ARO 28058.4-PH

2

Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images

INTERIM TECHNICAL REPORT

(TR-91-6-70)

C. C. Chu and J. K. Aggarwal

March 1991

U. S. ARMY RESEARCH OFFICE

CONTRACT No. DAAL03-91-G-0050

THE UNIVERSITY OF TEXAS AT AUSTIN

COMPUTER AND VISION RESEARCH CENTER

AUSTIN, TEXAS 78712

Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images

INTERIM TECHNICAL REPORT

(TR-91-6-70)

C. C. Chu and J K. Aggarwal

March 1991

U. S. ARMY RESEARCH OFFICE

CONTRACT No. DAAL03-91-G-0050

Computer and Vision Research Center

The University of Texas at Austin

Austin, Texas 78712

512/471-3259

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1991

3. REPORT TYPE AND DATES COVERED
Interim Technical Report

4. TITLE AND SUBTITLE
Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images

5. FUNDING NUMBERS
DAAL03-91-G-0050

6. AUTHOR(S)
Chen-Chau Chu and J. K. Aggarwal

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Computer and Vision Research Center
The University of Texas
Austin, Texas 78712-1084

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-6-70

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 28258.4-PH

11. SUPPLEMENTARY NOTES
The view, opinions and/or findinga contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This paper presents a knowledge-based system to interpret registered laser radar and thermal images. The object is to detect and recognize man-made objects at kilometer range in outdoor scenes. The multisensor fusion approach is applied to various sensing modalities (range, intensity, velocity, and thermal) to improve both image segmentation and interpretation. The ability to use multiple sensors greatly helps an intelligent platform to understand and interact with its environment. The knowledge-based interpretation system, AIMS, is constructed using KEE and Lisp. Low-level attributes of image segments (regions) are computed by the segmentation modules and then converted into the KEE format. The interpretation system applies forward chaining in a bottom-up fashion to derive object-level interpretations from data bases generated by low-level processing modules. Segments are grouped into objects and then objects are classified into pre-defined categories. AIMS employs a two-tiered software structure. The efficiency of AIMS is enhanced by transferring non-symbolic processing tasks to a concurrent service manager (program). Therefore, tasks with different characteristics are executed using different software tools and methodologies.

14. SUBJECT TERMS
computer vision, laser radar, thermal imaging, expert system multi-sensor fusion, image interpretation

15. NUMBER OF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Multi-Sensor Image Interpretation Using Ladar Radar and Thermal Images

INTERIM TECHNICAL REPORT

Chen Chau Chu and J. K. Aggarwal

March 1991

U. S. ARMY RESEARCH OFFICE

CONTRACT NO. DAAL03-91-G-0050

Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712
512/471-3259

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

Multi-Sensor Image Interpretation Using Laser Radar and Thermal Images*

Chen-Chau Chu, student, and J. K. Aggarwal

Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712
E-mail: jka@emx.cc.utexas.edu

Abstract

This paper presents a knowledge-based system to interpret registered laser radar and thermal images. The objective is to detect and recognize man-made objects at kilometer range in outdoor scenes. The *multi-sensor fusion* approach is applied to various sensing modalities (range, intensity, velocity, and thermal) to improve both image segmentation and interpretation. The ability to use multiple sensors greatly helps an intelligent platform to understand and interact with its environment. The knowledge-based interpretation system, AIMS, is constructed using *KEE* and Lisp. Low-level attributes of image segments (regions) are computed by the segmentation modules and then converted to the *KEE* format. The interpretation system applies forward chaining in a bottom-up fashion to derive object-level interpretations from data bases generated by low-level processing modules. Segments are grouped into objects and then objects are classified into pre-defined categories. AIMS employs a two-tiered software structure. The efficiency of AIMS is enhanced by transferring non-symbolic processing tasks to a concurrent service manager (program). Therefore, tasks with different characteristics are executed using different software tools and methodologies. The interaction between the high and low level modules and the reasoning rules enable AIMS to tolerate errors by verifing segmentation and improving initial interpretation incrementally. Experimental results using real data are presented.

AI Topic: machine vision, image interpretation, intelligent robotics.
Domain Area: detection and recognition of man-made objects in outdoor scene.
Language/Tool: *KEE* /Lisp/C.
Status: under development.
Effort: 2 man-years.
Impact: Enhances image understanding capability of intelligent robots by using AI and multiple sensing modalities.

* This research was supported in part by the DoD Joint Service Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-86-C-0045, and in part by the Army Research Office under contract DAAL03-87-K-0089.

1 Introduction

This paper reports a prototype system to interpret ground-based, kilometer-range laser radar (ladar or lidar [1]) and infrared images. The goal of the system is to detect and recognize man-made objects (MMO) in outdoor rural scenes. The complete system consists of two building blocks: (1) the segmentation modules for all low-level processing, and (2) an interpretation system for high-level reasoning. The focus of this paper is the interpretation system: AIMS (Automatic Interpretation system using Multiple Sensors) [2]. The MMOs in our test images are mostly vehicles, such as trucks. The background is composed of vegetation, ground, and sky. However, the capability of AIMS is not limited to this specific domain. For example, the system may also be used for robot navigation, remote sensing, and other tasks that require the capability of image understanding using multiple sensing modalities. Our system applies the *multi-sensor fusion* (MSF) approach to integrate information derived from multiple modalities to improve both image segmentation (by pixel-level sensor fusion) and image interpretation (by object-level sensor fusion). Different sensors provide not only different types of information, but also multiple observations of the same information through different channels. Therefore, vision systems based on MSF can provide better performance than that of mono-sensor vision systems. MSF applies toward not only different sensors, but also different processing techniques because no single sensor and no single technique is sufficient under all circumstances.

Most vision problems, especially those at the intermediate level (segmentation and perceptual grouping) and the top level (recognition and interpretation), while often seemed trivial to human, can neither be formulated as analytical optimization problems nor by rigorous mathematics alone. The interpretation of multi-sensory images is even more difficult because many sensors provide images very different from video intensity images we perceive and utilize in daily life. The difficulties in image processing and the dissimilarities between the sensors pose major problems to the effective utilization of all information. Therefore, intelligent systems that interpret multi-sensory images automatically can provide valuable assistance to human experts and empower robotic systems to accomplish a wider range of missions. However, robust algorithms for high-level vision tasks, such as image interpretation, have not yet been established.

1.1 Knowledge-Based Systems in Vision

Techniques derived from artificial intelligence research, such as knowledge-based systems (KBS) and inexact reasoning, may provide solutions to machine vision in general [3] and to sensor fusion in particular [4]. The KBS approach has been applied to various machine vision tasks, including image segmentation, object recognition, and scene interpretation for video, thermal [5], and indoor range images [6,7]. However, indoor range data are usually much more precise than data from outdoor range imaging because of the much shorter distances involved. Among various applications of ladar [1], it can be used as a ground-to-ground, long-distance sensing device. Figure 5 shows an example of ladar images. Ladar range data and thermal images have been used jointly to detect targets in the field [8]. Recently, XTRS, a target recognition system that uses ladar images has been reported [9]. Though the above mentioned systems have met some degrees of success, they have not rigorously applied MSF to enhance system performance. For example, in XTRS, two subsystems, one region-based and the other contour-based, work in parallel but *not cooperatively*. Therefore, the interpretation module in each subsystem does not have complete low-level information. Besides, most laser-based systems use only the range channel provided by the laser ranging devices.

In comparison, AIMS uses all the available modalities in an *integrated* fashion. Our ladar images have three inherently registered components: range, intensity, and velocity. The thermal images are manually registered with the ladar images. Each modality provides different but complementary information: 3D geometry and object surface structure are extracted from range data; intensity data provide object surface reflectivity information; velocity data indicate moving targets; thermal images provide information about object temperature and thermal capacitance. Segmentation information derived from all data channels using various segmentation techniques are integrated into a single segmentation map (low-level integration) before the interpretation starts. AIMS uses the integrated segmentation map and other information from all information channels in the form of consistent interpretation hypotheses and increased confidence factors (high-level integration). Hence, AIMS has complete information of the scene rather than just partial information from a single source or a single feature extractor. AIMS is designed with the KBS technique for its ability to (1) separate interpretation knowledge and the inference mechanism, (2) handle inexact reasoning, and (3) emplying both forward chaining for a data-driven, bottom-up approach and backward chaining for a focused search.

1.2 System Overview

Figure 1 shows the overall structure of our system. The segmentation modules are written in C, while the reasoning modules are built using *KEE*[†] and Lisp. *KEE* is a commercial package for expert system shell development. It provides the inference engine and the rule parser in AIMS. *KEE* uses *frame* [10] for knowledge representation and encourages object-oriented programming. The image segmentation modules execute low-level tasks using minimal knowledge about the problem domain. They are divided into six groups of different functions: (1) noise removal; (2) image segmentation by surface fitting; (3) segmentation by the statistics of pixel values; (4) segmentation by histogram analysis and thresholding; (5) integration of segmentation maps, and (6) database generation.

AIMS includes four major components: (1) the inference mechanism provided by *KEE*; (2) the rule bases and supplementary Lisp code, which contain the knowledge for image interpretation; (3) the data bases, which are produced by the database generator; and (4) the *service manager*, which executes numerical and graphics tasks for AIMS. The interpretation process starts by checking attributes extracted by the image segmentation modules. It then labels each segment as part of a man-made object or as the natural background (BG) based on these parameters. Next, segments are grouped into *objects* based on several criteria. Image interpretation rules then generate hypotheses of object interpretations. The hypotheses are strengthened or weakened by examining more evidence.

2 Data Characteristics and Image Segmentation

2.1 Laser Radar Data

Ladar discerns more structural details of distant objects because of its short wavelength. The random refraction and reflection of laser light in the atmosphere and on the object surfaces generate *speckle noise*. This noise is significant in long-distance

[†] *KEE* is a trade mark of IntelliCorp.

out-door range imaging but virtually non-existent in indoor range imaging [6,7]. It is difficult, if not impossible, to reason about ladar images at the pixel level because of the speckle noise. Therefore, good segmentation is a crucial intermediate stage before image interpretation. In addition, how the images are segmented is closely related to how they are interpreted. We apply two segmentation methods, *surface fitting* and *image statistics*, to ladar data in AIMS. The surface fitting method is designed to highlight object surface geometry, while the image statistics method is used to detect differences in object surface reflectivity. A complete discussion of the segmentation algorithms and their performances using ladar data is reported in [11].

Most man-made objects are made of surfaces representable by patches of low-order surfaces. This assumption is practically true when the distance to an object is large compared to its body dimensions, as it is in our task domain. Therefore, only planar surfaces are used. The surface fitting-based segmentation algorithm employs a region-growing approach. Surfaces are fitted to segments and segments grow as long as the fitting error is within a pre-determined bound. Different object surface materials may generate different speckle patterns, which in turn generate different standard deviations (SD) of pixel values. The differences of local mean and SD are used for segmentation. The statistical approach is also applicable to range and velocity data. For example, the average range value for a segment is a good estimation of its distance to the sensor.

2.2 Thermal Image Characteristics and Segmentation

The pixel values in thermal (infrared or IR) images are usually dominated by the thermal properties of different materials, such as the thermal capacitance and the heat sink/source distinction. Some of these properties can differentiate object surface materials and, hence, indicate the existence of MMO's. However, IR images usually have lower spatial resolution and contrast than video intensity images. These properties result in extra problems for the segmentation and recognition. A popular approach for IR segmentation is background/target thresholding using the histogram, assuming that pixel values consist of a bimodal distribution. The IR images used in this research satisfy this assumption. The targets usually occupy less than 20% of the total number of image pixels and exhibit higher temperatures than that of the background, which is mostly vegetation.

A segmentation scheme is designed based on such observations. By the *Central Limit Theorem*, one can assume that all the different thermal characteristics of background vegetation result in a Gaussian distribution of pixel values. This Gaussian bell is located at the lower-end of the histogram because hardly anything is *cooler* than the background vegetation (except shadows and the sky). The peak of this Gaussian bell is also the peak of the entire histogram because the background dominates the entire image. The peak of the histogram and its standard deviation σ is determined by solving $0.5 = exp(\frac{-(x-\mu)^2}{2\sigma^2})$, where x is the 3db width of the Gaussian distribution. In a Gaussian distribution, the mean μ is the same as the *mode* of the distribution and, hence, is easily determined as the peak of the histogram. Note that μ is not determined as the average of the entire thermal image. All pixels with gray values covered by the range of $[0, \mu+\sigma]$ are classified as background; all pixels with gray values in the range of $[\mu+3\sigma, 255]$ are considered as MMO. Pixels with gray values in-between are then determined by their proximity to classified pixels. However, only regions large enough are established as segments.

2.3 The Integration of Segmentation and Database Generation

Different methods operating on multiple data sources generate different segmentation maps. These maps may have errors and possibly contradict one another. Integration from multiple sources enhances the signal to noise ratio. Therefore, errors and inconsistencies are expected to be reduced in the integration process. It is helpful to apply different weights on various input segmentation maps because there may be significant differences in the quality and reliability of different segmentation methods and data sources. For examples, segmentation from velocity images containing moving targets should be given larger weights than those that do not. Edge information, if available, may also be integrated into the segmentation map as a cue for region separation.

The output of the low-level integration module is a new segmentation map in which all segments are large and their contours compact (determined by thresholds). The current implementation of this integration module [11,12] is domain-independent. The integration module works with both region-oriented segmentation and edge detection modules [12]. In general, range data are not as noisy as their intensity counterparts and, therefore, are given higher weights. Velocity data provide useful segmentation information only if moving targets are in the scene. Therefore, the weight on velocity segmentation depends on the segmentation outcome of individual images. A set of utility programs collect the values for various attributes using original images and the integrated segmentation map. These data are converted to the representation format of *KEE* by the database generator, and the database is then transferred to AIMS as the basis for the interpretation [2].

3 The Design of the Knowledge-Based System

The interpretation strategy of our work follows the three-step paradigm of Clancey's Heuristic Classification [13]. First, numerical parameters are converted into qualitative descriptors. Second, these descriptors are used to generate intermediate classifications of segments as man-made objects or background. Third, segments are grouped into objects and these objects are further classified into one of the pre-defined categories. Figure 2 shows the block diagram for AIMS and its operation.

3.1 Knowledge Sources and Representation

Man-made objects and natural backgrounds have different features. These differences are reflected in different modalities in various forms. Expert knowledge is needed to detect such differences and to recognize the detected objects. Five types of knowledge sources are used to construct rules:

1. imaging geometry and device parameters (knowledge which is dependent on the hardware but not on the imaged scene);
2. numerical measurements for each segment (knowledge derived from pixel values under the guidance of various segmentation maps), such as region size and average temperature value in a region;
3. neighborhood relationships in the segmentation maps (knowledge derived from the segmentation maps, but independent of image pixel values);
4. models of possible objects (knowledge derived from potential targets); and
5. general heuristics (knowledge derived from known facts in the task domain and common sense).

Several different frame structures are defined to record information about the imaging devices, segments, and models of potential targets. Some attributes contain *active values*, or *demons*, which fire corresponding procedures (additional Lisp codes) when certain operations are performed on the selected slots. For example, when two rules generate two different *interpretations* (a symbolic attribute) of a target, both interpretations may be accepted and are stored in the order of the strengths of the hypotheses. The data structures representing scene contents are organized as two levels: *segments* and *objects*. The segment frames are used to represent subparts, while the object frames are used to represent a group of segments. The segment frames correspond to individual segments (areas) in the integrated segmentation map. The object frames are built as a higher-level structure during the grouping stage in the reasoning process. Grouping is necessary to correct the potential problem of *over-segmentation* in the segmentation stage. An object containing a single segment *is* a representation overhead. However, such overhead is necessary because the reasoning process is bottom-up and the *segment-* level representation is built first.

3.2 Hypothesis Integration and Confidence Factors

Each *KEE* rule posts one or more hypotheses expressed as a quadruple (*segment/object, attribute, value, confidence factor*). The hypothesis(es) is stored in the specified segment frame and slot as a pair (value, confidence factor). The *confidence factor* (CF) is a real number between -1.0 and 1.0. The CF denotes the degree of disbelief (negative number) or belief (positive number) of the associated hypothesis. The CF is used to handle inexact reasoning as opposed to logic resolution, for which everything is exactly true or false. Certain low-level numerical attributes, such as the bounding rectangles and the size of a segment, are computed without using the CF.

The CF value determined by a rule usually changes with one or more selected parameters (Figure 3). This is necessary for two reasons. First, rules are not equally effective under all circumstances. A rule may generate the same hypothesis with different CFs for segments with different attribute values. Second, thresholding is usually used to transform quantitative descriptors into qualitative (symbolic) descriptors in KBS. The nonlinearity introduced in this way is not always desirable, especially during the intermediate stage of reasoning. Furthermore, it is difficult to choose a set of fixed thresholds which perform well in various conditions. Modifying the CFs dynamically as a continuous function reduces the rigidness of fixed thresholds.

Our work assigns the CFs *empirically* in the interpretation rules. Multiple hypotheses concerning the same attributes of the same object are combined in a way similar to MYCIN [14]. The CF combination rule is

$$Combine(a,b) = \begin{cases} 1-(1-a)(1-b) & \text{if } a > 0 \text{ and } b > 0 \\ -Combine(-a,-b) & \text{if } a < 0 \text{ and } b < 0 \\ (a+b)/(|a|+|b|) & \text{otherwise.} \end{cases} \quad (1)$$

The above combination rule provides satisfactory results, although it is based on heuristics as much as on probability theories. In our work, we have not adopted the methods that use detailed mathematical modeling, such as the Dempster-Shafer theory [15]. The reason is that several important assumptions in such probabilistic models are unlikely to be true in real situations. For example, it is very difficult (1) to get precise measurements of the probabilities (*a priori* or *a posteriori*) associated with all events; (2) to claim the statistics from a limited data set (i.e., training) as a reliable estimation of the underlying distribution function; and (3) to verify the independence between events. If these assumptions are unconfirmed, using complicated mathematic models does not deliver the promised optimality.

3.3 Rule Bases and the Reasoning Process

The rules in AIMS are organized into five groups: (1) pre-processing and system initialization, (2) coarse classification of segments into MMO/BG, (3) segment grouping, (4) classification of BG segments/objects, and (5) classification of MMO segments/objects. These groups of rules are sequentially invoked in forward chaining (FC). At any given time, only one group of rules is active in the *match-resolve-fire* cycle. However, stages (4) and (5) can operate in parallel. The conflict resolution strategies in AIMS are *rule weighting* and *FIFO*. The partition of rule bases reduces the matching overhead of rule selection, and provides indirect control over the breadth-first search implied in FC. Backward chaining (BC) rules will be added in the future to adopt the *hypothesize-and-verify* approach for focused searches. Thus, when a hypothesis with a strong confidence is posted, AIMS can switch into the BC mode to verify that hypothesis. The rule groups are described below:

1. The *pre-processing* module handles the differences between individual segmentation maps and integrated segmentation maps. Rules in this group also compute low-level attribute values and place them into correct slots. This module contains largely numerical tasks whose functionalities are gradually shifted to the database generator and the service manager.

2. The *MMO/BG distinction* is made based on various attributes and numerical parameters, such as the surface temperature, the surface fitting coefficients, the SD of range values, etc. We find that this binary decision of `MMO/BG` is always made correctly with high CF values.

   Example:
   IF (segment A is relatively hot)
   AND (segment A has compact contour)
   THEN (segment A is an MMO, confidence = Conf(temperature,shape)).

3. The *grouping of segments into objects* depends on the neighborhood relationship, the `MMO/BG` classification, the difference in distance, and the object contour analysis. Only segments of the same `MMO/BG` type can be grouped together. Thermal image segmentation usually helps the grouping process because thermal images are usually under-segmented due to the lack of contrast.

4. The *classification of BG* uses the velocity of an object, the position of a segment/object within the image frame, the SD of range values, and other attributes to classify BG segments into `SKY`, `TREE`, and `GROUND`. For example, `GROUND` is usually at the lower part of the image, though not always. Therefore, being planar and (surface normal) pointing upward are more important criteria.

   Example:
   IF (segment A is type `BG`)
   AND (segment A is relatively cool)
   AND (segment A can be fit by a (planar) surface)
   AND (the surface normal of segment A points upward)
   THEN (segment A is `GROUND` with a confidence of 0.9).

5. The *classification of MMOs* into BULLETIN BOARD, TANK, APC, JEEP, and TRUCK relies mostly on shape and size analysis. Rules that recognize targets in more general articulations are under development. However, based on dz/dy (surface gradient), the surface fitting error, and the knowledge of target body dimensions, it is possible to estimate the rotation of an object and to determine whether the target is viewed from broad-side.

   Example:
   IF (segment A is of type MMO)
   AND (segment A has a width of less than 4m)
   AND (segment A is no taller than 2m)
   THEN (segment A is a JEEP with a confidence of 0.8).

3.4 The Service Manager

Despite the flexibility of *KEE*, three major issues are identified as its weak spots: (1) execution efficiency, (2) low-level data access during high-level reasoning, and (3) interface capability and feedback to low-level processes. Lisp-based development systems, such as *KEE*, are convenient tools to execute symbolic reasoning tasks and to handle explicitly-encoded knowledge. However, these systems usually do so at the price of software overhead. The slowdown occurs for two main reasons. First, most such packages are built on multiple layers of software and, therefore, are very inefficient. Though *KEE* provides an extensive set of *primitives* (functions) for parameter access and program control, some functions can be implemented more efficiently using Lisp or C code. Second, image interpretation is not a task that consists solely of symbolic processing. For example, some rules may need the body dimensions, the average velocity, and the symmetry of the object's body contour for recognition. Moreover, not all the data can be conveniently stored/accessed in the *frame* paradigm. In general, accessing image pixel values and data files is difficult to implement directly using *KEE* primitives. Lisp code may be used, but it is not as efficient as C code running on general-purpose hardware. The graphics component of *KEE* does not provide the capability or flexibility needed by AIMS. Therefore, we have to implement our own graphics interface to control the graphics hardware.

Our solution to the above problems is a program, the *service manager* (Figure 4), which runs concurrently with AIMS. The purpose of the service manager is to help AIMS run low-level tasks efficiently on the designated development platform. AIMS sends a message to the service manager for the desired service; then the service manager interprets and executes the commands and feeds the results back. The functions of the service manager include numerical-intensive subroutines, color graphics, image file I/O, and the access of low-level data, such as pixel values and segmentation maps, etc. These operations can be written as supplemental Lisp code called from within *KEE*; however, Lisp code runs slowly for these tasks and lacks the flexibility of C in controlling the I/O and peripherals. Thus, when the system grows larger, such inefficiency degrades the system performance significantly and slows down the development process.

The interaction between low- and high-level processes is helpful for the interpretation system. The database generator provides the feed-forward interface from the low-level process to the symbolic reasoning process. The service manager provides the feedback path from the symbolic process to the low-level, numerical process. The efficiency of the service manager enables rule designers to use more complicated tests (the *IF* part) and to take more complicated actions in the conclusion (the *THEN* part) of rules. In addition to calculating numerical parameters, rules can be constructed to direct the segmentation modules to refine earlier segmentation results.

Using this service manager provides a good trade-off during the implementation of AIMS, because it cuts short the development cycle and facilitates more testing. On one hand, such hybrid software structure accelerates the software development. On the other hand, AIMS still keeps most knowledge (the interpretation rules) in a symbolic, explicit format independent of the inference mechanism. The contents of some well-understood *KEE* rules are gradually replaced by Lisp code for efficiency, while the form of the rules is not changed. At the same time, the Lisp code is gradually replaced by a task assignment to the service manager program. Thus, the benefit of a high-level expert system shell is mostly preserved and the problem of slower operation is reduced. The choice of a specific expert system shell (or to build one from scratch) depends many factors, such as available resources (e.g., man-years) and project requirements (e.g., run-time efficiency). The choice of a programming language is less dominant, since all of them have equivalent description power.

4 Experimental Results

Figure 5 contains the original ladar range, intensity, velocity, and registered thermal image. The scene shows a single 5-ton truck, 910m from the ladar sensor, heading to the right but not moving. The top image in Figure 6 is the integrated segmentation map with region boundaries in white contours overlaid on the range image. White regions are detected targets, and black areas are segments which do not have a high-confidence interpretation hypothesis. Some of the black areas are actually classified as GROUND or SKY. However, the confidence factors for such classifications fall below a threshold (0.4) and are considered too weak to report. Light gray marks GROUND and dark gray marks SKY.

The integrated segmentation correctly marks the entire truck as a single segment, and the interpretation module classifies the segment as MMO. The segment is further recognized as a TRUCK. The hot engine block is detected and (together with shape analysis of target contour) confirms the heading of the truck. The approximate body dimensions of the truck (length and height) are estimated from the bounding rectangle, dz/dy, and the spatial resolution of the ladar receiver (0.05 millirad). The dz/dy gradient is also used to estimate the rotations of the target as 16°, compared to the documented value of 24°. Since the truck is 910 meters away and the data are noisy, the rotation estimate can not be very accurate. The length of the truck is estimated as 5.68 meters, and this length can be used to verify the target recognition hypothesis. Note that our segmentation and integration algorithms favor compact regions. Therefore, gun barrels, antennas, and exhaust pipes are not always preserved. However, this behavior can be changed by modifying the integration algorithms to preserve linear features and long, pipe-like regions.

The system is implemented on an IBM RT PC running AIX. The data collection and database generation modules between the segmentation modules and AIMS take about two minutes of CPU time. AIMS takes about 30 minutes of wall-clock-time, (actually 19 minutes of CPU time due to extensive memory swapping) to interpret one set of images. The CPU time is expected to be longer if the C-based service manager is not employed and, hence, all the low-level processing tasks must be coded as *KEE* and Lisp functions. An experiment to implement 25% of all the interpretation rules in C with a primitive FC engine on a 64-node AT&T PIXEL machine accelerates the WCT by a factor of about 200.

5 Conclusion

A knowledge-based system (AIMS) for integrated laser radar (ladar) and thermal image interpretation is presented. It performs well on real images to detect and recognize man-made objects. AIMS consists of rule-based reasoning modules, and requires the segmentation modules to provide input data. The *multi-sensor fusion* (MSF) approach is applied at both the segmentation and the reasoning levels. The low-level integration module fuses segmentation cues from multiple sources to generate an improved segmentation map. The additional information provided by MSF is vital because of the significant loss of information in the transformation from a 3D world to 2D images and various forms of noise. AIMS uses forward chaining to drive the interpretation process in a bottom-up fashion. The reasoning process follows the order of data abstraction, heuristic classification (target detection), and refinement/verification (target recognition). The software structure of AIMS is a hybrid. Therefore, tasks at different levels of the machine vision paradigm are executed using different software tools and methodologies.

The performance of the system indicates both the power of the MSF approach and the suitability of using knowledge-based systems to pursue MSF. This assertion may be examined from three perspectives: (1) Multiple sensing modalities provide different and complementary information about the scene. The complexity of the MSF-based system is high and no known algorithm manages the information effectively. (2) The integration of segmentation maps provides high-quality segmentation, which is essential for intelligent image interpretation. (3) The high-level integration of interpretation knowledge from different knowledge sources and different sensing modalities produces better scene interpretation. The reasoning system integrates high-level information from multiple modalities in the form of consistent interpretation hypotheses and increased confidence factors.

AIMS has been developed over four years by one half-time researcher to reach its current status and is under further development. The current system is just a prototype and recognizes only a small number of objects. It must acquire additional knowledge to work on more difficult problems. When the problem domain changes, different sets of object models and recognition rules have to be built, and probably different sets of features are needed. Currently, 3D models of potential targets are under further development to improve target recognition in different viewing directions. These models are constructed as another knowledge base in *KEE* format system such that the knowledge of AIMS remains explicitly encoded and separated from the inference mechanism.

Acknowledgement

The authors would like to thank the Army Electro-Optical and Night Vision Center for providing the test images.

References

[1] C. G. Bachman, *Laser Radar Systems and Techniques,* Artech House, Dedham Massachusetts, 1979.

[2] C. Chu and J. K. Aggarwal, "The interpretation of laser radar images by a knowledge-based system," to appear in *Journal of Machine Vision and Applications.*

[3] L. G. Shapiro, "The role of AI in computer vision," *The Second Conference on Artificial Intelligence Applications,* Miami Beach, Florida, December 11-13, 1985, pp. 76-81.

[4] T. D. Garvey, "Survey of AI approaches to the integration of information," *Proceedings of SPIE*, vol. 782, May 1987, pp. 68-82.

[5] T. M. Silberberg, "Context dependent target recognition," *The Proceedings of DARPA Image Understanding Workshop,* Los Angeles, California, February 1987, pp. 313-320.

[6] A. K. Jain and R. Hoffman, "Evidence-based recognition of 3-D objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-10, no. 6, November 1988, pp. 783-802.

[7] T. J. Fan, G. Medioni, and R. Nevatia, "3D object recognition using surface descriptions," *Proc. DARPA Image Understanding Workshop,* April 1988, pp. 383-397.

[8] C. W. Tong, S. K. Rogers, J. P. Mills, and M. K. Kabrisky, "Multisensor data fusion of laser radar and forward looking infrared (FLIR) for target segmentation and enhancement," *Proceedings of SPIE*, vol. 782, May 1987, pp. 10-19.

[9] D. E. Dudgeon, J. G. Verly, and R. L. Delanoy, "An experimental target recognition system for laser radar imagery," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 479-506.

[10] E. Rich, *Artificial Intelligence,* McGraw-Hill, New York, New York, 1983.

[11] C. Chu, N. Nandhakumar and J. K. Aggarwal, "Image segmentation using laser radar data," to appear in *Pattern Recognition.*

[12] C. Chu and J. K. Aggarwal, "The integration of region and edge-based segmentation," to appear in *The Proceedings of the Second International Conference on Computer Vision, Osaka, Japan, December 1990.*

[13] W. J. Clancey, "Heuristic Classification," *Artificial Intelligence,* vol. 27, pp. 289-350.

[14] B. G. Buchanan and E. H. Shortliffe, *Rule-Based Expert Systems,* Addison-Wesley, Massachusetts, 1984.

[15] G. Shafer, *A Mathematical Theory of Evidence,* Princeton University Press, 1976.

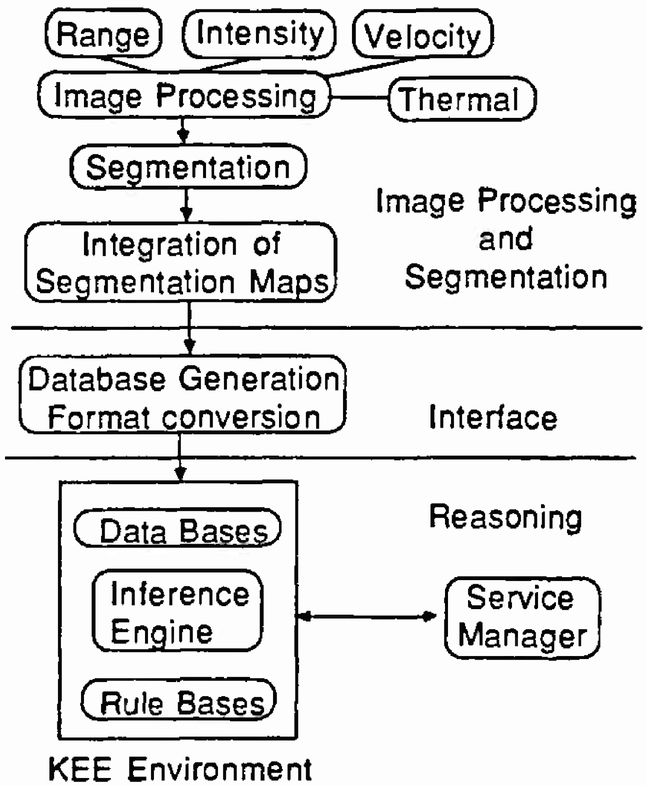

Figure 1: System overview

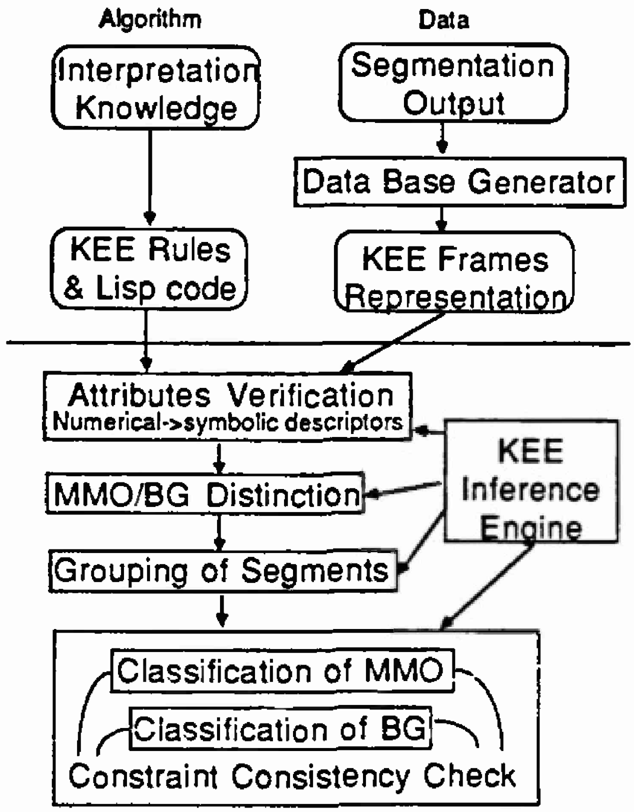

Figure 2: Block diagram for the interpretation system

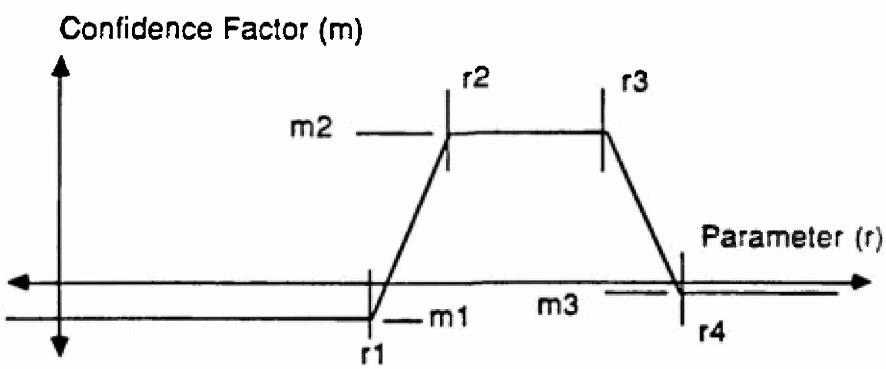

Figure 3: Parametrically changing confidence factors

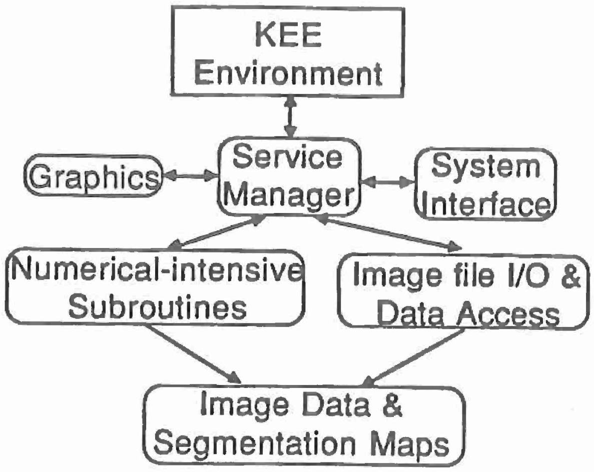

Figure 4: The Service Manager

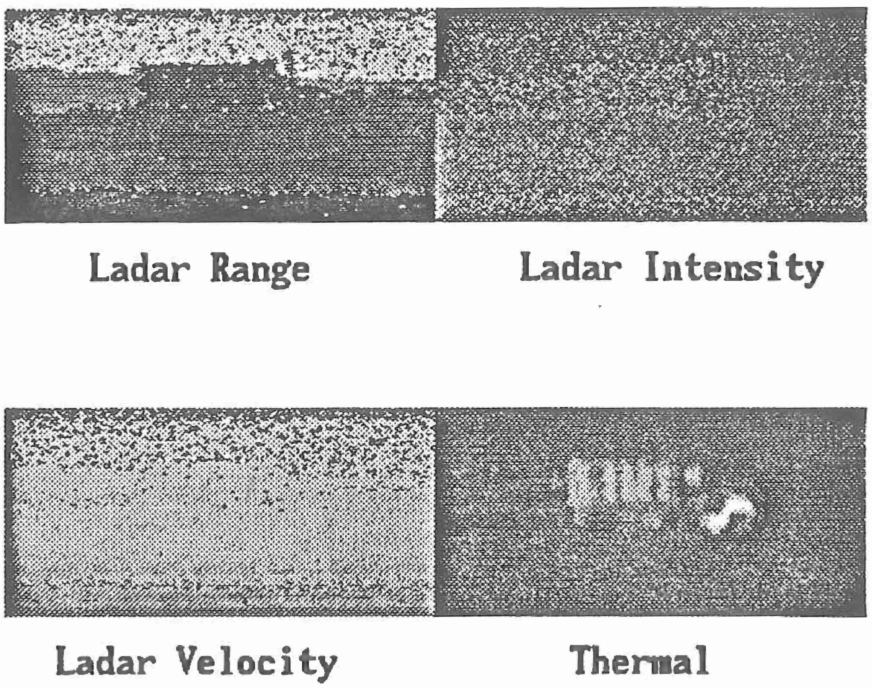

Figure 5: Source images for a 5-ton truck

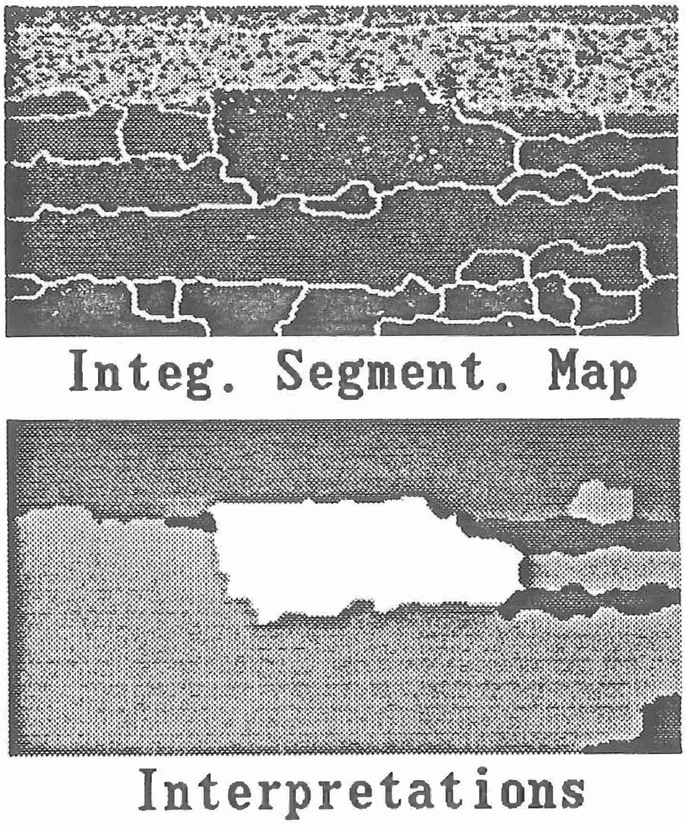

Figure 6: The segmentation and the interpretation