Multidimensional Automatic Target Recognition System Evaluation

Paul J. Kolodzy

■ We are developing an evaluation facility that includes an electronic terrain board (ETB) to provide an effective test environment for automatic target recognition (ATR) systems. The input to the ETB, which is a high-performance computer graphics workstation, is very high-resolution data (15 cm in 3-D) taken with pixel registration in the modalities of interest (laser radar, passive IR, and visible). The ETB contains sensor and target models so that measured imagery can be modified for sensitivity analyses. In addition, the evaluation facility contains a reconfigurable suite of ATR algorithms that can be interfaced to real and synthetic data for developing and testing ATR modules.

A first-generation hybrid-architecture (statistical, model based, and neural network) ATR system is currently operating on multidimensional (laser radar range, intensity and passive IR) sensor, synthetic, and hybrid databases to provide performance and validation results. A recent study determined the sensor requirements necessary for target classification and identification of eight vehicles under various view aspects, resolutions, and signal strengths.

This article presents a description of the infrared airborne radar used to gather sensor data, a discussion of sensor fusion and the hybrid ATR measurement system, and a review of the ATR evaluation facility. This article also discusses the computer manipulation and generation of laser-radar and passive-IR sensor imagery and the processing modules used for target detection and recognition. We give results of processing real and synthetic imagery with the ATR system, with an emphasis on interpreting results with respect to sensor design.

The battlefield scenario continues to grow in complexity as the use of high-resolution sensors and precision strike weapons has forced the increased use of concealment and camouflage technology to improve vehicle survivability. The advent of multidimensional sensors that trade individual sensor performance for aggregate system performance and automatic target recognition (ATR) systems that can assist in or automatically identify targets also are a threat to vehicle survivability. The understanding of multidimensional sensors, the algorithms that are used to process their data, and the manner in which they are evaluated is necessary to determine their suitability for military applications.

Unfortunately, the testing and acceptance of ATR systems for military applications has proven elusive. On one hand, many researchers are concerned that not enough information exists in one sensor modality to build an ATR system that performs effectively against targets in natural and man-made clutter. On the other hand, the use of multisensor information to solve this vexing problem is relatively recent, and the results are limited. Although we have strong indications that several sensor modalities are better than one for target identification, no convincing database of evidence exists.

At Lincoln Laboratory we have constructed a flyable multisensor measurement system to evaluate the use of single and multiple sensor modalities for search-and-identification applications. This article describes the measurement system, which includes a forward-looking suite of sensors, a down-looking suite of sensors, and an MMW sensor. We also describe an ATR system for processing laser radar range and intensity imagery as well as other sensor modalities.

Testing the ATR system to quantify the performance limits of the multisensor measurement system is an important step in the development of useful benchmarks and the definition of radar requirements. This article examines the performance tests we have developed and provides a summary of test results for spatial extent, image quality, and 3-D recognition requirements. An ATR evaluation facility is currently under development to provide an effective test environment for ATR systems. The inputs to the facility, which is a high-performance computer graphics workstation and data-processing engine, are very high-resolution data (15 cm in 3-D) taken with pixel registration in the modalities of interest (laser range, intensity passive IR, and visible) and stored in databases. An electronic terrain board (ETB) combines the databases with sensor and target models to modify the measured imagery for ATR sensitivity analyses.

The Infrared Airborne Radar

The Infrared Airborne Radar (IRAR) is a flyable multisensor measurement system that consists of a set of active and passive infrared (IR) and active millimeter-wave (MMW) sensors. This system is installed in a Gulfstream G-1 twin turboprop test aircraft used by Lincoln Laboratory; Figure 1 illustrates the locations of these sensors in the aircraft. We are especially interested in the ability of the multisensor measurement system to detect targets autonomously (i.e., without human interaction with the measurement system).

In the forward-looking sensor suite, the active laser

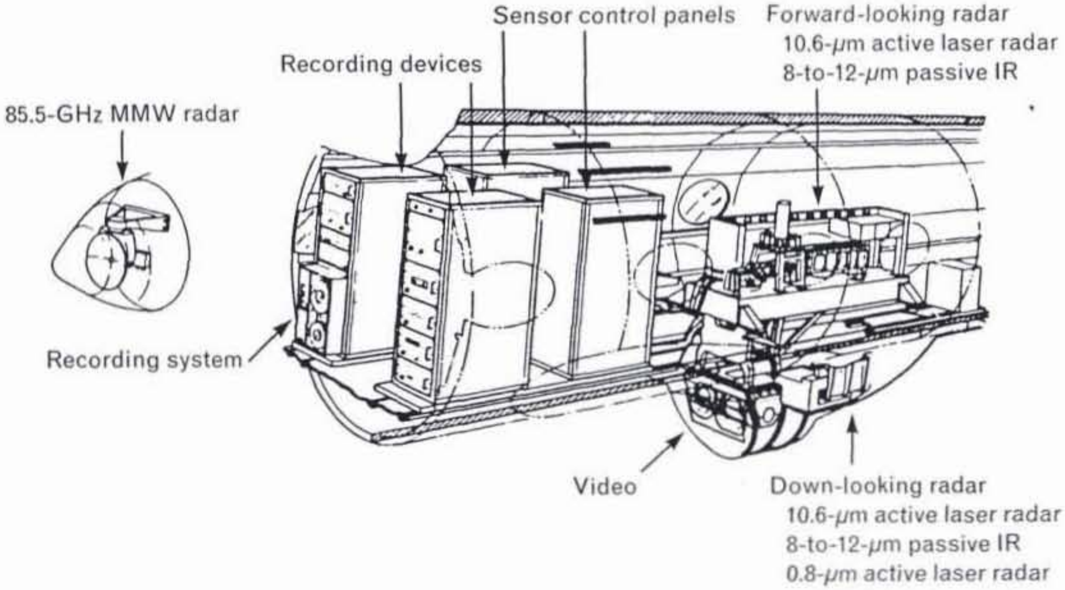


FIGURE 1. Schematic diagram of the multisensor measurement system on the Gulfstream G-1 aircraft, showing the location of each individual sensor system. The two sensor suites—forward-looking and down-looking—are located in the aft section of the aircraft, the recording system and electronic racks are located in the midsection, and the antenna for the MMW radar is located in the nose. The forward-looking sensor suite is mounted on an optical table and then relayed through a pod on the fuselage. The down-looking sensor suite is housed entirely in the pod aft of the forward-looking sensor system.

radar sensor measures absolute range with a precision of 1 m while the passive-IR sensor measures the thermal intensity of the target and scene in the 8-to-12-μm band. The down-looking sensor suite, which is a multispectral active-passive sensor, has the ability to measure relative range with a precision of 15 cm, as well as the ability to measure passive-IR thermal intensity. In addition, an MMW real-aperture measurement system developed by General Dynamics of Pomona, California, is installed in the aircraft. This MMW sensor measures absolute range with a resolution of 0.5 m, and is slaved to cover the same search area as the forward-looking sensor.

All the IRAR sensors reside on board the aircraft platform. The heart of the IRAR system is located in the center section of the aircraft. A radome extends down from the center of the aircraft, allowing the laser beam of the forward-looking sensor to exit through a germanium window on the left side of the radome. An additional window immediately to the right of the germanium window is used by the measurement system's boresighted color television camera, which is used to point the laser beam manually and to record a live sequence of the measured scene.

The radome was modified so that the down-looking sensor could be placed immediately behind the forward-looking laser-radar pointing-mirror assembly and look straight down; the scan direction of the down-looking sensor is therefore always perpendicular to the longitudinal axis of the aircraft. The MMW system is sufficiently small so that the 1-ft diameter radar dish and the gimbal mount are totally enclosed within the nose cone of the aircraft.

Forward-Looking Laser Radar

The transmitter in the forward-looking sensor is an RF-excited, water-cooled, CO_2 waveguide laser operating at 10.6 μm. In the pulsed mode, the transmitter laser provides a nominal 25-nsec pulsewidth at approximately 3-W average power at a pulse-repetition frequency of 20 kHz. In CW operation, the laser can provide power in excess of 30 W.

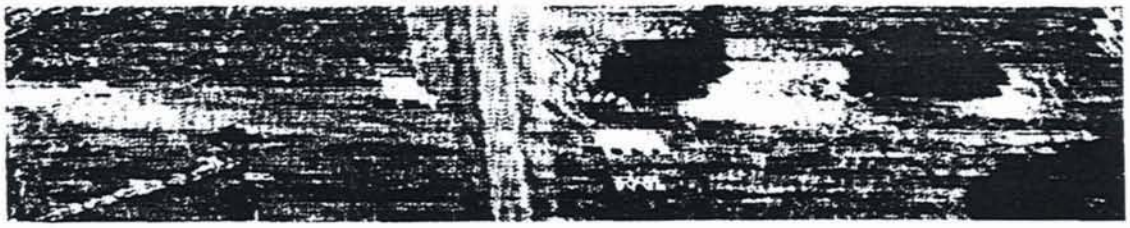
A 5-in diameter, afocal, Ritchey-Chretien telescope functions both as the transmit and receive aperture of the sensor to produce a 200-μrad diameter beam (100-μrad resolution). The sensor uses two linear 12-element arrays of HgCdTe photovoltaic detectors: one array for the active measurements and one for the passive measurements. Registration of the active and passive measurements is always assured because both arrays share the common telescope.

In the present configuration, the two arrays are oriented vertically to provide a 10° azimuthal coverage at 2.5 scans/sec in linescan mode. In a separate framing mode (25.6 mrad by 12.0 mrad), the scanning mirrors operate at 20 frames/sec; when the passive channel is enabled, however, the recording rate is reduced to 10 frames/sec because of recorder limitations. Television images from the boresighted TV camera are digitized and stored on computer tapes. Table 1 shows selected system parameters for the forward-looking sensor.

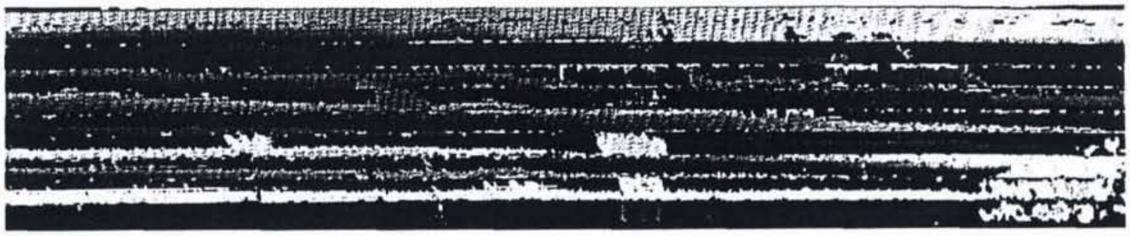
Figure 2 is an example of a laser radar range image and a passive-IR image made simultaneously by the forward-looking sensor. Two features in these images are particularly interesting with respect to data fusion and scene understanding: (1) the road that traverses vertically in the center of the scene is clearly visible in the passive-IR image in Figure 2(b) but invisible in the range image in Figure 2(a) because the road is at the same elevation as the local ground plane, and (2) although a tank (at the center left of the scene) has a negative passive-IR contrast with the background, it has positive range contrast in the active laser radar image. We can overcome the measurement limita-

Table 1. Forward-Looking Laser Radar System Parameters

CO_2 laser	
Wavelength	10.6 μm
Nominal power, CW	30 W
Pulsed, average	3 W
Number of detectors	12
Telescope aperture	13 cm
Instantaneous field of view	0.2 mrad
Range sampling interval	1.1 m



(a)



(b)

FIGURE 2. (a) Passive-IR imagery and (b) laser radar range imagery taken simultaneously at Stockbridge, New York, by the forward-looking sensor. The passive-IR image in part *a* is coded by thermal intensity, so that warmer objects such as vehicles are brighter than cold objects. The range image in part *b* is coded by color to distinguish objects at different distances from the viewer.

tions of each individual sensor by fusing the information from the two sensors to provide enhanced detection capability.

Millimeter-Wave Radar

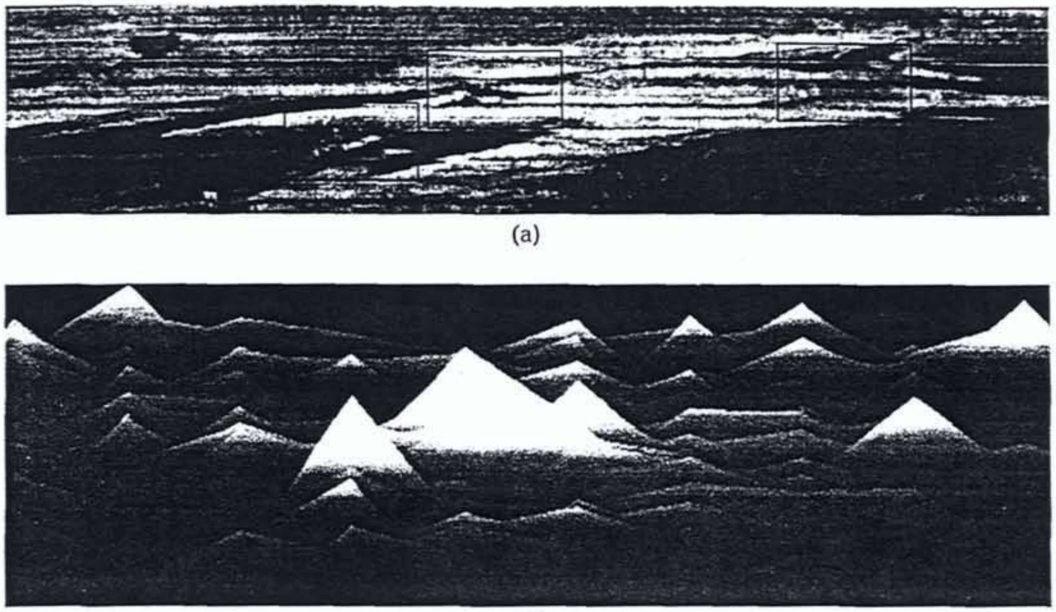
To investigate the advantages of combining the output of two or more diverse sensors, we added the General Dynamics 85.5-GHz real-aperture MMW radar to the forward-looking sensor suite. This radar has low cross-range resolution and high line-of-sight resolution, and operates at 3.5 mm. Table 2 lists the operating characteristics of this radar. The MMW antenna is mounted in the nose section of the aircraft and is boresighted to the IRAR sensor suite during pointing-mode operation.

The cross-range resolution of the MMW radar is such that a 10° azimuthal field of regard is stored as 15 intensity-versus-range profiles on each scan. The oversampling that occurs in the down-range dimension is then used to enhance the processing statistics for detection. The modulation characteristics of the sensor are such that the line-of-sight range resolution is 1.7 ft, while data are sampled at approximately half this value, thus providing the potential for excellent range resolution on the target.

Figure 3 illustrates the range resolution of the MMW radar in combination with a passive-IR imaging sensor. Three logging trucks in the passive-IR image in Figure 3(a) are each highlighted by a box. The environmental conditions at the time the data were taken are responsible for the low passive-IR con-

Table 2. Millimeter-Wave Radar System Parameters

Operating frequency	85.5 GHz
Transmitter power	15 mW
Modulation format	FMCW
Antenna diameter	12 in
Antenna beamwidth	0.76°, one way 0.57°, two way
Range resolution	1.67 ft
PRF	1600 Hz
Noise figure	20 dB, including system losses



(a)

(b)

FIGURE 3. (a) Passive-IR imagery and (b) boresighted MMW radar imagery. The MMW radar data are displayed as a 3-D plot of down-range, cross-range, and thermal-intensity values. The three logging trucks indicated by boxes in the passive-IR image correspond to four of the five highest MMW radar intensity peaks. Part *b* shows two peaks for the one truck in the center of part *a* because distinct returns were obtained from both the truck cab and the truck bed.

trast. If the MMW radar signal is displayed as a 3-D image (cross-range, down-range, and intensity), however, as shown in Figure 3(b), then four of the five highest intensity peaks shown in the figure correspond to radar returns from target locations. Two peaks are determined in Figure 3(b) for the truck in the center of Figure 3(a) because we obtained strong distinct returns from both the truck cab and the truck bed.

Down-Looking Laser Radar

The multispectral active-passive down-looking sensor is a compact multiple-channel system that employs two lasers for active detection and a single passive detection channel. This sensor is configured with a 10.6-μm amplitude-modulated continuous wave (AMCW) CO_2 laser and a 0.8-μm AMCW AlGaAs diode laser for the two active channels, which are coregistered with an 8-to-12-μm passive detection channel.

The system was designed with 1-mrad angular resolution to provide a 15-cm cube on the target from an optimal measurement height of 150 m. The active-channel lasers are modulated at 15 MHz to provide an AMCW waveform that translates to a 10-m range ambiguity but provides 15-cm precision (i.e., the range values are produced from 0 to 10 m in 15-cm increments and they fold over at the range boundaries). Thus these measurements are relative range measurements with 15-cm precision, as compared with the absolute range measurements of the forward-looking sensor. Table 3 lists selected parameters of the multispectral active-passive down-looking sensor, and Figure 4 shows five separate images produced by this sensor during a flyover of the USS *Connole*.

The multispectral down-looking sensor has two characteristics of interest for the development and testing of ATR systems: (1) the viewing aspect allows the imaging of objects in clutter that are not generally seen by forward-looking sensors, and (2) the highly precise range imagery gives us the capability to transform the observed scene to a variety of viewing aspects. Figure 5 illustrates this process. Figure 5(a) contains a photograph of a truck that is camouflaged

Table 3. Down-Looking Laser Radar System Parameters

Angular resolution	0.5 mrad, *x* and *y* axes
Range precision	15 cm
Range ambiguity interval	10 m
Altitude range	400 ft to 1300 ft
Ground coverage	2000 ft at 1000 ft

by netting and parked on a dirt road in a forest. Figure 5(b), which is the down-looking range image, clearly shows the road and the truck, with the height of the truck above the road encoded in color. Figure 5(c) is a computer-transformed forward-looking range image of the camouflaged truck from a viewpoint that is just above the road. In this way, a down-looking view can be used to develop or test algorithms for a forward-looking or near-forward-looking sensor through the use of coordinate transformations. A more detailed description of how down-looking data can be utilized for a variety of ATR evaluation tasks is given in the section entitled "The ATR Evaluation Facility."

Sensor Fusion

Figure 3 illustrates the possible benefits of fusing MMW radar imagery and passive-IR imagery. This figure demonstrates that the MMW radar image can be used to indicate areas of interest in a coregistered passive-IR image. Other techniques that incorporate the detection lists from both sensors usually fuse the lists by an OR or AND procedure; i.e., the target

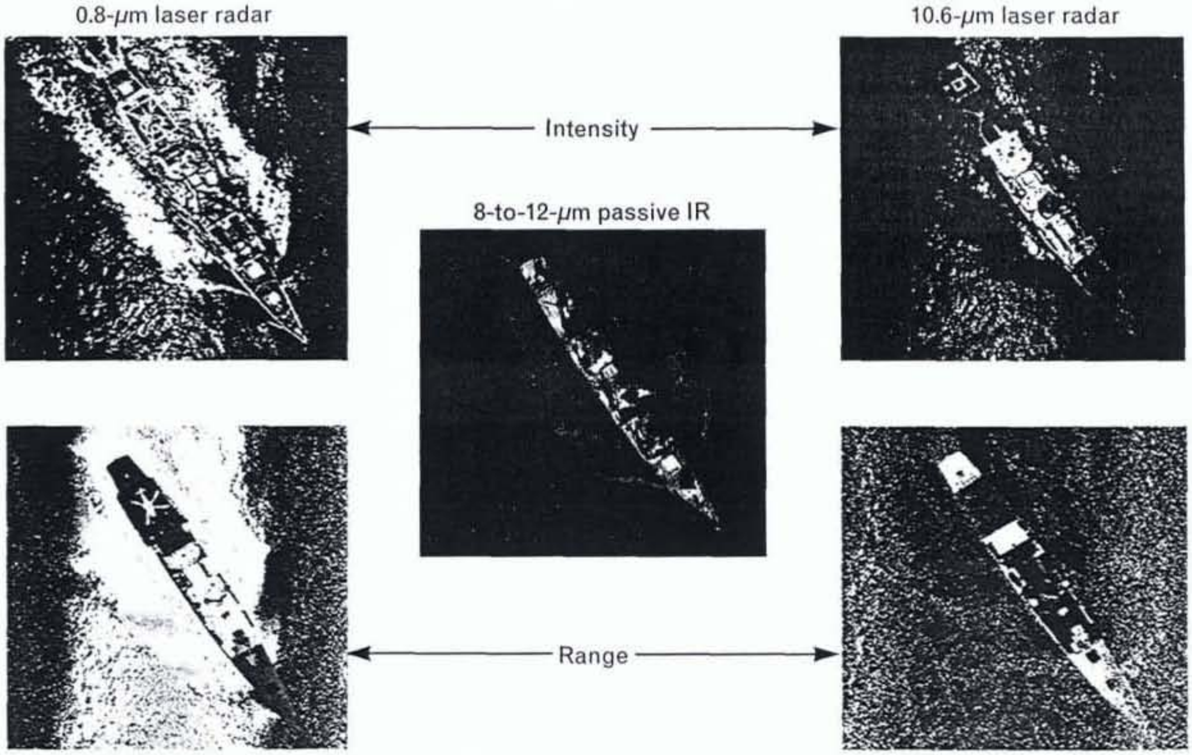


FIGURE 4. Example of imagery produced by the multispectral active-passive down-looking sensor during a flyover of the USS *Connole*. This sensor produces coregistered laser radar range and laser intensity images for wavelengths of 0.8 μm and 10.6 μm, as well as an 8-to-12-μm passive-IR thermal-intensity image. Note the parked helicopter near the stern of the ship in each of the sensor domains as well as the depiction of the ship's wake.

FIGURE 5. (a) Optical photograph of a truck covered with camouflage netting on a road in a forest. (b) The relative range image of the truck as determined by the multispectral down-looking sensor. (c) The 3-D spatial transformed image illustrates the relative range image data in part *b* as viewed from a depression angle similar to that of the optical photograph in part *a*.

must be detected at least on one list (OR) or on all lists (AND). The OR procedure produces a higher likelihood of detection at the expense of a high false-alarm rate. On the other hand, the AND procedure has a low false-alarm rate at the expense of a lower likelihood of detection. The next section describes an AND procedure that fuses sensor data to create a range-passive histogram, and the following section describes a maximum-likelihood fusion estimate for object detection.

Range-Passive-IR Histogram

Target cueing and detection can be accomplished with range data alone, with a range-only histogram, or with a range-passive-IR histogram (which is created by using an AND operation to fuse range and passive-IR data registered at the pixel level) [1]. The range-only histogram is a 3-D mapping of the number of occurrences of a range value plotted in a coordinate system of cross-range versus down-range. The histogram is calculated by scanning the range image pixel by pixel and adding one count to the histogram bin that corresponds to the pixel azimuth and the pixel down-range value:

$$H_{RR}(az, rng) = \sum_{el} U(az, el),$$

where

$$U(az, el) = \begin{cases} 1 & \text{if } R(az, el) = rng \\ 0 & \text{otherwise} \end{cases},$$

and where $R(az, el)$ is the absolute range of that pixel, and *rng* is the specific range value. Peaks in the 3-D range-only histogram indicate regions of significant vertical extent, or *verticality*, in the image, and the magnitude of the peak represents the vertical surface area of an object in the image. The property of verticality is effective for finding targets in open terrain; it produces a large number of false alarms, however, when applied in wooded areas.

The passive-IR thermal intensity can be used as a discriminant to separate trees from man-made targets that have a significant positive thermal signature. Pixel-level fusion of the range image data and the passive-IR image data is possible because each pixel of the range and passive-IR images is collocated. Each passive-IR pixel can be registered, according to its associated range value, to compute what we define as a *range-passive-IR histogram*.

The range-passive-IR histogram is a 3-D mapping of the sum of the passive-IR intensities plotted in cross-range versus down-range coordinates derived from the pixel-registered range image. Figure 6 shows an example of a range-passive-IR histogram. In Figure 6(a), a passive-IR intensity histogram is calculated for each column, which corresponds to a particular cross-range value that uses both the range image to provide the coordinates for the histogram and the pixel-registered passive-IR image for the intensity values. An azimuth value is selected, and then we scan the range image pixel by pixel along that azimuth column, where the range value for each pixel selects the histogram range bin. The corresponding passive-IR intensity value in Figure 6(b) is then added to that histogram bin. In this way, a three-dimensional histogram (cross-range, range, passive-IR intensity) is created, as shown in Figure 6(c). Peaks in the histogram indicate objects with vertical extent (i.e., trees, buildings, and vehicles) and with sufficient thermal contrast with respect to the background (i.e., running engines, heated buildings).

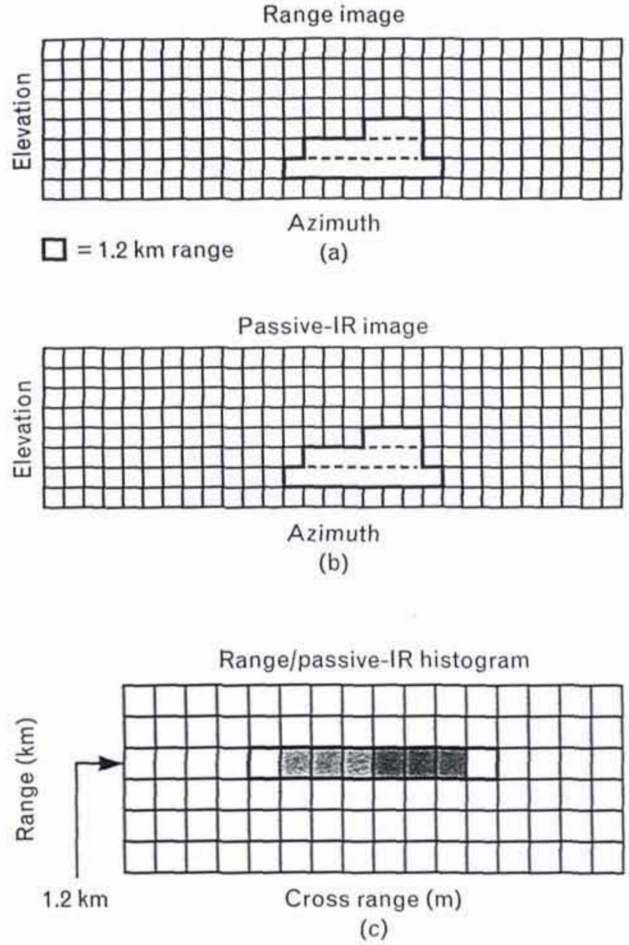


FIGURE 6. Schematic diagram of how the range image and passive-IR image are mapped into a range-passive-IR histogram. (a) An azimuth value is selected, and the range image is scanned pixel by pixel along that azimuth column; the range value for each pixel selects the histogram range bin. (b) The passive-IR intensity value from the corresponding passive-IR image column is then added to the histogram bin. (c) In this way a three-dimensional range-passive-IR histogram (cross-range, range, passive-IR intensity) is created.

This calculation is written as

$$H_{RP}(az, rng) = \sum_{el} P(az, el) \times U(az, el),$$

where $U(az, el)$ is as defined previously and $P(az, el)$ is its processed passive-IR intensity. Peaks in the range-passive histogram indicate regions of vertical extent that have positive thermal contrast.

Figure 7 shows how the range-passive histogram algorithm was applied to an IRAR linescan scene taken at Fort Devens, Massachusetts. The linescan scene contains the passive-IR image and laser radar range image of three trucks and a motor generator set. The vehicles were not in operation; their thermal signature is due entirely to solar heating. Figure 8 shown the resulting range-passive histogram. The three largest peaks correspond to the three trucks in the scene. For each peak, the truck position is now localized in cross-range and down-range. This example clearly shows the value of fusing multiple sensor domains at the pixel level with an AND operation, which improves the probability of detection and lowers the probability of false alarms.

Theoretical Study of Active-Passive Detection of Multipixel Targets

Research into the development of a quasi-optimal, single-sensor detection processor for multipixel laser radar was done by M. Mark [2] and resulted in the generation of receiver operating-characteristic curves for this processor. Mark used a generalized-likelihood ratio test to estimate unknown parameters for a maximum-likelihood estimate. Computer simulations with benign synthetic scenes, generated with uniform laser intensity, range, and passive-IR values for target and background, were used to provide performance measures. Recent extensions of this work to multiple sensor modalities (laser radar range and laser intensity,

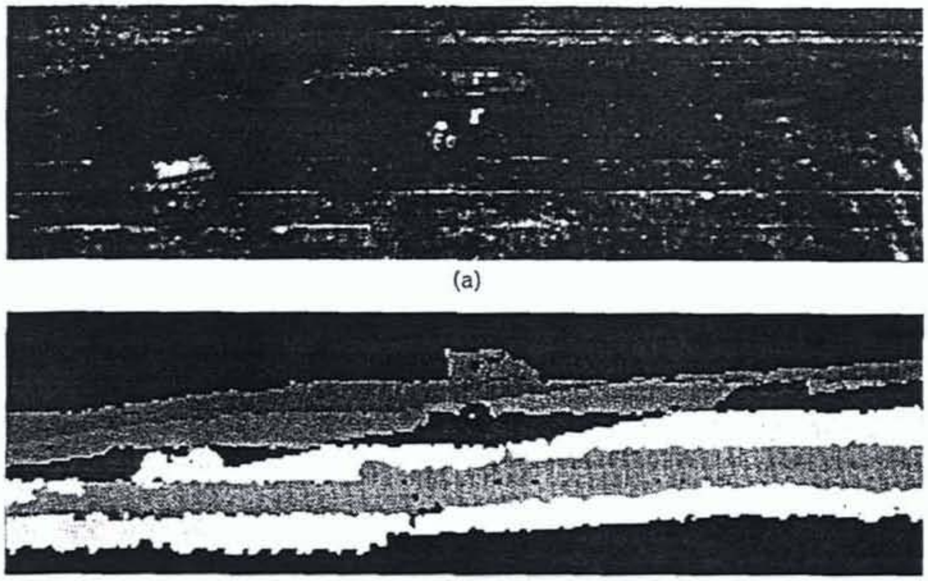


FIGURE 7. A scene containing three trucks and a motor generator as imaged by (a) the passive-IR sensor and (b) the laser radar range sensor in the forward-looking sensor suite. The trucks and generator are clearly visible in the center of the passive-IR image. The laser radar range image depicts the objects as silhouettes standing out of the sloping terrain and in the same location as in the passive-IR image.

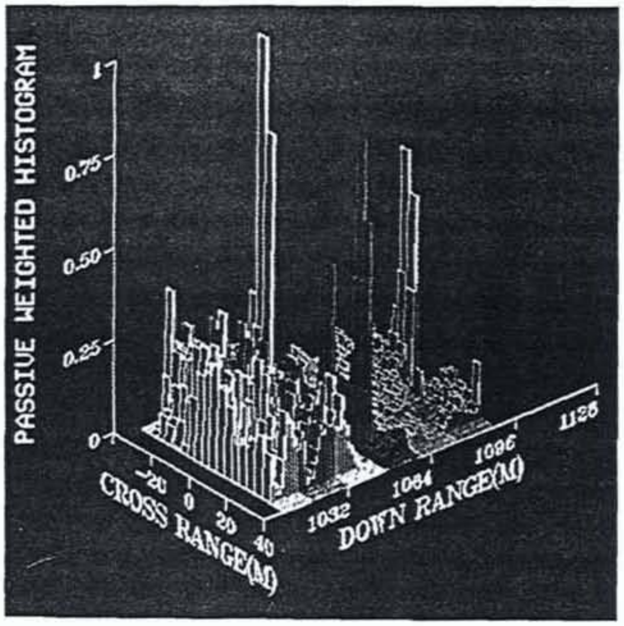


FIGURE 8. The result of processing the data in Figure 7 with the range-passive-IR histogram. The down-range values are color coded in the same manner as the laser radar range image in Figure 7. The four highest peaks correspond to the three trucks in the scene. These peaks would cue a classification processor to a region of interest.

passive-IR thermal intensity) were accomplished by S. Hannon and J. Shapiro [3]. The results of these computer simulations, which were later confirmed by experimental data, indicate that for a specified operating power such as the probability of detection and the probability of false alarm, the required sensor signal-to-noise ratios were relaxed for a multisensor measurement system over a single sensor system.

Figure 9 depicts the sensor/target requirements for a 10-pixel target (2 pixels by 5 pixels) on a 1000-pixel image (20 pixels by 50 pixels). The target size can be scaled to simulate a tank-sized vehicle at a distance of approximately 5 km with a sensor field of view (given a 6° depression angle) of 15,000 m^2. Figure 9 indicates the sensor requirements for detecting 99% of tank-sized vehicles at 5 km with a false-alarm rate of 10^{-3}, or of 0.1 km^{-2}. Because the simulations were done on idealized scenes, however, the results are not directly transferable to a specific sensor design. The trends still indicate a reduction for either passive-IR signal-to-noise ratios (SNR) or laser radar carrier-to-noise ratios (CNR) when a combination of two sensors is employed.

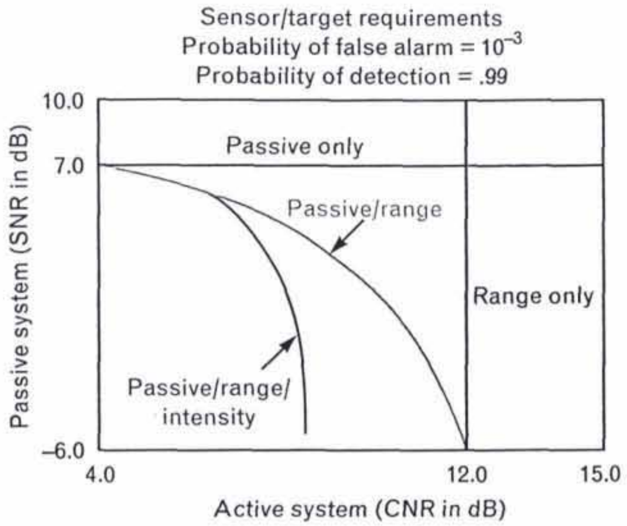


FIGURE 9. Sensor/target requirements for multipixel target detection using the generalized likelihood ratio test for single and multiple sensor modalities to detect a tank-size target at 5 km with a probability of detection of 0.99 and a false-alarm rate of 10^{-3} per image or 0.1 km^{-2} [3]. A 7-db SNR is required for a passive-only sensor system, and a 12-dB CNR is required for a laser radar range-only sensor system. The SNR and CNR requirements are relaxed for a combined passive-range sensor system or a passive-range-intensity sensor system.

Hybrid ATR System

We have developed ATR processing modules for the primary sensor groups described previously; these groups are laser radar intensity, range, passive-IR thermal intensity, and MMW. Although the individual processing modules can vary among sensor groups, the general processing structure has the same sequence of stages: cleanup, detection, segmentation, feature extraction, invariant mapping, and classification. The general ATR system was originally developed to operate on laser radar range and intensity imagery, and the results presented in this article are based on this imagery. Figure 10 illustrates the processing modules for the range-imagery recognition system; this system is described in more detail below.

Modular ATR System Concept

The unambiguous range image is first processed by the cleanup stage to reduce data anomalies and enhance the image. The cleanup stage attempts to reconstruct the most probable input image that would produce the measured sensor image. This reconstruction clarifies the image appearance, and makes the returns from the various objects in the scene appear more continuous and complete by reducing sensor and scene artifacts such as dropouts and anomalies.

Next, the enhanced image is processed by the detection stage to identify correctly sized regions of constant range as potential targets. The detection stage extracts these regions from the background clutter and removes the ground plane. The detected target at the output of this stage is a silhouette consisting of multiple fragments and rough boundaries.

The multiple fragments are combined by the segmentation stage into a complete, smooth, filled silhouette. The completed silhouette is then separated by the feature extraction stage into feature regions (e.g., barrel, turret, body, and tread for a tank). For this article, the entire target silhouette is considered the single feature. The silhouette is then mapped by the invariant-mapping stage into an abstract pattern that is invariant to translation, rotation, and scale within the sensor field of view. This invariant pattern is processed by the classification stage, which initially learns to cluster the invariant maps into groups and then, after the training cycle, classifies the input data with respect to its learned categories.

Image Cleanup

To provide adequate recognition performance in a noisy environment, the cleanup stage must be capable of using prior knowledge to restore measured images. We present here an image-restoration model that quantitatively incorporates prior knowledge of the measurement process and scene. The model is based on a Bayesian formulation using Markov random fields, as introduced by S. Geman and D. Geman [4]. The processing is massively parallel because the Markov-random-field assumption allows the image to be decoupled into a large number of connected local neighborhoods, each of which can be processed independently. The local-neighbor information is spread out in time such that a global image restoration is effected when the image-restoration system reaches a steady state.

Real-time image restoration is possible by using

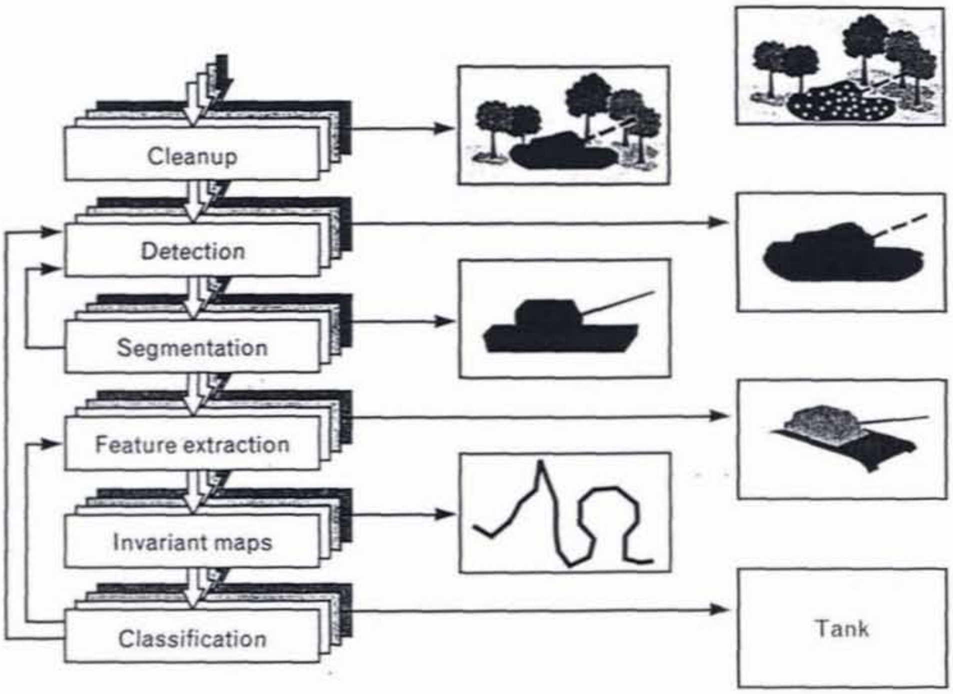


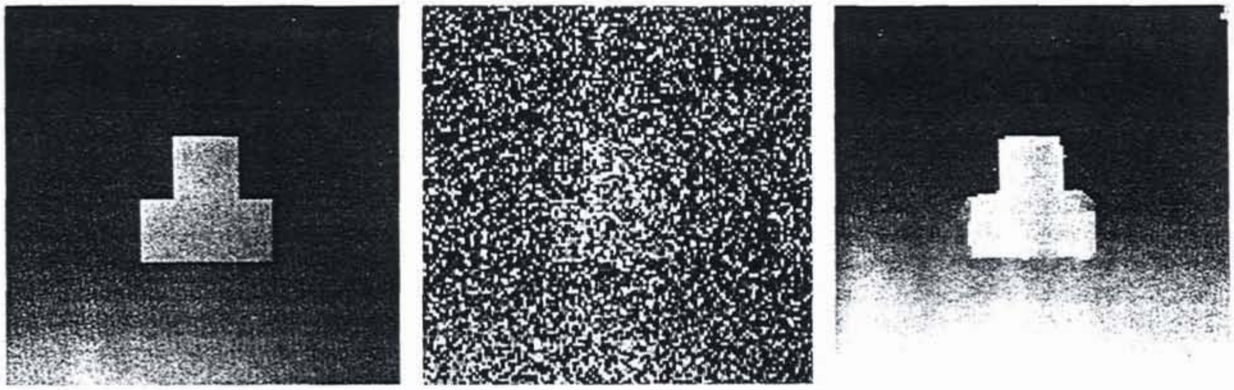
FIGURE 10. The six processing modules of the range imagery-recognition system: cleanup of sensor artifacts, detection of potential targets, segmenting targets to improve image characteristics, extraction of relevant features, invariant mapping of features to remove translation and rotation effects, and the classification of features into target categories.

the model with a massively parallel single-instruction multiple-data (SIMD) computer such as the Connection Machine or a direct hardware implementation on a custom microprocessor. A more detailed description of the image-cleanup process is given in this issue in the article by Murali M. Menon entitled "An Efficient MRF Image-Restoration Technique Using Deterministic Scale-Based Optimization."

We applied the image-cleanup process to a simple synthetic image corrupted with noise according to a measurement model described in the literature [3]. The noise does not have a Gaussian distribution and is based on realistic sensor measurements. The original noise-free synthetic image has a simple geometric shape at a constant pixel value, with a background that linearly increases in pixel value from the top of the image to the bottom. Figure 11(a) shows the uncorrupted image, while Figure 11(b) shows the image with 70% of the pixels corrupted with noise. The original image has 256 gray levels, and the noise spans the entire range of possible pixel values. Except for a few discrepancies at the boundary, the restoration shown in Figure 11(c) is nearly perfect, especially the recovery of the sloping background.

Target Detection

The detection stage of the ATR processing system extracts target-like regions from the enhanced range image produced in the cleanup stage. The process occurs in three phases: (1) regions of interest are selected, (2) target-like objects are detected, and (3) objects are extracted from the scene. Regions of interest are located by using range-only or range-passive-IR histograms, as previously described in the article. The peaks of these histograms indicate regions of significant vertical extent (i.e., constant range with varying elevation), or a significant thermal signature with some vertical extent. The selected regions are searched for areas of constant range that have range contrast with their neighbors and are similar in size



(a) (b) (c)

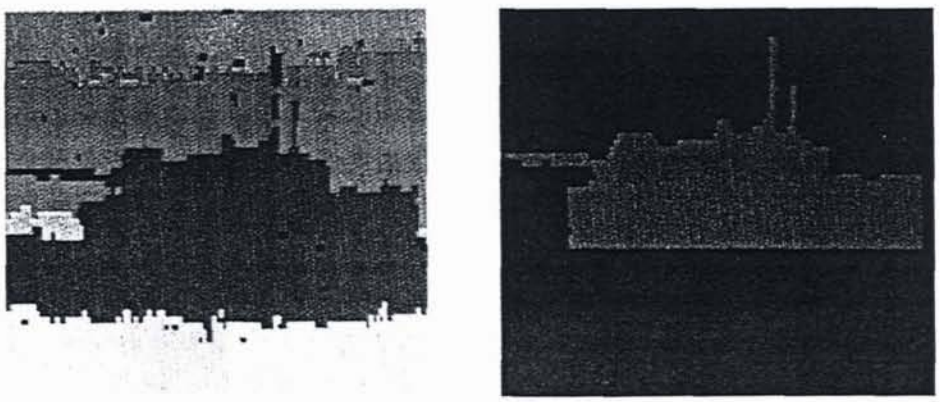
FIGURE 11. The effect of processing a synthetic laser radar range image of a geometrical object with the image-cleanup neural network. (a) The original noise-free image, (b) the image with 70% of the pixels corrupted with noise, and (c) the restoration of the original image from the corrupted image with only a few discrepancies at the image boundary.

(both in absolute height and width) to a target of interest. The target-like region is then separated from the background by selecting only the pixels with that range value. The object is then extracted from this selected image by computing and removing the ground plane.

Figure 12(a) shows the initial range image of an M48 tank at 700 m and the subsequent detection result that was formed by using the previously described range-only histogram and removing the ground plane. Figure 12(b) shows the M48 tank after the cleanup stage and detection stage of processing.

Segmentation

The segmentation stage of the ATR system smooths the boundaries and completes the fragments of the detected potential target. The boundary-contour system (BCS), a subsystem of a visual processing theory developed by S. Grossberg and E. Mingolla [5], is used to generate the perceived segmentation of the potential target, with respect to illuminance contrasts. The BCS system consists of two stages: an oriented-contrast (OC) filter and a cooperative-competitive (CC) loop. The OC filter measures local luminance



(a) (b)

FIGURE 12. (a) The initial range image of an M48 tank; (b) the subsequent detection result that was formed by using the range-only histogram and removing the ground plane.

differences, or edges, within an image at a number of different orientations. This filter models the orientation-selective cells discovered by D. Hubel and T. Wiesel [6] in the human visual system. These oriented edge strengths are then allowed to compete and cooperate with one another in the CC loop to generate the perceived boundary contours.

The four layers of the CC loop consist of two competitive layers, one cooperative layer, and a feedback layer. The first competitive layer thins and sharpens boundaries within the image by allowing competition for dominance in the final boundary segmentation between neighboring edge strengths of the same orientation. The second competitive layer straightens jagged or noisy boundaries by allowing competition between edge-contrast strengths with differing orientations at the same location. The cooperative layer completes and connects boundaries by allowing edges of like orientation to cooperate over a distance in the image. The feedback layer introduces into the system any new boundaries formed by the cooperative layer.

The OC filter is implemented by convolving a set of orientationally tuned digital filters with the input image. The CC loop is modeled by using a set of four coupled nonlinear differential equations for each orientation and location within an image. Input to the CC loop is static; therefore, the boundary is complete when the system of differential equations is in equilibrium.

The BCS algorithm has been previously applied to laser radar imagery as reported by Kolodzy et al. [7] and by E. Van Allen [8, 9]. Figure 13 shows an example of BCS processing on the range image of an M48 tank. The input range silhouette in the upper left of the figure is the input to the segmentation stage. The range-silhouette image is sampled to obtain the oriented contrast strengths by using the OC filter in each of twelve orientations. These oriented contrast strengths are then processed by the CC loop to produce twelve new images, which are compressed into a single image by using one of two methods: (1) compute the maximum contrast strength of any orientation at each pixel location (upper right of Figure 13), or (2) sum the contrast strengths across all the orientations at each pixel location (lower left of Figure 13). The compressed image is then filled to form a completed and smoothed silhouette of the potential target for classification. The specific filled image shown in the lower right of Figure 13 is the result of using the summing method for compressing the results of the CC loop.

Feature Extraction

In general, the filled and smoothed image provided by the segmentation stage of the ATR system is then used to extract relevant features for classification. Many different feature domains (images or vectors) such as image geometry, object parts, fractal dimensions, distance of hot spots from the central locations, and the Hough transform can be used and are part of ongoing research.

In particular, model-based systems have been developed at Lincoln Laboratory to parse images into geometric features (such as circles, squares, rectangles) and subsequently classify those features into target features (such as barrel, hull, turret). These model-based systems are discussed in the article by J. Verly et al. entitled "Machine Intelligence Technology for Automatic Target Recognition" [10]. The use of model-based systems for feature extraction in this ATR system has been evaluated previously by D. Dudgeon et al. [11], and they are not discussed further in this article. For our purposes, the pixel image provided by the segmentation stage is used as the feature for classification.

Invariance Mapping

The segmented silhouette is spatially mapped to eliminate translation, rotation, and scale variations prior to classification. An invariant silhouette is therefore built directly into the classifier memory to form a single compact representation for the target. This invariance reduces the number of stored patterns from one pattern for each of several terrain angles and ranges to a single stored pattern, which reduces memory requirements and search times and improves efficiency. Invariance can be obtained by using the following process: (1) locate the segmented silhouette in the field of view, (2) detect the silhouette edges, and (3) spatially map the silhouette edges. The resultant abstract pattern has the desired invariance and is processed di-

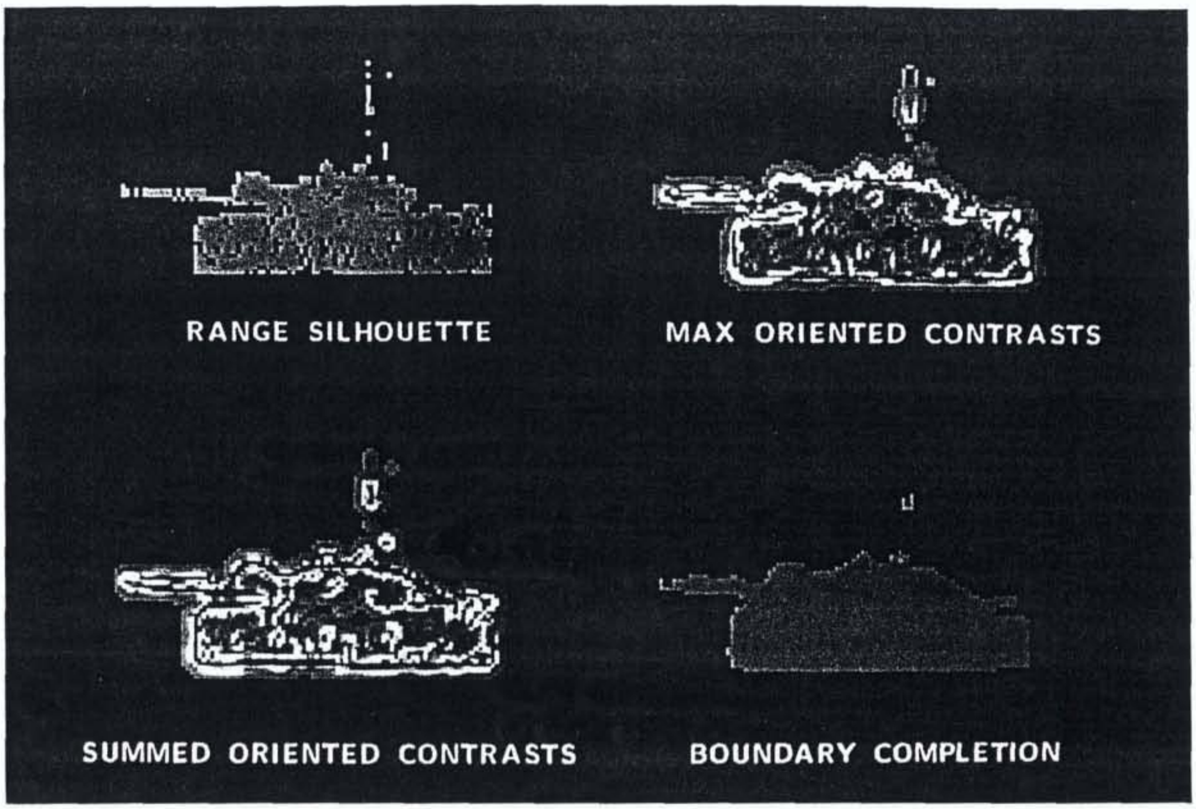


FIGURE 13. Segmentation of a laser radar image of an M48 tank with the boundary-contour system (BCS) showing smoothed boundaries and connected segments. The image in the upper left is the original segmented range silhouette. The result of applying the BCS using the maximum-contrast edge-strength method is shown in the upper right and the result using the summed-contrast edge-strength method is shown in the lower left. The summed-contrast edge-strength result was then filled and is shown in the lower right.

rectly by the classifier in the next stage.

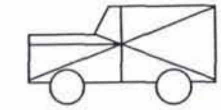
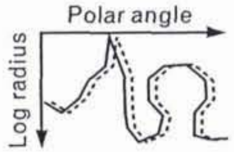
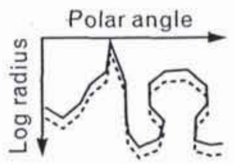
The target silhouette is located in the plane of the field of view by calculating a position-weighted sum, or *centroid*, of its pixel intensities. The centroid of the segmented silhouette is then used as the origin in the spatial mapping that follows.

The silhouette is next detected for edge strengths. The edge-detection algorithm uses contrast-sensitive oriented elliptical receptive fields. In this approach, the receptive fields are passed over the image to sum the pixel energy present in the area centered around each pixel. The major axes of the elliptical receptive fields are oriented in as many as twelve directions to calculate edge strength as a function of orientation. The output at each pixel in the edge image is the output value of the strongest receptive field orientation at that pixel. Because edge detection by receptive field processing is computationally intensive, there is a trade-off between orientation accuracy and processing time. The work described in this article was satisfactorily accomplished by using four orientations for the receptive fields.

The spatial-mapping function provides target rotation and scale invariance within the plane of the field of view. The function used in this work is a log-polar mapping of the edge-strength image about its centroid, as shown in Table 4. The log-polar mapping is biologically inspired by the visual field mapping of the human visual cortex, as demonstrated by E. Schwartz [12].

The log-polar mapping transforms the edge-detected range slice to a log-radius, polar-angle coordinate system by using the centroid of the silhouette as its reference point. Determining the location of the image centroid is important because the accuracy of the log-polar mapping is often highly sensitive to errors in the position of the centroid. Tests using noisy simulated imagery, however, have shown the

Table 4. Invariant Mapping of Silhouette Edge Strengths for Translation, Rotation, and Scale Invariance in the Sensor Field of View

Translation	Pixel centroid locates origin for mapping log radius and polar angle	
Rotation	Rotation in field of view is mapped to shift in polar angle	
Scale	Scale (or range) is mapped to shift in log radius	

centroid calculation and log-polar mapping to have robust behavior.

The log-polar mapping around the centroid of the target maps rotation in the field of view to a shift in polar angle, while it maps range to the target as a shift in log radius. This mapping is insensitive to rotation and scale variations; cross-correlation with an unrotated, unscaled log-polar mapping gives an estimate of the amount of rotation and scaling present in the detected silhouette.

Another method for making the log-polar mapping invariant to rotation and scale in the field of view is to calculate the magnitude of its Fourier transform. The shift property of the Fourier transform eliminates the rotation and scaling shifts of the log-polar mapping by treating the mapping as a periodic function, as reported for laser radar range imagery by Kolodzy et al. [7]. While this method has merit, it will not be discussed further here; it is the subject of other research [13].

Classification

In the final stage of the ATR system, a neural network is used to classify the abstract invariant maps into potential target categories. The adaptive resonance theory (ART) developed by G. Carpenter and S. Grossberg [14] defines a class of unsupervised neural network classifiers that cluster an N-dimensional input vector into a finite number of stable categories. This clustering is a necessity if large training sets are to be used.

Supervised networks and/or model-based systems require exact knowledge of the target, or *ground truth*, for each exemplar. For most large systems, thousands (or millions) of training frames would need to be ground truthed, which is a daunting if not impractical task. The ART-2 network, which is illustrated in Figure 14, is basically a two-level correlation classifier; this algorithm has been discussed by R. Lippmann [15] and Menon and Kolodzy [16] to be similar to the K-means clustering algorithm.

The ART-2 network is different from early ART structures because it is designed to classify analog, rather than binary, input patterns [17]. This analog capability requires a robust structure that pays strict attention to memory stability. The ART-2 network classifies and stores patterns in the following manner. The first layer (F1) normalizes the input with respect to the feedback signal from the second layer (F2), becoming a *short-term memory* (STM) trace. This trace activates nodes in the second layer proportional to the magnitude of its correlation with the corresponding

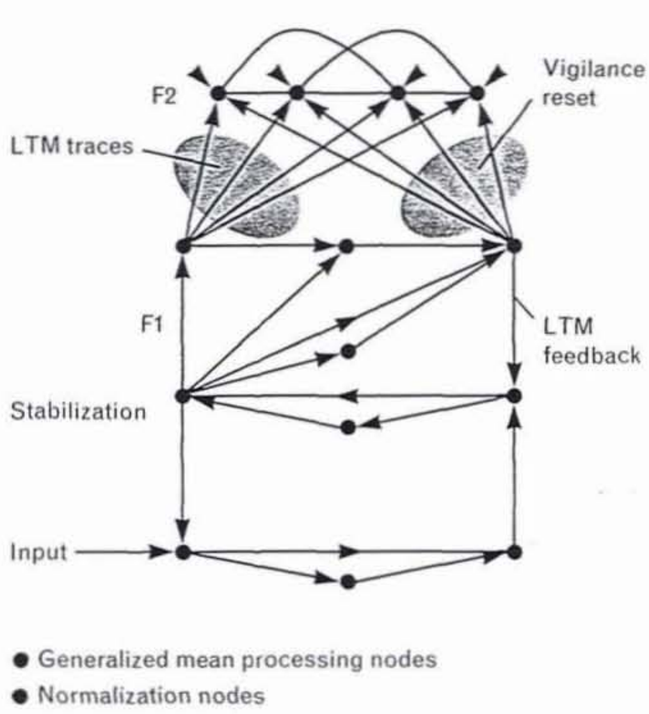


FIGURE 14. Processing diagram of the ART-2 neural network. The input image is presented and stabilized by the node-level processing in the F1 layer; the result is then correlated with stored long-term memory (LTM) patterns in the F2 layer. If the resultant correlation is not large enough with respect to the vigilance parameter, a reset signal is propagated as feedback down to the F1 layer. If none of the LTM patterns are sufficient, then a new LTM pattern is formed.

stored memory patterns, or *long-term memory* (LTM) traces. If the degree of match between the normalized input STM trace and the LTM trace associated with the most highly activated node in F2 passes a vigilance parameter, the STM trace in F1 is learned onto the LTM trace, thus storing the differences between these memory traces. Should a mismatch occur, a reset signal causes the input pattern to select another LTM category. If no existing LTM category can be found that matches the input pattern, a new category is created, which illustrates the ability of the ART-2 network to respond to a novel signal. Generally, similar patterns are categorized together because of high interpattern correlation, and these patterns continually activate the same category node in F2.

The final step of the classifier training is to associate target labels with LTM traces. Each LTM trace for an individual target is provided a label unique to that target. Multiple LTM traces for an individual target are formed because of either different views or statistical variability of the target. For example, a tank at a head-on perspective looks different from a tank at a broadside perspective, and thus would form two different categories. Also, at a low SNR the signal could change significantly enough to cause the classifier to form a new category if it is presented with a new noise structure.

Interpretation of Results

The performance of an ATR system can be indicated either by the score of each individual processing module or by the overall system score. This article uses the overall system score as a measure of results, with a higher concentration on the capabilities of the classifier. For supervised classifiers, the performance is commonly measured by the number of correct responses of the system when it is given a test set of input images. The ATR system presented here incorporates an unsupervised classifier, which uses a larger variety of performance measures. This article uses a scoring method based on the number and population distribution of categories created by the classifier during training.

Unsupervised classifiers are generally clustering algorithms that group input feature vectors into a finite number of categories. A user-defined distance metric is used to determine whether an input vector is to be clustered, or *matched*, with an existing category. The number of categories produced (given a specific training set) indicates the ability of the classifier to *generalize*. A classifier responding with more than one category for an object is not unreasonable if the features the ATR is extracting change significantly. For the ATR system presented here, this change in features occurs for the log-polar map when the vehicles are rotated out of plane. A classifier that requires only a few categories to perform recognition is desirable.

Two measures of a classifier's ability to generalize are currently being used: (1) the number of categories required for a given training set, and (2) the number of populated categories formed for the same training set. The differences between these two measures are found in the interpretation of sparsely populated clusters or categories. For example, if 100 inputs produces five categories populated by 95, 2, 1, 1, and 1 examples, respectively, then either five separate categories or one single category with five incorrect responses are necessary.

The trade-offs of these measures are identical to those for fielded ATR systems: performance versus hardware requirements. If every vehicle needs to be recognized, then all possible variations, including those categories individually populated, or *outliers* (those variations whose characteristics are rarely viewed), are required to be modeled and retained in the ATR system. It is possible, as shown by the 100 input examples above, that a significant reduction in the hardware requirements (i.e., memory) of the system can be obtained by allowing a certain reduction in recognition capability.

Evaluations Using Laser Radar Imagery

We have investigated the ability of this ATR system to classify laser radar range imagery of various military targets correctly. This system has been tested on a limited amount of imagery obtained with ground-based sensors built by the Opto-Radar Systems group at Lincoln Laboratory. The results of these tests are presented below.

The full capabilities (and deficiencies) of an ATR system, however, must also be determined, and this determination is possible only through exhaustive testing requiring large amounts of sensor data. Many conditions can be tested to determine the capabilities of the ATR system; we used three conditions: (1) CNR, (2) out-of-plane rotation, and (3) number of pixels on target. Unfortunately, the amount of sensor data required to test these three conditions thoroughly by using real sensor data is prohibitive in both time and cost. The use of synthetic imagery to place bounds on system capabilities is the logical alternative.

Synthetic laser radar range target imagery was generated by using Environmental Research Institute of Michigan (ERIM) wire-frame models of a variety of military and non-military vehicles. Background imagery was generated by using a flat ground plane projected with the attack angle of the sensor. Perfect object extraction from the ground plane was assumed for this study. Sensor statistics (e.g., noise) were added by using the laser radar range and intensity models developed by J. Shapiro et al. [18, 3 , 4]. This procedure was used to generate targets for the three test conditions listed above.

Ground-Based Sensor Data and Results

In 1981 at Camp Edwards, Massachusetts, the Opto-Radar Systems group recorded a large database of laser radar imagery of three vehicles—an M-48 tank, an M-113 armored personnel carrier (APC), and an M-110 howitzer. These vehicles were recorded at five orientations with five range backgrounds by using a transportable ground-based laser radar sensor that was the forerunner of the airborne IRAR system. This ground-based sensor allowed us to create a versatile database for testing ATR system performance. Each of the five background scenes consisted of sky, trees,

Table 5. Classification Results for Three-Target Database of 18 Images

	Number of Categories Formed	*False Alarms*	*ART-2 Vigilance*
BCS filled silhouettes	5	1	0.785
BCS maximum edge strengths	5	0	0.694
BCS summed edge strengths	4	0	0.718
Raw extracted target silhouette	3	2	0.790

or hillside, which we created by changing the location of the ground-based sensor relative to the target.

We selected an 18-frame image subset of the Camp Edwards database and processed this image subset through the ATR system. This image subset consisted of three frames of three targets at 750 m and 1000 m in range. The 750-m imagery had a sky background that provided infinite range contrast between the target and background. The 1000-m imagery had a hillside background that had almost no range contrast because of the high depression angle between the sensor and the target; many pixels in this imagery were only one range count different from the target.

The detection algorithm of the ATR system lo-

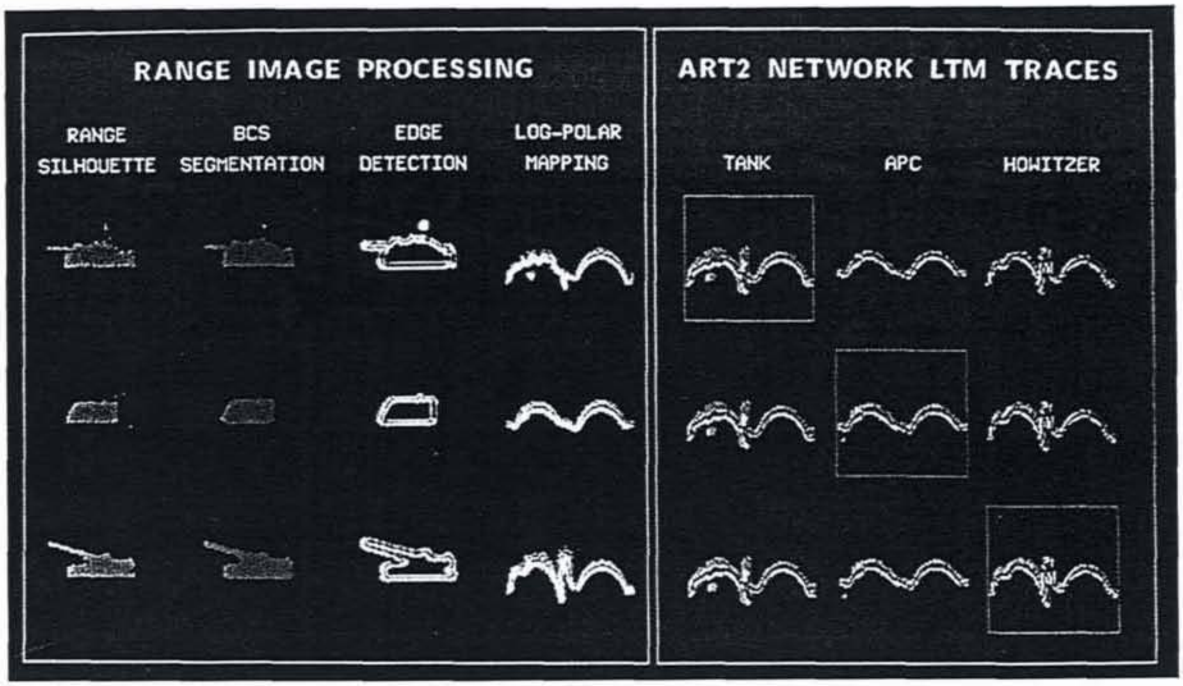


FIGURE 15. Classification of laser radar range imagery into stable recognition categories by using the ART-2 neural network. The range silhouette is shown for three input images—a tank, armored personnel carrier (APC), and howitzer—followed by the resultant image from the segmentation stage. The edge image is then computed, followed by the result of the log-polar mapping. The right side of the figure shows the three LTM patterns with a red box outlining the matched LTM category for the corresponding input. Note that the LTM patterns are not identical to the input log-polar patterns, because they are an aggregate of all the inputs classified with an individual LTM.

cated and extracted 100% of the targets in the test set. This result was not unexpected for the 750-m imagery. There was significant range contrast in the scene, so verticality measurements alone could be used for detection and the size filter was not required to extract the target. In the 1000-m imagery, however, the background was often only one range count different from the target, which required the size filter to extract accurately the region of interest defined as the target. A detection rate of 100% for the 1000-m targets demonstrated the robust behavior of the detection stage of the ATR system.

The ART-2 neural-network classification stage of the ATR system properly classified 95% of the targets into five stable recognition categories, as listed in Table 5. Sixteen targets formed four categories (specifically, six tanks, five APCs, three 750-m howitzers and two 1000-m howitzers), one 1000-m howitzer formed its own category, and one 1000-m APC was erroneously classified as a tank and counted as a false alarm. This performance is acceptable after careful examination of the imagery. The tanks and APCs formed relatively consistent invariant patterns for classification. The detected howitzers, however, were not classified consistently because the detection stage either included part of the ground plane or it removed part of the target body.

Figure 15 shows a sample classification result. The images in the left column of the figure are the detected silhouettes determined by using the range imagery of a tank, APC, and howitzer from the detection stage of the ATR system. These silhouettes are processed by the segmentation stage, the edge strengths are computed, and then the edge-strength images are transformed into the log-polar domain, as shown in the next three columns of the figure. The right side of the figure indicates the three LTM traces created by the ATR system after processing the nine 750-m image frames. The red-box highlight indicates the LTM trace with which that particular input image on the left is matched.

Classification performance was investigated for a set of variations to the baseline ATR system. The baseline system uses the edge images computed from the filled silhouettes produced by the segmentation stage as the input to the log-polar map. The filled silhouettes are produced by using the summed-edge compression method. The variations investigated were

Table 6. The Effect of CNR on the Number of Categories Formed for ATR System*

CNR (dB)	*Percent Anomalies*	*With Image Cleanup*	*Without Image Cleanup*
100	0.0	8	8
35	0.2	8	8
30	0.6	8	17
25	1.9	8	26
19	7.3	8	
16	13.9	8	
13	25.2	8	
10	42.3	8	
7	63.1	26	

* Tests included eight vehicles (jeeps, trucks, armored personnel carriers, and tanks) both with and without image cleanup.

the use of the maximum-edge image, the summed-edge image, and the edge image computed from the target silhouette produced by the detection stage. Each of the variations is related to a reduction of processing by either eliminating part of or the entire segmentation stage.

We describe the results of these variations to the baseline system in terms of the number of categories formed and the number of false alarms (false classifications) produced. The goal is to reduce both the number of categories (i.e., produce better generalization of the data) and the number of false alarms. The results given in Table 5 indicate that both the maximum-edge-strength image and the summed-edge-strength image eliminate the false alarms while the summed-edge-strength image also reduces the number of categories. The target-silhouette image further reduces the number of categories while sacrificing false-alarm performance.

These preliminary results indicate that the classification results are sensitive to the algorithms used in the processing stages prior to the classification stage. Additional results of tests using a larger database are required before we can conclude that summed edge strengths should be used exclusively as the input to the log-polar map.

Effect of CNR on Synthetic Broadside Target Recognition

In the first test we evaluated the effect of CNR on the recognition of broadside targets. We performed two individual experiments to determine the number of categories formed without image cleanup and the number of categories formed with image cleanup. Table 6 shows the results of recognizing eight broadside vehicles (two jeeps, two trucks, two tanks, and two APCs) that are synthetically generated with a sensor of 100-μrad angular resolution imaged at a distance of 750 m.

Without image enhancement, a high ART-2 classifier vigilance value was required to separate the eight vehicles. This high value forced the classifier to form multiple categories for each vehicle at a CNR value of 30 dB. The same result is obtained when image en-

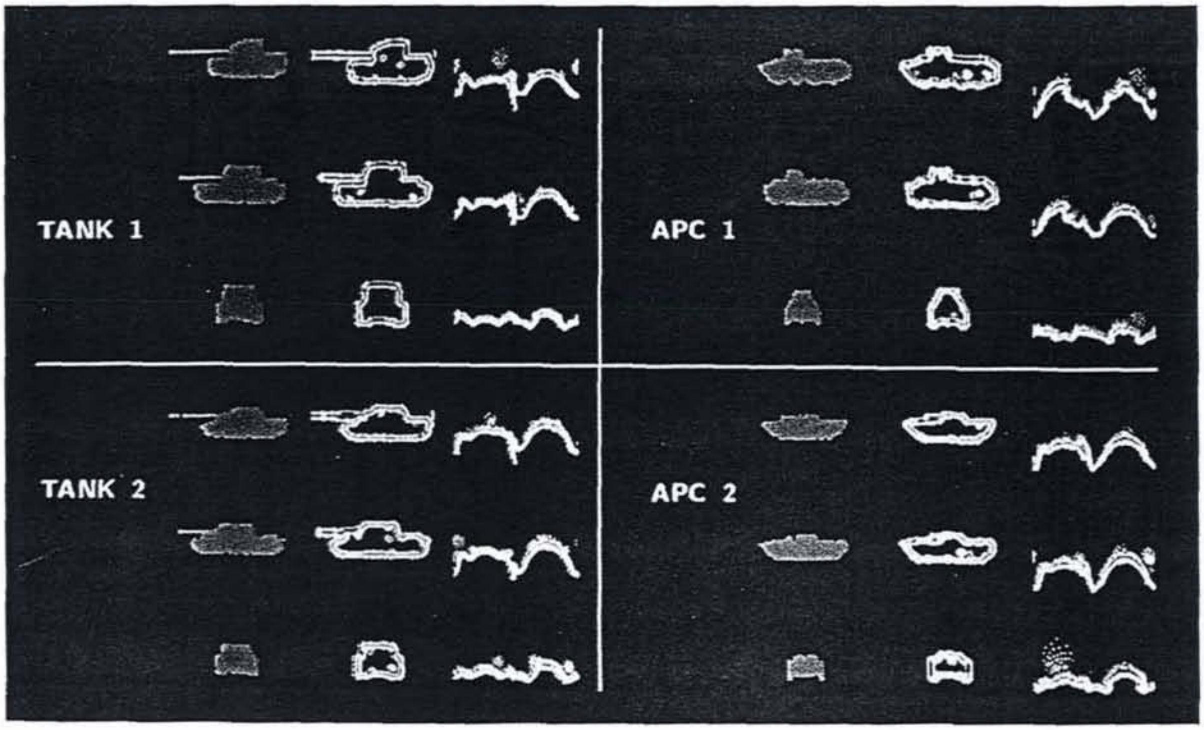


FIGURE 16. Log-polar maps of tanks and APCs rotated out of plane. The log-polar map of the two tanks are similar for the broadside and near-broadside views but different for the head-on view. The same similarities and differences exist between the log-polar maps for the two APCs.

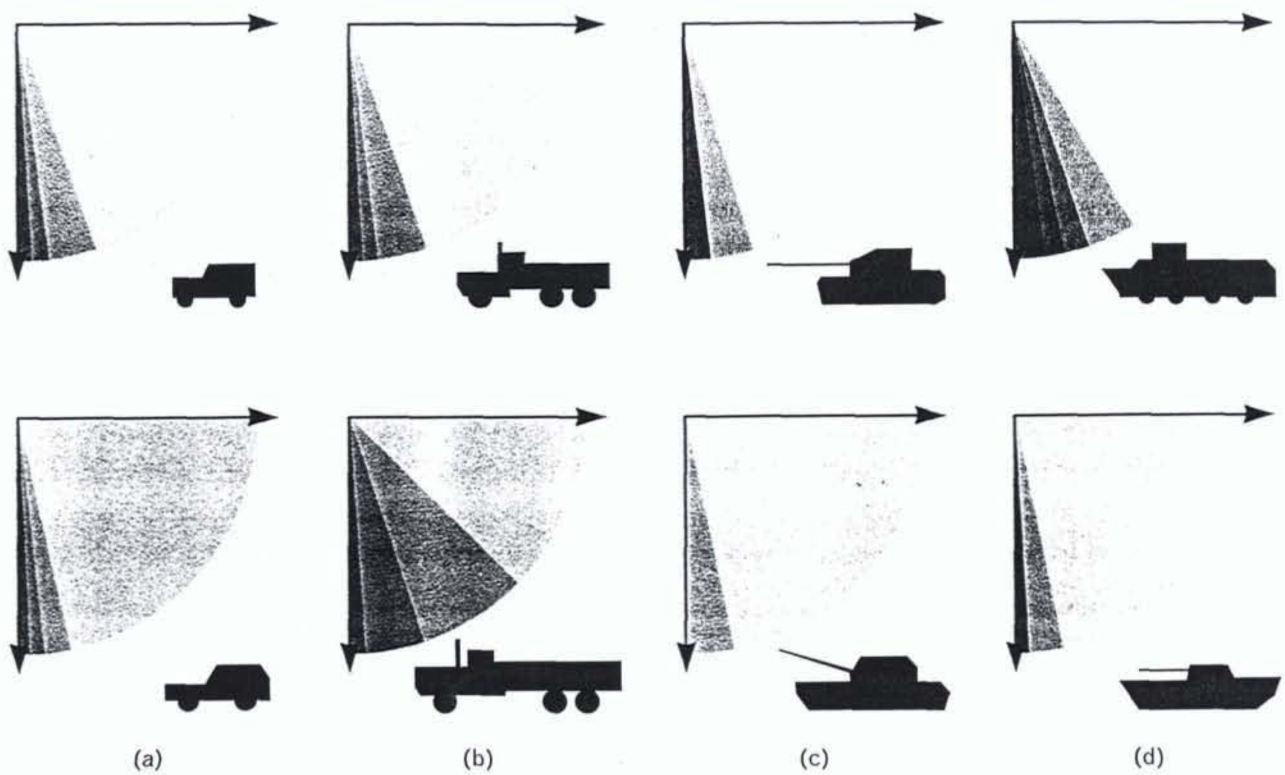


FIGURE 17. Approximate angular extent of each category for recognition of log-polar maps of eight vehicles with out-of-plane rotation; the vehicles are (a) two jeeps, (b) two trucks, (c) two tanks, and (d) two APCs. A total of 31 categories are formed. Each category and its angular extent is depicted by the shaded patterns in the figure. Each vehicle requires only a single category from the broadside view up to 45° of head on or greater. The majority of the categories are in the last 15° from near head on to head on because the log-polar maps change the greatest in that region.

hancement is included, in the form of the Bayesian preprocessor, at a CNR value of 7 dB. A typical operational sensor value of 19 dB at a distance of 1000 m indicates that image cleanup is a necessity. For further details of the experiment and results, see the report by S. Rak [19].

Out-of-Plane Rotation Recognition

A second test was performed to provide insight into the number of independent categories necessary to distinguish eight vehicles rotated out of plane from broadside to head on (a 90° rotation) [20]. When matched filters are used for recognition, we commonly create filters for every 5° of arc. This test was to provide experimental evidence for the number of categories necessary for recognition. Again, we used the same ATR system with the log-polar maps that we used with the ground-based sensor data.

A visual depiction of the information passed to the classifier indicates that input to the log-polar maps from broadside to 50° of head on are similar, whereas the maps near head on change radically. Figure 16 shows the log-polar maps for two tanks and two APCs at broadside, 50°, and head-on orientations. Visually, Figure 16 indicates that more categories are necessary for the near head-on orientations while only a few categories are needed for the near-broadside orientations.

The same test was performed with the eight vehicles rotated from broadside to head on in 1° increments, which created 90 inputs per vehicle. The 720 aggregate inputs were then used to train the classifier, which determined that only 31 categories were necessary to distinguish the eight vehicles at any orientation from broadside to head on. Figure 17 indicates the approximate angular extent of each of the 31 categories. Some vehicles require more categories than others. A general trend seen in this figure is that only one category is necessary for each vehicle to distinguish the vehicles from broadside to approximately 45° of head on. This result agrees with the intuitive understanding we have when viewing the log-polar maps.

Resolution Requirements and the Johnson Criterion

A final test of the ATR system is the comparison between the criteria indicated by J. Johnson [20] and the resolution requirements for recognition and identification. Johnson's work focused on determining the imaging requirements of a sensor to produce a level of discrimination and recognition for human observers. The work consisted of psychovisual experiments on U.S. Army personnel by using image intensifier imagery that is similar in quality to passive-IR imagery. The personnel were shown images of various vehicles at various resolutions and asked to identify the vehicles. The *Johnson criterion* is the number of pixels in a vehicle's minimum dimension (usually height) that is required for a 50% probability of correctly identifying the vehicle.

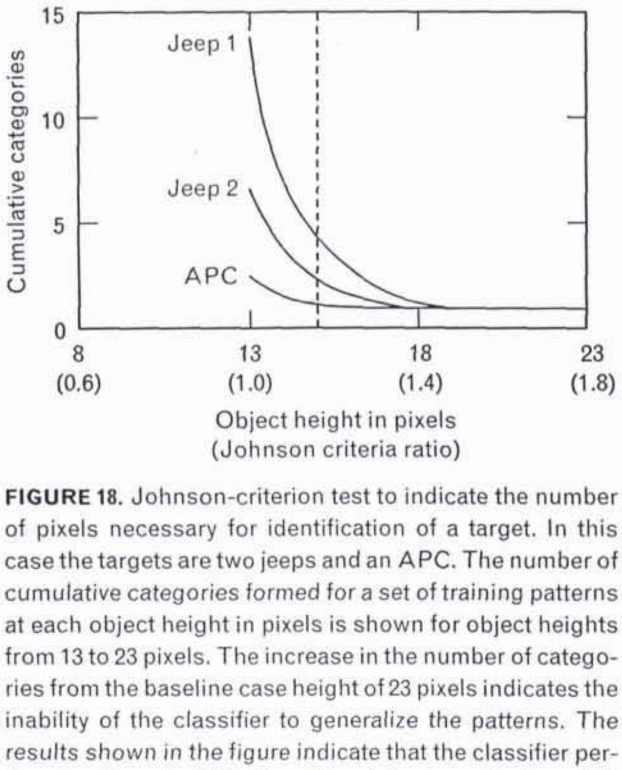


FIGURE 18. Johnson-criterion test to indicate the number of pixels necessary for identification of a target. In this case the targets are two jeeps and an APC. The number of cumulative categories formed for a set of training patterns at each object height in pixels is shown for object heights from 13 to 23 pixels. The increase in the number of categories from the baseline case height of 23 pixels indicates the inability of the classifier to generalize the patterns. The results shown in the figure indicate that the classifier performs well up to 20% above the Johnson criterion of 13 pixels, as indicated by the dashed line in the figure.

We performed an experiment with the spatial extent of each pixel as the variable; this experiment was identical to the one on the effect of CNR described above. Three broadside vehicles were used (two jeeps and an APC) for the training, and the number of pixels in the minimum dimension were varied from 13 to 23. Figure 18 indicates the number of categories formed as a function of the number of pixels in the minimum dimension. For complex vehicle outlines such as the jeeps, the classifier performs well up to 20% greater than the Johnson criterion for identification. For a much simpler vehicle such as the APC, the classifier is more robust and can still identify the vehicle at the Johnson criterion. For more details on the methodology and interpretation of results, see the report by Rak [19].

The ATR Evaluation Facility

Military applications require the use of ATR systems in both semi-autonomous and autonomous modes (in a semi-autonomous mode we believe in the recognition capabilities of the ATR system enough for a user to apply the results, while in autonomous mode we let the system act on the results on its own). The testing and acceptance of ATR systems for these military applications has proven to be difficult. The resources necessary to provide useful test results are usually overburdening. Either we must use large amounts of real sensor imagery, sometimes in multiple sensor modalities, for each given mission scenario, or we must use synthetically generated data. The real sensor imagery requires expensive and time-consuming efforts to gather the data, while the synthetic imagery places an inherent trust in the validity of the sensor and target models used to generate the synthetic data. The recent development, however, of inexpensive computer graphics workstations and data-processing engines has begun to change the emphasis from measurement missions to computer-generated data.

We are currently developing an ATR evaluation facility (AEF) that exploits the recent developments in computer graphics and data processing to provide an effective test environment. This facility merges high-resolution data, an electronic terrain board (ETB) that combines sensor data with synthetic targets and sensor models, and an ATR system that is under evaluation. The high-resolution data are taken in the modalities of interest (laser, passive IR, and visible) and stored in databases. The ETB uses the databases along with the sensor and target models to modify the measured imagery for ATR-system sensitivity analyses. This section describes the facility as well as current research on its development.

Description of the ATR Evaluation Facility

The AEF merges existing sensor data in multiple modalities with synthetic data from sensor and target computer models. Figure 19 shows the conceptual flow of information in the AEF. The airborne IRAR sensor suite, which is described earlier in this article, collects high-resolution imagery in laser intensity,

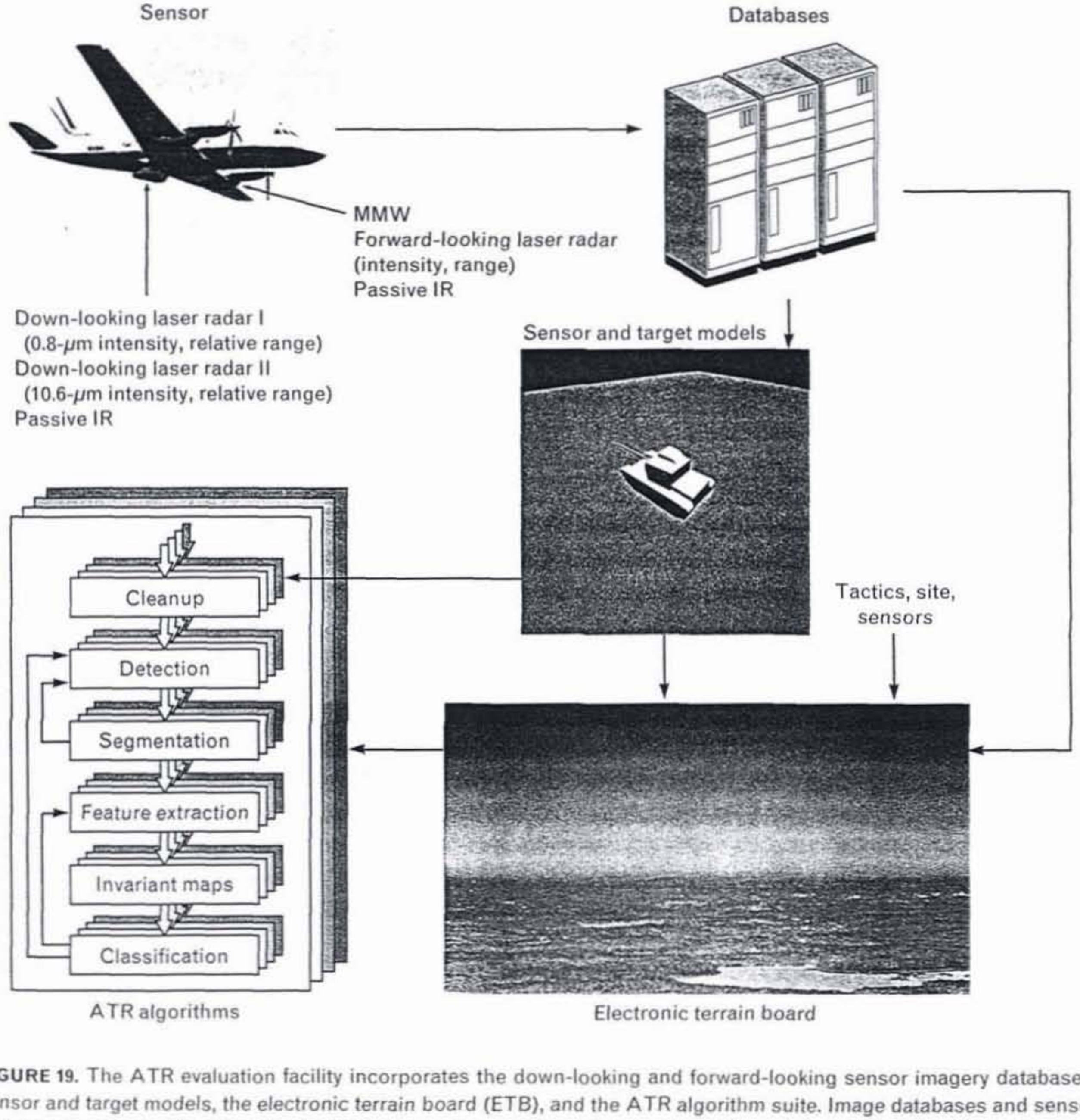


FIGURE 19. The ATR evaluation facility incorporates the down-looking and forward-looking sensor imagery databases, sensor and target models, the electronic terrain board (ETB), and the ATR algorithm suite. Image databases and sensor and target models are fused within the ETB, which allows us to modify target models and vary the measured backgrounds.

range, passive IR, and MMW in a variety of wavebands and view aspects. These data are stored in large databases that are used to refine the synthetic data created from sensor and target models.

The modeling efforts and the databases are merged in the ETB. Most false alarms and missed detections as well as missed classifications of targets are due to the variability of background clutter signals. Model-based systems and trainable recognition systems are developed by using limited target signatures only; unfortunately, these systems do not develop internal models for backgrounds as well. Therefore, we must find a way to merge target signatures, which are predominantly models, with background clutter.

It is difficult, however, to model background signals because of their variable and unpredictable nature. The background models are therefore the weakest link of a completely synthetic sensor image. The combination of measured background imagery with the more well-defined synthetic target models circumvents this problem.

Terrain Database

The down-looking laser radar sensor described earlier provides high-resolution range imagery. This imagery is a 2½-D representation of the actual terrain and precludes the existence of speckle noise indicative of intensity images. The 2½-D imagery contains the range of the first object or part of object that is interrogated for each pixel. Therefore, any part of an object at a further range or area occluded is not represented in the data. The 2½-D notation indicates that a fu!! 3-D image is produced, although the way we view the scene from above appears as if a blanket were covering the objects in the scene. Only the highest point of a pixel that is interrogated is recorded; any part of an object at a longer range or in an occluded area is not represented in the data. For example, a ball in midair viewed from above is represented as a hemisphere on top of a cylinder because no information is available on the space below the ball. Techniques for combining multiple views are being investigated to alleviate this current limitation.

The down-looking sensor simultaneously measures range as well as laser intensity and passive IR. The existence of 2½-D range imagery lends each of the sensor domains to coordinate transformations. As described earlier, the ability to transform the range data and subsequent pixel-registered passive-IR data allows the sensor imagery to be used to train and test ATR systems with many viewing aspects. The specific method used for the coordinate transformation can have a dramatic effect both on the requirements for computation and, more importantly, on the quality of the resultant image.

Traditionally, Euler angles have been used to represent coordinate transformations, and these coordinate transformations can be expressed as 3-by-3 rotation matrices. Because the computer graphics community commonly uses rotation matrices, most of the specialized hardware developed to perform coordinate transformations employs this method. This choice has been motivated primarily by the fact that translation and scaling as well as rotation can be represented by one matrix. The same transformations, however, that can be performed by a matrix can be performed with fewer operations by using quaternions [21]

An important consideration in the choice between matrices and quaternions for coordinate transformations occurs when we interpolate between two orientations. Rotation matrices are not well defined for interpolations, because rotations are carried out by three successive rotations about three fixed axes. Because these successive rotations are not commutative, changing the order of the rotations produces different results, which introduces a significant problem known as *gimbal lock*. This problem occurs when the interaction of two rotations aligns two of the three rotation axes and causes a loss in one degree of rotational freedom. Quaternions are free of this problem because the cross-product interaction between successive rotations is preserved [21]. Because of this rotational stability, the aerospace industry for many years has preferred quaternions over matrices defined by Euler angles for spacecraft applications.

Because most computer graphics workstations have hardware that is specifically designed to implement matrix transformations, we must continue to maintain all viewing parameters in matrix form. The AEF system is designed to perform all interpolations by using quaternions, which are then converted to matrix form for rendering.

To illustrate the use of quaternions for interpolating rotations we must first define what a quaternion is and how it is used to perform a rotation. A quaternion consists of two components—a scalar part and a vector part. Consider a quaternion $q = [s, \mathbf{v}]$, where s is a scalar, and $\mathbf{v}$ is a vector of three elements. In quaternion algebra, addition is defined as

$$q_1 + q_2 = \left[(s_1 + s_2), (\mathbf{v}_1 + \mathbf{v}_2)\right],$$

and multiplication is defined as

$$q_1 q_2 = \left[(s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2), (s_1 \mathbf{v}_1 + s_2 \mathbf{v}_2 + \mathbf{v}_1 \times \mathbf{v}_2)\right],$$

where $\mathbf{v}_1 \cdot \mathbf{v}_2$ is the vector dot product and $\mathbf{v}_1 \times \mathbf{v}_2$ is the vector cross product.

Before we can define rotations using quaternions we need to define the inverse operation

$$q^{-1} = \frac{1}{\|q\|^2}[s, -\mathbf{v}],$$

where

$$\|q\|^2 = s^2 + \mathbf{v} \cdot \mathbf{v}.$$

To rotate a point $\mathbf{p}$ we embed it into a quaternion as $[0, \mathbf{p}]$. Rotation is then defined as

$$v' = Rot(v) = qvq^{-1},$$

where q and q^{-1} are unit quaternions.

One consideration associated with the use of quaternions for coordinate transformations is that rotations are performed on the unit four-dimensional hypersphere, so that, as a result, simple linear interpolation between two orientations gives unequal rotations through the range of orientation values. The unequal rotations occur because the great arc of a unit hypersphere is the spherical equivalent of a line, and the linear interpolation steps fall on unequal portions of the line. These unequal rotations must be compensated for to give a smooth set of intermediate transformations. All of the interpolations between specified positions are performed by using unit quaternions and spherical linear interpolations, and then compensating for unequal rotations [21].

Figure 20 shows an example of the transformation of down-looking range data into various viewing perspectives. Starting with the range data shown in Figure 4, the range data are transformed and displayed as sand-colored video data and synthetically generated laser radar range data in a viewing sequence typical of a target interrogation. In effect, this series of transformations is like an observing eye on a flying carpet; it begins at a long standoff distance at a high altitude, it detects a possible target, it dives to a lower altitude, and it flies along the road to the target. This series of transformations demonstrates how down-looking imagery can be used to train and test an ATR system with many viewing aspects.

Synthetic Laser Radar Imagery

An important element in the ETB is the combination of synthetic imagery and modified sensor imagery. Synthetic imagery is derived from target and background models applied with the appropriate sensor statistics. In some cases, actual sensor imagery can be modified to degrade the quality of the imagery for test purposes. Both of these cases provide the additional flexibility necessary for ATR evaluation. This section describes the methodology used in creating or modifying laser radar imagery.

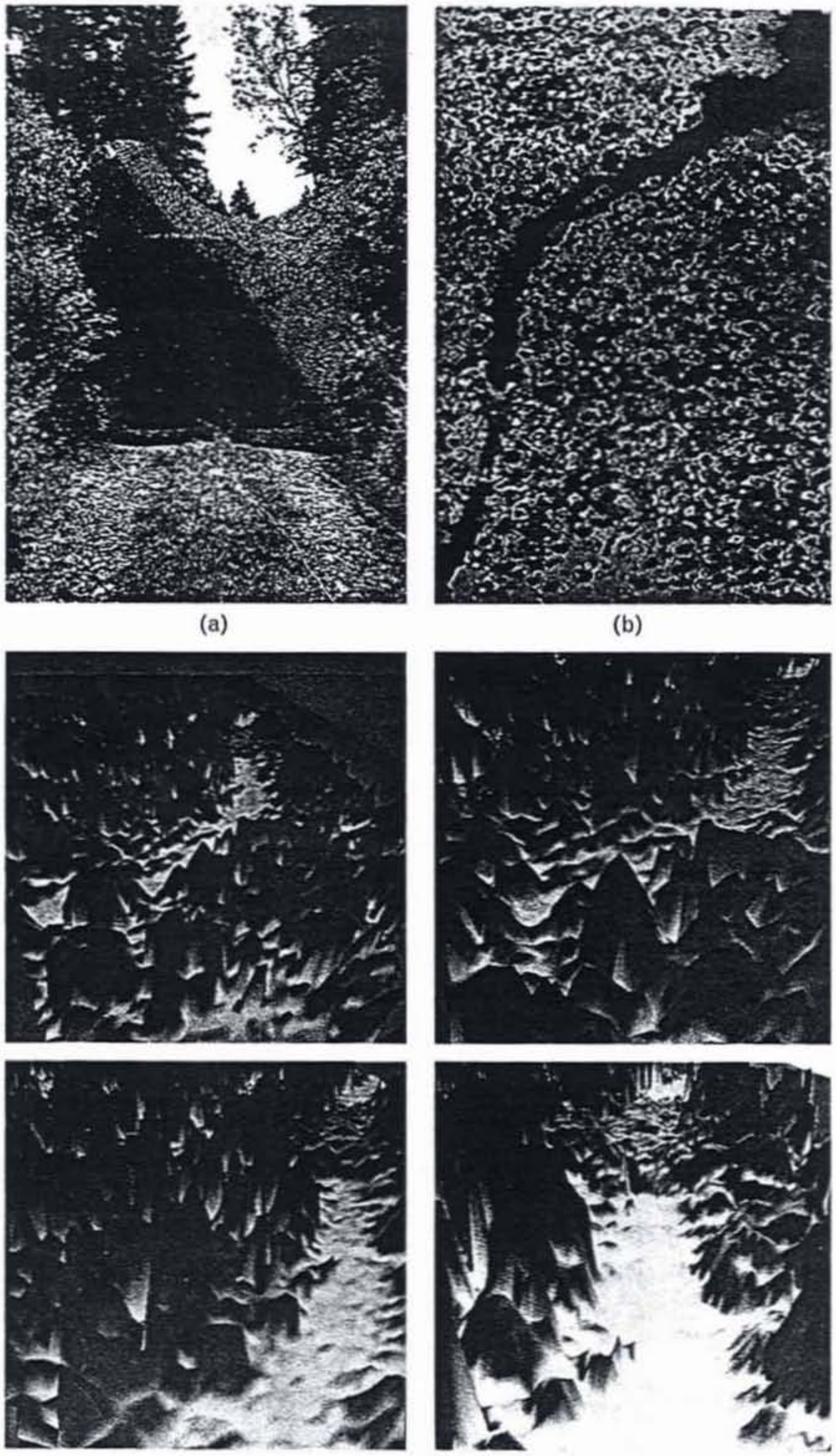
The statistics describing a monostatic pulsed ranging laser radar employing heterodyne detection are described by Hannon and Shapiro [3], and were used to develop a model for laser radar range data. This laser radar model requires that we select the range and CNR value for every pixel as well as the number of range bins Q available to the signal processor. The probabilities of an anomaly (equally distributed across $Q-1$ range bins) and for the correct range value are given by

$$P_{I_c} = 1 - e^{-\frac{I^2}{CNR+1}}$$

and

$$P_{I_w} = (1 - e^{-I^2})^{Q-1},$$

where P_{I_c} and P_{I_w} are the detected intensity probabilities from the correct range bins and the wrong



(a) (b) (c)

FIGURE 20. Down-looking laser radar data is transformed into a three-dimensional terrain-map view. (a) Photographic ground truth of a camouflaged truck, (b) down-looking laser radar image, and (c) a sequence of four views that were formed by using 3-D transformed laser radar imagery to "fly over" the target.

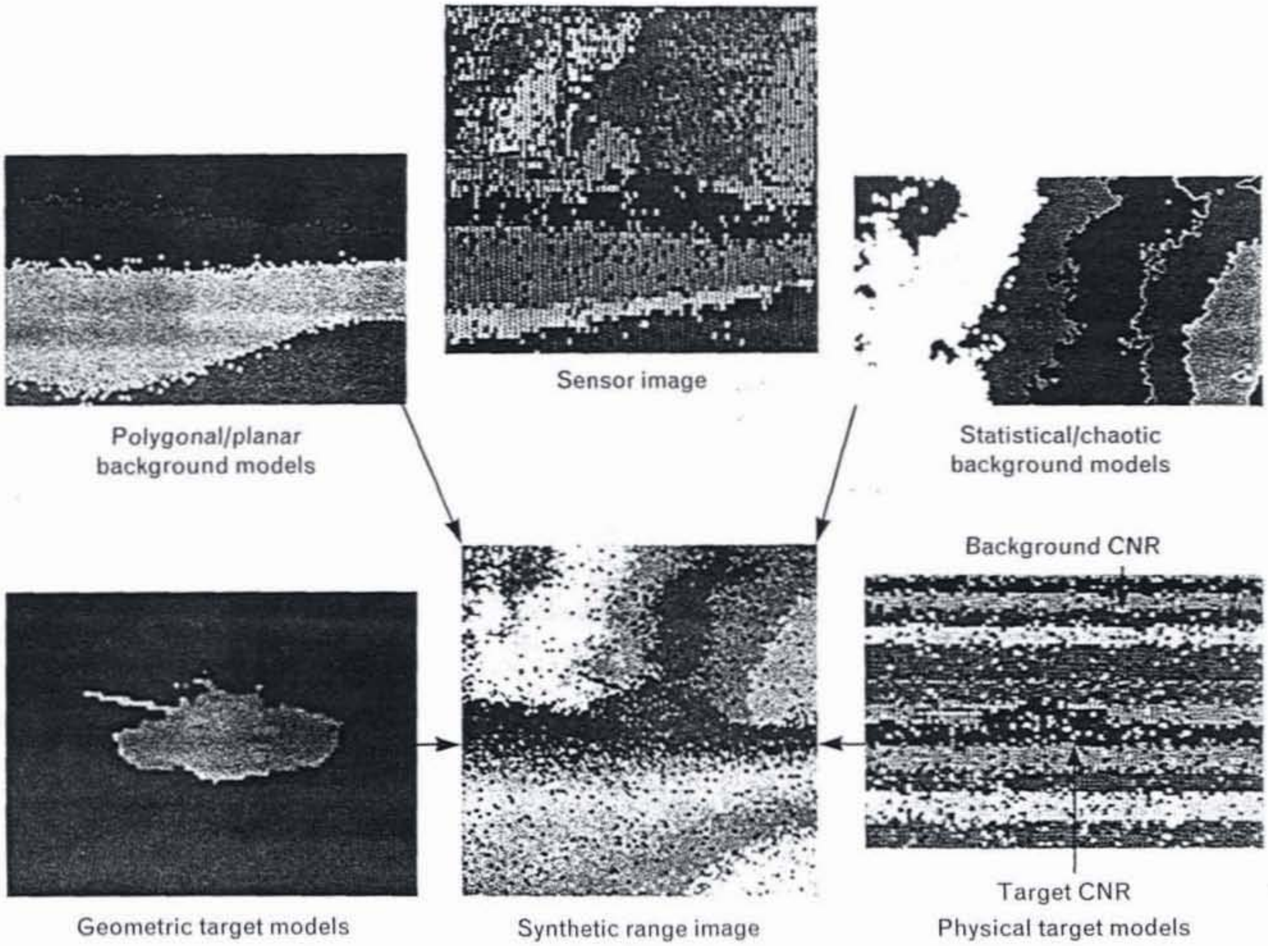


FIGURE 21. Scene decomposition and synthesis of a laser radar range image. The original sensor image (upper center) is decomposed into polygonal background (upper left); statistical, chaotic, or fractal background (upper right); facet target models (lower left); and physical target-background parameters (lower right). A new laser radar image is then synthesized (lower center).

range bins, respectively. These two probabilities are used in conjunction with a random-number generator to provide two random draws. The maximum value of the two random draws is selected as the intensity *I*.

The application of the described statistics can be demonstrated through an example. Figure 21 depicts the decomposition of a sensor laser radar range image into four primary parts. The background and target models constitute three parts: polygonal background models for relatively uniform terrain; statistical, chaotic, or fractal models for fragmented terrains such as foliage; and the geometric target models generated from wireframe or facet libraries (ERIM) or solid geometry libraries (U.S. Army Ballistic Research Laboratory). The physical target models, which are described by the CNR for each pixel, are used to compute the intensity and range value by using the Hannon laser radar model [3]. In the example shown, a variety of fractal dimensions were attempted and a best visual fit was selected for the foliage. A uniform CNR value of 17 dB (which is a typical value for imaging generic terrain by the IRAR sensor at 700 m) was used to generate the synthetic range image. The generated image is visually similar to the original sensor image.

Summary

A flyable, multisensor system has the ability to measure a combination of range, Doppler, laser intensity, and thermal signatures in both the forward-looking and down-looking aspects. Statistical advantages for incorporating multidimensional information exist for target-detection applications using theoretical analyses and heuristic algorithms. The use of multiple sensor modalities also provides some hope to address the vexing issues of ATR.

A modular, hybrid ATR system has been described that fuses statistical, model-based, and neural network processing structures. The system has been tested on laser radar range imagery as well as synthetic range imagery incorporating pulsed laser radar statistics. Results created by using the synthetic imagery indicate that target identification can occur in imagery with over 50% of the pixels corrupted by noise. Tests with out-of-plane rotated vehicles indicate that a finite number of nonuniform angularly spaced projections can be learned by the system to provide target identification. The current system can also provide identification with spatial resolution as low as 20% above the Johnson criteria.

To continue to test and evaluate complicated ATR systems, an ATR evaluation facility is being constructed to provide real, synthetic, and hybrid sensor image input to a selected ATR. This facility uses the available high-resolution down-looking laser radar range imagery and high-fidelity target models to generate the various operational scenarios.

Acknowledgments

The work described in this article is the combined effort of a large group of scientists and technical support personnel. The author acknowledges the efforts of R. Hull, T. Quist, and S. Prutzer for the development of the multisensor measurement system, and the efforts of D. Biron, E. Van Allen, S. Rak, M. Menon, and J. Baum for the development of the ATR system and the subsequent algorithms. The author also thanks A. Gschwendtner, P. DiCaprio, and H. Thomas for their contributions to the ATR evaluation facility.

This work was sponsored by the Advanced Research Projects Agency, the Balanced Technology Initiative Program, and the U.S. Air Force under an Electronic Systems Division contract.

REFERENCES

1. D. Biron, R. Hull, and T. Quist, private communication.
2. M.B. Mark, *Multipixel, Multidimensional Laser Radar System Performance*, Ph.D. thesis, MIT Dept. of Electrical Engineering and Computer Science (1986).
3. S.M. Hannon and J.H. Shapiro, "Laser Radar Target Detection with a Multipixel Joint Range-Intensity Processor," *SPIE* **999**, 162 (1989).
4. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Recognition Mach. Intell.* **PAMI-6**, 721 (1984).
5. S. Grossberg and E. Mingolla, "Neural Dynamics of Perception Grouping: Textures, Boundaries, and Emergent Segmentations," *Perception and Psychophysics* **38**, 141 (1985).
6. D. Hubel and T. Wiesel, "Functional Analysis of Macaque Monkey Visual Cortex," *Proc. Royal Soc. of London (B)* **198**, 1 (1977).
7. P.J. Kolodzy, M.M. Menon, and E.V. Allen, private communication.
8. E.J. Van Allen and P.J. Kolodzy, "Application of a Boundary Contour Neural Network to Illusions and Infrared Sensor Imagery," *IEEE First Intl. Conf. on Neural Networks* **4**, *San Diego, 21–24 June 1987*, p. 193.
9. E.J. Van Allen, private communication.
10. J.G. Verly, R.L. Delanoy, and D.E. Dudgeon, "Machine Intelligence Technology for Automatic Target Recognition," *Linc. Lab. J.* **2**, 277 (1989).
11. D.E. Dudgeon, P.J. Kolodzy, C. Mehanien, M.M. Menon, S.J. Rak, E.J. Van Allen, J.G. Verly, and R.L. Delanoy, private communication.
12. E. Schwartz, "Computational Anatomy and Functional Architecture of Striate Cortex: A Spatial Mapping Approach to Perceptual Coding," *Vision Research* **20**, 645 (1980).
13. S.J. Rak, D.G. Biron, P.J. Kolodzy, M.M. Menon, and E.J. Van Allen, private communication.
14. G. Carpenter and S. Grossberg, "Neural Dynamics of Category Learning and Recognition: Attention, Memory Consolidation, and Amnesia," in *Brain Structure, Learning, and Memory*, eds. J. Davis, R. Newburgh, and E. Wegman (AAA Symposium Series, 1985), pp. 1–49.
15. R. Lippmann, "An Introduction to Computing with Neural Networks," *IEEE ASSP Mag.* **3**, 4 (Apr. 1987).
16. M.M. Menon and P.J. Kolodzy, "A Comparative Study of Neural Network Classifiers," presented at the First Intl. Neural Networks Symp. (INNS), Sept. 1988.
17. G. Carpenter and S. Grossberg, "Self-Organization of Pattern Recognition Codes for Analog Input Patterns," *Appl. Opt.* **26**, 4919 (Dec. 1987).
18. J.H. Shapiro, R.W. Reinhold, and D. Park, "Performance Analyses for Peak-Detecting Laser Radars," *SPIE* **663**, 38 (1986).
19. S. Rak and P. Kolodzy, "Performance of a Neural Network Based 3-D ATR System," *Project Report NN-6*, Lincoln Laboratory (May 1991).
20. J. Johnson, "Analysis of Image Forming Systems," in *Proc. Image Intensifier Symp.* (U.S. Army ERDL, Fort Belvoir, VA, Oct. 1958).
21. K. Shoemake, "Quaternion Calculus for Animation," in "Math for SIGGRAPH," ACM SIGGRAPH '89 Course Notes, pp. 187–205 (Boston, MA), July 31–August 4, 1989.

PAUL J. KOLODZY
is assistant leader of the Opto-Radar Systems group. He received a B.S. degree from Purdue University, and M.S. and Ph.D. degrees from Case Western Reserve University, all in chemical engineering. His current areas of research include neuromorphic systems, automatic target recognition systems, and advanced distributed simulation technology. In 1983, he was a visiting scientist working in the area of laser remote measurement systems at Risø National Laboratory in Roskilde, Denmark. He was cochairman of the Simulation/Emulation Tools & Techniques Panel of the ARPA Neural Network National Study, and he is a member of the Sensor Fusion program committee. Paul has been at Lincoln Laboratory since 1986.