ESC-TR-92-206

AD-A266 640

2

Technical Report
979

Pixel-Level Fusion Using "Interest" Images

R.L. Delanoy
J.G. Verly
D.E. Dudgeon

26 April 1993

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Prepared for the Advanced Research Projects Agency under Air Force Contract F19628-90-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Advanced Research Projects Agency under Air Force Contract F19628-90-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Directorate of Contracted Support Management

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

PIXEL-LEVEL FUSION USING "INTEREST" IMAGES

R.L. DELANOY
J.G. VERLY
D.E. DUDGEON
Group 21

TECHNICAL REPORT 979

26 APRIL 1993

LEXINGTON MASSACHUSETTS

ABSTRACT

A simple general-purpose means of representing image sensory data at the pixel level is presented in the context of target detection. One or more interest operators (i.e., feature detectors) are applied to one or more sensory images. The output of each operator is an *interest image* in which high pixel values are assigned to locations where features are found that are indicative of the object (or target) being sought. Any relevant sensory data that are transformed into interest images can then be fused in a pixel-wise manner by means as simple as computing averages, maxima or minima. The approach has been implemented in the detection module of an automatic target recognition system built at Lincoln Laboratory.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. INTRODUCTION

Object detection relies upon the availability of features that are both selectively indicative of the object being sought and discernible in available sensor modalities. A single feature is seldom sufficiently reliable to ensure detecting all instances of a given object and is almost never robust enough to discriminate objects from surrounding clutter. Thus arises the relevance of sensor/data fusion to object detection and identification. Given a need for multiple features, possibly visible in different sensory modalities, how is such information to be assimilated at the pixel level?

This problem has been approached with the belief that fusion can be simplified once a "common denominator" is adopted. In this case, *interest* will be a metric that may apply to any feature, regardless of image source. And, because interest can be quantified, fusion is accomplished by simple arithmetic or fuzzy logic operations.

In this paper, interest will be discussed as a fusable quantity and as the basis of implementing selective attention for automatic target recognition (ATR). Portions of the experimental target recognition system (XTRS) will be presented to illustrate the ideas.

2. INTEREST AND SELECTIVE ATTENTION

Martin and Aggarwal [1] observed that most dynamic scene analysis systems have followed a hierarchical approach analogous to the stages of human perception. The lowest level, called *peripheral*, acts as a perceptual filter, which attracts attention only to those regions in an input image that are most crucial to processing. For dynamic scene analysis in particular, such regions are usually where changes have occurred from one scene to the next. Peripheral processing should use simple, computationally inexpensive techniques. The next level, called *attentive*, more liberally expends computer resources on selected, interesting regions of the input image. It converts gray-level pixel values to symbolic descriptions of the scene and scene dynamics. The third and highest level of processing, called *cognitive*, attempts to interpret the scene, given the symbolic descriptions of the scene and world-specific knowledge. It can request more information from the attentive level, but in the organization described by Martin and Aggarwal, the cognitive level does not communicate instructions to the peripheral level.

Not surprisingly, the hierarchy of steps in most constructed ATR systems parallels the levels described by Martin and Aggarwal. In this case, the peripheral level is called detection, the attentive level is usually called extraction or discrimination, and the cognitive level is analogous to identification, such as model-based matching.

Jain and Haynes [2] later pointed out that the task of the peripheral level (i.e., target detection) is to localize areas of *interest*. Such areas are then processed by attentive processes. In other words, limited perceptual resources are allocated only where the most interesting features appear in a scene. By quantifying interest, one can prioritize how and where resources are to be allocated. Previous applications have used interest to guide selective attention to differences between images. Moravec [3] created what he called an *interest operator* to measure the distinctness of one part of an image from neighboring parts. This operator was used to highlight image features either for correlation between an image pair, such as would be generated at sequential time intervals, or for stereoscopic pairs. Similarly, Jain and Haynes applied the concept of interest to edge motion in dynamic scene analysis.

In this report it is maintained that interest is also applicable to guiding target detection in single images. But in the absence of motion, how should interest be assigned? The answer turns out to be dependent on goals and context. A feature is interesting only if it is selectively indicative of some intended target and is observable in available sensor modalities. Clearly, these things can change with mission or environmental conditions. Therefore, diverging from Martin and Aggarwal, we feel that target detection (peripheral level) should have access to world knowledge and feedback control.

3. INTEREST IMAGES

Interest is a quantifiable attribute of an image location. An intrinsic image that maps such interest values across the visual space is defined as an *interest image*. In general, any algorithm that generates an interest image will be called a *feature detector*. Depending upon the goal, some sensor imaging modalities may provide ready-to-use interest images. For example, vehicles typically have edges and smooth surfaces that are good reflectors, which makes laser radar intensity images or synthetic aperture radar (SAR) images possible interest images. Other potential ready-to-use interest images are passive infrared (FLIR) images (hot objects are generally more interesting than those at ambient temperatures) and millimeter wave radar or laser radar doppler images (moving objects are usually more interesting than stationary ones).

In contrast, range values do not directly confer any knowledge of vehicle location. Consequently, unprocessed range images generally cannot be used as interest images. However, numerous transformations can be applied to a range image to create interest images. For example, if an object class being sought typically exhibits rod-shaped appendages, an algorithm that creates high scores wherever rod-shaped features are located in a range image will also create an interest image. Other interest images might be created through the application of various shape matching techniques, such as ones that identify blobs of a particular height or width, specific corner angles, parallel lines, or shapes with long aspect ratios. Range images can also be used to extract surfaces at arbitrary orientations, such as vertical surfaces.

Using verticality, interest can be shown to be goal dependent: The amount of interest assigned to vertical surfaces might be high if the object sought is a vehicle, but low if the object sought is a road. Similarly, context dependency can be shown: Verticality is an adequate indicator of vehicle presence provided that the contrasting terrain is relatively flat. In an urban scene containing numerous vertical surfaces, a verticality image would be useless as an interest image. Consequently, using interest images as a means of guiding selective attention requires a degree of flexibility in choosing which interest images are relevant to the particular goal and environmental context.

4. EXAMPLE INTEREST IMAGES

These ideas are illustrated using range and intensity images generated by a forward-looking 10.6 μm CO_2 laser radar. The images are of armored vehicles in various orientations, which are located from 0.7 to 1.6 km from the sensor. An example range and intensity image pair is shown in Figure 1. The following four feature detectors were found to be suited for the target set and available imagery.

217961-1

(a) (b)

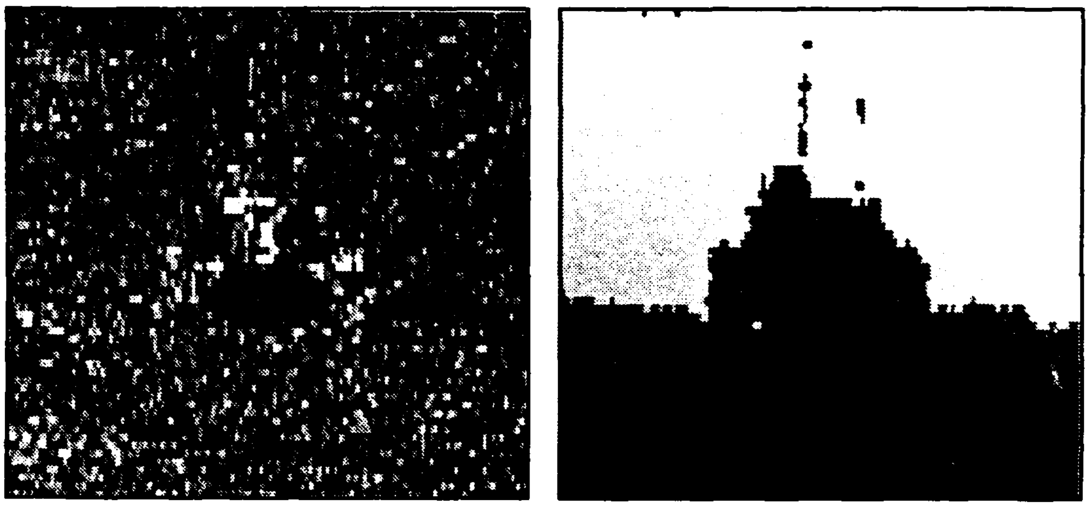

Figure 1. Example laser radar range and intensity images of an M48 tank (frontal view).

4.1 INTENSITY

As already discussed, intensity images can be an appropriate interest image for military vehicles. As shown in Figure 1, numerous bright pixels are found on the turret and fenders of the tank. However, note the dark area between the tank treads under the hull. This area of low energy return is due to the geometry of the front of the vehicle; the plate of metal is angled such that the laser energy is directed to the ground and not back to the sensor. Such large patches of low intensity are just as indicative of targets as bright returns. Consequently, the intensity interest image can be further improved by the following *reflection* mechanism. Intensity image values are replaced by the average of pixels within a 5 × 5 window in order to remove isolated dark spots. Then, pixel values of the averaged image more than 1.5 standard deviations below the image mean value are set to a value proportionately above the mean. Both raw and enhanced intensity images can be seen in Figure 2 for two targets, an armored personnel carrier (APC) oriented 45 degrees to the sensor and the same tank image as is shown in Figure 1. Note that the low intensity area corresponding to the side of the APC becomes a high interest area in the enhanced interest image.

Other sensor modalities can exhibit a similar bimodal distribution of interest values. For example, in passive IR imagery, objects that have considerable heat capacities (such as those containing large masses of metal) tend to be hotter than the ambient temperature in the evening, but colder in the morning. Either extreme might be indicative of armored vehicles.

4.2 HEIGHT-LIMITED VERTICALITY

Range imagery allows the calculation of surface orientation in three-dimensional space at each pixel in the image. Vertical surfaces are useful in the context of detecting targets in forward-looking range imagery, since targets of interest sitting on nearly horizontal terrain will typically exhibit at least one nearly vertical surface.

Verticality is estimated within a single column of pixels.[1] The resulting verticality image is a map of the heights (in meters) of columns of pixels having the the same approximate range value. Such a verticality image is shown in Figure 3(b). Notice, however, that very tall vertical surfaces, such as the utility poles, can overwhelm shorter vertical surfaces such as the tank at the base of the leftmost utility pole. Consequently, any verticality value greater than 3.6 m (the height is specific to the targets of interest, in this case the armored vehicles) is reset to a lower value, which results in an interest image selective for vertical surfaces around 3.6 m high [see Figure 3(c)]. The use of height-limited verticality as a means of target detection was demonstrated in a report by Otazo and Tung [4].

4.3 BODY INTEREST IMAGE

Another feature of the armored vehicles in our target set that distinguishes them from surrounding clutter is their width. No matter what perspective they are viewed from, the horizontal length of the unoccluded vehicles lies between 2.5 and 8.0 m.

An interest image of horizontal rows of pixels with lengths between these two values is a fairly simple exercise of mathematical morphology (MM) [5,6]. The result of closing the image R with a horizontal SE 2.5 m long is subtracted from R closed with a horizontal SE 8.0 m long, or

$$BI = R^{H_{8.0}} - R^{H_{2.5}} \quad . \tag{1}$$

[1] In general, verticality is computed by fitting a planar surface to image range values and cannot be determined using a one-pixel-wide vertical window. However, this approximate approach is acceptable for small depression angles and for small fields of view.

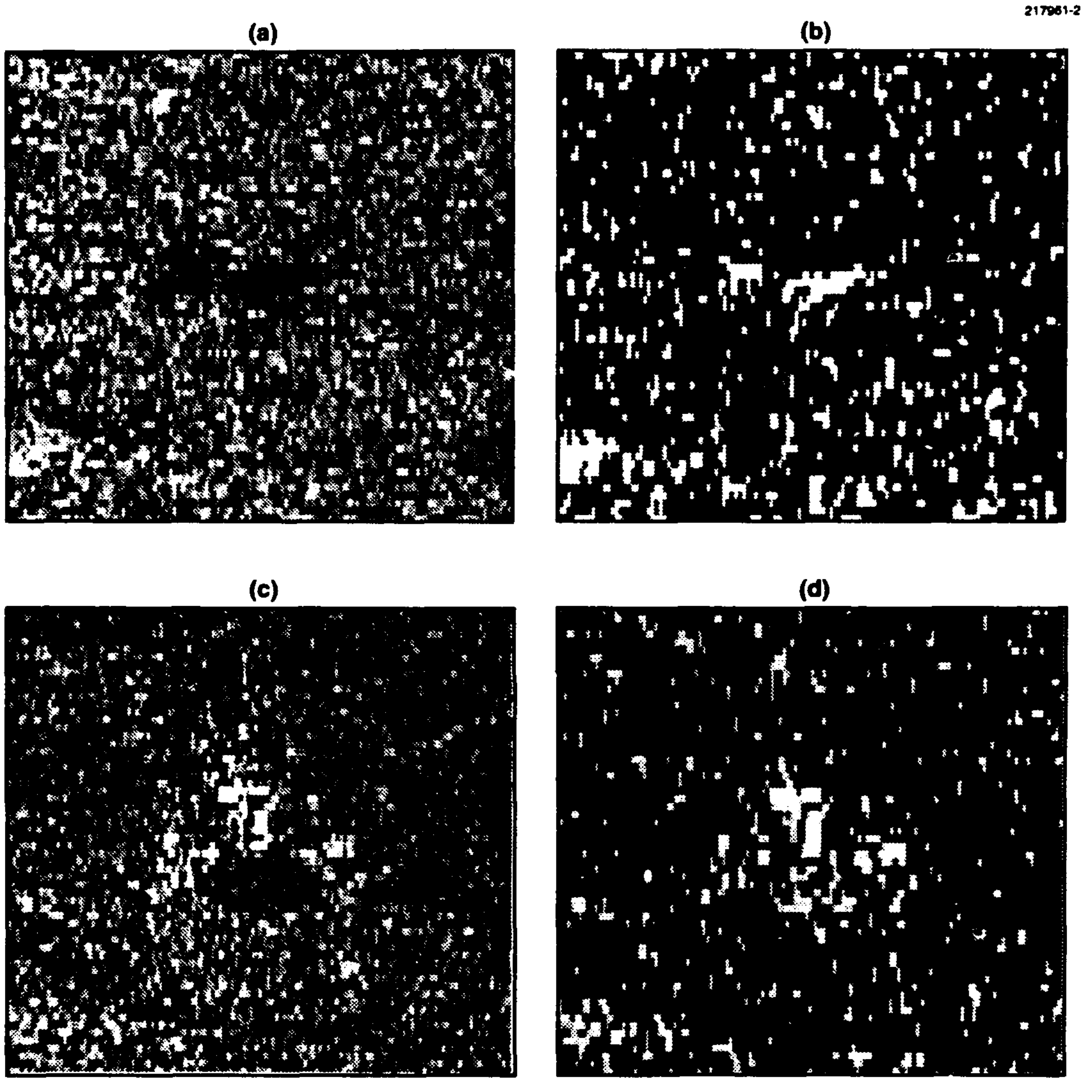

Figure 2. Raw and enhanced intensity images of an M113 armored personnel carrier oriented 45 degrees to the sensor [(a) and (b)] and a frontal view of an M48 tank [(c) and (d)].

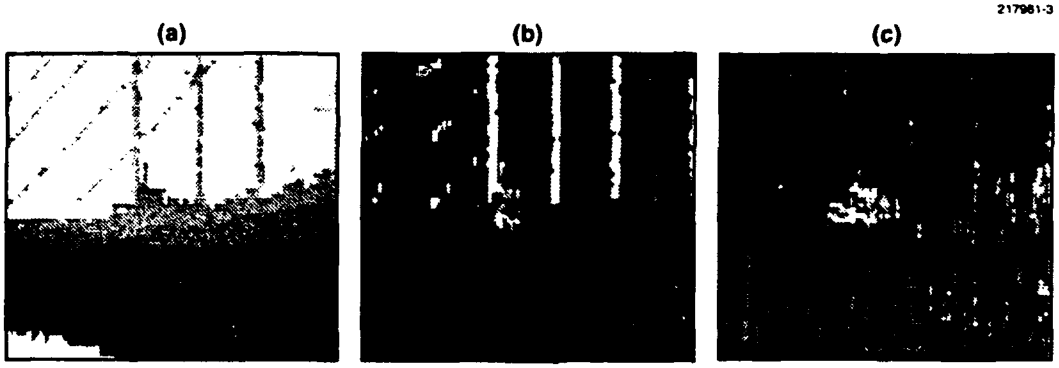

Figure 3. Height-limited verticality interest image: (a) range image, (b) verticality, and (c) height-limited verticality.

Note that as the range varies from pixel to pixel; the lengths (in pixels) of $H_{8.0}$ and $H_{2.5}$ must also vary. Details of mechanisms and problems associated with techniques of *adaptive morphology* will be discussed elsewhere [7]. The steps for generating the body interest image are illustrated in Figure 4.

4.4 ROD INTEREST IMAGE

The most indicative features of tanks and howitzers are often guns and antennas. Consequently, an image reflecting the locations of long thin shapes can also serve as an interest image. To find approximately vertical rods, the original range image is subtracted from the MM closing of R with a short horizontal structuring element: $R^H - R$. Similarly, approximately horizontal rods can be detected by $R^V - R$. The rod interest image RI is the union of these two intermediate results: $RI = (R^H - R) \cup (R^V - R)$. An example rod interest image is shown in Figure 5(d).

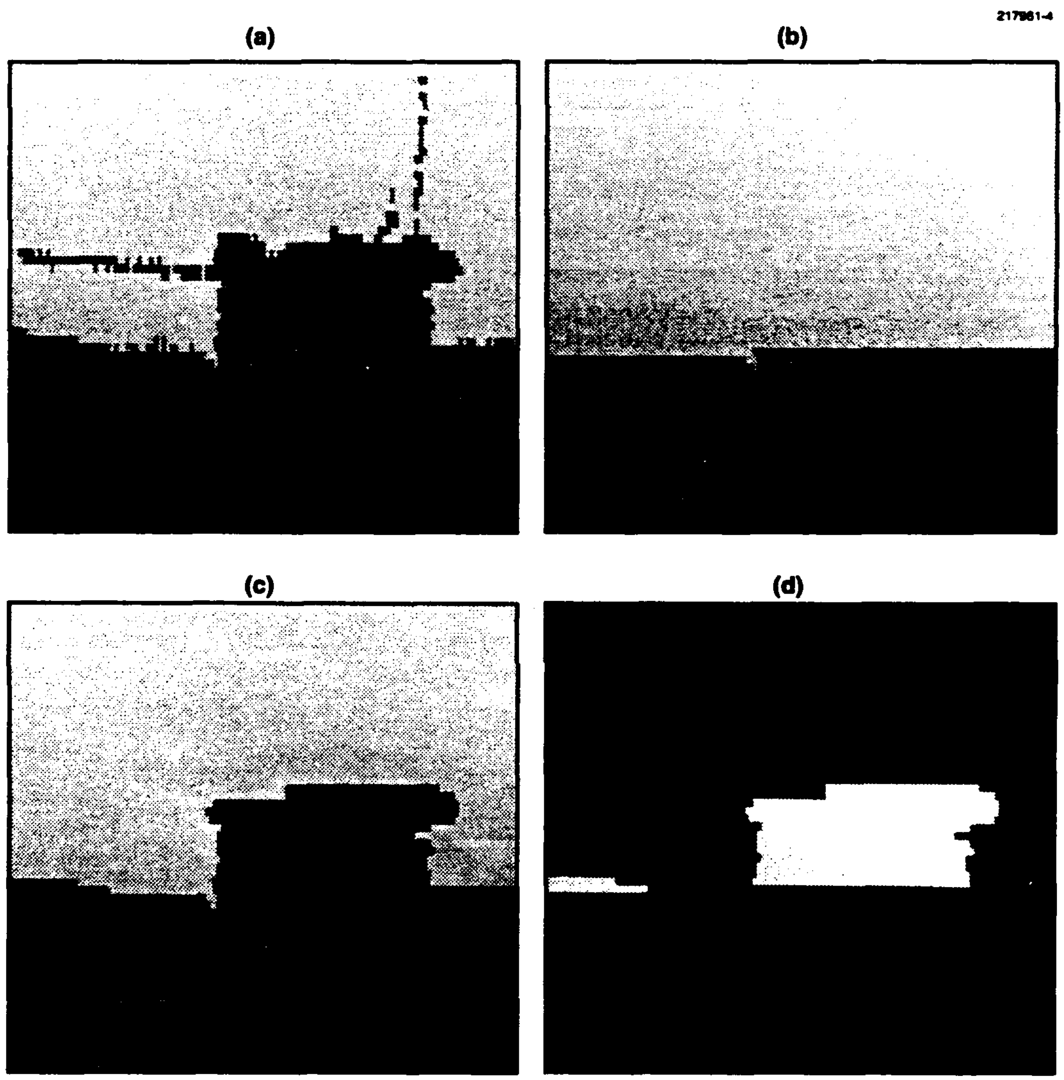

Figure 4. Body interest image: (a) original range image R, (b) long close: $R^{H_{6.0}}$, (c) short close: $R^{H_{2.5}}$, and (d) final body interest image: $R^{H_{6.0}} - R^{H_{2.5}}$.

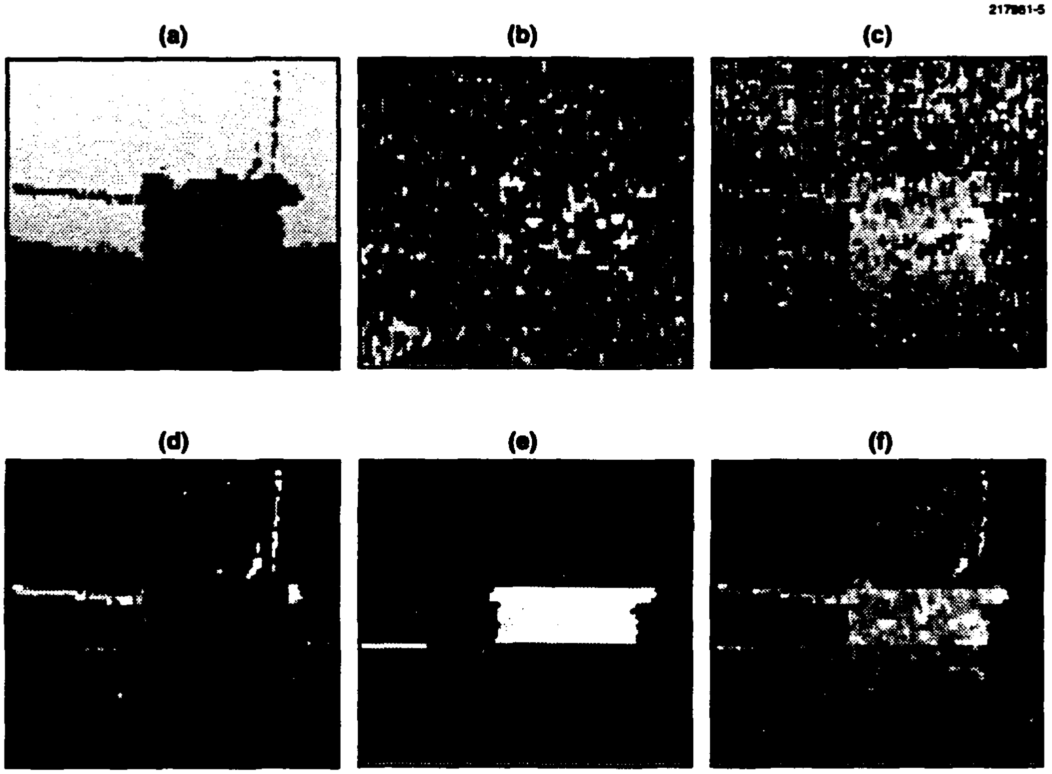

Figure 5. Fusion example: (a) original range image, (b) enhanced intensity, (c) height-limited verticality, (d) rod image, (e) body image, and (f) combined interest image.

5. FUSION OF INTEREST IMAGES

Here, interest is treated as a dimensionless, scalar quantity. By doing so, any combination of interest images can result in another interest image. Consequently, fusion is not restricted to a single step; already fused interest images can be fused with other interest images.

To conveniently combine interest images, they all must be scaled to a common range, such as real numbers in the range [0,1] or integers in the range [0, 255]. Once in this form, thinking of each interest value in terms of a fuzzy set membership may be useful [8]. Fuzzy set theory is a means of representing uncertainty or partial belief that an entity is a member of some set or class of objects. In the case of interest images, the interest (or membership) value indicates the degree of match between an image pattern and the feature being detected. Thus, in this sense, an interest image is a map of feature membership. Fusion can then be accomplished in a pixel-wise manner using tools of fuzzy logic, such as *fuzzy-and* (minimum), *fuzzy-or* (maximum), and *fuzzy-not* (subtraction from 1.0). Thus far, simple rules of fuzzy logic have been used for combining interest images. For example, in the version of XTRS developed for down-looking GaAs laser radar imagery, interest images are combined using fuzzy-or. However, arbitrarily complex rules of fuzzy logic can be also be used, perhaps reflecting knowledge about the interrelationship of the visibility of features.

The pixel-wise combining of interest images need not be limited to operations of fuzzy logic. In fact, arbitrary operators can be applied to all the interest values corresponding to a given pixel. In most of the experiments presented here, simple operations were used. For example, in the existing forward-looking version of XTRS, pixels are averaged across all interest images to generate the combined interest image. Although it has not been necessary to do so, numeric combinations of interest values can also be arbitrarily complex. For example, values from different interest images might be differentially weighted or conditionally included in the average. An example of all four interest images discussed above and the resulting averaged interest image is shown in Figure 5.

It is important to note that the individual feature detectors shown in the preceding section are each not particularly robust indicators of the intended targets. They do not highlight all instances of targets and do generate many false detections. But by combining interest images, these feature detectors essentially vote on the possibility that a target exists at the given location. Not surprisingly, combining the outputs of feature detectors in this manner results in better performance than any of the four feature detectors taken individually.

6. CLUSTERING AND WINDOW SELECTION

Given an interest image, target detection becomes a process of looking for local concentrations of high interest values. Generally, this entails some kind of clustering technique since a single target typically generates a constellation of high interest values. In XTRS, a window is created for each cluster, which defines the location and size of subimages to be generated for subsequent processing. The window also keeps a record of processing steps applied to the window. Such a record is essential for guiding feedback control.

For the version of XTRS dealing with forward-looking laser radar imagery, clustering is performed in two steps. The first step divides the range image into *range segments*, which are image regions each consisting of all pixels having values in a contiguous span of range bins. Range segments are selected by finding clusters of n contiguous range bins with high interest values. The value of n varies from 3 to 7, depending on the range precision of the sensor. Figure 6 shows an example of an image of range segments and the range and interest images from which it was derived. The range segment with the highest concentration of interest is shown in white, with successively darker shades of gray denoting less interesting segments.

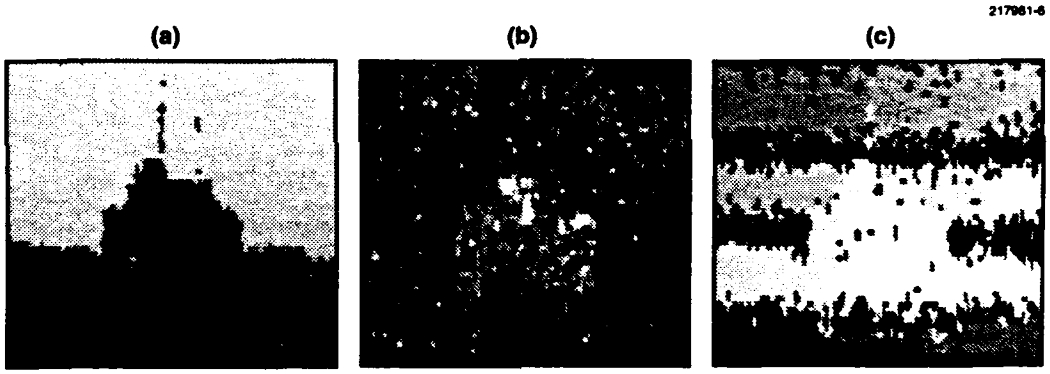

Figure 6. Range segmentation example: (a) range image, (b) combined interest image, and (c) image of eight range segments

When the object of interest dominates the field of view, the range segment image described sufficiently divides the image area. However, when the object occupies a small part of the field of view, and especially when other extraneous, noninteresting objects are present, it becomes necessary to further subdivide each range segment. Briefly, each range segment is subdivided into a nonoverlapping, contiguous array of *scoring boxes* whose sizes are determined by the size of the target and the average distance to the range segment (range segments closer to the sensor are divided into fewer, larger scoring boxes). All scoring boxes within a range segment are assigned a score equal to the the average of interest scores of pixels within the associated range segment. The highest scoring windows across all range segments are prioritized by score and passed to the extraction module for subsequent processing. An example of range segmentation, followed by window segmentation, is shown in Figure 7.

An alternative mechanism of clustering is used in the version of XTRS that deals with down-looking laser radar imagery. In this case, all pixels are in the same approximate span of range bins, making range segments unnecessary. Instead, clustering is done using MM dilation. First, the input interest image is converted to binary form by setting all above threshold interest values to 1 and setting all other pixels to 0. The binary image is then dilated using a circular structuring element whose diameter is about half the length of the vehicles in the set of intended targets. Windows are defined for each remaining disjoint region and are ranked by the highest interest value within each region.

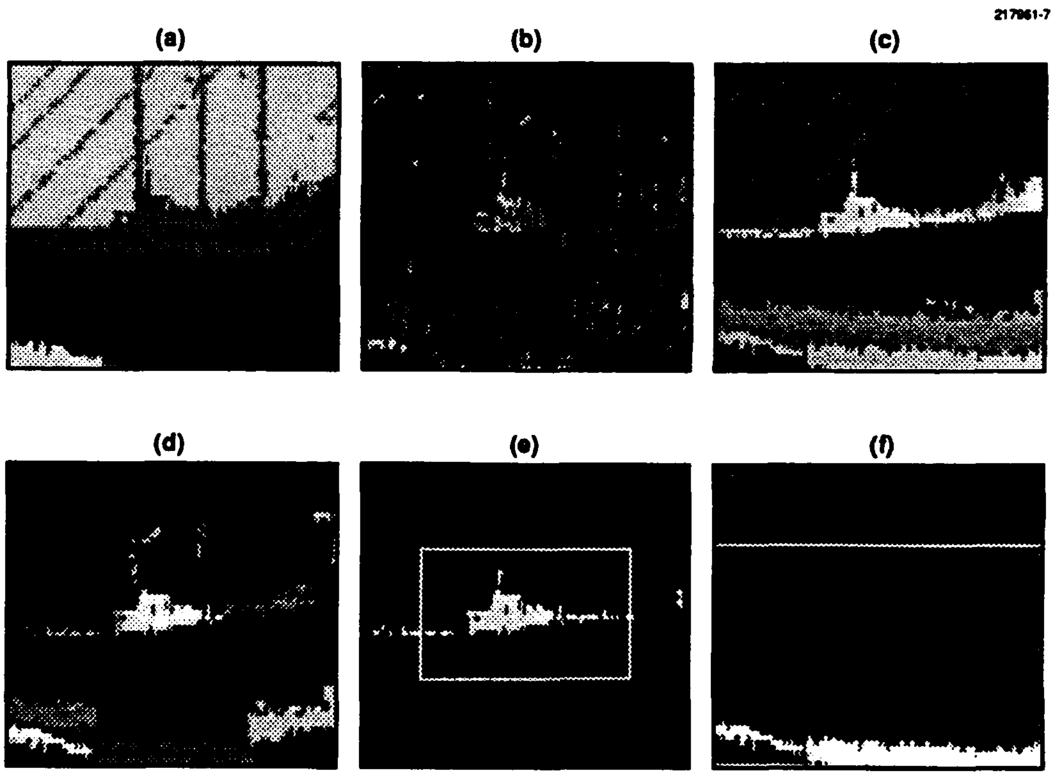

Figure 7. Window segmentation example: (a) range image, (b) combined interest image, (c) range segments, (d) highest ranking window segments, and (e, f) selected windows.

7. MODULE STRUCTURE

Each module in XTRS is governed by two rule-based experts (see Figure 8). The first one is called the *parameter selection expert.* Given knowledge of goal and context, this process evaluates the input stream of data and chooses parameters (algorithms, tunable coefficients, data structures) from a library for execution by the module engine. The second rule-based expert, called the *feedback expert*, evaluates the output of the module engine. Depending on the results of the evaluation, control may be passed on to the next module or may cause processing to be repeated with a different algorithm or a change of some tunable coefficient. As is suggested in Figure 8, there also exists a global feedback expert that controls long loop feedback from one module to another.

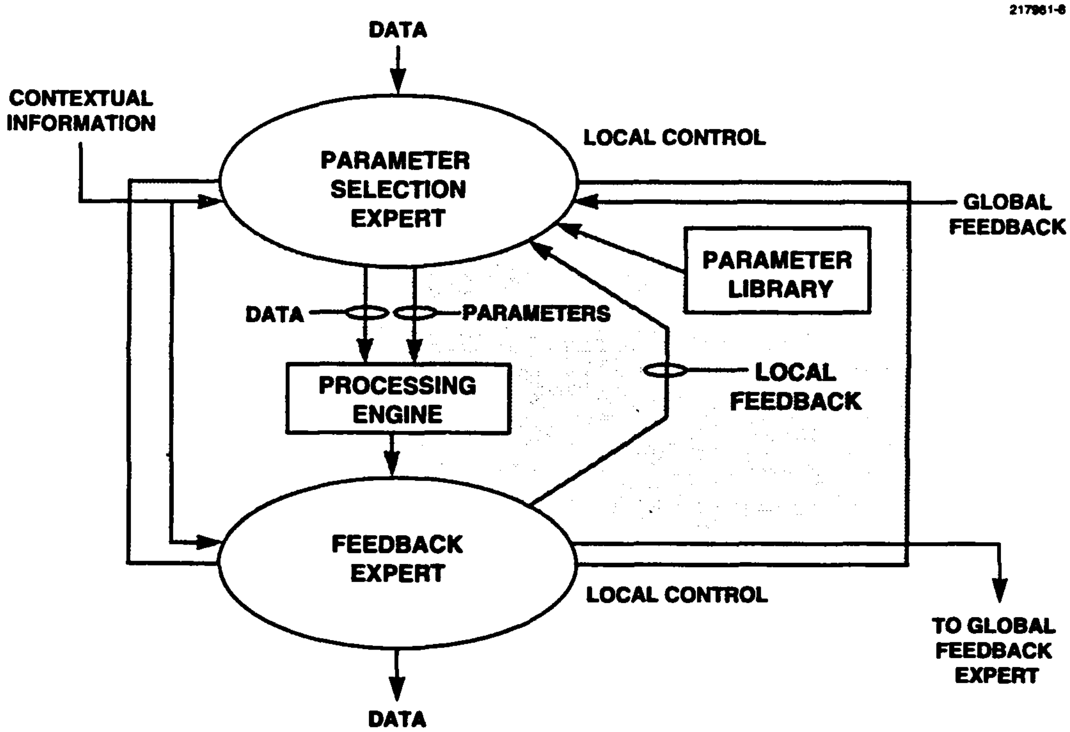

Figure 8. Generic module structure of XTRS.

For the detection module, the parameter selection expert selects the set of feature detectors that should be run by the detection engine. The interest images computed by the detection engine are then combined using simple rules of combination that vary with different versions of XTRS. Searching the combined interest image for clusters of high interest values generates a set of window specifications for subsequent extraction. Finally, the feedback expert determines the relative contributions of each feature detector to the overall interest score of a selected window. This information is used by the parameter selection expert of the extraction module to create a list of hypothesized target identities and to select appropriate extraction algorithms for each window.

8. CONCLUSIONS

Interest can serve as a "common denominator" for pixel-level fusion for target detection. Fusion can be as simple as pixel-wise averaging across interest images, or can be arbitrarily complex using fuzzy logic or arithmetic rules of combination. And by changing the set of interest images or the rule of combination, the combined interest image can dynamically alter its feature content to adapt to changes in mission or environment.

Given any interest image as input, target detection can be done by some variation of techniques for finding clusters of high interest values (examples of which have been presented above). And although not proven, the output of probably any feature detector can be configured as an interest image. If this premise is true, then it follows that universal detection engines, based on interest images, can be built that will work for any combination of features or targets, and for any combination of sensory modalities.

REFERENCES

1. W.N. Martin and J.K. Aggarwal, "Survey - dynamic scene analysis," *Comput. Graphics Image Proc.* 7, 356–374 (1978).
2. R. Jain and S. Haynes, "Imprecision in computer vision," *Computer* 15, 39–48 (1982).
3. H.P. Moravec, "Towards automatic visual obstacle avoidance," *Proc. 5th IJCAI* (1977) p. 584.
4. J.J. Otazo and E.W. Tung (private communication, 1981).
5. T.R. Esselman and J.G. Verly, "Applications of mathematical morphology to range imagery," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 797 (23 December 1987).
6. T.R. Esselman and J.G. Verly, "Feature extraction from range imagery using mathematical morphology," *Visual Commun. Image Proc. II* SPIE 845, 233–240 (1987).
7. J.G. Verly and R.L. Delanoy, "Adaptive mathematical morphology," accepted for publication by *IEEE Trans. Image Proc.*
8. L.A. Zadeh, "Fuzzy Sets," *Inf. Control* 8, 338–353 (1965).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
26 April 1993

3. REPORT TYPE AND DATES COVERED
Technical Report

4. TITLE AND SUBTITLE
Pixel-Level Fusion Using "Interest" Images

5. FUNDING NUMBERS
C — F19628-90-C-0002

6. AUTHOR(S)
Richard L. Delanoy, Jacques G. Verly, and Dan E. Dudgeon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-979

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
ARPA/ASTO
3701 N. Fairfax Drive
Arlington, VA 22203

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ESC-TR-92-206

11. SUPPLEMENTARY NOTES
None

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

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

A simple general-purpose means of representing image sensory data at the pixel level is presented in the context of target detection. One or more interest operators (i.e., feature detectors) are applied to one or more sensory images. The output of each operator is an *interest image* in which high pixel values are assigned to locations where features are found that are indicative of the object (or target) being sought. Any relevant sensory data that are transformed into interest images can then be fused in a pixel-wise manner by means as simple as computing averages, maxima or minima. The approach has been implemented in the detection module of an automatic target recognition system built at Lincoln Laboratory.

14. SUBJECT TERMS
data fusion; automatic target recognition; mathematical morphology; machine intelligence; low-level computer vision; fuzzy set theory; laser radar

15. NUMBER OF PAGES
32

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	Same as Report