Aided and Automatic Target Recognition Based Upon Sensory Inputs From Image Forming Systems

James A. Ratches, C.P. Walters, Rudolf G. Buser, and B. D. Guenther

Abstract—This paper systematically reviews 10 years of research that several Army Laboratories conducted in object recognition algorithms, processors, and evaluation techniques. In the military, object recognition is applied to the discrimination of military targets, ranging from human-aided to autonomous operations, and is called Automatic Target Recognition (ATR). The research described here has been concentrated in human-aided target recognition applications, but some attention has been paid to automatic processes. Definitions and performance metrics that have been developed are described along with performance data showing the present state-of-the-art. The effects of signal-to-noise and clutter parameters are indicated in the data. Multisensor fusion and model-based algorithms are discussed as the latest techniques under consideration by the military research community. The results demonstrate that useful performance can be achieved, and tools are evolving to understand and improve the performance under real-world conditions. The referenced research strongly indicates the need for the development of image science, as described in the paper, to support the theoretical underpinnings of ATR.

Index Terms—Automatic Target Recognition, ATR, imaging sensors, image processing, aided target acquisition, multisensors, sensor fusion, ATR algorithms, performance metrics, databases.

1 Introduction

The oldest imaging forming system used by humans has been our visual system. The high-resolution eye sensor, when coupled to the brain via the optic nerve, is the ultimate objective in sensor, signal processor design. In spite of the elegant design of the human eye–brain system, it has serious shortcomings. For example, the eye is limited in the wavelengths to which it is sensitive, it does not have capability at extended range, it does not see well at night, it does not do well in transmission-attenuated atmospheres, and it can be tricked rather easily. In response to these limitations, humans have developed devices to view our environment far beyond our sensing system that allow us to accomplish increasingly complex tasks. However, the additional data have begun to overwhelm our ability to quickly process all the information and make decisions based upon these data.

In medical imaging, a wide range of modalities has improved diagnostic capabilities. Also, our ability to acquire imagery has increased: An echo planar image can be generated in 100 msec, and four computerized tomography scan images can be generated in a second. Medical cost containment prevents an increase in the number of physicians, but the increasing quantity of images and image types increases the need for more radiologists. The human need for processing help is also found in automatic fingerprint and face recognition, manufacturing controls and inventory screening, and robotics.

In the military, sensors have been developed for viewing the battlefield at night and during obscuring weather, allowing 24-hour, "all-weather" performance. Image intensifiers, thermal imaging, high-resolution television, and lasers are prime examples of the technologies employed by the military. The data from these sensors pour in along with demands on the soldier to make rapid decisions. The soldier, like the radiologist, is overloaded with information from a vast array of sensors while responding to demands of life-threatening dimensions. The soldier needs to efficiently use all sensor information and requires image processing to aid the decision-making process. These requirements are the origin of the concept of Automatic (or aided) Target Recognition (ATR) in the military and Guided (or computer-aided) Diagnostics in the medical community. ATR is a generic term used to describe various automated and semiautomated functions carried out on imaging sensor data to perform operations ranging from the simple—cuing a human observer to a potential target, to the complex—autonomous object acquisition and identification.

ATR is the machine function of detecting, classifying, recognizing, and/or identifying an object without the need of human intervention. In the military, the most sophisticated example of ATR is the fire-and-forget, lock-on-after-launch missile. Here, an ATR would recognize the candidate targets in the scene after it has been launched, select the target of choice, track the target during the flight, make final aim point selection, and conduct terminal guidance to the target. In autonomous applications, an image may

- *J.A. Ratches, C.P. Walters, and R.G. Buser are with U.S. Army Communications—Electronics Command, Research, Development and Engineering Center, Night Vision and Electronic Sensors Directorate, Ft. Belvoir, VA 22060-5806. E-mail: jratches@nvl.army.mil.*
- *B.D. Guenther is with U.S. Army Research Office, Research Triangle Park, NC 27709-2211.*

Manuscript received 3 Oct. 1996. Recommended for acceptance by R. Kasturi. For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 105307.

never be displayed, as in the case of a fire-and-forget missile. Acceptable autonomous operation is still an unattainable goal of ATR in the military and is not yet an accepted mode of operation in the medical community.

The less ambitious, current objective is to produce Aided Target Recognition, i.e., the subset of ATR in which a human interacts with the system and makes some of the decisions. The sensor-processor system recognizes the object of interest and makes some annotations as to its classification. The annotation may be derived using information from a nonimaging sensor, such as a radar, from some measure of functionality, such as an acoustic Doppler measurement of blood flow, Global Positioning System (GPS), or from processing of the image data. The human confirms that the annotated object has been correctly identified. The benefit of such an arrangement is that the human can accomplish other functions or handle an increased workload with ATR.

In this paper, we will present a systematic review of the development and evaluation of Aided Target Recognition carried out by several Army Laboratories over the preceding 10 years. Earlier surveys can be found in [1] and [2]. We begin by describing the current performance metrics utilized for evaluation and the image databases utilized for the creation of ATR algorithms. Over 75 ATR implementations have been evaluated by the Army since the early 1980s. The large number of implementations, and proprietary restrictions placed on some by their developers, forces us to limit our discussion to general classes of algorithms.

A great deal can be learned from the results described here without detailed descriptions of the algorithms. Descriptions of the general approaches taken in the development of the specific algorithms can be found in the texts of Marr [3], Pratt [4], and Zhou and Chellappa [5]. A recent bibliography has been prepared by Rosenfeld [6]. As will be seen in the historical evolution, early measurements discovered that target boundary estimation techniques were not robust because they exhibited sensitivity to noise, aspect, target variation, and partial obscuration. Current model-based, target-recognition techniques, based on relationships among edges and vertices, have been found to be more robust.

It seems obvious that with the increase of available imaging modalities, a combination of image data from a variety of sensors should improve the performance of an ATR system. However, it is not necessarily obvious how one should perform Multisensor Data Fusion. In the medical community, success has already been obtained through the fusion of functional and anatomical information. In the military, as will be described here, fusion of range and spatial information has led to increased performance.

The paper concludes with a description of future directions of Army research based upon the shortcomings of current ATR implementations discovered in the past decade by Army Laboratories.

2 PERFORMANCE METRICS

The primary figures of merit for ATR systems are based on their measurable ability to detect, classify, recognize, and identify targets from background clutter and system noise. Nearly 40 years ago, John Johnson [7] determined that a person

- could detect an object given a resolution of one line pair across the minimum dimension of the object,
- could recognize it given four line pairs, and
- could identify it given eight line pairs.

This was found to be true whether the person used the unaided eye or other sensors, such as night-vision goggles or thermal imaging systems. The definitions for these metrics (developed for the human operator) as originally proposed in Johnson [7] and Ratches [8] are in common usage in the military community:

1) *Probability of detection* is the probability of correctly discriminating an object in the image from background and system noise.
2) *Probability of classification* is the probability of correctly determining the class of a detected target. In the case of Army tactical target acquisition, this means telling if the target is tracked or wheeled.
3) *Probability of recognition* is the probability of correctly determining the class membership of the target. Again, for Army tactical targets, is the tracked vehicle a tank, an armored personnel carrier, or a self propelled gun?
4) *Probability of identification* is the probability of correctly determining the exact identity of the target, e.g., for automobiles, is it a Ford, Chevrolet, or Plymouth?
5) *False alarm probability* is the probability of an error in detection. The units of false-alarm rate are false alarms per square degree in object space. (A square degree is approximately one square meter at a range of 60 meters.)

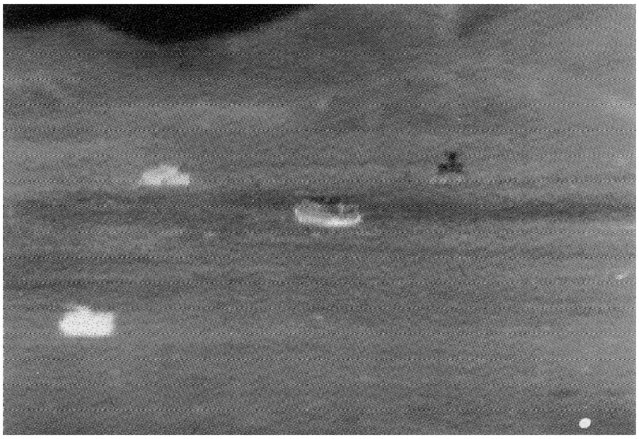
Using these levels of performance, several metrics have been used to evaluate the algorithms developed for ATR.

1) *Signal-to-noise*: The objects in a scene under consideration are corrupted by system noise and coherent noise (speckle). Included in this metric is contrast, which is the difference between the intensity from the target, I_t, and the background, I_b, normalized by the average background intensity:
$$\frac{\left(I_t - I_b\right)}{I_b}. \tag{1}$$
2) *Receiver-Operator-Curve* (ROC): The performance of a system degrades as the signal-to-noise ratio decreases. It is useful to plot detection probability against false-alarm rate as a function of signal-to-noise ratio to produce an ROC similar to plots used in radar design.
3) *Confusion matrix*: This is a 2D array that indicates the identity assigned to an object by one of several ATR systems under comparison. For example, how often was a Ford confused with a Chevrolet or a Plymouth. In the case of detection, the confusion matrix reduces to the classical false-alarm rate, i.e., the rate at which an ATR declares a noise spike or background clutter object as a target.
4) *Consistency*: This is a measure (see Walters [9]) of how often a given ATR algorithm implementation gives

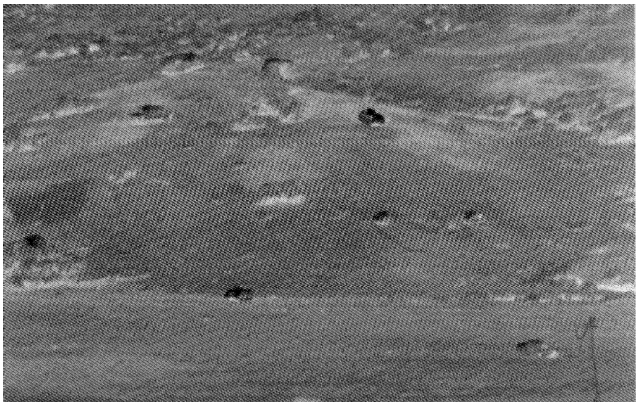
the same declaration for successive image frames having the same scene content. The difference between frames is noise from the atmosphere, the sensor electronics, or display properties.

One would hope that the above definition of terms could be expanded to produce some predictive parameters, but obvious additions to the metrics have not led to good predictive capabilities. Two examples will illustrate the attempts at expansion of the metrics. Futile attempts have been made to find relatively smooth curves similar to the ROC for the other levels of performance, such as recognition. A second example is the inability to successfully quantify clutter. *Clutter* is a term borrowed from radar that has been expanded from the radar definition to include all signals in a scene of no interest to the observer; for example, trees are of no interest to an observer looking for automobiles. Currently, clutter is classified by trained observers as low, medium, or high, but we have been unable to quantify a scene's clutter content. Fig. 1 shows three typical scenes that are examples of the three classes of clutter.

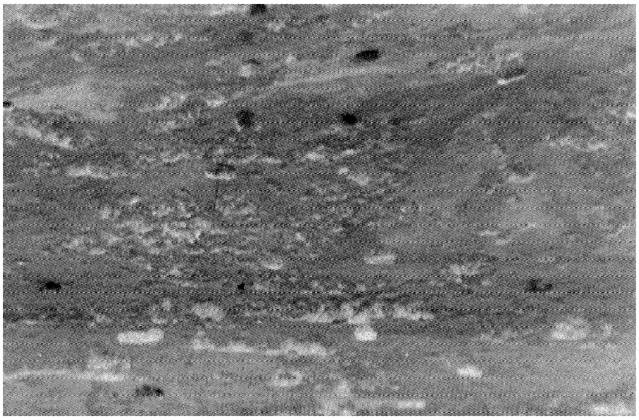
Attempts to create additional image metrics based on information theoretic approaches by Clark et al. [10], [11] have not met with success. The failure in our ability to expand the available metrics is probably due to the enormous variability of the input scene [12]. Backgrounds and targets can undergo wide variations in a sensor image due to weather, time of day, season, geographic location, target history, occlusion, and tactical condition. Increased scene variability can also result from camouflage and competing background clutter objects. Recent attempts at addressing the thermodynamic variation in targets in a rigorous manner have been suggested by Lanterman et al. [13] and Cooper et al. [14].



(a)



(b)



(c)

Fig. 1. Simulated infrared imagery representing clutter conditions. (a) Low. (b) Medium. (c) High.

3 DATABASES

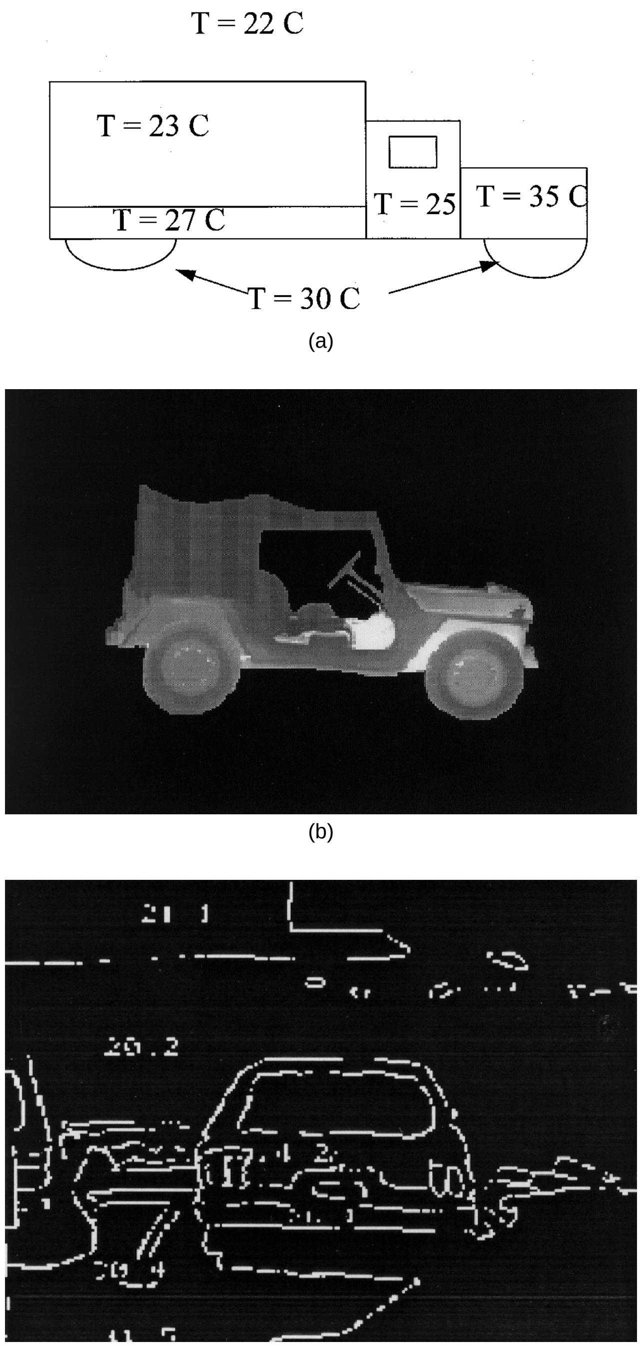
The lack of a theory that can be used to design an ATR algorithm and from which performance can be predicted before implementation means that experiments must be run on the algorithm implementation using input data that span the expected scene variability. These databases are also used by the algorithm developer to provide the understanding of the physics of the scene and potential targets in order to suggest, implement, and improve new algorithms. In addition to a database for designing and training ATR algorithms, a sequestered set of data is needed to test the performance of the algorithms and to measure their performance using the metrics defined above.

There is a hierarchy of sensor image databases that has evolved over time in the defense community that meets various levels of requirements for sensor and ATR development. The complexity of modern signature databases is due to the increased sophistication of modern sensors that can require, for example, pixel-registered signatures in a variety of simultaneous spectral regions. Table 1 shows the hierarchy of sophistication that has evolved historically in the military community based on the signature description, the environment under which the signatures are acquired, and the potential application. The complexity of the data library increases as one moves down the table. Cost has forced us to turn to synthetic scene generation and sensor simulation to generate data for the lower table entries.

Figs. 2–6 show representative examples of this hierarchy of target and background information, as well as the data types available to the Department of Defense laboratories. The data are not of the quality normally seen in image processing publications but are the quality normally seen in operational scenarios. Early data were in the form of simple thermal maps of a target taken manually with a Barnes PRT-5. Each measurement is an individual reading, with a radiometer aimed at a particular location on the target. Fig. 2a

TABLE 1
SIGNATURE DATA

Signature Description	Environment	Military Applications
Vis/NIR Contrast, DT, Radar Cross Section, Acoustical Intensity	Vehicles in Day and Night	• Sensor Range Prediction • Signature Suppression • Scenario Modeling
Low Resolution, Calibrated Images	Vehicles at Specific Geographic Locations (Time of Day, All Seasons)	• ID Range Prediction • Aim Point Selection
Target and Background Images	Vehicles at Variety of Aspects Selected Geographic Locations	• Dynamic Search Modeling • ATR Development • Camouflage
High Resolution, Multispectral Images (e.g., FLIR and MMW/SAR)	Vehicles at Many Aspects Wide Range of Locations and Climates	• Sensor Fusion • Combat ID • Horizontal System Integration
High Resolution, Multispectral Video	All Vehicles All Locations and Weather Many Aspects Related to Digital Map	• Sensor and ATR CAD • Virtual Reality
Real Time Imagery High-Resolution Multispectral with Tactile and Acoustic Data	All	• Virtual Reality Development



(a)

(b)

(c)

Fig. 2. (a) Temperature map collected by hand with a nonimaging thermal radiometer. (b) Thermogram taken with an imaging radiometer. (c) A modern thermal map generated by a computer and an imaging radiometer.

is a schematic of some of those data. Fig. 2b is an example of data from a modern imaging radiometer, which gives a picture of a target with image brightness corresponding to the temperature of the object. A computer-generated thermal map of data collected with an imaging radiometer is shown in Fig. 2c.

Fig. 3 contains simultaneous images of a target in the passive infrared spectral region (Fig. 3a) and a laser radar (LADAR)[1] image (Fig. 3b).

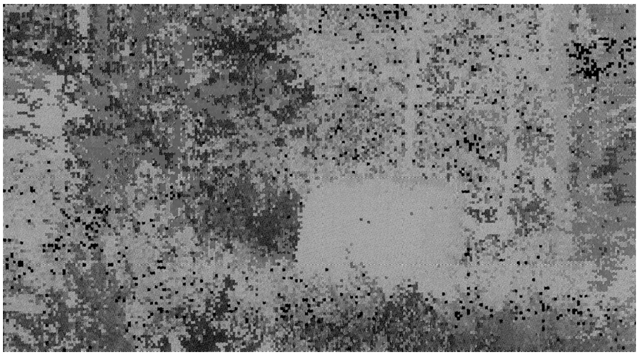
Fig. 4 shows an image produced by a Forward Looking Infrared (FLIR) sensor in a wide field (Fig. 4a) and electronically zoomed portion of the image (Fig. 4b). Shown, in Fig. 5, is millimeter wave (MMW) radar scans (Fig. 5a) and high-range resolution range data (Fig. 5b) of the same region shown in Fig. 4. The images form a multispectral set that meet current system requirements of large field-of-regard imagery for detection and high-resolution images for target recognition and identification.[2] Finally, Fig. 6 shows a typical synthetic aperture radar image of an airport.

4 ATR ALGORITHMS

It is convenient, although not always accurate, to think of an ATR implementation as shown in Fig. 7. The scene is imaged by some sensor(s) and converted into a signal to be processed. The processor may be analog or digital, based on optical or electronic designs. The typical implementation of the processor is a digital electronic processor. The processor performs operations that permit detection of an object of interest and segmentation of that object from the rest of the image. The processor may also receive information from other sources, such as range from a laser rangefinder, position coordinates from GPS, and weather data. The other

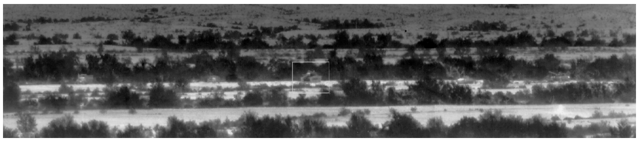
1. The kind of laser radar shown here is an incoherent laser range profiler. Range to each pixel in the scene is measured and usually color coded for presentation to an observer.
2. A database standard for the Department of Defense has been created by the DOD ATR Working Group.

(a)

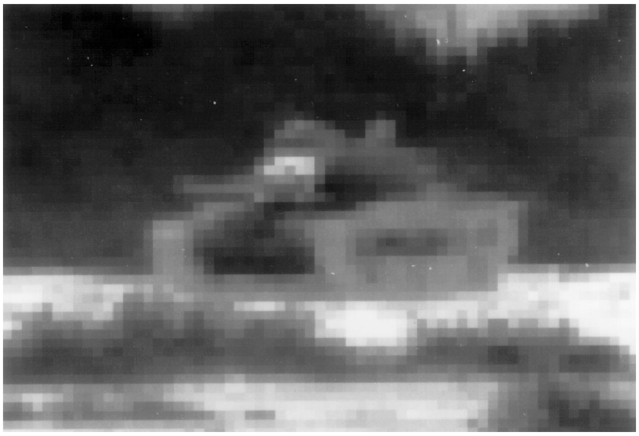


(b)

Fig. 3. (a) Conventional image in the passive infrared spectral region. (b) A laser radar image of the same region as (a).



(a)



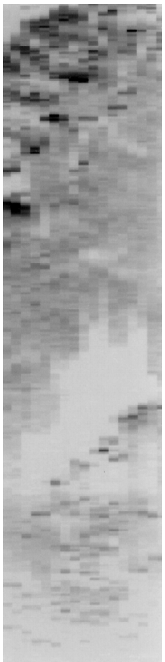
(b)

Fig. 4. (a) A wide field of view infrared image. Note the boxed region. (b) A magnified display of the boxed region in the wide field of view.

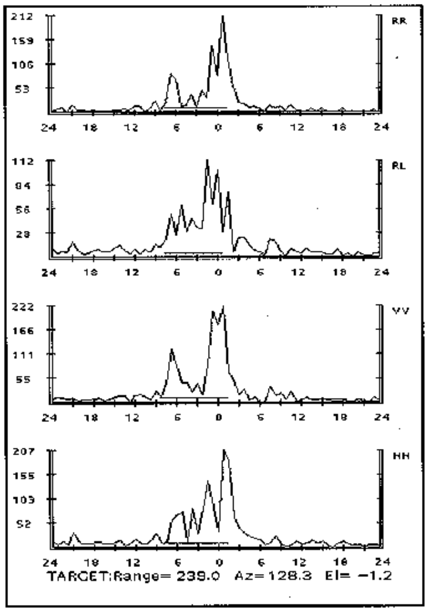
sources of data might provide contextual information that can be used to reject some possible objects of interest. Features are extracted from the segmented target to reduce the processing load for the decision-making step. Various levels of discrimination are then performed: classification, recognition, and identification. The location and description of one or more targets of interest are annotated on the final image to the operator.

The selection of features is key to the higher order discriminations that occur after segmentation. Most formal algorithms attempt to code the processes believed to be occurring in humans [15], [16]. Fundamental ATR classification is predicated on the transformation of the problem from the image space to a feature space. Feature space is made up of a set of basis vectors that correspond to certain defined measurable quantities on the candidate targets that have been segmented in the image. The feature basis set may not be complete or orthogonal. Historically, feature basis selection has been ad hoc, with examples such as perimeter length, height-to-length ratio, area-to-perimeter ratio, and moments generally based on the early work of Rosenfeld [17]. The target is then represented in feature space as a vector whose components are the values of the defined quantities that have been measured on the segmented object. The location of the vector in feature space determines the identification of that object as a target or not based upon an identification of the region of feature space as "target space" through the use of algorithm training via a set of "typical" data.

Algorithms of the early 1980s were heuristic. Typically, detection was based on some sort of threshold, determined by the contrast of an object compared to the local background in an arbitrary box drawn around the object [18]. The second action in the process was a series of steps globally referred to as segmentation. The first step in the segmentation process was typically performed by running one of the standard-textbook edge-finding operators on the region that had been detected [19], [20]. A comparison of segmentation processes can be found in Markham [21]. The next step logically connected the edge segments and filled in the gaps to form a continuous line around the presumptive target. Finally, the region was converted to a binary image by assigning a high-bit value to all pixels inside of the line. Features were then calculated on the segmented area. The calculated values were actually a vector in feature space that could be subsequently used for object sorting and so on. Classification was usually the highest level of discrimination and was based upon some sort of statistical classifier, e.g., Bayesian [22], k-nearest neighbor, or Parzen [23], [24]. Performance of these early ATR systems was found to be marginal in government testing in facilities such as those described in Section 6. Detection in low clutter did not exceed 70 percent, and recognition was little better than that obtained by random guessing. False-alarm rates in all but the most benign clutter were unacceptable. An example of an annotated FLIR imaged scene generated during testing is shown in Fig. 8. The annotations contain proper detection and classification events along with false alarms.



(a)



(b)

Fig. 5. (a) A 2D display of radar range data of the object imaged in Fig. 4. (b) A high-resolution range profile of the object imaged in Fig. 4.

Fig. 6. A typical synthetic aperture radar image.

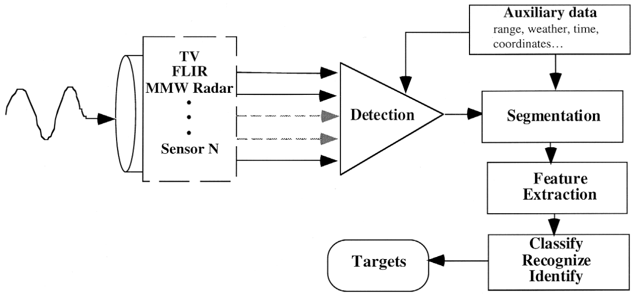


Fig. 7. The typical processing steps for an ATR from the sensor to target discrimination.

The performance shortcomings of the early processors can be attributed to the ad hoc basis for the choice of target features. The statistical distribution of these features was measured or assumed, and thresholds were chosen for statistical classifiers. Guesses based upon intuition were made by the algorithm designer as to what features needed to be calculated that would permit separation of targets from background and each other with high probability. No understanding or analysis of the scene content or the physics behind the image formation was used or available. ATR performance degraded significantly when new targets or different environmental conditions were encountered beyond the set used to train the algorithms. In addition, the implementation of the algorithms on serial processors limited the resultant performance, and information extracted in one algorithm suite component could not be shared among the various other components of the suite, e.g., frame-to-frame information or segmentor- and classifier-extracted data could not be shared. These first, threshold-based statistical ATR algorithms were not robust, as is shown in Fig. 13.

In the late 1980s and into the 1990s, a new generation of algorithms was developed that did not necessarily follow the traditional sequential processing paradigm outlined in Fig. 7. These algorithms used knowledge-based systems or template-matching approaches [25], [26], [27]. The operation of this class of algorithms can be divided into two stages: a region-of-interest (ROI) generation stage and a target identification stage. The task of the ROI stage is to locate all target-sized objects above some minimum contrast in the image. This is accomplished by convolving a simple double-window filter [28] with the image. Regions may also be expanded by merging neighboring regions with similar characteristics. Examples of region growing can be found in [29], [30], and [31]. Typically, the ROIs produced by this stage are then subjected to a template matcher in which the contents of the inner window are compared to stored templates of the target set, after ad-

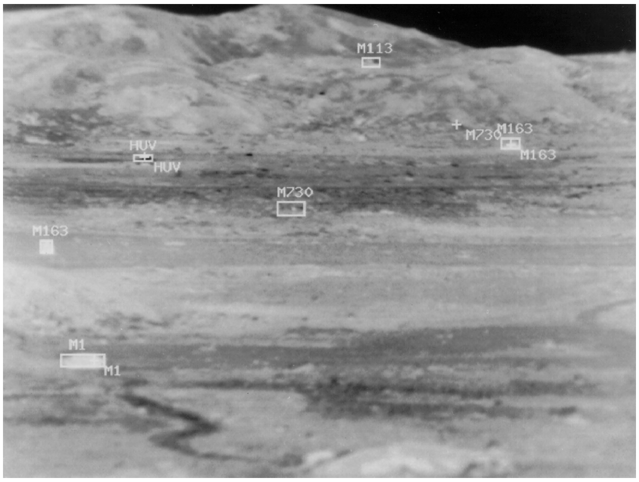


Fig. 8. Example of a simulated infrared scene with annotations made by an ATR of detections (boxes) and recognition decisions.

justment for pose and scale. The best match, usually in a mean-squared-error sense, is identified as the object in the ROI. Each match between an ROI and a template results in a score that can be subjected to a thresholding procedure for false-alarm reduction.

Performance of the advanced ATR systems has shown a significant improvement in government-controlled testing (see Fig. 13). Detection has increased to the 80 percent level in low to medium clutter conditions. However, the false-alarm rate is still high. The major improvement made in performance has been in classification and recognition. In testing of classification and recognition, two tests of increasing difficulty have been used:

- a three-class problem requiring the sorting of objects between a tank (M60), an armored personnel carrier (M113), and an Army truck (M35) and
- an eight-class discrimination problem resulting in eight possible responses to target identity.

Actual data are shown in Section 6 on measured performance.

The newest approach for increasing performance of algorithms is through the use of independent information, such as that available from multisensors, as well as integration of spatial and temporal information. The use of a model-based approach integrated with more human-like, perceptual processing neural networks is now under evaluation as a possible approach for information integration. For a survey of neural networks applied to ATR, see Roth [32].

5 Processor Architectures and Hardware

5.1 Architectures

ATR implementations can contain a wide variety of possible image-processing functions for automated processing. The subsections below indicate the range of algorithms and approaches that are found in the ATR systems we have examined.

5.1.1 Image Enhancement/Restoration

Image modifications that are currently considered desirable are automatic focusing, zoom, automatic gain and level (dynamic range) adjustment, histogram equalization, image contrast enhancement, false color, median filtering, and sharpening filters. Both automated and interactive enhancement implementations are currently being investigated. Image restoration becomes critical when the sensor is an undersampled staring array of detectors.

5.1.2 Bandwidth Compression/Decompression

On the modern digital battlefield and, surprisingly, in the modern operating room, imagery overwhelms the bandwidth of the available communication links. It is reasonable to associate the functions of image compression with the sensor processor, since the compression routine may affect the ATR implementation. A few compression routines are currently available, such as the Joint Photographics Expert Group (JPEG) routine based upon Discrete Cosine Transforms (DCT). Although not yet commercially available, compression techniques based on fractals and wavelets are being developed. We will not report on compression and its affects on ATR in this article.

5.1.3 Cuers/MTI

Cuing of a target is the first step in the classical ATR function hierarchy. One method of performing this function is to do Moving Target Indication through frame-to-frame correlations.

5.1.4 Trackers

Once a target or targets have been detected, tracking must be carried out for most fire control solutions. Trackers must be intelligent: For example, they must be able to track through clutter and obscurations, reacquire lost targets, and select aim points. Often, more than one target needs to be tracked. The operation of trackers and tracking algorithms will not be discussed further in this paper.

5.1.5 Recognizers/ID

Mistakes in identification in both medical and military images can have disastrous effects. Fratricide on the battlefield during Operation Desert Storm pointed out the need for a reliable method of identification on the battlefield. The ultimate solution to this battlefield problem will probably have ATR as part of a sophisticated situational awareness system. In medical applications, automated screening of such diagnostic tools as Pap smears would do much to lower medical cost but will gain acceptance only if the occurrence of false positives is at a very low level.

5.1.6 Sensor Fusion

The importance of fusion of information from multiple sensors will be discussed in a later section.

5.2 Hardware

The fact that ATR is a difficult problem can be illustrated by considering the number of possible computations that must be carried out in order to process real-time images. Consider a standard TV picture with approximately 100,000 pixels and five bits of gray scale. Both the size of the picture in pixels and the dynamic range are very conservative for advanced electro-optical imaging systems. However, for this example, there are $10^{150,000}$ combinations of pixel values that are possible and may be necessary for an ATR to make decisions. Although many pixels may be inconsequential and many combinations physically unrealizable, a huge library of possible candidate decisions must be considered by an ATR processor. In a dynamic scene where the image may change at TV frame rates (30 frames per second), there are potentially 10^8 digital values per second that must be processed. Obviously, ATR requires an enormous amount of processing power, speed, and memory and necessitates smart software strategies.

The early history of military ATR emphasized hardware development to supply greater computing power for the algorithms to manipulate the sensor data. Fig. 9 shows the growth in processor capability over the recent past and the projection into the near future. The computation rate in giga-floating-point operations per second is shown as a function of chronology. Single-processor growth is plotted along with that of parallel processing systems with N components. Aladdin is a Department of Defense (DOD)/Defense Advanced Research Projects Agency-developed parallel processor designed into a miniature, modular, high-density package for applications on small platforms, such as missile seekers. A typical Aladdin design will execute 500 million instructions per second and can be contained in a $2.5 \times 4.5 \times 6$ inch package. This processor can perform ATR functions on a 128×128 pixel image at 30 frames per second.

Current sensor design has been heavily influenced by algorithm limitations. Second-generation FLIR focal plane array specifications were proposed such that the sensor output data would be computer "friendly." Second-generation focal plane arrays were required to have square pixels, to be oversampled in two dimensions, to provide baseline restoration, to utilize a noninterlaced scan, to preserve scene dynamic range, and to have time delay and integration of signal from successive pixels in order to increase signal-to-noise ratio. These design characteristics were implemented in order to present more stable, highly sampled, and high signal-to-noise imagery with reduced image artifacts for algorithm compatibility.

New infrared focal plane arrays for FLIR are being designed and fabricated with a high degree of processing on the focal plane (smart focal plane arrays), resulting in a sensor that more closely resembles the eye and may provide a better approximation to human eye–brain performance.

The present and future trend for DOD development of ATR is through leverage of commercial hardware evolution and concentration on more processing capability with massively parallel architectures. The current view is that the hardware cannot be specified until we have a better understanding of the algorithms. Therefore, the emphasis is to generate processor architectures and algorithms that have significant processing potential and improved performance and that can be hosted on processor hardware utilizing commercial computer technology.

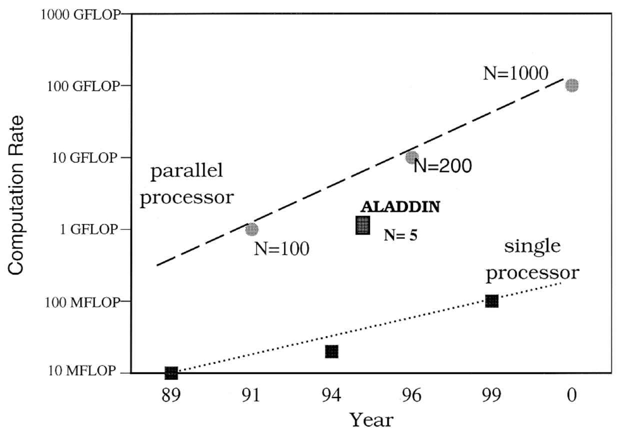


Fig. 9. Projected growth of processor computation rate as a function of chronology for single and parallel processors.

6 MEASURED PERFORMANCE

The primary factors that affect performance of ATR are signal-to-noise ratio, pixels on target, and target-like clutter. In order to assess the performance of the various ATR implementations and algorithms, a center for ATR evaluation has been created at the U.S. Army's Communications—Electronics Command Night Vision and Electronic Sensors Directorate (NVESD) at Ft. Belvoir, Virginia. At this center, ATR implementations involving man-in-the-loop are evaluated under the controlled environment of the laboratory. A physical terrain board has been employed in order to control input target parameters, scene characteristics, and atmospheric environment. With a 400:1 physical terrain board, shown in Fig. 10, the background texture and clutter can be repeated, and targets of controlled contrast and signature can be examined.

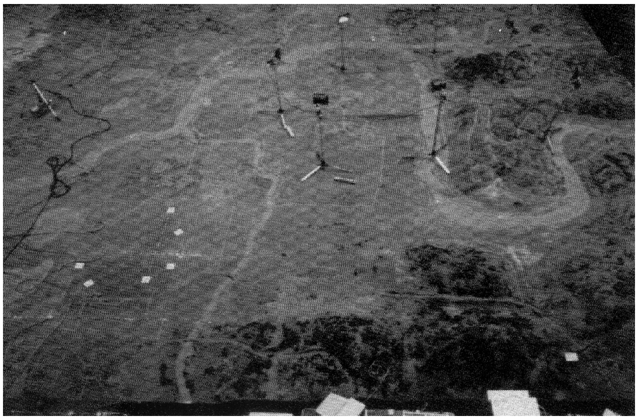


Fig. 10. Night vision 400:1 physical terrain board.

Almost all DOD-developed processors have been tested at least once in this facility since the early 1980s. Documented and quantitative data have been reported on the performance of ATR systems that show improvement over the years this facility has been operational. Much of the data appears in the classified literature and was gathered, for the most part, by C.P. Walters of NVESD.

Development is well along in the transformation of this terrain board capability into a fully computerized version. This electronic terrain board will generate multispectral imagery synthetically, filter the images with the sensor of interest representation, process the resulting images with any algorithm that is being tested, and present the final annotated image for a human operator or simulate closure on a target for missiles. This capability is essential for the scientific testing of the multisensor candidates of the future.

Fig. 11 shows the relationship between the probability of detection and the local signal-to-noise ratio (SNR) for eight of the ATR systems evaluated at NVESD. These evaluations were conducted using digital imagery taken from the NVESD physical terrain board. The algorithms cover a wide range of approaches: relational template matching, two versions of distance measuring with edge probing, template matching with triple window detection, a statistical feature-based approach, two early model-based (computational model matching) approaches, and an analog optical correlator that located the peak in the correlation signal (Flannery et al. [33], [34]). Note that it is common to find a knee in the curves between SNR values of four and seven. The best of these algorithms, in the sense of fewest false alarms for a given high probability of detection (about .80), are those that show a precipitous drop in detections below an SNR of about four or five. The curves with the high probability at the lower SNRs are those that have had their thresholds adjusted to give this result, but with high false alarms. The signal-to-noise ratio in these experiments combines both contrast and sensor noise. It is defined as the difference between the digital voltage in a target image and the local background digital voltage (the contrast) ratioed to the root mean square noise voltage due to the sensor. The results shown are averaged over many targets and aspects.

The imagery used in these experiments was taken from the NVESD physical terrain board. The sensor used in obtaining these results is a silicon TV that records reflected ambient radiation in the visible to near infrared spectral regions (0.4 to 1.2 micrometers). When the video is inverted, i.e., black-to-white and white-to-black, the resulting imagery, subjectively, looks like FLIR imagery. It is important to note that the targets and background rocks and trees were painted by hand to emulate the thermal conditions of real targets, rocks, and trees as viewed through a thermal imager. Evaluation in the simulated world, in this case using physical terrain board data, was not intended to test the absolute performance of an algorithm. Instead, it was intended to provide a baseline against which progress within an algorithm or differences between algorithms could be scientifically assessed. The physical terrain board data were always tested using algorithms designed to operate on thermal imagery. A number of experiments have demonstrated that performance of an ATR against "real" thermal imagery and equivalent physical terrain board imagery (same targets, same range, same thermal condition, same target array, same aspect ratio) was statistically identical.

The data in Fig. 11 show a not unexpected trend. The detection probability rises with the SNR and approaches a limiting value, dependent upon the detection algorithm. Of some surprise, however, is that each curve, which represents a different algorithm, has a knee around an SNR of five, independent of algorithm. There appears to be no theoretical rationale for an SNR of this value nor any intuitive reason based on experience with other signal processing techniques. It must remain a goal of further research in

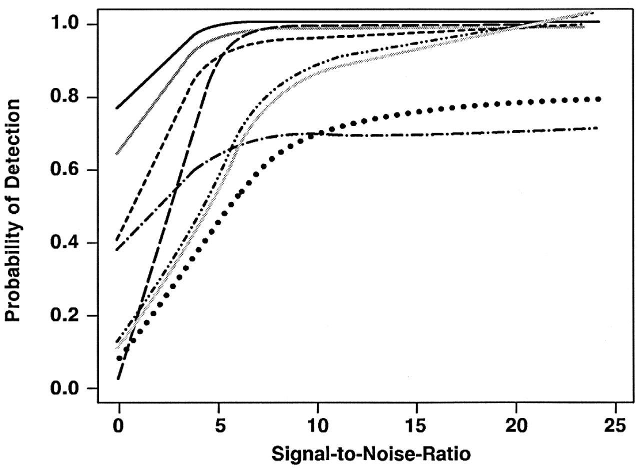


Fig. 11. Laboratory-measured detection performance of eight ATR algorithms as a function of signal-to-noise ratio.

image science to explain this experimental observation. It is important to note that these data do not include the false-alarm rate and, hence, is not representative of an ROC. The ATR algorithms used to generate the data in Fig. 11 give a wide range in false-alarm rate at the same detection rates.

Fig. 12 shows the affect of sensor resolution on the probability of detection at several clutter levels for a modern template-matching algorithm. These results also come from an evaluation using the NVESD physical terrain board imagery. The algorithm generated a match score as a result of comparing the inner window ROI data with stored templates of 10 different targets. The score was subjected to a threshold to eliminate false alarms. There are several interesting affects demonstrated in these results. First, note the general trend of decreased detection performance as resolution decreased (image pixels subtend larger amounts of the real world). Second, note that in high clutter, there is decreased detection performance both at high and low resolution. The decreased performance at high resolution is due to increased uncertainty about the boundaries of the target. At high resolution, uncertainty over where the target ends and a background object begins as well as uncertainty over subareas within the target results in confusion of the algorithms. Generally, the impact of clutter is to degrade the maximum detection probability by up to 40 percent, as can be seen in the figure. The clutter specification is arbitrary and subjective due to the lack of a scientific definition of a clutter metric. The clutter types represented in the data in the figure correspond to the clutter examples shown in Fig. 1. Not shown in Fig. 12 is the fact that clutter has a very severe affect on false-alarm rate.

While an ATR algorithm may exhibit performance levels below the human's in terms of probability of detection, it is tireless, and the speed with which it performs the functions is often many times faster than a human can. The speed enhancement coupled with a human engaged in the final decision should result in the execution of more tasks or tasks of greater complexity, but currently a formal metric for this enhancement has not been created.

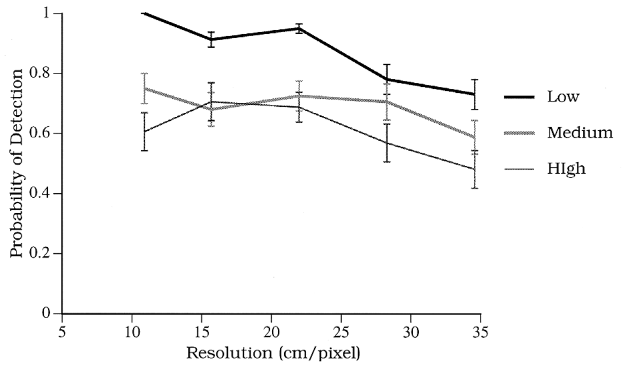


Fig. 12. Laboratory-measured detection performance for an ATR as a function of resolution (range) for three different subjective clutter levels.

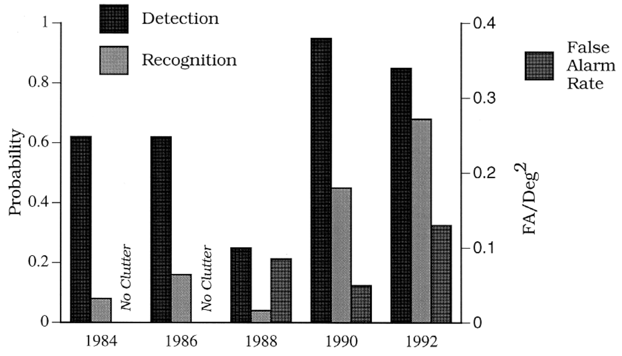


Fig. 13. Performance progress for ATR systems performing the three-class sorting task for the condition of medium-contrast targets in low clutter. There was no clutter in the imagery used in 1984 and 1986.

The question of how well an ATR implementation performs compared to a human was addressed by Walters. The algorithms tested included a variety of approaches, ranging from a classic segmentation / statistical classifier to several template-matching approaches and even included an optical processor (an optical correlator using digitally segmented images). Table 2 shows the results from extensive testing with the same imagery given to humans and ATR systems. The clutter level was relatively benign. The probabilities reported were the maximum and minimum values for either the humans or the two groups of ATR implementations performing object discrimination: one ATR implementation based on assignment to three classes and one based on assignment to eight classes. Probability of discrimination is conditioned on correct detection. The human testing was a forced-choice experiment, i.e., “don’t know” was not an acceptable answer.

The results indicate that for the conditions and the ATR implementations tested, the detection capabilities of the humans and automatic systems were approximately equal, whereas the false-alarm rate was significantly greater for the ATR implementations, even in this low-clutter situation. The performances for humans and ATR implementations for the three-class discrimination problem were essentially equal. Eight-class discrimination was significantly more difficult for the ATR implementations than for the humans.

Table 3 is the confusion matrix produced during these experiments of Walters. This table shows the assignment a human or ATR gave to a detected object compared to the object’s actual identity. Confusion matrices for detection are convenient ways for comparing the information contained in ROC curves. The results in Table 3 demonstrate that the humans outperform the tested ATR implementations on all target types in the eight-class problem. The human and the ATR implementations do not confuse the same targets. It is possible that humans and ATR implementations are not using the same target features to perform recognition. These types of experiments may provide insight into the design of autonomous devices.

One of the great values of the performance data gained from testing ATR implementations at the NVESD facility is the historical perspective that can be obtained. The progress made in ATR performance over the years can easily be documented because of the controlled conditions under which the testing is done. The same input scenes under the same conditions can be repeated on new algorithms, making comparisons possible of performance on different ATR implementations separated by years.

Fig. 13 is an example of an analysis of performance over an extended period. Probabilities of detection and recognition are plotted along the left-hand vertical axis, and false alarms per square degree in object space are on the right-hand axis. The horizontal axis represents chronological time. It can be seen that in the mid to late 1980s, detection performance was good, but three-class recognition was essentially chance probability. The algorithms tested in 1984 and 1986 were traditional statistical classifiers. Starting in 1988, template-matching approaches were evaluated in clutter environments that presented a significant challenge to the algorithms. By 1992, the evaluations included both more advanced template-matching approaches and primitive model-based algorithms. In 1984 and 1986, the testing was done only in backgrounds with no clutter because of the high false-alarm rate. As the newer, template-matching algorithms and model-based algorithms were designed, significant improvements in recognition and false-alarm rate were made. Further significant improvements in detection, recognition, and false-alarm rates are expected with the use of multisensor fusion, e.g., FLIR and MMW radar, and is the subject of many papers in the classified literature.

7 MULTISENSOR FUSION

As improvements in sensors have been realized through development of imaging technologies, the optical limits of sensor resolution have been reached, and the second-generation FLIR is a prime example. The size of the infrared detectors are at the diffraction limit of the optical objective, and the sensitivity is close to Background Limited Performance. Hence, more advanced sensors will provide only small optical improvements over present capability. We should be able to obtain more dramatic gains in performance through the use of independent information, such as that available from multisensors, as well as integration of spatial and temporal information. In order to make large

TABLE 2
OVERALL PERFORMANCE FOR HUMANS AND ATR SYSTEMS

		Probability of detection	False alarm rate	Probability of 8-class discrimination	Probability of 3-class discrimination
ATR systems					
	Max	0.869	13.323	0.353	0.732
	Min	0.604	3.532	0.268	0.541
	Mean	0.688	8.195	0.289	0.705
Human observers					
	Max	0.833	0.9	0.814	0.798
	Min	0.52	0.017	0.298	0.343
	Mean	0.683	0.234	0.586	0.663

TABLE 3
CONFUSION MATRIX FOR HUMANS, THREE- AND EIGHT-CLASS DISCRIMINATIONS

Ground truth	System	M60	M113	M35	M1	M2	M998	M163	M730
M60	3-Class	0.67	0.21	0.12	NA	NA	NA	NA	NA
	8-Class	0.17	0.12	0.05	0.13	0.08	0.13	0.19	0.11
	Human	0.77	0	0	0.07	0.09	0.02	0.02	0.01
M113	3-Class	0.08	0.66	0.27	NA	NA	NA	NA	NA
	8-Class	0.02	0.25	0.12	0.07	0.02	0.24	0.19	0.09
	Human	0.02	0.65	0.03	0.03	0.08	0.11	0.08	0.01
M35	3-Class	0.18	0.21	0.61	NA	NA	NA	NA	NA
	8-Class	0.12	0.07	0.36	0.11	0.1	0.12	0.05	0.08
	Human	0.01	0.06	0.85	0.02	0.01	0.01	0.01	0.02
M1	8-Class	0.04	0.2	0.15	0.43	0.06	0.09	0.04	0
	Human	0.03	0.03	0.01	0.89	0.02	0.03	0	0
M2	8-Class	0.12	0.14	0.13	0.16	0.27	0.07	0.04	0.07
	Human	0.16	0.04	0	0.01	0.58	0.02	0.14	0.05
M998	8-Class	0.02	0.17	0.09	0.05	0.09	0.42	0.13	0.04
	Human	0.03	0.09	0.01	0.04	0.11	0.62	0.08	0.04
M163	8-Class	0.1	0.25	0.02	0.02	0	0.23	0.27	0.12
	Human	0.02	0.19	0.01	0.02	0.25	0.04	0.44	0.02
M730	8-Class	0.15	0.21	0	0.1	0	0.01	0.16	0.37
	Human	0	0	0	0	0.03	0.02	0.03	0.91

increases in performance—for example, in false-alarm reduction in higher clutter—use must be made of new, independent information. Independent parameters not now available in 2D imagery would be those derivable from information from other sensor modalities or from nonimaging sources.

Range information is a prime example of a quantity that would complement a 2D image. Range from a laser rangefinder or a radar can give critical new information to a processor trying to extract targets from FLIR or TV images. It is necessary that the laser or radar returns be registered on the imagery, resulting in a 3D image. Range data can also be extremely valuable in sizing ATR processing windows around potential targets that permit rejection of false alarms for objects too small or too large to be a target at the range specified.

The data from more than one sensor can be fused by several techniques: information/data, pixel, feature, and decision-level fusion. Data fusion refers to the incorporation of target data from several sources, e.g., imaging sensor, GPS, and digital map. Pixel fusion involves the overlay of pixels from two sensors to form an image, e.g., TV and FLIR. Feature fusion correlates independent feature information from two or more sensors prior to making a decision. Decision fusion is a voting scheme in which each sensor is polled as to the presence of targets. Other methods of utilizing multisensor data corresponding to higher level descriptions of images are expected to yield better performance from fusion. One of many possible formalisms that is being pursued is the establishment of militarily relevant perceptual building blocks for recognition similar to geons [35] as well as application of Biederman's Recognition-by-Components theory [36] to the problem of extracting information from images that are important to recognizing the object. Nair and Aggarwal [37], [38] have pursued a recognition by object parts for target recognition in FLIR images based on computer vision concepts. Similarly, Lanterman et al. [39] describe a pattern theoretic approach to fusing MMW radar and FLIR imagery to perform detection, recognition, and tracking.

A series of recent tests by the Army has demonstrated the gains in automated performance that can be realized by combining data from several sensors. In these tests, a modern FLIR imaging sensor (operating from eight to 12 µm), a LADAR, and an MMW (35-GHz) radar were used in the modes discussed in the following paragraphs.

Imagery was taken from an airborne platform by an FLIR along with pixel-registered range and polarimetric data obtained using a 35-GHz radar and LADAR range profiles. This imagery was then processed by an ATR that used a feature fusion approach. In one test, the ATR implementation's performance was evaluated using three data classes: FLIR only, fused LADAR/FLIR, and fused MMW/FLIR data. The imagery used in the experiment was gathered at a variety of times in the diurnal cycle. NVESD scientists, using data taken in a desert environment, reported significant improvements in detection and recognition performance with dramatic improvements in false-alarm rate. Fig. 14 shows the probability of detection versus normalized false-alarm rate for the three data classes. There is almost an order of magnitude improvement in false-alarm rate for fusion over single-sensor data.

Other tests using imagery from a region of farmland and wooded lots [40] show the overall benefit of feature fusion. In Fig. 15, the probabilities of detection and recognition for three data classes of FLIR only, MMW radar only, and fused MMW/FLIR data display some improvement. However, the false-alarm rate in this experiment dropped significantly for the conditions encountered in the test. This data should not interpreted as demonstrating that false alarms have been eliminated, but rather that there is experimental evidence that sensor fusion can provide significant reductions in false-alarm rates, e.g., an order of magnitude in some cases, over single-sensor performance. The improvement in false-alarm rate due to sensor fusion is not surprising but the extent of that improvement was unexpected. More details on this set of experiments are in the classified literature.

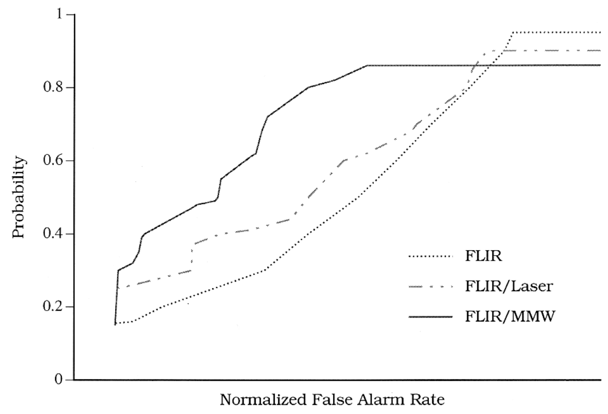


Fig. 14. Detection performance versus false-alarm rate for a single (FLIR) sensor and multisensors.

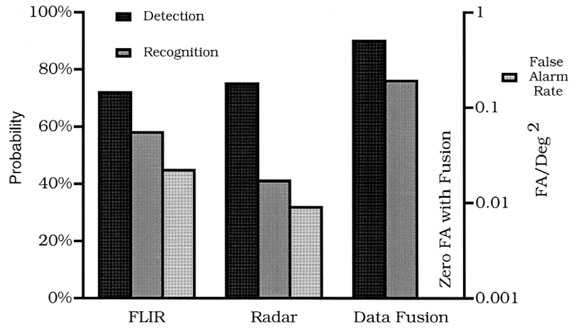


Fig. 15. Detection, recognition, and false-alarm performance for single sensors (FLIR or radar) and sensor fusion.

These first-reported field experiments of sensor fusion have shown dramatic performance improvement over single-sensor processing performance. Other combinations of multiple sensors will need to be investigated to determine what, if any gains can be obtained utilizing other sensor technologies, such as acoustic detection, near-infrared image intensifiers, and visible television imagery. In addition to sensor fusion, data fusion must be investigated to determine the merits of nonimage information for decision making. Examples of sources for data fusion are weather descriptors, map coordinates, and operational/contextual information. Complete descriptions of sensor fusion activities in the military and university communities have recently appeared in government reports and [41], [42].

8 FUTURE DIRECTIONS

There are nearly an infinite number of possible targets and backgrounds in an imaged scene. There are a large number of interesting targets in military applications alone. These various targets may have a variety of signatures in the various spectral regions. Infrared and other passive imaging modalities generate signatures that vary with time of day, season, solar loading, history, geographical location, operational status, aspect of presentation, friction, atmospheric conditions, and so on. In addition, the target may be partially obscured, have low contrast, be contained in a low-resolution image, be camouflaged, and so on. The background can have none or any amount of target-like clutter objects that compete with the real targets for attention.

The essence of understanding performance of ATR implementations is an understanding of image science, by which we mean the mathematical descriptions of targets and clutter that enable understanding of the fundamentals of the image content. The hindrance to developing an understanding is associated with the variability of the objects that must be detected and recognized in a plethora of scene backgrounds. The quantitative relationship between the physics of the scene and what information can be extracted from it by some form of processing does not exist. Adequate signature models or scene descriptions that can be used as a starting point for quantifying the input to an ATR implementation do not exist. Once a formulation has been created, the sensor effects could easily be introduced, since sensor phenomenology is well-understood. However, even a technique for bounding the performance of an ATR based upon scene content would be a significant step forward. One of the first steps is the bounding techniques for estimating target pose shown by Grenander et al. [43] using Hilbert-Schmidt bounds for estimators on matrix Lie groups.

Probably the key problem associated with understanding and predicting ATR performance is the scientific understanding and quantification of clutter as it affects detection. Understanding and quantifying clutter as it affects performance is a problem that has defied solution for many years. It was a problem when considering human performance but was largely ignored by designing experiments with small false-alarm rates. Human observers were told to respond only when they had a high certainty of the existence of a target and its identity. We do not have this option with ATR testing.

There have been several attempts to quantify clutter. Schmeider and Weathersby [44] used a target minus background radiance to root mean square clutter radiance ratio with limited success, and Shirvaikar and Trivedi [45] used gray-level co-occurrence matrices to define a texture-based image clutter metric. The key to clutter quantification has to do with how competitive the clutter objects are to targets. For example, clutter in the U.S. desert in infrared imagery is high when searching for men but is extremely low when looking for vehicles. This is because the cacti look like men in the infrared scene, but there is little that looks like an automobile. Hence, any metric that purports to quantify clutter must be related to the target of interest, and that metric will significantly affect the design of detection algorithms for separating targets from the background. A clutter metric devoid of a target relationship will not work. With the advancement of computer processing, some computing-intensive technique might prove successful. For example, the convolving of targets across a scene to generate correlations between the target and scene objects might provide a reasonable clutter metric. Some of the approaches used to evaluate image complexity or image quality may also offer useful metrics.

Once a clutter metric and a method for describing targets and backgrounds are in hand, the next order of business is the generation of a performance model for the sensor-ATR system. Such a model would allow the design and optimization of the sensor and the processor for given scenarios. Sensitivity analyses based upon validated models that could show the affects on ROC curves and confusion matrices due to sensor and algorithm variations would be an enormous benefit to the development community. Skeptics in the ATR community suggest that the lack of clutter metrics coupled with the wide variability of algorithm concepts make the modeling problem intractable. Whether true or false, the establishment of bounding techniques for ATR performance will provide significant benefit to the system designers, by showing how close a particular algorithm implementation is to the theoretical bound of performance.

Limited progress in establishing a theoretical foundation for ATR has been due to the fact that the military requirements have driven the community to ad hoc and heuristic approaches to algorithm development. Lanterman et al. [46] and Grenander and Miller [47] have shown recent progress in applying pattern theoretic approaches to scene understanding in medical and military applications. However, now that the need for a rigorous pursuit of an image science has been recognized, it is clear that scientific experiments must be performed that supply data for guidance of theory, verification, and validation of models. Absolutely essential to the experimental approach is a canonical set of imagery, covering all spectral regions of interest, that is broadly available to the research community.[3] In the past, such important data have been severely restricted due to security classification guidelines. In addition to more representative imagery, more human performance data are required in order to make comparisons between human and automated performance. ATR performance is not expected to be better than human performance in terms of probabilities and false-alarm rate, but ATR is expected to perform significantly faster. The level at which a human performs must be accurately known, so that the trade-off can be made by systems developers between the accuracy of the human and the speed of processors.

As discussed above, recent experiments indicate that multisensor integration may provide the level of ATR required by various applications. The utilization of independent parameters from the scene appears to give performance improvements that make ATR possible. In medical imaging, multisensor fusion is already used to provide contextual information to the radiologist. Hand in hand with the employment of multisensor input data must come investigation of new paradigms, such as model-based algorithms. Model-based algorithms contain libraries of models of the targets for scenarios of interest. Target models coupled with environmental affects models presumably can

3. A data base of military relevant targets is now available to the academic community at http: / /cis@cis.wust.edu.

represent any state in which the target can occur. Then, the images that the sensor produces are compared to the library models until a match occurs with some level of confidence. This model-based approach shown pictorially in Fig. 16 can be compared to the previous approach represented in Fig. 7. The model-based approach addresses several functions simultaneously, such as recognition, classification, and identification, whereas the older approach carries out these operations serially. This image processing, model-based concept for ATR is very familiar to the computer vision community [48] ,[49]. For a more detailed example of a model-based approach that fuses range and intensity information, see Verly et al. [50].[4]

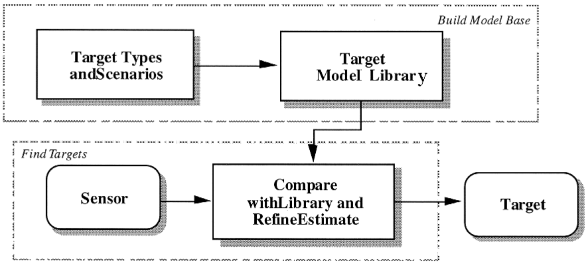


Fig. 16. Model-based algorithm concept.

The building of the model templates from the canonical database of targets is essential to the generation of model-based algorithm development. Fig. 17 illustrates one possible procedure for the creation of the required library. From a validated 3D computer-aided design model, 2D silhouettes are fabricated by the computer. The image features of the silhouettes are sampled to form clusters of targets with common features. The clusters are then probed for edges, vertices, and other qualities that distinguish the 2D representations of the targets. A search tree is constructed that can be interrogated in a logical process to arrive at the correct identity of an object. For an example of this approach to relational template matching, see Kramer et al. [52].

The description of model-based algorithm development highlights the necessity of real-world imagery or realistic synthetic scene generation. The validity of imagery taken in the real world is absolute. However, the environment in which the target exists can make the database useless when applications change. Synthetic scene generation provided in an electronic terrain board, on the other hand, can give an infinite variation of conditions, but the validity of the scene must always be established. Established scene metrics would allow validation of synthetic scenes without the need to continually acquire real-world imagery. Even with a good theoretical basis from image science, synthetic scene generation is hampered by the very high data rates required for interactive simulation. Let us assume that the synthetic image must have 10 times the sensor resolution before sampling by the sensor. For a modern FLIR with 480 $\times$ 1,280 pixels and 30 frames per second, the scene generator must produce 1.5×10^8 pixels per second.

4. For definitions of the differences among image processing, pattern recognition, and computer vision, see Nandhakumar and Aggarwal [51].

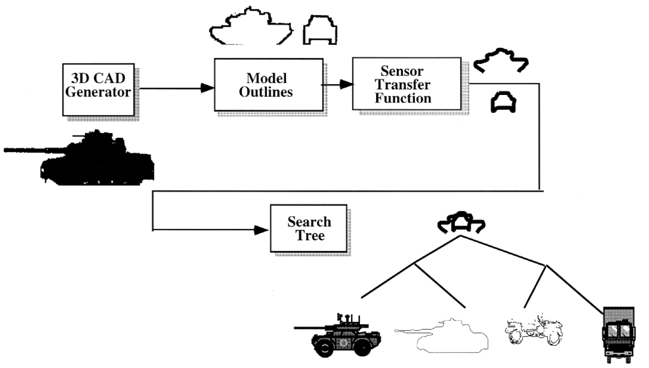


Fig. 17. ATR relational template-matching concept for designing decision search trees from sampling target models filtered by the sensor.

9 Conclusions

As seen in this survey of the state-of-the-art of ATR technology as measured by Army Laboratories, the present state-of-the-art of ATR systems is still far from imitating the performance of the eye–brain combination except for a few select ATR implementations. For some low to medium clutter scenarios in which the number of target possibilities is not very great, the target acquisition performance by the ATR is good enough to be useful to the military. The useful ATR systems are realizable in hardware that host algorithm implementations that exist and have been demonstrated against militarily relevant scenes. However, current ATR algorithms cannot be accurately modeled; there is little correlation between known image metrics and observed ATR performance.

Experimental results justify the pursuit of sensor fusion by the military in order to realize significant improvements in performance. Improvements imply maintaining probabilities of detection, classification, recognition, and identification of larger classes of targets while reducing false alarms in higher clutter levels. This means moving from one set of ROCs to another higher performance set by introducing new, independent information. The biological examples of using several senses to make decisions about scene content support the concept. However, the lack of theoretical guidelines from image science does not suggest how to proceed down this path.

This paper also suggests that the major shortcoming in the ATR technology base has to do with the lack of scientific underpinnings. The present indication that multisensor fusion will provide the enhanced performance was not based on any predictive scientific theory. It was based upon intuition and trial and error. A strong, active image science that provides image metrics—especially clutter measures, understanding of scene information, and models—and that performs experiments to generate data that lead to model formulation and validation is needed to indicate the most fruitful scientific endeavors for rapid progress in the technology area.

It is also contended, based upon the developments presented in this paper, that what are truly needed are new ideas on the information content of a scene and how to take advantage of it. It does not appear that a great increase in processor computing power is required. The major increases in performance reported here have resulted from the transition to model-based algorithms and the employment of multiple sensors. These are the kinds of new ideas and concepts that will generate significant leaps in ATR performance.

Finally, tools that will enable the ultimate solution to the ATR problems are maturing. Signature and background modeling and databases are being expanded with real-world imagery, and synthetic generation efforts are under way. Evaluation methodologies are continuing to evolve. Algorithms are evolving, but so far not in a revolutionary way. Rapid prototyping techniques for algorithm generation and processor hardware fabrication are expected to permit the quick assessment and implementation of new ATR improvements. As all these complementary activities come to fruition, the establishment of an organized ATR approach, at least in the DOD, will enable the efficient fielding of many applications. Military fielding of ATR can then set the stage for effective transition to civilian markets.

Acknowledgment

Rudolf G. Buser's photograph published with permission. U.S. Army photograph.

References

[1] B. Bhanu and T. Jones, "Image Understanding Research for Automatic Target Recognition," *IEEE AES Systems Magazine*, pp. 15-22, Oct. 1993.

[2] S. Chow and T. Jones, "Army's FLIR/ATR Evolution Path," *Proc. SPIE*, vol. 1,050, no. 6, 1989.

[3] Marr, D., *Vision*. San Francisco: W.H. Fueman, 1982.

[4] W.K. Pratt, *Digital Image Processing*. New York: John Wiley & Sons, 1978.

[5] Y.-T. Zhou and R. Chellappa, *Research Notes in Neural Computing Volume 5, Artificial Neural Networks for Computer Vision*. New York: Springer-Verlag, 1992.

[6] A. Rosenfeld, "Survey: Image Analysis and Computer Vision 1996," *Computer Vision and Image Understanding*, vol. 66, no. 1, pp. 33-93, 1997.

[7] L.M. Biberman, *Perception of Displayed Information*. New York: Plenum Press, 1973, pp. 183-187.

[8] J.A. Ratches, "Static Performance Model for Thermal Imaging Systems," *Optical Eng.*, vol. 15, no. 6, pp. 525-536, 1976.

[9] C.P. Walters, "Remove the Automatic Target Recognition Performance Evaluation Bottleneck: The CNVEO AUTOSPEC Facility," *Optical Eng.*, vol. 30, no. 3, pp. 247-253, 1991.

[10] L.G. Clark, L.I. Perlovsky, W.H, Schoendorf, C.P. Plum, and T.L. Keller, "Evaluation of FLIR Sensors for Automatic Target Recognition Using an Information-Theoretic Approach," *IEEE Proc. NAECON*, vol. 2, pp. 1,048-1,055, 1993.

[11] L.G. Clark and J.V. Velten, "Image Characterization for Automatic Target Recognition Algorithm Evaluations," *Optical Eng.*, vol. 30, no. 2, pp. 147-153, 1991.

[12] K. Augustyn, "A New Approach to Automatic Target Recognition," *IEEE Trans. AES*, vol. 28, no. 1, pp. 105-114, 1992.

[13] A. Lanterman, M.I. Miller, and D.L. Snyder, "Representations of Thermodynamic Variability in the Automated Understanding of FLIR Scenes," *Automatic Object Recognition 6, Proc. SPIE*, vol. 2,756, pp. 26-37, Apr. 1996.

[14] M. Cooper, U. Grenander, M. Miller, and A. Srivastava, "Accommodating Geometric and Thermodynamic Variability for Forward-Looking Infrared Sensors," *Algorithms for Synthetic Aperture Radar, Proc. SPIE*, vol. 3,070, 1997.

[15] J.R. Pomerantz, "Are Complex Visual Features Derived From Simple Ones?" *Formal Theories of Visual Perception*, E.L.J. Leeuwenberg and H.F.J.M. Buffort, eds. New York: John Wiley & Sons, 1978.

[16] J.R. Pomerantz, "Perceptual Organization in Information Processing," *Perceptual Organization*, M. Kubovy and J.R. Pomerantz, eds. Hillsdale, N.J.: Erlbaum, 1981, pp. 141-180.

[17] A. Rosenfeld and M. Thornton, "Edge and Curve Detection for Visual Scene Analysis," *IEEE Trans. Computers*, vol. 20, pp. 562-569, 1971.

[18] B.J. Schachter, A. Lev, S.W. Zucker, and A. Rosenfeld, "An Application of Relaxation Methods to Edge Reinforcement," *IEEE Trans. Systems, Man and Cybernetics*, vol. 7, pp. 813-816, 1977.

[19] D.K.W. Walters, "Computer Vision Model Based on Psychophysical Experiments," *Pattern Recognition by Humans and Machines, Volume 2: Visual Perception*, E.C. Schwab and H.C. Nusbaum, eds. New York: Academic Press, 1986.

[20] T. Poggio, "Early Vision," *Perspectives in Computing*, A. Rosenfeld, ed. New York: Academic Press, pp. 190-206, 1986.

[21] K.C. Marham, "Comparison of Segmentation Processes for Object Acquisition in Infrared Scenes," *IEE Proc. F, Radar and Signal Proc.*, vol. 136, no. 1, pp. 13-21, Feb. 1989.

[22] K. Fukunaga, *Introduction to Pattern Recognition*. New York: Academic Press, 1990.

[23] K. Fukunaga, M. Keinosuke, and D.M. Hummels, "Bayes Error Estimation Using Parzen and K-NN Procedures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, 1987.

[24] K. Fukunaga and R.R. Hayes, "The Reduced Parzen Classifier," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, pp. 423-425, Apr. 1989.

[25] J. Baras and A. LaVigne, "Convergence of a Neural Network Classifier," *Proc 29th IEEE Conf. Decision and Control*, pp. 1,735-1,740, Dec. 1990.

[26] *Vector Quantization and Signal Processing*, A. Gersho and R. Gray, eds. New York: Academic Press, 1991.

[27] K.S. Fu, "Syntactic Models in Pattern Recognition and Applications," *Pattern Recognition in Practice*, E.S. Gelsema and L.N. Kanal, eds. Amsterdam: North-Holland, 1980, pp. 303-314.

[28] P.J. Burt, "Smart Sensing in Machine Vision," *Machine Vision: Algorithms, Architectures, and Systems*. New York: Academic Press, 1988.

[29] D. Nair and J.K. Aggerwal, "A Focused Target Segmentation Paradigm," *Proc. Fourth European Conf. Computer Vision*, vol. 1, pp. 579-588, Cambridge, England, 1996.

[30] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884-900, Sept. 1996.

[31] J. Kitrosser, "An Evaluation of a Region-Growing Algorithm Performing Detection for Automatic Target Recognition," *J. Imaging Science and Technology*, vol. 38, no. 4, pp. 311-320, July/Aug. 1994.

[32] M.W. Roth, "Survey of Neural Network Technology for Automatic Target Recognition," *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 28-43, Mar. 1990.

[33] D. Flannery, J. Loomis, M. Milkovich, and P. Kellin, "Applications of Binary Phase-Only Correlation to Machine Vision," *Optical Eng.*, vol. 27, no. 4, pp. 309-320, 1988.

[34] D. Flannery, J. Loomis, and M. Milkovich, "Transform-Ratio Ternary Phase-Amplitude Filter Formation for Improved Correlation Discrimination," *Applied Optics*, vol. 27, no. 19, pp. 4,079-4,083, 1988.

[35] J.E. Hummel and I. Biederman, "Dynamic Binding in a Neural Network for Shape Recognition," *Psychological Rev.*, vol. 99, no. 3, pp. 480-517, 1992.

[36] I. Biederman, "Recognition-by-Components: A Theory of Human Image Understanding," *Psychological Rev.*, vol. 94, no. 2, pp. 115-147, 1987.

[37] D. Nair and J.K. Aggarwal, "Hierarchical, Modular Architectures for Object Recognition by Parts," *Proc. 13th Int'l Conf. Pattern Recognition*, Vienna, pp. A601-A606, 1996.

[38] D. Nair and J.K. Aggarwal, "Robust Automatic Target Recognition in Second Generation FLIR Images," *Proc. IEEE Workshop Applications of Computer Vision*, pp. 194-201, Sarasota, Fla., 1996.

[39] A. Lanterman, M.I. Miller, and D.L. Snyder, "The Unification of Detection, Tracking, and Recognition for Millimeter Wave and Infrared Sensors," *Radar/Ladar Processing, Proc. SPIE*, vol. 2,562, pp. 150-161, July 1995.

[40] D. Reago, E. Lloyd, W. Gercken, E. Jones, M. Gahler, and T. Bui, "Fusion of 2nd Generation Forward Looking Infrared, Millimeter Wave Radar and High Pulse Rate Laser Sensors for Tactical Target Acquisition," *Proc. 19th Army Science Conf.*, DTIC No. AD-B197265, Orlando, Fla., June 1994.

[41] *Multisensor Fusion for Computer Vision*, J.K. Aggarwal, ed. New York: Springer-Verlag, 1991.

[42] J.R. Maheshkumar, V. Veesana, S.S. Iyengar, and R. Brooks, "New Computational Techniques for Complementary Sensor Interpretations in Detection-Localization Systems," *Optical Eng.*, vol. 35, no. 3, 1996.

[43] U. Grenander, M.J. Miller, and A. Srivastava, "Hilbert-Schmidt Lower Bounds for Estimators on Matrix Lie Groups," unpublished paper.

[44] D.E. Schmieder and M.R. Weathersby, "Detection Performance in Clutter With Variable Resolution," *IEEE Trans. AES*, vol. 19, no. 4, pp. 622-630, 1983.

[45] M.V. Shirvailkan and M.M. Trivedi, "Developing Texture-Based Image Clutter Measures for Object Detection," *Optical Eng.*, vol. 31, no. 2, pp. 2,628-2,639, Dec. 1992.

[46] A. Lanterman, M.I. Miller, and D.L. Snyder, "Jump-Diffusion Processes for the Automated Understanding of FLIR Scenes," *Automatic Object Recognition 4, Proc. SPIE*, vol. 2,234, pp. 416-427, Apr. 1994.

[47] U. Grenander and M.I. Miller, "Representation of Knowledge in Complex Systems," *J. Royal Statistical Soc.*, vol. 56, no. 4, pp. 569-603, 1994.

[48] F. Arman and J.K. Aggarwal, "Model-Based Object Recognition in Dense-Range Images—a Review," *Comm. ACM*, vol. 25, no. 1, pp. 5-43, Mar. 1993.

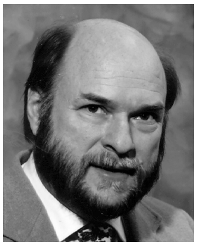
[49] F. Arman and J.K. Aggarwal, "CAD-Based Vision: Object Recognition in Cluttered Range Images Using Recognition Strategies," *Computer Vision Graphics and Image Processing: Image Understanding J.*, vol. 58, no. 1, pp. 33-48, 1993.

[50] J. Verly, R.L. Delanoy, and D.E. Dudgeon, "Model-Based System for Automatic Target Recognition From Forward-Looking Laser-Radar Imagery," *Optical Eng.*, vol. 31, no. 12, pp. 2,540-2,552, Dec. 1992.

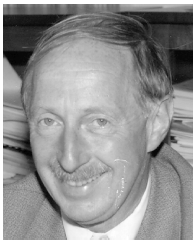
[51] N. Nandhakumar and J.K. Aggarwal, "Multisensory Computer Vision," *Advances in Computers,* vol. 34, p. 60, 1992.

[52] A. Kramer, D. Perschbacher, R. Johnston, and T. Kipp, "Relational Template Matching Algorithm for FLIR Automatic Target Recognition," *SPIE Symp. Aerospace and Remote Sensing*, pp. 29-37, 1993.

James A. Ratches received a BS from Trinity College in Hartford, Connecticut, in 1964 and an MS and PhD in physics from Worcester Polytechnic Institute in 1966 and 1969, respectively. Dr. Ratches is the chief scientist at the U.S. Army's Communications-Electronics Command Night Vision and Electro-Optics Directorate (NVESD). In his 28 years at NVESD, he has been a major contributor to the areas of modeling, simulation, characterization, and evaluation of electro-optical systems. His performance model for thermal imaging systems became the standard for the entire infrared community. As director of the Visionics Division of NVESD, he reoriented the modeling, simulation, and evaluation activities from emphasis on human to machine performance. As chief scientist, he has become a spokesman for the military technical ATR community.



Clarence P. Walters received his BA and MA in psychology from Pacific Lutheran University in 1967 and 1970, respectively, and the PhD in experimental psychology from George Washington University in 1974. Dr. Walters was a postdoctoral fellow at Purdue University, specializing in the neuroanatomy of olfactory and visual systems. He is a senior research psychologist at the U.S. Army's Communications-Electronics Command Night Vision and Electro-Optics Directorate (NVESD). He has been at NVESD since 1982, where he has been the technical leader in developing the methodologies for the scientific assessment of ATR performance. He has applied his training in experimental psychology to the field of image understanding and ATR. Previously, for 10 years, he was an assistant professor of psychology at George Mason University, specializing in perception, experimental methods, and sensory processing.



Rudolf G. Buser received his doctorate in natural sciences from the University of Munich, Germany, in 1957. After working for a short time at Siemens in advanced digital and analog computer concepts and as an assistant professor of physics at the university, Dr. Buser came to the United States in 1958. He became a research leader with the Army at Ft. Monmouth, New Jersey, performing basic research in lasers and enabling the transition of this technology to Army laser rangefinders, designators, radars, combat identification, countermeasures, and aided surveillance and target acquisition systems. In the 1970s, he became a member of the Army's Night Vision Laboratory at Ft. Belvoir as director of its Laser Division. Dr. Buser subsequently became the director of the entire laboratory from 1989 until 1966. As director of the Army's sensor program, he was responsible for the development and transition of optical, thermal imaging, image intensification, image processing, radar, platform protection, sensor modeling and simulation, and laser technologies. Dr. Buser has been a leader in the technical national/international communities, participating in numerous prestigious panels and as a member of the American Physical Society and as a fellow of the Optical Society.

Bob D. Guenther received the BS in physics from Baylor University and the MS and PhD in physics from the University of Missouri—Columbia in 1960, 1963, and 1968, respectively. Dr. Guenther was employed by the U.S. Army Missile Command from 1968 to 1979. From 1968 through 1972, he was involved in research in solid state physics. From 1973 to 1975, he was the Army representation to the Air Force High Energy Laser Program and was involved in the technical investigation of new laser concepts. Beginning in 1975, he was engaged in the field of optics, specializing in the areas of holography, optical processing, and image processing. He was the director of the Physics Division of the U.S. Army Research Office and is currently in the Physics Department of Duke University. Dr. Guenther is a fellow of the Optical Society of America and the American Physical Society.