Dynamic Model Matching For Target Recognition From a Mobile Platform*

Hatem Nasr and Bir Bhanu

Honeywell Systems and Research Center
3660 Technology Drive
Minneapolis, Minnesota 55418

ABSTRACT

A novel technique called *dynamic model matching* (DMM) is presented for target recognition from a moving platform such as an autonomous combat vehicle. The DMM technique overcomes major limitations in present model-based target recognition techniques that use a single, static target model, and therefore cannot account for continuous changes in the target's appearance caused by varying range and perspective. DMM addresses this problem by combining a moving camera model, 3-D object models, spatial models, and expected range and perspective to generate multiple 2-D image models for matching. DMM also generates recognition strategies that can emphasize different object features at varying ranges. DMM operates within a larger system for landmark recognition based on the perception, reasoning, action, and expectation paradigm called PREACTE. Results are presented on a number of test sites using color video data obtained from the autonomous land vehicle.

1. INTRODUCTION

Target recognition from a mobile platform such as an autonomous combat vehicle in outdoor scenarios presents one of the most challenging problems of the machine vision community. It requires the ability to recognize targets from varying ranges and perspectives under changing environmental conditions. Earlier approaches emphasized the need for rotation-invariant and range-independent target models.[1,2] It was soon evident, however, that these models are weak because they have few parameters and cannot adequately handle different aspects at different ranges. Weak segmentation methods further aggravate the recognition problem.

Landmark recognition is a typical application of target recognition for autonomous vehicles. It is used to update the land navigation system that accrues a significant amount of error after the vehicle traverses long distances, which is typically the case in surveillance search and rescue missions. The vision system of the autonomous vehicle is required to recognize the landmarks as the vehicle approaches from the road or on terrain.

We have developed an expectation-driven, knowledge-based landmark recognition system called PREACTE that uses a priori, map and landmark knowledge, spatial reasoning, and a novel dynamic model matching (DMM) technique. PREACTE's mission is to predict and recognize landmarks and targets as the vehicle approaches them from different perspective angles at varying ranges. Once the landmarks have been recognized, they are associated with specific map coordinates, which are then compared to the land navigation system readings, and subsequent corrections are made. Landmarks of interest include buildings, gates, poles, and other man-made objects.

DMM departs from previous model-based and prediction-based vision systems[3,4] by addressing the following requirements:

- Target models are dynamic.
- Different targets require different representation and modeling techniques.
- Single targets require hybrid models.
- At different ranges, different matching and recognition plans need to be performed.

DMM generates and matches target landmark and map site descriptions dynamically based on different ranges and perspectives.

These descriptions are a collection of spatial, feature, geometric, and semantic models. From a given (or approximated) range and view angle, and using a priori map information, 3-D landmark models, and the camera model, PREACTE generates predictions about the individual landmark location in the 2-D image. The parameters of all models are a function of range and view angle. As the vehicle approaches the expected landmark, the image content changes, which in turn requires updating the search and match strategies. Landmark recognition in this framework is divided into three stages: detection, recognition, and verification. At far ranges, only "detection" of distinguishing landmark features is possible, whereas at close ranges, recognition and verification are more feasible, since more details of objects are observable.

In the following sections we present a brief description of PREACTE, details of DMM, and show results on real imagery. More details on PREACTE can be found in Nasr and Bhanu.[5,6]

2. CONCEPTUAL APPROACH

The task of visual landmark recognition in the autonomous combat vehicle scenario can be categorized as uninformed or informed. In the uninformed case, given a map representation, the vision system attempts to attach specific landmark labels to image objects of an arbitrary observed scene and infers the location of the vehicle on the map (world). In this case, image to map registration and spatial or topological information about the observed objects is typically used to infer their identity and the location of the robot on the map as a result. In the informed case, while the task is the same as before, there is a priori knowledge (with a certain level of certainty) of the past location of the robot on the map and its velocity. It is the informed case that is of interest in this paper.

Figure 1 illustrates the overall approach to PREACTE's landmark recognition task. It is a top-down, expectation-driven approach, whereby an expected site model (ESM) on the map is generated based on extensive domain-dependent knowledge of the current (or projected) location of the vehicle on the map and its velocity. The ESM contains models of the expected map site and its landmarks. These models provide the hypotheses to be verified by a sequence of images acquired at predicted times t, given the velocity of the vehicle and the distance between the current site and the predicted one. Figure 2 illustrates this concept. As shown, map site models introduce spatial constraints on the locations and distributions of landmarks, using a "road" model as a reference. Spatial constraints greatly reduce the search space while attempting to find a correspondence between the image regions and a model. This mapping is usually many-to-one in complex outdoor scenes because of imperfect segmentation.

The ESM is dynamic in the sense that the expectations and descriptions of different landmarks are based on different ranges and view angles. Multiple and hybrid landmark models are used to generate landmark descriptions as the robot approaches a landmark, leading to multiple model/image matching steps. This is what is referred to as dynamic model matching (DMM). The landmark descriptions are based on spatial, feature, geometric, and semantic models. There are two types of expectations: range dependent and range independent. Range-dependent

*This research was supported by DARPA under Contract No. DACA76-86-C-0017.

expectations are landmark features such as size, length, width, volume, etc. Range-independent ones include color, perimeter squared over area, length over width, shape, etc.

Different landmarks require different strategies and plans for detection and recognition at different ranges. For example, a yellow gate has a distinctive color feature that can be used to cue the landmark recognition process and reduce the search space. A telephone pole, on the other hand, requires the emphasis of the length/width feature.

In PREACTE, given the vehicle position in the map and an acquired image, PREACTE performs the following steps:

1. Generate 2-D descriptions from 3-D models for each landmark expected in the image.

2. Find the focus of attention areas (FOAAs) in the 2-D image for each expected landmark.

3. Generate the recognition plan to search for each landmark, which includes what features will be used for each landmark and at what range.

4. Generate the ESM at that range and aspect angle.

5. Search for regions in the FOAA of the segmented image that best match the features in the model.

6. Search for lines in the FOAA in the line image that best match the lines generated from the 3-D geometric model (this step is performed at close ranges where details can be observed).

7. Match expected landmark features with region attributes, and compute matching confidences for all landmarks.

8. Correct the approximated range by using the size differences of the suspected landmark in the current and previous frames.

9. Compute the uncertainty about the map site location based on the previous and current matching results.

2.1. MAP/LANDMARK KNOWLEDGE BASE

Extensive map knowledge and landmark models are fundamental to this recognition task. Our map representation relies heavily on declarative and explicit knowledge instead of procedural methods on relational databases. The map is represented as a quadtree, which in turn is represented in a hierarchical relational network. All map primitives are represented in a schema structure. The map dimensions are characterized by their cartographic coordinates. This schema representation provides an object-oriented computational environment that supports the inheritance of properties by different map primitives and allows modular and flexible means for searching the map knowledge base. The map sites between which the vehicle traverses have been surveyed and characterized by site numbers. A large database of information is available about these sites. This includes approximate latitude, longitude, elevation, distance between sites, terrain descriptions, landmark labels contained in a site, etc. Such site information is represented in a SITE schema, with corresponding slots. Slots names include HAS_LANDMARKS, NEXT_SITE, LOCATION, etc.

Each map site that contains landmarks of interest has an explicitly stored spatial model, which describes in 3-D the location of the landmarks relative to the road and to each other. By using a detailed camera model, range, and azimuth angle, we can generate 2-D views of the landmarks as shown in Figure 3. These views contain symbolic and numeric descriptions of the landmarks and their parts.

Given a priori knowledge of the vehicle's current location on the map space and its velocity, it is possible to predict the upcoming site that will be traversed through the explicit representation of map knowledge. The ESM contains information about the predicted (x,y) location of a given landmark and its associated FOAA, which is an expanded area around the predicted location of the object.

2.2. OBJECT MODELING

Landmark predictions are based on stored map information, object models, and the camera model. Each landmark has a hybrid model that includes spatial, feature, geometric, and semantic information. Figure 4 illustrates this hybrid model representation for a yellow gate; this model also includes:

- Map location
- Expected (x,y) location in the image
- Location with respect to the road (i.e., left or right) and approximate distance
- Location in 3-D

The feature-based model includes information about local features, such as color, texture, intensity, size, length, width, shape, elongation, perimeter squared over area, linearity, etc. The values of most of the range-dependent features, such as the size, length, width, etc., are obtained from the generated geometric model at that given range and azimuth angle. Range-independent feature values are obtained from visual observations and training data. Different parts of the yellow gate are represented in a semantic network.

3. DYNAMIC MODEL MATCHING

Each landmark has a number of dynamic models, as shown in Figure 1. The predicted landmark appearance is a function of the estimated range and view angle to the object. The range and view angle are initially estimated from prior locations of the vehicle, map information, and velocity; they can be corrected based on recognition results. The landmark recognition task is performed dynamically at a sampled clock rate. Different geometric models are used for different landmarks; for example, telephone poles can be best represented as generalized cylinders, whereby buildings are better represented as wire frames. The different representations require the extraction of different image features.

There are three basic steps to the landmark recognition process after generating the prediction of the next expected site and its associated landmarks. These are 1) landmark detection, 2) landmark recognition, and 3) map site verification and landmark position update on the map. At each stage, different sets of features are used.

Detection is a focus-of-attention stage; it occurs at ranges, say, greater than 35m. Very few details of landmarks (such as structure) can be observed; only dominant characteristics can be observed, such as color, size, elongation, straight lines, etc. From the map knowledge base, spatial information can be extracted, such as position of the landmarks with respect to the road (left or right) and position (in a 2-D image) with respect to each other (above, below, or between). So, using spatial knowledge abstracted in terms of spatial models and some dominant feature models, landmarks can be detected, but not recognized with a relatively high degree of confidence. However, this varies from one landmark to another; because some landmarks are larger than others, it may be possible to recognize them at such distances.

The second step, landmark recognition, occurs at closer ranges, say, 35 to 10m. At these ranges, most objects show more details and structure. Segmentation is more reliable, which makes it possible to extract lines and vertices. This in turn makes it possible to use detailed geometric models based on representations, such as generalized cylinders, wire frames, and winged edge, depending on the landmarks. Nevertheless, feature- and spatial-based information is still used prior to matching the geometric model to image content, because it greatly reduces the search space. We should note here that the feature and spatial models used in the first step are updated, because obviously the landmarks are perceived differently in the 2-D image at short ranges.

The third step is a verification stage that occurs at very close ranges. At this stage, PREACTE confirms or denies the existence of the landmarks and the map site location to the vehicle. Since subparts can be identified at close ranges for some landmarks, semantic models can be used to produce a higher degree of confidence in the recognition process. Some landmarks may partly disappear from the field of view (FOV) at this range. This information about the potential disappearance of objects from the FOV is obtained from the 3-D model of the landmark, the camera model, and the range.

Recognition plans are explicitly stated in the landmark model for different ranges, as shown below:

```
(defvar yellow-gate
  (make-instance     'object
    :name            'yellow-gate
    :parts           (list y-g-west-wing  y-g-east-wing)
    :geo-location    '(392967.4  1050687.7)
    :plan            '((40 15 detection) (15 8 recognition) (8 0
                       verification))
    :detection       '(color)
    :recognition     '(color length width area p2_over_area shape)
    :verification    '(color length width area p2_over_area  shape
                       lines) ))
```

Once the FOAA for a landmark is determined from the predicted model, all regions from the segmented image are matched against the landmark. More details on this matching technique can be found in Nasr and Bhanu.[5,6]

We compute the uncertainty U_S at each map site location in the following manner:

$$U_s = (U_{s-1} + \alpha D) * \prod_{i=1}^{m} \frac{0.5}{E(1_i)_{max}}$$

where U_S is the uncertainty at site s, U_{S-1} is the uncertainty at the previous site, L is the average accumulated error or uncertainty per kilometer of the vehicle navigation system, α is the number of kilometers traveled between the previous and the current site, and $E(l_i)_S$ is the evidence accumulated about landmark l_i at site s. U_S has a minimum value of zero, which indicates the lowest uncertainty and is the value at the starting point. The upper limit of U_S can be controlled by a threshold value and a normalization factor.

4. RESULTS

PREACTE and DMM were tested on a number of images collected by the vehicle. The PREACTE system was implemented on the Symbolics 3670. The image processing software was implemented in C on the VAX 11/750, and the image data were collected at 30 frames/sec. In this test, the robot started at map site 105 and headed south at 10 km/hr (see Figure 5). The objective of the test was to predict and recognize landmarks that were close to the road over a sequence of frames. Figures 6 through 20 show different stages of landmark recognition at different map locations. The figures show dynamic models generated by PREACTE at varying ranges. They also show how PREACTE changes recognition strategies at different stages of detection, recognition, and verification. In addition, the figures show the computed site uncertainty at each stage. The site uncertainty fluctuates depending on the landmark recognition results and the distance the vehicle has traveled.

In the future, we will extend this approach to the general situation in which the robot may be traveling through terrain and must determine precisely where it is on the map by using landmark recognition.

REFERENCES

1. P.J. Besl and R.C. Jain, "Three-Dimensional Object Recognition," ACM Computing Surveys, Vol. 17, No. 1, March 1985, pp. 75-145.
2. B. Bhanu, Editor, "CAD-Based Robot Vision," *IEEE Computer*, Special Issue on CAD-Based Robot Vision, August 1987.
3. R.T. Chin and C.R. Dyer, "Model-Based Recognition in Robot Vision," ACM Computing Surveys, Vol. 18, No. 1, March 1986, pp. 67-108.
4. R. Horaud and T. Skordas, "Model-Based Strategy Planning for Recognizing Partially Occluded Parts," *IEEE Computer*, Special Issue on CAD-Based Robot Vision, August 1987, pp. 58-64.
5. H. Nasr and B. Bhanu, "Landmark Recognition for Autonomous Mobile Robots," *Proceedings of the 1988 IEEE International Conference on Robotics and Automation*, April 1988.
6. H. Nasr, B. Bhanu, and S. Schaffer, "Guiding an Autonomous Land Vehicle Using Knowledge-Based Landmark Recognition," *Proceedings of the DARPA Image Understanding Workshop*, February 1987, pp. 432-439.

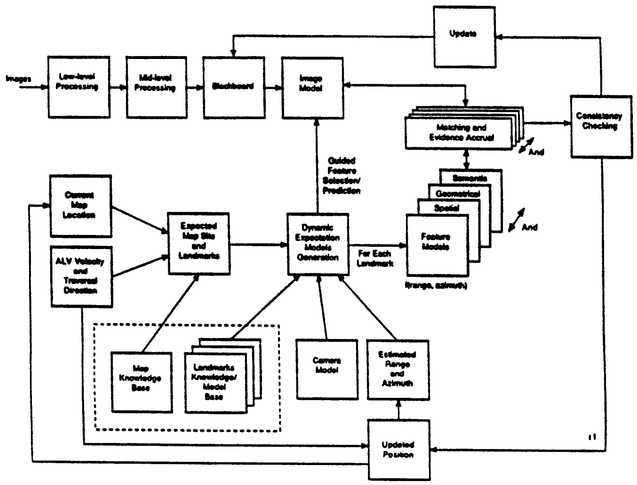


Figure 1. Detailed conceptual approach of PREACTE and DMM.

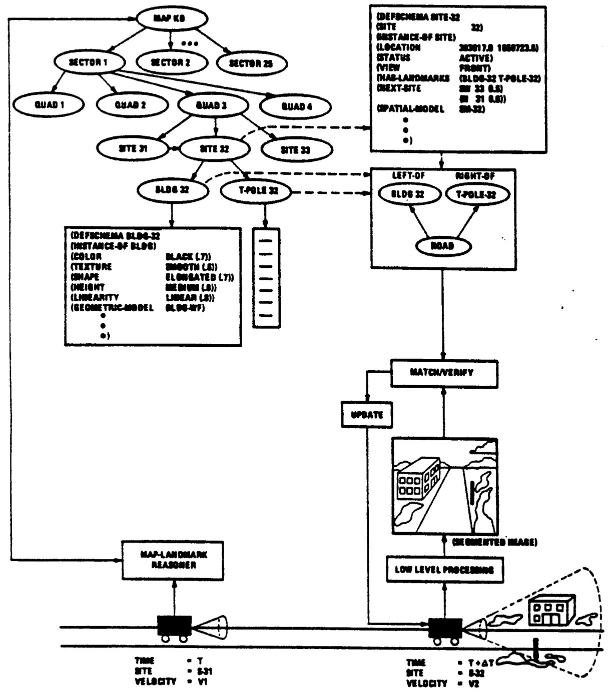


Figure 2. A graphic illustration of PREACTE's landmark recognition and map/landmark representation.

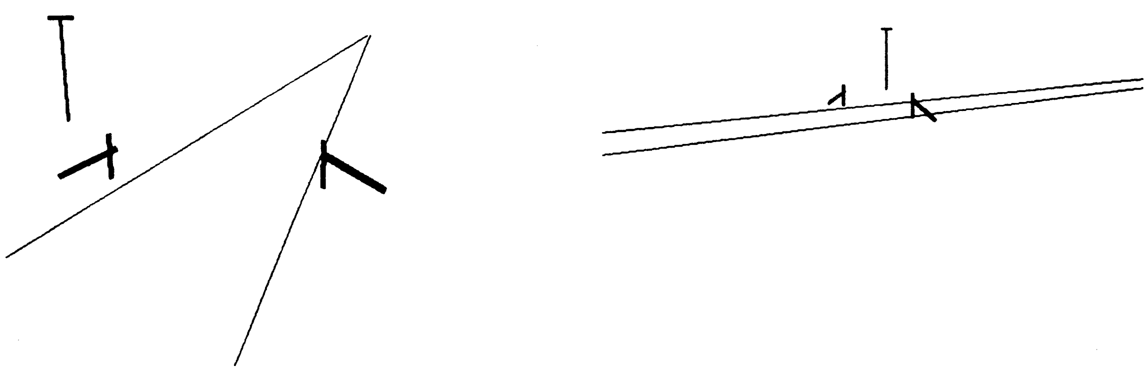


Figure 3. 2-D projections from different view angles and ranges.

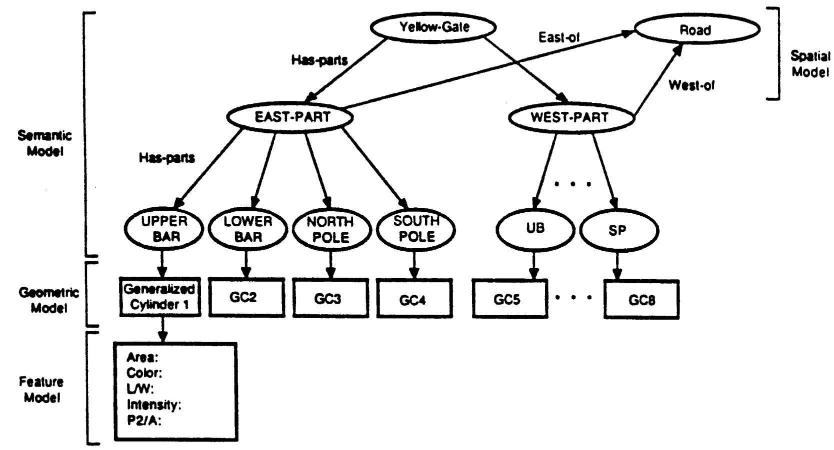


Figure 4. Hybrid model of the yellow gate landmark.

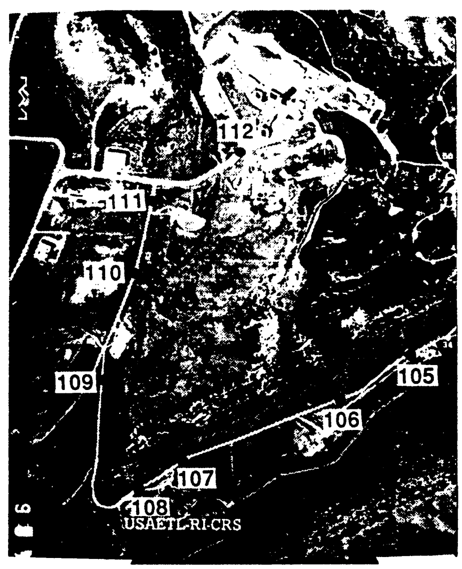


Figure 5. Aerial map photograph with selected sites for landmark recognition.

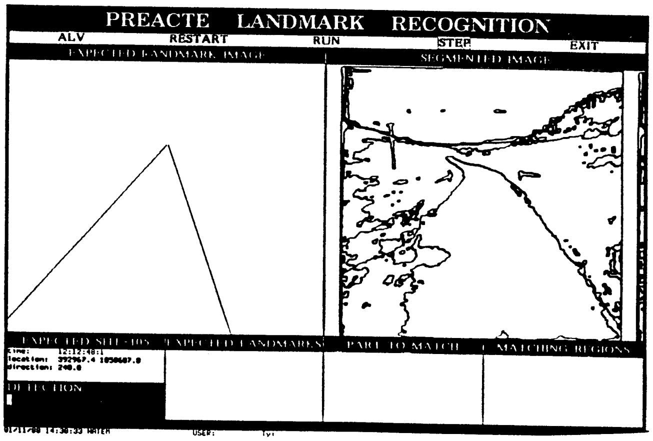


Figure 6. Site 105 is the next predicted map site. It contains a gate and a telephone pole. PREACTE projects a road model of the scene and an image is processed and segmented.

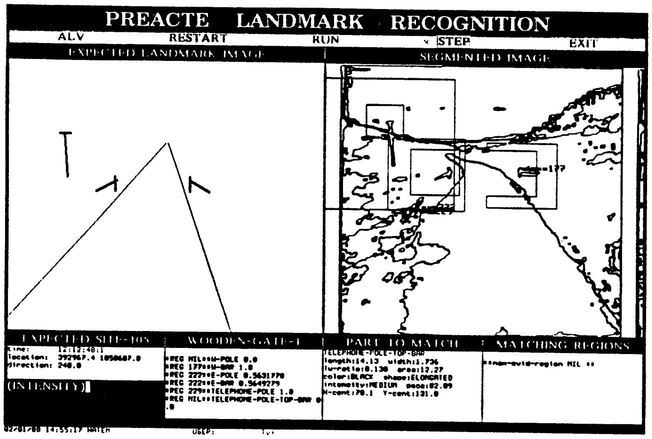


Figure 7. A 2-D model of the expected site is generated at the predicted range, and matching occurs. The "PART TO MATCH" pane shows descriptions of specific landmark parts as matching occurs. The intensity feature is emphasized in the lower left corner.

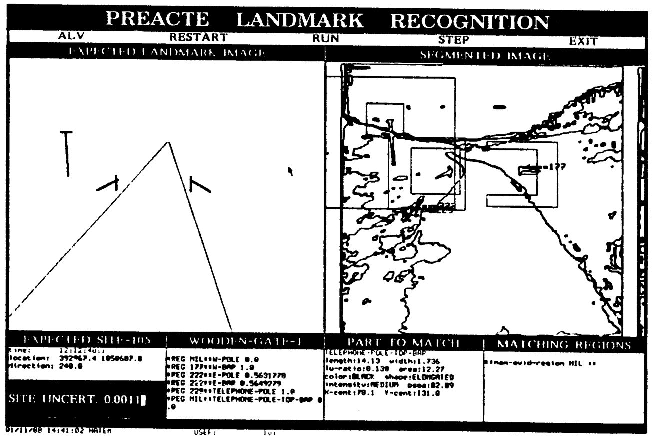


Figure 8. End of detection stage, with site uncertainty computed.

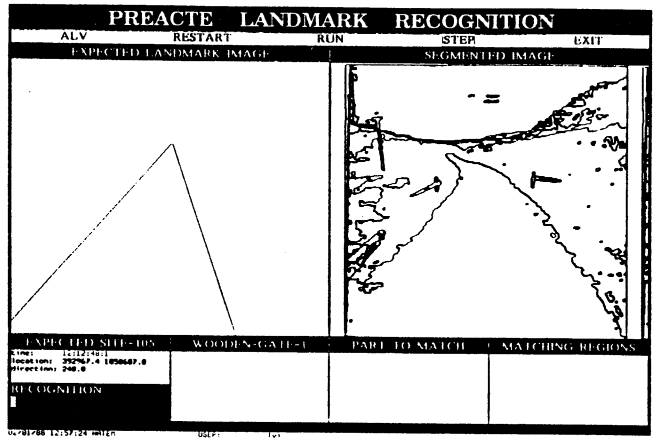


Figure 9. Beginning of recognition stage. New images are processed and a road model is projected.

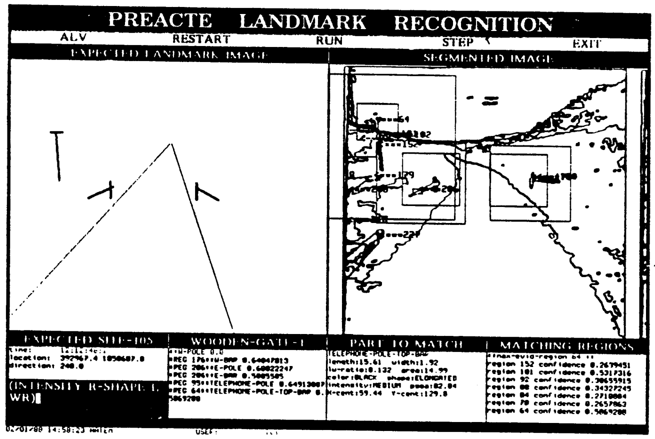


Figure 10. A new 2-D model of the scene is generated. More gate parts are identified. The rectangles over the segmented image indicate the FOAAs. The new model emphasizes a different set of features: intensity, length to width ratio, and a shape measure.

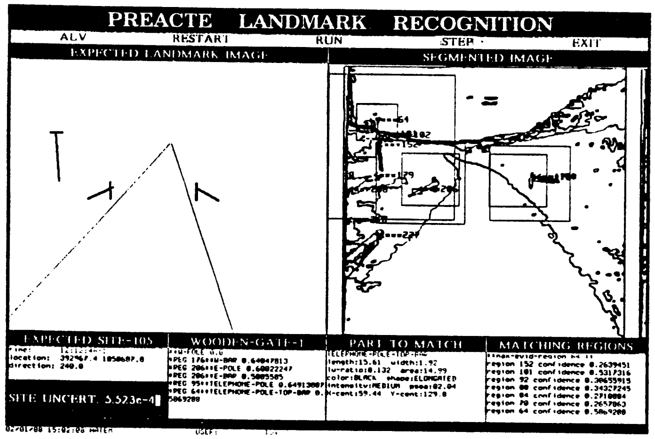


Figure 11. Site uncertainty has decreased because of additional positive evidences about the landmark.

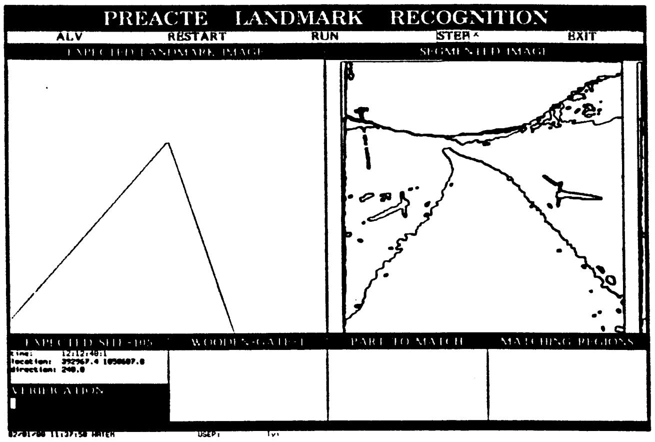


Figure 12. End of recognition stage and beginning of verification stage.

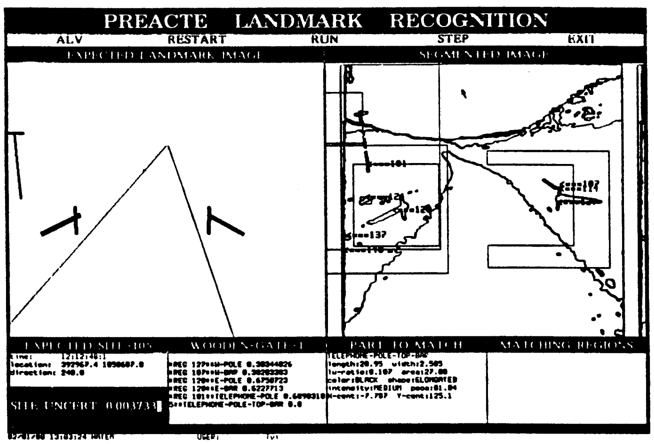


Figure 13. Site uncertainty is computed at this verification stage. The uncertainty has slightly increased because of higher matching requirements.

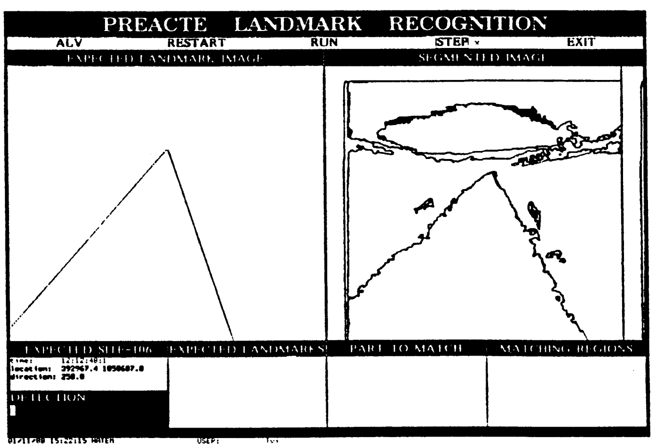


Figure 14. The vehicle arrives at a new site, which contains a yellow gate.

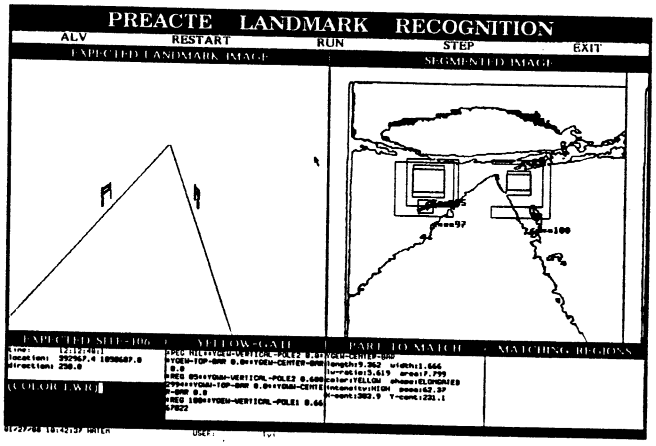


Figure 15. Only the left part of the gate is detected. The right part of the gate falls mostly outside the FOAA.

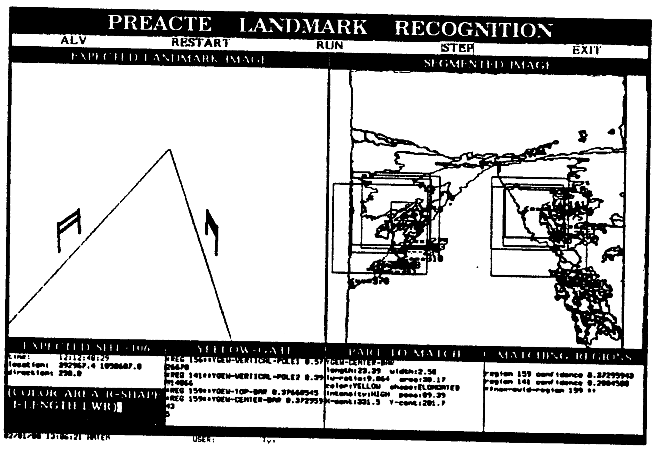


Figure 16. Five features are emphasized at this recognition stage: color, size, shape, length, and LWR.

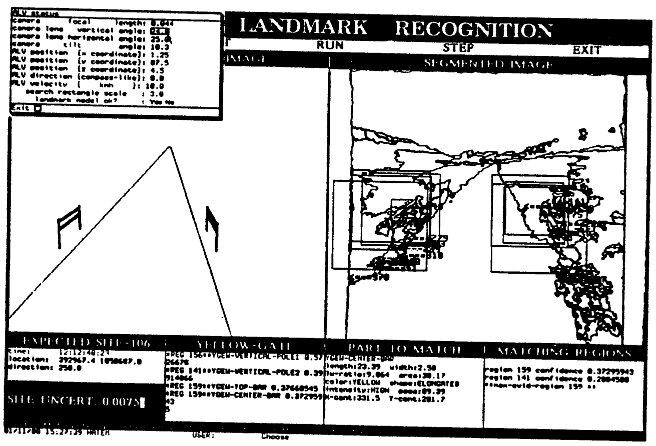


Figure 17. The box in the upper left corner shows the different camera parameters, which can be modified automatically or manually.

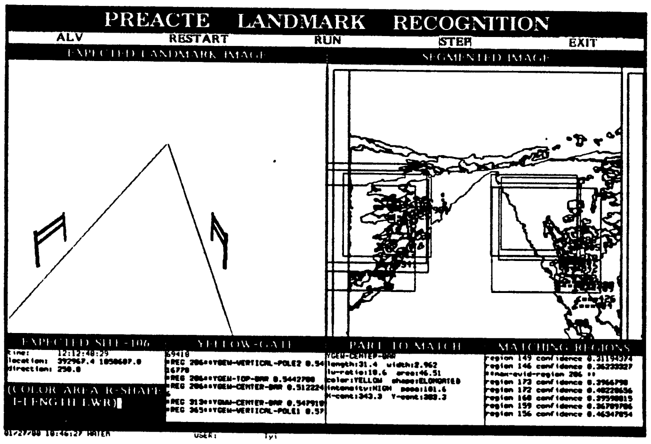


Figure 18. The gate at closer range, still at a recognition stage. Matching results have degraded because of bad segmentation results. The "Matching Regions" pane shows the candidate regions for a given part of the gate.

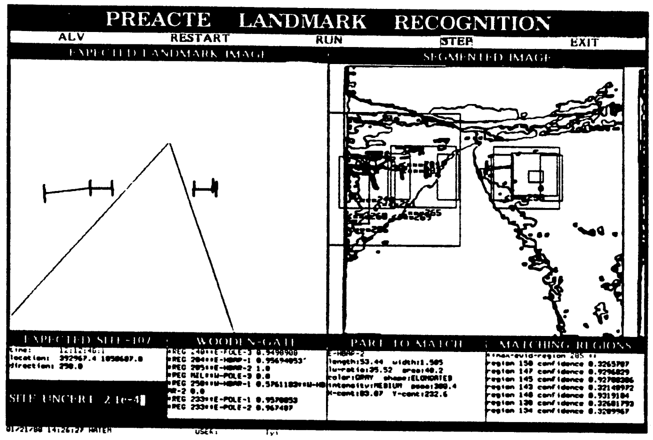


Figure 19. The vehicle arrives at another site that contains another gate.

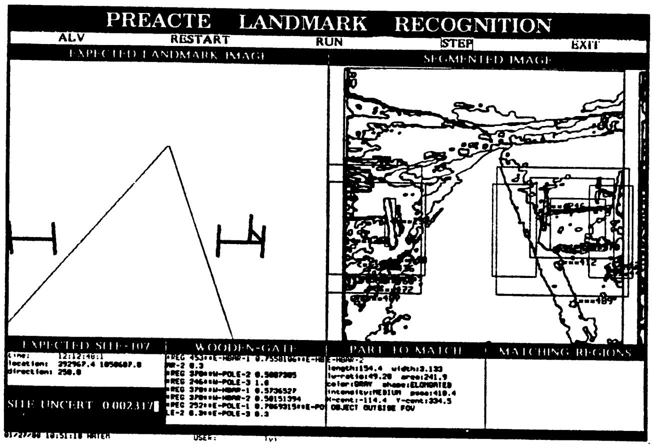
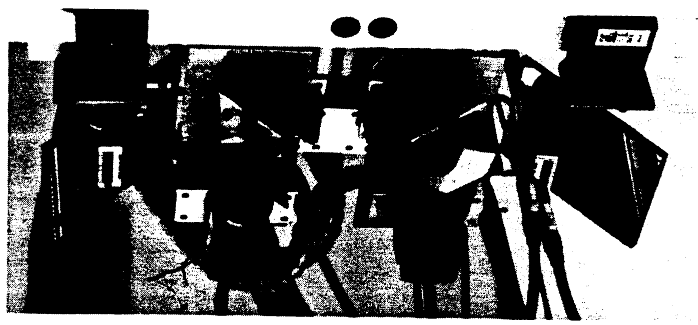


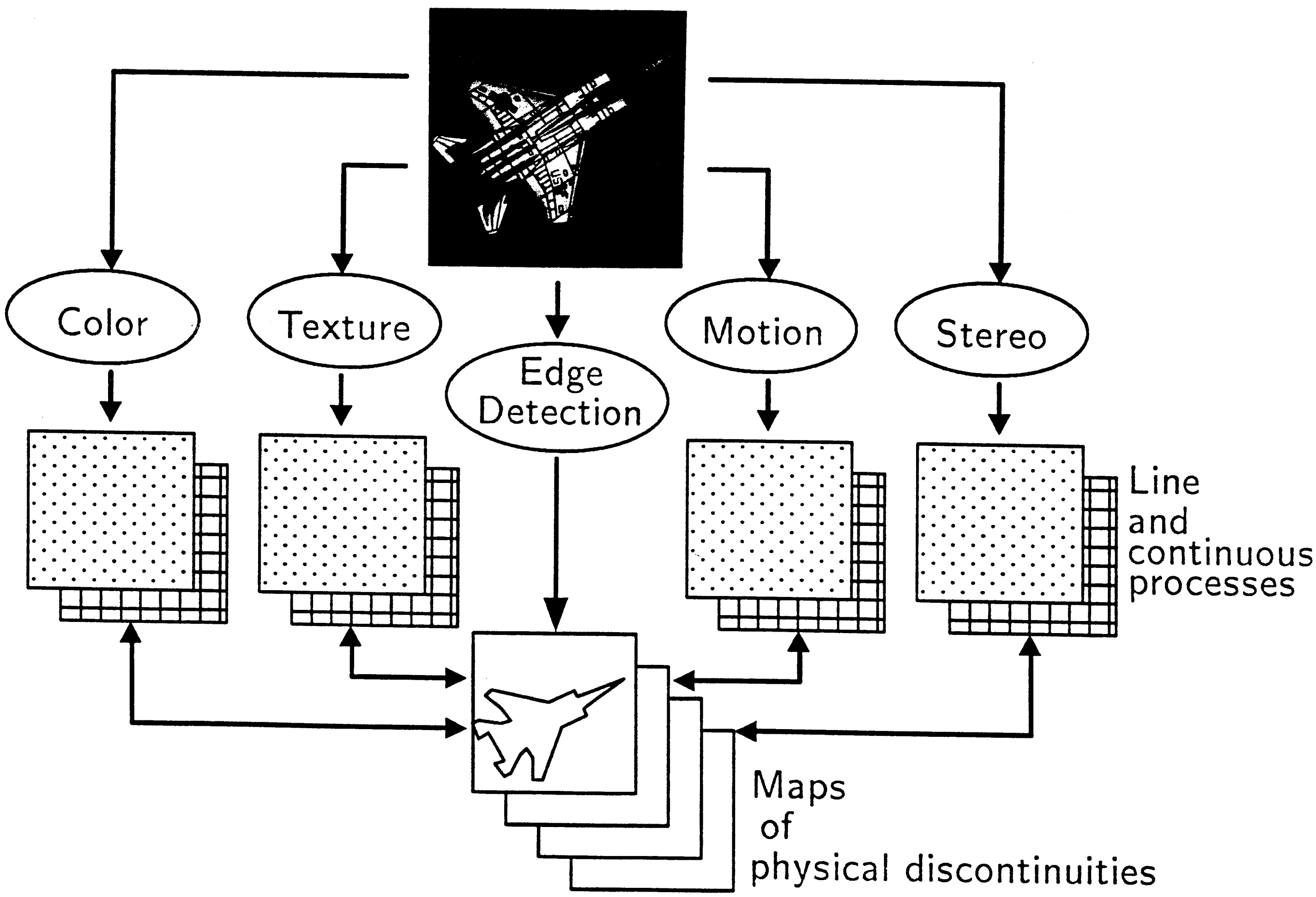
Figure 20. The gate model at very close range. The uncertainty value is still at an acceptable range. DMM predicts that some gate parts are outside the field of view and avoids matching them.

Volume II

PROCEEDINGS:

Image Understanding Workshop





Sponsored by:

Defense Advanced Research Projects Agency
Information Science and Technology Office

DARPA

April 1988

Image Understanding Workshop

Proceedings of a Workshop
Held at
Cambridge, Massachusetts

April 6-8, 1988

Volume II

Sponsored by:

Defense Advanced Research Projects Agency
Information Science and Technology Office

This document contains copies of reports prepared for the DARPA Image Understanding Workshop. Included are results from both the basic and strategic computing programs within DARPA/ISTO sponsored projects.