EVALUATION OF AUTOMATIC TARGET RECOGNITION ALGORITHMS

Bir Bhanu

Ford Aerospace & Communications Corporation
Newport Beach, California 92660

ABSTRACT

In this paper we briefly review the techniques used to solve the automatic target recognition (ATR) problem. Emphasis is placed on the algorithmic and implementation approaches. The evaluation of ATR algorithms such as target detection, segmentation, feature evaluation and classification are discussed in detail and several new quantitative criteria are suggested. The evaluation approach is discussed and various problems encountered in the evaluation of algorithms are addressed. Strategies to be used in the data base design are outlined. New techniques such as the use of semantic and structural information, hierarchical reasoning in the classification and incorporation of multisensor in the ATR systems are also presented.

I. INTRODUCTION

One of the key components of the present and future defense weapon systems to be used in autonomous vehicle missions is the Automatic Target Recognition (ATR) system. Some examples are Low Altitude Navigation Targeting InfraRed for Night (LANTIRN) system, cruise missile and remotely piloted vehicle (RPV) applications such as the Acquila RPV [1,2]. The algorithmic components of an ATR system can be decomposed into preprocessing, detection, segmentation, classification, tracking, prioritization and aimpoint selection (Fig. 1). The goal is to effect these considerations in real time. The problem domain requires the tools of image processing, image analysis, pattern recognition and artificial intelligence. Research work in this area has been going on for the last 20 years, but it has been only recently that sophisticated algorithms and microprocessor, VLSI and VHSIC technology [2] have become available so that with the improvement in infrared and laser sensor technology it is now feasible to accomplish the objectives of an ATR system with success. However, the area of evaluation of ATR algorithms is in its infancy. Algorithms are tested on a very limited data base and very good classification performances have been reported. However, in practice these results have been only partially successful. Some of the key reasons for this are the nonrepeatibility of the target signature, experience with a very limited data base, obscured targets, very little use of available information present in the image such as context, structure, range etc. If use is made of these diverse sources of information, then it is expected that the target signature characteristics can be extracted reliably. In this presentation, assumption of an ATR system with a FLIR input image is made implicitly. The term autocuer is used interchangeably for an ATR system. In section II we review the techniques, which have been used to solve the target recognition problem with emphasis on the commonality in the approaches used at various levels in preprocessing, detection, segmentation, classification etc. Evaluation of ATR algorithms is dicussed in detail in section III. Several new quantitative criteria are discussed. Major problems in the evaluation of ATR algorithms and their possible solutions are also presented. Collection of the data base, its requirements and organization are addressed in section IV. Use of semantic, structural and statistical information in a hierarchical classifier in a state of the art multisensor system is presented in section V. Finally section VI presents the conclusions of the paper.

II. AUTOMATIC TARGET RECOGNITION ALGORITHMS

PREPROCESSING : This step is generally carried out with the objective of improving target contrast and reducing noise and clutter contrast present in the image. It is usually accomplished by a local filter such as a median filter. The techniques which are effective use high pass filter for edge crispening and locally variable scaling for contrast stretching. Lo [3] has compared the performance of variable threshold zonal filter, unsharp masking, PATS [4], histogram equalization, statistical difference operators and constant variance technique. He found that the first three filters give the best enhancement results on FLIR images.

TARGET DETECTION : It is the process of localizing those areas in the image where a potential target is likely to be present. In some techniques such as "superslice," [5] localization and segmentation are inseparable. Burton and Benning [6] and Schachter [7] present an evaluation of some target detection methods. Burton and Benning [6] and Politopoulos [8] use double window filter. This filter is based on the contrast between the target and its immediate background. It consists of two non-operlapping rectangle windows, assuming the inner window surrounds the target and the outér window contains background. Range is used to control the window sizes [6,9]. The metric used to determine the likelihood of a target being localized at a pixel is different in [6] and [8]. No statistical model is used in their experimental design. Schachter [7] uses a a simple statistical model and estimates the probability density function of the random variables designating target and background windows. Minor and Sklansky [10] use a spoke filter which is an extension of Hough circle detector. It examines locally edge magnitude and direction. It is an eight-spoke digital mask. It needs preprocessing which includes intensity normalization, dc notch filter, and edge detector. Bhanu et. al. [9] use intensity , edge and range information. Mitchell and Lutton [11] use intensity and

texture measures. Rubin and Tseng [12] use linear discriminant function of image local features to obtain interest points in the image. Target is assumed to be in the neighborhood of these points. Tisdale [13] uses gradient operators. Soland and Narendra [4] extract the "object intervals" along each scan line in the edge image. These lines are concatenated to obtain the outline of each object. In mode seeker technique [14] a pixel is iteratively replaced by the average of a selected set of its neighbors. The neighbors are chosen which belong to the same histogram peak as the given pixel. Upon termination two spiked histogram is obtained, one for the target and the other for background. The global extension of this method called "superspike" gives good results. Border following algorithms and pyramid approach do not perform well [7].
SEGMENTATION : Once a potential target is localized, it is extracted from the background as accurately as possible. "Superslice" algorithm [5] assumes that the objects are distinct from their surroundings by the presence of edges and different thresholds may be required to extract different objects in the same scene. First it computes a thinned edge image. It then uses several thresholds and finds the borders of the connected components in the image. The segmentation is obtained by selecting the threshold such that edge and border coincidence is maximized. Minor and Sklansky [10] use edge direction in addition to edge magnitude like "superslice." These techniques have problems when there are bimodal targets i.e., there is a significant intensity distribution across the target. Danker and Rosenfeld [15], Rosenfeld et. al. [16], and Bhanu et. al. [9,17] use two class (target and background) and three class (target, background and clutter) relaxation methods. These are iterative parallel schemes which make use of contextual information to remove any inconsistency and ambiguity in the labeling of pixels. Hartley et.al. [18] compare "superslice", pyramid approaches, relaxation and mode seekers. Chen and Yen [19] use Fisher's linear discriminant for the purpose of segmentation by pixel classification. Their theoretical error probability computation for the % of pixels misclassified compares favorably with the experimental resuls. Local statistics and edge information are commonly used.
FEATURES COMPUTATION, SELECTION AND CLASSIFICATION : After the segmentation, a set of features is computed for each object. The reliablity of these features is essential for target classification. Commonly used features are geometric, topological, projection, spectral and Hu's moments. Semantic features such as geographic, temporal context and environmental have been used only to a very limited extent. In addition to the specific restrictions imposed by the classification techniques (e.g. linear and quadratic classifiers), the desirable properties of the features are the invariance with respect to geometry (rotation, scale and translation), computational efficiency and extractibility. Evaluation of the robustness of features for classification with respect to different segmentation results has not been thoroughly investigated.

Normally feature selection is done to obtain computational efficiency and reduce memory requirements of the classifier. Non-extensive sequential feature selection techniques are not optimal. It is desired to do feature selection on uncorrelated features since statistical dependence or even conditional independence among the features may cause the best subset not to include individually the best feature. Feature selection in the context of the ATR problem has been done informally by histogram examination [5,20], Bhattacharyya measure [9], Kolmogorov-Smirnov test [20], exhaustive scheme [4], physical reasoning and linear regression techniques etc [21].

Classification has been mostly done by a K-nearest neighbor algorithm [4,9,12], using projections [11], linear and quadratic classifiers [5], tree based classifiers [4,5] or using clustering techniques. In the tree based designs first clutter is isolated from the target types and then a statistical technique is used for classification. Genearlly, in these studies there has been no statistical analysis of the performance and the data base was limited. Interframe analysis has been used to improve the classification performance.
PRIORITIZATION, TRACKING AND AIMPONT SELECTION : Prioritization is the procees of associating the priorities to the targets in the field-of-view. This information is prestored, which is normally based on the the type of the target and the probability of its correct classification. Once the targets are prioritized, they are handed off to a tracker. Noting that any detected scene change from frame to frame is potentially a target, both signal and symbolic based approaches have been used [22,23]. Since various trackers based on correlation, feature, intensity and contrast complement each other and minimize loss-of-lock, Dorrough et. al. [24] consider a multi-mode tracker consisting of an intensity centroid tracker, edge tracker and correlation tracker. The tracker has two modes: tracking and coasting. In the tracking mode target motion analysis is done and checks are made to determine if the target continues to be tracked. If it is not, the tracker enters the coasting mode and the target characteristics before the loss-of-lock are used to reacquire the target. Gilbert et. al. [25,26] use adaptive statistical clustering and projection based classification approaches [11] to identify and track objects. Aimpoint selection requires the determination of the critical aimpoint of a target. A stored feature vector corresponding to the target class and aspect (obtained from the previous classification step) is used for aimpoint designation. In scenarios involving a missile approaching the target, the maintennance of the aimpoint is important. It is carried out by using correlation, feature matching etc.

III. EVALUATION OF ATR ALGORITHMS

The work in the area of the evaluation of ATR algorithms is in its infancy. The assumptions that an image or data at the input of an ATR system should satisfy are often not stated explicitly and there has been a general absence of models. Furthermore, reliability and accuracy of the results of different techniques used to evaluate the autocuer algorithms depend upon the image data base. In order to be able to quantify the data base according to the type of images that it contains, measures of the image information content are needed. In the following first some image information content measures are presented and then the specific characteristics of FLIR images are discussed which form the basis of development of evaluation techniques for ATR system and help to identify representative sets of real test images for algorithm evaluation purposes.

Information Measures : A FLIR image can consist of tactical targets, background characteristics, noise and clutter. Target refers to an object which an ATR system must detect, recognize and classify. Clutter is defined as an object which resembles a target, but is not a target. Noise is characterized by a random intensity distribution over the image caused by the sensor and image formation process. Following are some quantitative information measures.

(1) Edge Points Measure : The edge points are characterized by high contrast. The potential target-like objects are usually present in the neighborhood of these points. Information content of an image can thus be measured by finding the points in an image at which the magnitude of an edge operator is greater than a certain threshold. Several measures are defined in terms of these edge points such as :

(a) Number of edges per unit area

(b) Information theoretic measure, I, present in the image

$$I = \text{Log}_2 \frac{1}{p} \text{ bits}$$

where, P is the probability of possible pictures made up of edge points. For example, if in an image of 500x600 pixels with 8 bits per pixel, there are 10% edge pixels, then the information content of the image is,

$$I = \text{Log}_2 256^{(3\text{X}10^5)} \text{ X } \frac{1}{10} = 2.45 \text{ X } 10^5 \text{ bits}$$

(c) Compute the probability p at an edge point pixel to be white $p_i(1)$ or black ($p_i(2) = 1-p_i(1)$), then obtain the entropy given as:

$$H = - \sum_i \sum_{\ell=1}^{2} p_i(\ell) \log p_i(\ell) \text{ , i} \varepsilon \text{ edge point}$$

H represents the degree of uncertainty present in the image.

(2) Uniformity and Structural Measures : These criteria measure the consistency of a pixel with respect to its neighbors and reveal the homogeneity of the regions. Also, they provide structural information about the image. One such measure is,

$$U = \sum_x \sum_y [f(x,y) - \overline{f}(x,y)]^2$$

where, f(x,y) is the gray level value at a pixel in the image and $\overline{f}(x,y)$ is the average gray level value in a 3x3 window centered about the pixel (x,y).

Other criteria which measure structural information are linear and cultural features present in the image. Measures such as co-occurrence matrices provide texture information and are useful as general scene information characterizing criteria. The number of target-like objects in an image which are not targets characterize the amount of clutter present in the image.

Characteristics of FLIR Images : In order to achieve optimum performance of an ATR system it is essential that its design make maximum possible use of the specific characteristics of FLIR images. For example, segmentation techniques should be based on models which take into account the real aspects of FLIR scenes so that the resulting information derived from these models be useful in predicting realistic preformance measures for the technique. It would be useful to distinguish the characteristics originating in the image formation process (sensor noise and signal transfer characteristics) and in the actual physical scene. Following are the characteristics and modelling approaches to FLIR images which are important to the general autocueing problem.

(1) The simplest model of a FLIR image is a two class, black and white intensity definition. Target is brighter or darker than the immediate background.

(2) In many situations the simple model of (1) may not be sufficient to extract the target if part of the target is hotter and part is colder than the background. This is a common occurrence in FLIR images. In such cases, the image model should make use of the context and some shape information about the targets (knowledge based system).

(3) When multiple targets are present in the image they may be found occluding each other. Thus, separation of individual targets may be difficult. The scene modelling should incorporate occlusion considerations [27].

(4) Size of the target is an important parameter in the ATR system design. As the range increases, the target occupies a reduced number of pixels in the image. If the number of pixels on the target becomes very small, the target may dissolve itself into the background. Thus, range information may be of crucial importance to the scene model.

(5) Targets may be obscured by or be partially hidden in smoke, dust and shadows. Again, the scene modelling should accordingly account for these effects.

(6) Often it happens that target boundaries are poorly defined and buried in the background. In such cases, the use of textural, structural and contextual scene information may be useful.

(7) Frame to frame analysis is important for scene modelling. The objective of such an analysis is to reduce the amount of computation and increase the classification accuracy by requiring repeated consistency of the classification decision.

The identification of the information measures and characteristics help in the following ways :

(a) Develop practical and realistic algorithms and techniques to evaluate the existing ATR systems,

(b) Analyze various components of an ATR system,

(c) Create a synthetic data base in the laboratory which incorporates the realism of the FLIR scene,

(d) Obtain a subset of images from the data base which can be used as a representative set in testing ATR systems and finally,

(e) Specify requirements for the data which help generate data bases of FLIR images which are increasingly representative of the real world.

Evaluation Approach : An ATR system should be evaluated on the basis of the task it is able to perform in a given operating environment keeping into consideration such factors as sensor type, resolution of data, type of objects, complexity and information content of the scene. The design of the system should be such that each of its components makes maximum use of the input data characteristics and its goals are in conformity with the end result of reducing the classification errors and false alarm. The classifier is not expected to perform adequately if the feature set on which it operates is poorly formed. The basic philosophy for the development of evaluation techniques is based on the fact that to obtain the optimum performance, it is essential to obtain the maximum attainable performance from each of the components of the ATR system. Thus it is necessary that each component has its own quantitative figures of merit against which it can be individually evaluated. However, since the ultimate goal of the ATR system is correct classification and the intermediate steps, such as, preprocessing, segmentation, feature selection, classification etc. are subservient to that goal and not an end in themselves, it is logical that each of the components must not only be evaluated with respect to its own figures of merit but also against its effect on the overall classification. Thus classification performance of the system is evaluated using an overall black box approach, together with the evaluation of each of the components and its effect on the classification results. For example detection and segmentation techniques are evaluated on the basis of how well they are able to locate and preserve the shape of the target. The assumptions inherently involved in the development of the detection and segmentation algorithms are to be checked against the input data and their effect on the output results. A segmentation technique based only on the concept of contrast will not be able to produce the target boundaries faithfully. This in turn provides valuable information to the algorithm developer guiding him in a reevaluation of his assumptions so that they are compatible with the realistic imagery and thus consistently good results can be obtained. This allows one to evaluate an ATR system efficiently, to identify better system components among various ATR systems, to understand them in depth and design better systems in the future.

Three parameters which give the overall performance of an ATR system are: probability of target detection, probability of classification and false alarm per frame. Probability of target detection (classification) is defined as the ratio of total number of targets correctly detected (classified) in the testing set divided by the total number of targets in the testing set. False alarm per frame is defined as the total number of false alarms divided by the total number of images in the testing set. It is to be noted that when using these definitions we need not restrict ourselves to the static image analysis. The target may have been detected using motion analysis and other contextual cues. For example Fig. 2 shows an air-to-ground image. The potential targets detected by using a double gate filter are pointed by arrows [8]. However, it did not detect the moving target near the top of the image. In Fig. 3 we show two consecutive frames of the same scenario shown in Fig. 2. Using optical flow field analysis [23], we are able to detect the moving target. In Fig. 3(c) the predicted and the detected true targets are shown in boxes. Fig. 4 still shows another example where a large number of targets have been detected [8]. However, the three true targets are along the "road side" as pointed by arrows. Thus making use of context, detection and classification performance can be improved. In the following, quantitative figures of merit are presented for preprocessing, detection, segmentation, features computation and selection, classification and tracking. These are based on the system concept. The effect of the system on the input is measured by its output.

Performance Criteria for ATR Components :

The preprocessing operations condition the incoming data stream to reduce the sensor and/or environment-dependent perturbations. Examples of preprocessing functions are noise suppression, dc restoration, focus control, histogram equalization, offset adjustments etc. The figure of merit depends upon the operation. As an example a median filter is supposed to maintain the sharpness of the edges. To evaluate this filter we apply an edge operator to the image and determine the % of the number of edges which are present in the thresholded input image and the thresholded output edge image as the figure of merit.

The purpose of the detection operation is to localize the area where a potential target is likely to be present. If a target is missed in the process of localization, then it will be missed altogether. Several performance measures are : (a) Probability of target detection, a target is said to have been detected if its centroid lies within a small window (its dimensions are function of the true target size) centered at the centroid of the true target. (b) The ratio of the localized area to the image area, it measures the computational efficiency of the algorithm. (c) The number of localized areas to the actual number of targets present in the image. This will affect the false alarm rate.

The objective of segmentation is to extract the target from the background after it has been detected. Segmentation techniques are to be evaluated on the basis of how well the implicit or explicit model in the technique is able to predict the performance. Some quantitative measures are: (a) Number of target pixels misclassified with respect to the true target, (b) Correlation coefficient between the true and extracted target, (c) Mean square error between the true and extracted target, (d) Object-to-background contrast and intensity difference, Bhattacharyya distance between the true target and clutter objects, (e) Shape measure to estimate the shape difference between the true and extracted targets by using shape numbers [28].

Evaluation of features and their clustering is important in the design of ATR algorithms. Computational efficiency and accuracy of the features and their discriminatory power to distinguish different targets are also important, since the classification results depend upon the accuracy and reliability of the features. For example, Hu's moments which are invariant with respect to size, position and orientation have commonly been used. However, they are not contrast invariant and the contrast change of an image introduces a nonlinear scaling effect. Resolution is affected by scale, so that the invariant moments are no longer strictly invariant under rotation and scale changes. To examine the segmentation effects, the features of the segmented target and the true target are compared by measuring the distance between them and evaluating different features by carrying out a variance analysis. To evaluate the clustering of features a clustering fidelity criterion such as, $\beta = Tr(S_b) \cdot Tr(S_w)$, where S_b and S_w are between-cluster and within-cluster scattering matrices, is used. The behavior of β is such that it passes through a maximum at the intrinsic number of clusters and at the maximum the ratio of $Tr(S_b)/Tr(s_w)$ is exactly 1. The maximum of β can be determined by incrementing the number of clusters until a decrease is detected. This allows one to obtain the number of clusters inherent in the data. Departure of the number of clusters thus obtained from the known number of clusters tells about the quality of the features.

It is important to determine if clustering is really present or it is a statistical artifact because a clustering method always finds clusters, whether or not they are real. Furthermore, it may impose a particular structure on the data rather than find the actual structure present. For example, for the patterns lying along two long parallel lines, a mean-squared algorithm will probably cut the lines rather than group the patterns on each line. Different techniques are likely to give different solutions unless the data are very clearly structured. Following are some methods that allow one to decide if the clusters are really present in the data.

1) Use several clustering techniques (such as partition and hierarchical approaches) based on different assumptions on the same set of data. This is necessary since no one technique can be judged to be the best in all situations. Only the clusters produced by all or majority of methods are accepted. If all the techniques produce very similar results, then it would justify the presence of some structure in the data. The presence of different solutions suggests there is no clear cut cluster structure.

2) Partition the data randomly into two sets and cluster each half independently using a clustering technique. Cluster assignment in the partitioned samples should be similar to that of the entire sample, if the clusters are stable and not an artifact of the sample set.

3) Test the randomness of the data. Investigate the tendency of the random subsets of the data set to cluster. Measure the number of clustered points. If the data has some inherent structure, then it is expected that the number of random points will be much smaller than the number of clustered points. The occurrence of stable clusters tells about the structure present in the data.

The performance of a classifier is measured by finding the probability of classification and false alarm rate. It is intimately related with the feature set. It is also a function of the data base and its size. The data base is divided into a training set and an independent testing set. There are several approaches to evaluate how well a classifier is trained. Two such approaches are:

(1) Partition the training set into two disjoint sets. Obtain the decision rule from the training set; estimate its accuracy using the test set. Now the training and test set are interchanged and the accuracy is estimated. The discrepency between the error rates obtained in the two cases is an estimate of the training of the classifier.

(2) In the second approach called the "leaving-one-out," one sample is removed from the K total samples and the remaining samples are used to derive a decision rule and test it on an isolated sample. The average of K error rates provide an estimate of how well a classifier is trained [9].

Reliability of a classifier is a function of the clustering quality of the features. It is measured by finding the stability of the feature clustering. Stability is obtained by finding the difference of the sum of the squares of the distances between the reference cluster centers and the cluster centers actually obtained in the "leaving-one-out" technique. The efficiency of the K-nearest neighbor algorithm is also an important consideration in the evaluation of the classifier. Branch and bound and k-d tree approaches have been used.

In practice an interframe analysis is carried out to improve the performance of an ATR system. Here a tracker is normally used. The basic performance measures of of a tracker are centroid drift, jitter, probability of loss-of-lock and reacquisition. Other measures are the ability of the tracker to perform adequately in the presence of platform motion and target obscuration.

IV. DATA COLLECTION, ITS REQUIREMENTS AND ORGANIZATION OF DATA BASE

One of the key requirements in the development of algorithms for ATR system is the availability of the data base on which algorithms are to be evaluted. The interaction of algorithms and data base is shown in Fig. 5. Note that models for algorithm evaluation drive the requirements for data base, its collection and organization. It is generally assumed that most of the information about the data environment, for which a classifier is supposed to be a part of, is contained in the labeled samples (for other approches see [29]). As a result the performance of the classifier depends upon the feature set of the training data set. The classification performance is measured by applying the classification algorithm on the labeled samples and noting the results. Therefore, the training data should be fairly representative of the real world-scenario in which an ATR system is to be operated. Since it is impossible to obtain images exhibiting all scenarios and conditions which may exist in real life situations, the training data should be generated using real images which exhibit the typical characteristics of the FLIR scenes as have been outlined in section III. Synthetic data is initially useful in the design and evaluation of the individual components of the system. The intermediate data is estimated using parametric image models as shown in Fig. 6. However, the classifier performance should be ultimately evaluated on the real data not on the synthetically generated data. Also the number of samples in different classes in the training set should be in proportion to their most likely outcome in practice.

Collection of the data base should be such that it meets the requirements for which this data is to be used. It is useful to associate a header with each image and record the data in some standard format. Data should provide the basic ground truth information for each image. Such information includes file name, bits/pixel, time and date, scene location, number of targets, locatiion of targets, image field of view, target size, range, depression angle, temperature, humidity, sensor height, sensor type, atmospheric conditions, background information etc.

The requirements of the data base are tied to the specific scenario in which the cuer is supposed to be operated. It should include the images exhibiting the varying characteristics as outlined above together with operational requirements such as terrain conditions, different aspects, climatic and atmospheric conditions, day and night conditions, false targets, decoys, cold targets, burning objects, partially occluded targets, targets in groups, targets on roads etc. It should also include imagery exhibiting the variety of likely clutter to be present.

Organization of the data can be done parametrically or scenariowise or a combination of both. Parametrically specification can be made as to the type and size of the target, certain aspects, specific bandwidth of range, depression angle, height, day/night conditions etc. With respect to scenarios it may be interesting to see examples of closing in or poping up type sequence of images. A combination of these will allow us to select any desired set of images.

The size of the data base is a function of specific cuer scenario requirements. It should have sufficient number of samples for each aspect, class and scenario (10 to 200) to train a classifier so as to obtain statistically meaningful results. Theoretical bounds as a guide to the amount of data for a given performance of a recognizer using models are obtained using inequalities of probability theory.

V. MULTISENSOR TARGET RECOGNITION SYSTEM

In this section we present an example of the state of the art ATR system [24]. The schematic diagram of this multisensor ATR system is shown in Fig. 7 . It uses semantic, structural and statistical information in a hierarchical manner. FLIR (8-12 micron), LADAR (10.6 micron), and millimeter wave (3.2 mm) sensors operate in a synergistic manner to obtain the best performance, even in the presence of adverse conditions. The sensed image is preprocessed to enhance the quality of the image or restrict the area of the image which is to be searched. Then the targets are localized and segmented from the background. Their features are computed and preliminary information regarding their classification is obtained. At the same time, interframe analysis and tracking are carried out which not only help to reduce the amount of computation involved in the detection and segmentation of the targets in subsequent frames, but also improve the classification results. The slant range (distance of the target from the center of the field of view) information is used during target detection and in improving the classification results. This in turn aids in the aimpoint selection process. When the FLIR and LADAR sensors are not operational, a millimeter wave sensor is used to determine the relevant information regarding the target. This information is then used for aimpoint selection.

Using the target's intensity, texture and slant range an intelligent preprocessing step is carried out. This step reduces the image area to be searched to detect the target by over 80%, and thus it greatly reduces the amount of computations for initial target detection [9]. This percentage can be raised by use of the data from the other sensors. An efficient relaxation scheme is used to extract the target from the background while preserving the gross boundary of the target. The technique allows control over segmentation through the use of 3 parameters. It has been thoroughly evaluated with respect to size of the target, contrast and signal-to-noise ratio [30]. As an example Fig. 8 shows the segmentation of a ship. Note that there is a significant intensity variation across the targets. Part of the target image is whiter and part darker (target is partly hot and partly cold).

A number of shape, geometric, moment and gray scale features are computed after the signature of the target has been extracted from the background. Features which are important in discriminating between different types of target are automatically selected. Feature selection is carried out by using the Bhattacharrya distance, K-means algorithm and the discrete K-L transform. The selected feature set has been found to be more robust and comprehensive than the techniques based only on features such as moments [9]. These features can be correlated with those obtained from other sensors. The classification scheme consists of a number of linear or quadratic classifiers with ancillary clustering techniques. These classifiers are designed to work on a reduced set of features after a feature selection process. They also allow the design of tree classifier such as [31], where at each nonterminal node a K-means algorithm and Bhattacharrya measure are used. A classifier giving the least error rate is selected. The classifier controls the detection and segmentation of the target image. The K-nearest neighbor algorithm is implemented using an efficient tree approach.

The final assignment of a target to one of several classes is done using the classification results from two sources of information. (a) Classification of a target using the classification results on several frames of the input data by an interframe analysis.(b) Classification of a target using the range image from the LADAR sensor. The target is finally classified by using a simple logic whose inputs are (a) and (b). In order to improve significantly the performance of the target recognition system, it is essential to utilize both multi-frame information and a multi-mode tracker with look-ahead capability. The multi-mode tracker allows for the use of centroid, correlation and target feature tracking for maintaining target track.

Input to the aimpoint selection block comes from the millimeter wave sensor, semantic information and interframe analysis blocks. These provide classification and aspect information. Target prioritization, mission and guidance information is stored as a feature vector for each target class with several aspects per class. The aspect data obtained from the sensors is used to appropriately select a feature vector from the set of prestored information on that class of target. From this, an aimpoint is computed on the actual target. Normally, moments are used to select the predefined aim points of an acquired target.

The semantic information block has several functions. It takes the input from the LADAR sensor and generates information regarding the type and aspect of the target. This helps to reduce aimpoint selection time and improves target classification accuracy. It takes its input from the interframe analysis and multimode tracking block. It also allows the implementation and processing of any specific information regarding target type and mission scenario. For example, the assumed scenario may imply the need for processing only moving target information relative to several other classes. When the range is small (100-500) meters there are two options for target classification using FLIR data. Either the FLIR image can be minified with respect to some standard range so that inside structure of the target is not visible in the image or structural information present in the image can be used. A simple example of structure utilization is shown in Fig. 9. Image in Fig. 9(a) has two targets, a tank and a truck. Fig. 9(b) shows the results of a zero crossing edge operator and Fig. 9(c) shows the zero corssing details of the tank target. Importatnt parameters in the classification are size of the wheels and their relative positions. Circles corresponding to the wheels are obtained using least squares, RANSAC paradigm or Hough transform. The classification result with confidence measure is passed to the following classification block in hierarchy. Structural information becomes very useful for missiles with fiber optics, where all the computing is done at the ground base.

VI. CONCLUSIONS

In this paper an overview of the automatic target recognition problem is presented. The major problems in the evaluation of ATR algorithms are discussed and some solutions are suggested. The evaluation approaches are discussed in detail and several quantitative performance criteria are presented. Information measures of the image and some typical characteristics of FLIR images allow one to evaluate the scene as to the difficulty in detecting targets and set the requirements for the collection and organizaton of data base. As the sensor and VHSIC technology is progressing and data base collection efforts are undertaken, algorithms are being improved by utilizing semantic and structural information and hierarchical reasoning in a multi-sensor system. This will help to obtain better performance of an ATR system. Other available knowledge and techniques which should be incorporated in the algorithm design are the use of context and integration of information from diverse sources such as moving and stationary targets, location of grounds targets and horizon, targets moving in group, efficient focussing mechanism based on hypotheses testing and competition and cooperation among these hypotheses for locating the targets from frame to frame, use of map information, efficient search strategies for the storage and retrieval of the data in classification, mission specific information such as the type of targets, composite classifiers [32] and the classifier designs for different learning environment scenarios from supervised to vicissitudinous. The multi-sensor system presented here is an attempt to make use of some of these concepts and techniques. Decision theoretic multiattribute approaches are helpful in the development of evaluation methodologies. It is anticipated that a single algorithm may not be able to provide optimum performance under all scenarios. It is desired to use different algorithm designs implemented such that they are compatible with the ATR system in which they are to be used. Depending upon the requirements the desired design can be selected to achieve maximum performance. All this leads to the development of recognition expert systems with flexible control strategies.

REFERENCES

[1] W.B. Scott, "Lantirn System Flight Testing Begins," Aviation Week and Space Technology, September 20, 1982, pp. 117-1[2625.

[2] D.F.Barbe, "VHSIC Systems and Technology," IEEE Computer, Feb. 1981, pp. 13-22.

[3] C.M. Lo, "Forward Looking Infrared (FLIR) Image Enhancement for the Automatic Target Cuer System," Image Processing for Missile Guidance, Proc. SPIE 238, pp. 91-102.

[4] D.E. Soland and P.M. Narendra, "Prototype Automatic Target Screener," Smart Sensors, Proc. SPIE 178, pp. 175-184, April 1979.

[5] D.L. Milgram and A. Rosenfeld, "Algorithms and Hardware Technology for Image Recognition," Final Report to U.S. Army Nignt Vision & Electro-Optics Lab., Fort Belvoir, Virginia, March 31, 1978.

[6] M. Burton and C. Benning, "Comparison of Imaging Infrared Detection Algorithms," Infrared Technology for Target Detection and Classification, Proc. SPIE 302, pp. 26-32.

[7] B.J. Schachter, "A survey and Evaluation of FLIR Target Dectection/Segemntation Algorithms," Proc. DARPA Image Undestanding Workshop, September 1982.

[8] A.S. Politopoulos, "An Algorithm for the Extraction of Target-Like Objects in Cluttered FLIR Imagery," IEEE Aerospace & Electronic Sys. Soc. Newsletter, Nov. 1980, pp. 23-37.

[9] B. Bhanu, A.S. Politopoulos and B.A. Parvin, " Intelligent Autocueing of Tactical Targets in FLIR Images," To be presented at the IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition, 1983.

[10] L.G. Minor and J. Sklansky, "Detection and Segmentation of Blobs in Infrared Images, IEEE Trans. Systems, Man and Cyber., March 1981, pp. 216-232.

[11] O.R. Mitchell and S.M. Lutton, "Segmentation and Classification of Targets in FLIR Images," Image Understanding Systems and Industrial Applications, Proc. SPIE Vol. 155, pp. 83-90, 1978.

[12] L.M. Rubin and D.Y. Tseng, "Automatic Target Cueing," Proc. 4th Int. Joint Conference on Pattern

Recognition , Nov. 1978, pp. 774-779.
[13] G.E. Tisdale, "The AUTO-Q Digital Image Processor for Target Acquisition, tracking, and Handoff," Paper presented at IRIS'80, Seattle, Washington, May 1980.
[14] K.A. Narayanan and A. Rosenfeld, "Image Smoothing by Local Use of Global Information," IEEE Trans. Sys., Man and Cyber., Vol. 11, 1981, pp. 826-831.
[15] A.J. Danker and A.Rosenfeld, "Blob Detection by Relaxation, IEEE Trans. Systems, Man and Cyber., Jan 1981, pp. 79-92.
[16] A.Rosenfeld et al., "Blob Extraction by Relaxation," Proc. DARPA Image Understanding Workshop, April 1979.
[17] B. Bhanu and O.D. Faugeras, "Segmentation of Images Having Unimodal Distributions," IEEE Trans. on Pattern Analysis and Machine Intelligence, July 1982, pp. 408-419.
[18] R.L. Hartley, L.J. Kitchen, C.Y. Wang and A. Rosenfeld , "Segmentation of FLIR Images: A Comparative Study," Proc. DARPA Image Undestanding Workshop, September 1982.
[19] C.H. Chen and C. Yen, "Object Isolation in FLIR Images Using Fisher's Linear Discriminant," Pattern Recognition, Vol. 15, no. 3, pp. 153-159, 1982.
[20] E.M. Rounds, T. King and D. Steffy, "Feature Extraction from Forward Looking Infrared (FLIR) Imagery," Image Processing for Missile Guidance, Proc. SPIE 238, pp. 126-135.
[21] F.W. Smith and M.H. Wright, "Automatic Ship Photo Interpretation by the Method of Moments," IEEE Trans. on Computers, September 1971, pp. 1089-1095.
[22] I.L. Ayala et.al., "Moving Target Tracking Using Symbolic Registration," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-4, Sept. 1982, pp. 515-520.
[23] R.D. Holben, "An MTI (moving target identification) algorithm for Passive Sensors," Proc. IEEE NAECON, 1980, pp. 114-121.
[24] D.C. Dorrough, V.F. Pizzurro, B. Bhanu and J.R. Pasek, "A Multi-Sensor Multi-Mode Tracker Approach to Missile Ship Targeting," Proc. Tri-Service Workshop on Missile Ship Targeting, Naval Research Laboratory, Washington, DC, August 10-12, 1982.
[25] A.L. Gilbert, "Video Data Conversion and Real-Time Tracking," IEEE Computer, August 1981, pp. 50-56.
[26] A.L. Gilbert et. al., "A Real-Time Video Tracking System," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 2, Jan. 1980, pp. 47-56.
[27] B. Bhanu, " Recognition of Occluded Objects," To be presented at the 8th Int. Joint Conf. on Artificial Intelligence., Karlsrhue, West Germany, Aug. 8-12, 1983.
[28] E. Bribiesca and A. Guzman, "How to Describe Pure Form and How to Measure Differences in Shapes Using Shape Numbers," Proc. IEEE Conf. on Pattern Recognition and Image Processing, 1979, pp. 427-436.
[29] B.V. Dasarathy, "Information Processing for Target Recognition from Autonomous Vehicles," Electro-Optical Technology for Autonomous Vehicles, Proc. SPIE, Vol. 219, pp. 86-93, 1980.
[30] B. Parvin and B. Bhanu, "Segmentation of Images Using a Relaxation Technique,"To be presented at the IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition, 1983.
[31] Y.K. Lin and K.S. Fu, "Automatic Classification of Cervical Cells Using a Binary Tree Classifier," Pattern Recognition, Vol. 16, No. 1, pp. 69-80, 1983.
[32] B.V. Dasarathy and B.V. Sheela, "A composite Classifier Design: Concepts and Methodology," Proc. IEEE, Vol. 67, May 1979, pp. 708-713.

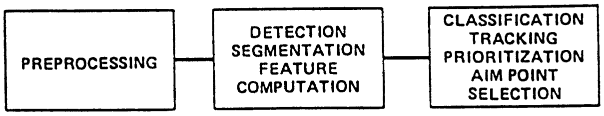
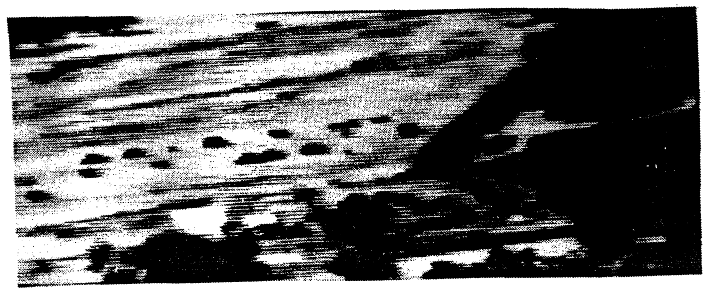


Fig. 1. Basic block diagram of a typical ATR System

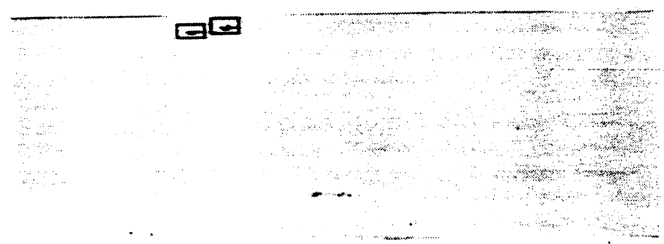
Fig. 2. Detection of targets using a double gate filter. The moving target near the top is not detected.

Fig. 4. Use of contextual information - along the "road side" to detect and classify targets



(a) Frame 1

(b) Frame 2



(c) Predicted and detected targets are shown in boxes

Fig. 3. Use of optical flow field analysis for detecting a moving target near the top of the images

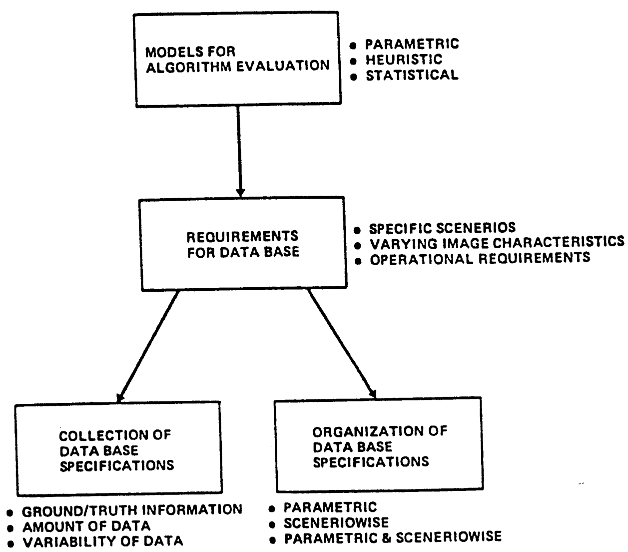


Fig. 5. The interaction of algorithms and data base

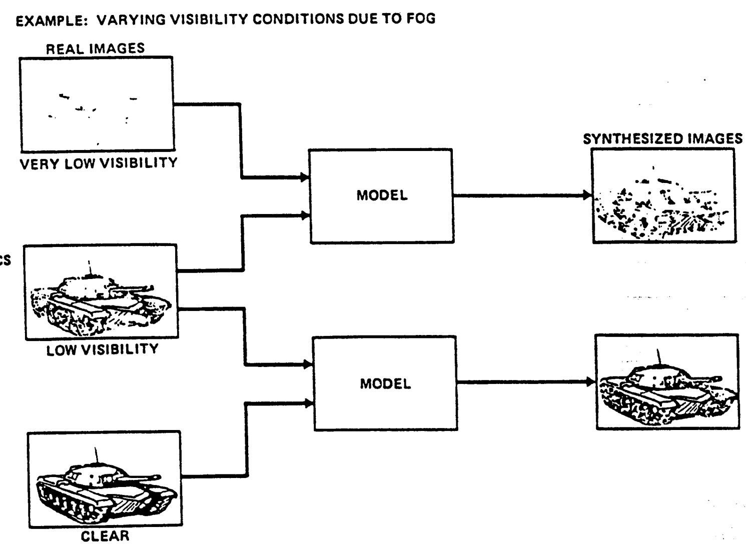


Fig. 6. Estimation of intermediate data using parametric image modelling

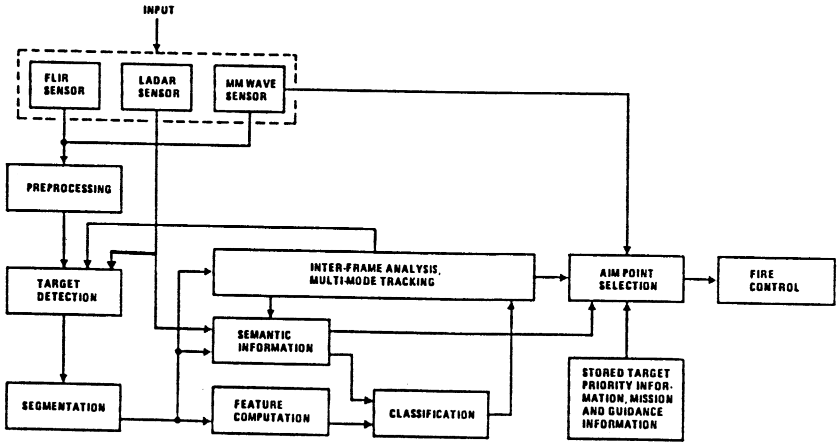
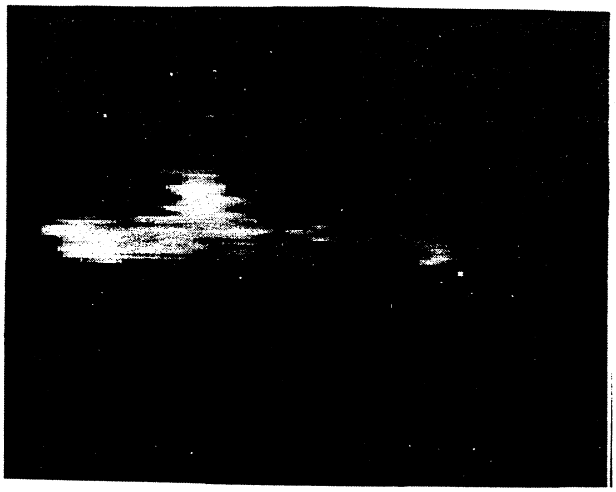
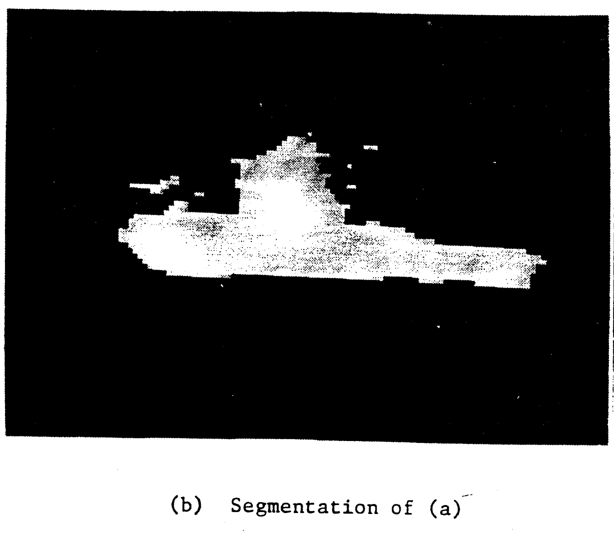


Fig. 7. The schematic diagram of the multi-sensor ATR system



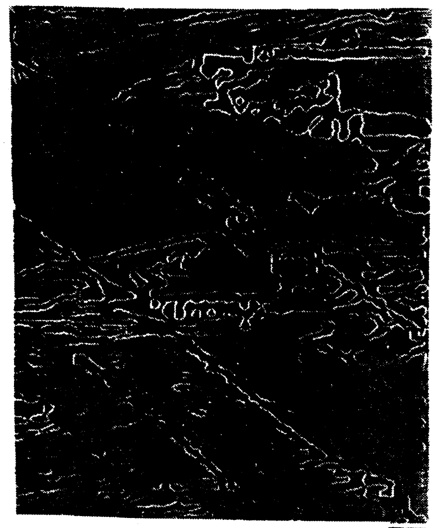
(a) FLIR image of a ship target



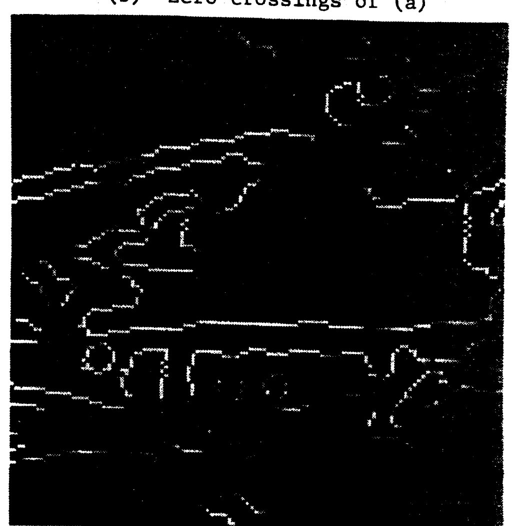
(b) Segmentation of (a)

Fig. 8. Segmentation of a ship target using relaxation

(a) FLIR Image



(b) Zero crossings of (a)



(c) Zero crossings of the tank target

Fig. 9. Use of structural information

BIR BHANU

PROCEEDINGS

Of SPIE-The International Society for Optical Engineering

Volume 435

Architecture and Algorithms for Digital Image Processing

André Oosterlinck, Per-Erik Danielsson
Chairmen/Editors

August 25-26, 1983
San Diego, California

BIR BHANU

Proceedings of SPIE—The International Society for Optical Engineering

Volume 435

Architecture and Algorithms for Digital Image Processing

André Oosterlinck, Per-Erik Danielsson
Chairmen/Editors

August 25-26, 1983
San Diego, California

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Telex 46-7053

SPIE (The Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to advancing engineering and scientific applications of optical, electro-optical, and photo-electronic instrumentation, systems, and technology.