92-33

TACTICAL WEAPON
GACIAC
GUIDANCE & CONTROL
INFORMATION ANALYSIS CENTER

GACIAC PR 92-01

PROCEEDINGS OF THE SECOND AUTOMATIC TARGET RECOGNIZER SYSTEMS AND TECHNOLOGY CONFERENCE

Volume I - Unclassified

17-18 March 1992

Conducted at:

Center for Night Vision and Electro-Optics
Ft. Belvoir, Virginia

Sponsored by:

Defense Advanced Research Projects Agency (DARPA)
DoD Working Group on Automatic Target Recognition (DoD WGATR)
Automatic Target Recognizer Working Group (ATRWG)
Joint Services Guidance and Control Committee (JSGCC)

Published by GACIAC
IIT Research Institute
10 West 35th Street
Chicago, Illinois 60616-3799

Presented at the 2nd Automatic Target Recognizer Systems and Technology Conference
17-18 March 1992 GACIAC PR 92-01

MODEL BASED ATR: ALGORITHMS BASED ON REDUCED TARGET MODELS, LEARNING AND PROBING[1]

13 February 1992

John S. Baras[2] and David C. MacEnany
AIMS, Inc.
6159 Executive Blvd.
Rockville, MD 20852

ABSTRACT

Two new model based ATR algorithms are presented. The algorithms employ economic target models, recent mathematical developments in the extraction and reduction of target silhouettes from noisy images, and on-line probing for clutter reduction and target classification. The target models used employ a small, carefully selected set of characteristic viewpoints and local features. The design of the algorithms are described and their performance evaluated using synthetic FLIR data. The two algorithms are: the Probing Algorithm and the 4-Way Distance Algorithm. Both algorithms have an early phase which is primarily statistical, and a late phase which is primarily geometric, although the two phases are tightly coupled, with frequent revisitation of the image data used to increase decision confidence. In the Probing Algorithm novel on-line probing is performed using stored target geometric models to efficiently implement a generalized maximum likelihood principle. The Probing Algorithm self-adapts to find the most appropriate target location, range and viewpoint geometry and then performs decision-making based on selected maximum likelihood estimates and on-line computed, optimal probing sets. The 4-Way Distance Algorithm performs decision making based on a classification tree logic utilizing matching scores computed on four components of the target silhouette.

1.0 INTRODUCTION

Most traditional ATR methods are based on pattern matching techniques utilizing "global features" such as: area, perimeter, moments of inertia, Fourier coefficients of silhouettes, etc. These methods are *global parameter matching* approaches to ATR. The generic idea is to compute a vector of global "features" from each candidate target model, and from the sensory data, and compare them in order to find the best matching target model. Template matching and correlation are such methods as well. The fundamental difficulties with ATR methods based on global features are well known: non-robustness under varying background conditions, deterioration of performance with clutter and occlusion or other kinds of "noise" and spurious data. Our approach utilizes *model-based techniques* employing *local features* and their interrelationships. By concentrating on local features, we can establish appropriate correspondences between the sensory data and the target model, even in the presence of *occlusion*, or in the presence of *spurious data*.

Our approach to model-based ATR algorithms differs significantly from traditional methods. Our methodology employs systematic techniques for constructing economic model hierarchies from CAD, synthetic, or field imagery databases (see [6] in this proceedings). Our algorithms are of the hypothesize-and-test type. Detection and classification are addressed together and linked via feedback, which generates iterations to revisit image sensor data. We emphasize local features including feature selection, ranking, matching and probing. We utilize both statistical and geometric features. We consider in particular both the detectability of a local feature by a sensor (i.e., its saliency) as well as the stability of features with respect to variations in range and viewing angle. We also employ learning

[1] Work partially supported by contract No. DAAB07-90-C-F425 from U.S. Army CECOM, through the C²NVEO. COTR: Ms. T. Kipp C²NVEO.

[2] This author is also affiliated with the Electrical Engineering Department and the Systems Research Center, University of Maryland, College Park, MD 20742

algorithms for combining evidence and for off-line design of ATR logic. The economical target models developed provide the basis for a uniform data representation for many different sensors including FLIR, LADAR, VISION and mm-Wave RADAR. Similarly, the algorithms developed here apply equally well to sensors other than FLIR and hold great promise for multi-sensor fusion in model-based ATR.

All portions of both algorithms have been developed with maximum parallelism in mind. Considering the current satisfactory execution time of both algorithms on standard sequential machines, such as SUN Sparcstations, their implementation on multiprocessors will permit extremely efficient real-time operation. By-products of this development effort include sophisticated software tools and utilities for learning clustering algorithm design, decision tree design, target model reduction, probing design and ATR algorithm design, evaluation and parameter tuning.

In developing our ATR algorithms we benefited from careful review of results in perceptual research [12]. The fundamental ATR problem considered in our work is that of recognizing a target from a collection of targets in a single two-dimensional image. The sensor viewpoint is not known, range is provided but is only a rough estimate to the center field-of-view, the contrast can be quite varied and target-like clutter exists in the scene. Furthermore we are interested in algorithms that work well even when parts of the target are occluded. Since humans can perform these functions efficiently and in real-time, it stands to reason that results from perceptual research (psychophysics) can provide powerful hints and suggestions for the developments and organization of such algorithms. For example, recent research has demonstrated [12] that visual recognition can commonly be achieved from a two-dimensional image without any preliminary reconstruction of depth information or surface orientation from the visual input. This is achieved by the process of *perceptual organization*, in which relations are found directly among the two-dimensional features of an image. Perceptual organization refers to the basic capability of the human visual system to derive relevant groupings and structures from an image without prior knowledge of its contents. For example, people can immediately detect symmetry, clustering, collinearity, parallelism, connectivity and repetitive textures.

Another important hint from perceptual research is the use of hypothesize-and-test strategies, including the accumulation of evidence which is consistent with several interpretations; each with different degrees of confidence. In addition, line-drawings are important for recognition, even when coarse and approximate. Furthermore, perceptual results emphasize the significance of examining curves at different resolutions. Apparently humans pay particular attention to properties of curves which remain stable across different "resolutions." The simple polygonal approximations of target silhouettes across different resolutions (scales) that we use in our algorithms and models are very much in the spirit of these ideas [6].

Pursuing the latter point a little further, quite often we need to determine boundary or internal edges based on points that we have found in the sensor data. Perceptual research again suggests that the notions of continuity (smoothness) across several resolutions, as well as characteristic curvature points are important for recognition. The weak continuity methods employed in our algorithms in various forms (fast and inexpensive, more detailed, multi-resolution) are in complete agreement with these findings.

One of the most powerful and general sources of information constraining the sensor image data arises from properties of the image projection process which maps a three-dimensional scene into a two-dimensional image. Under the very reasonable assumption that the viewpoint of the sensor is independent of the objects in the scene, then one can show [12] that only certain classes of image relations are likely to occur more often than by chance. These classes of relations are those that remain stable over a range of viewpoints. By extrapolation to FLIR-based ATR systems, these ideas permit us to exploit the high degree of redundancy in the sensor data, since only a few matches are required to determine the viewpoint and we can then make accurate predictions for the locations in the image of other local target features. We also exploit spatial correspondences between data and target features. In particular, we emphasize testing correspondences to decide if they are *consistent with the projection of the target from a single viewpoint*. Since the set of feature correspondences is typically greatly overdetermined, it is possible to make reliable decisions even in the presence of many missing features or occlusion.

For example, parallelism and the presence of equal spacings between a series of collinear features are properties that are preserved over all viewpoints except where perspective effects are significant. Since many targets occupy only small visual angles, such relations provide efficient probes (tests) for ATR, in the sense that they can be used to identify reasonable areas of interest for further processing. Perceptual ideas lead to grouping of local data features based on the so-called *non-accidentalness* of the grouping. For example the *oriented contrast probe* [4], developed and used by us and briefly described in this report, is partly motivated by the perceptual consideration that it is

highly unlikely that a pair of parallel edges of some strength are due to random noise or clutter; most likely they are due to a man-made object or a target.

Hoffman [10] and others [1], [14], have emphasized the maxima, zeros, or minima of the curvature for the purposes of partitioning the silhouette curve into salient parts. The most frequently cited source of evidence for the saliency of these features has been the work of Attneave [2]. This has led to statements that maxima of curvature are the most perceptually significant features of curves. However, as shown in [12], even if we chose the points that are as far as possible removed from the points of discontinuity chosen by Attneave (i.e. choose half points between the original set of points), the drawing still remains easily recognizable. As a matter of fact a more correct point of view is to consider the stability (invariance) of these points of three dimensional curves under projection onto the image plane from different viewpoints. Curvature maxima and minima do not have these invariance properties, but curvature inflection points, discontinuities in tangent, zero curvature points and curve terminations do remain stable [11]. These considerations suggest that curvature inflection points and ends of straight segments are the best choices for segmenting a three-dimensional curve: the segments can be used as the corresponding *local features*. Therefore the *recommended local features are: convexities, concavities and straight lines.* These are precisely the target silhouette features we use in our algorithms.

An important difficulty in ATR algorithm development is the great variability of viewpoint. The commonly held belief is that since spatial information in the image is highly dependent upon viewpoint, any strong predictions about the spatial appearance of a target are likely to apply to only a relatively small subrange of the possible viewpoints. Therefore, the argument goes, a complete search must enumerate over the various discrete ranges of viewpoints that need to be considered. The seemingly large size of this search space has been a major factor leading researchers to avoid searching over the range of viewpoints and instead direct their efforts into the derivation of three-dimensional structure from the image. We have shown [4] that the set of target silhouettes taken from every two degrees, over the entire range of 0 to 180 degrees, naturally clusters into a small number of equivalent classes. The fact that the sensor data are noisy further reinforces the argument. We have developed algorithms that automatically compute these clusterings. This reduces dramatically the search space and it is used in an essential way both in our algorithms for off-line model construction and in the ATR algorithms themselves.

In selecting local features we have considered the following characteristics: geometric characteristics of silhouettes, local curvature maxima and minima, curvature inflection points, zero curvature points and segments, neighborhoods around curvature extrema, proximity to other features, local contrast statistics, texture information, "sensor visibility," target fingerprints,[3] portions of polygonal approximations to boundaries and related symbolic representations and "information content" of a local feature.

Good ranking of local features can reduce substantially the complexity of an on-line ATR algorithm, its time performance, and increase its recognition performance. Such rankings can be used to provide feature indexing, which has proved to be very helpful [9] with large object libraries. The terms *significance* or *saliency* of a feature are used to provide a notion of ranking for features. Methods for ranking local features that we have experimented with include: invariance with respect to resolution (scale), invariance with respect to viewpoint, location and range, detectability in noisy data and target competitive clutter, discrimination power (among several targets) and computational complexity.

Such methods can be further extended to provide extremely efficient indexing methods of local features or groups of features. These techniques hold great promise for extending our algorithms to large libraries of target models, without commensurate increase of the computational complexity.

2.0 ALGORITHMIC CONSTRUCTION OF ASPECT GRAPHS

As discussed in the introduction, we have successfully addressed the problem of reducing the target model representations with respect to viewpoint variations. Our method first determines viewpoint equivalence classes according to a distortion measure and then determines the significant silhouette features that remain invariant in the viewpoint ranges determined by the equivalence classes. In this *iterative process*, refinements can result, causing further subdivision of a class. The algorithms can work across multiple resolutions because our target silhouettes are reduced to polygons over various resolutions. Furthermore, we can compare characteristic silhouettes from each equivalence

[3] Target fingerprints, or scale-space diagrams [15], are diagrams of local features *vs.* scale, *e.g.*, the curvature primal sketch in [1]; see also [6].

class of each target to test if the resulting cluster has acceptable classification performance. Consistent with established terminology in computer vision [5] we call the resulting viewpoint equivalence classes *aspects*. If we arrange the aspects as nodes in a graph we can connect them with links to obtain what is called the *aspect graph* of the target. The links denote the appearances of new features, or the disappearances of features relating the various aspects. Thus we have constructed a semantic net (or relational graph) representation of the viewpoint variation of a target silhouette. This precompiled object is used to guide our ATR algorithm *on-line*. This is the first time that a *systematic, quantitative* method has been developed to reduce viewpoint information so dramatically.

The notion of aspect allows us to group together sets of target features which can be "viewed" (sensed) from a single viewpoint. This in turn allows us to group together viewpoints which observe the same aspect. Further, we want to emphasize that this representation can also be viewed as a *quantization of the space of sensing operations*. Indeed this representation allows us to easily determine the target features which can be observed by a particular sensor from a particular viewpoint relative to the target. This is done by determining which aspect will be viewed by the viewpoint, and then looking up the target features which are associated with that aspect. This representation is a key element in our approach.

In our system we characterize aspects in terms of observable features. As we change viewpoint, specific features appear and disappear from the view of the sensor. We call such appearances and disappearances of features *events*. The aspect graph is a graph representation of this quantization. Each node of the aspect graph represents a set of features (the aspect) observable from the same viewpoint. To each node we associate the set of viewpoints from which the aspect can be observed. Arcs in the aspect graph connect nodes with adjacent viewpoints. Each arc linking two aspects is characterized by the event(s) that take the one group of observed features to the other. Finally with each node in the graph we associate a *principal viewpoint*.

Aspect graphs of target silhouettes can be generated algorithmically or by exhaustive examination of the object. Our method is algorithmic. The given data are target silhouettes over a set of aspect and elevation angles, obtained either from CAD data or field data. Using the methods of [6] we construct the corresponding set of characteristic polygons $\mathcal{P}^{T_i}_{s,r}$, computed (automatically) at the same sensitivity and scale, and representing the silhouette of the same target T_i. Using sophisticated vector quantization techniques we cluster these representations into equivalence classes, corresponding to the nodes of the aspect graph. The algorithm simultaneously selects a representative viewpoint for each node, by selecting the most representative of the corresponding polygonal approximations of the *turning functions* (see [6]). We have experimented with various notions of "centroid" so as to select the best turning function corresponding to that aspect.

When we perform matching against portions of sensor data we use the corresponding polygonalizations of the of the representative viewpoints from each aspect. This speeds up the matching process considerably. We also have at the same time a substantial reduction of the set considered for matching. This set consists of all the aspects for each target. Our experience to date is that for each target we need a small set of aspects to cover the entire viewpoint range for low elevation, which was the case of interest motivating this work.[4] An example of viewpoint equivalence classes for an M60 tank, M35 truck and an M113 APC is shown in Figure 1. We expect a commensurate increase to cover elevation variations as well. We note that aspect graphs can be constructed for different values of the resolution parameter to cover target model variation with respect to range. Let $\mathcal{AG}^{T_i}_r, \mathcal{AG}^{T_i}_{r'}$ denote two aspect graphs of the same target T_i, at resolutions r, r', where r' represents the finer resolution. It is clear that the two graphs are tightly coupled: the aspects of the finer resolution aspect graph are subsets of the coarser resolution aspect graph. On the other hand the events of the coarser resolution graph are subevents of the finer resolution graph, since the finer resolution uses more local features of the target. We have also developed algorithms to automatically implement this quantization processing as well [6]. An aspect graph for an APC at coarse resolution is given in Figure 2.

Finally we can compare aspect graphs for different targets, on the basis of the local features grouped under each aspect (*i.e.*, node of the graph). A careful comparative analysis reveals possible rankings of the local features or groups of local features which can be used to guide the ATR search.

3.0 EARLY PHASE

The task of the early phase is primarily that of identifying regions of interest (ROIs). It accomplishes this by

[4] Three aspects were needed to cover the range [0°, 90°] for noiseless recognition of CAD models and high contrast targets in synthetic FLIR; this was initially extended to five aspects for general targets in synthetic FLIR imagery and later extended to seven due to the fast operation of the on-line probing algorithm and since more aspects don't hurt recognition performance, they only hurt time performance.

first locating points of interest (POIs) in the image using the so-called *vertical contrast probe* (VCP) which searches for locations in the image which are flanked to the left and right by vertically oriented lines of contrast separated by a distance which is roughly consistent with the known widths between the left and right side edges of the targets at the various aspects for a given set of range estimates. This probe clearly takes advantage of the iso-oriented target assumption of the *trim2* database (*i.e.*, targets are not "tilted") but through the "deadband" concept it can tolerate tilts of about $\pm 20°$ [4]. The vertical contrast probe is a special case of the more general *oriented contrast probe* [4] which forms the basis for a more restrictive *4-way contrast probe*.

The idea of the VCP is to employ a primitive geometrical constraint to quickly locate candidate points of interest (POIs) in the image. Around each of these POIs a target box is constructed using the *minimum bi-variance edge* (MbVE) algorithm which employs the separation distance found by the VCP and a novel stochastic-geometric line finder [3] [4]. The MbVE algorithm is a fast, $O(n)$-complexity, edge-finding algorithm which we developed for its speed, statistical properties and relation to weak continuity functionals [7], [13].

Each target box so constructed tightly hugs the target if present and provides a reference to define an outer concentric target box which is precisely the ROI. In Figure 3(a) we show the arrangement of target boxes at a POI. Figure 3(b) illustrates rough silhouettes extracted from *trim2* images which are used to construct ROIs. We note that these silhouettes are indeed rough and emphasize that *our probing algorithms employ no on-line edge extraction or segmentation for the purposes of matching*. The edges shown in Figure 3(b) are only used to automatically construct an ROI which is well-suited for the classifier processing to follow. Figure 3(c) shows the on-line geometry of such an automatically constructed ROI consisting of the inner target box and concentric boundary region which are used to compute statistical features.

The next step is to run a statistical classifier on the ROI to decide if it is non-target-like clutter or target-like. Notice that we have used low-level geometry constraints as a first step in identifying POIs worth further consideration, as opposed to wasting time computing classifier statistics over ROIs which lack such geometrical characteristics. In addition, the MbVE computation defines a statistically optimal inner-outer split of the ROI adaptively, so that the classifier can estimate means and variances, *etc.*, with sufficient statistical significance over a minimal set of pixels.

Four statistical features are calculated with reference to the ROI, namely, the contrast-to-noise ratio between the two, the fractal dimension of the inner box, the Kullback-Liebler distance between the two and the target variance. These four statistical features together with the height and width of the inner target box and the mean *edge strength* provided by the MbVE algorithm together form a seven-long feature vector which is classified using the Learning Vector Quantization (LVQ) decision rule. LVQ is a nonparametric classifier which "learns" its decision partition rule by examples from a database. The LVQ classifier was trained using the *trim2* database to learn the difference between actual (simulated FLIR) targets and clutter based on the seven-long feature vector and the *trim2* ground truth. The end result of training the algorithm is a partition of this seven-dimensional space into labeled cells called Voronoi cells; the cell labels are *target* or *clutter*. During on-line classification, the seven-long feature vector calculated for each ROI falls into some Voronoi cell of the trained classifier and then the ROI is classified as either *target* or *clutter* according to the label of that Voronoi cell. ROIs classified as clutter are dropped from any further processing. For this reason, the optimization involved in the off-line LVQ training procedure is not designed to achieve a very low false alarm rate but instead emphasizes high probability of detection as shown in Table 1.

Loosely speaking, the LVQ classifier is trained to reject obvious, non-target-like clutter. An ROI that is classified as *target* is passed on to the late phase along with a confidence that it is a target. The confidence is derived for each Voronoi cell during training and quantifies the fact that some cells have more discrimination power than others.

4.0 LATE PHASE OF THE PROBING ALGORITHM

The late phase of the Probing Algorithm (PA) emphasizes target recognition proper and the rejection of target-like clutter. This phase relies heavily on probing methods to estimate the viewpoint and perform the classification and rejection. First, the viewpoint is estimated, where the viewpoint is a triple consisting of the range to the target, the position of the target (i.e., the location of the centroid of the silhouette) and the aspect angle of the target. The viewpoint is estimated using a model-based maximum likelihood estimation procedure, taking a *separation of classification and estimation* philosophy. This is the most computationally intensive part of the algorithm, with complexity linear in the number of targets. Because of this, we conceptually isolated viewpoint probing from the rest of the algorithm and performed the required optimization using a coarse-to-fine (suboptimal) maximization, the idea being that future improvements in the mathematical organization of the estimation procedure can be implemented

immediately, independent of the workings of the remainder of the algorithm. We have recently implemented two such improvements that result in an order of magnitude speed-up over the code delivered earlier to C^{2}NVEO. Furthermore, it is important to emphasize that even the basic viewpoint estimation module can be given an entirely parallel implementation using standard digital hardware resulting in efficient real-time operation. In addition, we have obtained preliminary results which indicate that these probing algorithms may be implemented using an optical processor, if necessary.

The PA is a model-based algorithm which uses stored silhouette models. Hence, an important question addressed was how many different representative aspect angles for each target are needed to discriminate between the three military targets in target-like clutter with high recognition performance. It is here that our results on aspect graphs and related target model reduction resulted in substantial computational savings. We found that five aspects per target were sufficient to cover the range $[0°, 90°]$ at 1.5km to center field-of-view, although later this was increased to seven for better overall performance in target-like clutter (i.e., thirteen for $[0°, 180°]$). Again, it is important to emphasize that this small number of viewpoints can be easily handled in real-time by standard parallel hardware.

A significant proportion of the computations performed by the viewpoint estimation probe (VEP) is directly related to the maximization over the target model scale, or equivalently, target range. Again, this complexity was added in order to address C^{2}NVEO requirement that the algorithm tolerate about a ± 0.6km range uncertainty. However, it is obvious that in an operational FLIR, range uncertainty will be more effectively addressed via multi-sensor fusion in which range estimation is performed using LADAR or mm-wave sensor data.

For each ROI passed on from the early phase, the VEP algorithm starts off with an on-line estimate of the mean scale and centroid for each possible target class. The mean scale is provided by the rough range to center field-of-view, while the mean centroid is taken to be the centroid of the ROI. The probing for the true scale (range to target) and true centroid is performed in a neighborhood of this mean scale and mean centroid. The algorithm also probes for the true aspect. The end result is three (for the *trim2* database) viewpoint estimate triples, say G_t, G_a, G_v, one for each target class: tank, APC , truck (van). These viewpoint triples have the usual maximum likelihood interpretation: If a tank is present within the current ROI, then the best guess according to the VEP is that its aspect, range and location is given by G_t; If an APC is present within the current ROI, then the best guess according to the VEP is that its aspect, range and location is given by G_a; and analogously for the truck (van). Based on these three estimates, one can derive an overall maximum likelihood decision for the target conditioned on the hypothesis that the ROI contains a target, however, for improved recognition performance we developed another type of probe, discussed next, in order to implement a generalized maximum likelihood decision rule.

In the next stage of processing, the three viewpoint triples estimated by the maximum likelihood viewpoint estimation probe are passed on to the class decision probe (CDP) which exploits the relative juxtapositions, ranges and aspects of the three hypothesized target silhouettes, as determined by G_t, G_a and G_v, in order to discriminate between the active target hypotheses. Because the optimization performed by the VEP takes place over a finite, eight-dimensional set, the number of combinations is finite but large, over 300 million. Of course, through the approximate coarse-to-fine optimization employed during the viewpoint probing only one-ten thousandth of these combinations are examined in any given ROI. Nevertheless, since these combinations are not known *a priori*, the number of combinations cannot be reduced off-line. Hence, an off-line optimization for the optimal probe sets involving the entire *trim2* database was considered out of the question, especially since even a single change in the parametrization of the search neighborhoods would entail another enormous off-line optimization. Such an optimization was deemed to be practically unacceptable in an actual military target recognition environment, especially given the lessons learned in the Gulf War. Instead, we developed a fast method to compute the optimal probe sets *on-line* using computational geometric ideas. These probe sets permit the algorithm to resolve a "best" decision as to the true contents of the ROI, conditioned on the hypothesis that the ROI contains one of the three targets. The optimality of this decision is of course constrained by the information provided by the silhouette modeling and we stress that much better decisions will result from the proper use of full internal boundary models [6].

The basic probing performed by the VEP is to combine the contributions of each local statistic computed on-line over the optimal probing sets. During this process each viewpoint triple is ranked according to this statistic and the triple with the highest statistic value defines the maximum likelihood estimate for each target class. The most complex probing in the PA takes place during the class decision probing. This process is guided by the stored target silhouette models which determine the geometry and location of the probing sets. The probing sets are then

superimposed on the image data in order to compute the local statistics. Here, the local statistics can be interpreted as saliencies (or fidelities) of the hypotheses generated by the stored target silhouettes and the optimal probing sets. Consider one of the hypothesis pairs passed to the CDP after the viewpoint estimation probing. This involves two viewpoint triplets, say G_t and G_a, which in general one can think of as representing two silhouettes of two different sizes, in two different positions and at two different aspects in the vicinity of the ROI. This part of the algorithm can also be interpreted in a maximum likelihood sense. An asymmetric test is performed, using optimal on-line probing, designed to maximize the discrimination power of the statistic by optimizing the probing sets.

After the CDP, the viewpoint triplet of the "winner" is passed on to the target-like clutter probe (TCP) for a final "target *versus* clutter" test. The final test employs a combination of four parameters from matching measures based on probing, which are combined using piecewise linear discriminant functions designed to maximize clutter rejection without reducing target detection and recognition. This discriminant makes the final target/clutter decision for each ROI. The final step of the algorithm is to resolve decisions which are spatially inconsistent, for example when two target calls result in two different targets supposedly occupying the same physical space.

5.0 LATE PHASE OF THE 4-WAY DISTANCE ALGORITHM

The late phase of the 4-Way Distance Algorithm (DA), like the PA, also emphasizes target recognition proper and the rejection of target-like clutter. The approach used in the DA relies on the use of a polygonal distance metric for recognition purposes and decision logic designed off-line using classification trees. The DA was developed in order to explore the utilization of classification trees in learning imagery environments when appropriate databases exist. It was expected (and experiments and tests validated this) to have inferior performance compared to the PA when faced with a totally unknown database or scenario.

Decision or *classification trees* (CT) partition the decision space by performing clustering. CTs are useful in problems with high dimensionality possessing a mixture of data types and nonstandard data structures and involving sequential decision making, hence their utility in ATR: Existing databases can be used to construct CTs. Following [8] the construction of CTs involves three steps: the selection of the set of questions Q, a rule for selecting the best split and a criterion for choosing the right sized tree.

Binary CTs are constructed by repeated splits of subsets of the measurement data X into two descendant subsets. A learning sample $\mathcal{L}$ is used simultaneously to construct the tree and provide estimates of performance. In designing the "optimal" split s for a node t, we consider the decrease in the impurity of the node t. More precisely let

$$I(t) = -\sum_j p(j|t) \log\ p(j|t),$$

denote the measure of the impurity of node t (the entropy of the node); where $p(j|t)$ denote the proportions of each class in node t. If t_L, t_R denote the two nodes splitting t the goodness of the split s is given by

$$\Delta I(s,t) = I(t) - p_L I(t_L) - p_R I(t_R).$$

The optimal split s is selected so as to maximize the above decrease in impurity. The methods developed in [8] first develop a tree and then prune it to a preferred size by considering estimates of misclassification rates as a function of the tree depth. An easier to compute impurity index with many good properties is the *Gini index*

$$I(t) = \sum_{i \neq j} p(i|t)p(j|t).$$

As stated in [8], the actual design problem is the clever selection of the questions. In CTs used for ATR the actual design problem is the selection of the features to use in order to split the nodes. One can examine off-line many features and select the optimal. One can also design the optimal size CT off-line provided adequate data are available. The attractive promise of course is that a well designed CT will reduce the on-line time requirements when it is used for ATR against real data. However, the required Monte Carlo simulations and related computations can be quite extensive for a large set of targets. Instead of constructing such trees based entirely on Monte Carlo simulations, we advocate their construction through the use of geometric and physical properties of the target and of the scene, in order to select appropriate features to split nodes. For instance, the ranking of local features, can provide appropriate

input for this. The major disadvantage of CTs is that they are relatively inflexible in that changes in the target set can cause major recomputations.

The 4-way distance method decomposes each target model into four parts called the south, east, north and west, which represent the bottom, right side, top and left side of the target, respectively. Matching is done in terms of these subtemplates in order to provide a simple 4-way immunity to target obscuration. This decomposition requires that the matching process respect constraints such as *target consistency*, *relation consistency* and *viewpoint consistency*. Target consistency means that each of the four model parts must come from the same target template; relational consistency, for example, means that the east subtemplate is truly in the east, etc., as well as satisfying other geometric relations between the local templates; and, viewpoint consistency means that all four parts are consistent with a certain viewpoint of the target. Recognition is done using a decision tree which is designed off-line. The consistency constraints just described are embedded in the design implementation of the DA via decision trees. The matching is done using a polygonal distance metric on a second set of minimum bi-variance edges computed within the ROI.

6.0 EXPERIMENTS, TESTS AND PERFORMANCE

We have described above two model-based ATR algorithms utilizing novel techniques. The emphasis of the work has been to concentrate on a bounded problem so as to initiate a systematic, methodological development of model-based ATR algorithms while avoiding reliance upon heuristics. In particular, the effort was directed to address economic target model development, model complexity, efficient algorithm development, algorithm complexity, algorithm performance and the relationship of model complexity to algorithm complexity and performance. As a consequence of these objectives, the algorithms were designed so as to use only target silhouette information at this phase. The performance achieved in tests and experiments is remarkable given the fact that primarily only target silhouette data are used and that range has great uncertainty. The use of internal boundaries, thermal characteristics, *etc.*, is planned for future stages and it is bound to increase dramatically the performance.

The problem definition provided to us by C^2NVEO had the following key ingredients: Three-class/flat-earth scenarios as in C^2NVEO *trim2* database; Target classes consisting of the M60 tank, the M113 APC and the M35 truck (or van); 2° camera depression angle. Only the rough range to center field-of-view was assumed as input with the nominal ranges to center field-of-view being 1.5km, 2.5km, 3.5km and 4.5km as in the *trim2* database. Many ATR algorithms require as input some *a priori* estimate of the clutter level within the input image; our approach, however, does not require this information as it employs a probing-based, on-line estimator of the clutter level that works with essentially 100% accuracy. Through a self-scaling feature the PA can handle input range bins from 1km to 5.5km. This scaling feature was included at a significant computational cost in order to satisfy the constraint that the algorithm tolerate rough range estimates to center field-of-view ($\pm$0.6km) on input. Both algorithms were designed under the *iso-oriented* target assumption, i.e., that neither the targets nor the camera are "tilted" as is true in the *trim2* database; including "tilt" and other orientation uncertainty is a straightforward extension of the methodology. The PA was designed with a modest immunity to target obscuration: certain types of obscuration involving up to approximately 40% of the target silhouette boundary can be withstood by the PA.

The PA algorithm is divided into two phases called the *early phase* and the *late phase*. The objective of the early phase is to reject *non-target-like clutter* and to locate candidate *regions of interest* (ROIs) for subsequent classification within the late phase. The late phase classifies the ROIs passed to it as either *target-like clutter* or as *target*. If a given ROI is classified as *target*, then the PA also provides the target class decision and calculates a viewpoint estimate consisting of a position estimate for the centroid of the target, a range estimate for the target and an aspect angle estimate. The PA currently emphasizes primarily silhouette boundary information, although more fully incorporating internal boundary details is a relatively straightforward extension of the basic methodology and should result in substantial improvement in recognition performance and false alarm suppression. In the current version of the PA, internal boundary information is mainly used to identify so-called "signature targets," *e.g.*, a tank with a cold turret and a hot chassis. The PA was initially designed to accept parameter sets categorized by two clutter levels, high and low, and four ranges. This resulted in eight different sets of parameters to fine-tune, although as it turned out, the final settings of the most important parameters were practically identical for all eight combinations of range and clutter. For example, the clutter rejection probe employed by the PA uses the same four filter parameters and four thresholding parameters across all ranges and clutter levels. This indicates that the parameterizations chosen are in a certain statistical sense "minimal" and robust.

The DA algorithm also is divided into an early and a late phase. The early phase is essentially identical to that of the PA algorithm, while the late phase employs polygonal distance matching and decision tree logic. The DA employs substantial learning and can be trained to learn a database. On the other hand, the PA algorithm is adaptive and does not require learning of training for recognition. It is expected that the probing algorithm will perform better on previously unseen databases. On the other hand, a well-trained DA can provide a calibration of the performance of the PA.

Table 2 gives typical values for pixels on strong to medium contrast targets for database comparison purposes. In Tables 3–5 we provide typical results for the two algorithms on the C^{2}NVEO *trim2* terrain board data. In Table 3, we see that the 4-Way Distance algorithm can indeed achieve extremely good performance via over-training. The tree trained for the results in Table 3 had over 200 nodes, so this represents a significant amount of learning. This table therefore provides a benchmark, a close to optimal performance measure, against which one can assess the relative performance of the PA or the absolute difficultly of the database. Table 4 shows that the performance of the probing algorithm compares favorably against this benchmark. Note in particular the very good results indicated in the confusion matrices. Table 5 gives the results of the 4-Way Distance algorithm without optimized learning and we see the concommitant loss of performance.

Based on our partial inspection of tests performed by C^{2}NVEO on an unknown database it appears that the performance of the PA reported in Table 4 has held up. In part we attempted to account for the fact that almost one-third of the targets in these tests were tilted, a condition not addressed in the design of the PA.

A proprietary analysis utility for the PA which visually displays the algorithm in operation has been built. In Figure 4 through Figure 9 we show some representative "screendumps" during part of an on-line session involving part of the late phase and a thermally uniform head-on tank from the *trim2* database (cota:0531). Figure 4 through Figure 6 show the VEP in operation as it probes for the optimal viewpoint triple. The upper left window shows the target and associated probing sets. The two central windows show the current (left) and current best target for each class (right) and their associated "two-dimensional" probing statistics. The two bottom windows show the one dimensional probing statistics. Figure 7 and Figure 8 show screendumps of the CDP in action at two phases of the class decision probing; the windows have the analogous interpretation. Figure 9 gives a close-up of the optimal probing sets computed on-line for this engagement. This analysis utility facilitates improvements in the Probing Algorithm's capabilities and performance.

7.0 CONCLUSIONS AND EXTENSIONS

In conclusion, we have described an approach to ATR algorithm design which is model-based and relies extensively on advanced probing methods. From the early phase to the late phase the resulting algorithms expend computational time on an ROI proportional to the difficulty of the region, however the main parts of the algorithms are entirely parallelizable and involve only simple local computations. Of the two algorithms, the probing algorithm is the most promising and has produced the best results to date in experiments and tests. It is also robust and displays graceful degradation when unknown or only partial data are encountered.

Future areas of investigation include expansion to a six to ten class problem, handling obscuration in an even more robust way, including tilt estimation within the viewpoint estimation probe, and extension to multiple sensors as well as improving the organization, speed, detection performance and recognition power of the algorithms. Regarding extension to multisensor ATR, we would like to point out the similarity of target silhouette representation described in [6] for FLIR imagery, with that of target silhouette representation in LADAR data, and in target range profiles in mm-wave data. The former is obvious, the latter can be seen easily from the data in Figure 10. Figures 10(a), (b), (c) display mm-wave radar range profiles for a tank, APC and truck, respectively, from the VISION I database. The resulting piecewise constant waveforms are quite similar with the polygonal approximation to the target silhouette when viewed in turning function space [6] as Figure10(d) clearly illustrates. This commonality of data representation implies that our scale-space polygonalization methods, clustering, learning and probing algorithms can be easily modified so as to apply on data from other sensors as well as in producing efficient feature based fusion in multi-sensor ATR algorithm development.

Acknowledgments The authors would like to thank Dr. V. Mirelli and Ms. T. Kipp for their helpful criticism and many useful suggestions throughout the development of the algorithms described here, especially with respect to the late phase probing algorithms.

REFERENCES

1. H. Asada and M. Brady, "The Curvature Primal Sketch," *IEEE Patt. Analysis and Machine Intell.* vol. PAMI-8, no. 1, pp. 2-14, 1986.
2. F. Attneave, "Some Informational Aspects of Visual Perception," *Psych. Rev.*, 61, 1954, pp. 183-193.
3. J.S. Baras, D.C. MacEnany and A. LaVigna, "Automatic Target Recognition Algorithms," Quarterly Progress Report on Contract DAAB07-90-C-F425, AIMS-TR-90-7, AIMS INC., December 1990.
4. J.S. Baras, D.C. MacEnany and A. LaVigna, "Automatic Target Recognition Algorithms," Quarterly Progress Report on Contract DAAB07-90-C-F425, AIMS-TR-90-7, AIMS INC., April 1991.
5. K. Bowyer, D. Eggert, J.Stewman, and L. Stark, "Developing the Aspect Graph Representation for Use in Image Understanding", *Proc. DARPA Image Understanding Workshop 1989*, pp.831-849.
6. D.C. MacEnany and J.S. Baras, "Scale-Space Polygonalization of Target Silhouettes and Applications to Model-Based ATR," *Proceedings of the Second ATR Systems and Technology Conference.* Feb. 1992.
7. A. Blake and A. Zisserman, *Visual Reconstruction* MIT Press, 1987.
8. L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone, *Classification and Regression Trees*, Wadsworth and Brooks, 1984.
9. W. Eric L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, The MIT Press, 1990.
10. D. D. Hoffman, "Representing Shapes for Visual Recognition", Ph. D. Thesis, MIT, 1983.
11. D.P. Huttenlocher and S. Ullman, "Object Recognition Using Allignment", *Proc. DARPA Image Understanding Workshop, 1987*, pp. 370-380.
12. D. G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, 1985.
13. D. Mumford and J. Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Comm. on Pure and Appl. Math.* Vol. XLII,pp. 577-685, 1989.
14. W. A. Perkins, "A Model-Based Vision System for Industrial Parts", *IEEE Trans. Comp.*, 27(2), 1978, pp. 126-143.
15. A. Witkin, "Scale-space filtering", in *From Pixels to Predicates* A. Pentland (Ed.) Ablex Publishing, pp. 5-19 1986.

Probability of Target Detection from Early Phase

Range	Clutter1	Clutter2	CLutter 3
1500	1.00	1.00	0.98
2500	0.98	0.92	0.98
3500	0.97	0.99	0.97
4500	0.97	0.96	0.94

Table 1

Number of Pixels on Template

Range	M60-0 deg	M60-90 deg	M113-0 deg	M113-90 deg	M35-0 deg	M35-90 deg
1500	1302	2220	640	1045	864	1771
2500	396	714	192	363	289	616
3500	176	336	135	276	120	280
4500	108	184	81	171	72	147

Table 2

4-Way Distance Algorithm (over-trained) (*trim2* database)

Probabilities of Detection and False Alarms for all Ranges and Clutter Classes

Prob. Detection	False Alarm/sq.deg	Range	Clutter Class
0.92	0.01	1500	1
0.87	0.00	2500	1
0.76	0.02	3500	1
0.58	0.63	4500	1
0.81	0.24	1500	2
0.73	0.22	2500	2
0.69	2.62	3500	2
0.65	2.29	4500	2
0.74	0.42	1500	3
0.74	0.21	2500	3
0.74	2.27	3500	3
0.70	2.84	4500	3

Confusion Matrix for all Clutter Classes at 1500m

	Tank	APC	Truck
Tank	0.80	0.06	0.09
APC	0.13	0.87	0.07
Truck	0.07	0.07	0.84

Confusion Matrix for all Clutter Classes at 2500m

	Tank	APC	Truck
Tank	0.81	0.09	0.03
APC	0.09	0.82	0.01
Truck	0.10	0.09	0.96

Confusion Matrix for all Clutter Classes at 3500m

	Tank	APC	Truck
Tank	0.69	0.10	0.12
APC	0.17	0.76	0.17
Truck	0.14	0.14	0.72

Confusion Matrix for all Clutter Classes at 4500m

	Tank	APC	Truck
Tank	0.82	0.08	0.25
APC	0.09	0.79	0.14
Truck	0.09	0.13	0.61

Table 3

Probing Algorithm (*trim2* database)

Probabilities of Detection and False Alarms for all Ranges and Clutter Classes

Prob. Detection	False Alarm/sq.deg	Range	Clutter Class
0.85	0.00	1500	1
0.87	0.01	2500	1
0.71	0.05	3500	1
0.54	0.27	4500	1
0.77	0.29	1500	2
0.70	0.26	2500	2
0.65	3.50	3500	2
0.53	1.48	4500	2
0.67	0.89	1500	3
0.82	0.42	2500	3
0.72	2.43	3500	3
0.57	2.25	4500	3

Confusion Matrix for all Clutter Classes at 1500m

	Tank	APC	Truck
Tank	0.74	0.03	0.11
APC	0.22	0.89	0.13
Truck	0.05	0.08	0.77

Confusion Matrix for all Clutter Classes at 2500m

	Tank	APC	Truck
Tank	0.83	0.07	0.12
APC	0.07	0.82	0.01
Truck	0.10	0.11	0.87

Confusion Matrix for all Clutter Classes at 3500m

	Tank	APC	Truck
Tank	0.73	0.05	0.15
APC	0.19	0.80	0.10
Truck	0.09	0.15	0.75

Confusion Matrix for all Clutter Classes at 4500m

	Tank	APC	Truck
Tank	0.82	0.14	0.14
APC	0.06	0.64	0.07
Truck	0.12	0.22	0.79

Table 4

4-Way Distance Algorithm (*trim2* database)

Probabilities of Detection and False Alarms for all Ranges and Clutter Classes

Prob. Detection	False Alarm/sq.deg	Range	Clutter Class
0.83	0.02	1500	1
0.87	0.00	2500	1
0.76	0.02	3500	1
0.58	0.69	4500	1
0.71	0.38	1500	2
0.69	0.41	2500	2
0.68	2.35	3500	2
0.62	2.17	4500	2
0.64	0.46	1500	3
0.62	0.49	2500	3
0.69	2.00	3500	3
0.70	2.55	4500	3

Confusion Matrix for all Clutter Classes at 1500m

	Tank	APC	Truck
Tank	0.81	0.04	0.08
APC	0.12	0.89	0.05
Truck	0.07	0.07	0.87

Confusion Matrix for all Clutter Classes at 2500m

	Tank	APC	Truck
Tank	0.86	0.09	0.02
APC	0.10	0.87	0.01
Truck	0.04	0.04	0.98

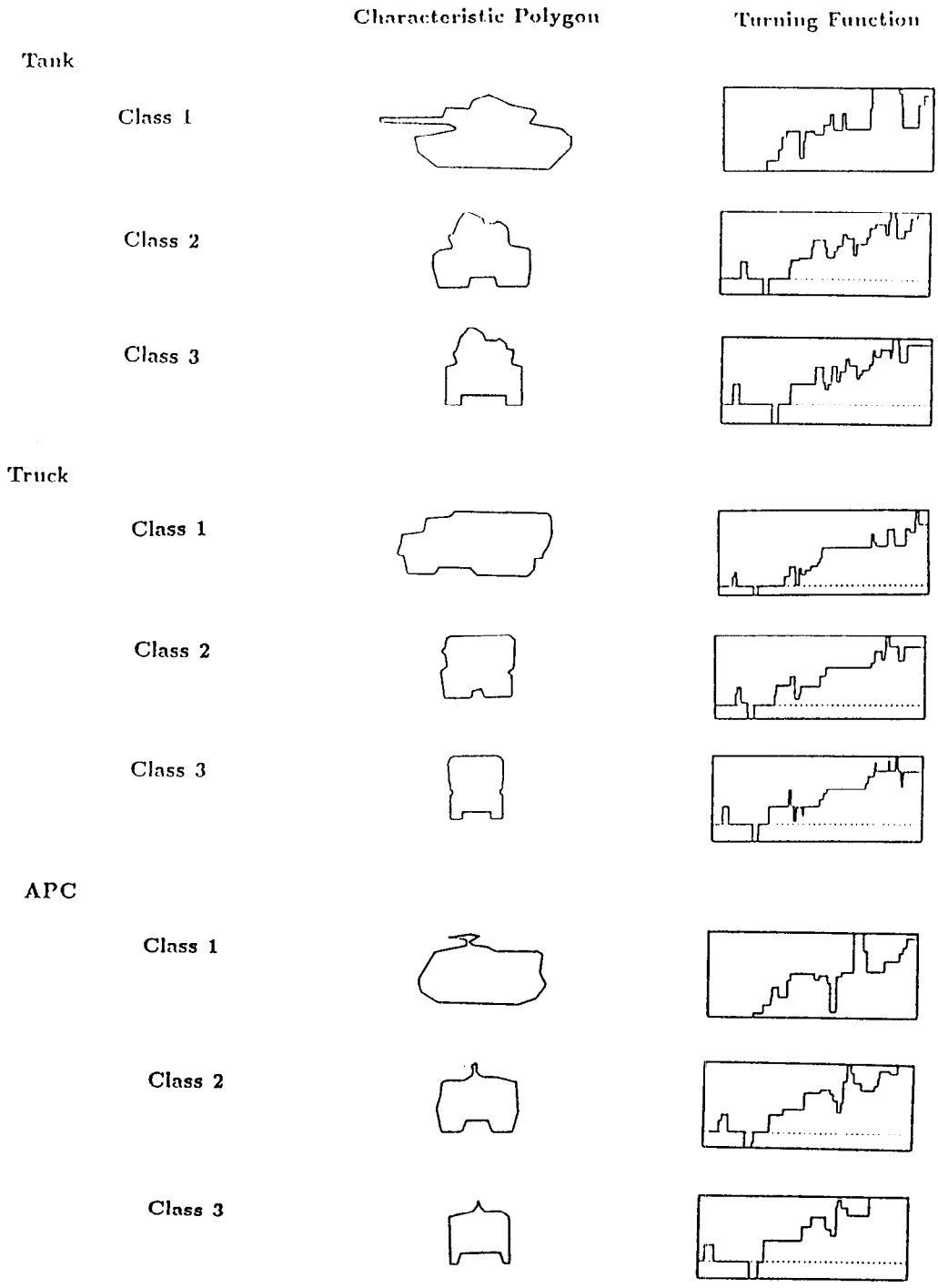
Confusion Matrix for all Clutter Classes at 3500m

	Tank	APC	Truck
Tank	0.38	0.08	0.09
APC	0.22	0.61	0.20
Truck	0.40	0.31	0.71

Confusion Matrix for all Clutter Classes at 4500m

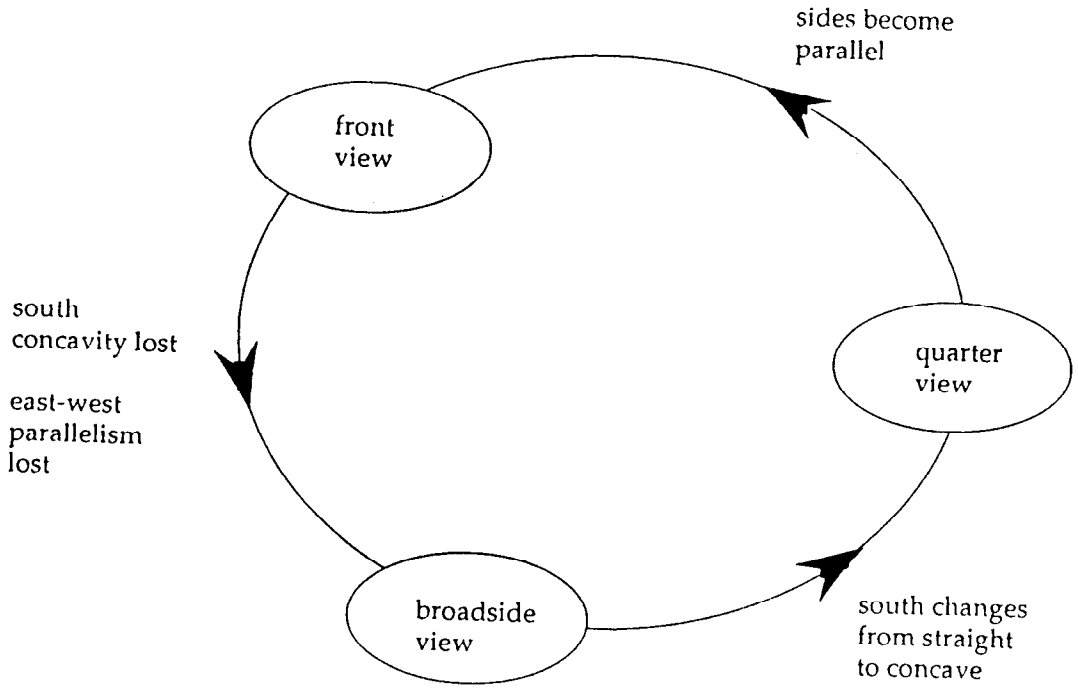
	Tank	APC	Truck
Tank	0.83	0.08	0.25
APC	0.08	0.79	0.14
Truck	0.08	0.12	0.61

Table 5



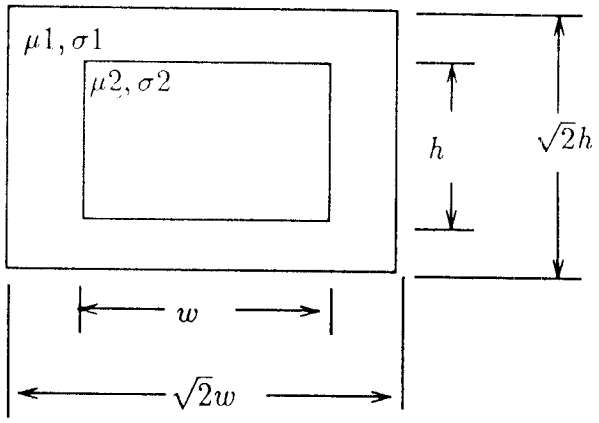
Aspect clusterings of characteristic polygons for M113, M60 and M35.

Figure 1



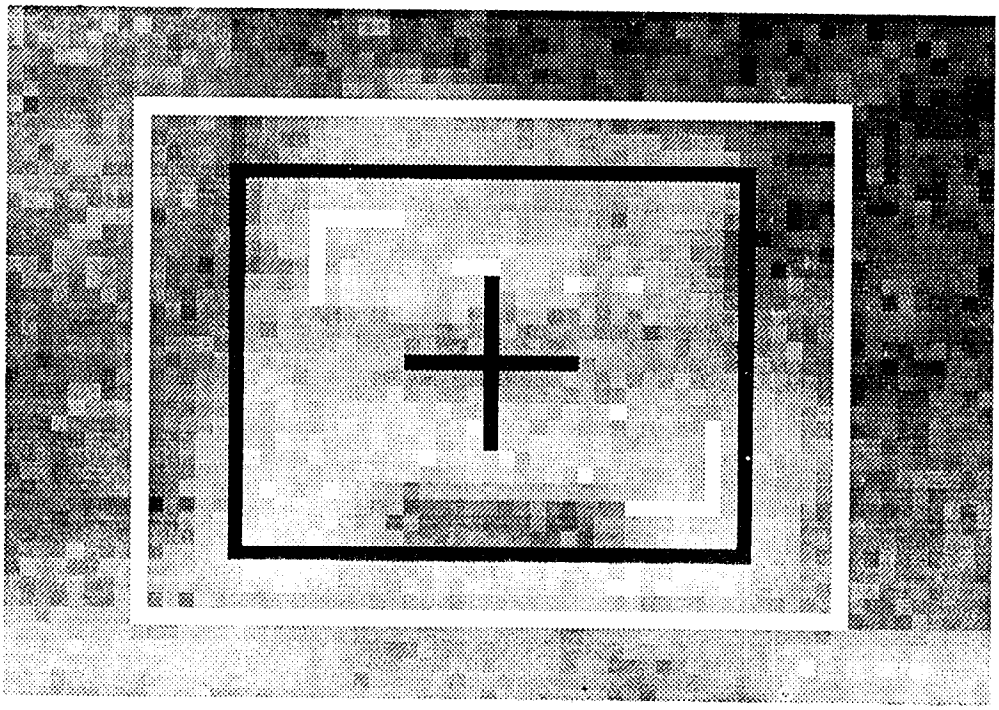
Aspect graph for APC at course resolution.

Figure 2



(a)

(b)



(c)

Figure 3

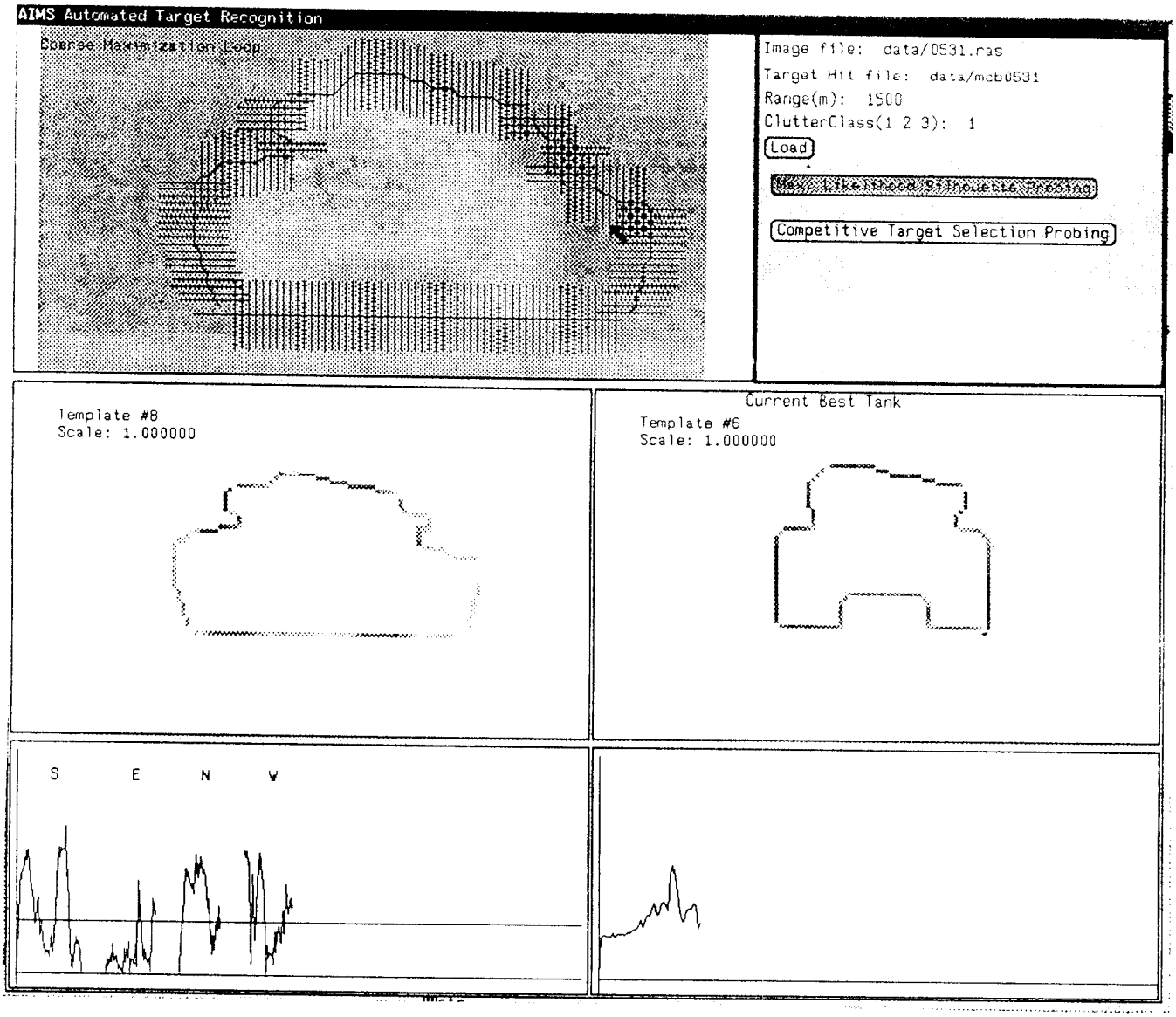

AIMS Automated Target Recognition
Coarse Maximization Loop
Image file: data/0531.ras
Target Hit file: data/mcb0531
Range(m): 1500
ClutterClass(1 2 3): 1
Load
Max. Likelihood Silhouette Probing
Competitive Target Selection Probing
Template #8
Scale: 1.000000
Current Best Tank
Template #6
Scale: 1.000000
S
E
N
W

Figure 4

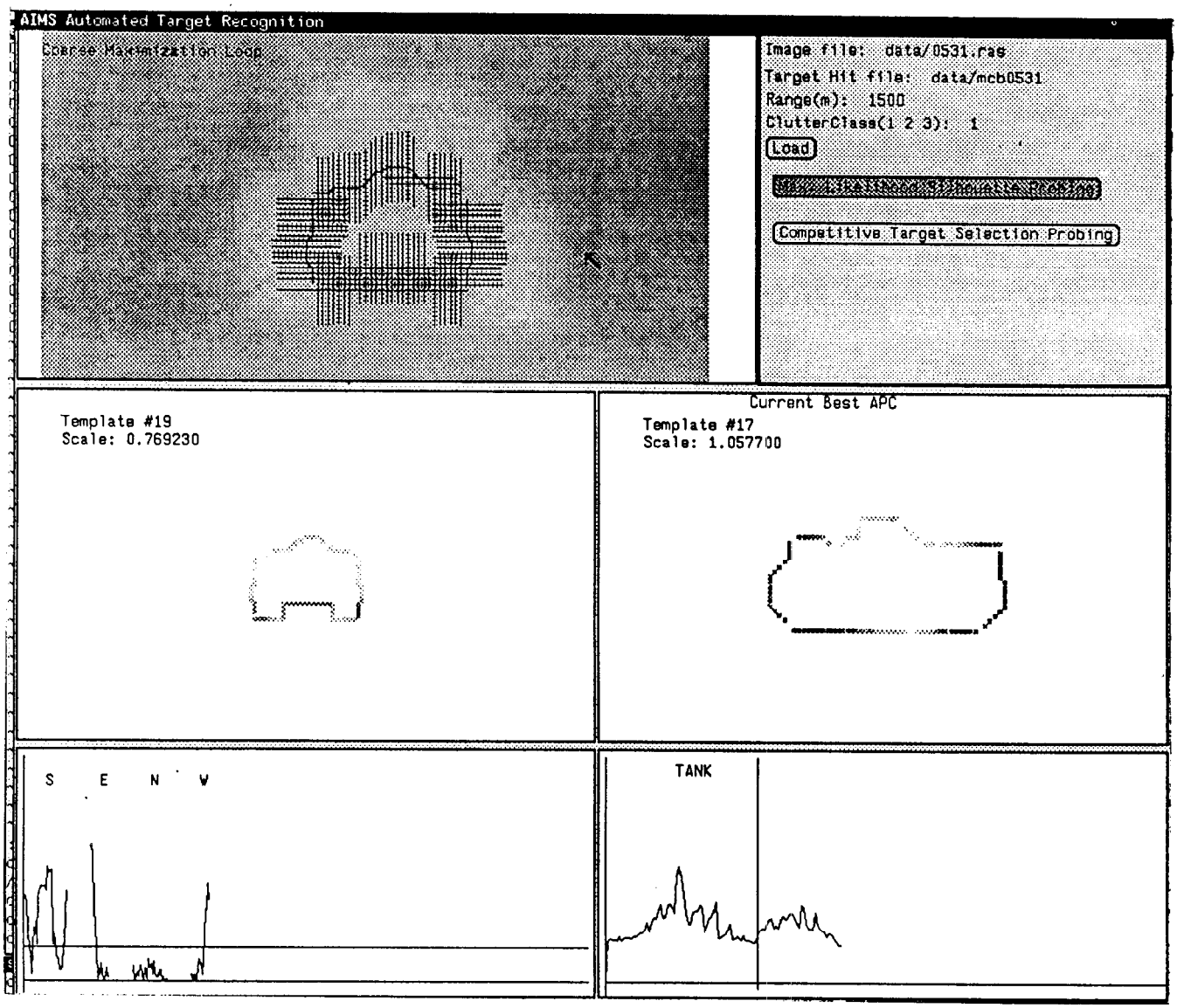

AIMS Automated Target Recognition
Coarse Maximization Loop
Image file: data/0531.ras
Target Hit file: data/mcb0531
Range(m): 1500
ClutterClass(1 2 3): 1
Load
Competitive Target Selection Probing
Template #19
Scale: 0.769230
Current Best APC
Template #17
Scale: 1.057700
S E N W
TANK

Figure 5

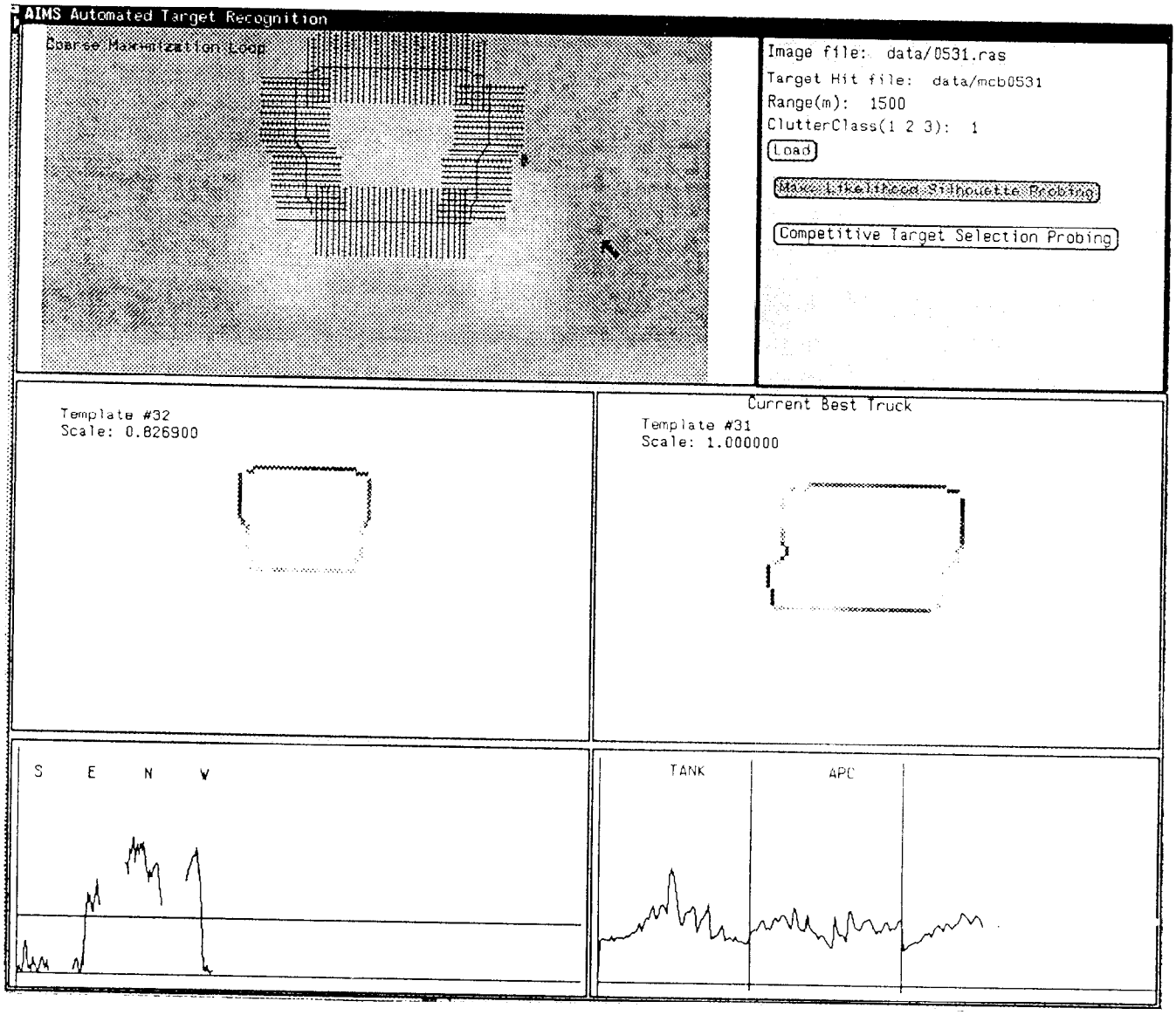

AIMS Automated Target Recognition
Coarse Maximization Loop
Image file: data/0531.ras
Target Hit file: data/mcb0531
Range(m): 1500
ClutterClass(1 2 3): 1
Load
Max. Likelihood Silhouette Probing
Competitive Target Selection Probing
Template #32
Scale: 0.826900
Current Best Truck
Template #31
Scale: 1.000000
S E N W
TANK
APC

Figure 6

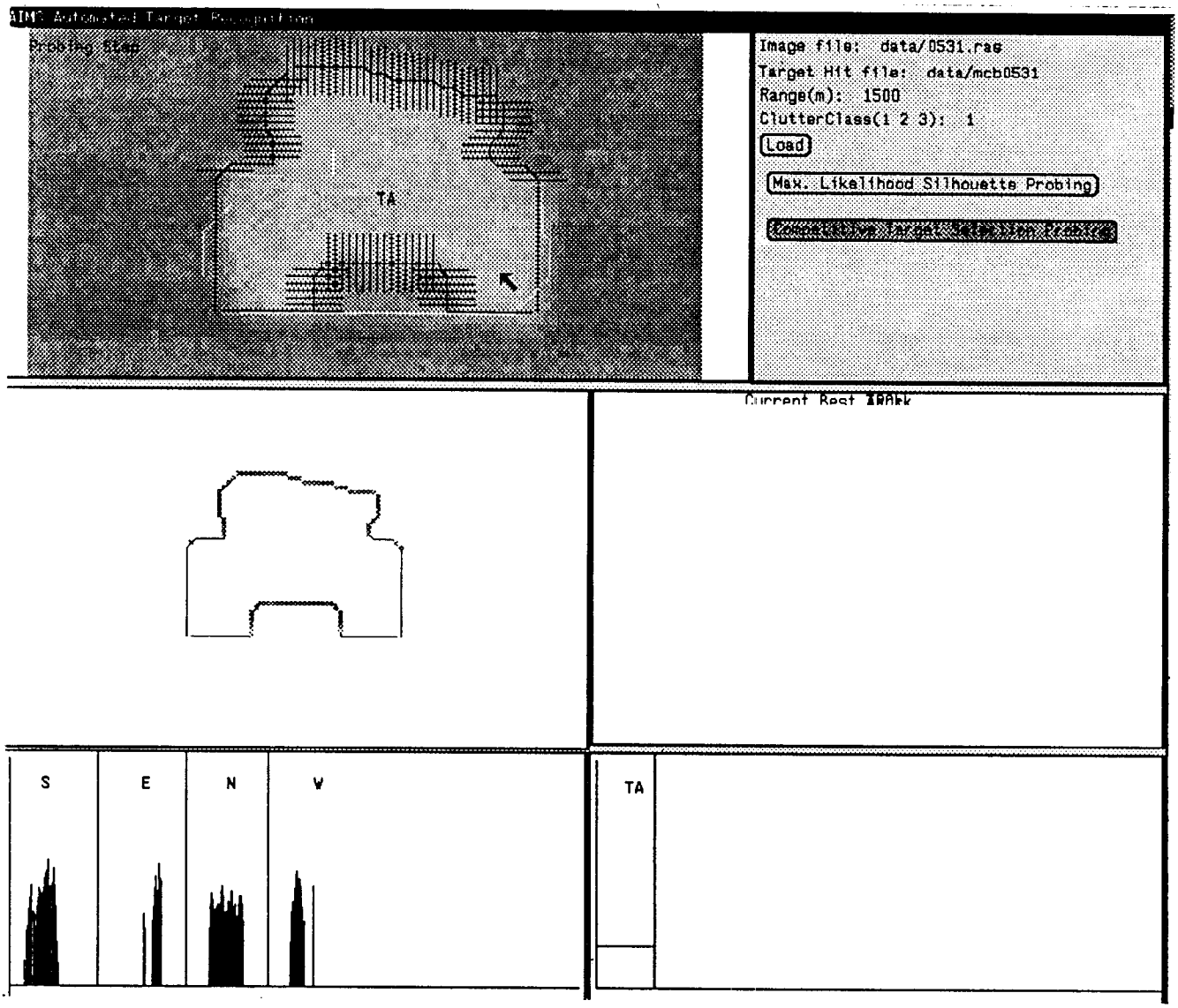


Image file: data/0531.ras
Target Hit file: data/mcb0531
Range(m): 1500
ClutterClass(1 2 3): 1
Load
Max. Likelihood Silhouette Probing
TA
S
E
N
W
TA

Figure 7

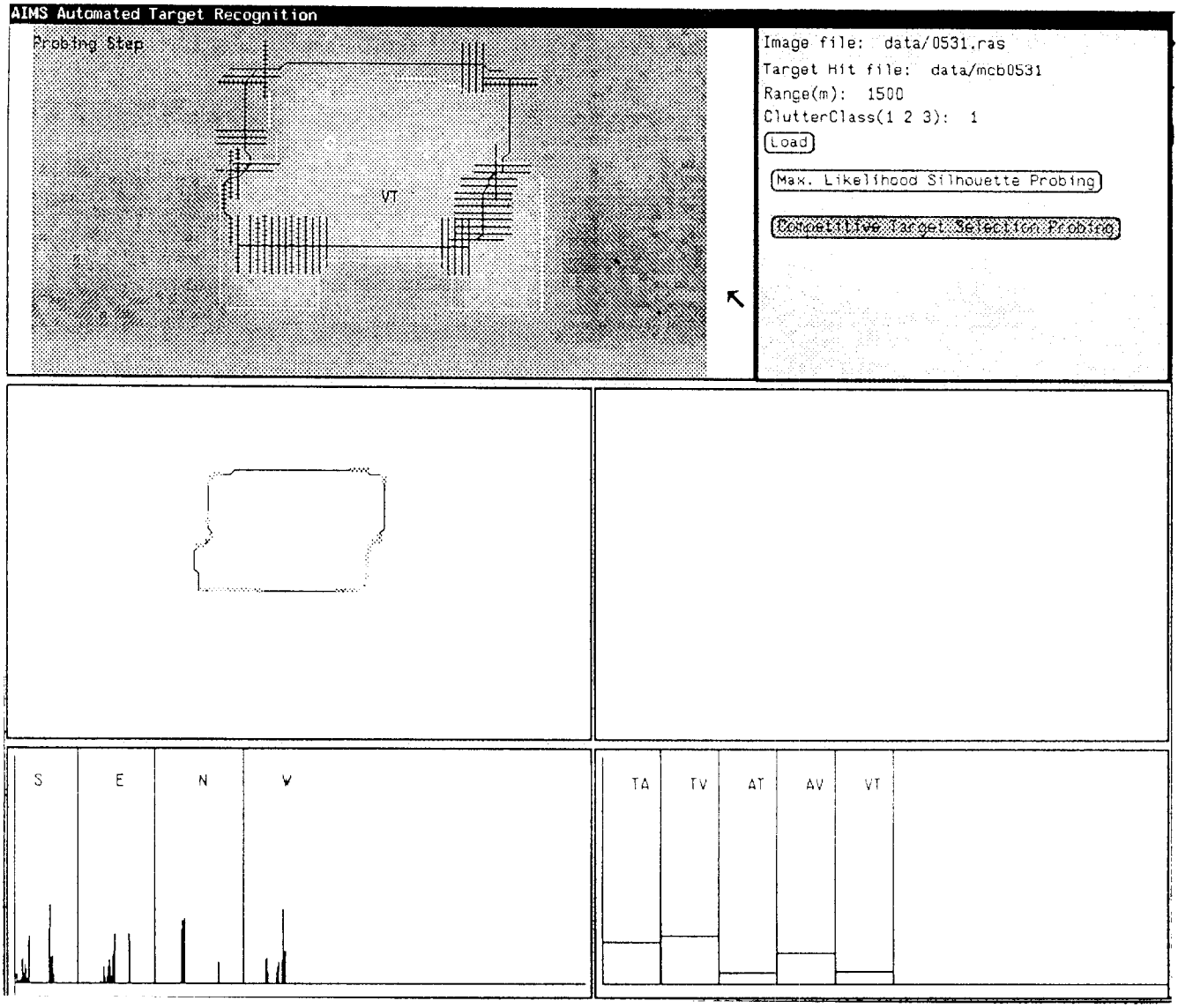

AIMS Automated Target Recognition
Probing Step
VT
Image file: data/0531.ras
Target Hit file: data/mcb0531
Range(m): 1500
ClutterClass(1 2 3): 1
Load
Max. Likelihood Silhouette Probing
Competitive Target Selection Probing
S
E
N
W
TA
TV
AT
AV
VT

Figure 8

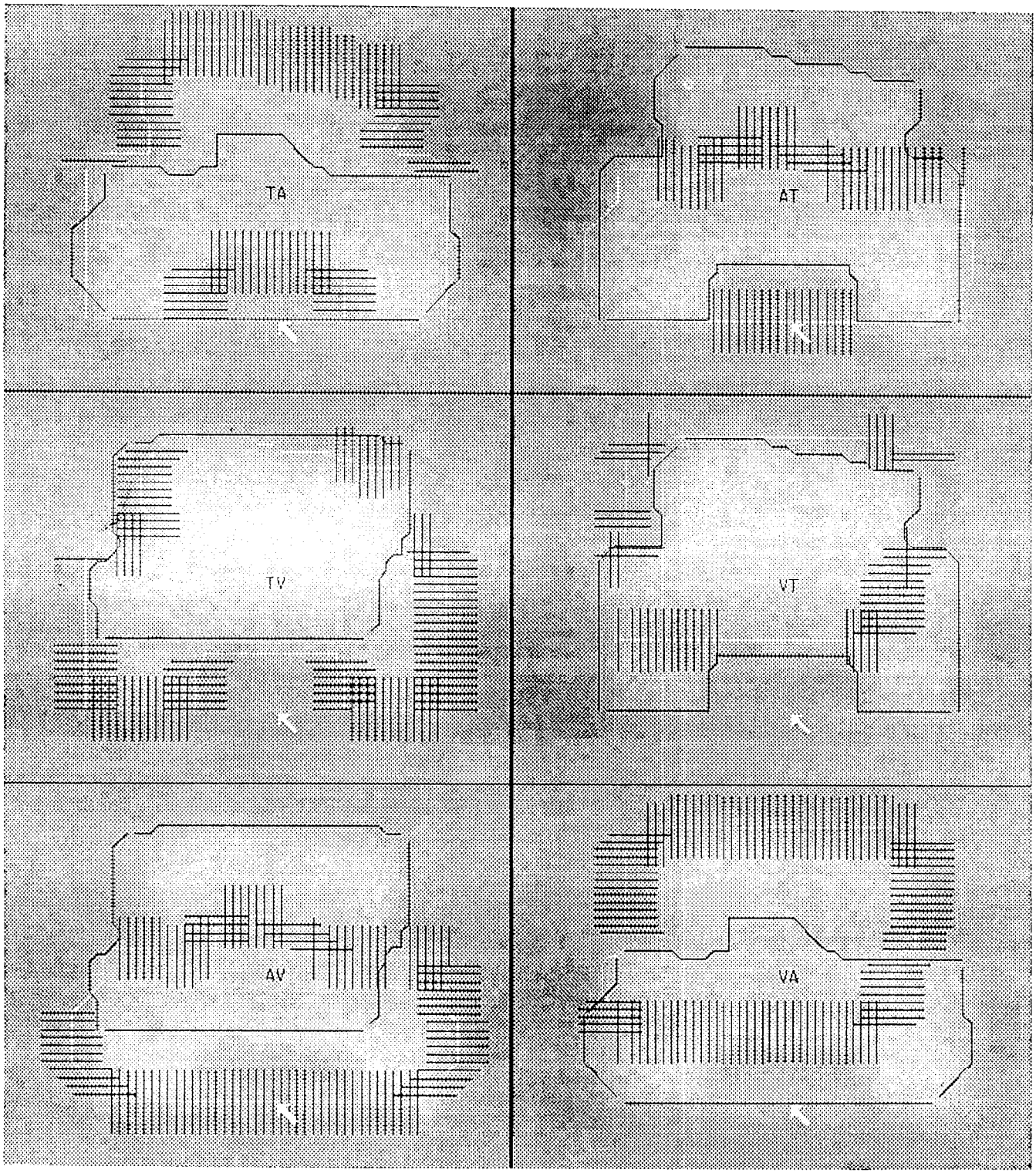
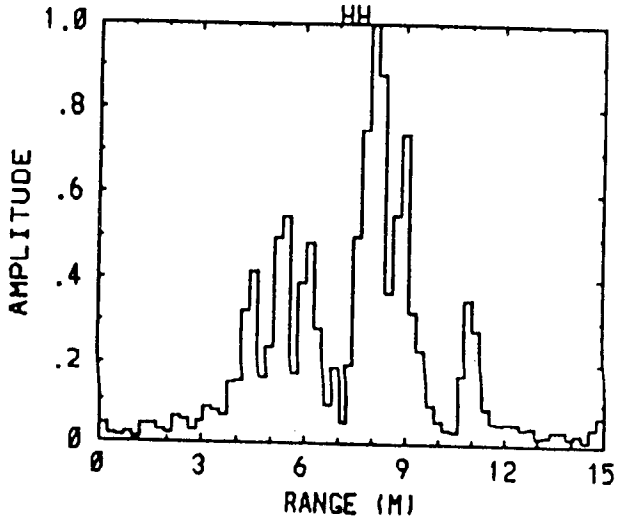
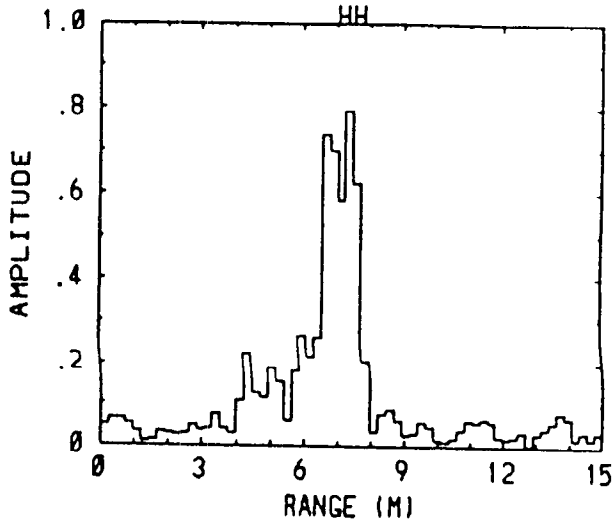


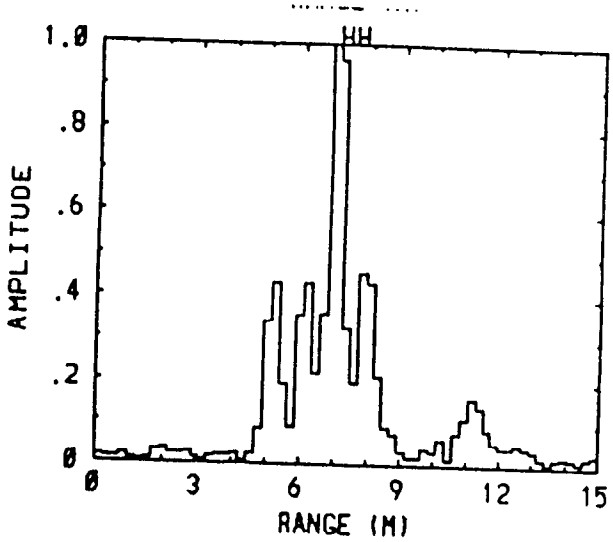
Figure 9



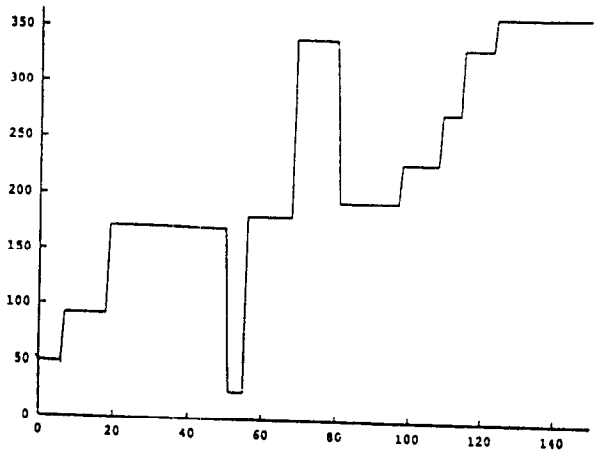
(a)



(b)



(c)



(d)

Figure 10