative and efficient way.

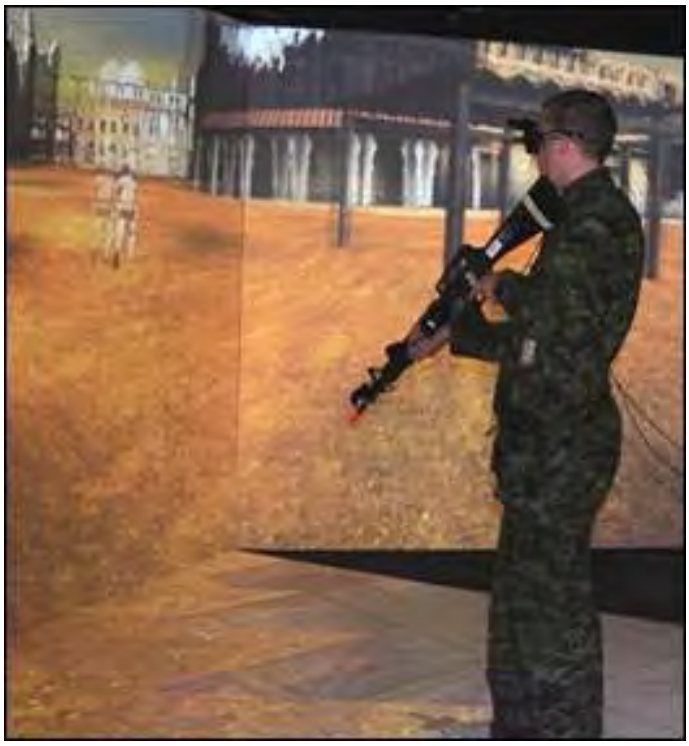


Figure 2-2: A Soldier Training in a Semi-Urban Environment Using a CAVE System (DRDC).

2.2 CHANGE DETECTION

Change detection consists in comparing 3D models of a given geographic area acquired at different points in time and highlighting the differences between the models. Depending on the resolution of the models, different types of change can be detected. Lower resolution models acquired using satellite data and/or high altitude airborne data may allow detection of construction and destruction of buildings and changes in terrain surfaces (newly constructed roads, new holes, collapsed terrain). Higher resolution models generated using low altitude airborne data and ground-based data may allow detection of vehicles, boats, man-made objects and IEDs. Change detection can be used to perform tasks such as battle damage assessment and urban development monitoring.

In Figure 2-3, an example is shown from a field trial conducted by FOI as part of a counter-IED project in 2009. In this trial, a section of a road (about 1.5 km long) was scanned twice with a mobile laser scanner system and a couple of changes (dummy IED objects) were introduced between the two measurements. By analyzing the spatial characteristics, geometrical changes corresponding to the IED objects could be identified. A main problem however, is that geometrical changes occur due to natural variations, such as non-identical sampling patterns, leaves moving in the wind, different viewing angles, etc. This indicates that post-processing, complementary sensor data and/or a priori information/assumption about the expected placement of the IEDs are probably needed in order to reduce the false alarm rates to acceptable levels.

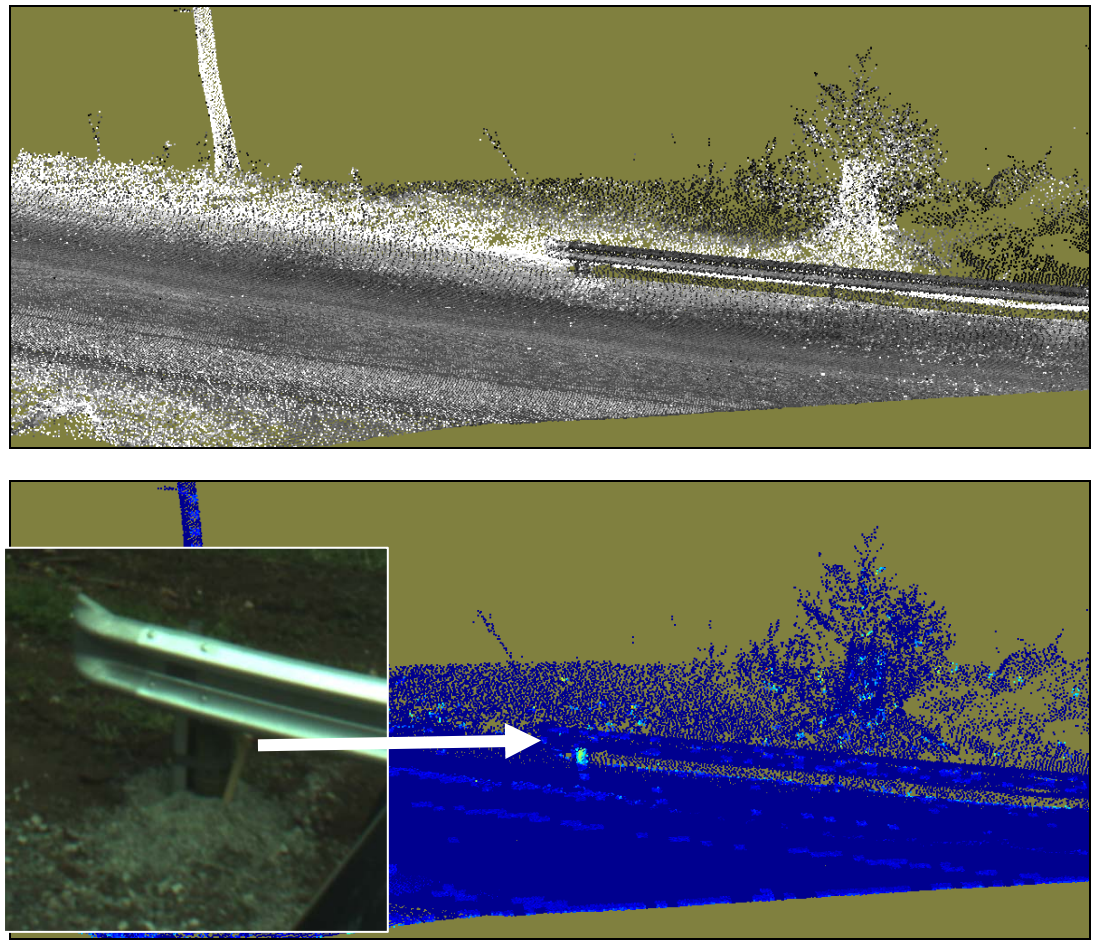
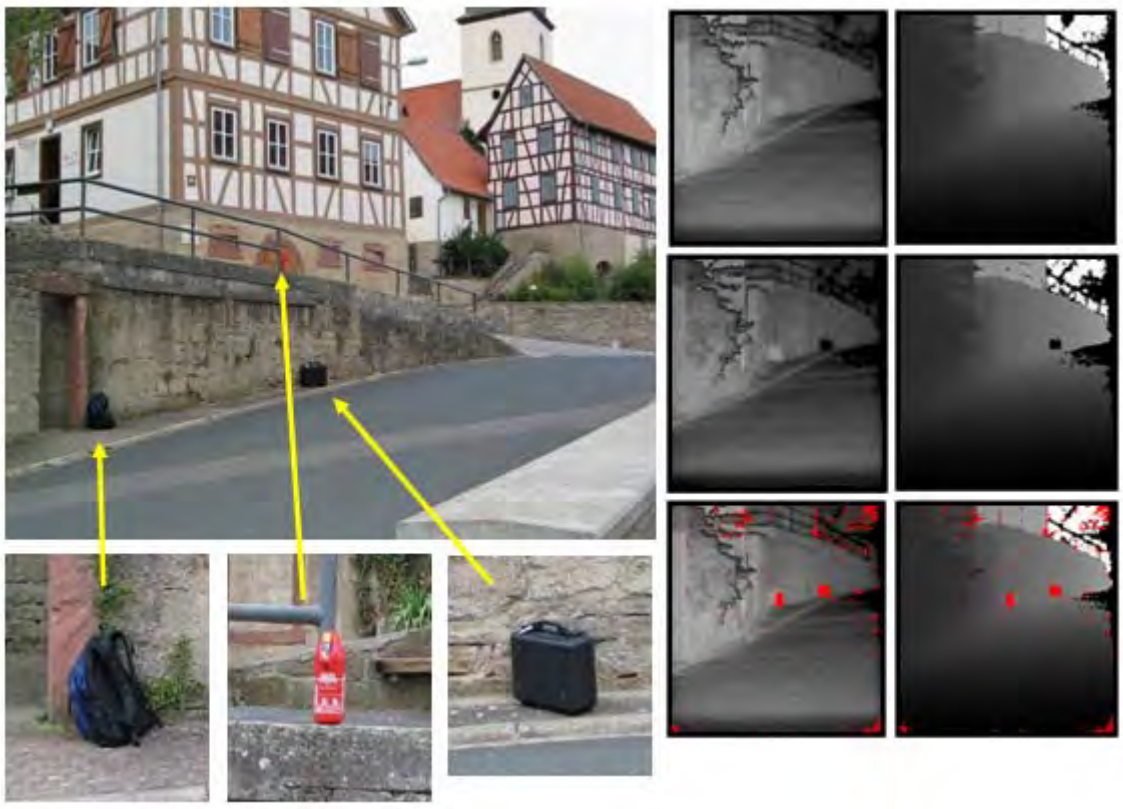


Figure 2-3: Example of Change Detection in 3D LIDAR Data for IED Detection Purposes. Top: A snapshot of part of the point cloud (greyscale level corresponds to laser intensity). Bottom: Result after nearest neighbour analysis, where the most recently collected point cloud is compared with a reference point cloud. The brighter pixels (cyan colour) indicate the presence of the (dummy) IED. Also note the changes detected in the vegetation next to the road.

Another example is shown in Figure 2-4 where an urban scene was imaged at two different times using a flash LADAR system with potential IED objects placed in the scene prior to performing the second acquisition. The objects of interest placed in the scene included a fire extinguisher, a briefcase and a backpack. A change detection process based on comparison of laser range and intensity data was applied to the scene in order to highlight the differences due to the addition of the objects in the scene. As can be seen in Figure 2-4, the objects positions are correctly highlighted. However, as discussed in the previous example, other factors such as moving leaves influence the final result. Post-processing and filtering techniques are therefore required to effectively identify in the scene the differences that actually correspond to the introduced objects from the differences that are caused by natural changes in the environment.



a) b)

Figure 2-4: Change Detection in Urban Setup – a) Different objects were placed in an urban scene; and b) By comparing the intensity data (left) and the range data (right) before and after adding the objects to the scene, changes can be highlighted (shown in red).

2.3 SENSOR SIMULATION

Given a 3D model of the terrain, synthetic sensor data can be produced through physics-based simulation. Such data can be used to predict or assess the performance of a certain sensor in various conditions in order to estimate the effective operational range of the sensor, predict the signature of potential targets, supply automatic data analysis tools (e.g., ATD/ATR) with synthetic training data, etc. Normally, in order to enable accurate simulation, the geometry of the 3D model has to be complemented with additional attributes such as material properties. Figure 2-5 and Figure 2-6 show some examples of simulated sensor data from FOI.

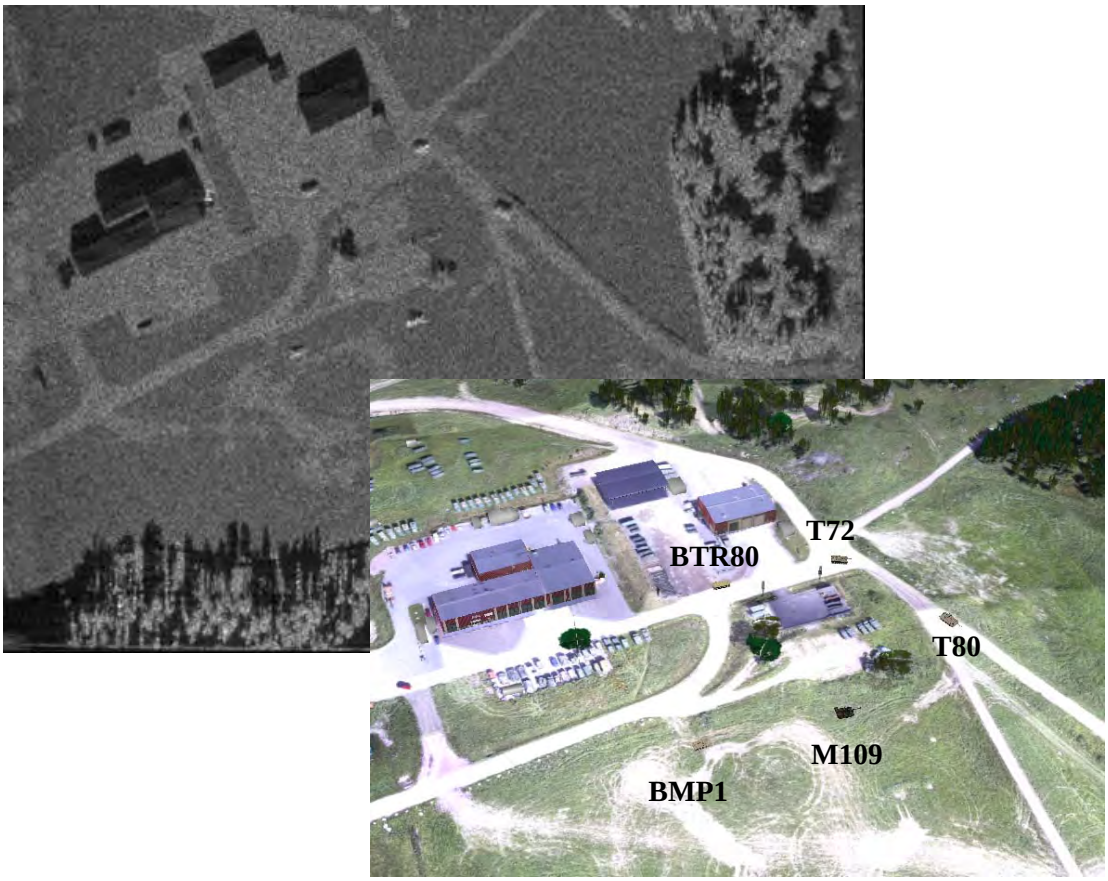


Figure 2-5: Simulated SAR Image and a View of the 3D Model Including Five Military Vehicles (FOI).

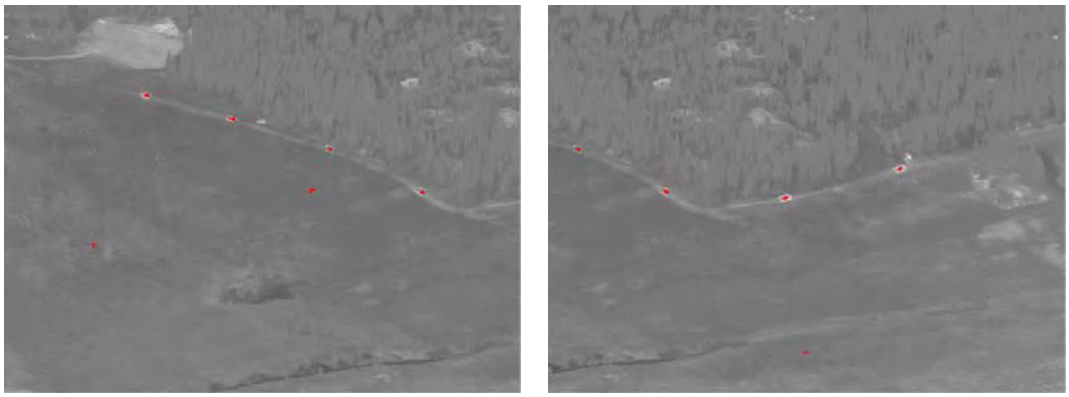


Figure 2-6: Snapshots from a Sequence of Simulated IR Images. The output from an automatic detection algorithm is marked with red pixels (FOI).

2.4 AUTONOMOUS NAVIGATION

Unmanned vehicles which have the capability of operating with autonomy are valuable assets to the military, both in wartime and in response to natural disasters. The ability of vehicles to effectively navigate and perform designated tasks within extremely harsh or dangerous environments and without the

requirement of operator intervention, provides a distinct advantage over conventional operations. A typical mission would be for a vehicle, perhaps after receiving only high level commands from a human operator, to perform certain tasks while moving from point A to a point B through a cluttered, unknown, highly dynamic urban environment. Possible objectives of the mission could be to:

1) Gather intelligence or provide surveillance along the preliminary route;
2) Provide a detailed map, or perhaps identifying changes to a previously provided base-map of the area;
3) Detect and identify potential targets (geometric and otherwise) along the route; or
4) Precisely deliver a payload.

Other objectives may be for example, to autonomously land rotorcraft to resupply war fighters or perform search and rescue operations. It is easy to imagine many variations and extensions of these scenarios which may also rely upon line of sight and navigational stealth within the urban setting. Moreover, these operations may be performed by solitary agents or collaboratively.

There are numerous technical challenges that must be addressed in order for unmanned vehicles to operate autonomously in cluttered urban environments. An autonomous aerial vehicle, for example, will encounter a diverse range of obstructions in such an environment including buildings, trees, power poles and lines, road signs, and moving objects such as people or other vehicles. Flight of a small vehicle through an urban warfare environment will involve frequent blockage of communication with other vehicles/ground stations, periodic obscuration of GPS signals, and jamming from enemy forces. This means that reliable communications from or with other vehicles and satellites cannot be assumed and, therefore, a UAV under these circumstances cannot rely on external sensing and processing resources for extended periods of time. Naturally, an autonomous vehicle should always take advantage of information from other sources when available, but must independently sense its surroundings and conduct path planning exclusively with onboard assets when external information is not available. Therefore, the onboard sensor suite and the ability to utilize sensor data for obstacle avoidance and path planning are critical in order for the autonomous vehicle to operate effectively in an urban canyon.

Another critical technology for autonomous navigation is the ability to synthesize sensor data for Guidance, Navigation, and Control (GNC), including obstacle avoidance manoeuvres. As briefly described earlier, the user provides high level commands to the autonomous vehicle, such as, for example, monitor specific objects in a loiter pattern, search along specific paths, or follow some predefined search logic. In each case the guidance and control system must satisfy the mission objective while taking action to avoid sensed obstacles in its path. In the case of aerial vehicles, flight control algorithms must necessarily plan ahead for more than one maneuver, so that required collision avoidance tactics do not deleteriously affect the vehicle's later options and prevent it from recovering to complete its mission. The control system of a small autonomous aircraft must be able to accomplish these tasks in the presence of local disturbances such as wind gusts and simultaneously deal with complex, uncertain dynamics. The (near-) optimal control algorithms for 3D flight path planning are computationally expensive and require up-to-date reconstructed geometry from the current 'sensing horizon' of the vehicle's on-board sensors. The overall flight control problem is formulated as a *receding horizon control* optimization [2],[3],[4], which entails extremizing a performance functional over a moving window in time to determine the optimal control inputs. Receding horizon control represents an extremely general framework for addressing the problem of autonomous flight in urban environments. Most importantly, it allows for the incorporation of multiple constraints in the optimization problem so that obstacle avoidance can be achieved by enforcing constraints on the aircraft states. Although still computationally intensive and subject to instabilities, the receding horizon controller formulates the equations of motion with reduced-order representations of the non-linear aircraft-sensor dynamics and solves many small-scale optimization problems with updated constraints based on new information such as updated estimates of obstacles and the current reconstructed geometry of its environment.

Necessarily, autonomous navigation requires the development of complex 3D models of urban terrain in real time from data, either generated progressively or updated, depending upon the collection context, by sensors onboard the vehicle. The sensors typically provide ranging estimates of structures and obstacles in the operating environment, which may be derived from passive cameras (using optical flow or stereographic methods such as Structure from Motion as described in Section 3.1.2) or active laser measurements (LIDAR). At this stage of assimilation the measurements are registered relative to the vehicle and are available for reactive control and emergency avoidance. To register the range data within a global map, transformations must be performed after accurate geolocation and pose estimations, using either an internal navigation system, involving GPS and internal measurements (e.g., accelerometers, gyros, and magnetometers), simultaneous localization and mapping algorithms (*SLAM*, [5],[6],[7]), or hybrid combinations of the two. As the sensor data is assimilated into an obstacle map in three dimensions, the vehicle must then adaptively modify its path planning and motion control mechanisms, using receding adaptive control.

Chapter 3 – SENSOR SYSTEMS FOR GENERATING 3D POINT CLOUDS

This chapter provides a review of the different sensors and sensing techniques that can be used to collect 3D point clouds and aid the extraction of features. Two different physical processes have been exploited to measure the range and consequently determine 3D position: triangulation and time-of-flight. Time-of-flight sensors are active (electromagnetic or acoustic) and measure the time taken for a pulse to travel from the source to the feature of interest and back to the sensor. Triangulation is based on geometry and can be active or passive; it requires accurate registration of two positions separated ideally by a large baseline. As a general rule of thumb triangulation can provide accurate range information to approximately 20 times the baseline and as such is of more value for operation at relatively short range. Active systems have greater applicability to operation at longer range but with added complexity.

3.1 PASSIVE SENSING

Stereo electro-optic imaging sensors can be used to collect imagery which can be registered and used to extract range information. Although stereo imaging cameras are becoming commercially available [8] they are not designed to extract point clouds but rather for direct 3D display using stereo imaging systems exploiting polarization, colour (anaglyphs) or viewing angle (autostereoscopic).

A passive stereo imaging system has a choice of sensors with visible and near-infrared cameras (CCDs or CMOS) offering the best angular resolution but unless amplified by using for example electron multiplication (EMCCD) then these sensors can only be operated in the day-time under solar illumination. Infrared sensors operating in the midwave (3 – 5 μm) or longwave (8 – 12 μm) offer greater day/night capability and superior penetration through atmospheric obscurants such as fog and mist. However practical demonstrations of infrared stereo are much less common, simply because of the increased cost of the sensors. Moving to longer wavelengths there is potential to use RADAR imaging to provide an all weather capability and even image inside buildings; progress in these areas is discussed below.

The automatic extraction of features can benefit from the different imaging modes. The simplest discriminant is of course intensity which can be used to segment different features in a video stream. Colour (or frequency) is another common discriminant with RGB cameras in the visible waveband widely used and multi-band infrared cameras becoming available. Extending the spectral theme, hyperspectral cameras can discriminate fine spectral differences and aid the process of feature extraction. Polarisation is a further process that can be exploited. Natural illumination is polarized and depending on surface structure and angle of reflection many objects will retain this polarization. It is well known that polarisation can be used to discriminate between man-made objects such as vehicles and natural objects such as bushes and trees. However possibly the most powerful discriminate for automatic feature extract is position. Position in two planes (or angles) is easily measured using a focal plane array but the third dimension (range) is a much more difficult challenge and is the topic for the remainder of this chapter.

3.1.1 3D Point Clouds from 2D Images

Most passive systems work analogously to the human visual system. The location of 3D objects is derived from parallax, i.e., small changes in the direction from sensors to object. Since depth information is lost when taking a picture, one cannot exploit a single image without additional information. A single sensor has to move in order to determine the 3D structure by triangulation or one needs a second sensor. Passive imaging sensors measure directions, not distances. Therefore, the objects can be reconstructed only up to a spatial similarity transformation, which is specified by seven degrees of freedom (three translation parameters, three rotation parameters and one scale parameter). Using two or more synchronized cameras

mounted on a stereo rig offers the possibility to introduce the scale by measuring the length of the base line used.

For the orientation of the images and the determination of 3D coordinates of objects, mathematical models of the scene and of the sensor(s) are needed. Usually these models assume that the observed scene is static and that the extracted 3D points lie on opaque, non-specular surfaces. Concerning the camera, one typically assumes a straight-line-preserving projection, that is, straight lines in the world, e.g., building edges, are mapped to straight line segments in the image. Furthermore, it is assumed that a unique projection centre exists, so that the light rays between object and image points pass through a single point, i.e., perspective geometry is present, cf. Figure 3-1. To ensure the compliance of this camera model, it is advisable to perform sensor calibration either by a lab calibration or, if possible, by self-calibration on the fly.

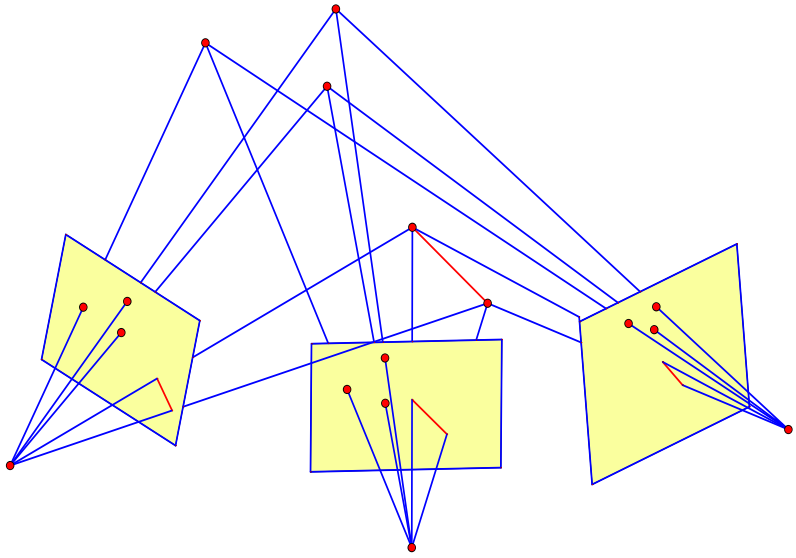


Figure 3-1: Triangulation Principle Depicted Using Three Images Captured from Two Different Viewpoints.

Approaches that automatically acquire 3D points and/or 3D line segments from sequences of images at unknown locations, using projective geometry and automatic correspondences, are available (see standard textbooks [9], [10], [11] for instance). For convenience, the presentation in what follows is restricted to 3D points. The general procedure is first to establish correspondences between image features of consecutive images of a sequence. These correspondences are then the base for the subsequent parameter estimation step that determines the positions of the 3D points and the poses of the moving camera simultaneously. Figure 3-1 depicts the situation for two images captured from two different viewpoints. Given the poses for the two camera viewpoints, the 3D points can be determined by triangulation (spatial forward sectioning, corresponding rays intersect at the 3D point sought). Conversely, given the positions of the 3D points, the poses of the camera can be determined by spatial resection. Since neither the camera poses nor the 3D point position are known, both parameter groups have to be determined simultaneously. Techniques to do this are known as Structure from Motion (SfM) in computer vision, Simultaneous Localization And Mapping (SLAM) in robotics and bundle adjustment in photogrammetry.

Basically, the correspondences of image features can be supplied in two ways: Given automatically extracted image points in one image (interest or salient points), the corresponding image points in a second image can be determined by feature tracking (see, e.g., [12]). These methods require images taken from viewpoints close to each other (short baseline) in order for the automatic correspondences to work, which makes them noise-sensitive and numerically unstable. Alternatively, the image features can be extracted in all images

separately and then the correspondences established by matching. These methods (see [13]) have attracted much attention especially for wide-base-line applications.

Although bundle adjustment yields an optimal solution, it is not suitable for large data sets due to limitations in computational resources. A workaround is the application of sliding window approaches for filtering [14], which allows online processing. Figure 3-2 shows three representative images from an infrared image sequence captured by an airborne platform (for details refer to Section 4.4). The extracted and plotted interest points correspond to hot and cold spots respectively [15],[16]. The estimation results for the corresponding point cloud and the camera poses are shown in Figure 3-3.

Figure 3-2: Three Representative Infrared Images Showing an Urban Area with Plotted Interest Points.

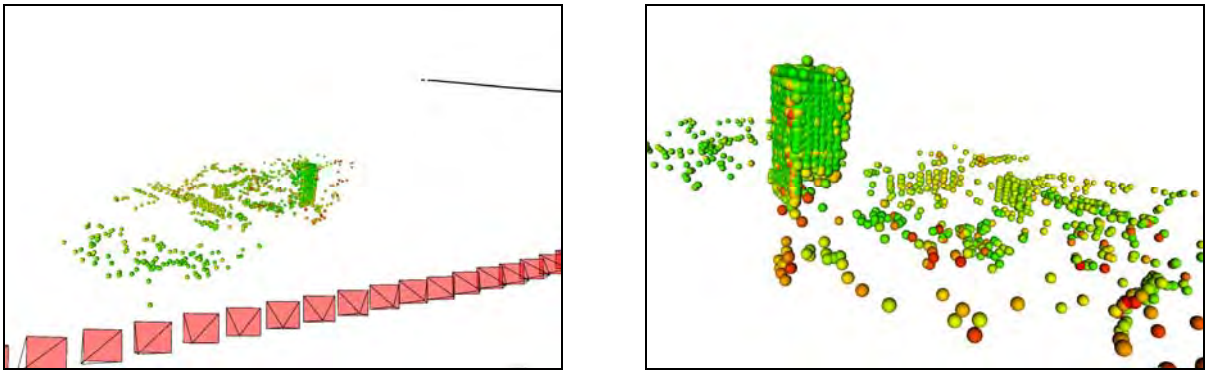


Figure 3-3: Two Views of the Reconstructed Urban Scene. The colours of the 3D points refer to positional precision (green = more precise, yellow = less precise, red = still less precise); the tetrahedra in the foreground depict the estimated poses of the used camera.

The spatial distribution of the 3D points depends on the image contents. Usually one gets a sparse representation with a heterogeneous distributed point cloud. This cloud of 3D points has no structural information among the points. Specifically, there is no topological information to represent connectivity among points. In geometric modelling, it is necessary to convert this cloud of points into a surface representation such as a mesh model. (cf. Section 4.3). Approaches to obtain a dense surface representation have gained much attention recently; cf. [17] for instance. For derivation of these digital surface models or elevation models, one needs smoothness assumption about the surface.

3.2 ACTIVE SENSING

Time-of-flight sensors are the most common active sensors but before describing these it is worth noting that triangulation can also be used in active mode to measure range. If the baseline between the emitter and detector are accurately know then the angle of arrival can be measured and the range calculated.

An excellent review of active triangulation and a comparison against time-of-flight systems is given in reference [18]. The advantage over purely passive systems (structure from motion) is that the point of interest is clearly defined by the laser spot, which can be scanned to construct a 3D image. However the range precision deteriorates with the square of the range to the target so these systems are best suited to relatively short ranges (out to 30 m) and the system complexity makes the purely passive system more attractive. For operation out to many kilometers and to remove the correspondence and disparity problems associated with stereo vision [19] time-of-flight systems need to be employed.

3.2.1 Time-of-Flight Systems

The pulsed TOF LADAR system transmits a laser pulse of short duration to the target of interest and measures the time for the reflected pulse to return to the sensor. Given a measured round-trip time for the pulse, t and the speed of light, c, the range to the object point is calculated using:

$$R = \frac{ct}{2} \tag{3.1}$$

The important system parameters can be understood by considering the energy incident on the target and relating it to the energy returned to the imager. The basic physics behind a TOF system is described by the laser RADAR equation [20]:

$$E_R = \frac{4KE_T T}{\pi \phi^2 R^2} . \Gamma . \frac{T}{4\pi R^2} . \frac{\pi D^2 \eta_R}{4} \tag{3.2}$$

where the energy returned to the sensor (E_R), is dependent on the four processes captured by the four parts of equation 3.2. The first part describes the energy density at the target, which is dependent on the transmitted energy (E_T), the atmospheric transmission (T), the beam divergence (ϕ – beam width in radians) and the range, R. K is a factor to account for the beam profile at the point of interest within the beam. The second physical process is the reflection of the radiation from the target and this is captured by the laser cross-section, Γ. The scattered light propagates back towards the sensor with an area that expands as R^2. Finally the reflected energy is intercepted by the objective lens of the sensor (Diameter, D). η_R is the receiver efficiency and includes the losses associated with system transmission and those dependent on the detection technique, which could include loss in coherence or polarisation if appropriate.

The equation above is used if the target is small with respect to the field of view of the sensor. However, in the case of interest here, the laser divergence is chosen to match the detector field of view. For a scanned system with a single detector the laser beam quality is important and a diffraction-limited beam may be needed. In a flash illumination system with a 2D array of pixels the laser beam can have poorer spatial mode quality. If Lambertian (isotropic) scattering is assumed and the laser and sensor are coincident then equation 1 can be reduced to:

$$E_R = \frac{E_T T^2 \rho D^2 \eta_R}{4R^2} \tag{3.3}$$

where ρ is the total hemispherical target reflectance. This equation clarifies the dependence on the square of the atmospheric transmittance, objective diameter and the inverse square dependence upon range.

If a detector array is used then the energy in a pulse is shared between the pixels, which necessitates high energy, whereas for a single detector system a high repetition rate source is needed to provide a rapid scan. Typically the laser power for both systems will be comparable but delivered in a different format.

3.2.1.1 2-Dimensional (Flying Point) Scanning Systems (LIDAR)

Light detection and ranging (LIDAR) or laser detection and ranging (LADAR) sensors have been developed that exploit various different detection and modulation processes. Direct detection LIDAR is by far the most common but frequency and amplitude modulation techniques have also been developed and demonstrated to great effect.

Direct detection systems transmit a short laser pulse (~3 ns) and digitise the returned signal; the time delay is used to measure the range to the target. With a fully digitised return signal processing techniques can be used to extract not just the range to the target but the structure within the pixel. For example first and last pulse return can be used to measure the height of a forest canopy and the depth of the forest floor (for further discussion see Section 3.2.2). Typical performance figures for some airborne LIDAR systems are tabulated in Table 3-1. Modern airborne systems operate at a pulse repetition frequency of around 100 kHz, where the repetition frequency is reduced as the range is extended. Typical systems operate with an average power of a few watts and can achieve a range out to a few kilometres dependent on the reflectivity of the target material. Generally, the trend is towards eye-safe wavelengths around 1.5 μm in order to operate with higher power and extend the range. Range precision down to a few centimeters is possible with absolute accuracy of ~10 cm more realistic. Beam divergence and pointing accuracy typically limit the angular resolution to 0.1 mrad (10 cm at 1 km).

Table 3-1: Provides an Overview of Commercial Airborne LIDAR Systems and Achievable System Performance. The table is based on information collated in reference [21].

Manufacturer	Airborne Hydrography	Leica Geosystems	Optec	Riegl	Topo Sys GmbH
System	Hawk Eye II [22]	ALS60 [21]	Orion [23]	LMS – Q680 [24]	Falcon II [25]
Wavelength	532 nm	1064 nm	1064/1541 nm	1550 nm	1560 nm
Pulselength	5 ns	5 ns	< 7 ns	< 4 ns	5 ns
Pulse Frequency	64 kHz	20 – 200 kHz	50 – 200 kHz	80 – 400 kHz	83 kHz
Echoes/Pulse	4	4	4	Full waveform at 1 GHz	9
Range Precision	3 cm	3 – 4 cm	< 5 cm	4 cm	1 cm
Spatial Precision at 1 km Range	50 cm	20 cm	25 cm	10 cm	20 cm
Operational Height	500 m (typical)	200 m to 5 km	200 m / 1 km / 2.5 km	1.0 – 2.5 km	1.6 km

One of the rapidly growing markets for LIDAR technology is ground-based mobile mapping whereby a LIDAR system is mounted on a moving vehicle and operated in a 360 degree scanning mode. Many mobile services [26] and systems are becoming available such as Lynx for Optech [23] and HDL-64E from Velodyne [27]. These systems offer vertical coverage up to ~30 degrees with a distance accuracy of ~2 cm out to a range of approximately 100 m. They are ideal for imaging the façade of buildings and other areas inaccessible to airborne survey. At the cheaper end of the market the LD-LRS series of sensors from SICK [28] are very popular and were used extensively in the DARPA urban grand challenge for autonomous navigation.

Frequency modulated systems measure range by impressing a linear frequency chirp upon the transmitted beam then mix the returned signal with a local replica to determine the frequency shift and hence the range.

This technique has been demonstrated to great effect with 1.5 μm fibre laser technology. Erbium fibre laser technology is attractive for LIDAR applications as it promises a compact, reliable laser source however direct detection systems need short pulses and the high peak power is incompatible with the small fibre core. QinetiQ in the UK avoided this problem be developed a chirped pulse system that achieves excellent range precision with a relatively long frequency chirped laser pulse. This approach avoids the high peak powers that would otherwise damage the fibre allowing them to operate at considerably longer range than traditional LIDAR systems [29]. A representative image collected at a range of 18 km is shown in Figure 3-4.

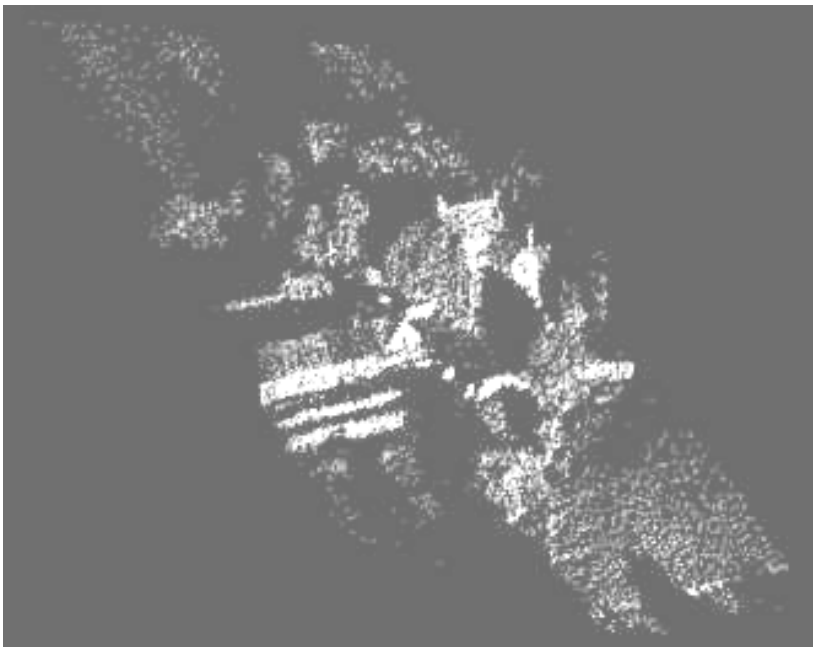


Figure 3-4: Farm House and Outbuildings at 18 km (Ground-Ground Demonstration) – Reproduced with the Permission of QinetiQ Limited.

Amplitude modulated systems transmit a beam in which the signal level is modulated as shown in Figure 3-5. The phase change, θ between the transmitted beam and received signal is measured and the range to the target is calculated using:

$$R = \frac{\theta\lambda}{4\pi} \tag{3.4}$$

where λ is the wavelength associated with the modulation (rather than the carrier). LIDARs based on amplitude modulation provide excellent results for short range operation but ambiguous results are obtained at range intervals equivalent to a 360 degree phase change. For example with a modulation frequency of 10 MHz unambiguous results are obtained out a range of 15 m ($\lambda/2$ as the beam must travel to and from the target) but targets at 5 m and 20 m will be ambiguous. Various techniques have been developed to mitigate the range ambiguity such as the use of multiple frequencies whereby the range ambiguity is extended to the lowest common multiple. Focal plane arrays based on the amplitude modulation technique are becoming widely available and will be discussed below.

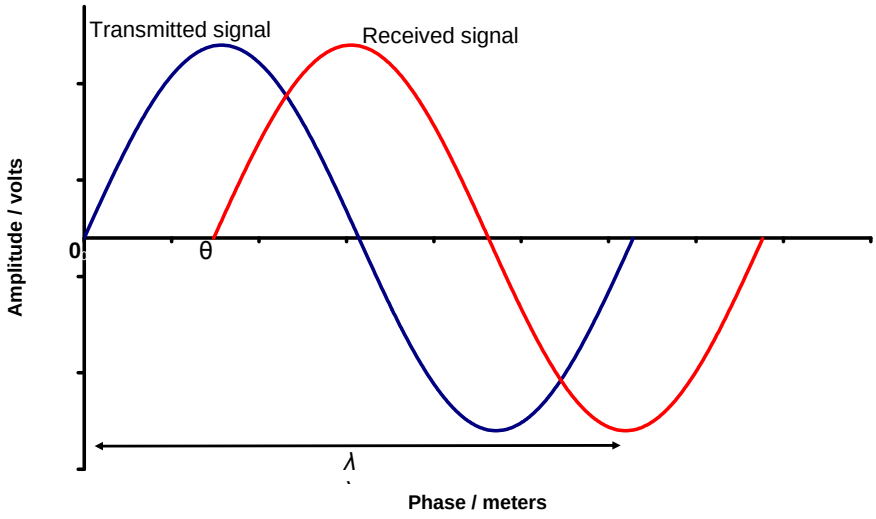


Figure 3-5: Transmitted (blue) and Received Signals in an Amplitude Modulated LIDAR System.

3.2.1.2 1-Dimensional Scanning Systems

A natural progression from the two dimensional LIDAR scanning is to scan in only one dimension and use a sensor array. This mode of operation has the potential to achieve a significant increase in the area coverage rate from these airborne scanning systems. A 1-dimensional scanning system can be realised in a number of different ways:

- A linear array of detectors and scanning either azimuth or elevation.
- A gated focal-plane array in which the position of the gate can be scanned with respect to transmitted laser pulse.

In the gated imager a short laser pulse illuminates the target and a gated camera (possibly intensified) is used to capture range slices. Andersen et al. [30] have used laser pulses of 500 ps duration at 532 nm and a gated and intensified camera to capture facial profiles. Relative range accuracy down to 1 mm was achieved at distances up to 500 m.

There are a number of examples of temporal scanning to construct a 3D point cloud however very few examples of LIDAR systems operating with linear detector arrays in a line scan-mode. This is quite surprising as there are many technical challenges in moving from a traditional single-point LIDAR system to a staring 3D imager and a line-scan system could simplify many of these problems. Reducing the physical size of timing circuitry such that it can fit into a pixel is a significant challenge. In a linear array there is extra space that can be used to control the timing circuitry and extract the signal from the pixel. Increasing the laser energy from that required for spot illumination to burst illumination is also quite challenging. Illumination over a strip could be more easily achieved. The linear array only makes sense if the other dimension is provided by platform or sensor motion but in many of the applications of interest (airborne LIDAR, mobile mapping) this is exactly how the system would be operated.

3.2.1.3 3D Staring Arrays (Flash LIDAR)

In a 3D staring array each pixel in the array captures the range to the target. This is challenge both for the illumination source which must illuminate a large area and for the focal plane array designer who must include timing circuitry within each tiny pixel.

One of the first and most impressive demonstrations was achieved under the DARPA JIGSAW programme [31] which demonstrated 3D imaging of a vehicle under a thick forest canopy by exploiting the small gaps between the leaves and image registration from multiple viewing angles. The JIGSAW sensor is based on a 32 x 32 focal plane array in which each detector is operated in the Geiger avalanche mode such that a small signal (approaching a single photon) returned from the scene triggers the avalanche process and stops the timer. The pixel size at 30 μm diameter is relatively small but the pixel pitch at 100 μm means that the sensor has a poor fill factor. To compensate for the low fill factor the laser illuminator consists of an array of beams (32 x 32) generated from a single beam using a diffractive optic. The system operates with a frequency doubled Nd:YAG laser operating at 532 nm with pulse duration of 300 ps and repetition rate of 16 kHz (~16 million pixels/sec). A range resolution of 40 cm has been demonstrated.

The JIGSAW programme achieved very impressive results but from relatively short ranges (~200 m). Other groups have focussed on a higher density of pixels on the focal plane array and operation at longer range, which generally means operating with an eye-safe laser source.

ASC (Advanced Scientific Concepts Inc [33]) has developed a 3D staring array camera system (TigerEye 3D). The advanced Readout Integrated Circuit (ROIC) incorporates both analogue and digital processing on each pixel to capture the leading edge of the returned signal and sample the returned waveform. Each pixel can achieve 15 cm range precision reducing to a few centimetres after processing the waveform. The ROIC has been hybridised with InGaAs avalanche photodiodes and operates in the eye-safe waveband at ~1.5 μm. The detector array has been implemented in a 128 x 128 array on a 100 μm pitch. The active pixels are smaller than the pitch at 20 μm to reduce the noise so a lenslet array is used to improve the quantum efficiency. This removes the requirement for an array of laser beams and the system can be used with burst illumination laser. The detector array has been implemented in the camera system shown in Figure 3-6 and used to collect 3D movies, an example image is also shown in Figure 3-6. The system operates out 1 km range using pulses of 7 mJ at 30 Hz [32] which corresponds to a transmitted power of 0.2 W for 500 x 10^3 voxels/sec. A comparison of scanning LIDAR versus 3D flash LIDAR has concluded that the flash imaging can be up to 4 times quicker [34] for equivalent transmitted power levels.

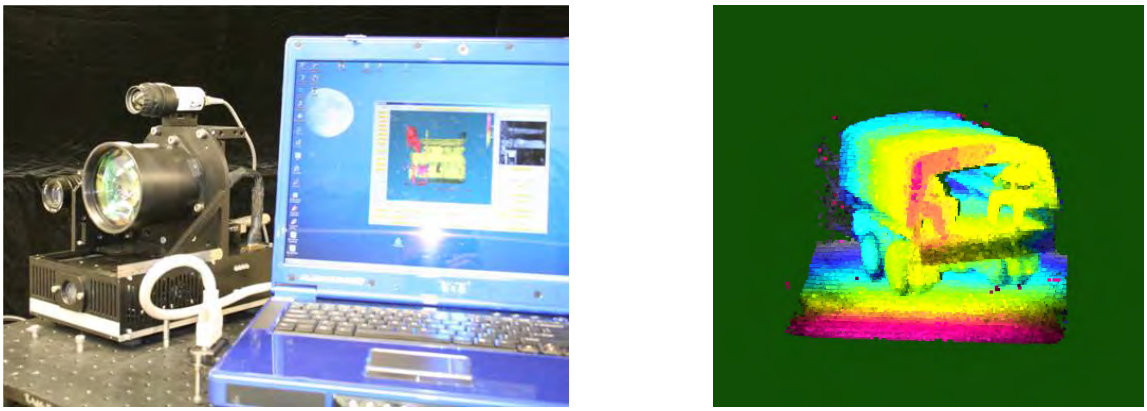


Figure 3-6: 3D Flash Illumination Imaging System and Sample 3D Image of a Truck.

Selex in the UK has been developing a 3D time-of-flight focal plane array based on avalanche photodiodes in MCT [35],[36]. The goal is to develop a long range 3D imaging capability that can supplement thermal imaging systems and provide enhanced resolution at long range. MCT was chosen as the diode material because it can be operated in avalanche gain mode in the eye-safe waveband with relatively low voltages and it has a very high electron-to-hole mobility ratio which provides nearly ideal (noise-free) avalanche gain. Operation in avalanche mode is needed to achieve the range performance and achieve compatibility with airborne laser designator technology. The requirements drive the specification towards small pixels and

operation with laser pulses typically 10 – 20 ns. A novel timing circuit has been developed to enable a demonstration with pixels of 24 μm diameter.

A focal plane array has been developed in half TV format (320 x 256) camera, integrated with a big-sky laser and demonstrated in outdoor trials. The timing circuitry on each pixel responds to the centre-of-gravity of the laser pulse and range precision of less than 30 cm can be achieved. A picture of the 3D laser imaging rig and an example of the imagery collected is shown in Figure 3-7. The current technology is based on Gen II (loop-hole) CMT technology but research is progressing towards Gen III technology which will enable larger format arrays (full TV) and multi-functional performance; mid-wave thermal imaging and 3D range measurements on a single focal plane array.

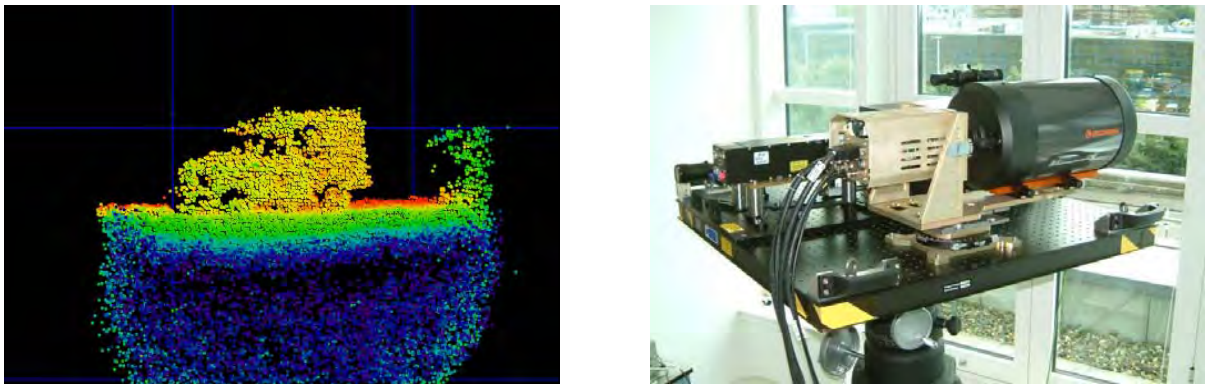


Figure 3-7: UK 3D Laser Imaging Camera System and Sample Imagery of Land Rover.

There has been a flurry of development activity on 3D cameras based on amplitude modulated time-of-flight focal plane arrays. Detector array developments, that allow the array response to be modulated at the same frequency as the illumination source, have greatly simplified the system design allowing relatively large arrays (VGA) with small pixels (40 μm) to be produced. Typical modulation frequencies of ~15 MHz are used to achieve range accuracy below 1 cm out to a depth of approximately 7 m. The technology is focused towards the vehicle safety and the computer gaming launch and current cameras operate with typical integration times of 10 ms and over a wide field-of-view (up to 60°). Of particular note is the acquisition of 3DV systems but Microsoft [37] which has plans to market a new 3D gaming environment with its Xbox 360 called Kinect [38] with integrated RGB camera and microphone array. This 3D imaging technology is designed for short range operation but with suitable modifications to the optics, modulation frequency and illumination source it should be feasible to exploit this new class of sensors for ground-based urban mapping at longer range.

Table 3-2: Comparison of Performance Parameters for Amplitude-Modulated Camera Systems.

Company	3DV Systems	Canesta	Mesa-imaging	PMD Technologies	Primesense
Camera	Zcam		SR4000	Camcube 2.0	PrimeSensor
Array Size	640 x 480	160 x 120	176 x 144	204 x 204	640 x 480
Pixel Pitch			40 µm	60 µm	
Operational Range	0.5 – 2.5 m	0.3 – 1.5 cm	0.8 – 8.0 m	7 m	0.8 – 3.5 m
Modulation Frequency			15 MHz	18 – 21 MHz	
Resolution	1 – 2 cm	0.3 – 1.5 cm	1.5 cm	3 mm	1 cm
Frame Rate	60 fps		54 fps	25 fps	
Illumination Wavelength		870 nm	850 nm	870 nm	

3.2.1.4 Through-Wall Synthetic Aperture RADAR

The principle of operation of a Through-Wall Synthetic Aperture RADAR (TWSAR) consists in transmitting a RADAR signal towards a scene of interest, a building in this case, and collecting the signal that is reflected back from the wall, and from the targets and clutter behind the wall. One type of setup consists in using separate transmit and receive antennas. These antennas can be mounted on a rail system (Figure 3-8a) or on a vehicle in a side-looking configuration (Figure 3-8b). As the system is moved (along the rail or by driving the vehicle), a chirp signal is periodically transmitted through the transmit antenna towards the wall. The returns collected by the receive antenna are summed coherently to form a synthetic aperture.

a) **b)**

Figure 3-8: TWSAR Prototype – a) Mounted on Rail; and b) Mounted on a Mobile Lab.

Through-wall RADARs have the potential of mapping interior room layout, including the location of walls, doors and furniture. They could provide information on the in-wall structure and detect objects of interest concealed in buildings such as persons and arm caches. However, imaging through walls presents many technical challenges. Through-wall RADARs must cope with signal transmission and reception through inhomogeneous media and require a large dynamic range. The later requirement results from a desire to detect weak signals coming from behind the wall in the presence of direct, strong signal reflections from the external wall surface.

Typical synthetic aperture RADAR systems produce 2D images that provide range and azimuth information. Figure 3-9 shows one such image. It was acquired by driving the vehicle-mounted TWSAR testbed shown in Figure 3-8b along a building wall [39].

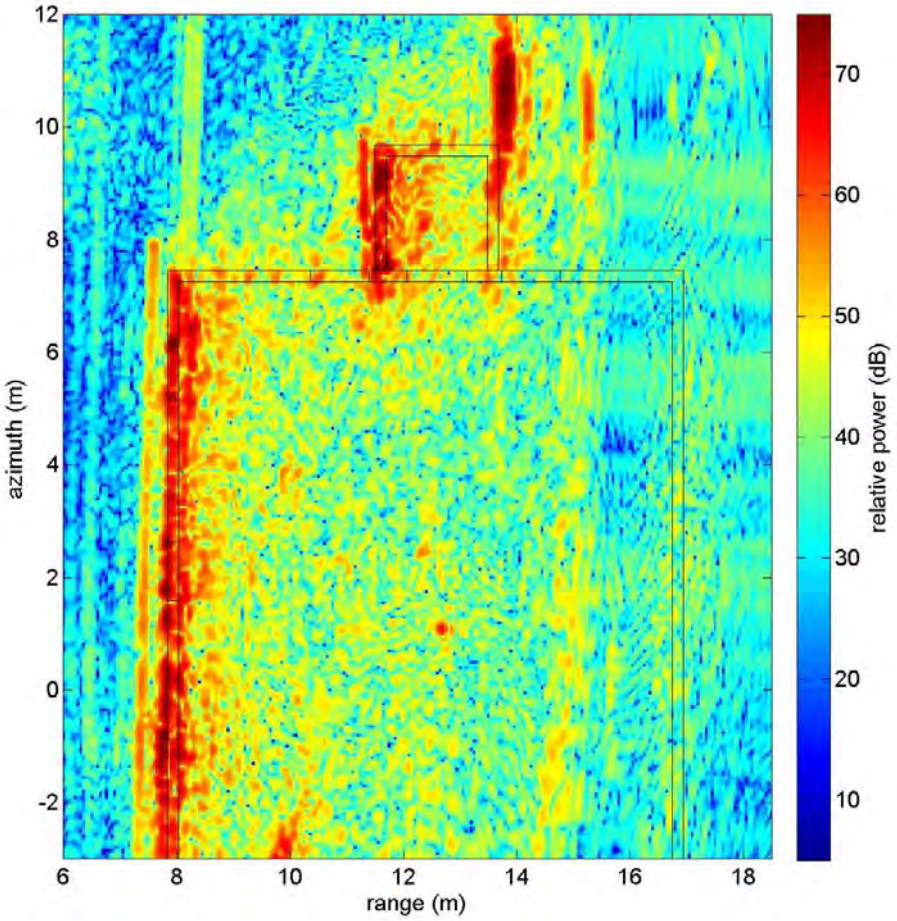


Figure 3-9: 2D TWSAR Image of a Wood Building with Building Layout Outlined in Black.

Research is ongoing to develop 3D TWSAR systems capable of producing 3D images that would provide information in the elevation dimension. Efforts are also made to elaborate wall compensation strategies and to develop tools for target detection and classification [40].

3.2.2 Estimating Distances through Waveform Analysis

In this section we address some issues related to measuring range with LIDAR systems. We will consider two aspects: signal processing techniques for detecting pulses in received waveform signals, and practical issues related to LIDAR measurements of an urban scene.

3.2.2.1 Detecting a Pulse

Here, we will give a short discussion about the basic signal processing aspects on how a LIDAR system operates in terms of extracting pulses and estimating range from the received laser light. As most LIDAR systems used for 3D urban mapping operate using the time-of-flight principle, we will limit the discussion here to such systems. For a comprehensive overview on state-of-the-art within this topic, as well as specifications of several commercial LIDAR systems refer to [41].

In order to describe the principle of operation, first assume that a laser pulse is emitted at time t_0. Furthermore, assume that a sufficiently large proportion of the laser light is reflected off at least one surface in the path of the laser beam and back to the detector. The reflected pulse reaches the detector at time $t_0+\Delta t$. Given the speed of light c, the range can be estimated through $d = c\Delta t$. Obviously, in order to obtain good range data it is crucial that Δt is estimated accurately.

A number of approaches are proposed for solving this task, of which three seem to be especially frequent: peak detection, constant-fraction detection and correlation-based detection. The first two techniques typically work satisfactorily if the Signal to Noise Ratio (SNR) is good, which is why one of them is often found in single-pixel (scanning) systems. As the noise level increases, however, the performance of these techniques degrade and the benefits of using the more computationally demanding correlation-based techniques such as the matched filter become more prominent [42]. In the matched filter, the received signal is correlated with the emitted signal which generally increases the robustness of the range estimates in the presence of noise.

In practice the received signal is a sampled and quantized version of the actual physical signal. The simplest way to estimate Δt is to select the position of a sample as the true position, be it the peak value or the first value above a certain threshold. However, since the true range value is likely to lie somewhere between two samples, this kind of approach will produce an expected range error that is proportional to the distance corresponding to the time between two samples. In order to increase the resolution and accuracy of the range estimate beyond the sampling resolution, some additional computations can – or should – be performed, e.g., through local weighting.

The above-mentioned techniques are based on the implicit assumption that each pulse in the signal is the result of only one reflecting surface, i.e., the pulse corresponds to one range value. In practice, this means that the reflecting surfaces in the path of the laser beam must be well separated in space, so that there is no significant overlap between their respective contributions. Now, assume that there are two reflecting surfaces and that the distance between them is small compared to the pulse width. The result is that the pulses reflected off these surfaces will merge into a significantly broader pulse. Obviously, treating that pulse as if only reflection was present is suboptimal. Not only will there be only range estimate, but the range estimate itself is likely to be quite uncertain. Typically, it will be somewhere between the true surfaces, producing virtual “ghost” points in the void between the surfaces. Such an unwanted effect calls for a more careful analysis of the signal. A possible remedy is to remove points stemming from signals that manifest significant pulse broadening; as the received pulse being significantly broadened compared to the emitted pulse indicates that the laser light has hit many surface elements on its way through the atmosphere (e.g., through a tree canopy). Conversely, if a laser pulse hits a single surface that is perpendicular to the direction of the laser beam, its shape is (more or less) preserved and we could expect the range estimate to be more accurate. In Figure 3-10, an example of pulse broadening is shown.

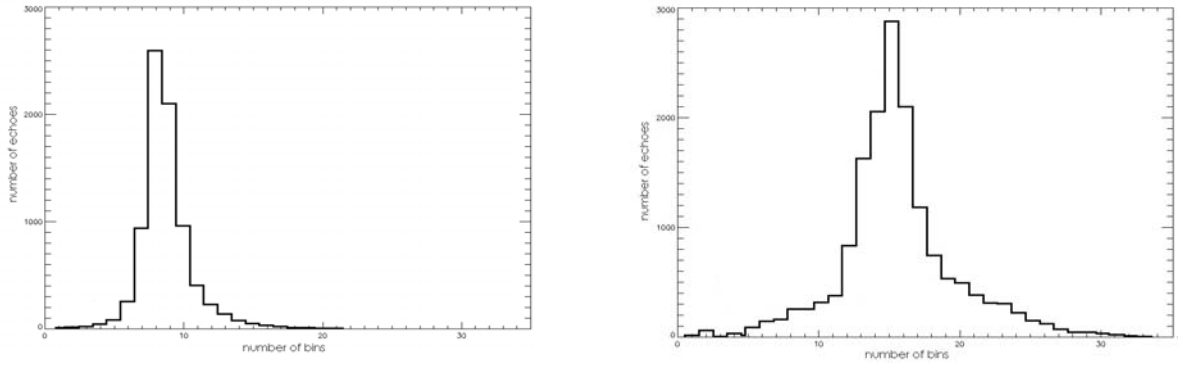


Figure 3-10: Histogram of the Width of Detected Pulses in an Area Containing Buildings (left) and Vegetation (right). Note the general broadening of pulses in the vegetation area due to the superposition of many small reflections from the leaves.

Note also the pulse broadening even occurs at slant incidence towards a single surface. Whether this broadening effect is large enough to cause trouble in practice depends on the size of the footprint of the laser beam and the incidence angle.

Pulses that have been significantly broadened can be subject to more refined analysis aiming at extracting information "hidden" in the signal. Two very useful techniques for this purpose are pulse fitting and deconvolution. The goal in pulse fitting is to find an optimal approximation of the signal using a set of basis functions. In other words, the signal is decomposed into a number of components that (hopefully) correspond to the contributions from the individual surfaces. The basis function is typically defined as the shape of the emitted pulse, or in practice, a nice function sufficiently similar to it.

The received signal $y(t)$ can be described as the emitted signal $s(t)$ convolved with a function that represents the scene $x(t)$. Whereas pulse fitting aims at approximating the signal at hand with a number of basis functions, the essence of *deconvolution* is basically to estimate what the received signal would have looked like if an ideal pulse was emitted, i.e., to estimate $x(t)$. If we are successful in doing so, the resulting signal will contain sharp spikes each of which corresponds to the position of the reflecting surfaces (Figure 3-11). The peaks may then be detected with one of the pulse detection techniques outlined above.

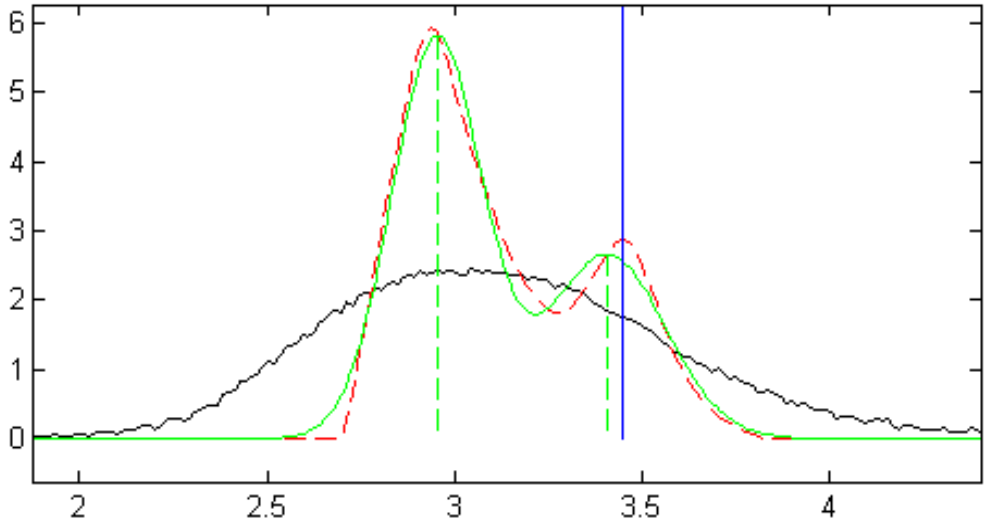


Figure 3-11: Deconvolution Example. The original signal (black) consists of two contributions: one from a hard surface (whose position is shown by the blue vertical line) and one from partially obscuring vegetation. The red dashed curve shows the result after deconvolving the original signal with the function of the emitted pulse. The green curve is the result of fitting two Gaussian functions to the red curve.

3.2.2.2 Measuring Distance in Practice

So far in this section we have discussed various signal processing tools that we have at our disposal for estimating range in the individual waveform signals, from simple peak detection to pulse fitting and deconvolution. Together with data about the pose (position and orientation) of the sensor, obtained from GPS/IMU measurements, we can describe the 3D position of each reflecting surface element in a global coordinate system, resulting in a (huge) point cloud that can be further analyzed (Chapter 4).

Obviously, a prerequisite for any pulse detection technique is that there is a decent signal to work with in the first place; the surface must reflect enough laser light back to the receiver to exceed the noise level. In practice there are factors that sometimes hamper the collection of good data.

One problem concerns certain materials that absorb quite strongly at the laser wavelength and thus often cause data dropouts. Another problem is caused by specular surfaces who act like mirrors reflecting a large portion of the incoming laser light in another, undetectable direction. This can be seen in regions with water, where the number of laser detected hits is generally quite low, especially if the water surface is calm (i.e., mirror-like). Furthermore, internal reflections in certain materials may lead to phase shifts that can result in an attenuation or extinction of the light.

Above, we mentioned the problem of resolving surfaces that lie close to each other. This is typically the case in vegetation and virtually no commercially available system allows for the extraction of individual leaves or branches of trees or bushes. However, LIDAR systems have an inherent ability to "see through" semi-transparent structures, e.g., vegetation, which in this respect often gives them an advantage over passive imaging. For the same reason that the sky can often be discerned through the canopy from underneath a tree, the LIDAR system can "see" the ground from above the tree top. And unlike passive systems that need the same spot on the ground to be visible in multiple images in order to produce a range estimate, the LIDAR system needs only one view on the same spot. Hence, provided that the amount of light reflected from the ground back to the detector is above the noise level, it is fully possible to accurately measure the ground level in vegetation areas.

3.3 REGISTRATION OF 3D POINT CLOUDS

Given the task to create a detailed and accurate 3D model about a certain object (or an entire scene, for that matter), it is desirable to have access to as many accurate and complete measurements of the object as possible. The ability to combine multiple data sets corresponding to the geographical area into one would allow us to obtain a more complete "view" of the object. A typical example from mapping of urban terrain is that data from an airborne platform may be combined with data from terrestrial measurements to capture both the roof and the walls. However, in order to be able to combine several data sets, they have to be aligned with one another. The first idea that may come to mind is to accurately measure the pose (position and orientation) of each sensor and then simply superpose the two datasets using the pose information. In practice, it may be quite risky to rely on such an approach, for at least two reasons. First, the alignment of the data sets can never become better than the accuracy of the pose estimates of the sensors. Second, the accuracy of the pose estimate may be subject to random and significant variations over time, e.g., depending on the quality of the GPS signal, which will make it difficult to assess the quality of the alignment at a given time. Luckily, there are data processing techniques to help in this matter, known as *registration* techniques. The goal of the registration process is to establish the geometrical relationship between multiple data sets so that all data can be represented in a common coordinate system.

As an example, assume that we have a primary point cloud **P** and a secondary point cloud **S** that we want to align with **P**. Normally we will treat **P** and **S** as rigid bodies. Then the registration problem is *to find the rotation matrix **R** and the translation vector **t** that brings* **S** *into the best possible alignment with* **P**.

In the early 1990's, Besl and McKay [43] presented a registration method known as the Iterative Closest Point (ICP), that has become very popular and is probably the most well-known registration method to date. Its main principle of operation is quite straightforward:

1) Pair each point in **S** with the closest point in **P**;
2) Compute the optimal transformation that minimizes the error (MSE) between the paired points; and
3) Apply the optimal transformation to **S** and update the MSE.

These three steps are iterated until a minimum (or a satisfactorily small) MSE is reached. Although very frequently encountered, ICP has a number of shortcomings. For example, it has an inherent tendency of creating a large overlap between **P** and **S** (as that tends to reduce the MSE) which is problematic if only the

parts of the data that actually overlap are relatively small. Since the advent of the original ICP, a number of improvement have been proposed in the literature, e.g., concerning different distance measurement, strategies for closest point search, false match rejection and motion estimation [44].

Another recently suggested solution to the 3D registration problem is based on the normal distribution transform, NDT [45]. Basically, in the 3D-NDT [46] approach the point clouds are divided into a number of cells, or voxels, each of which is described by a normal distribution function representing the probability of finding a data point at a certain position. Not only does this approach mitigate the problem of handling huge data sets (as only a few parameters are needed for each voxel), but it does not depend on computationally demanding nearest-neighbour search of the ICP algorithm.

A registration example is shown in Figure 3-12, where FLASH LIDAR data from Bonnland (part of the SET-118 data pool) have been stitched together with the ICP algorithm. The final result is a joint data set where the position of each LIDAR point is expressed relative to a reference frame (here, the very first frame). The green dots show the estimated position of the LIDAR sensor for each of the acquired frames.

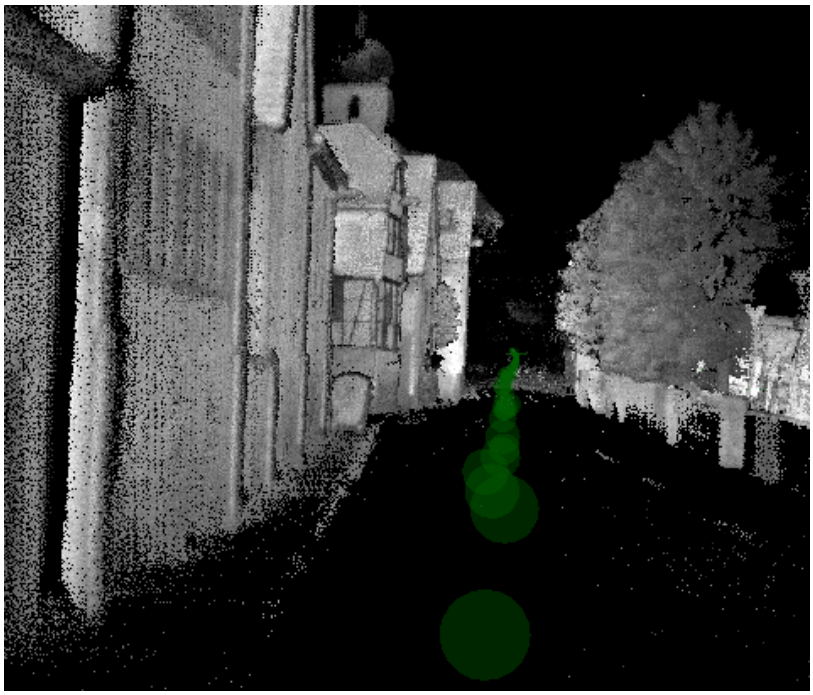


Figure 3-12: A Number of FLASH LIDAR Data Sets Collected from a Car while Driving Down a Street in Bonnland, Registered with the ICP Algorithm. The green dots show the estimated position of the sensor for each frame.

3.4 GEOREFERENCING OF 3D POINT CLOUDS OR MODELS

3D models may be acquired at different time and using different types of sensors. When it is required to represent a set of models in a common reference frame, one solution consists in georeferencing the models. Georeferencing means registering a 3D model against a global coordinates system so as to determine its position on Earth.

One method to georeference a 3D model relies on control points. A control point is a fixed point on Earth (such as a building corner) whose spatial coordinates are measured precisely. To compute the position of a 3D model in a global coordinates system, it is first required to survey the area where the model will be

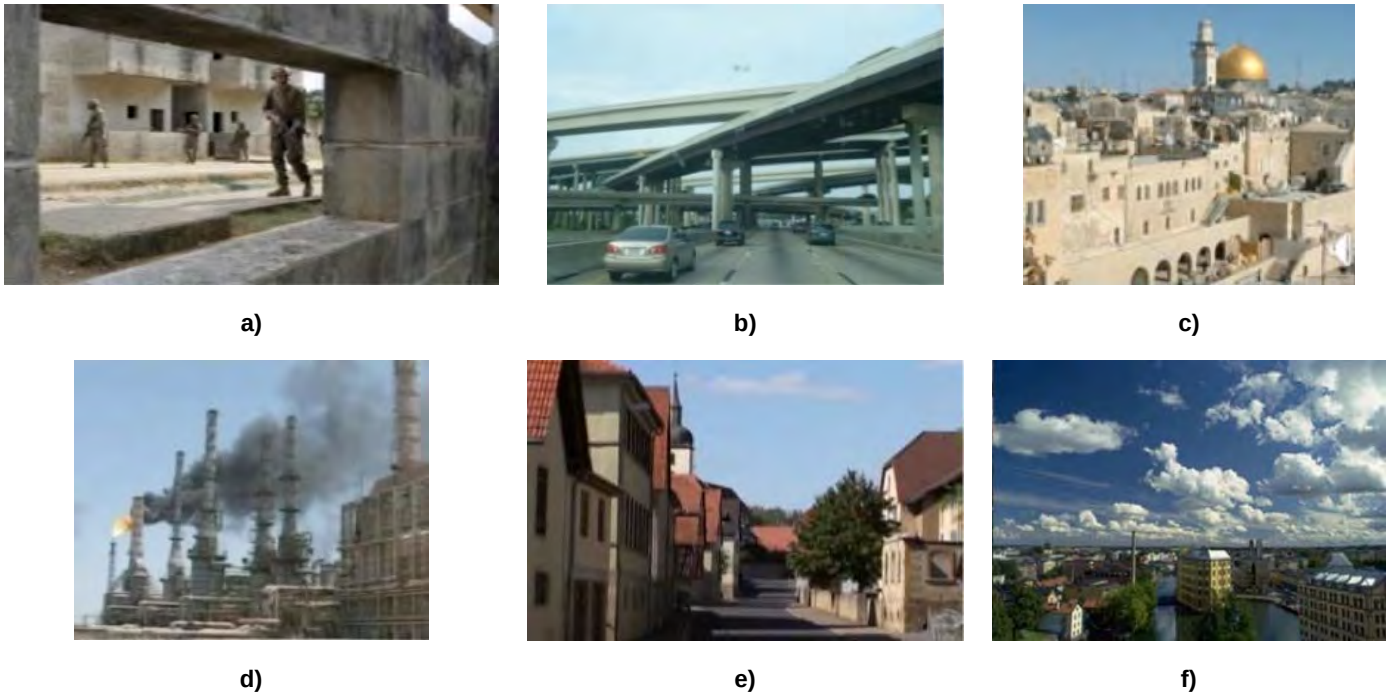
acquired to get the precise location of a certain number of control points. Correspondences can then be established between control points and their counterpart in the 3D model. The geographic position of the model is obtained by minimizing the position error over all the correspondence pairs. The accuracy of the georeferencing process can be quantified by evaluating the residuals of the minimization procedure.

Another possible solution to georeference a 3D model consists in using an integrated geo-positioning system to compute the trajectory of the vehicle mounted with the sensors used to acquire the model. Typical integrated geo-positioning systems generally combine the use of one or more GPS receivers, a Differential Global Positioning System (DGPS) receiver, an Inertial Measurement Unit (IMU) and a wheel odometry measurement device, all mounted on the acquisition vehicle. During data collection, the position of the vehicle is recorded by the geo-positioning system. Afterwards, the vehicle's trajectory is optimized and filtered using post processing software. Interpolation techniques and coordinates transformations are then applied to compute the global location of each of the acquired 3D points, which results in a georeference a 3D point cloud. This georeferencing method requires both the sensor data and position data to be accurately timestamped.

The position of a georeferenced 3D model is usually expressed using the Global Positioning System (GPS) where a geographic location is referenced by its longitude, latitude and altitude. Models can also be positioned using the Universal Transverse Mercator (UTM) coordinate systems where a location is referenced by its UTM zone and its northing and easting coordinate pair.

Chapter 4 – FROM POINT CLOUDS TO 3D MODELS

Effective modelling of 3D urban terrains by assimilation of sensed point cloud data is challenging and an active area of research. Typically the sensed objects and environment require models of complex geometry and topology which cannot be formulated in terms of common functional surfaces. Urban features such as tree canopies, overpasses, utility lines, elevated walkways, tunnels, and building interiors (Figure 4-1) require more general 3D representations that allow multi-valent height fields and higher genus topologies.



a) **b)** **c)**

d) **e)** **f)**

Figure 4-1: Examples of Urban Terrain Including: a) Building passages; b) Overpasses; c) Buildings and archways with complicated topology; d) Industrial complex of machinery, piping and utilities; and e)-f) Man-made structures with passageways, vehicles, and vegetation in Bonnland (left) and Norrköping (right) where data sets were contributed by SET-118 members.

Military applications require timely processing of large data sets. The processing framework should be able to:

i) Efficiently update the fitted geometry as new point cloud data is received;

ii) Naturally incorporate multi-scale versions of the fitted surfaces;

iii) Locally encode meta information; and

iv) Identify geometric features in a form suitable for database organization to enable fast queries for feature classification and matching such as those described in Section 4.3.

Additional applications may require real-time line of sight calculations of viewsheds for mission planning and autonomous navigation. In this chapter, requirements for representation are detailed and a brief review of the approach of both explicit and implicit methods is presented, along with their respective advantages and disadvantages. In Section 4.2, the pros and cons of cube and tetrahedral-based hierarchical representations are described in terms of the requirements expected of the representations. Section 4.3 provides a detailed description of classification of urban geometric objects. Section 4.4 illustrates a range of sample point cloud data sets which were contributed by NATO SET-118 member organizations and subsequently processed

using various assimilation algorithms to reconstruct approximations to the sensed surfaces. This is followed in Section 4.5 by three examples of model instantiations by processing 3D sensed data collections generated by other electro-optical sensors: 3D road extraction using stereo image pairs (Section 4.5.1), building reconstruction from multi-aspect high-resolution InSAR data (Section 4.5.2), and the use of photo interpretation for height estimation in terrain maps (Section 4.5.3).

4.1 CONCEPTS AND REQUIREMENTS

Representations for 3D urban terrains which hold the most useful potential for supporting military operations such as those outlined in previous sections, should have a framework which incorporates the following features:

a) Representations of the geometry and the topology of the terrain should be accurately assimilated directly from highly non-uniformly distributed point-cloud data, preferably without human intervention. Issues such as missing data should be addressed and reflected in the representations.

b) Representations used in fitting the point cloud should provide compression ratios of up to 100:1 using natural geometric primitives. Encoding/decoding of terrain should be efficient, scale linearly with the data, and use representations for efficient bitstream transmission in a client-server framework.

c) Computational effort of the processing algorithms and their derivative products, such as classification and line of sight, should scale linearly with the data. Moreover, algorithms should be amenable to parallelization.

d) Metrics appropriate for geometric modelling and applications should be naturally incorporated into the surface fitting algorithms. Rate-distortion estimates, i.e., trade-offs between compression and metric-accuracy should be provided for common classes of urban terrain. Local estimates of quality of fit should be available as local metadata.

e) Multi-resolution versions of the terrain representations should provide models with increasing geometric and topological detail as a user drills down locally.

f) Representations should incorporate dynamic learning or re-learning of 3D urban terrain to efficiently update models as supplementary or replacement data becomes available.

g) Representations should provide fast line-of-sight and viewsheds for complex 3D terrains.

h) Data structures used in processing should be hierarchical and naturally formulated to allow efficient look-up and search algorithms, both locally and through multiple resolutions.

We provide a several examples to illustrate the significance of each of these requirements. The first example (see Figure 4-2) is a 2.5D LIDAR data collection provided by US Army Corps of Engineers Topographical Engineering Center derived from a flyover of Baltimore, MD. Since the data has been processed to lie on a uniformly spaced grid with pixel values determining a height map, the surface may be considered as a univalent functional surface with values at a pixel position determined by the intensity.

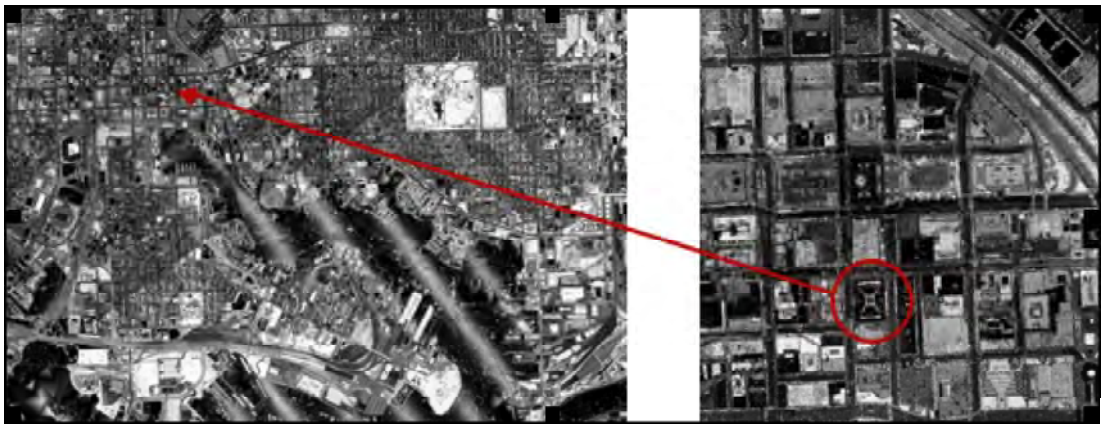


Figure 4-2: Requirement of Multi-Scale Localization. A rendering of a portion of a Baltimore LIDAR data set as an image (left) with a zoomed view in the vicinity of the domed building (right).

Although this point cloud does not have the full character of topology and geometry as might be collected by other more exhaustive LIDAR imaging, it still exhibits properties of large scale data collections over large coverages that require hierarchical processing methods in order to quickly provide views with varying resolutions at user-defined locations. One representative local drill-down around a domed building (circled in red in Figure 4-2) is provided in Figure 4-3a) by zooming in to a high resolution of the point cloud. In this image the colouring of points is according to height for assist in providing perspective for visualization purposes only. Due to the pose of the sensing platform the sides of the buildings are typically obscured and do not provide LIDAR returns, which results in missing data from the vertical walls in a full 3D model. If care is not taken in the surface reconstruction to respect data sparsity and identify regions with missing data, then over fitting of the data is likely to result, as illustrated in Figure 4-3e).

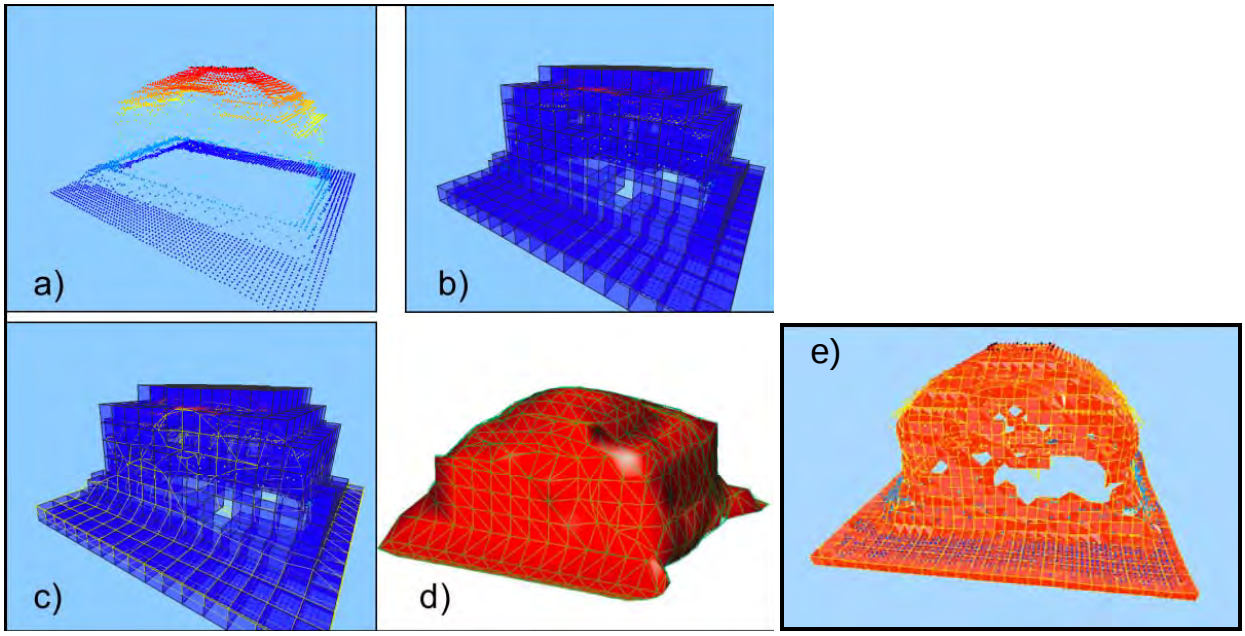


Figure 4-3: Domed Building Circled in Figure 4-2: a) Zoomed view of the point cloud in the vicinity of a domed building, indicating missing data obscured by airborne collection; b) Local oct-cubes; c) Curves indicating implicit distance field to point cloud interpolated from values on vertices of local oct-cubes; d) Fitted surface provided by implicit level surface for distance field; and e) Effect of over-resolving the oct-cubes and selecting a level set parameter too small for the data resolution which results in overfitting the data.

As we have remarked, the raw Baltimore LIDAR data collection was processed in order to provide an estimated height value for each pixel coordinate. For many applications the data may be required to be assimilated with updates of redundantly sampled regions with the hope of providing better estimation (see Figure 4-4).

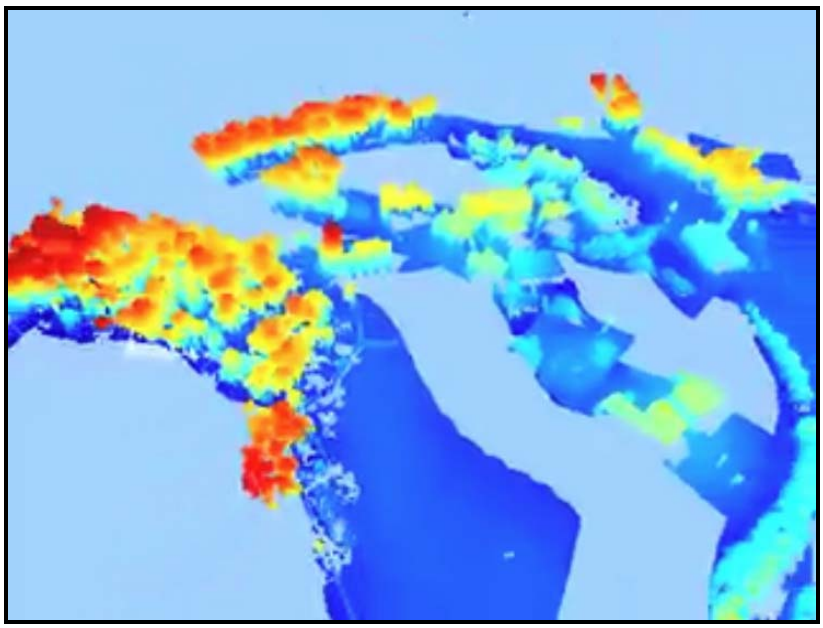


Figure 4-4: Eglin[1] LIDAR Flyover with Overlapping Swaths.

Other applications, such as autonomous navigation of unknown or modified terrain, require on-line adaptive assimilation of the terrain. Figure 4-5 illustrates on-line learning of a 2.5D functional surface in mid-acquisition of the overflight LIDAR data pictured in Figure 4-4. Statistical estimation of surface heights should be updated as the overlapping regions are acquired.

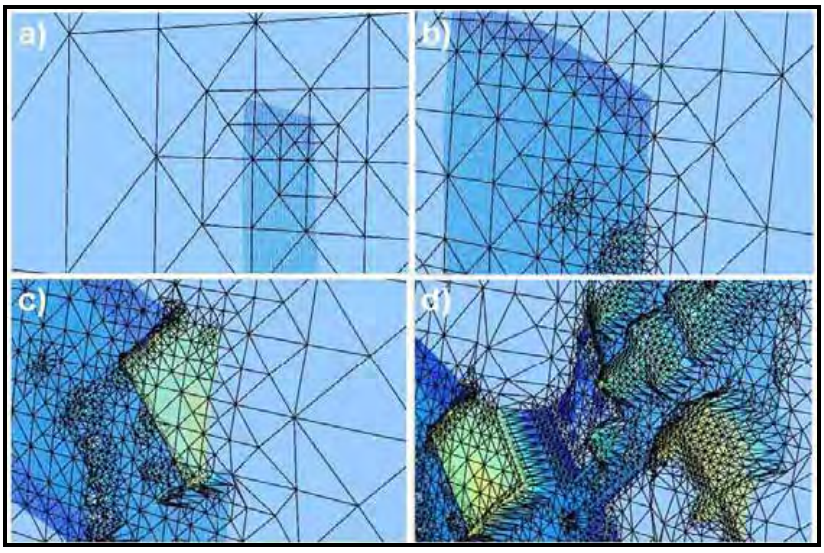
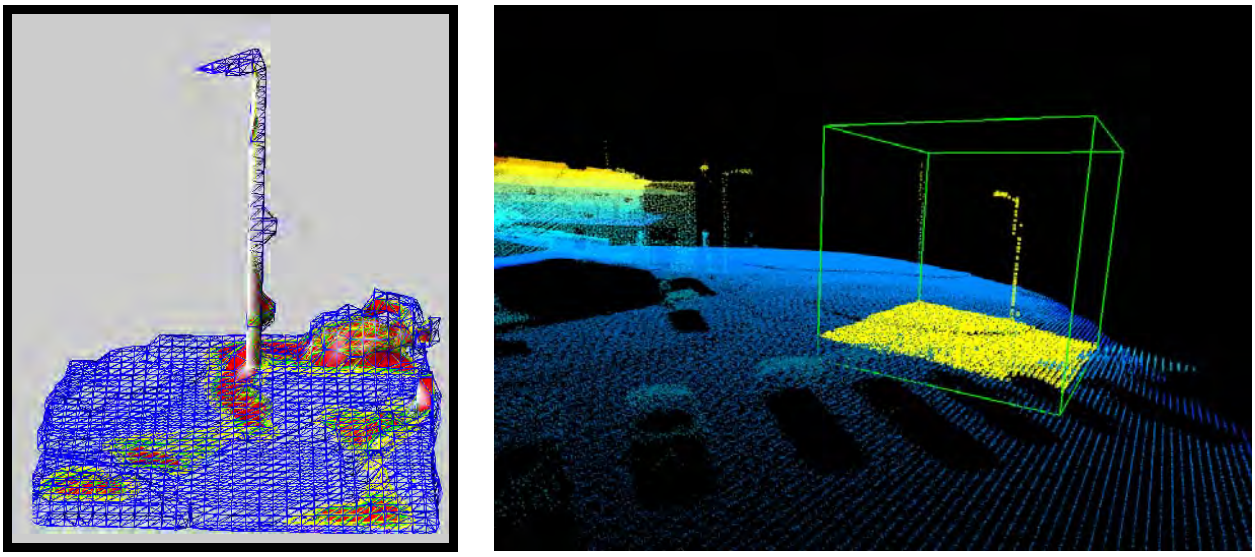


Figure 4-5: Image Sequence of Adaptive Sub-Division Builds a) – d) for On-Line Surface Reconstruction from Eglin[1] LIDAR Overflight Data.

[1] AFRL/MNG VEAA Data Set #1 of Eglin AFB, with permission.

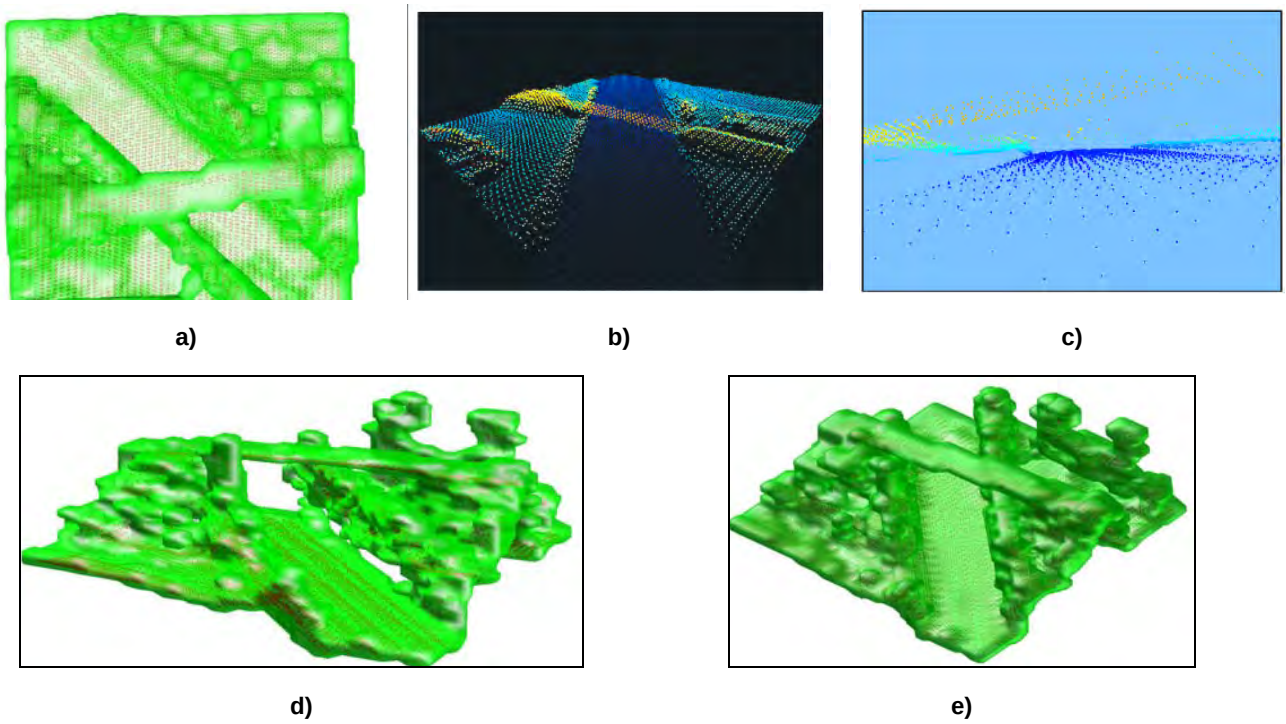
Although the reconstructed geometry in Figure 4-5 provides a coarse approximation of the urban terrain and is satisfactory for some applications, finer resolution is available and may be finely modelled as shown in Figure 4-6. This is an example of simple topology, but non-functional 3D univalent surface.



a) b)

Figure 4-6: a) Close Up of Parking Lot Light Pole from Eglin LIDAR Overflight Data. Missing data can also be observed; and b) Fine Scale Surface Reconstruction Using Level Sets is Used to Fit the Point Cloud Data and is Able to Resolve Additional Detail.

To provide an example involving somewhat more complicated topology, we use publically available overflight LIDAR data prepared by the International Society for Photogrammetry and Remote Sensing Commission III – Photogrammetric Computer Vision Working Group [100]. The data shown in Figure 4-7a) was obtained by modifying these data so as to extend the roadway underneath the overpass to fill in the occluded portion. The resulting data has similar sensor characteristics, but is a genus 1 topological object with more complicated geometry due to the vegetation near the intersection.

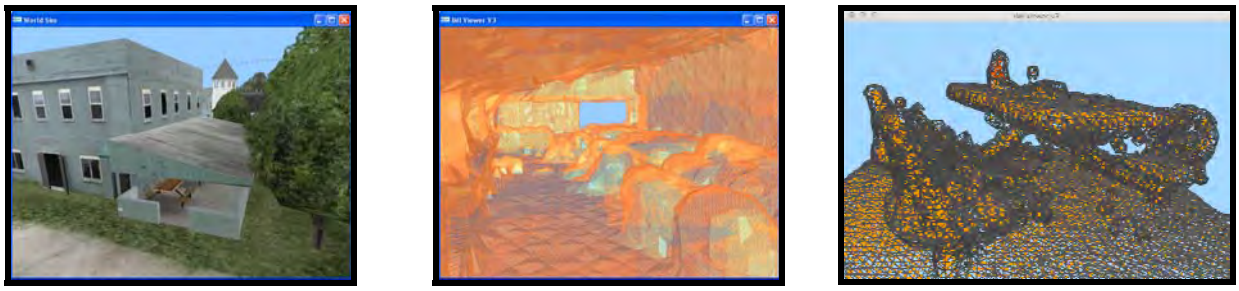


a) b) c)

d) e)

Figure 4-7: a) Original LIDAR Data for Underpass; b) Close Up View from a Different Perspective than the Imaging Sensor, in particular, Looking Under the Underpass; c) Top View of Intersection; d) Reconstructed Surface of Underpass and Neighbouring Vegetation and Trees; and e) Alternate View Showing Overfitting of the Sparse Data Along the Roadway Borders.

A more complicated topology and geometry is shown in Figure 4-8 in which the reconstructed surfaces of synthetic flash LIDAR are rendered. This data was generated by the South Carolina LIDAR simulator using a sequence of ten flash images.

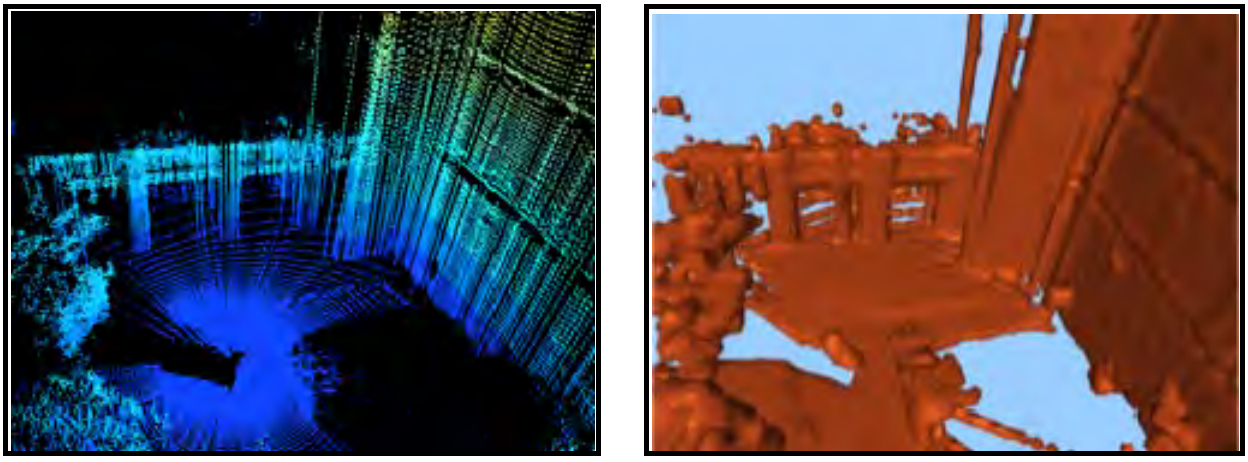


a) b) c)

Figure 4-8: a) Simulated Flash LIDAR Data of Patio Area – Ft. Benning; b) Coarse Learned Surface from Point Clouds Resulting from 10 Registered 128 x 128 Flash LIDAR Images; and c) Finer Resolution of Isosurface of Point Cloud of Table Constructed Using Unsigned Distance Field from Point Cloud in b).

The final example of this section illustrates LIDAR data generated in more typical situations by ground-based platforms. In this case a line scanning Hokuyo LIDAR is mounted on a vertically oriented bar which is rotated about the axis of the bar to produce a relatively dense point cloud of a court yard near Hancock Hall on the campus of Virginia Polytechnic Institute and State University in Blacksburg, Virginia, USA. This data exhibits many of the problems described earlier in this section. In particular, the topology and

geometries are complex and difficult to assimilate into a high quality, reliable surface fit (Figure 4-9). Difficulties arise in large part due to ambiguities of missing or occluded data since information of the acquisition process, such as time-stamped geolocation of the sensor, is not utilized.



a) b)

Figure 4-9: a) Raw LIDAR Data Generated by Hokuyo Line Scanning Sensors Which Were Rotated at a Fixed Position about a Vertically Aligned Axis; and b) Surface Reconstruction Using a Level Set of an Unsigned Distance Captures Some Important Topology, Such as Building Faces and Columns, and Demonstrates Complexities Arising Both from the Sensed Surfaces as well as Processing Artefacts.

4.2 MODEL REPRESENTATIONS

Surface modelling of scattered, highly heterogeneously distributed point clouds presents many challenges. Previously the approaches to fitting geometric structure to this type of data can be classified as either *explicit* or *implicit* methods. The selection of which method is used typically depends upon the end applications and the properties of the point cloud. In this section a brief description of each methodology is presented, followed by examples and discussion of the pros and cons. Ideally, models for terrain surfaces should seamlessly incorporate both explicit and implicit representations into hybrid models, exploiting the advantages of both of these viewpoints while minimizing their drawbacks.

4.2.1 Explicit Surface Representations

An explicit representation usually takes the form of a parametric representation of the surface. For 2.5D surfaces the parameterization is simply of the form $z = f(x,y)$ where z is height at position (x,y). In terrain applications, however, more intricate parameterizations are required in order to model urban terrains with overhanging structures, open building entrances and interiors, vegetation, and similar features. These more general parameterizations take the form $(x(u,v),y(u,v),z(u,v))$ where (u,v) ranges over the parameter space. A simple example of a non-functional parameterized surface would be that of a sphere of radius ρ:

$$\left(\rho\sin(u)\cos(v),\rho\sin(u)\sin(v),\rho\cos(u)\right), \quad \text{where } 0 \le u \le \pi, 0 \le v \le 2\pi\,.$$

The departure here from usual terrain models (DTED, DEM) is that explicit representations are taken to be locally the graph of a function in *some* coordinate system which is not assumed to be the standard (x,y,z) Euclidean system. This draws on methods and concepts from computer aided geometric design (CAGD) which involves invaluable, but computationally expensive, processing tools in industrial design [47].

Various research journals such as Elsevier's *Computer Aided Geometric Design* are dedicated particularly to the mathematical and algorithmic issues of this subject. These local coordinate systems are typically estimated by the local structure of the point cloud by determining the local density and variation of the point cloud. The most natural and efficient approach to determine local structure is to *divide and conquer* using hierarchical coarse-to-fine partitioning of the data into occupancy cells. This also has the advantage of representing the data at different scales and allows for local statistics to be easily computed and accessed. Two methods of partitioning into occupancy cells are described later in this sub-section: octrees of cubes and bin-trees of tetrahedra. Various criteria can be used to determine when there has been sufficient partitioning to estimate a local orientation (or projection) for which the surface can be represented locally as a functional surface. A common popular approach to determine the orientation is to use the covariance matrix and corresponding principal component analysis associated with that segment of the point cloud. Once an orientation is estimated for a local segment, a similarity transformation is applied to rotate back to Euclidean coordinates (x,y) so that standard surface fitting methods, using splines for example, can be used. The resulting surface fit is then rotated back to the original local coordinate system as a local parameterized surface fit of the point cloud. These local versions are all stitched together into a global parameterized representation of the desired form $(x(u,v),y(u,v),z(u,v))$.

Next we describe two methods to quickly analyze and cluster a 3D point cloud into occupancy cells.

4.2.1.1 Octree Representation

An octree is a hierarchical tree data structure. It consists of a cubic root node that is recursively sub-divided into 8 child nodes of increasing resolution at each level of the tree. It is a structure similar to a 3D grid but with the added advantage of being multi-resolution. Figure 4-10 shows an example of an octree with corresponding tree representation, where levels are represented with different shades of blue.

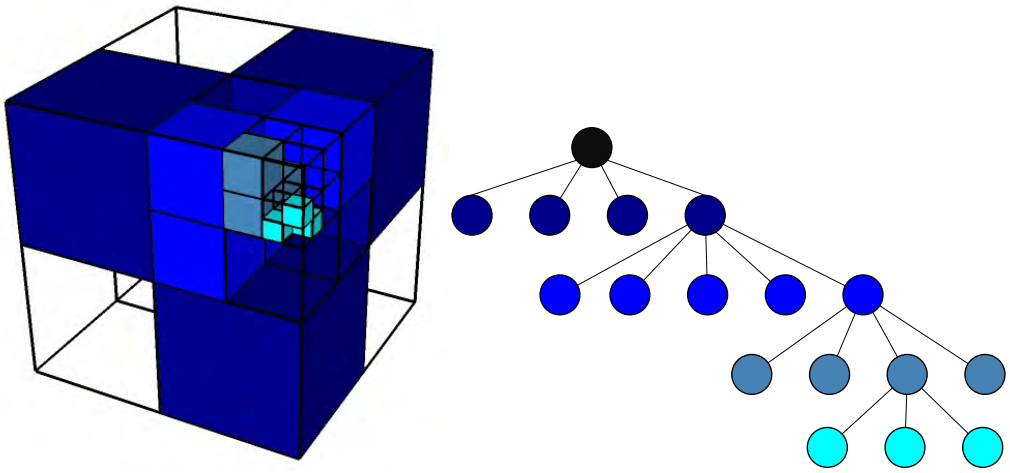


Figure 4-10: Octree with the Corresponding Tree Representation.

Octrees are often used to optimize and accelerate the display of 3D point clouds. In this case, a root node whose size is sufficient to contain the point cloud is recursively sub-divided into child nodes until the number of points contained by each leaf node is below a threshold. This type of representation eases operations such as ray tracing.

Octrees may also be used as dynamic structures to store the occupancy state of the environment as it is measured [48]. For this application, each node is assigned the occupancy probability of the volume it encompasses. As new data is acquired, new nodes are created on demand and existing nodes see their

occupancy probability updated. The maximal resolution of the octree (size of terminal nodes) is selected in accordance to the required model precision.

One advantage of octrees is that nodes are only created for parts of the environment where information is available, incidentally lowering memory requirements. For comparison, when using a 3D grid, one has to generate all the cells even if some of them are unused. In addition, octrees may be compressed easily to lower their resolution and memory footprint. On the other hand, octrees have a more complex structure and navigation between nodes might prove challenging for some applications.

4.2.1.2 Tetrahedral Occupancy Cells

Another common approach is to instead use binary sub-division and geometric simplexes to partition the data. In 3D the simplexes are tetrahedra, geometric primitive polyhedra, consisting of 4 non-coplanar vertices and 4 triangular faces (Figure 4-11). Tetrahedra are natural objects for 3D geometry. They are defined by their faces, triangles which current graphics card hardware use as a basis for visualization and acceleration of numerical computations. The ideas for tetrahedral occupancy cells are similar to the cube-based hierarchical structures, except the initial cube is first partitioned into 6 tetrahedra which will be the base nodes (see Figure 4-11). Each of these tetrahedra has the main diagonal of the cube as its longest edge. As described below, further sub-division of each node (i.e., tetrahedra) is performed by splitting the longest edge of the tetrahedral which results in two children tetrahedra. The octree structure from hierarchical cubes is then replaced by a binary tree structure of tetrahedra. The surface boundary of the union of the occupied tetrahedra at any stage of sub-division can be thought of as a rough, 0-th order approximation of the surface of the point cloud. For higher order approximation, the different tetrahedra must be organized and linked together in case some sort of stitching is necessary.

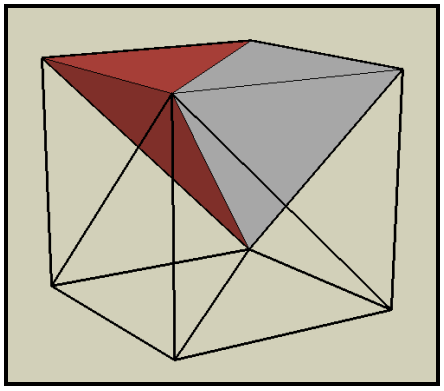


Figure 4-11: Cube Partitioned into Six Tetrahedral Simplices.

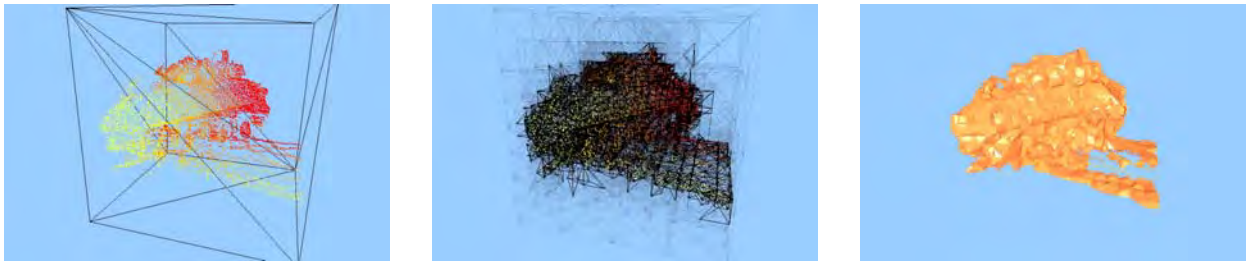
This will require additional structure since trees do not guarantee that nodes that are spatially close are necessarily close in traversing the tree. Moreover, when using adaptive sub-division schemes, discontinuities along the boundary of the nodes in the form of hanging vertices are introduced and must be handled. For example, when using a marching cubes scheme on a multi-resolution octree, one must take special care to ensure that the surface produced is continuous between neighbouring cubes that potentially differ greatly in size or by their appearance in the depth of the octree. Note that using this scheme on all six tetrahedra derived from the cube, the levels of an octree sub-division emerge from every three uniform sub-divisions of the tetrahedra trees. That is, if both the octree scheme and the tetrahedra scheme start from the same cube, the same set of vertices will be used by the leaf nodes in both schemes.

At this point the tree of tetrahedra has a frontier, or set of leaf nodes, which can still fall victim to the same discontinuity problems mentioned previously. To prevent this, a conforming step is performed by sub-

dividing additional tetrahedra while building the tree. To sub-divide a single tetrahedron the longest edge is selected and a new vertex is added at its midpoint. This new vertex is then connected to the two vertices not on the longest edge giving two children tetrahedra. As the new vertex generally introduces a discontinuity, i.e., a node in the middle of an edge of a neighbouring tetrahedron which shared the original edge, we additionally sub-divide all other tetrahedra that share this split edge. The additional sub-divisions result in a smooth transition from the finer regions of the frontier to the coarser regions. This allows algorithms like *marching tetrahedra* to be used without concern for special cases along boundaries. Also, since this scheme is simply a collection of binary trees, connections between nodes to allow for fast traversal of nodes at any level in the tree can be added, including links for traversal of the varying levels of the frontier, or leaf, nodes. Thus nodes that are spatially close are ensured to be close in the tree structure.

With the additional links between the nodes of the tree, locally stored statistics in the structure of each node can easily be distributed for analysis to neighbouring nodes without global searches. Moreover, if new data arrives, then the local information can be easily updated, and splitting criteria reapplied for possible finer resolution based upon the new data. Particular statistics of interest would be the barycentre and covariance matrix of that portion of the point cloud within the cell, since they indicate spatially and structurally similar data. The statistics or data stored in each node are application dependent and do not necessarily have to be related to the splitting criteria. With this information the splitting and stopping criteria can be modified to create trees that provide varying types and amounts of information. For example, if only occupancy information is of interest, a simple threshold on the number of points per node may be used. If instead the space should be sub-divided until each node contains roughly planes or lines, then PCA can be performed on a node individually or a collection of neighbours. Also, local curvature information can be computed and easily compared to neighbouring nodes or even coarser or finer nodes.

As mentioned earlier in this section, multi-resolution tetrahedra or cubes could be used as the support for local piecewise polynomial surface fits of the point cloud by using methods from Computer Aided Geometric Design (CAGD). This is in fact being implemented and tested by Texas A&M University and Rice University personnel of the ARO MURI in the Wavelet Streaming Surface Reconstruction (WSSR) algorithm [102]. For the purposes of this report we focus on how to integrate techniques from the next section on implicit surfaces for hybrid versions of point cloud assimilation.



a) b) c)

Figure 4-12: a) Point Cloud of Wright-Patterson ASC Flash LIDAR Data of DRDC Valcartier Armoured Personnel Carrier Rendered by Range Colouring; b) Partitioned into Occupied and Force-Split Tetrahedra; and c) Rough Surface Representation Formed by Surfaces of Leaf Node Tetrahedra.

4.2.1.3 Comparison of Oct-Cubes and Tetrahedral Bin-Trees

Before proceeding to the discussion of implicit surfaces, a brief comparison is provided for octree-cube and binary tree – tetrahedron occupancy structures. As is commonly the case, there is a play off in each case between progressively storing information into structures for quickly determining neighbours (memory drain) and re-computation of local properties to traverse the tree structures.

Oct-cubes use octrees for their structures, but can utilize very simple indexing for parent/child and spatial neighbours, which results in minimal overhead in terms of memory for pointers and structures and faster implementation. On the negative side, oct-cubes are non-conforming and are less reliable when integrated in implicit surface algorithms which produce level sets and surfaces.

Bin-trees used for tetrahedra are simple to encode, but the gain in speed for processing operations require additional overhead in memory (in addition to pointers for parents and children, pointers also used for face neighbours and vertex neighbours for leaf nodes). Implementation effort is also substantially more involved. However, multi-level tetrahedra are more efficient for graphics hardware computations and for more efficiently capturing a point cloud in 3D. As such the local statistics for tetrahedra should on average be more reliable.

4.2.2 Implicit Surface Representations

In an implicit representation in 3D, the surface is described as the set of points which are related implicitly through their relationship in satisfying an equation of the form:

$$F(x,y,z)=0\,.$$

Thus, the surface can be viewed as the boundary of the level set of a function F defined on a domain in R^3. There are many functions F which yield the same surface. One particularly useful choice is the function $\phi = F_d$ which gives the *signed* distance of the point (x,y,z) to the surface. The sign will be positive if the point is exterior to the surface and negative otherwise. This choice is particularly useful for improved approximation of surface fits of finely distributed point clouds, and in line of sight calculations and autonomous navigation.

A point (x,y,z) is in the *line of sight* from (x_0,y_0,z_0) if the distance function ϕ does not change sign on the line segment connecting these two points. In autonomous navigation, a fly zone can correspond to the buffer region $\phi \geq \delta$ where $\delta > 0$ is some prescribed tolerance based upon the vehicle's control and operational characteristics.

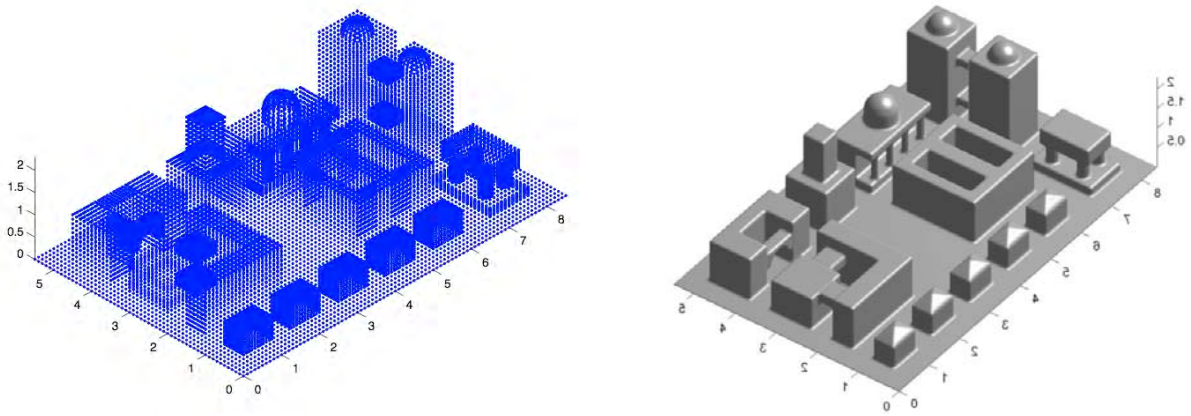
The most essential step in an implicit representation of a surface is a separation and classification of the whole space domain into two regions: interior and exterior. The boundary between these two regions is the surface. For representation, compression and reconstruction, one only needs to know and store information near the boundary which fits the point cloud. Of course the interior region can and will generally, be multiply connected, e.g., many buildings separate from each other, all defined simply through one scalar function.

Other possibilities than the distance function are possible for the function F to provide an implicit fit to a point cloud or which would have a prescribed set as a surface interior. Variational problems can be formulated in which a sensed point cloud acts as a constrained attractor to construct a surface fit to unorganized data. ϕ can be thought to act as a potential function for the point cloud. The objective is to find a local minimiser of an energy functional that behaves like a minimal surface or an elastic membrane variably attracted to the data set. The potential force and the surface tension are balanced in a variational equation where the solution ϕ must satisfy a corresponding Euler-Lagrange equation [49] of the form:

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| \left[\nabla d(x) \cdot \frac{\nabla \phi}{|\nabla \phi|} + \frac{1}{p} d(x) \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} \right].$$

Here, the first term in brackets corresponds to the attraction by the distance field d and the second term corresponds to a minimal surface regularization weighted by the distance function d to the point cloud. Recall that this last term consisting of the divergence of the unit normal of ϕ is just the curvature.

Implicit models are a natural choice for the modelling of 3D urban terrain because of their flexibility and robustness in dealing with complicated topology. Every object can be regarded as solid or volumetric. Hence one can mark regions that are inside structures or underground as interior. The whole complementary region is marked as exterior. Line-of-sight information can be used to distinguish interior and exterior regions. Non-genus-0 topology poses no extra difficulty. Other information, such as connectedness, can also be incorporated to compute the correct topology [50]. Usually the interior region is composed of multiple disconnected components which are associated with the interior of different objects. Many operations and manipulations become very simple using implicit methods, for example, Boolean operations, finding intersections, visibility and path planning [51],[52] – see Figure 4-13.



a) b)

Figure 4-13: a) Synthetic Point Cloud of a Simulated Urban Terrain with Moderately Complex Topology; and b) Level Set Reconstruction of Sensed Surface Using Variational Surface Fitting.

The advantages of implicit representation in image processing are now well demonstrated. They lead to state-of-the-art denoising and deblurring algorithms [53]. There are several compelling reasons to utilize implicit representations in terrain processing. The first is captured in the remarks above on line of sight and navigation. The second has to do with efficient encoding. Some surfaces or portions of surfaces are much easier to describe in implicit form. This is the case for most man-made structures. Consider for example the cylindrical surface $x^2+y^2 = r^2$, $0 \leq z \leq L$, which could correspond to a portion of a telephone pole or light pole. Other man-made structures often have similar simple representations. In implicit form this surface can be described by very few bits. Its description by explicit methods (for example approximation of its level curves by piecewise linear functions) would be much more costly. Another point is that implicit representations more easily describe the topology of the surface and connectivity regions.

4.2.3 Hybrid Methods

There is no simple answer to the question as to which of explicit or implicit methods are superior in general. Hierarchical occupancy cells and the analysis stages of explicit methods are very fast and scale linearly with the data. Implicit methods have advantages of fast line of sight, high quality denoising and deblurring, but typically require dense data and computationally expensive processing. Thus, an ideal framework would be an integrated system which has the benefits of both: the potential to utilize the speed and localized analysis of explicit methods while exploiting level set properties of implicit methods. This is the main focus of the ARO funded Multi-University Research Initiative entitled “Dynamic Modeling of 3D Urban Terrain” [49],[53],[101],[102]. One general resource for rural and urban terrains is maintained at the Virtual Terrain website [103]. There are detailed links which provide background for usage of terrain and cultural models; ontologies; data sources and formats; methodologies for data acquisition and fusion, classification and geometry building; and references for commercial software.

4.3 CLASSIFICATION OF TERRAIN OBJECTS

Many applications, such as simulations, require explicit information about the types of the individual objects in the scene. Clearly, the sensor readings themselves do not carry any such high-level information but it has to be added afterwards in a *classification* step, either manually or – as will be the topic here – through automatic data analysis.

From a modeling point of view, the classification functionality is an important component, as it allows for creating application-specific model instantiations, in which each classified segment is replaced with a 3D model component. This means that whence we have classified a particular segment into a certain class, say, a tree, we can replace the original data points stemming from the tree with a 3D model of a tree, chosen so as to be relevant in the application at hand.

4.3.1 Bare-Earth Extraction

The ability to estimate the height of the ground, or bare-earth, surface is typically of fundamental importance for creating an explicit 3D model of a particular scene. Such a ground model, often referred to as a *Digital Terrain Model*, or DTM, is the base on which other model components – building, trees, cars, lamp posts, trash bins, etc. – are placed in order to form the final 3D model. From a signal processing point of view, having access to a DTM facilitates the scene analysis significantly as it limits the space in which to look for, and expect, certain kinds of objects. In addition, the DTM in itself is often used in many geotechnical applications, e.g., to assess risks of flooding, landslides, etc. An example of a DTM created with a flexible surface matching technique is presented in [55]. In principle, all existing DTM estimation techniques make use of one fundamental assumption: *ground data points have low elevation values*. Hence, the first step in ground modeling is to establish a set of very probable ground points based on the expected nature of the ground in the particular region, see Figure 4-14. Then, based on an initial set of ground points, the ground level at other locations in the region can be estimated through either successively classifying more points as ground and/or interpolating (or extrapolating) the ground level in regions void of classified ground points. Overviews of different DTM estimation techniques can be found in [56],[57].

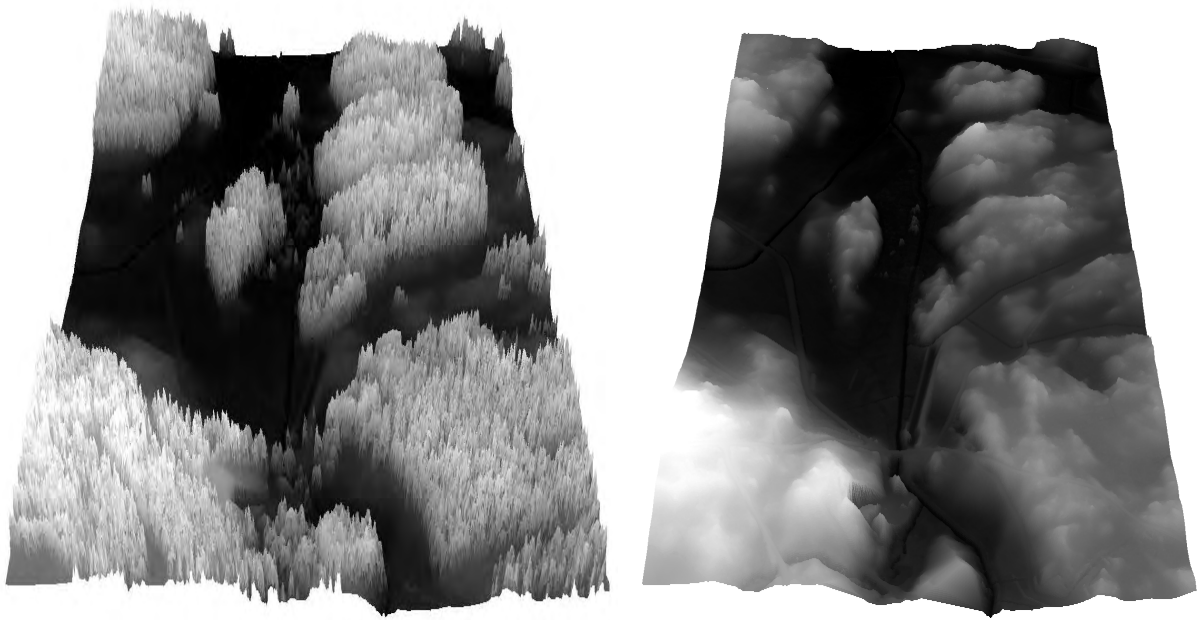


Figure 4-14: (Left) Elevation Data (Digital Surface Model); and (Right) Ground Surface (Digital Terrain Model).

4.3.2 Classification of Above-Ground Objects

We now turn to the task of classifying the objects that are located upon the ground, here referred to as *above-ground objects*. Analogous to the problem of distinguishing between ground and non-ground data points, we are seeking a set of *characteristics* or *features* that can be used to separate the classes. In order for the classification to succeed, the features for a particular class should be different from other classes and not vary too much within the class. Put differently, good features should display *high inter-class variation* and *low intra-class variation*. What features can or should be used depend on a lot of circumstances, e.g., what types of objects are considered, what accuracy/detail level is required and what sensor data are available. In the literature, many different features have been proposed and it is beyond the scope of this report to go into details concerning specifics of their nature [57]. Nevertheless, the features are designed so as to capture certain relevant underlying physical properties of the objects, and the variation among those properties is more tractable than that among the plethora of suggested specific features and algorithms.

Here we will focus on two especially common ingredients in 3D urban models – buildings and trees – and give some examples of physical properties that are exploited for classifying these objects. In practice, several properties are often combined to achieve robustness to changes in appearance of objects within the same class. With continuously increasing resolution and accuracy of airborne range/3D measurement systems, it is likely that the future will behold more efforts on automatic recognition of smaller and finer objects, such as fire posts, wells, walls, antennas, pedestrians, vehicles, etc.

4.3.2.1 Buildings

- **Relatively Large –** Many (most) buildings are large compared to other objects expected to appear in the scene. This means that a classifier using only a simple size feature would probably do a good job on classifying the major buildings in the scene.

- **Regularly Shaped Footprints –** Most buildings have a regular shape, in the sense that they consist of (linear) wall segments that are oriented in a structured fashion. However, not all regular buildings appear as regular structures in the data. This can be due to occlusion effects (e.g., trees hanging over the building, thus altering its apparent shape) or insufficient data density (causing walls to appear as jagged). Obviously, the smaller the building and the heavier the occlusion, the more difficult it is to use the edge information for robust classification. In fact, at some point, it becomes difficult to distinguish the footprint of a building from that of tree.

- **Smooth Roof Segments –** We expect the geometry of roofs to be smoother than that of vegetation. In fact, building roofs can often be successfully described as with a set of geometrical primitives, e.g., flat surfaces. Obviously other roof shapes exist, such as conical or dome-shaped, that can be treated following the same line of principle. From a signal processing point of view, the smoothness/flatness can be quantified by computing the residual of fitting a plane to a set of points, by estimating local height variations, by local statistics (e.g., PCA, covariance), etc.

- **Homogeneous Objects –** A LIDAR sensor can make use of the fact that buildings are typically homogeneous objects. Among other things, this implies that it is impenetrable for the laser light and hence that we do not expect to find hits from the ground within the building, the exceptions being skylights or courtyards. This property can also be used for segmentation of laser data into homogeneous regions (see Figure 4-15). Note the problems that occur at the border of the buildings, as the roof typically extrudes from the wall. Nevertheless, the interior of the building is typically void of ground hits.

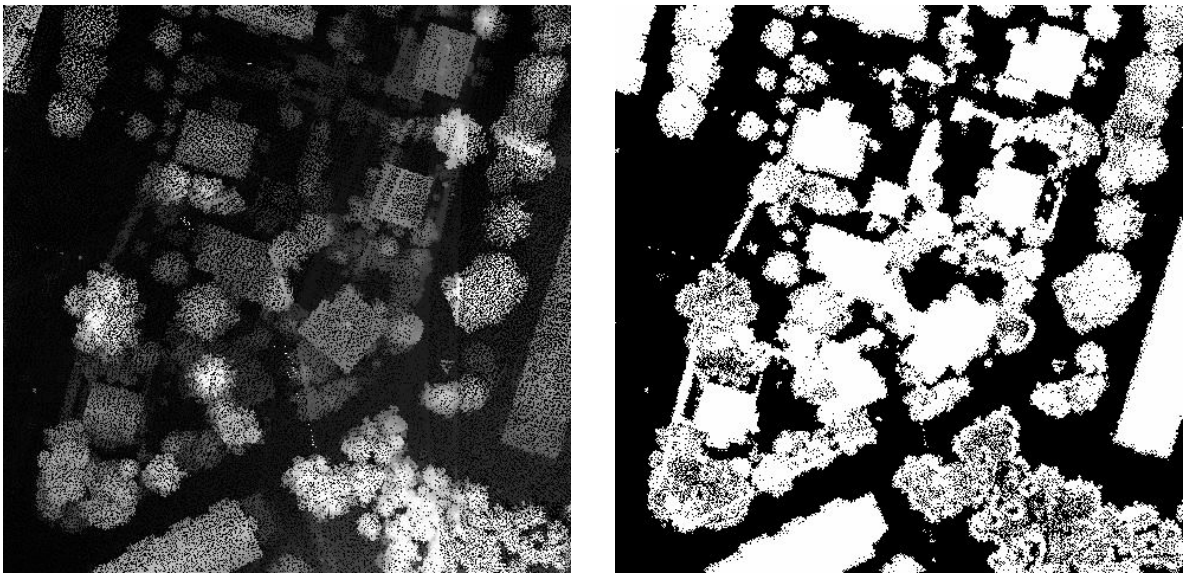


Figure 4-15: (Left) Elevation Image; and (Right) Black Pixels Indicate where Ground Hits have Occurred. Note the difference between the buildings and the trees in this respect.

4.3.2.2 Trees

- **Inhomogeneous Objects (many small reflecting surfaces at different depths)** – Trees are often only partially obscuring the ground beneath them, and as a consequence there are often laser hits from both the tree canopy and the ground beneath, see Figure 4-15. In addition, the vegetation areas display seemingly random height variations which can be quantified through Principal Components Analysis, a height variance filter, etc. As mentioned earlier in this report, some modern LIDAR systems provide the possibility of collecting the entire received waveform. This waveform is the superposition of all the reflections back to the detector within the laser beam footprint. Since trees consist of many small reflecting surfaces at different distances from the laser system, pulses reflected off trees generally broaden considerably. This could prove very useful for distinguishing between trees and hard surfaces [57].

- **Local Height Maxima** – If sampled densely enough, trees typically exhibit the shape of local height hills, especially after low-pass filtering of the height data. Finding such hills makes it possible to detect single trees. In passive imagery, the height hills may cast shadows on their neighborhood that can be detected to support the detection of the height hills in the spatial domain.

- **Spectral Characteristics Often Different from Those of Buildings** – Since vegetation and building roofs typically consists of different materials, their spectral appearance is expected to differ. For example, trees with leaves often display a prominent green color whereas buildings roofs do not. The difference is often obvious within the RGB domain, but may become emphasized if more spectral bands are used, e.g., NIR (near infra-red). Hyperspectral data have become commercially available, which further increases the class-separating potential of the spectral data. Obviously, the success of classification based on spectral properties hinges on managing spectral variations caused by seasonal changes (leaves vs. no leaves, snow, etc.), shadows, sun irradiation, angular-dependent reflectance functions, sensor noise, etc. Ideally, if one could compensate for all such effects, the result would be a spectral response curve that could be matched to a spectral library. In practice, however, the most efficient way of classifying spectral information data is often to train a classifier on a representative portion of the data at hand and have it learn the spectral characteristics of the region under study.

4.3.3 Classifier Examples

At FOI the problem of automatic detection of buildings and trees has been studied for the last decade. As a result, three individual classifiers have been developed that are all based on different techniques for segmentation and classification, each with its own merits and disadvantages. One technique [58],[59] segments the data using multiple echo information and classifies the resulting segments using shape and height variation features, including Hough transform-based features for recognizing regular building footprints. Another technique [59] aims at first segmenting the data into regions containing no interior ground hits, using the fact that laser beams often penetrate canopies and hits the ground below, while being stopped effectively by most roof materials. Large segments that contain flat surfaces (i.e., roofs) are then marked as potential buildings. A third technique [60] performs a connected component (region growing) segmentation of the data and relies on Principal Component Analysis to distinguish planar segments (roofs) from other types of objects.

In order to get a more detailed understanding of the performance, a limited region was selected and analyzed thoroughly, object by object. Table 4-1 contains an example of a summary of such an analysis for one particular building detection result (basically all techniques displayed similar results). The purpose of this misclassification analysis was to learn more about the cause of the errors produced by the algorithm with respect to ground truth (here, a city cadastral map). After all, there are many elements that affect the overall performance – from data acquisition, platform positioning and alignment of swaths to segmentation and classification. Among other things, the results indicated that many of the undetected buildings were (at least partially) occluded, and hence that those objects had not been accurately measured with the LIDAR system. Furthermore, other buildings had only very few hits on the roof, as of consequence of roof material that caused data dropouts due to very unfavorable reflectance properties, while other buildings were simply too small so that no structures could be detected reliably from the sparse returns. A large portion of the erroneously detected objects had a regular geometry that made them look like buildings, only they were not actually buildings according to the cadastral map, but containers, tents, etc. In short, this analysis indicates that the dataset itself is a main limiting factor here.

Table 4-1: Detailed Analysis of Erroneously Classified Objects, as Compared to Ground Truth.

False Negatives	#	%
Number of Undetected Buildings	**40**	
Occluded building	14	35.0%
Very few LIDAR samples	8	20.0%
Small/low building	10	25.0%
Building not present in LIDAR data	3	7.5%
Only part of building present within image	5	12.5%
False Positives		
Number of Erroneously Detected Buildings	**50**	
Natural objects (trees, bushes, water, etc.)	20	40.0%
Human-made, regular objects (vehicles, containers, etc.)	30	60.0%

4.3.4 Data and Information Fusion

As mentioned above, the quality of the data itself, for example in terms of sampling density, range noise, occlusion, etc., is an important limiting factor. Thus, what would really increase overall performance is richer data that carry more information about the objects of the urban terrain. As richer sensor data is starting to appear frequently on the market, e.g., LIDAR waveform, simultaneous spatial and spectral measurements,

geolocation, fusion of terrestrial and airborne data, structure-from-motion data, hyperspectral data, SAR imagery, etc., the future will probably hold an increased amount of integration of different kinds of data, at the expense of reduced development efforts concerning "LIDAR only"-based mapping techniques.

In practice, classification of objects in urban terrain based on geometrical properties generally enables quite good performance, provided that the original data are dense and accurate enough to capture fine details of the sought-after objects. However, there will most often, if not always, be flaws in the data due to occlusion, range noise, poor resolution, dropouts, etc., that will lead to erroneous decisions based on the particular data. In order to improve the chances of making correct classifications, various information cues can be used and combined, referred to as *fusion*. Having access to raw data enables us to design classifiers that combine data through *data*, or *signal-level, fusion*, e.g., LIDAR data and aerial imagery that allows to exploit both spatial and spectral features in the classification process.

With a rapidly increasing number of commercial software and services for automatic analysis of sensor data, it is likely that the end-users will not always have the sensor data at their disposal, let alone have enough knowledge of how to design robust features and classifiers. In addition, the users may have complementary information such as already existing maps, old classification results, etc. In such cases, the user would benefit from fusing data on a higher level, e.g., through majority voting among several classifiers, which is an example of *decision-level fusion*. In Table 4-2 the results of an information fusion (through majority voting) are presented.

Table 4-2: The Table Shows the Results of Information (Decision-Level) Fusion of Individual Results Through Majority Voting. Intentionally, none of the techniques were optimized for the current dataset. This is most notable for the techniques referred to as "2003" (high FAR) and "2006" (low P_d).

Voting Methods \ Data Test Sets	2003	2006	2007	Fusion
P_d	0.830	0.701	0.811	0.801
Number of detected known objects [km^{-2}]	205	173	199	197
FAR [km^{-2}]	168	44	65	54

4.3.5 Building Reconstruction from Airborne LIDAR Data

Many (most) 3D modeling or visualization software packages require that 3D objects are represented in a vector format, such as CityGML, kml, 3ds, OpenFlight, etc., rather than by a collection of 3D points or an implicit model.

Basically, the modeling can be performed at an arbitrary level of detail, from very simple and coarse "shoe box" models to highly detailed 3D models capturing every significant geometrical feature of the building. If the sampling density is high compared to the geometry variations within the building, which is often the case with modern high-resolution sensor systems, it is often possible to apply automatic algorithms for creating a realistic 3D model for each individual building.

The general approach to the polygon modelling problem is to fit a set of predetermined geometrical primitives to the data and by doing so determining the primitives and the topology that best describe the building. This is how most existing building modelling techniques work, although they may differ with respect to the specific techniques used to determine the primitives and the post-processing applied. By far the most common geometrical primitives are planar surfaces and line segments. See Figure 4-16 for an illustration of some key steps in the process of creating a 3D model from elevation data.

One can distinguish between two main classes of building modelling approaches: *bottom-up* and *top-down*. Techniques in the bottom-up class start by gradually extracting local primitives from the data that are eventually combined into a 3D model. Such techniques are inherently capable of modelling any

building that can be described with the set of primitives. In practice however, the performance is often limited by the quality of the data – sampling density with respect to the spatial variation of the building roof, noise, occlusion, etc. In addition, a good 3D modelling result obviously requires that the geometrical primitives (often planar surfaces, ridges, break lines, etc.) match the real characteristics of the building. Trying to model a cylindrical building using only piecewise linear wall segments is clearly not ideal but could still result in a decent-looking and useful model, depending on the requirements of the application at hand.

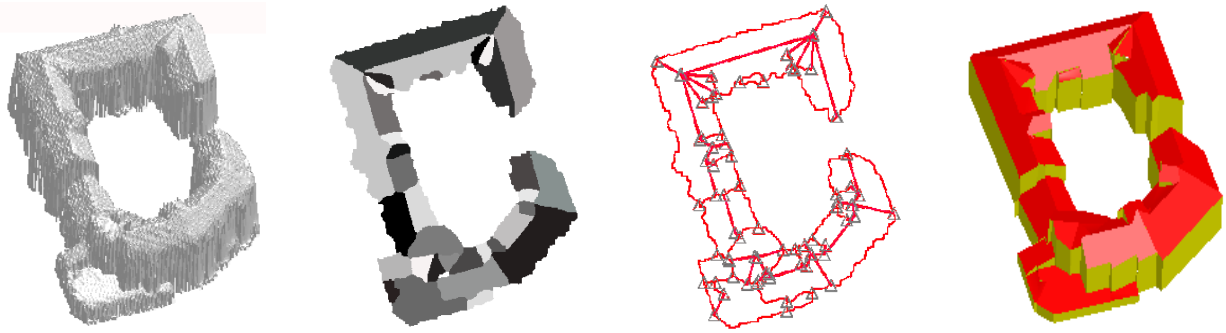


Figure 4-16: Illustration of the Process of Creating a 3D Building Model from Airborne LIDAR Data. From left to right: original elevation data belonging to the building; detected geometrical primitives (here, planar patches); detected keypoints and connectivity lines extracted from topology analysis; and the final 3D model.

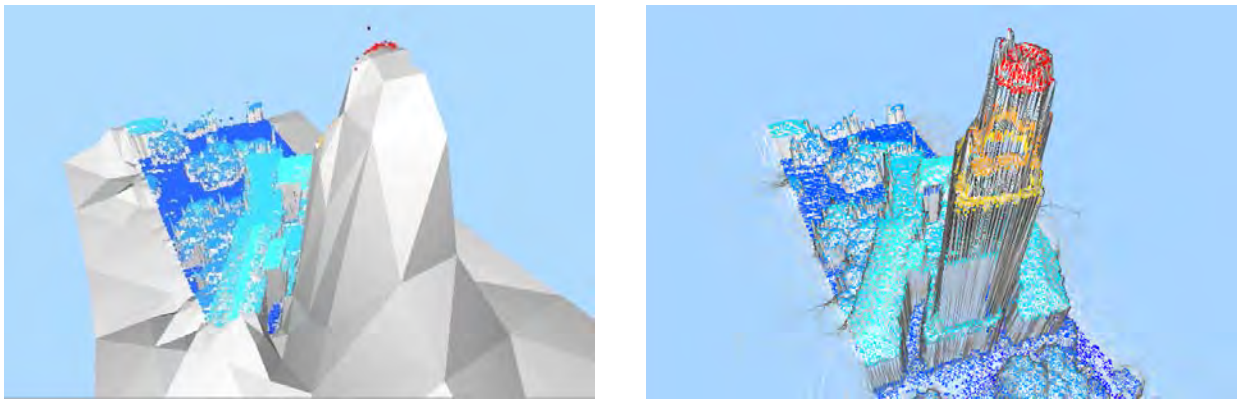
The core in the top-down approach is hypothesis testing – a pre-defined set of prototype building models, possibly with some tuneable parameters, is fitted to the data and the one yielding the highest score is selected as the winner. The use of prototypes in the matching process ensures nice-looking results but restricts the variations in the final building 3D models to the diversity of the prototype library. In addition, the computational complexity increases with the number of prototypes.

4.4 SAMPLE PROCESSED SET-118 DATA SETS

NATO SET-118 members have contributed collections of point clouds as data for use in testing various processing methods such as ability to handle non-uniformly spaced data, significant noise levels, to distinguish occlusions from missing data, and to accommodate complex geometry and topology.

4.4.1 WP-AFRL: PILAR (Polarimetric Imaging Laser RADAR)

2.5D LIDAR data was provided by Wright Patterson personnel (US AFRL – RY JM) from their line scanning airborne system [97]. The data was processed using South Carolina's on-line adaptive partitioning algorithm [98],[4], while progressively updating the surface as data is acquired. This is illustrated in Figure 4-17. In part a) data (point cloud coloured by height) has not yet been received in the foreground, but the gridding adjusts as data is acquired in order to better reflect the geometry. In this case the surface is a functional surface with simple topology. The final surface is shown in part b).



a) b)

Figure 4-17: a) Gridding of PILAR Data Swath is Adaptively Refined in Order to Estimate Heights of Local Surface Patch to Fit the Point Cloud; and b) Surface Statistics are Progressively Updated On-Line as Data is Received.

4.4.2 WP-AFRL: Flash LIDAR Data

Flash LIDAR data (see Section 3.2.1 for a detailed description) was provided by Wright Patterson (US AFRL – RY JM) personnel as part of SET-118's Bonnland data acquisition effort in November 2007. This data set was collected from a ground vehicle and is heterogeneously rich with some areas providing highly detailed point clouds (Figure 4-18), while others exhibit many of the problems common to imaging in urban settings, such as distinguishing missing returns, entrances, and occlusions in analyzing the data (Figure 4-20).

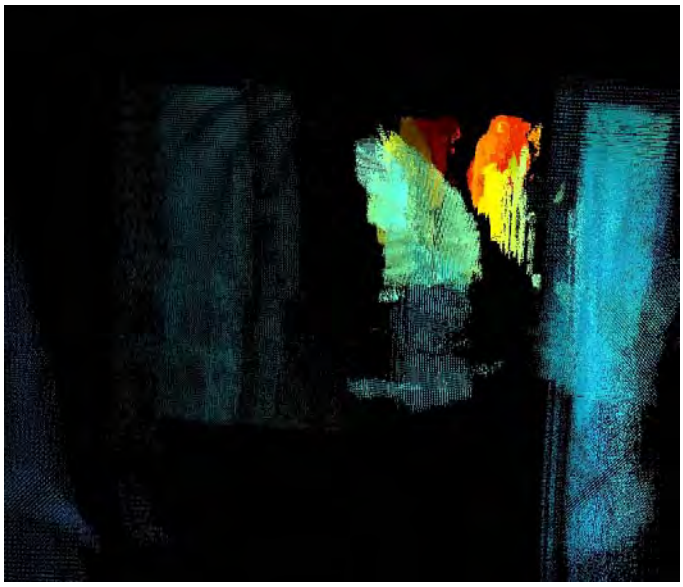


Figure 4-18: Rendering of Post-Assembled Bonnland Point Cloud, Time-Coded by Colour from Blue to Red, from a View Looking up the Roadway Haupstrasse.

In Figure 4-19, we have overlaid upon a frame, taken from the Bonnland video at the time of collection, the point cloud from an approximate viewpoint as the flash LIDAR in order to provide rough reference. Figure 4-20 illustrates natural problems which arise in surface reconstruction. Figure 4-20a) is a zoomed-out

view the full point cloud collection, with colour rendering used to denote acquisition time (red to blue) of the flash LIDAR images, which were later registered using an ICP algorithm. The LIDAR sensor was pointing forward as the platform moved toward the Northeast up the roadway of Haupstrasse. Figure 4-20b) is a rendering of two buildings along the right-hand side of the acquisition pathway. Figure 4-20c) is the result of surface reconstruction of the point clouds of the two buildings using a multi-resolution distance field. The holes in the point clouds (and therefore the reconstructed surfaces) are the result of both occlusions from the push-broom effect of a fixed sensor orientation on the platform as it moves forward from left to right and of data sparsity. In this particular case, the holes in the left hand sides of the surfaces cannot be distinguished between an entrance way, missing data, or occlusion without knowing the exact sensor state (position and pose) at the instant of acquisition. In the absence of further assumptions on the data or access to metadata information, such as time coding of the sensor state, inherent ambiguities such as distinguishing between occlusions and missing data will be unresolved.

a) **b)** **c)**

Figure 4-19: Referencing the LIDAR Point Cloud to Video from the Sensor Platform – a) Photo extracted from video; b) Portion of point cloud overlaying photo; and c) Rendering of that portion of the point cloud imaged from the first two building in the left foreground.

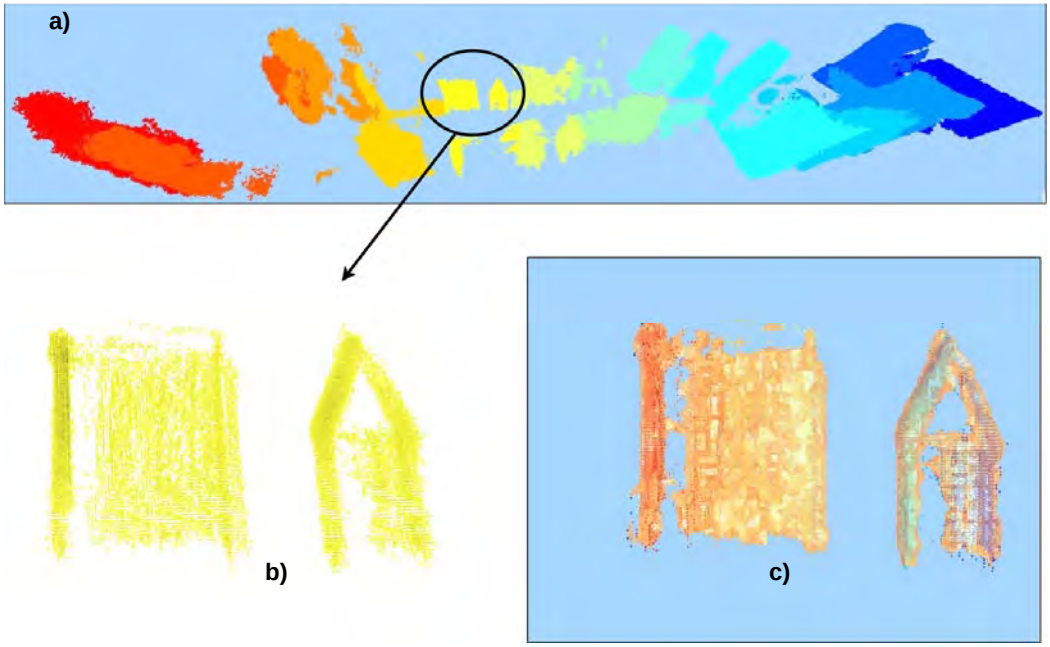


Figure 4-20: Bonnland Data Features and Processing – a) Zoomed out view of full extent of point cloud collection, again with colour rendering used to denote acquisition time of the flash LIDAR images, which were later registered using an ICP algorithm; b) Zoomed-in view of two buildings along the roadway; c) Surface reconstruction of the two buildings using a multi-resolution distance field.

4.4.3 FOI's Norrköping Collection

SET-118 members from Sweden's FOI and University of South Carolina in the U.S. have collaborated on testing promising segmentation and surface reconstruction ideas which have been developed in an ARO supported MURI. This data was fully described in the previous sub-section (Section 4.3), along with a complete description of the analysis and subsequent classification of urban objects, such as bare earth, waterways, and above ground objects such as various types of buildings, vehicles, roads, paths, power lines, trees, and other vegetation. In this brief sub-section, only the automatic processing of the Norrköping data by SET-118 participants is addressed, in order to test the effectiveness of automatic segmentation and reconstruction algorithms.

Two regions were cropped from the Norrköping data set. The point cloud for the cropped region #1 is shown in Figure 4-21a) where the *iterated Mahalanobis* algorithm [99] has been used to 'cluster' points and classify them by a colour map. The algorithm is based on a multi-resolution version of statistical Principal Components and used to determine local coordinates and corresponding distances in order to identify outliers to a given segment. The algorithm iteratively grows the current segment capturing all points in the cloud which are judged to be close in this distance. The algorithm may be implemented with computational complexity O(*N* log *N)*. Figure 4-21b) is a surface reconstruction of using adaptive partitioning and mathematical learning of non-parametric point distributions [98]. Figure 4-21c)-d) are given to illustrate the limits of this algorithm for cropped region #2 in the absence of additional meta information. In Figure 4-21c) the points in the point cloud are again classified by colour, according to the local Mahalanobis distance. The magenta coloured set of a sloped roof is well defined from its neighbouring segments. In Figure 4-21d) however, the magenta coloured collection of points is not able to separate a roof-top from neighbouring trees. Principal Component analysis, by comparing eigenvalues, is able to individually distinguish trees (as was the case in Figure 4-21c)), so it is not clear if an adjustment of the algorithm, especially in its multi-resolution instantiation, will be able to execute the separation.

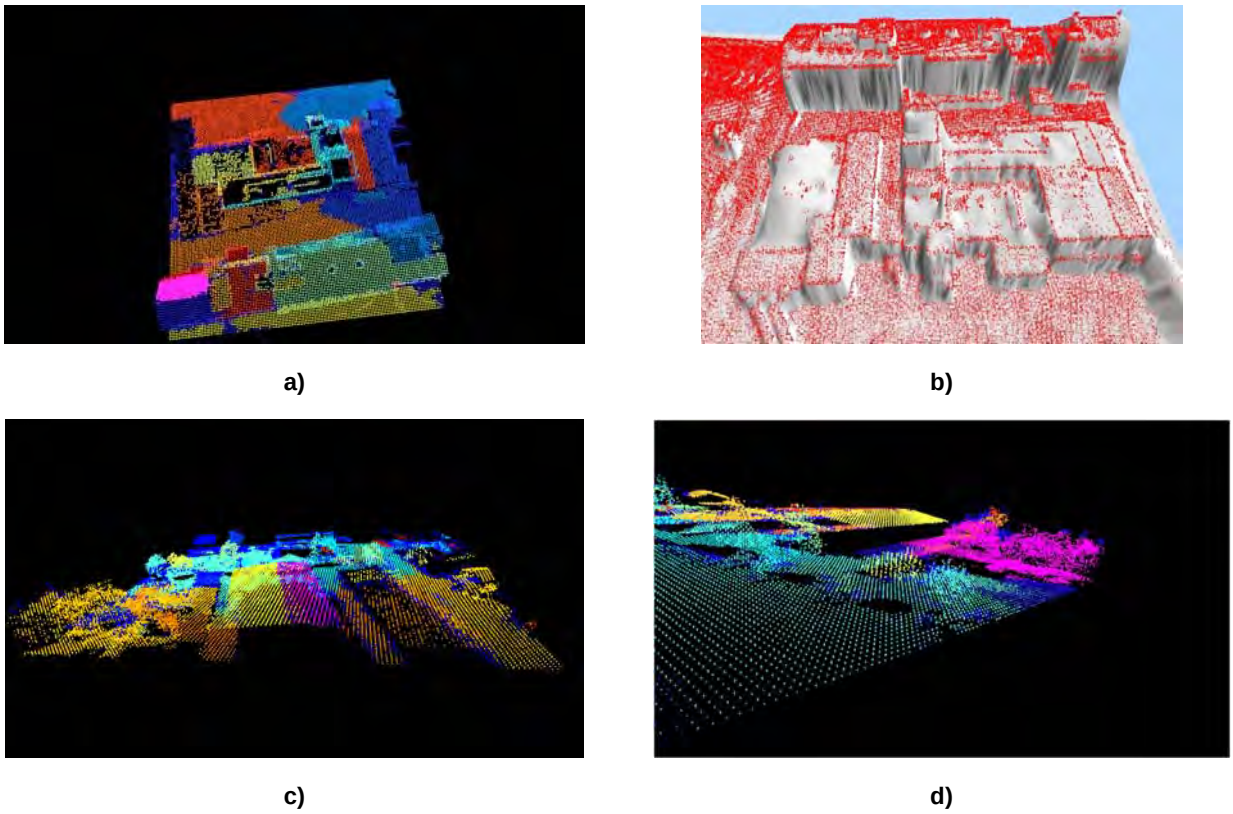


Figure 4-21: Segmentation and Surface Reconstruction of Cropped Portions of Norrköping Data Set – a) Point cloud for cropped portion #1 which is segmented using the iterated Mahalanobis algorithm with different segments classified by colour; b) Surface reconstruction using multi-resolution learning algorithm for surface heights for point cloud – after a rotation of about 180°; c) Point cloud for cropped region #2 with sample segment – a slanted roof – rendered in magenta; and d) Point cloud for cropped region #2 demonstrating the difficulty of segmenting merged features, in this case a building roof and adjoining trees (rendered in magenta).

Figure 4-22 shows an example of automatic terrain classification for the Norrköping data. The classes are ground (black), buildings (pink/red), vegetation (green), unknown (yellow) and water (blue). The classification is based on geometrical features except for the water class which relies on the fact that water regions manifest themselves as large regions void of laser hits.

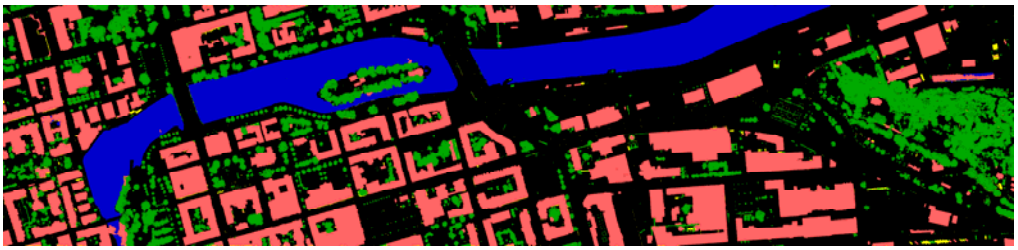


Figure 4-22: Results from Automatic Object Classification for a Part of Norrköping.

In Figure 4-23, some of the 3D building models for the city of Norrköping have been put on the bare-earth model and colorized using an aerial photo. Prior to the 3D modelling of the buildings, they were detected using algorithms discussed in Section 4.3.3.

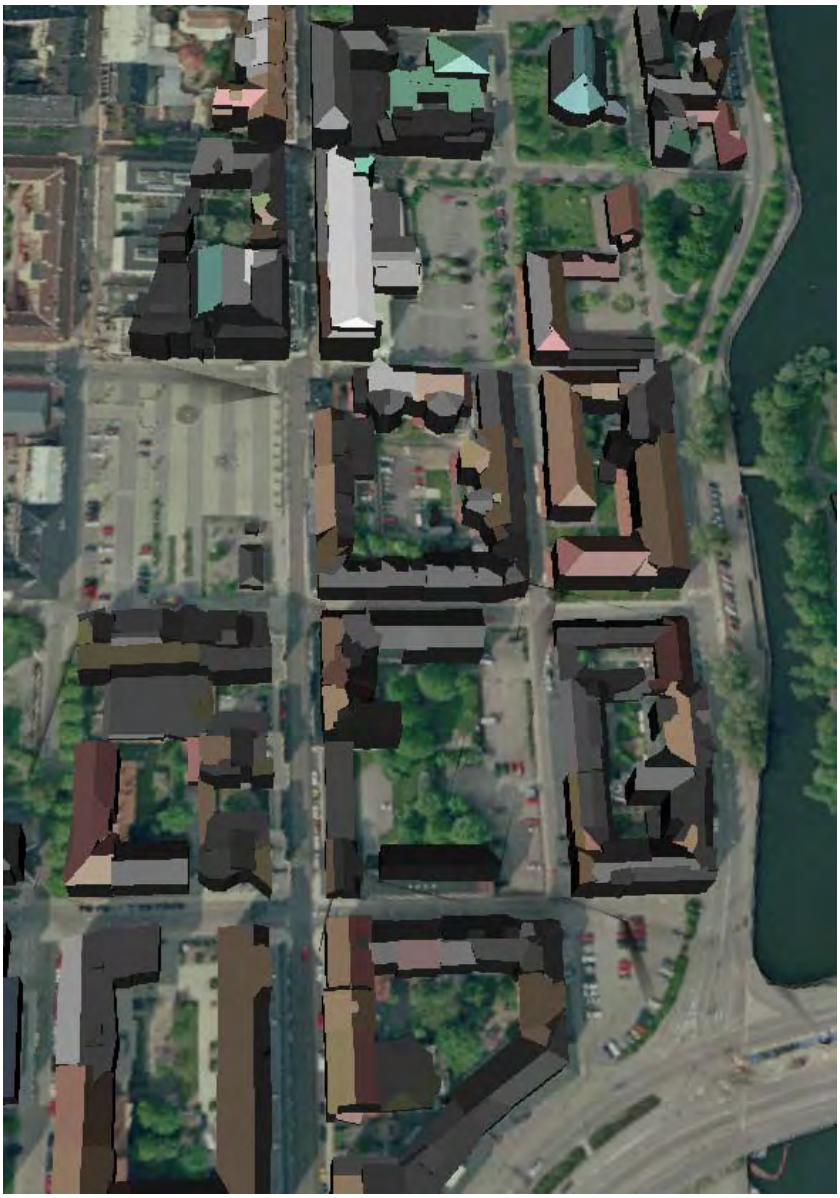


Figure 4-23: 3D Building Reconstruction of a Cropped Portion of Norrköping Data.

4.4.4 UK – Defence Science and Technology Laboratory's Burst Illumination LIDAR

The original objective of the Mahalanobis distance formulation was to determine distances between normally distributed multi-variate distributions, and to determine outliers for each. This feature of the iterated Mahalanobis algorithm is consequently very effective for denoising raw sensor data as a pre-processing stage prior to geometric reconstruction. Figure 4-24a)-d) present the results of such an algorithm [99] applied to burst illumination LIDAR data provided by NATO SET-118 collaborators from the Defence Science and Technology Laboratory. Figure 4-24a) is a photo of the imaged scene, with parts b)-c) representing side and front views of the sign in the foreground. In these images, magenta is used to colour the residual set after the iterated segmentation. Figure 4-24d) is a rendering of the surfaces constructed from the segmented sub-clouds along with the original data. Figure 4-24e)-f) show similar side and front views for the vehicle in the background of the photo in part a). The side views of the target board and vehicle indicate the potential of the method.

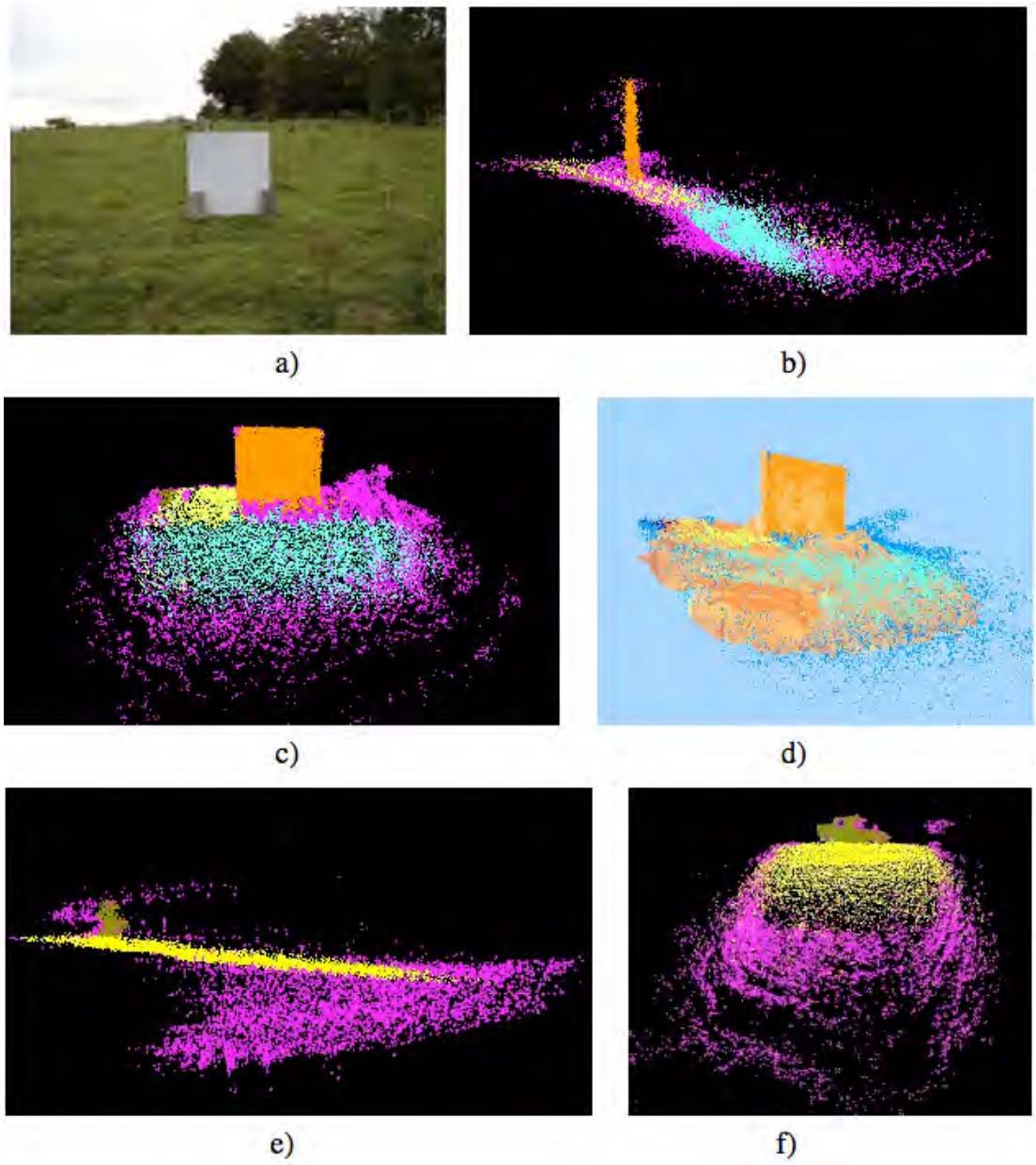


Figure 4-24: Iterated Mahalanobis Segmentation Applied to Burst Illumination LIDAR Data Provided by DSTL-UK NATO SET-118 Collaboration – a) Photo of test scene with target sign in foreground and vehicle in background; b) Side view of segmented sub-clouds around target sign with 'residual label' magenta; c) Front view of segmented sub-clouds around target sign; d) Reconstructed surfaces around target along with point cloud; e) Side view of background vehicle; f) Front view of background vehicle.

4.4.5 Processing of FOM – Urban SLAM Data Set

German (FGAN-FOM) members of SET-118 provided a 3D point cloud generated from vision-based video using SLAM (Simultaneous localization and mapping). The video was taken from a helicopter hovering around a building. Individual frames were processed for feature points using a corner detector. These points were correlated in adjacent frames using structure from motion techniques from computer vision in order to estimate their 3D positioning, as well as the camera state. Figure 4-25a) is a photo of the imaged building and its surroundings. This was a single frame extracted from the video with superimposed

feature points from that frame. Figure 4-25b) shows the full SLAM generated point cloud along with a surface reconstruction using an implicit fit of the heterogeneous sparse point collection. Figure 4-25c) is a rendering of the point cloud after segmentation, where distinct segmented components are labelled with colours (for example, magenta for the front face, orange, yellow, light blue for other components). Figure 4-25d) shows the result of fitting a surface to each component individually and then performing the corresponding Boolean operation to the collection of segments. This provides an acceleration of basic operations such as line of sight and viewsheds of the geometry.

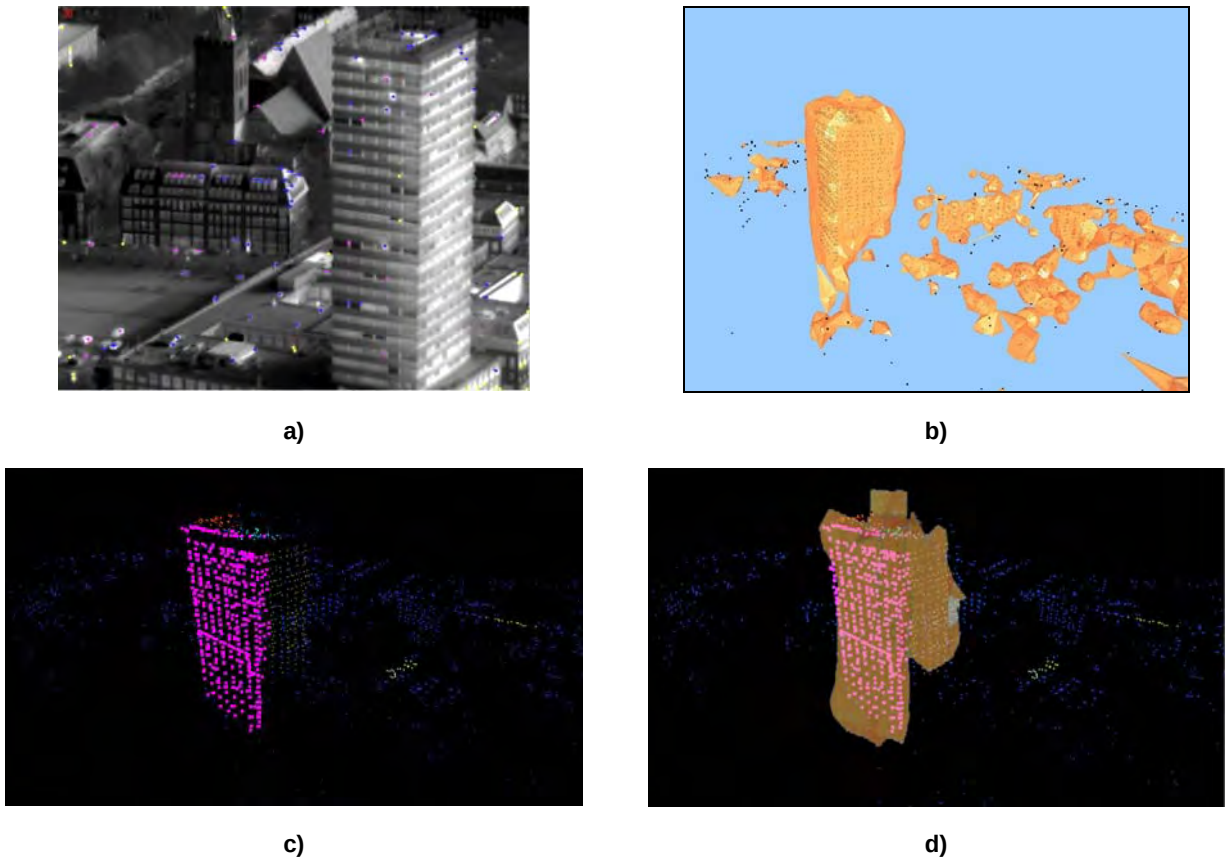


Figure 4-25: Segmentation and Surface Reconstruction of FOM SLAM Data – a) Photo of building with overlaid feature points; b) Point cloud generated from SLAM applied to sequence of feature points from aerial video with approximating multi-resolution surface fit using unsigned distance; c) Iterated Mahalanobis processing of SLAM generated point cloud, where different segments are labelled by colour; and d) Reconstruction of surface by fitting several segments individually and unifying the surfaces.

4.4.6 Defence Research and Development Canada, Valcartier LIDAR Data Collection

SET-118 members from DRDC-Valcartier provided high quality registered LIDAR data acquired by a ground level vehicle. The data set features occlusions, missing data, multiple returns and moderately complex topology. The Mahalanobis procedure was applied to the data set and individual features produced as segments were reconstructed. A sample of the point cloud, the Mahalanobis segmentation, and a resulting surface reconstruction of one segment is displayed in Figure 4-26. Image in Figure 4-26a) depicts the original LIDAR point cloud. Figure 4-26b) is a rendering of the iterated Mahalanobis-segmented sub-clouds labelled by various colours, with Person 2 coloured as magenta. A multi-resolution surface reconstruction is shown for these two sub-clouds in Figure 4-26c)-d), respectively.

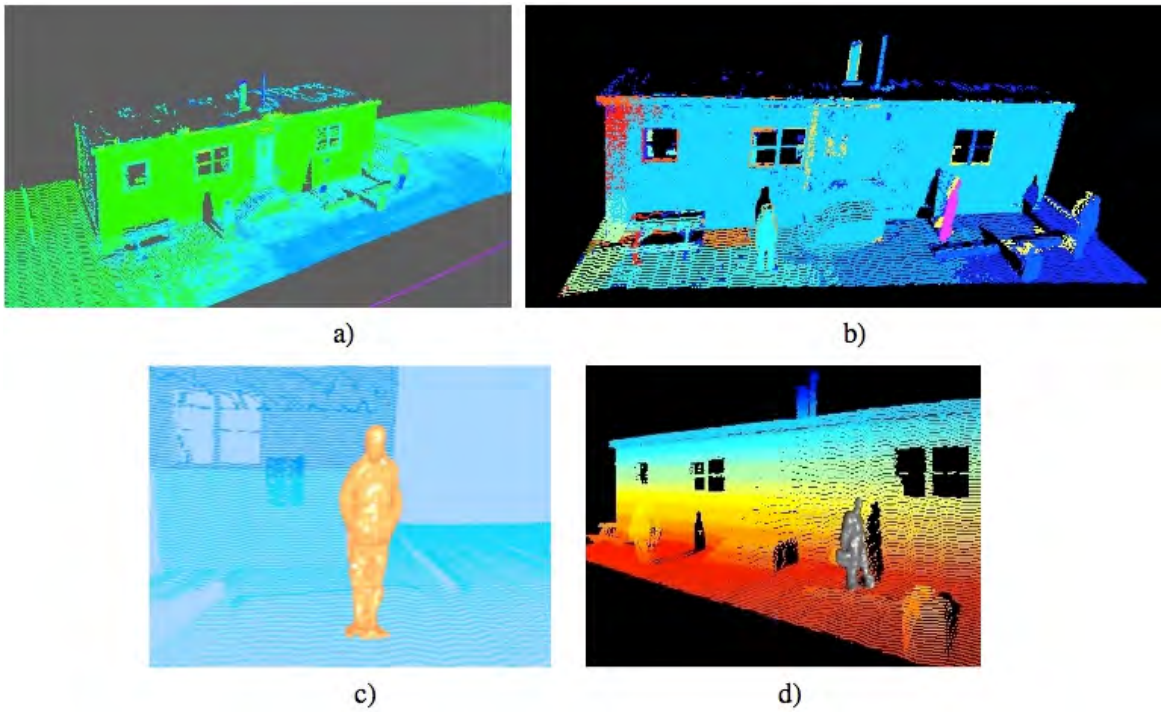


Figure 4-26: Iterated Mahalanobis Segmentation of Scanning LIDAR Data Provided by DRDC Valcartier NATO SET-118 Collaboration – a) LIDAR Point Cloud; b) Segmented sub-clouds labelled by colour; c) Surface reconstruction for person 1; and d) Surface reconstruction for person 2.

4.5 MODEL INSTANTIATION

4.5.1 Semi-Automatic 3D Extraction of Roads

Most of the precision targeting facilities need accurate geospatial information, including vector 3D models of urban infrastructures, which should be extracted from sensing datasets (e.g., satellite imagery) with well-defined accuracy requirements. In a typical 3D reconstruction of an operating scenario, linear features as roads or rivers are the ones most suitable for semi-automatic extraction.

There are different approaches to the problem of "road extraction" from satellite images, depending on the requirements and constraints based upon the end applications. Typically, two approaches are used:

- Automatic extraction where the user usually selects an area of interest and the system extracts all the visible roads in that area; and
- Semi-automatic approach where the user selects the road to extract, by selecting few points.

For our purposes we have adopted a semi-automatic approach, giving the user full control on the road to extract. In particular, the user has to select the start and the end points of the road he wants to extract, and then the software should generate from a stereo-pair of images a list of three-dimensional georeferenced triangles, which precisely model the surface of the road. Furthermore, we require the following constraints:

- To use only RGB data from the image stereo-pair, without any other data (LIDAR information, multi-spectral images, etc.).
- To use a Commercial Off-The-Shelf (COTS) photogrammetric tool, on a Windows-based standard personal computer, for basic stereo-pair handling.

- The final output shall be a vector model composed of a triangle strip, which approximates the road surface with a specific maximum error.

The extraction process is based on the following steps:

- Acquisition of the inputs: the first and last point of the road, the widths in these points (boundaries).
- Detection of the "centreline" of the road.
- Detection of the two sides of the road, using the "Ribbon Snake" or the "Ziplock Snake" algorithm.
- Altitude determination, using cross-correlation of patterns in the stereo-pair.

In the following paragraphs, each step is further detailed.

4.5.1.1 Road Boundary Detection

The first phase of our extraction algorithm requires the user to choose the first and last point of the road he wants to extract. Since the algorithm also needs the road width at the two extremities (the boundary widths), we help the user during this estimation process: as the user moves the mouse cursor along the road, we calculate automatically and visualize in real time the width of the road (see Figure 4-27).

Figure 4-27: Boundary Detection – the Roads to Extract (left) and the Detected Boundaries (right).

At the beginning of the extraction process, the mouse cursor changes in two concentric circles (as shown in Figure 4-27: the bigger one indicates the maximum search area for width detection, the smaller one indicates the minimum threshold to avoid disturbing factors (i.e., cars or middle lines). The radii of the two circles can be freely modified by the user.

The road width is detected by means of the *Canny edge filter* [61] and the *Hough transform line detector* [62]: the Canny edge filter gives us the capability to extract the edges of the road and the Hough transform line detector transforms the edges in a set of lines. Finally, we developed an algorithm (based on some built-up rules) that finds the best couple of lines and calculates the width and the orientation of the road.

4.5.1.2 Centreline Detection

The next phase of the road extraction process aims to determine the road *centreline*. This task is performed by means of the *Steger curvilinear feature detection* [63]. We put the input image into a filter chain, and we

pass the output of the chain to the Steger algorithm (see Figure 4-28). We use a locally-adapted contrast enhancement *Wallis filter* [64] in the chain, to enhance the contrast between road and surroundings.

Figure 4-28: Output of the "Steger Line Detector" Algorithm.

The segments found by the curvilinear feature detection process are elaborated by a *best path finding* algorithm to determine the correct road centreline, using a *gap filling* feature to minimize the effect of partial or total occlusions due to trees, cars, etc. (see Figure 4-29).

Figure 4-29: Road Centreline Calculated by the Path Finding Algorithm.

4.5.1.3 Road Modelling by the "Snake" Algorithm

We model the road using "ribbons" whose sides correspond to the road boundaries. We have chosen the *Ribbon Snake* algorithm to perform this task. A Ribbon Snake [65] is a time-dependent curve defined parametrically as:

$$\vec{v}(s,t) = (x(s,t), y(s,t), w(s,t)), \quad 0 \le s \le 1.$$

Such representation implies that each $v(s_0, t_0)$ is characterized by its width $2w(s_0, t_0)$ and the location of its centre $(x(s_0, t_0), y(s_0, t_0))$. All the centre points compose the centreline of the ribbon.

The snake deforms itself as the time progresses, to minimise an energy functional composed of three terms:

- A geometrical component which controls the snake tension.
- A geometrical component which controls the snake rigidity.
- The image contribution (the magnitude of image gradient). This is responsible for the snake attachment to the road contour.

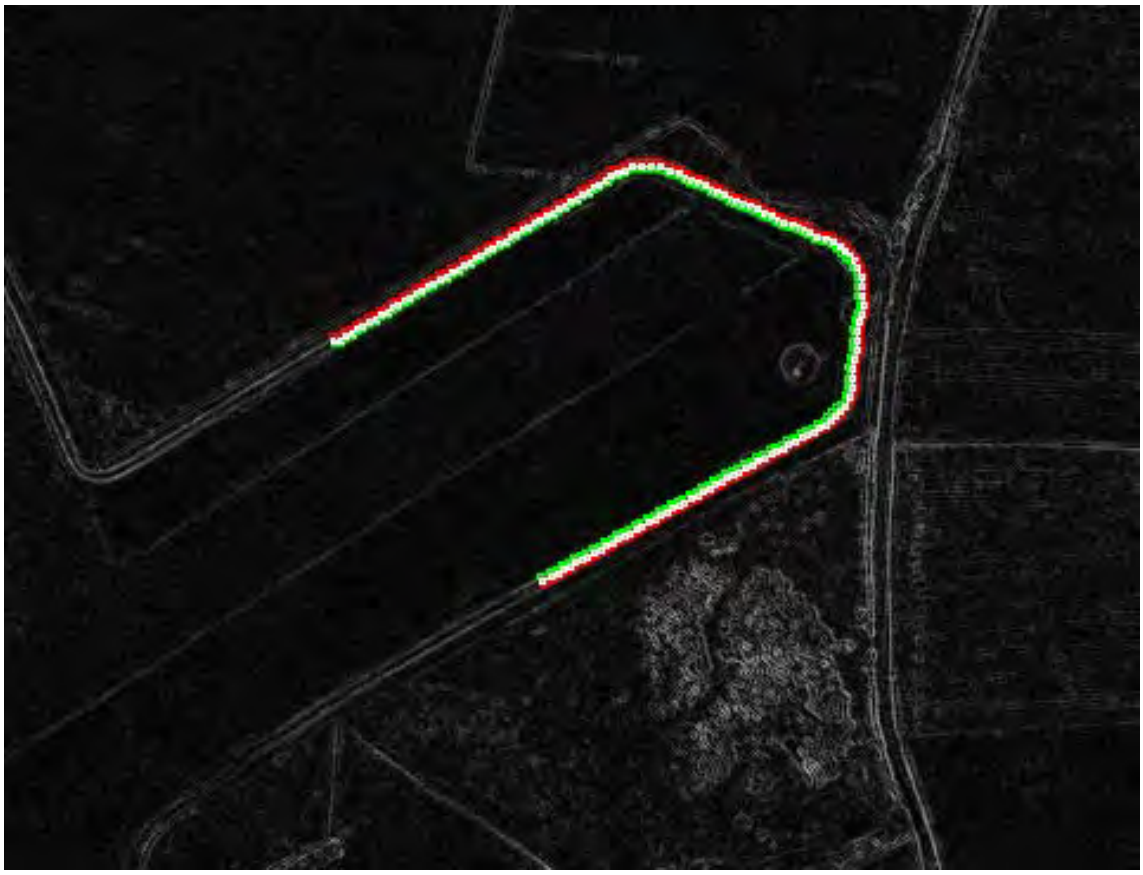


Figure 4-30: Snake Initialization by Centreline and Boundary Widths.

We use the centreline detected in the previous step and the two boundary widths given by the user, as the initial state for the Ribbon Snake. We use the Dynamic Programming implementation for the optimization procedure: it is an iterative algorithm that stops when the total energy of the snake is minimized (considering the energies from all the points). The resulting snake is passed to a *decimation* algorithm that deletes some points, depending on their collinearity properties.

4.5.1.4 Road Modelling by the "Ziplock Snake" Algorithm

In certain cases, the road centreline could not be found, due to strong total occlusions, like building shadows or forests. So, we developed an alternative algorithm to use only in these cases. It does not use the centreline data, but a modification of the original Ribbon Snake algorithm, named "Ziplock Snake" [66]. In this version, the snake is optimized taking into account first the image information near its extremities and then, progressively, toward its centre.

As shown in the image (see Figure 4-31), the snake is composed by two parts, the active part, in which each point is influenced by all the energy contributions, and a passive part, in which the image contribution is disabled. The force boundaries are moved individually, one vertex at the time. A force boundary is advanced when we can verify that the motion of the corresponding active section has stabilized.

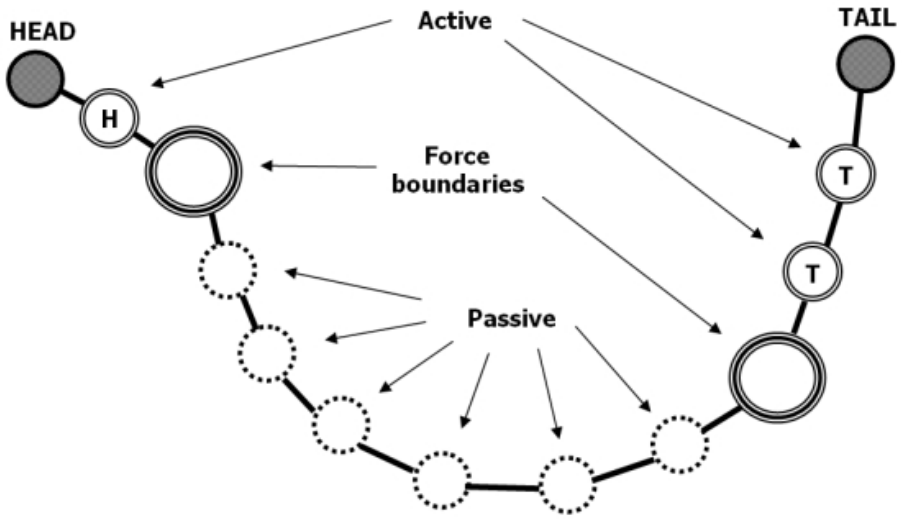
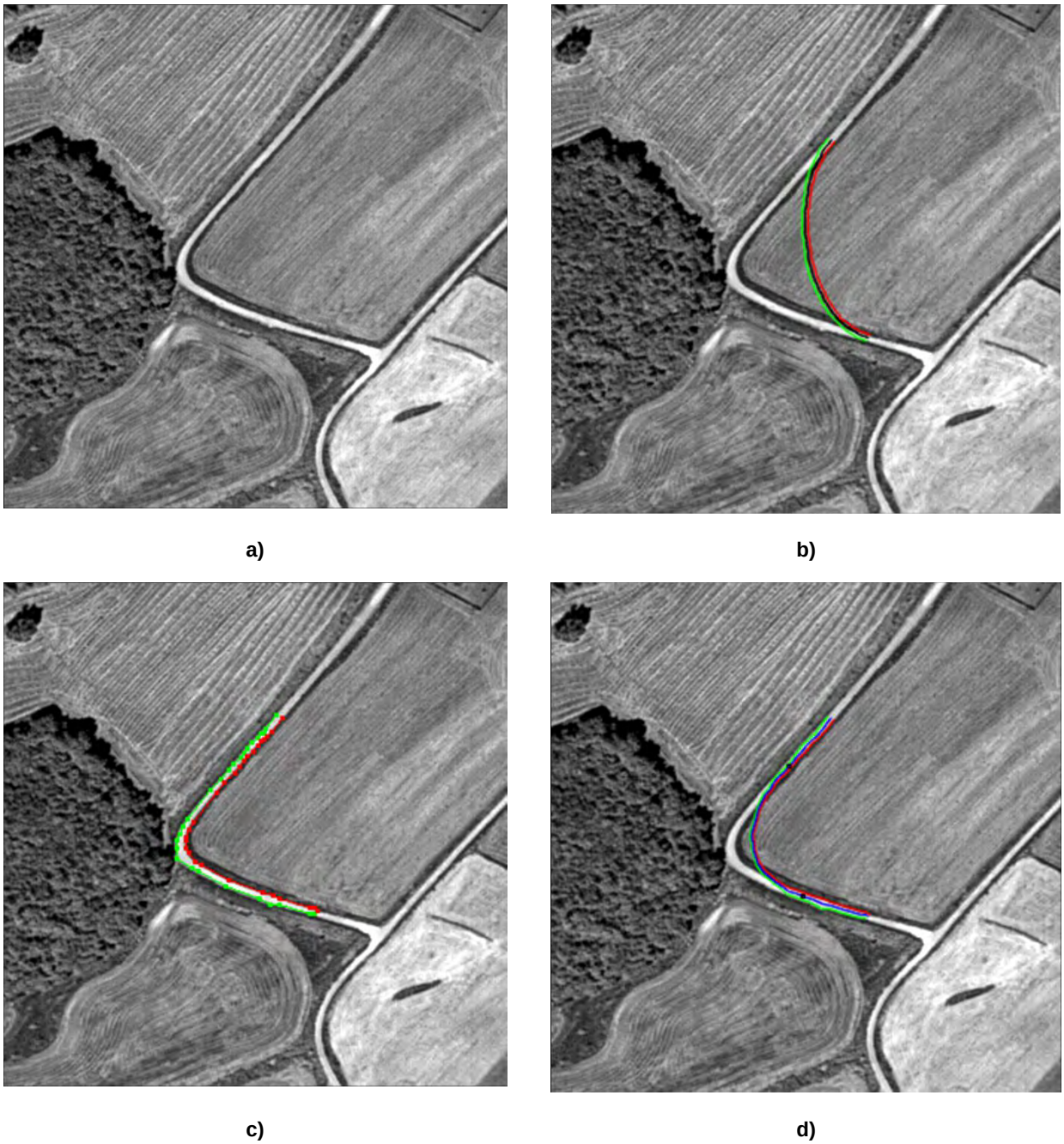


Figure 4-31: Ziplock Snake Elements.

We can see the ziplock approach in this sequence of images (see Figure 4-32): the snake is attached to the image contour in a way similar to a ziplock being closed. Due to the computational expense of the computation required, one should use ziplock snakes to model portions of roads shorter than the ones modeled by the centreline method.



a) **b)**

c) **d)**

Figure 4-32: Ziplock Snake – a) Input image; b) Snake initialization; c) Snake iteration; and d) Final result.

4.5.1.5 Altitude Determination from Stereo-Couples

- The points we have extracted from the images during the previous steps are georeferenced in latitude and longitude, but their altitude is approximated. In order to assign the right altitude value to each point, we perform the calculation applying a cross-correlation algorithm to the stereo-couple: as the algorithm recognizes the same pattern in the two images, it is possible to determine the correct elevation of the point by triangulation. These altitude values have to be checked, since they could be affected, in some cases, by cross-correlation errors (i.e., due to shadows and/or discrepancies between

the times in which the images were taken). The coherence of these altitude values is automatically verified by the software against a set of criteria and, if necessary, the points in the neighbourhood are analyzed until an acceptable result is reached.

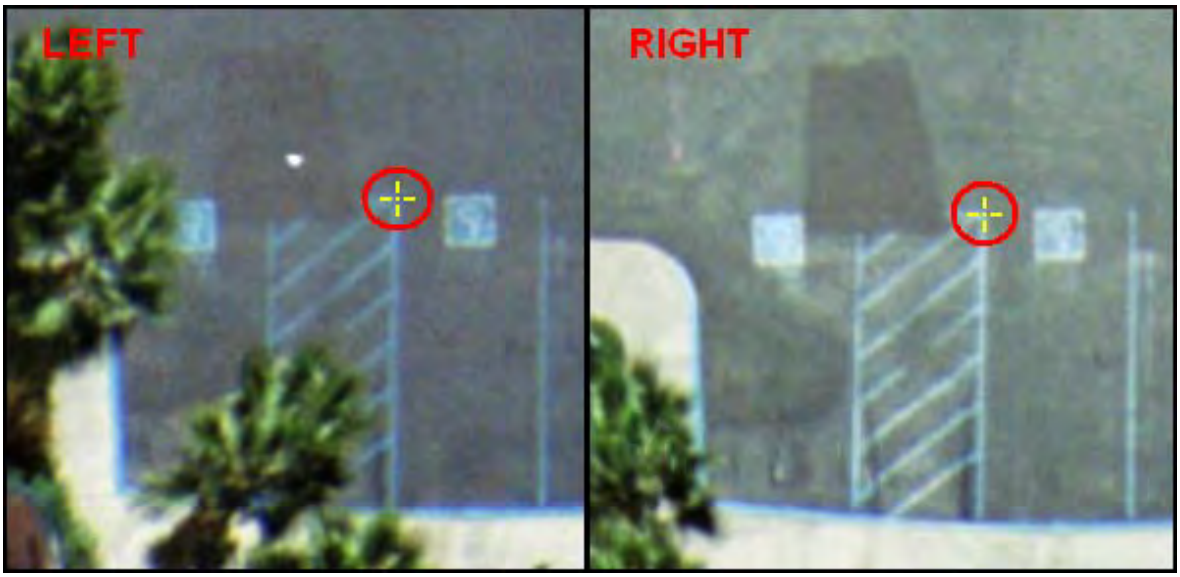


Figure 4-33: Cross-Correlation Example in a Stereo-Pair.

4.5.1.6 Results and Validation

The polygonal models for three test cases are shown in Figure 4-34, Figure 4-35 and Figure 4-36. For each test case, the reference model and the automatically generated road are shown. The reference models were manually extracted and checked for the best accuracy, using three Ikonos satellite stereo-pairs (GSD about 1 meter). An automatic test was performed on the three couples of road models, in order to calculate the horizontal and vertical rms (Root Mean Square) differences between the reference model and the automatic extracted road. Such differences were calculated on both the two sides of each road, using a sampling step equal to 0.01 meters: this means that we have calculated 100 samples for each meter of the reference road. The optimal sampling rate was determined by the stability analysis depicted in Figure 4-37: the rms values reach the convergence when the number of samples per meter is greater than 80.

Figure 4-34: Reference Model #1 (up) and Automatically Extracted Road #1 (above).

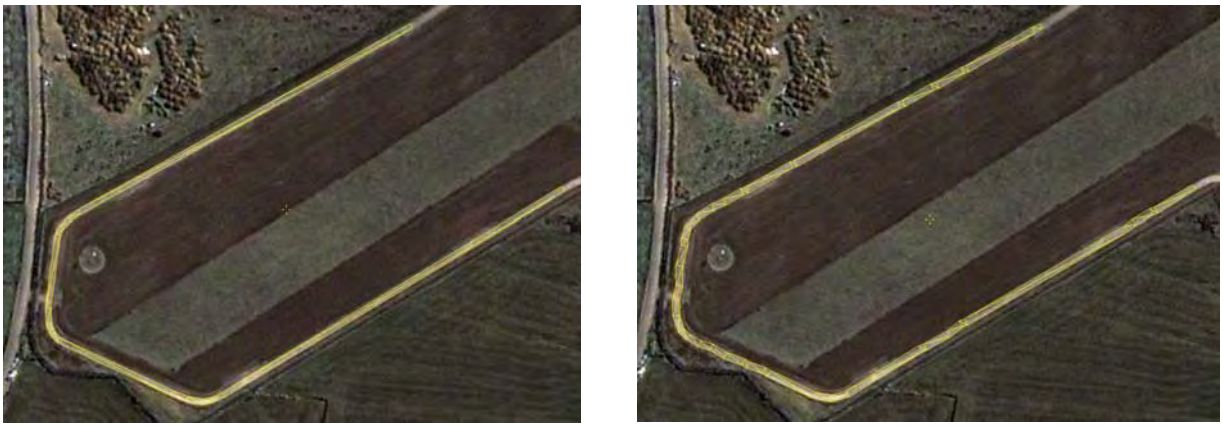


Figure 4-35: Reference Model #2 (left) and Automatically Extracted Road #2 (right).

Figure 4-36: Reference Model #3 (left) and Automatically Extracted Road #3 (right).

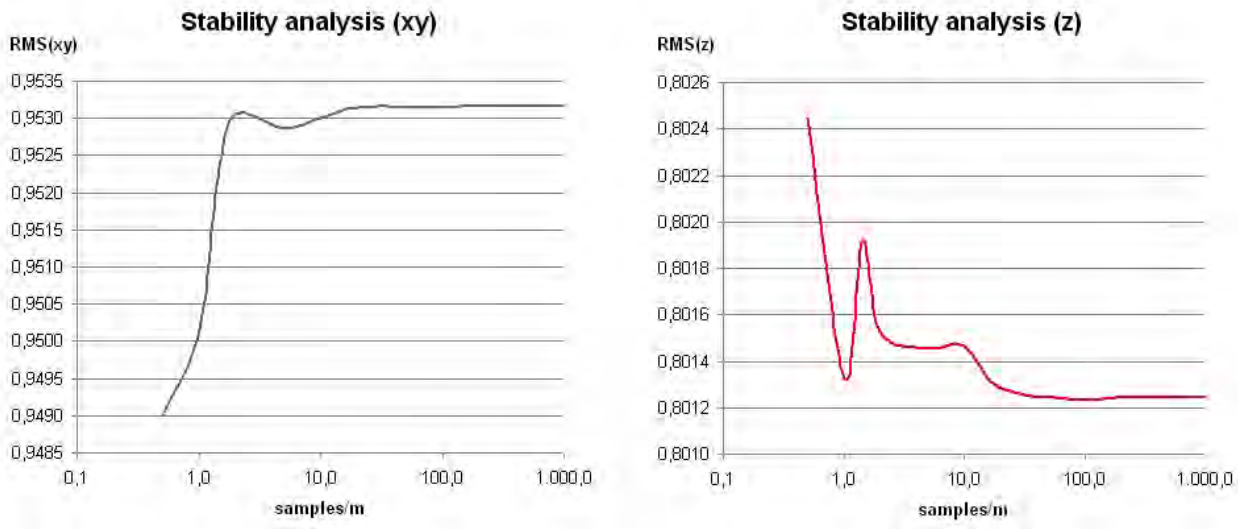


Figure 4-37: Stability Analysis for the xy-Plane and the z-Axis.

In the following table, the horizontal and vertical rms differences are shown, for three test cases (see Table 4-3).

Table 4-3: Horizontal and Vertical rms Differences for Three Test Cases.

Test Case	Horizontal rms (xy-Plane)	Vertical rms (z-Axis)
Road Model #1	0.953 m	0.801 m
Road Model #2	0.619 m	0.218 m
Road Model #3	0.582 m	0.439 m

4.5.2 Building Reconstruction from Multi-Aspect High-Resolution InSAR Data

In Sections 4.1 – 4.3 of this chapter, reconstructions based upon LIDAR sensor data were described. In this section we will give a brief overview of existing reconstruction approaches considering multi-aspect SAR or InSAR data. Subsequently, the magnitude and interferometric phase signature of flat- and gable-roofed buildings under orthogonal viewing directions is analyzed. In the last section our reconstruction approach is described and results based on high-resolution InSAR data are presented.

4.5.2.1 State-of-the-Art

A variety of building reconstruction methods based upon various sensor data have lately been presented in literature. In the following, only the group of approaches based on multi-aspect data is considered and shortly discussed.

4.5.2.1.1 Building Reconstruction through Shadow Analysis from Multi-Aspect SAR Data

The extraction of building dimension by analyzing the building shadow based on a single SAR magnitude image was presented in Bennett and Blacknell [67]. Extending this approach to multi-aspect SAR magnitude, images some ambiguities in the shadow interpretation can be solved and more robust reconstruction results are achievable [68],[69],[70]. The drawback of these methods is given by the shadow analysis itself, which assumes suburban area, no interferences with other objects and flat terrain.

4.5.2.1.2 Building Reconstruction from Multi-Aspect Polarimetric SAR Data

Multi-aspect polarimetric SAR data are considered for building reconstruction by Xu and Jin [71]. First, parallel line segments are extracted from the multi-aspect data by means of a local Hough transform. Second, parallelograms are generated, which contain the layover area of the buildings. Subsequently, a classification takes place to discriminate between parallelograms of direct reflection from the ones of double-bounce signal propagation. In the last step, a maximum likelihood method is used to fuse the multi-aspect information and to reconstruct buildings three-dimensionally. The main demand of this approach is that buildings are isolated, which is difficult in dense urban areas and for higher buildings.

4.5.2.1.3 Building Reconstruction from Multi-Aspect InSAR Data

In Bolter and Leberl [72] the detection and reconstruction of buildings is based on multi-aspect InSAR data by considering InSAR height and coherence. First, the maximum height value of all aspects is chosen and the resulting height map is smoothed with a median filter. Afterwards, potential building regions are generated and minimum bounding rectangles are fitted to these regions. The differentiation between buildings and other elevated vegetation is based on coherence and height map. In a last step, simple building models with either a flat roof or a symmetric gabled roof are fit to the segmented building regions. This approach was further extended in [73] including information from corresponding SAR magnitude data, especially exploitation of building shadows. For these approaches, problems arise if buildings are not isolated or if they are higher than the ambiguity height of the InSAR acquisition.

4.5.2.1.4 Iterative Building Reconstruction Using Multi-Aspect InSAR Data

An iterative building reconstruction approach from multi-aspect InSAR data was presented by Soergel [74]. Based on speckle reduced magnitude images, primitive objects such as edge and line structures are extracted from the slant range data. Subsequently, building hypotheses are set up by generating parallelograms from primitive objects for every aspect. Thereafter, the hypotheses are projected from slant range to ground range geometry in order to fuse the multi-aspect hypotheses. First, the buildings are reconstructed as elevated objects with flat, gabled, or pent roofs as well as right-angled footprints. In order to overcome occlusion effects, building candidates from multiple aspects of the same scene are fused. The resulting building

hypotheses are used as input for a simulation to detect layover and shadow regions. The comparison between this simulated SAR data and the original SAR data shows differences in case of false detections, which are subsequently eliminated by generation of new building hypothesis. The entire procedure is repeated iteratively and is expected to converge towards a description of the real 3D scene. The main drawback of this work is a minimum size of reconstructible buildings. Due to the fact, that the first building hypotheses are assembled in the slant range geometry, very small buildings (e.g., 10 m x 10 m x 5 m – width x length x height) are undetected by this approach. Hence, in our approach [75] the assembly of building hypotheses is realized in the ground range geometry. Furthermore, along with the magnitude signature, an analysis of the interferometric phase signature is considered during the reconstruction approach, which will be described in much more detail in the following sections.

4.5.2.2 Signature of Buildings in High Resolution InSAR Data

In this sub-section, the InSAR signature of flat- and gable-roofed buildings under orthogonal illumination directions is shown and analysed. Especially the signature parts relevant for our building reconstruction are characterized in the magnitude and interferometric phase signatures.

4.5.2.2.1 Magnitude Signature of Buildings

The appearance of buildings is characterized by the side-looking viewing geometry of the SAR sensors and range measurements. In the first row of Figure 4-38 the optical signature and, in the second, the typical SAR signature parts – layover, corner, roof and shadow are schematically shown, whereby the grey scale values represent the expected magnitude value in the SAR data. The layover results from the integration of a different backscatter signals (e.g., ground, building wall and roof) into the same range cell. The corner is caused by double bounce reflection between ground and building wall. The roof area is characterized by a single backscattering of the building roof. The ground behind the building is partly occluded and appears then as dark shadow region.

In the third row of Figure 4-38, the real SAR magnitude signature of the two buildings is given by considering the orthogonal illumination directions. The signature of the flat-roofed building changes radically between both data sets. The single backscattering of roof area is not observable in the second data set. Similar effects are observable for the gable-roofed building. The layover area is sub-divided in two parts, which are characterized by different backscatter contributors. The first part, the bright line, results from sloped roof, wall, and ground, and the second darker part from wall and ground only.

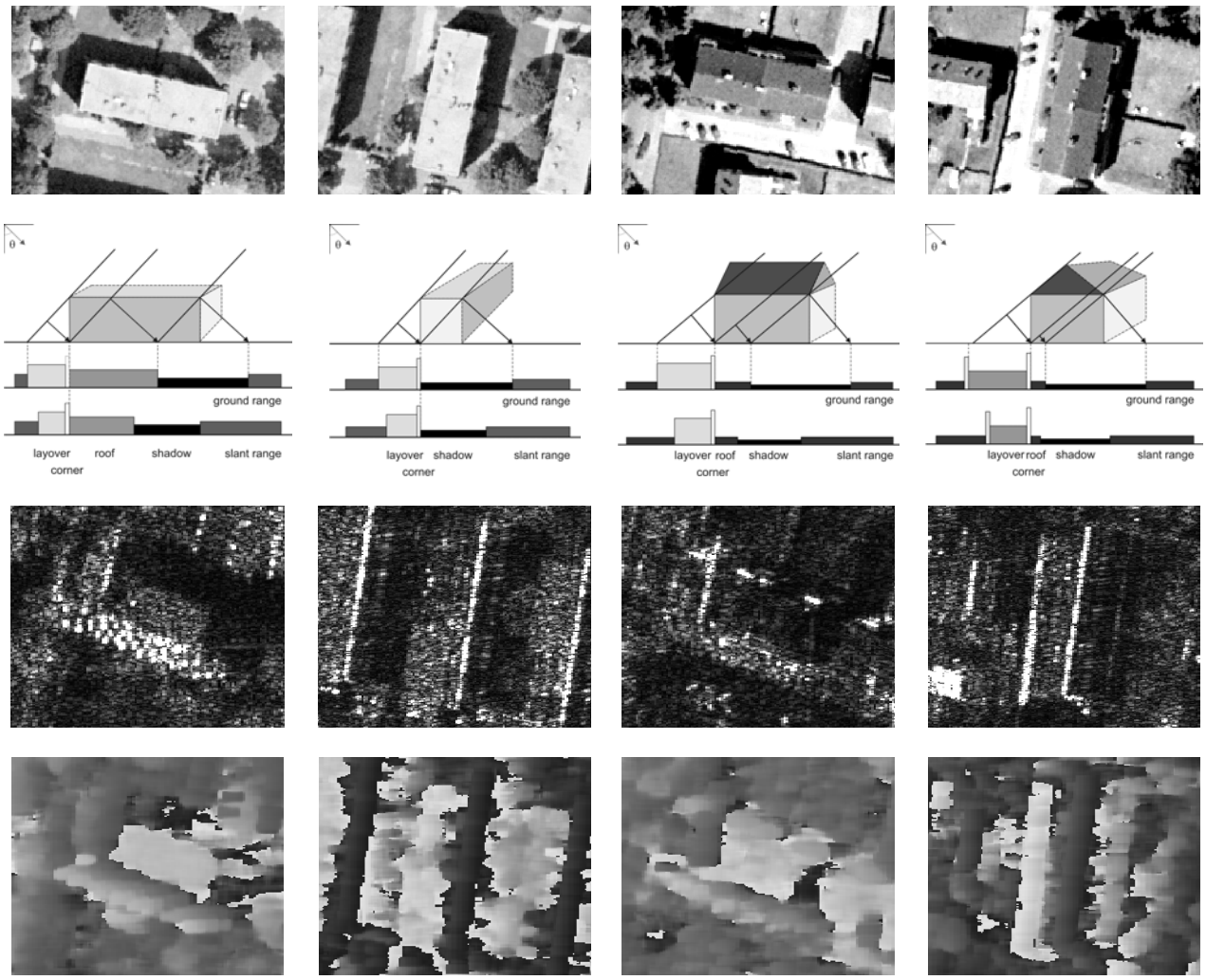


Figure 4-38: Appearance of Flat- and Gable-Roofed Buildings Under Orthogonal Illumination Conditions – Optical image (first row); Schematic view (second row); SAR magnitude data (third row); and InSAR phase data (fourth row).

Next to the illumination direction, also the building dimensions and the off-nadir angle θ influence the building signature, which was discussed for flat-roofed buildings in [75] and for gable-roofed buildings in [77]. For our purpose of small building detection, the signature analysis pointed out that the corner lines and double lines in the magnitude data are the most stable and dominant features. Hence, our building recognition and reconstruction is based primarily on bright lines segmentation.

4.5.2.2.2 Interferometric Phase Signature of Buildings

Beside the magnitude pattern, the interferometric phase signature of buildings is characterized by layover, multi-bounce reflection, direct reflection from the roof and shadow, too. In Figure 4-38 (fourth row) the variation of the InSAR phase signature due to different building types and illumination properties is shown. In general, the phase values of a single range cell result from a mixture of the signal of different contributors. Hence the final InSAR height of an image pixel is a function of the heights from all objects contributing signal (e.g., heights from terrain, building wall, and roof) to the particular range cell. The layover region in the InSAR phase data is also called front-porch region and shows a downward slope from close range to far range in slant range direction. For the flat-roofed building signature, this is caused

by the two constant (ground and roof) and one varying (wall) height contributor. The significant corner position in the magnitude profile shows in the phase profile a phase value nearly similar to local terrain phases. This is caused by the sum of the double-bounce reflections between ground and wall, which have the same signal run time as a direct reflection at building corner point. The subsequent single response of the building roof leads to a constant trend in the phase profile. From the shadow region no signal is received so that the phase is only characterized by noise. Focusing on the illumination direction, similar effects are observable between the magnitude and phase data. The orthogonal illumination configuration given in the second column leads to a dominant layover area, but no constant phase area from the building roof does exist due to the small building width.

The phase signature of the gable-roofed building under the given illumination direction (third column) is mainly dominated by the backscattering of ground and building wall. Reason for less developed response of the building roof is the roof orientation away from the sensor. In the orthogonal case (fourth column) the response of the roof dominates the layover signature. As a consequence, the layover maximum is much higher than for the other illumination direction. The shape of the layover phase profile is determined by the off-nadir angle, the eave, and the ridge height. For example, a strong steep slope leads to a high gradient in the phase profile. A more detailed analysis of flat- and gable-roofed building phase signatures is given in [76] and [77].

Summing up, for our purpose of small building detection, the signature analysis of InSAR data pointed out, that the corner lines and double lines in the magnitude data are the most stable and dominant feature for building detection. Hence, our building reconstruction is based primarily on bright lines segmentation. Furthermore, the layover region in the phase data contains a lot of geometric building information, which can be very helpful during the 3D reconstruction of gable-roofed buildings and the post-processing of flat-roofed buildings.

4.5.2.3 Building Reconstruction Approach

In the following, our approach of building reconstruction based on multi-aspect high-resolution InSAR data is described. The full workflow is given in Figure 4-39 and main processing steps are described briefly.

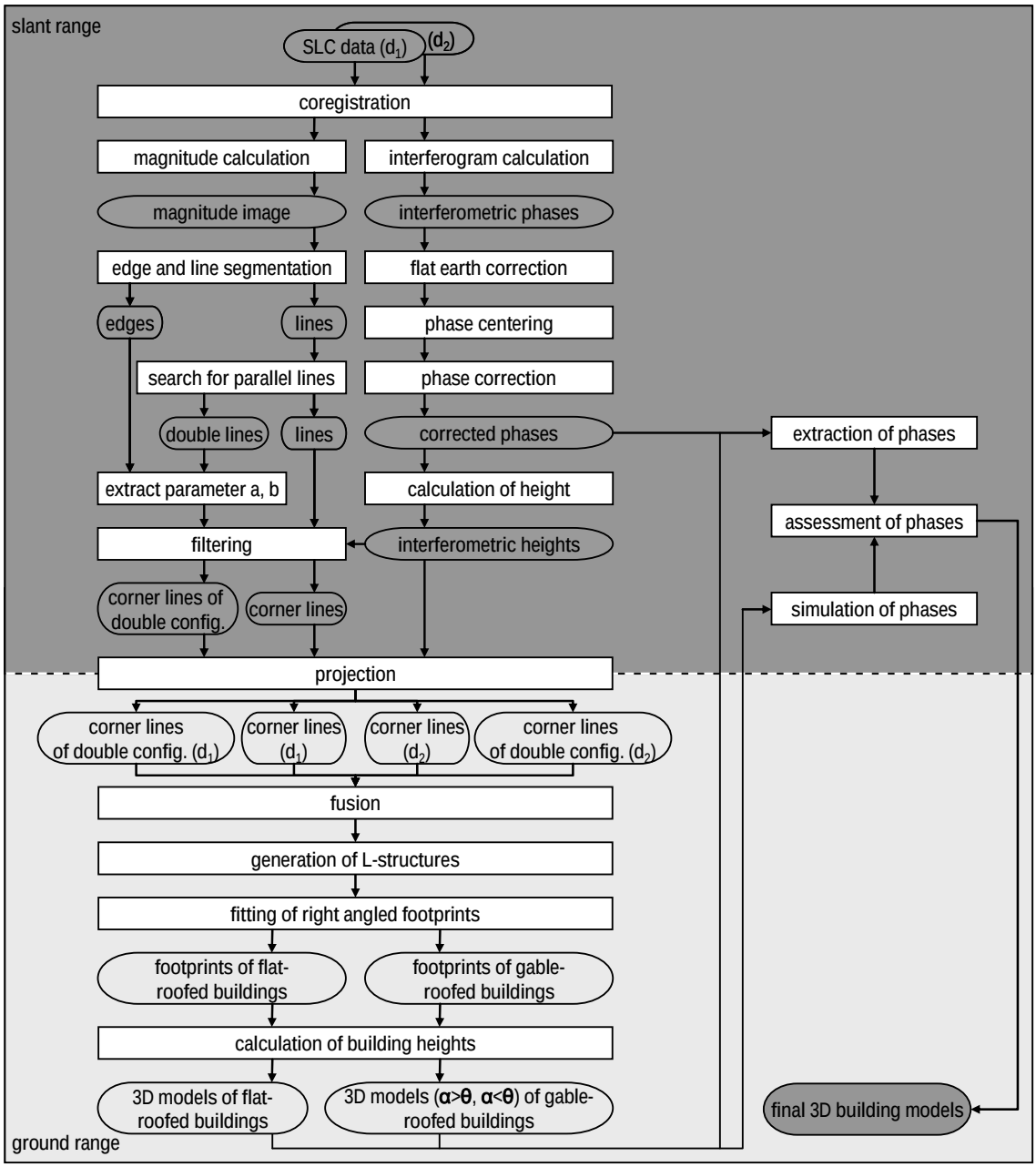


Figure 4-39: Workflow of Algorithm.

We assume as prior knowledge of the buildings that they have rectangular footprints, a minimum building coverage of 8 m x 8 m x 4 m (width x length x height), vertical walls and a flat or gable roof. The recorded InSAR data have to contain acquisitions from at least two aspects spanning an angle of 90 degree in the optimal case in order to benefit from complementary object information.

4.5.2.3.1 Pre-Processing

The pre-processing of the data starts in the slant range geometry and contains the sub-pixel registration of the interferometric image pairs and the calculation of the interferograms. This interferogram generation includes multi-look filtering, followed by flat earth compensation, phase centring, phase correction, and height calculation.

4.5.2.3.2 Extraction of Building Features

The extraction of building features contains:

- Segmentation of primitives.
- Extraction of gable-roofed building parameters.
- Filtering of primitives.
- Projection and fusion of primitives.

Based on the magnitude images, edge and line primitives are segmented by using an adapted detection operator to [78] and [79]. Subsequently, based on the segmented bright lines “double line” pairs are assembled and two geometric building parameters of gable-roofed buildings are calculated. More details are presented in [80]. Thereafter, the primitives are filtered by exploiting the characteristic interferometric heights at the corner line position. The final corner lines have to show a mean InSAR height value similar to mean terrain height. In the next step, the corner lines of each aspect are projected from the individual slant range to the common ground range geometry. The fusion of all multi-aspect primitives provides a common group of building primitives.

4.5.2.3.3 Generation of Building Hypotheses

The generation of building hypotheses contains the generation of building footprints and the calculation of building heights. The building footprints are generated from the common group of building primitives by assembling pairs of lines, which form an L-structure. Based upon this right angled footprints and the footprint hypotheses are passed. The subsequent height estimation considers the building roof type, whereby the results of “double line” segmentation are used to distinguish between flat- and gable-roofed building hypotheses. More detailed information as well as the equations for the height estimation are given in [80].

4.5.2.3.4 Post-Processing of Building Hypotheses

The final post-processing of the building hypotheses is realized by a detailed analysis of the building InSAR phases. This analysis is based on the comparison of real InSAR phases and simulated phases [76] received from the generated hypotheses. First, ambiguity problems in the reconstruction of gable-roofed buildings can be solved in this way. Second, the resulting oversized footprints caused by adjacent trees or fences can be corrected by the comparison step. A detailed description of both post-processing steps in given in [80]. After the assessment step the final 3D building results are created.

4.5.2.3.5 Results of Building Reconstruction

The results of the building reconstruction are presented in Figure 4-40. As test area the city of Dorsten (Germany), characterized mainly by residential flat- and gable-roofed buildings, was chosen. The InSAR data were acquired by the Intermap Technologies X-Band sensor AeS-1 [81]. The data show a spatial resolution of about 38 cm in range and 16 cm in azimuth direction and were taken by an orthogonal viewing configuration.

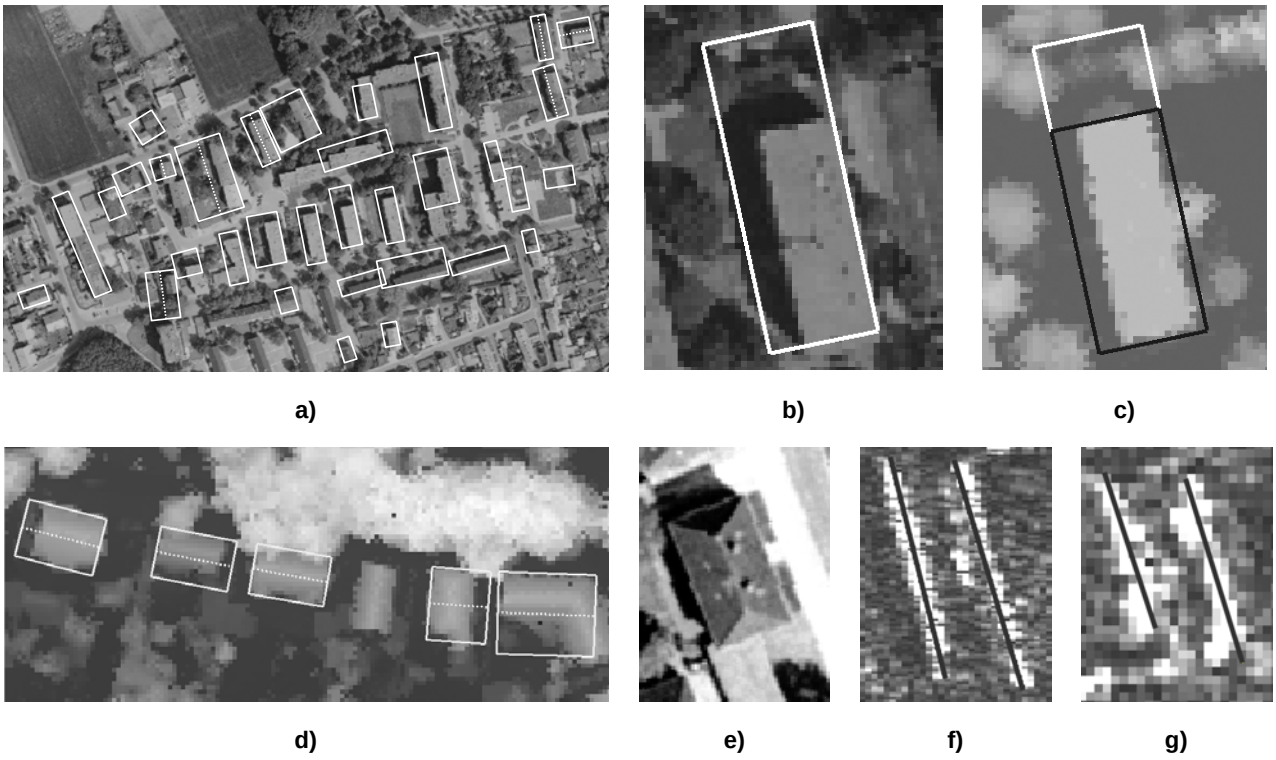


Figure 4-40: a) Results of Mainly Flat-Roofed Building Detection; b) Results of Gable-Roofed Building Detection; c) Oversized Building Hypothesis; d) Corrected Building Hypothesis; e) Optical Data of Hip-Roofed Building; f) Detection Result of AeS-1 Data; and g) TerraSAR-X Data.

In Figure 4-40 the results of the flat- and gable-roofed building reconstruction are shown. The majority of the buildings given in (a) are well detected and shaped. The less detection in the lower part of the scene is caused by missing overlap between the multi-aspect data. Furthermore, our building recognition failed if trees or buildings were located too closely to the building of interest. Some of the reconstructed footprints, especially the row of buildings in the middle of the scene, are larger than ground truth, due to too long segmented corner lines caused by signal contributions of adjacent trees or fences.

The second part (b) shows good results for the gable-roofed building detection, too. The ridge orientation of four buildings is well detected. The fifth building shows a hip roof instead of the assumed gable roof. The final 3D results of this building group are given in [77].

Focusing on the presented post-processing results, the oversized building footprint (c) is well corrected and given in (d). In detail, the building length is well corrected from 50.7 m down to 36.9 m (ground truth 36 m). Also the building height (from 9.8 m to 11.4 m – ground truth 13 m) and the height standard deviation (from 4.0 m to 3.3 m) inside the building footprint are strongly improved.

Summarizing the results, there is still room for improvement. On one hand, the processing improvement is possible to achieve better results after the post-processing step or to obtain a higher completeness of building recognition by combining more than only two aspects. On the other hand, the adaptation of this approach for high-resolution airborne data (e) to the new generation of high-resolution space borne data (f) is desirable, especially by regarding these first results of “double line” detection.

4.5.3 Extraction of 3D Information of Individual Objects with Monoscopic Techniques

Monoscopic techniques are a generic methodology best fitted to extract height information on isolated features such as buildings. The monoscopic technique is based on the photo-interpretation of a single

image instead of a pair of images utilised in the stereoscopic approach. For the specific purpose of 3D feature extraction, the third dimension can be derived from auxiliary information that is present in the image. The methods proposed to obtain the height of the 3D features provide an estimation of the height, whose precision depends on the conditions of the image, as it will be detailed in the following paragraphs.

Two methods have been analysed for the estimation of the features in the monoscopic photo-interpretation:

- The shadow of the buildings and the sun elevation. This information allows deriving the height of the buildings.
- Photo-interpretation of the type of building. In this case the height of the building cannot be measured directly in the image, but it is determined by estimation.

These methods utilise very simple algorithms but can be very useful for security applications, where it is more important to provide methods that allow a rapid 3D extraction than obtain extremely accurate results.

4.5.3.1 Estimation of Height of a Feature from the Shadow of the Feature

This method consists in measuring the shadow of a feature (normally a building or any other feature which is high enough to apply this method) in the image. The height of the feature is calculated as a function of the shadow length and sun elevation angle.

The sun elevation angle is usually included in the metadata accompanying the images. The calculation of the height of the feature is performed by solving the trigonometric problem shown in the next figure. This simple method allows extracting features that are high enough to project a shadow that can be measured with precision. The following figure also shows a QuickBird satellite image example from an airport control tower in Madrid.

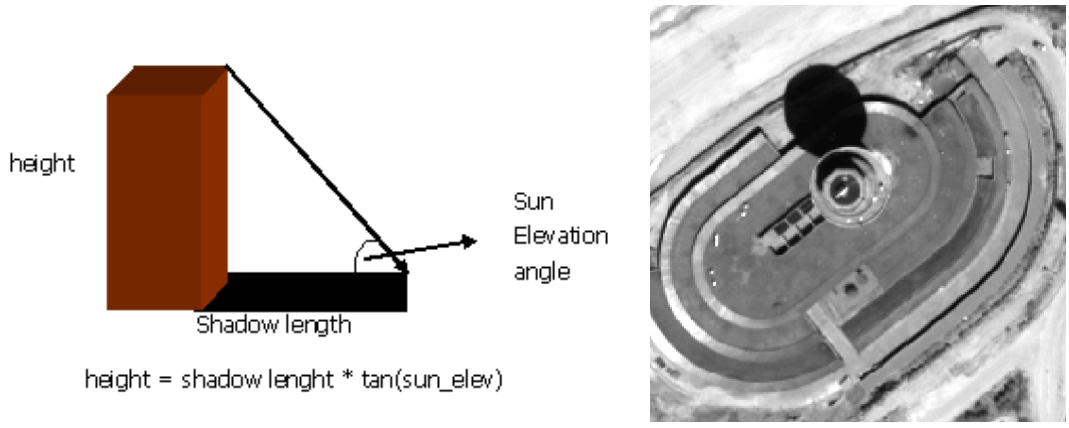


Figure 4-41: (Left) Height Calculation from the Shadow; and (Right) Exemplary Case of Application Under Ideal Conditions, i.e., Measurable Shadow, of an Airport Control Tower.

However, there are typically two situations which jeopardize the application of this technology, i.e., two ways in which shadow cannot be measured: when the shadow is in occultation or when we face a distributed or a too complex object.

When the image acquisition azimuth angle is very similar to the sun elevation, the shadow is occulted by the buildings. In this situation, the sun the target scene and the satellite are aligned and the shadow area is eclipsed by the 3D feature under observation. In the next figure we can find an example of this situation, where the shadow is mostly hidden by the buildings and scarcely can be measured.

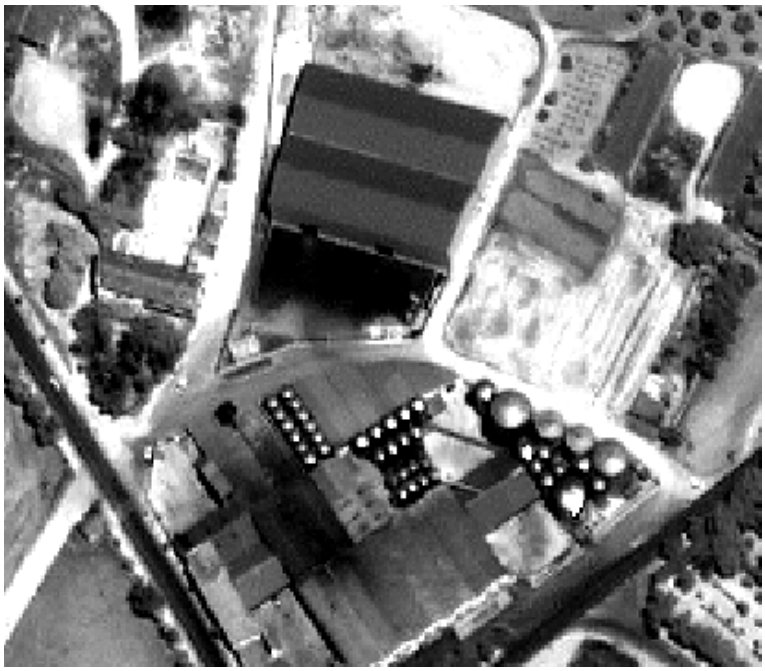


Figure 4-42: Satellite Image with No Visible Shadows.

When the 3D object is very complex and the shadow is not projected over flat terrain, but over another 3D object, the method is not applicable. This is the case when instead of analysing a simple and isolated feature, we analyse a cluster or a distributed object. The shadow is not flat and although it could eventually be measured, the basic geometry model is no longer valid. The following figure shows an example of a compact urban area. Since the houses are aggregated into blocks, information cannot be extracted on single buildings since there are no measurable projected shadows. It is only possible to extract information at a block level. The figure also shows the automatic borders extraction (a typical step used for the extraction of the shadows) where we can see the insurmountable difficulty to interpret individual shadows.

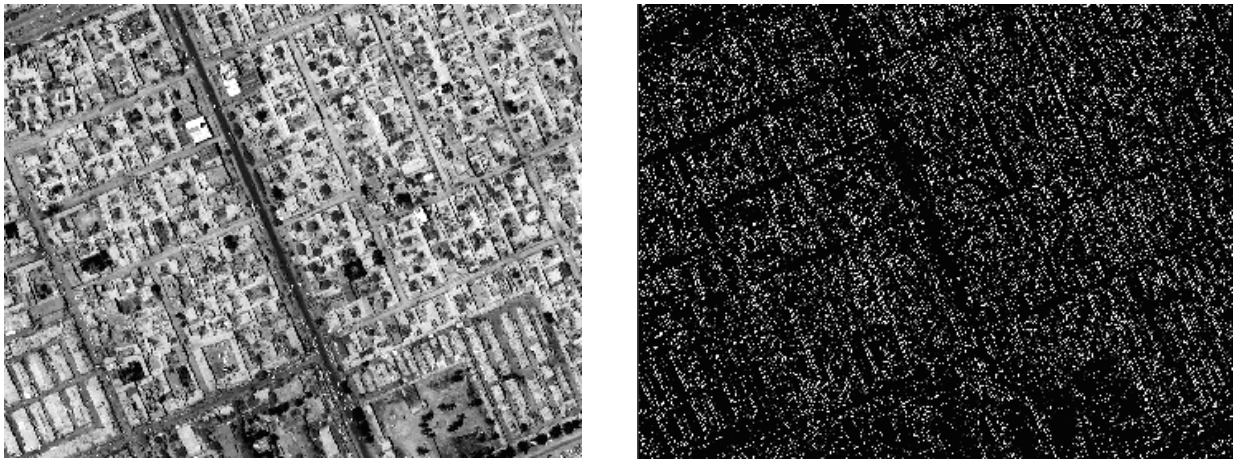


Figure 4-43: Exemplary Situation Where the Shape from Shading Method Cannot be Applied – Left: Satellite image showing complex 3D urban structure where the buildings are aggregated into blocks; Right: Border extraction showing the difficulty to identify shadows at building level. Ikonos image courtesy of SET-118 group.

4.5.3.2 Features Height Estimation by Means of Photo-Interpretation

A different approach when no other method can be applied due to the characteristics of the images (no stereo pair is available and the shadow is not visible) is the estimation of the height of the buildings by photo-interpretation of the shape of the building. The result of this method is only valid for applications not requiring a very accurate measure of the height. This is a fully man assisted method where a photointerpreter needs to identify the outline of the building and estimate the height regarding the shape of the building. For example, the building shown in the following picture can be estimated to have four floors, therefore the height can be estimated to be: 4 + (4 * 3) = 16 meters.

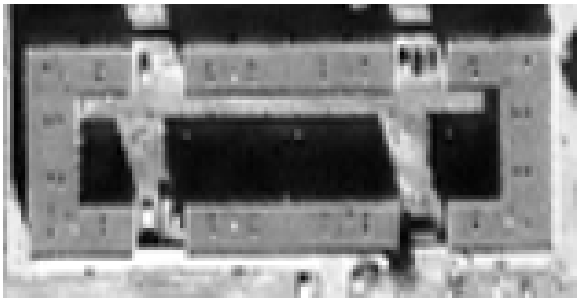


Figure 4-44: Building Example Where the Height is Extracted from Image Knowledge.

The error originated by this method can be very high but, when there is no other available information to estimate the height, it can be enough for certain applications, such as, for instance, the 3D scenarios for simulation applications.

4.5.3.3 Pros and Cons

The advantages of the monoscopic approach are the following:

- It is easily applicable to satellite images where very accurate sun angles measurements are available in the metadata.
- The 3D data extraction is very easy with this method and no specific training is required.
- Specific tools can be developed to improve the extraction of 3D features applying this method.

The disadvantages are:

- The accuracy of the extracted information is not very high; therefore this can limit some applications. This method may be acceptable for the generation of 3D scenarios for simulation where not very high accuracy is required.
- This method cannot be applied to all satellite images: only those images having specific conditions in the sun angles and acquisition angles.

As far as the photo-interpretation method concerned, it requires a previous knowledge of the structures of the buildings that appear in the images in order to be able to estimate the height. The accuracy of the height estimation is low, but it can be enough for applications that require rapid 3D extraction methods and where accuracy is not a relevant requirement to be satisfied.

Chapter 5 – ASSESSMENT OF 3D MODELS

The assessment of 3D urban models is currently accomplished largely either by application of classical metrics or by vision-based judgments made by individuals or groups of human beings. These metrics are useful for identifying deficiencies of acquisition, to compare different models and to check conformance with specifications [82]. These specifications are currently often geometric issues such as positional accuracy and completeness but will eventually include suitability for fulfilment of various human goals. Automatic evaluation of the quality together with the knowledge about the input data and relevant meta-information, for example, the purpose of the acquisition, are increasingly needed in decision processes about terrain and terrain models.

Classical mathematical metrics, especially the so-called Sobolev norms, which include the widely used quadratic or rms metric (the "root-mean-square" metric, also called the L_2 metric or norm), have often been used as measures of geometric approximation [83],[84],[85]. Weighted quadratic metrics have had considerable success in terrain modelling [86]. Recently, computational and analytical evidence for use of the Hausdorff metric [87], which is related to the L_∞ metric, has arisen. Overlap of volume bounded by terrain surfaces has been used with considerable success [88],[82] for urban terrain. In this chapter, we first review methods that are currently in use and then describe potential developments in the future. The classical metrics, widely used and useful as they are, generally do not provide good measures of characteristics that correspond to the goals of the human user/observer. In Section 5.1, we provide information about the procedures in current use. In Section 5.2, we provide information about human-goal-based metrics that could be created in the future.

5.1 CURRENT METRICS FOR QUANTITATIVE ASSESSMENT

5.1.1 L_p Metrics

The L_p metrics (norms) used to measure how well a function f (for example, a 2.5D surface in a model of urban terrain) approximates a function g ("ground truth") have been adopted from classical engineering. These metrics have the advantages of being based on well-understood theory and being applicable to a vast number of situations in classical physical science and engineering. In the L_p metric (see [89]), from which much of the material in this sub-section is taken), the difference d_p between two functions f and g on a domain D is:

$$d_p = \left\{ \int_D |f-g|^p \, \mathbf{d}D \right\}^{1/p} \tag{5.1}$$

when $1 \leq p < \infty$ and:

$$d_\infty = \sup_D |f-g| \tag{5.2}$$

when $p = \infty$. The widely used rms metric (quadratic metric) is a discrete version of the L_2 metric.

Consider a flat surface g such as depicted in Figure 5-1. Let f be the same flat surface in most places but have a long thin ridge (for example, a fence or thin wall), as depicted in Figure 5-2. For thin ridges, the L_p metric, $1 \leq p < \infty$, of the difference between f and g will be small. Nevertheless, in human perception, f is generally not considered a good approximation of g. One important aspect of human perception is visibility. The visibility between observer-target pairs on a surface f with a long thin ridge is very much different from the unobstructed visibility on the flat surface g. For $1 \leq p < \infty$, L_p metrics for the difference between f and

g can thus be small when the human-perceived difference between f and g is large. When g is a flat surface and f is the same surface with additional small-amplitude oscillation, the L_∞ metric of the difference between f and g is small when the human-perceived difference is large. The L_p metric, $1 \leq p \leq \infty$, of the difference can also be large when the human-perceived difference is small, for example, when g is a Heaviside function (one value on one side of a line or curve and some other value on the other side) and f is the same function but with a slight shift in the location of the discontinuity. Heaviside functions represent sides of buildings and cliffs and are common in urban and natural terrain. The L_p metrics and all commonly used classical metrics give equal weight to equal amounts of undershoot and overshoot. However, this equal weighting does not match human goals well because "too high" (overshoot) and "too low" (undershoot) are not opposites of each other when visibility, drainage, communication (radio, optical, etc.) and many other human goals are under consideration.

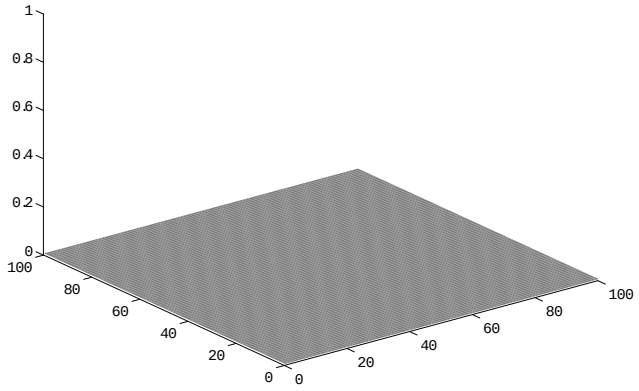


Figure 5-1: Flat Surface.

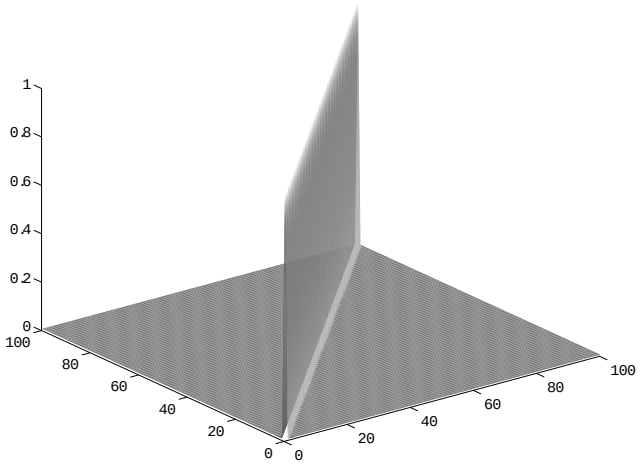


Figure 5-2: Flat Surface with Long Thin Ridge.

5.1.2 Hausdorff Metric

Metrics other than L_p metrics have been used to measure goodness-of-fit in geometric modelling. Petukhov [90] carried out analysis in the Hausdorff metric (the maximum of the minimum distances between the surfaces), but the subject area goes back further to Sendov and the Bulgarian school of approximation [96] To make this precise the maximum distortion $d_H(A,B)$ of two sets A and B is computed

as the larger of the distances of the further-most point of A from the set B and the furthermost point of B from the set A:

$$d_H(A,B) = \max(\max_{x\in A} dist(x,B), \max_{y\in B} dist(y,A))$$

Two surfaces S_1 and S_2 satisfying $d_H(S_1,S_2) < \varepsilon$ means that each point of S_1 lies within an ε - bubble around S_2 and similarly each point of S_2 lies within an ε - bubble around S_1. The relationship of the Hausdorff metric in 3D to the rms metric L_2 and the uniform metric L_∞ is illustrated for 2D cross-sections in Figure 5-3.

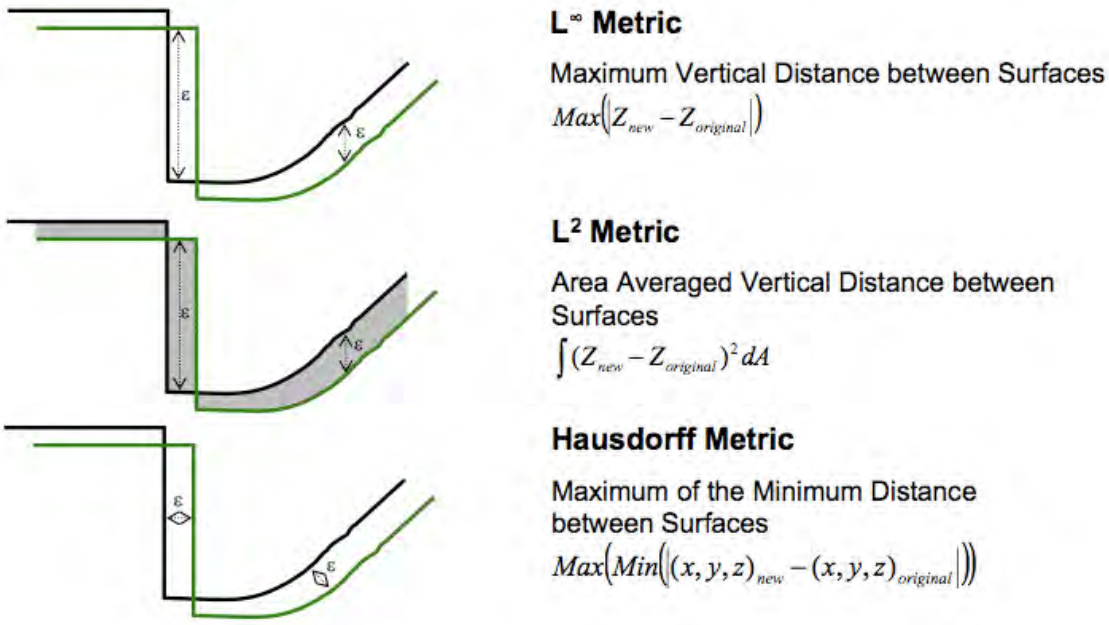


Figure 5-3: ε Distance Measurements between Black and Green Surfaces in the L_∞, L_2, and Hausdorff Metrics, Respectively. For clarity, only cross-sections of the two surfaces are illustrated.

Observe from this figure that, in the uniform metric (L_∞), the differences of the surfaces are measured only in the vertical direction. In this case, if a footprint of a building is perturbed, no matter how small, the error becomes the full height of the building. The rms metric computes an averaged value of these same vertical errors between the two surfaces, i.e., the square of the vertical errors is integrated over the region of interest. On the other hand, the Hausdorff metric estimates the worst distortion between the curves, but not biased to any direction. There are many variations of Hausdorff metrics, including averaged L_p versions of the closest distances between the surfaces for which an approximation theory has been developed [92]. A generalization, the Hausdorff-Gromov metric, allows for incorporation of isometric distortions [93] to be applied to either surface before computing the Hausdorff error which may be useful for SFM estimation. A related, but different distortion metric has been developed by Guibas and his collaborators [94]. This metric was named Earth Mover's Distance since it may be interpreted as the work required transporting mass from one place to another.

To illustrate with terrain mapping the effects of different metrics, we refer to a report by Thies et al. [95] where these three metrics are used as the distortion metrics in best approximating an urban terrain with a fixed budget and computing the corresponding regions of line of sight or "viewsheds". The computational experiment used a section of a high quality DTED terrain map of Baltimore as ground truth and computed the best approximate surfaces of the terrain by selection of the optimal 5,000 vertex heights to reduce the

respective metrics. Viewsheds for these three test surfaces, each produced by minimizing the corresponding metric, were computed at 20 uniformly spaced positions proceeding north along a main road. This urban terrain and route are illustrated in the cartoon of Figure 5-3. The line of sight region was then determined for each of the 20 locations along the road at a common height of 3 meters above the road surface with the results recorded in Figure 5-4. These results simulate a convoy proceeding along the road.

Figure 5-4: Baltimore Height Data Used to Estimate Distortions of Viewshed Using Different Metrics. The orange line depicts the path of a convoy traveling north on a main road.

In Figure 5-5, the resulting viewsheds are rendered for the rms (L_2), uniform (L_∞) and Hausdorff metrics. In these images, the areas coloured yellow are false positives, that is, points seen on the metric-approximated surface but are not visible from the convoy's route on the true surface. The areas coloured red are false negatives, that is, points that are not seen on the approximate surface but that are seen on the true surface. Grey denotes the region that can be seen on both the approximate and the true surface; black signifies the terrain that cannot be seen on either surface.

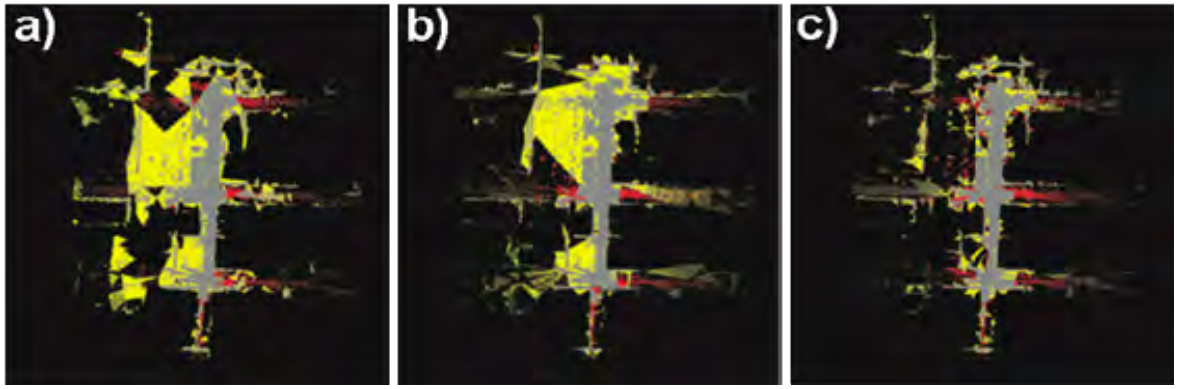


Figure 5-5: Correct and Erroneous Portions of Viewsheds for Convoy Route in Baltimore – a) L_2 Metric; b) L_∞ Metric; and c) Hausdorff Metric. See text for explanation.

5.1.3 Quality Measures Based on Overlap of Volume

Metrics such as the root mean square (rms) metric cause problems when applied to 3D terrain with complex structures [82] due not merely to lack of rationale for application of the rms metric in this situation but also to a point matching problem (which points on the models to be compared should indeed be compared with each other). In the literature, there exist a number of alternative quality measures. These have been derived for different tasks and can partially be converted in each other. Selection criteria for the quality measures are [88]:

1) The values of the quality measures should be *reliably computable* with *moderate technical effort*. Special cases that can occur with empty model sets should be avoided.

2) *Evaluations in 2D and 3D* should be possible without any change of methodology.

3) The quality measures should have a *limited range*. Measures with a range between zero and one can be interpreted immediately as percentages. They allow comparison of the results for different model assessments.

4) The quality measures should *be easily interpretable*. Concerning conformance with a specification, it should be possible to give a statement such as “10% of the urban terrain has not been acquired”.

5) The values of the quality measures should be *independent of the volume*. The volume to be rastered is defined by a bounding box for each data set. Since this cuboid is often aligned with the coordinate axes, the volume depends on the coordinate system. A complement to the data set or a rotation of the data set can change the volume.

6) The quality measures should be locally and global identical and their values locally and global computable. This allows for instance the comparison of a local measure value for a single building with the corresponding value for the entire model, which can be seen as an average value. It should be possible to carry out the interpretation of the results hierarchically. The global 2D and 3D measures can first be taken into consideration. If their values are insufficient for the fulfilling the need, local measure values can be used.

7) The quality measures should be invariant with respect to conjoint translations and rotations of the test and reference model.

The quality measures used in the following are based on a test model set T and a reference model set R. Quality measures in the literature that satisfy the criteria specified above include:

The *quality rate*

$$\rho_q = \frac{|T \cap R|}{|T \cup R|} \qquad \rho_q \in [0,1] \tag{5.3}$$

The value of the quality rate is independent of the assignment of the test and reference data set (symmetry). Its optimum value is 1.

The *detection rate*

$$\beta = 1 - \frac{|R \setminus T|}{|R|} \qquad \beta \in [0,1] \tag{5.4}$$

measures the proportion of buildings or building parts correctly detected. Its optimum value is 1.

The *branch factor*

$$\rho_b = \frac{|T \setminus R|}{|T \cap R|} \tag{5.5}$$

measures the proportion of objects falsely detected. Its optimum is 0.

The *miss factor*

$$\rho_m = \frac{|R \setminus T|}{|T \cap R|} \tag{5.6}$$

measures the proportion of objects not acquired. Its optimum is 0.

Measures (5.5) and (5.6) do not fulfil Criterion 3, but are nevertheless easily interpreted. These measures can be computed for whole scenes in 2D or 3D and can also be computed for individual buildings. The quality rate (5.3) can be computed with weights, for example, those resulting from a distance transformation.

To illustrate the applicability of these measures, Figure 5-6 shows two building models obtained by two different acquisition methods, cf. [88]. The assignment of test and reference is irrelevant for this example, since no information about the accuracies is a priori known. For the comparison of the test and reference data sets intersection and union sets have to be determined. In principle these calculations can be done vectorially. However, in the case of volume determinations, these calculations and the corresponding data structures are rather complex. A voxel-based approach, that is, a spatial enumeration in cells or voxels, was chosen to circumvent these problems and to suppress the effects of different topologies. The precision of the quality measures depends on the spatial resolution defined by the sizes of the volume cells. This approach has the advantage of allowing the evaluation in 2D (position) or in 3D (position and height) by using the same quantitative quality measures and methods.

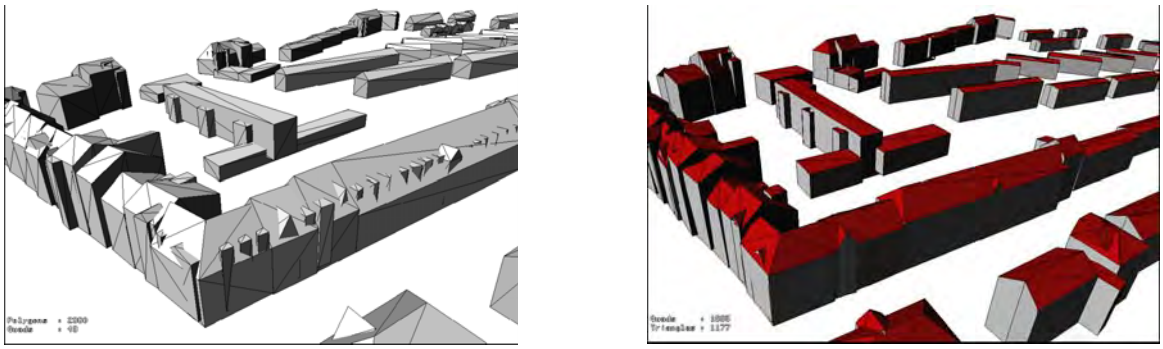


Figure 5-6: Building Models Obtained by Two Different Acquisition Methods.

Both data sets were rastered with a cell size of 0.5 m. After this screening, the differences, unions and intersections of the volumes were calculated and the connected components determined. Figure 5-7 shows an example of a screening and a difference set.

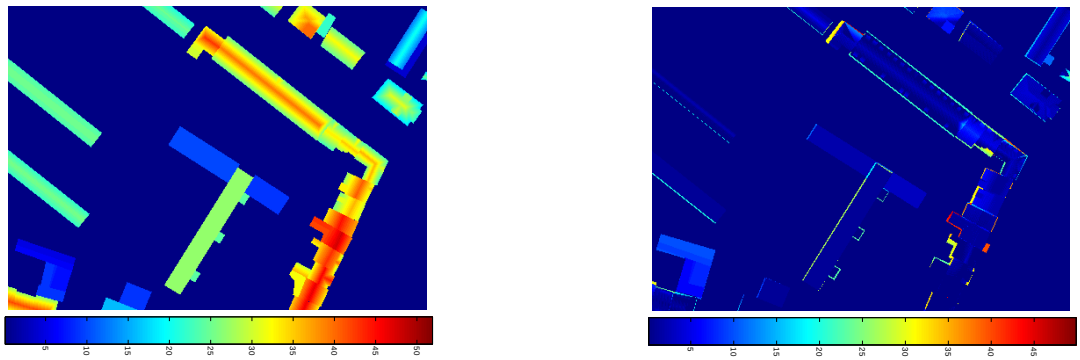


Figure 5-7: (Left) Detail of the Screened Test Data Set T and (Right) Difference Set T\R.

The quality measures listed above in this section, computed for this example for both 2D ("footprint") and 3D, are listed in Table 5-1.

Table 5-1: Quality Measures (Globally Calculated) for the Two Models of Figure 5-6.

Quantity	2D	3D	Range	Optimal Value
Quality Rate	0.87	0.75	[0,1]	1
Miss Factor	0.10	0.24	≥0	0
Branch Factor	0.05	0.10	≥ 0	0
Detection Rate	0.96	0.91	[0,1]	1

The evaluation can be carried out on a per-unit basis. Figure 5-8 shows the results for the quality rate computed for buildings or building parts.

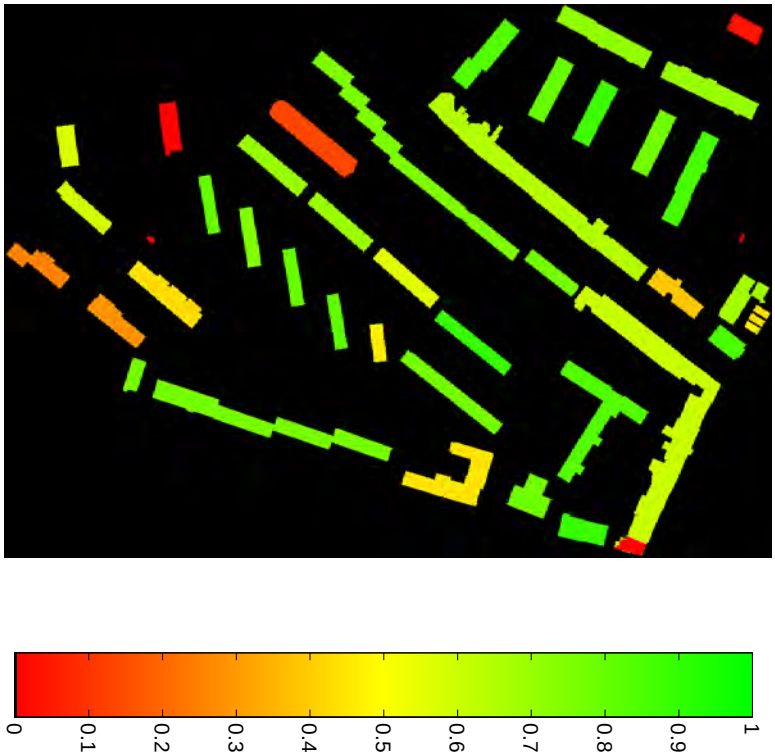


Figure 5-8: Diagnostics by Visualisation of the Local Quality Rates, Computed in 3D.

5.2 FUTURE HUMAN-GOAL-BASED METRICS FOR QUANTITATIVE ASSESSMENT

The metrics described above provide information about accuracy and other properties of models. However, these metrics are only incompletely linked with human goals. This is particularly true of the widely used L_p metrics.

If a metric for the difference between two functions is intended to be part of a human decision-making process, then this metric should be directly linked to the human goals in this process. Unfortunately, metrics that fully express what is important in human goals have not yet been developed. There are many different human goals, including accuracy of visibility, that could be bases for measuring the difference between two surfaces. In [89], from which much of the material in this section is taken, options for defining metrics that express accuracy of visibility and more general human goals are described.

The extent to which the visibility regions (line-of-sight regions) of two surfaces coincide is an issue related to a human goal that is important in many situations. We define here a visibility function and two metrics based on visibility. Let there be given a height function $\varphi(x,y)$ for (x,y) in some 2D domain D. Let there also be given two 3D domains, O and T, at the points of which "observers" and "targets", respectively, are located. For example, when observers and targets are humans, unmanned ground vehicles and unmanned aerial vehicles that are always on or above φ and always at or below a certain height H, the domains O and T would both be:

$$\{ (x , y , z) \mid (x , y) \in D , \varphi (x , y) \leq z \leq H \} \tag{5.7}$$

A target at a point $(\hat{x}, \hat{y}, \hat{z})$ in T is visible to an observer at a point (x,y,z) in O if the (open) line segment from (x,y,z) to $(\hat{x}, \hat{y}, \hat{z})$ is always above the surface φ, that is:

$$z + t(\hat{z} - z) > \varphi(x + t(\hat{x} - x), y + t(\hat{y} - y)), \quad 0 < t < 1 \tag{5.8}$$

In this case, we say that there is "line of sight" from (x,y,z) to $(\hat{x}, \hat{y}, \hat{z})$.

Define a point-to-point visibility function:

$$\begin{aligned} \mathrm{vis}_{\varphi}(x, y, z; \hat{x}, \hat{y}, \hat{z}) &= 1 \text{ if point } (\hat{x}, \hat{y}, \hat{z}) \text{ is visible from point } (x,y,z) \\ &= 0 \text{ if point } (\hat{x}, \hat{y}, \hat{z}) \text{ is not visible from point } (x,y,z) \end{aligned} \tag{5.9}$$

Based on this point-to-point visibility function, we can define a visibility metric V for the surface φ and for an observer at (x,y,z) in O to be the integral of vis_{φ} over T, that is:

$$\mathrm{V}_{\varphi;T}(x, y, z) = \iiint_{T} \mathrm{vis}_{\varphi}(x, y, z; \hat{x}, \hat{y}, \hat{z})\, \mathrm{d}\hat{x}\, \mathrm{d}\hat{y}\, \mathrm{d}\hat{z} \tag{5.10}$$

Let there be given two height functions, a model $f(x,y)$ and "ground truth" ("real terrain") $g(x,y)$. For an observer at (x,y,z), we define the difference-of-visibility or "DV" metric between functions f and g to be:

$$\mathbf{DV}_{f,g;T}(x, y, z) = \iiint_{T} \left| \begin{array}{l} \mathbf{vis}_{f}(x, y, z; \hat{x}, \hat{y}, \hat{z}) \\ -\mathbf{vis}_{g}(x, y, z; \hat{x}, \hat{y}, \hat{z}) \end{array} \right| \mathbf{d}\hat{x}\, \mathbf{d}\hat{y}\, \mathbf{d}\hat{z} \tag{5.11}$$

If the DV metric is small (large), visibility on f does (does not) closely approximate visibility on g. In practice, the ground truth g that appears in vis_g in the DV metric is rarely known completely. Only data points on and perhaps some other information (derivatives at some points, monotonicity, drainage patterns, etc.) about g are known. Hence, an exact DV metric will typically have to be replaced by an approximate DV metric.

There are many extensions. The point-to-point visibility function vis_{φ} assumes that all visible targets are equally visible, no matter how distant they are from the observer. With telescopic lenses, this assumption is reasonable in many circumstances. However, when telescopic lenses are not available (for example, in low-cost sensors) and when visibility decreases with distance (for example, due to haze or smog in the atmosphere), it is appropriate to introduce in the DV metric a weighting function w that decreases as the distance of the target from the observer increases, perhaps at different rates in different regions and directions. In this case, the point-to-point visibility function $\mathrm{vis}_{\varphi}(x, y, z; \hat{x}, \hat{y}, \hat{z})$ for a surface φ can be defined to be:

$$\begin{aligned} \mathrm{vis}_{\varphi}(x, y, z; \hat{x}, \hat{y}, \hat{z}) &= w(x, y, z; \hat{x}, \hat{y}, \hat{z}) \text{ if point } (\hat{x}, \hat{y}, \hat{z}) \text{ is visible from point } (x,y,z) \\ &= 0 \text{ if point } (\hat{x}, \hat{y}, \hat{z}) \text{ is not visible from point } (x,y,z) \end{aligned} \tag{5.12}$$

When a telescopic lens is not available but the atmosphere is clear, one could choose:

$$w(x, y, z; \hat{x}, \hat{y}, \hat{z}) = \frac{1}{(x - \hat{x})^2 + (y - \hat{y})^2 + (z - \hat{z})^2} \tag{5.13}$$

to express the fact that the visually recorded area of the target as seen by the observer is inversely proportional to the square of the distance from the observer to the target.

When the observer position is variable, it may be best to use "global" metrics. One can define a global visibility metric GV for a function φ to be the integral of V over O, that is:

$$\mathrm{GV}_{\varphi;O,T} = \iiint_O \left[\iiint_T \mathrm{vis}_\varphi(x,y,z;\hat{x},\hat{y},\hat{z})\, d\hat{x}\, d\hat{y}\, d\hat{z} \right] \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{5.14}$$

One can define a global difference-of-visibility metric GDV between functions f and g to be the integral of DV over O, that is:

$$\mathrm{GDV}_{f,g;O,T} = \iiint_O \left[\iiint_T \left| \begin{matrix} \mathrm{vis}_f(x,y,z;\hat{x},\hat{y},\hat{z}) \\ -\mathrm{vis}_g(x,y,z;\hat{x},\hat{y},\hat{z}) \end{matrix} \right| \mathrm{d}\hat{x}\, \mathrm{d}\hat{y}\, \mathrm{d}\hat{z} \right] \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{5.15}$$

As defined in (5.11) and (5.15), the DV and GDV metrics weight all observer and target points equally. In many cases, users may wish to weight some regions in O and T differently from other regions to reflect the importance of accurate modelling in those regions. This can be done by introducing weight functions in the integrands of the DV and GDV metrics. Analogous adjustments can, of course, be made in the integrands of the visibility metric V of (5.10) and the global visibility metric GV of (5.14).

The difference-of-visibility metrics DV and GDV treat false negatives (f predicts no visibility when g has visibility) and false positives (f predicts visibility when g does not have visibility) equally and provide sums of these two types of error. However, there are many circumstances in which one wishes to measure only one of these types of error, not both together. For example, for concealing unsightly civilian assets such as garbage dumps and for concealing many military assets, false negatives are serious while false positives are less so. On the other hand, for emplacing optical communication nodes under friendly conditions, false negatives are generally not serious. They merely increase costs marginally by reducing the options in the planning stage. However, false positives are serious because they result in constructing nodes in locations where the system cannot function and thereby increase costs dramatically. Difference-of-visibility functionals that measure only false negative error or only false positive error are easily constructed. For example, a difference-of-visibility metric for false negative error is:

$$\iiint_T \max\left\{ \begin{matrix} \mathbf{vis}_g(x,y,z;x,y,z) \\ -\mathbf{vis}_f(x,y,z;x,y,z) \end{matrix}, 0 \right\} \mathbf{d}\hat{x}\, \mathbf{d}\hat{y}\, \mathbf{d}\hat{z} \tag{5.16}$$

and a difference-of-visibility metric for false positive error is:

$$\iiint_T \max\left\{ \begin{matrix} \mathbf{vis}_f(x,y,z;x,y,z) \\ -\mathbf{vis}_g(x,y,z;x,y,z) \end{matrix}, 0 \right\} \mathbf{d}\hat{x}\, \mathbf{d}\hat{y}\, \mathbf{d}\hat{z} \tag{5.17}$$

The sum of these two errors is the total difference-of-visibility error DV.

We have worked with functions $f(x, y)$ and $g(x, y)$ that represent univalent height fields or "terrain skins". The extension of the techniques described here to fully 3D terrain (with, for example, space underneath bridges, interiors of buildings, subway tunnels, etc.) is computationally intensive but conceptually straightforward.

The emphasis here has been on visibility as a metric because:

1) Measuring visibility is a common and important military and human goal; and
2) Quantitative metrics for visibility can be defined in ways, such as those described above, that human beings recognize are directly connected with the human goal.

However, the approach here is not limited to visibility and extends to many other situations. Discovering the metrics by which human beings judge similarities and differences in all areas of human interest, the metrics by which they extract features and the metrics by which they understand perceptual cues will require long-term interdisciplinary research by mathematicians, statisticians, cognitive scientists and domain experts. While classical mathematical metrics may on occasion be found to be appropriate, it is likely that most of the metrics related to human goals will be quite different from the metrics that have been commonly used in the past.

The DV and GDV metrics are computationally intensive. Whether one can find computational procedures for calculating these metrics that have acceptably low computational cost is an open question. The DV and GDV metrics for determining the difference in visibility between exact terrain and an approximate model will generally have to be computed in some approximate manner. In cases where the computational cost of the metric is too large, the question of approximate ersatz metrics will arise. This question in turn leads to a question of metametrics, that is, “metrics of metrics” that measure how well one metric approximates another metric. In the past, smoothness has been virtually the only pattern used in approximation theory. However, there are equally strong patterns – ones not related to smoothness – in many classes of non-smooth functions. Human beings instinctively recognize cities and areas of cities by these patterns. These patterns are potential guideposts on the way to developing a fully non-linear approximation theory with non-smooth metrics of non-smooth functions on non-smooth manifolds that is applicable to terrain and many other modern areas of interest.

Besides being ways of determining how accurate the visibility of a given model f is, the DV and GDV metrics and other human-goal-based metrics are potential bases for generation of optimal models f, that is, of functions f that minimize the metrics over an appropriate function space or manifold. This minimization will be challenging, since the metrics need not be smooth. For example, the DV and GDV metrics are only C^0 continuous with respect to f even when f varies smoothly. Utilizing the structure of terrain in whatever algorithm is chosen will likely be essential for computational efficiency.

Chapter 6 – CONCLUSIONS AND RECOMMENDATIONS

In the present report, we have summarized recent research and development, including but not limited to that carried out by NATO SET-118, and outlined future directions in 3D modelling of urban terrain. NATO SET-118 has provided value by facilitating coordination of this research and development in many dimensions, including sensor systems for generating the data, analytical and computational techniques for creating models from the data and metrics for assessing the models.

The information presented in this report indicates that the scientific and administrative factors to support rapid progress in research and development of 3D modelling of urban terrain are all in place. What is now needed is coordination among organizations (international, national, academia, industry) and coordination of near-term development and long-term research.

6.1 RECOMMENDATIONS

Comprehensive, internationally coordinated research and development programs are required to promote rapid progress. Specific recommendations are as follows:

- NATO member countries, especially those that participated in SET-118, should fund research and development programs with the objective of producing one or more theoretically founded prototype systems for compressed full 3D modelling of urban terrain within 5 years and one or more working operational system within 10 years. Compression factors that are needed are two to three orders of magnitude greater than the compression factors of the order of 10 to 20 that are available today.
- NATO member countries should continue to provide broad support for academic and industrial efforts in 3D modelling of urban terrain as well as in linkage of 3D modelling with geographic information systems and other systems that include items/factors beyond geometry, topology and texture.
- The strongly interdisciplinary nature of 3D modelling of urban terrain should be reflected in all efforts supported by NATO member countries.

6.2 A PATH TO THE FUTURE

Due to the shift of defence operations from classical, symmetric scenarios to current-day asymmetric, urban scenarios, the need for more accurate and comprehensive 3D models of urban terrain has strongly increased and is likely to continue to increase in the near and medium-term future. The many different directions of research and development in 3D modelling of urban terrain are evidence of the richness of this field. While all of these directions are important and interesting in their own right, the task is now to coordinate and focus basic research and development efforts in areas that will quickly lead to creating a class of models with acceptable or better accuracy and computational speed. The bottom line in security 5 – 20 years from now will depend on the vigour with which this research and development is pursued today.

Chapter 7 – RTG PARTICIPANTS AND MEETINGS

7.1 RTG PARTICIPANTS

- Meidow, Jochen (chair), Fraunhofer IOSB, Germany
- Thönnessen, Ulrich (chair, retired), Fraunhofer IOSB, Germany
- Barber, David, Dstl, Geospatial Intelligence Team, Information Management Department, United Kingdom
- Bennett, Andrew, Dstl Sensors and Countermeasures Department, United Kingdom
- Carrasco Diaz, Daniel, Remote Sensing Department Indra Espacio, S.A., Spain
- Fournier, Jonathan, DRDC Valcartier, Tactical Surveillance and Reconnaissance Section, Canada
- Hansen, Wolfgang von, FGAN-FOM, Germany
- Lavery, John, U.S. Army Research Office, USA
- McEwan, Kenneth, Dstl, Sensors & Countermeasures Department, United Kingdom
- Pasquariello, Stefano, SELEX Sistemi Integrati, Italy
- Ricard, Benoît, DRDC Valcartier, Canada
- Richmond, Richard, Air Force Research Laboratory Sensors Directorate, USA
- Sharpley, Robert C., University of South Carolina, Department of Mathematics, Industrial Mathematics Institute, USA
- Tolt, Gustav, Swedish Defence Research Agency FOI, Sweden

7.2 MEETINGS

1) 2007, April 22-24, DRDC, Quebec City (CAN)
2) 2007, October 9-11, Indria Espacio, Madrid (ESP)
3) 2008, April 22-24, Elsag Datamat, Rome (ITA)
4) 2008, November 17-20, University of South Carolina, Columbia, SC (USA)
5) 2009, April 27-29, Ordnance Survey, Southampton (GBR)
6) 2009, November 3-5, Fraunhofer-FOM, Ettlingen (DEU)
7) 2010, April 20-22, Virginia Tech, Blacksburg, VA (USA)

Chapter 8 – REFERENCES

[1] H.A. Angel, L.J. Massel and P.G.S. Vilhena (2005) *Reconnaissance Information Transfer in Urban Operations. Contract Report*, DRDC Toronto CR 2005-044.

[2] J. Bellingham, Y. Kuwata and J. How (2003) *Stable Receding Horizon Trajectory Control for Complex Environments*. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Austin, TX, USA, August 2003.

[3] Y. Kuwata and J. How (2004) *Three Dimensional Receding Horizon Control for UAVs*. In AIAA Guidance, Navigation, and Control Conference, Providence, RI, USA, 16-19 August 2004.

[4] A. Kurdila, M. Nechyba, R. Lind, P. Ifju, P. Binev, W. Dahmen, R. DeVore and R. Sharpley (2004) *Vision-Based Control of Micro-Air-Vehicles: Progress and Problems in Estimation*. 43rd IEEE Conference on Decision and Control, Paradise Island, Bahamas, December 2004, pp. 1636-1642.

[5] R.C. Smith and P. Cheeseman (1986) *On the Representation and Estimation of Spatial Uncertainty*. The International Journal of Robotics Research **5** (4), pp. 56-68.

[6] R.C. Smith, M. Self and P. Cheeseman (1986) *Estimating Uncertain Spatial Relationships in Robotics*. In Proceedings of the Second Annual Conference on Uncertainty in Artificial Intelligence. UAI ‘86. University of Pennsylvania, Philadelphia, PA, USA: Elsevier, pp. 435-461.

[7] J.J. Leonard and H.F. Durrant-Whyte (1991) *Simultaneous map building and localization for an autonomous mobile robot.* Intelligent Robots and Systems’ 91.’Intelligence for Mechanical Systems, Proceedings IROS’91. IEEE/RSJ International Workshop, pp. 1442-1447.

[8] www.fujifilm.com/products/3d/camera.

[9] O. Faugeras and Q.-T. Luong (2001) *The Geometry of Multiple Images*. MIT Press, Cambridge.

[10] J.C. McGlone, E.M. Mikhail and J. Bethel (2004) *Manual of Photogrammetry*. American Society of Photogrammetry and Remote Sensing, 5th edition.

[11] R. Hartley and A. Zisserman (2000) *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge.

[12] B.T. Lucas and T. Kanade (1981) *An Iterative Image Registration Technique with an Application to Stereo Vision*. In Proc. of Image Understanding Workshop, pp. 212-130.

[13] D.G. Lowe (2004) *Distinctive Image Features from Scale-Invariant Key-points*. In International Journal on Computer Vision (IJCV), Vol. 60, No. 2, pp. 91-110.

[14] C. Beder and R. Steffen (2008) *Incremental estimation without specifying a-priori covariance matrices for the novel parameters*. In Proceedings of VLMP Workshop 2008, Anchorage, AK, USA, June 2008.

[15] C.G. Harris and M.J. Stephens (1988) *A combined corner and edge detector*. In Proceedings of Fourth Alvey Vision Conference, pp, 147-151.

[16] W. Förstner and E. Gülch (1987) *A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features*. In ISPRS Intercommission Workshop, Interlaken, June 1987.

[17] H. Hirschmüller (2005) *Accurate and efficient stereo processing by semi-global matching and mutual information*. In Proceeding of Computer Vision and Pattern Recognition (CVPR) 2005, Volume 2, pp. 807-814, IEEE.

[18] C. English, A. Deslauriers and I. Christie (2005) *The complementary nature of triangulation and ladar technologies*. In Laser radar technologies and applications X, 2005, Vol. 5791, pp. 29-41, SPIE.

[19] M. Adams (2002) *Coaxial range measurement – current trends for mobile robotic applications*, IEEE Sensors Journal, Vol. 2, No. 1, pp. 2-13.

[20] C.S. Fox (1996) *Active Electro-Optic Systems* (Volume 6). The Infrared and Electro-Optical Systems Handbook, SPIE Press.

[21] Airborne lidar Sensors, GIM International, Vol. 23, No. 2, pp. 16-19, 2009.

[22] http://airbornehydro.com/HawkEyeII/hawkeyeII.html.

[23] http://www.optech.ca/orion.htm.

[24] http://rieglusa.com/products/airborne/lms-q680/index.shtml.

[25] http://www.topsys.com.

[26] http://www.infoterra.co.uk/data_mob.php.

[27] http://www.velodyne.com.lidar.

[28] http://www.sick.com.

[29] G. Pearson, K. Ridley and D. Willets (2005) *Chirp-pulse-compression three-dimensional lidar imager with fibre optics*, Applied Optics, Vol. 44, No. 2, pp. 257-265.

[30] J. Anderson, J. Busek and H. Heisenberg (2005) *Long distance high accuracy 3-D laser radar and person identification*, SPIE Vol. 5791, pp. 9-16.

[31] R. Marino et al. (2005) *High-resolution 3D imaging laser radar flight test experiments*, SPIE Vol. 5791, pp. 138-151.

[32] R. Stettner et al. (2005) *Large format time-of-flight focal plane detector development*, SPIE Vol. 5791, pp. 288-292.

[33] http://www.advancescientificconcepts.com.

[34] R. Morrison et al. (2005) *Urban reconnaissance with an airborne laser radar*, SPIE Vol. 5791, pp. 1-8.

[35] I. Baker, P. Thorne and J. Henderson, (2006) *Advanced multifunctional detectors for laser-gated imaging applications*, SPIE Vol. 6206, Paper 08, pp. 1-10.

[36] I. Baker, D. Owton et al. (2008) *Advanced infrared detectors for multimode active and passive imaging applications*, SPIE Vol. 6940.

[37] http://www.engadget.com/photos/microsofts-project-natal-roots-revealed-3dv-systems.

[38] http://www.xbox.com/kinect.

[39] P. Sévigny, D.J. DiFilippo, T. Laneve, B. Chan, J. Fournier, S. Roy, B. Ricard and J. Maheux (2010) *Concept of operation and preliminary experimental results of the DRDC through-wall SAR system*. SPIE Defence & Security, [7669-6].

[40] B. Chan (2010) *Contourlet domain hidden Markov tree-based detection algorithm for DRDC through-wall synthetic aperture radar system (TWSAR) applications*. SPIE Defence & Security, Orlando, 5-9 April 2010 [7699-11].

[41] C. Mallet and F. Bretar (2009) *Full-waveform topographic lidar: State-of-the-art*, ISPRS Journal of Photogrammetry and Remote Sensing 64, pp. 1-16.

[42] C. Grönwall, O. Steinvall, F. Gustafsson and T. Chevalier (2007) *Influence of laser radar sensor parameters on range-measurement and shape fitting uncertainties*, Optical Engineering 46 (10), 106201, October 2007.

[43] P.J. Besl and N.D. McKay (1992) *A method for registration of 3-D shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (2), pp 239-256.

[44] Y. Liu (2004) *Improving ICP with easy implementation for free-form surface matching*, Pattern Recognition, Vol. 37, No. 2, pp. 211-226.

[45] P. Biber and W. Strasser (2003) *The normal distributions transform: A new approach to laser scan matching*, In Proceedings of the IEEE International Conference on Intelligent Robots and Systems IROS, Vol. 3, pp. 2743-2748.

[46] M. Magnusson, A. Lilienthal and T. Duckett (2007) *Scan Registration for Autonomous Mining Vehicles Using 3D-NDT*, Journal of Field Robotics 24(10), 803-827.

[47] G. Farin (2001) *Curves and Surfaces for CAGD*, Fifth Edition: A Practical Guide, Morgan Kaufmann Series.

[48] J. Fournier, B. Ricard and D. Laurendeau (2007) *Mapping and exploration of complex environments using persistent 3D model*. In Conference on Computer and Robot Vision, pp. 403-410.

[49] H. Zhao, S. Osher and R. Fedkiw (2001) *Fast Surface Reconstruction Using the Level Set Method*, in Proceedings of the IEEE Workshop on Variational and Level Set Methods in Computer Vision (VLSM 2001), July 2001.

[50] H. Zhao and S. Osher (2002) *Visualization, analysis and shape reconstuction of unorganized data sets*, Contributed chapter in S. Osher and N. Paragios, Editors, Geometric Level Set Methods in Imaging, Vision and Graphics, Springer.

[51] S.F. Frisken, R.N. Perry, A.P. Rockwood and T.R. Jones (2000) *Adaptively sampled distance fields: A general representation of shape for computer graphics*, Proc. 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000, pp. 249-254.

[52] Y.-H.R. Tsai, L.-T. Cheng, S. Osher, P. Burchard and G. Sapiro (2004) *Visibility and its dynamics in a PDE based implicit framework*, J. Computational Physics, **199** (2004), 260-290.

[53] L. Rudin, P.L. Lions and S. Osher (2003) *Multiplicative denoising and deblurring: theory and algorithms*. In Geometric level set methods in imaging, vision, and graphics. Editors S. Osher and N. Paragios. Springer-Verlag, New York, USA, pp. 103-119.

[54] S. Osher and R. Fedwick (2003) *Level set methods and dynamic implicit surfaces*, Applied Mathematical Sciences **153**, Springer-Verlag, New York, USA.

[55] M. Elmqvist (2002) *Ground surface estimation from airborne laser scanner data using active shape models*, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Vol. XXXIV, pp. 114-118, Part 3A, ISPRS Commission III Symposium, Graz.

[56] G. Sithole and G. Vosselman (2004) *Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds*, ISPRS Journal of Photogrammetry & Remote Sensing 59, pp. 85-101.

[57] J. Shan and C.K. Toth (2008) *Topographic Laser Ranging and Scanning: Principles and Processing*, CRC Press, Boca Raton, FL, USA.

[58] M. Brandin and R. Hamrén (2003) *Classification of Ground Objects Using Laser Radar Data*, Master Thesis, Linköping University, Sweden.

[59] G. Tolt, Å. Persson, J. Landgård and U. Söderman (2006) *Segmentation and classification of airborne laser scanner data for ground and building detection*, Proceedings of SPIE Defence and Security Symposium, Orlando, FL, USA, Vol. 6214, Laser Radar Technology and Applications XI.

[60] J. Ruhe and J. Nordin (2007) *Classification of Points Acquired by Airborne Laser Systems*, Master Thesis, Linköping University, Sweden.

[61] J. Canny (1986) *A computational approach to edge detection*. In Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-8(6):679-698.

[62] R.O. Duda and P.E. Hart (1972) *Use of the Hough Transformation to Detect Lines and Curves in Pictures.* In Communications of Association for Computing Machinery, 15(1):11-15.

[63] C. Steger (1998) *An Unbiased Detector of Curvilinear Structures*. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 113-125.

[64] R. Wallis (1976) *An approach to the space variant restoration and enhancement of images*. In Proceedings of Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey, CA, USA.

[65] P. Fua and Y. Leclerc (1990) *Model Driven Edge Detection.* In Machine Vision and Applications 3, pp. 45-56.

[66] W. Neuenschwander, P. Fua, G. Székely and O. Kübler (1997) *Ziplock Snakes*. Int. J. Computer Vision 26, 191-201.

[67] A.J. Bennett and D. Blacknell (2003) *The extraction of building dimensions from high resolution SAR imagery*. IEEE Proceedings of the International Radar Conference, pp. 182-187.

[68] C.P. Moate and L. Denton (2006) *SAR image delineation of multiple targets in close proximity*. Proceedings of 6th European Conference on Synthetic Aperture Radar, VDE Verlag.

[69] R.D. Hill, C.P. Moate and D. Blacknell (2006) *Urban scene analysis from SAR image sequences*. Proceedings of SPIE Algorithms for Synthetic Aperture Radar Imagery XIII, Vol. 6237.

[70] M. Jahangir, D. Blacknell, C.P. Moate and R.D. Hill (2007) *Extracting information from shadows.* In SAR imagery. IEEE Proceedings of the International Conference on Machine Vision, pp. 107-112.

[71] F. Xu and Y.Q. Jin (2007) *Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 45, No. 7, pp. 2336-2353.

[72] R. Bolter and F. Leberl (2000) *Detection and Reconstruction of Human Scale Features from High Resolution Interferometric SAR Data*. IEEE Proceedings of the International Conference on Pattern Recognition, pp. 291-294.

[73] R. Bolter (2001) *Buildings from SAR: Detection and Reconstruction of Buildings from Multiple View High Resolution Interferometric SAR Data*. Dissertation, University of Graz, Austria.

[74] U. Soergel (2003) *Iterative Verfahren zur Detektion und Rekonstruktion von Gebäuden in SAR- und InSAR-Daten.* Dissertation, Leibniz Universität Hannover, Germany.

[75] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen and U. Soergel (2007) *Building recognition from multi-aspect high resolution InSAR data in urban area*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 45, No. 11, pp. 3583-3593.

[76] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen and U. Soergel (2007) *InSAR Phase Profiles at Building Locations*. Proceeding of ISPRS Photogrammetric Image Analysis, Vol. 36, Part 3/W49A, pp. 203-208.

[77] A. Thiele, E. Cadario, K. Schulz and U. Soergel (2010) *Analysis of Gable-Roofed Building Signatures in Multi-Aspect InSAR Data*. IEEE Geoscience and Remote Sensing Letters, Vol. 7, No. 1, pp. 83-87.

[78] R. Touzi, A. Lopes and P. Bousquet (1988) *A Statistical and Geometrical Edge Detector for SAR Images*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 26, No. 6, pp. 764-773.

[79] F. Tupin, H. Maitre, J.F. Mangin, J.M. Nicolas and E. Pechersky (1998) *Detection of Linear Features in SAR Images: Application to Road Network Extraction*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 36, No. 2, pp. 434-453.

[80] A. Thiele, J.D. Wegner and U. Soergel (2010) *Building Reconstruction from Multi-Aspect InSAR Data*. In: Soergel (Ed.) Radar Remote Sensing of Urban Areas (Remote Sensing and Digital Image Processing), Springer, 1st Edition, ISBN: 978-9048137500, Chapter 8, pp. 187-214.

[81] M. Schwaebisch and J. Moreira (1999) *The high resolution airborne interferometric SAR AeS-1*. Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition, pp. 540-547.

[82] H.F. Schuster and U. Weidner (2003) *A New Approach towards Quantitative Quality Evaluation of 3D Buildings Models.* In: J. Schiewe, M. Hahn, M. Madden and M. Sester (Eds), Challenges in Geospatial Analysis, Integration and Visualization II, ISPRS Workshop, Stuttgart.

[83] A. Yu Bezhaev and V.A. Vasilenko (2001) *Variational Theory of Splines*, Kluwer Academic/Plenum Publishers, New York, USA.

[84] C. de Boor (2001) *A Practical Guide to Splines*, Revised Edition, Applied Mathematical Sciences, 27, Springer-Verlag, New York, USA.

[85] L.L. Schumaker (1981) *Spline Functions: Basic Theory*. Wiley-Interscience, New York, USA.

[86] M. Garland and P.S. Heckbert (1998) *Simplifying surfaces with color and texture using quadratic error metrics*, Proc. IEEE Visualization, October 19-20 1998, pp. 287-295.

[87] J. Munkres (1999) *Topology*, 2nd Ed., Prentice Hall, NJ, USA.

[88] J. Meidow and H.F. Schuster (2005) *Voxel-based Quality Evaluation of Photogrammetric Building Acquisitions*, ISPRS Workshop on Object Extraction for 3D City Models, Road Databases and Traffic Monitoring (CMRT05), pp. 117-124, Volume XXXXVI, Part 3/W24 of IAPRS.

[89] J.E. Lavery (2006) *Human-goal-based metrics for models of urban and natural terrain and for approximation theory*, Proc. 25th Army Science Conf. (November 2006), Department of the Army, Washington, DC, USA, PO-01.

[90] A. Petukhov (2002) *On approximation of generalized functions by convolutions in the Hausdorff metric*, East J. Approx. Th., 8, 59-94.

[91] Bl. Sendov (1990) *Hausdorff approximations*. Translated and revised from the Russian 1979 edition. Mathematics and its Applications (East European Series), 50. Kluwer Academic Publishers Group, Dordrecht.

[92] Bl. Sendov (1993) *Integral Hausdorff distance*, C.R. Acad. Bulgare Sci. **46** (1993), 21-24.

[93] M. Gromov (1999) *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, Vol. 152, Birkhäuser Boston, Inc., Boston, USA.

[94] Y. Rubner, C. Tomasi and L. Guibas (1998) *A Metric for Distributions with Applications to Image Databases*, Proceedings ICCV (1998), 59-66.

[95] A. Thies, B. Philips, P. Binev, R.A. DeVore, M. Hielsberg, L.S. Johnson, B. Karaivanov, B.A. Lane and R.C. Sharpley (2005) *Smooth, Piecewise-Polynomial Terrain Representation Using Nontraditional Metrics*, STTR Final Report, Schafer Corporation Contract No. W911NF-04-C-0060, U.S. Army Research Office, March 2005.

[96] P. Binev, W. Dahmen, R. DeVore and N. Dyn (2004) *Adaptive approximation of curves*, Approximation theory: a volume dedicated to Borislav Bojanov, 43-57, Prof. M. Drinov Acad. Publ. House, Sofia.

[97] R.D. Richmond and B.J. Evans (2005) *Polarimetric Imaging Laser Radar (PILAR) Program*. In Advanced Sensory Payloads for UAV. Meeting Proceedings RTO-MP-SET-092, Paper 19, 19-1 – 19-14). Neuilly-sur-Seine, France, RTO. Available from: http://www.rto.nato.int/abstracts.asp.

[98] P. Binev, A. Cohen, W. Dahmen, R. DeVore and V. Temlyakov (2005) *Universal algorithms for learning theory – Part I: piecewise constant functions*, *Journal* of Machine Learning Research **6**, 1297-1321.

[99] B. Karaivanov (2010) *Self-Adjusting Region Growing (SARG)* In Point Clouds and Their Polygonization, IMI Pre-print.

[100] International Society for Photogrammetry and Remote Sensing Commission III, Photogrammetric Computer Vision Working Group III / 3, *Processing of point clouds from laser scanners and other sensors*, 2004 – 2008, http://www.itc.nl/isprswgIII-3.

[101] R.A. DeVore, D. Freeman, M. Hielsberg, L. Owens and G. Petrova (2010) *An Algorithm for Surface Approximation and Encoding From 3D Point Cloud Data*, Pre-print.

[102] A. Waters, M. Narayan, R. Baraniuk and R. DeVore (2010) *Hausdorff-Based Tree Decomposition for Point Cloud Representation and Compression*, Pre-print.

[103] http://vterrain.org/.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-TR-SET-118 AC/323(SET-118)TP/393	ISBN 978-92-837-0144-6	UNCLASSIFIED/ UNLIMITED

5. Originator

Research and Technology Organisation
North Atlantic Treaty Organisation
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title

3D Modelling of Urban Terrain

7. Presented at/Sponsored by

Final Report of Task Group SET-118.

8. Author(s)/Editor(s)	9. Date
Multiple	September 2011
10. Author's/Editor's Address	**11. Pages**
Multiple	118

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

3D Modelling	Military applications	Registration of point clouds
Data sets	Model assessment	Sensor systems
Feature extraction	Model instantiation	Through-wall SAR
InSAR	Model representation	Urban terrain
LIDAR	Object classification	

14. Abstract

The NATO Research and Technology Organisation (RTO) Research Task Group (RTG) on "3D-Modelling of Urban Terrain" (SET-118) was formed to study various representation forms for objects in urban terrain, to inspect procedures for automatic scene reconstruction by exploiting the data of modern sensors, and to discuss potential assessment metrics and criteria. The technologies under investigation cover active and passive sensors, such as flash laser, video, or interferometric SAR. For the modelling and the automatic 3D scene reconstruction, the concepts and requirements are discussed along with the corresponding representation forms. Military applications that benefit from the availability of 3D models and thereby show the relevance of these models are itemized and discussed.

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS

RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rto.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

CANADA
DSIGRD2 – Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5, 2750 Ballerup

ESPAGNE
SDG TECIN / DGAM
C/ Arturo Soria 289
Madrid 28033

ESTONIE
Estonian Ministry of Defence
Estonian National Coordinator for NATO RTO
Sakala 1, Tallinn 15094

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALIE
General Secretariat of Defence and National Armaments Directorate
5th Department – Technological Research
Via XX Settembre 123, 00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment, Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

REPUBLIQUE TCHEQUE
LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

ROUMANIE
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353, Bucharest

ROYAUME-UNI
Dstl Knowledge and Information Services
Building 247
Porton Down
Salisbury SP4 0JQ

SLOVAQUIE
Akadémia ozbrojených síl gen.
M.R. Štefánika, Distribučné a informačné stredisko RTO
Demänová 393, Liptovský Mikuláš 6
031 06

SLOVENIE
Ministry of Defence
Central Registry for EU and NATO
Vojkova 55
1000 Ljubljana

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar
Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et 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)

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rto.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

CANADA
DRDKIM2 – Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street, 9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

DENMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESTONIA
Estonian Ministry of Defence
Estonian National Coordinator for NATO RTO
Sakala 1, Tallinn 15094

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALY
General Secretariat of Defence and National Armaments Directorate
5th Department – Technological Research
Via XX Settembre 123, 00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment, Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

ROMANIA
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6, 061353, Bucharest

SLOVAKIA
Akadémia ozbrojených síl gen.
M.R. Štefánika, Distribučné a informačné stredisko RTO
Demänová 393, Liptovský Mikuláš 6
031 06

SLOVENIA
Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN
SDG TECIN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge and Information Services
Building 247
Porton Down
Salisbury SP4 0JQ

UNITED STATES
NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and 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 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database or on CD-ROM)

ISBN 978-92-837-0144-6