LANDMARK RECOGNITION FOR AUTONOMOUS MOBILE ROBOTS*

Hatem Nasr and Bir Bhanu

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

A new approach for landmark recognition based on the perception, reasoning, action, and expectation (PREACTE) paradigm is presented for the navigation of autonomous mobile robots. PREACTE uses expectations to predict the appearance and disappearance of objects, thereby reducing computational complexity and locational uncertainty. It uses an innovative concept called dynamic model matching (DMM), which is based on the automatic generation of landmark description at different ranges and aspect angles and uses explicit knowledge about maps and landmarks. Map information is used to generate an expected site model (ESM) for search delimitation, given the location and velocity of the mobile robot. The landmark recognition vision system generates 2-D and 3-D scene models from the observed scene. The ESM hypotheses are verified by matching them to the image model. Experimental results that verify the performance of the PREACTE and DMM algorithms for real imagery are also presented.

1. Introduction

To accomplish missions such as surveillance and search and rescue, a mobile robot has to travel long distances. This results in a significant amount of positional error in the land navigation system. Landmark recognition is used to update the land navigation system by recognizing the observed objects in the scene and associating them with specific landmarks on a geographic map, thus enabling the robot to remain on course. Landmarks of interest include mostly man-made objects, such as telephone poles, storage tanks, buildings, houses, gates, etc.

Unlike previous related work, the paradigm of an intelligent agent (like an autonomous mobile robot) that we use here is based on a perception, reasoning, action, and expectation (PREACTE) cycle. We have developed an expectation-driven, knowledge-based landmark recognition system, called PREACTE, that uses a priori, map, and perceptual knowledge; spatial reasoning and knowledge aggregation; and novel dynamic model-matching (DMM) methods. In contrast to the work of Davis,[1] explicit knowledge about the map and landmarks is assumed to be given. This knowledge is used to generate an expected site model (ESM), given the robot's location and velocity. 3-D models of landmarks at a particular map site are stored in heterogeneous representations. Using these 3-D models, the vision system generates many 2-D scene models as a function of estimated range and azimuth angle. The ESM hypotheses are dynamically verified by matching them to the abstracted image models. This matching is accomplished by using grouping of segments (lines and regions) and spatial reasoning. Positive as well as negative evidences are used to verify the existence of each landmark in the scene. The system also provides feedback control to the low-level processes to permit parameter adaptation of the feature detection algorithms to changing illumination and environmental conditions.

PREACTE emphasizes model-based vision, which has been a popular paradigm in computer vision because it reduces computational complexity and requires no learning. Binford has summarized model-based vision work[2] and described several systems, including the work of Brooks on ACRONYM,[3] Hanson and Riseman's work on VISIONS,[4] and Nagao and Matsuyama's work on the analysis of complex aerial photographs.[5] McKeown et al. have used map- and domain-specific knowledge in SPAM rule-based systems for the interpretation of airport scenes in aerial images.[6]

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

Hwang has also used domain knowledge to guide interpretation of suburban house scenes in aerial imagery.[7] He has used a test-hypothesize-act sequence to generate many hypotheses, which are then integrated into a consistent interpretation. Bhanu has used several modeling and relaxation matching techniques for the recognition of 2-D and 3-D nonoccluded and occluded objects.[8-12]

In the DMM concept mentioned above, object/landmark descriptions are generated dynamically based on different ranges and view angles. 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 robot 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, "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 details of the PREACTE and DMM concepts and results on real images taken by an autonomous mobile robot.

2. Conceptual Approach

The task of visual landmark recognition in the autonomous mobile robot 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, 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 ESM on the map is generated based on extensive domain-dependent knowledge of the current (or projected) location of the robot 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 a predicted time t, given the velocity of the robot 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 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.

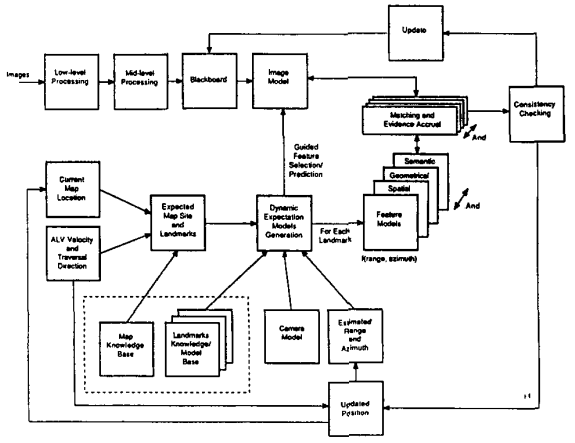

Figure 1. Detailed conceptual approach of PREACTE and DMM

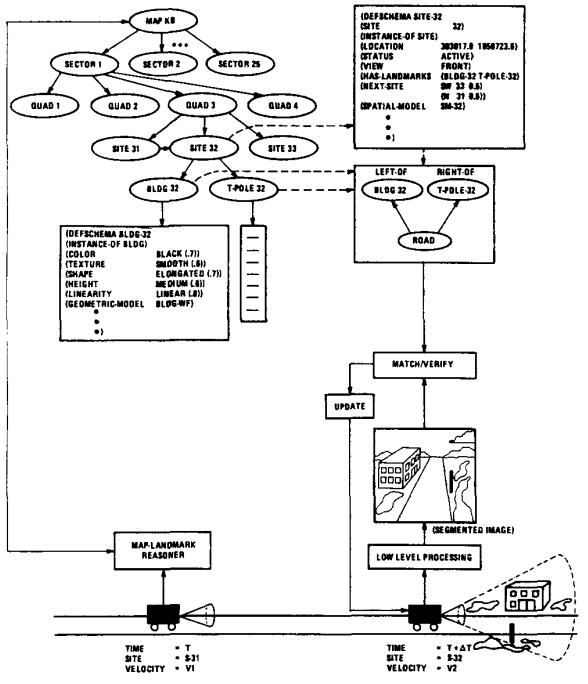

Figure 2. A graphic illustration of PREACTE's landmark recognition and map/landmark representation

In PREACTE, given an image, landmark recognition basically consists of 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 in a given map site
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 evidences 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

In the segmented image, features such as size, texture, color, etc. of each region provide independent evidence for the existence of an expected landmark. Evidence accrual is accomplished by an extension of a heuristic Bayesian formula,[13-14] which will be discussed in section 4. The heuristic formula is used to compute the certainty about a map site location based on the certainty of the previous site and the evidence of the existence of each landmark at the current site.

2.1. Map/Landmark Knowledge Base

Extensive map knowledge and landmark models are fundamental to the recognition task. Our map representation relies heavily on declarative and explicit knowledge instead of procedural methods on relational databases.[6] The map is represented as a quadtree, which in turn is represented in a hierarchical relational network. All map primitives are represented in a schematic structure. The map dimensions are characterized by their cartographic coordinates. This schematic 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.

Given a priori knowledge of the robot'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. Image Modeling

Following image segmentation, a number of image features are extracted for each region, such as color, length, size, perimeter, texture, minimum bounding rectangle (MBR), etc., as well as some derived features, such as elongation, linearity, compactness, etc. All image information is stored in a blackboard. Symbolic feature extraction is performed on some of the region-based features. So, instead of having area = 1500 pixels and

intensity = 52, we could have area = large and intensity = low. The symbolic characterization of the features using "relative" image information provides a better abstraction of the image and a framework for knowledge-based reasoning. On one hand, this has the advantage of making the feature space smaller and therefore easier to manipulate. On the other hand, it makes the feature space insensitive to feature variations in the image; this is why numeric features are also preserved.

Each set of region features is represented in a schematic structure instead of a feature vector. This schematic representation of regions does not have any conceptual justifications; however, it provides a compatible data structure with the landmark models in the knowledge base. Most of the region features have representative attributes in the landmark models. This allows symbolic pattern matching to be performed easily. Beyond that, it makes the reasoning process more traceable.

A critical region in the image is the road region, which is used as a reference in the image model. In most cases, the road is easily segmented out, assuming it is a "structured" road that provides good contrast (i.e., an asphalt or concrete road, not a dirt road). The road is represented in the model by its vertices and the approximate straight lines of its left and right borders.

2.3. Object Modeling

Landmark expectations 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 3 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

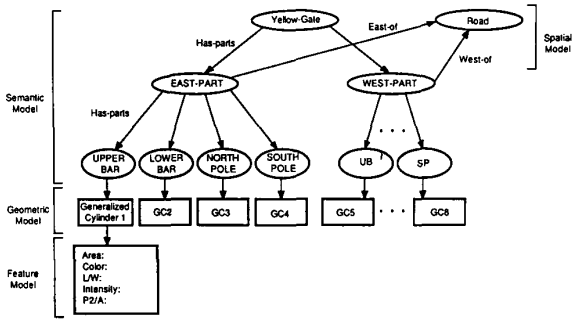

Figure 3. Hybrid model of the yellow gate landmark

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. The geometric model is landmark dependent, and its parameters are range dependent. Different parts of the yellow gate are represented in a semantic network. The geometry of the gate parts in the image is the result of the 3-D projection on a 2-D plane using a camera model, given a certain range and azimuth.

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 robot, 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 45m. 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, 20 to 45m. 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 edges, 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 robot. 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) ))
```

4. Evidence Accumulation and Map Location Uncertainty

Given a set of regions {R} in the image that satisfies the spatial constraints of the FOAA imposed by landmark l_i in the ESM (there is usually more than one corresponding region), we compute the evidence $E(l_i)$ using the FIND_EVIDENCE algorithm that each r_j in {R} yields. The r_j that results in $E(l_i)_{max}$ (provided it is a positive evidence) is considered the best-match candidate for l_i. Then the individual set of evidences $E(l_i)_{max}$ is aggregated, and the certainty level about the current map site location is computed.

The FIND_EVIDENCE algorithm considers that each landmark l_i in the ESM has a set of attributes $\{A_{i1}, ..., A_{ik}, ..., A_{in}\}$, each with a likelihood LH_{ik}. Each region r_j in {R} has a set of features $\{f_{j1}, ..., f_{jk}, ..., f_{jn}\}$. Note that A_{ik} and f_{jk} correspond to the same feature (in the model and the

image), such as color, size, texture, etc. Given these features, we compute the evidence that l_i is present in the image by using a heuristic Bayesian formula, given by:

$$P(l_i/f_{j1}, ..., f_{jk}, ..., f_{jn}) = \frac{P(l_i) * P(f_{j1}, ..., f_{jk}, ..., f_{jn}/l_i)}{P(f_{j1}, ..., f_{jk}, ..., f_{jn})} \quad (1)$$

By making the independence assumption among features, the above equation can be rewritten as:

$$P(l_i/f_{j1}, ..., f_{jk}, ..., f_{jn}) = \frac{P(l_i)*P(f_{j1}/l_i)*...*P(f_{jk}/l_i)*...*P(f_{jn}/l_i)}{P(f_{j1})*...*P(f_{jk})*...*P(f_{jn})} \quad (2)$$

$$= P(l_i) * \prod_{k=1}^{n} \frac{P(f_{jk}/l_i)}{P(f_{jk})}$$

where n is the number of features and $P(l_i)$ is the initial probability of a landmark being found at a given site. $P(l_i)$ is initially equal to 1 for all landmarks. For example, if texture can take either of the four values: coarse, smooth, regular, or irregular, then P (texture = smooth) = 1/4. Finally,

$$P(f_{jk}/l_i) = \begin{cases} LH_{ik} & \text{if } f_{jk} = A_{ik} \\ \frac{1 - LH_{ik}}{d(f_{jk}, A_{ik})} & \text{if } f_{jk} \neq A_{ik} \end{cases} \quad (3)$$

This is best explained through the following example. Assume two regions r_1 and r_2 in the image with different sizes (f_{jk}), SIZE (r_1) = SMALL and SIZE (r_2) = LARGE. Assume a model of landmark L, with the expected size (A_{ik}) to be LARGE and with a likelihood (LH_{ik}) of 0.7. The SIZE feature can take any of the following ordered values: {SMALL, MEDIUM, LARGE}. If r_2 is being matched to L, equation 3 yields 0.7 because $f_{jk} = A_{ik}$. On the other hand, if r_1 is being matched to L, then equation 3 yields (1-0.7)/2. The denominator 2 is used because LARGE is two unit distances (denoted by $d(f_{jk}, A_{ik})$) from SMALL. We rewrite equation 2 as:

$$P(l_i/f_{j1}, ..., f_{jk}, ..., f_{jn}) = P(l_i) * \prod_{j=1} I(f_{jk}/l_i) \quad (4)$$

where $I(f_{jk}/l_i)$ is the term within the product sign. The value of $I(f_{jk}/l_i)$ can be greater than 1, because the heuristic nature of the formulation does not reflect a probabilistic set of conditional events, as formulated in Bayes theory. Moreover, $P(l_i/f_{j1}, ..., f_{jk}, ..., f_{jn})$ can result in a very large number or a very small positive number. By taking the logarithm of both sides of equation 4 we have:

$$Log[P(l_i/f_{j1}, ..., f_{jk}, ..., f_{jn})] = Log[P(l_i)] + \frac{\sum_{j=1}^{n} Log[I(f_{jk}/l_i) * W_j]}{n} \quad (5)$$

where W_j is a normalization factor between 0 and 1.

Next we define the evidence terms E and e to be the logarithms of P and I*W, respectively, assuming $P(l_i)$ is initially equal to 1. So, the evidence formula can be written as follows:

$$E(l_i) = \frac{\sum_{j=1}^{n} e(f_{jk}/l_i)}{n} \quad (6)$$

The values of $E(l_i)$ fall between 0 and 1. If $E(l_i) > 0.50$, the evidence is "positive." On the other hand, if $E(l_i) < 0.3$, the evidence is interpreted as "negative" or "weak." Otherwise, $E(l_i)$ is characterized as "neutral."

4.1. Negative Evidences

An important feature incorporated into PREACTE is the use of positive as well as negative evidences to verify its expectations. There are many types of negative evidences that could be encountered during the hypothesis generation and verification process. One that is of particular interest to us is highly negative evidence (for example, < 0.3) about a "single" landmark in conjunction with very positive evidences about the other landmarks and a reasonable level of certainty about the previous site. This case may be caused by one or more of the following:

- Error in the dimension of the expectation zone
- Bad segmentation results
- Change in the expected view angle or range

In such a case, PREACTE would enlarge the expectation zone by a fixed margin and find the evidences introduced by the new set of regions. If this fails to produce an admissible set of evidences, then the expectation zone of the image is resegmented using a new set of parameters that are strictly object dependent.

4.2. Map Location Uncertainty

Even though landmark recognition is introduced to assist the autonomous robot's land navigation system, uncertainty is obviously attached to the results of the recognition system. We compute the uncertainty U_S at each site location in the following manner:

$$U_s = (U_{s-1} + \alpha D) * \prod_{i=1}^{m} \frac{0.5}{E(l_i)_{max}} \quad (7)$$

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 robot 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.

5. Results

We have implemented a prototype system in Common Lisp on the Symbolics 3670. The image processing software was implemented in C on the VAX 11/750. The Symbolics 3670 hosts all of the PREACTE software.

PREACTE was tested on a number of images collected by the robot. 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 4). The objective of the test was to predict and recognize landmarks that were close to the road over a sequence of frames. Figures 5 through 7 show landmark recognition of the yellow gate; parts of the gate were correctly identified at different ranges.

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. E. Davis, *Representing and Acquiring Geographic Knowledge*, Morgan Kaufman Publishers, Inc., 1986.

2. T.O. Binford, "Survey of Model-Based Image Analysis," *The International Journal of Robotics Research*, Vol. 1, Spring 1982, pp. 18-64.

3. R.A. Brooks, "Symbolic Reasoning among 3-D Models and 2-D Images," *Artificial Intelligence*, Vol. 17, 1981, pp. 285-348.

4. A.R. Hanson and E.M. Riseman, "VISIONS: A Computer System for Interpreting Scenes," in *Computer Vision Systems*, A.R. Hanson and E.M. Riseman (eds.), New York: Academic Press, 1978, pp. 303-333.

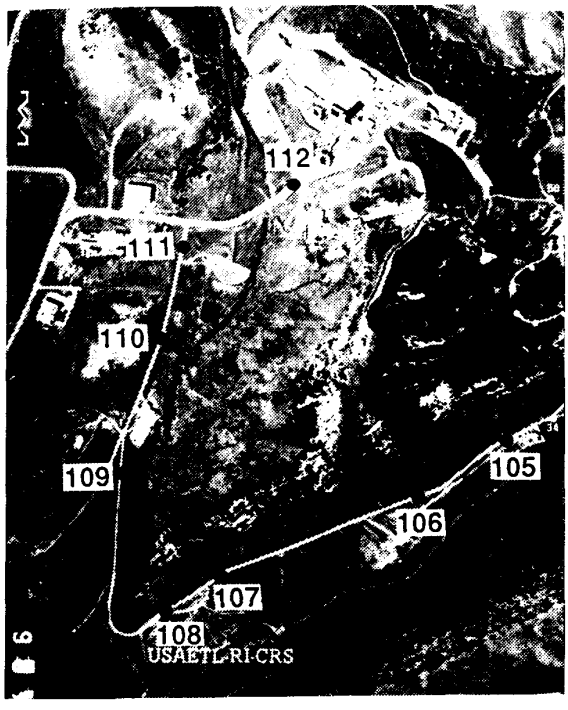

Figure 4. Aerial map photograph with selected sites for landmark recognition

5. M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, 1980.

6. D.M. McKeown, Jr., W.A. Harvey, Jr., and J. McDermott, "Rule-Based Interpretation of Aerial Imagery," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, September 1985, pp. 570-585.

7. S.V. Hwang, "Evidence Accumulation for Spatial Reasoning in Aerial Image Understanding," Ph.D. Thesis, Department of Computer Science, University of Maryland, College Park, 1984.

8. B. Bhanu, "Recognition of Occluded Objects," *Proc. of the 8th International Joint Conference on Artificial Intelligence*, Vol. IJCAI-83, Karlsruhe, West Germany, August 8-12, 1983, pp. 1136-1138.

9. B. Bhanu, "Representation and Shape Matching of 3-D Objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, May 1984, pp. 340-351.

10. B. Bhanu and O.D. Faugeras, "Shape Matching of Two-Dimensional Objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, March 1984, pp. 137-156.

11. B. Bhanu and T. Henderson, "CAGD-Based 3-D Vision," *Proc. of the IEEE International Conference on Robotics and Automation*, March 1985, pp. 411-417.

12. B. Bhanu and W. Burger, "DRIVE: Dynamic Reasoning Using Integrated Visual Evidences," *Proc. of the DARPA Image Understanding Workshop*, University of Southern California, February 1987.

13. E. Charniak, "The Bayesian Basis of Common Sense Reasoning in Medical Diagnosis," *Proc. of the American Association of Artificial Intelligence Conference 1983*, Vol. AAAI-83, pp. 70-73.

14. D. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Publishing Co., 1985.

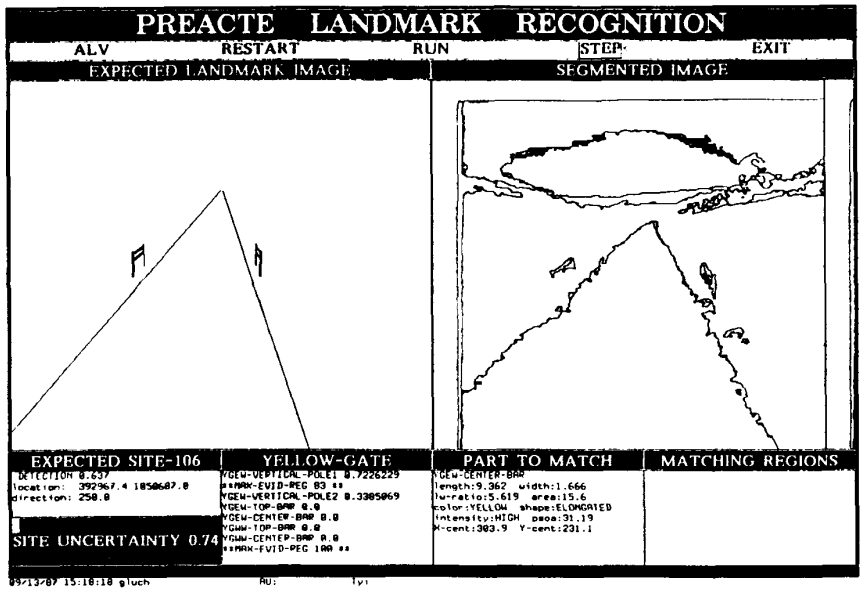

Figure 5. Expected model of the yellow gate (left); segmented image (right); detection results at a 20m range

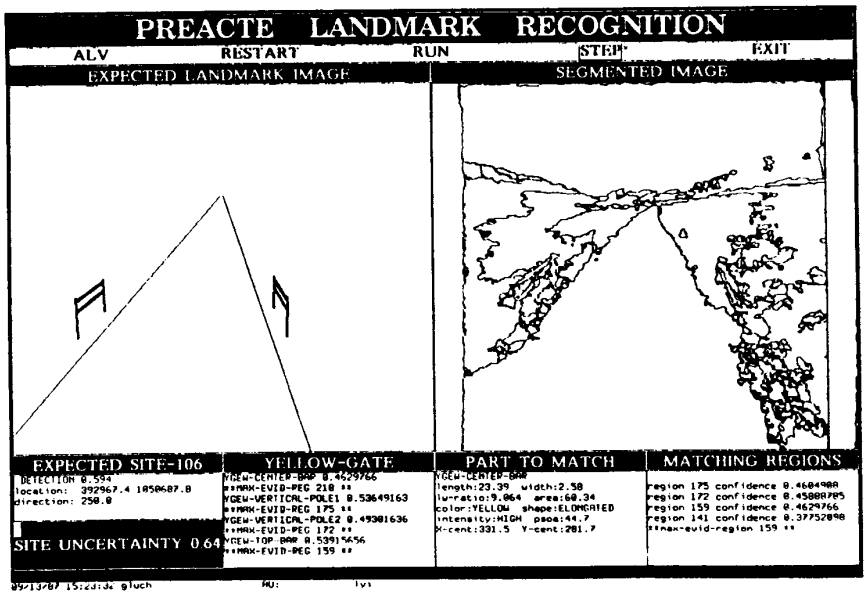

Figure 6. Updated gate model at a closer range of 12m with further detection results

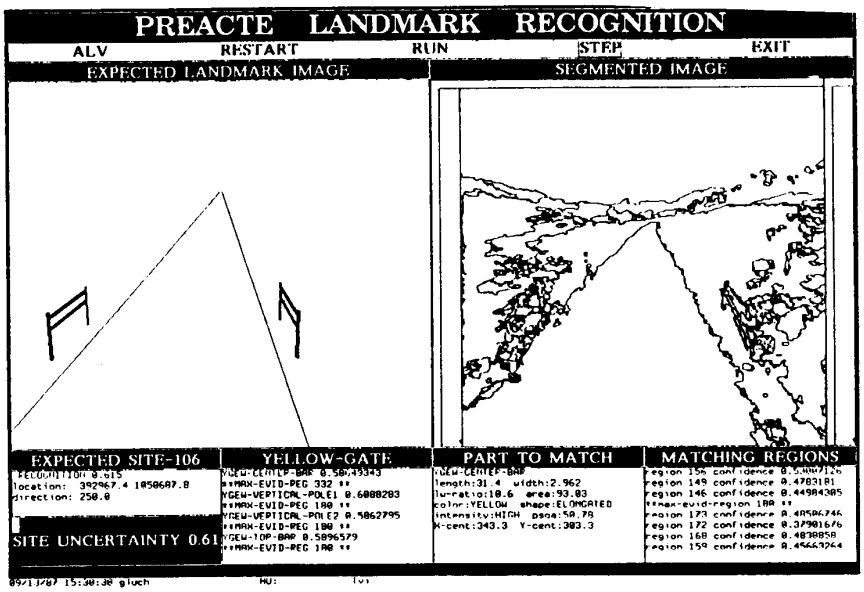

Figure 7. Recognition results at a range of 8m