Model-Based Object Recognition Using Laser Radar Range Imagery

Asuman E. Koksal[a], Jeffrey H. Shapiro*[a], and William M. Wells, III[b]

Massachusetts Institute of Technology
[a]Research Laboratory of Electronics, and Laboratory for Information and Decision Systems
[b]Artificial Intelligence Laboratory
Cambridge, MA 02139 USA

ABSTRACT

The combined effects of laser speckle and local oscillator shot noise degrade coherent laser radar range measurements. As a result, laser radar range imagery suffers from both uniformly-distributed range anomalies and Gaussian-distributed local range errors. Our goal in this research is to develop a target recognition system capable of recognizing military vehicles in range images provided by airborne laser radars. In particular, we will focus on using laser radar range imagery in a statistical model-based object recognition system. We shall present performance results for our object recognition system using laser radar data from the MIT Lincoln Laboratory Infrared Airborne Radar (IRAR) data release together with 3-D CAD models which account for the possible military targets that may be present on the site imaged by the laser radar.

Keywords: Automatic target recognition, laser radar, model-based object recognition.

1. INTRODUCTION

A coherent laser radar can produce range, intensity or Doppler images in 3-D pulsed imager mode.[1,2] The laser radar transmits a series of laser pulses, one for each pixel in a raster scan. The reflected light for each pixel then undergoes optical heterodyne detection followed by intermediate frequency (IF) filtering, video and peak detection, as shown in Fig. 1. The range image of some field of view is formed by measuring the time delay between the peaks of the transmitted and the detected waveforms. Laser radar range images are degraded by the combined effects of laser speckle and local oscillator shot noise. The former is due to rough-surfaced nature of encountered objects when compared to the laser wavelength, which causes constructive and destructive interference in the reflected light.[3] The latter is the fundamental noise involved in optical heterodyne detection.[4] Speckle degrades range images through range anomalies, which occur when a deep speckle fade combines with a strong noise peak resulting in a range measurement substantially different from the true range value.

Collectively, these degradation mechanisms suggest a statistical approach to laser radar image processing. There has been a long interest in the statistics of peak-detecting coherent laser radars. Combining theory, experiments and computer simulations, a significant degree of understanding about the characteristics of laser radar has been achieved.[5–8]

Recognizing 3-D objects from range imagery has received considerable attention in the last few years. However, most approaches followed so far are applicable to high resolution and high range precision images. In this paper we present a statistically optimum approach for doing model-based object recognition using low resolution, noise-degraded range images. Our objective is to build an end-to-end system that operates in an autonomous fashion using raw sensor images as its input and making a recognition decision at the output. We used a modular building-block approach, choosing to use modules that were faithful to the input-data statistics and which relied on explicitly-modeled uncertainties at each stage in the overall recognition system. Toward this end, we used an input stage comprising an essentially optimal noise-anomaly suppression scheme called the fast maximum-likelihood/expectation-maximization (ML/EM) algorithm.,[9] and for the object-recognition output stage we employed a model-based alignment/scoring technique known as posterior marginal pose estimation (PMPE).[10] We explored the recognition performance of our system by utilizing real laser radar range images from the MIT Lincoln Laboratory Infrared Airborne Radar (IRAR) Data Release provided by MIT Lincoln Laboratory.[11,12]

*Send correspondence to J.H. Shapiro, Email: jhs@mit.edu; Phone: 617-253-4179; FAX: 617-258-7336.

Research supported by U.S. Army Research Office Grant DAAH04-95-01-0494.

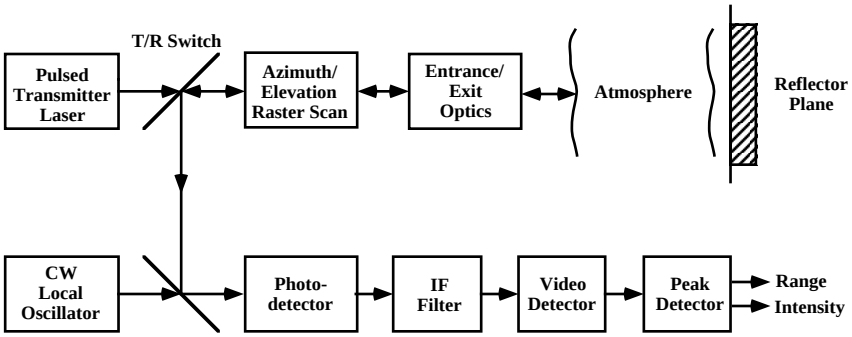

Figure 1. Block diagram for a coherent laser radar operated as a pulsed imager.

2. SYSTEM MODEL

In an object recognition system that uses an imaging sensor a significant source of target-signature uncertainty is the variability of target position and orientation with respect to the sensor. The position and orientation of an object in an image is usually referred to as its "pose." This makes the pose estimation problem a major focus in the recognition problem. Our goal, however, was to do recognition in the general sense, i.e., our algorithm was supposed to deduce the identity of the target in the image among possible candidates in a database. Therefore the first step in developing the system was forming a data library involving 3-D CAD models representing the military targets might have been present on the site imaged by the laser radar. The basic idea of model-based recognition is to score the match between each target model and the sensor image, declaring the highest-scoring model to represent the object that is present.

The key components of the overall object recognition system are summarized in Fig. 2. The noisy, low resolution laser radar range image is first processed to extract the finest scale information adequately supported by the raw observation data while simultaneously suppressing the range anomalies. The resulting preprocessed range image is then segmented to isolate target from background as accurately as possible. The purpose of the feature extraction module that follows is to select the essential characteristics needed to identify the target in the segmented image. In order to accurately model arbitrarily shaped objects, images are represented by edge-contour features. These edge-feature data are submitted to the recognition stage to determine whether or not the target in the image corresponds to an object known to the object recognition system.

The 3-D CAD models representing distinct objects that may be present in the scene undergo some off-line

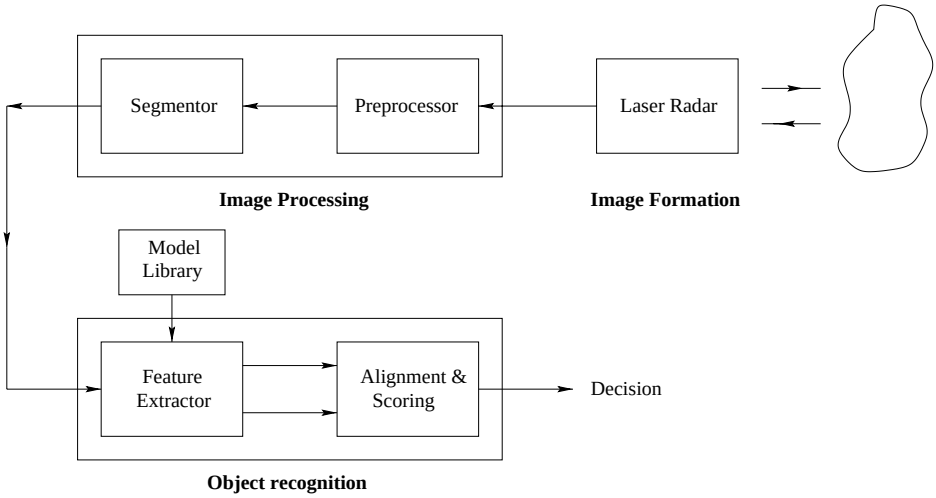

Figure 2. Block diagram for a model-based laser-radar object recognition system.

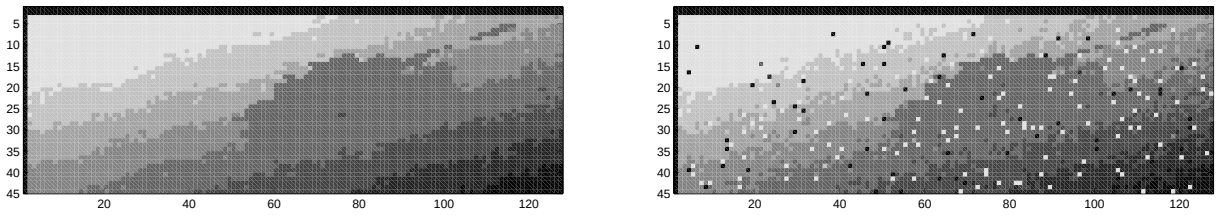

Figure 3. Left panel: high-CNR framing mode laser radar range image of an M60 tank in featureless sloping terrain. Gray scale is used to code range, with lighter colors being farther away from the sensor. Right panel: simulated laser radar range image created by using the left panel as the range truth and employing the statistics of moderate-CNR laser radar operation.

processing as part of the recognition system. Each model is first rendered to synthesize multiple 2-D object views spanning the space of out-of-plane object rotations. Each of these 2-D views is then subjected to the same feature-extraction procedure that is used on the segmented, preprocessed range image. In this way a model library of 2-D feature images is created for alignment and scoring against the sensor-data feature image.

The alignment step estimates the pose of the object in the image. Then the object model is projected into the image plane with the estimated pose and then compared with the actual image. The scores of alignment for each model are compared to make a recognition decision.

3. INPUT IMAGERY

The laser radar imagery used in this work has been collected by MIT Lincoln Laboratory, as part of the Infrared Airborne Radar (IRAR) program and is available via the IRAR Data release.[11,12] All of the data sets were collected from experimental ground-imaging sensors aboard an aircraft. We are mainly interested in processing forward-looking framing-mode range images. The left panel in Fig. 3 illustrates a sample range image containing an M60 army tank situated in a sloping but otherwise featureless terrain.

The range data is a 45×128 pixel image. The gray shade of each pixel represents the distance in range bins measured by the laser radar. According to the single-pixel statistical model for a coherent laser radar,[13] a range image should have some anomalous (very large range-error) pixels, with the remaining pixels having small Gaussian-distributed range errors. The left panel in Fig. 3 was taken under very high carrier-to-noise ratio (CNR) conditions, for which the number of anomalous pixels was essentially zero. Therefore, we have used this image as the source of range truth in our experiments, from which realistic, simulated raw images can be produced in accordance with proper range statistics. A simulated range image produced by this procedure is shown in the right panel of Fig. 3.

4. IMAGE PROCESSING

In any object recognition system, it is essential that the raw, noise-degraded input data be processed to identify the regions of interest that will be used in the recognition stage. This section discusses the preliminary steps employed to process laser radar range images. In particular, we will present the preprocessing and the segmentation steps. Each module will be used to process the range data shown in the right panel of Fig. 3, which was generated from the range truth given in the left panel of this figure.

4.1. Preprocessing

The laser radar range images are characterized by coarse range precision, added noise and an appreciable amount of anomalous pixels. These images first need to be preprocessed to reduce these sensor-dependent effects. A wide variety of approaches can be used to accomplish this purpose. However, in this research, *ad hoc* image enhancement schemes such as median filtering have been avoided since they do not rely on sensor physics. Instead the problem is formulated as an estimation problem based on the appropriate statistical model for the laser radar range images.

4.1.1. Laser Radar Statistical Model

Suppose that the measured data is a $J \times K$ pixel 3-D range image, $\{\, r_{jk} : 1 \le j \le J,\ 1 \le k \le K \,\}$, where the value of r_{jk} represents the range value of the pixel. The corresponding true range values are $\left\{\, r^*_{jk} : 1 \le j \le J,\ 1 \le k \le K \,\right\}$. Both the range data and the true range values are rearranged into $Q = JK$-dimensional column vectors, $\mathbf{r} = \{\, r_q,\ 1 \le q \le Q \,\}$ and $\mathbf{r}^* = \left\{\, r^*_q,\ 1 \le q \le Q \,\right\}$, respectively. For a Q-pixel range image of some field of view, the pixel spacing is assumed to be large enough so that the range measurements are statistically independent given their respective range truths. Thus, the joint probability density that $\mathbf{r} = \mathbf{R}$ occurs given that $\mathbf{r}^* = \mathbf{R}^*$ is given by the individual product of single pixel probability density functions,[13]

$$p_{\mathbf{r}|\mathbf{r}^*}(\,\mathbf{R} \mid \mathbf{R}^*\,) = \prod_{q=1}^{Q} \left[[1 - \Pr(A)] \frac{\exp\left(-\dfrac{(R_q - R^*_q)^2}{2(\delta R)^2} \right)}{\sqrt{2\pi(\delta R)^2}} + \Pr(A) \frac{1}{\Delta R} \right], \quad \text{for } R_{\min} \le R, R^* \le R_{\max}. \tag{1}$$

In Eq. 1: $\Pr(\mathrm{A})$ is the probability of anomaly, i.e., the probability that speckle and shot noise effects combine to yield a range measurement more than one range resolution cell from the true range; ΔR is the width of the radar's range uncertainty interval $\mathcal{R} \equiv [R_{\min}, R_{\max}]$; and δR is the local range accuracy, i.e., the root-mean-squared (rms) range error given that the data is not anomalous.

The first term on the right in Eq. 1 equals the probability that the measurement is not anomalous times a Gaussian probability density with mean equal to the true range value; this term represents the local range behavior. The second term on the right in Eq. 1 represents the global range behavior; it equals the probability that the pixel is anomalous times a uniform distribution over the entire range-uncertainty interval.

4.1.2. Maximum-Likelihood Range Profile Estimation

To find the optimal range estimate of the true range image, $\mathbf{r}^*$, given the measured data, $\mathbf{r} = \mathbf{R}$, we will use maximum-likelihood (ML) estimation,

$$\hat{\mathbf{r}}^*_{\mathrm{ML}}(\mathbf{R}) = \mathrm{argmax}_{\mathbf{R}^*} \left[p_{\mathbf{r}|\mathbf{r}^*}(\,\mathbf{R} \mid \mathbf{R}^*\,) \right]. \tag{2}$$

However, the joint probability density given in Eq. 1 implies that the ML estimate of the range image is the raw data itself. This means that range anomalies cannot be suppressed by this method without the injection of some regularity conditions into the process. We shall follow the work of Greer et al. in this regard.[9] They have developed a fast, numerically-robust ML estimation procedure for laser radar range images that uses a Haar-wavelet decomposition in conjunction with the expectation-maximization (EM) algorithm, which we summarize below.

Suppose that we construct a parametric representation for the true range vector, $\mathbf{r}^*$, by introducing a $Q \times Q$ orthogonal matrix $\mathbf{H}$ and defining a Q-D parameter vector via

$$\mathbf{x} = \mathbf{H}^T \mathbf{r}^*. \tag{3}$$

By assumption, the columns of $\mathbf{H}$, denoted $\{\, \phi_q : 1 \le q \le Q \,\}$, form an orthonormal basis for Q-D vector space. The regularity condition we shall impose is to assume that the true range is characterized by a parameter vector, $\mathbf{x}_P$, of length $P < Q$, comprising the first P dimensions of $\mathbf{x}$, i.e.,

$$\mathbf{r}^* = \mathbf{H}_P \mathbf{x}_P, \quad \text{where } \mathbf{H}_P = \left[\ \phi_1 \quad \phi_2 \quad \cdots \quad \phi_P\ \right]. \tag{4}$$

The next step is to find the ML estimate of the partial representation. The last $Q - P$ columns in $\mathbf{H}$ are not used in the characterization of the range truth, hence they need not be used in finding the range estimate. This provides a means of selecting the resolution, P, of the estimated range data. The full-resolution estimate is the best possible ML estimate. However, our purpose is not only maximizing the likelihood but also suppressing the anomalies. Thus, there is a tradeoff in selecting the resolution of the range estimate.

The ML estimate of the parameter vector satisfies,

$$\mathbf{H}_P^T \mathbf{W}_Q(\hat{\mathbf{x}}_{P\mathrm{ML}})(\mathbf{R} - \mathbf{H}_P \hat{\mathbf{x}}_{P\mathrm{ML}}) = \mathbf{0}, \tag{5}$$

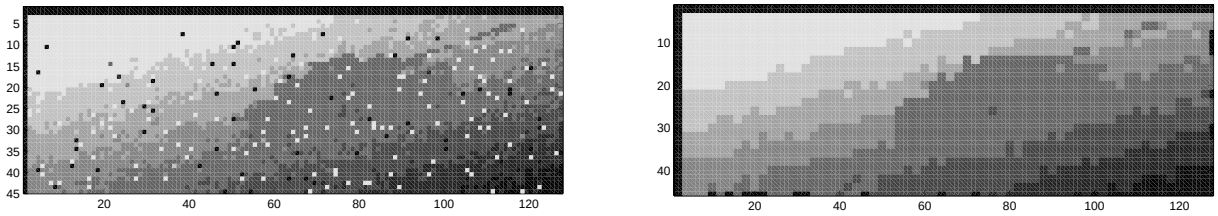

Figure 4. Left panel: simulated raw image with anomalies and local range errors. Right panel: Haar-wavelet ML fit to the left-panel image using 2×2 pixel blocks and showing near-complete suppression of range anomalies.

where

$$\mathbf{W}_Q(\mathbf{X}_P) \equiv \mathrm{diag}\,[\,\Pr(\,R_{jk} \text{ not anomalous} \mid \mathbf{X}_P\,)\,]\,. \tag{6}$$

This is a nonlinear weighted least-squares problem which can be solved using EM algorithm.[14] Starting from an initial estimate of the parameter vector, the EM algorithm produces a sequence of estimates—of monotonically increasing likelihoods—through an iterative sequence of expectation and maximization steps. Thus this procedure will reach the global maximum if its initial estimate is on the highest likelihood "hill."

A direct implementation of the maximization step at high resolution requires inversion of a $P \times P$ matrix with $P \gg 1$. Moreover, although this matrix is theoretically invertible, many of its elements are almost always subject to numerical underflow. Greer et al. circumvented both the computational complexity of inverting a large matrix *and* the lack of robustness arising from numerical underflow by using the Haar-wavelet basis to construct $\mathbf{H}_P$ and exploiting the special characteristics of this basis.[9] The resulting algorithm is the fast ML/EM algorithm. In essence it splits the Q-pixel range image into P blocks, each of size Q/P pixels. The range estimate for each block is the weighted sum of the pixel values in the block, where the weights are the probabilities that the associated pixels are not anomalous.

Figure 4 shows the Haar-wavelet ML fit to our raw range image using 2×2 pixel blocks. We see from this figure that the anomalous pixels have been successfully suppressed. It is also possible to find ML fits at different resolutions. Since the finest scale fit is successful in suppressing the anomalies, we have decided to use it as the input to the subsequent steps of our system.

4.2. Segmentation

In general, segmentation is employed to distinguish all possible object regions—both target and clutter—from each other and from the background. For the case of framing-mode laser radar range images, only a small area containing the target is imaged, therefore the purpose of the segmentation step in our system is only to isolate the target from the background. Segmentation is nevertheless an essential stage in the system, because the target's ground attachment is difficult to ascertain at the range resolutions available in our imagery. There are a number of techniques that can be used to solve this problem. The simplest approach is to fit a planar model to the background, and identify all pixels that lie above this plane as the target.

To estimate the background plane, we assume that the true range values of the pixels comprise a plane and that the measurement statistics are characterized by a mixture of Gaussian and uniform distributions. The uniform distribution describes the behavior of anomalous pixels, which in this case correspond to the target pixels. ML estimation is employed to find estimate the slopes and zero-intercept of the background plane. This problem is almost identical to the multiresolution range profiling problem.[13] The resulting nonlinear least-squares problem is solved by the EM algorithm. The target pixels are located using the final weights of the pixels provided by the expectation step of the EM algorithm. The weight of each pixel represents the conditional probability that the associated pixel is not anomalous. Hence, the pixels corresponding to the target are the low-weighted pixels and can be determined by comparing the weights against a threshold.

However, locating the low-weighted pixels does not cleanly segment the target from the background. This is because there are pixels lying at the bottom of the target whose range values are too close to those of the background for our planar profiler to identify as "anomalous," i.e., as being target. This is an important problem because our alignment procedure is edge-based and extracting correct target-boundary extraction is therefore crucial to the functioning of the alignment stage. Therefore we go one step further and try to find the target-background

intersection. We find the ML/EM planar range profile of the target-body pixels to determine the target plane. Then we compare the estimated target plane with the original preprocessed image to extract all pixels whose range values lie on this plane.

The left panel of Fig. 5 illustrates the pixels that lie on the estimated target plane. Because of low resolution nature of input imagery, the range values of the pixels at the bottom of the target are very close to those on the ground and the procedure described extracts a tiny ground-strip as being part of the object. We assume that the ground attachment lies at the center of this strip. To obtain the final segmented image, we rotate the ground plane gradient to elevation direction. We truncate along azimuth at the center line and along elevation at the target body limits. Then we rotate the ground plane gradient back to its original pose. The final segmented image is shown in the right panel of Fig. 5.

5. MODEL-BASED OBJECT RECOGNITION

The primary goal of our overall recognition system can be stated compactly as follows: find and evaluate the alignment of the image and the model data. For computational efficiency in the search process, both the laser radar image and the object model are represented by a set of features and matching is performed in this new domain. The model features are transformed by a certain pose and the corresponding alignment is scored statistically. The alignment module maximizes this score by adjusting the pose estimate. The final recognition decision is the model-library object with the highest aligned score.

In the general context, pose space is six-dimensional, involving three rotations and three translations. We handle the out-of-plane rotations by using, for each object in our 3-D CAD model library, a discrete set of 2-D views which are treated as distinct object models. The use of available range information further reduces the dimension of the search space by eliminating the need to search over the scale parameter.

Reduction of the pose space dimension is not only good for computational efficiency. It has a real benefit for object recognition. Models whose physical dimensions are very different from the object present in the sensor image will not align well with the sensor image, because the model-scaling will not be allowed in the alignment process. Were scaling allowed during alignment, the matching algorithm would be allowed to align the main image boundaries of the target with scaled versions of the model in process, no matter how disparate their true sizes were. For low resolution data, this alignment with scaling would seriously increase the probability of misclassification.

This section discusses the constituent subsystems of the object recognition process, in particular the feature extraction and the alignment/scoring steps.

5.1. Feature Extraction

The feature extraction step is employed to obtain compact information about the image and the object model for use in matching. Because edges contain a great deal of information about object shapes, we have chosed to use edge-based in our alignment process. It is important to emphasize at this point that the same feature extraction step is applied to both the preprocessed, segmented laser radar range image and the 2-D renderings of the 3-D CAD models. The isolated objects in both images are placed on zero-range backgrounds.

The feature extraction module consists of two steps: edge detection and feature selection. For both the sensor and CAD-model images, edges may simply be regarded as abrupt changes in the range values of the pixels. The edge pixels are identified via thresholding between neighboring pixels. A pixel is identified as an edge pixel if it has a

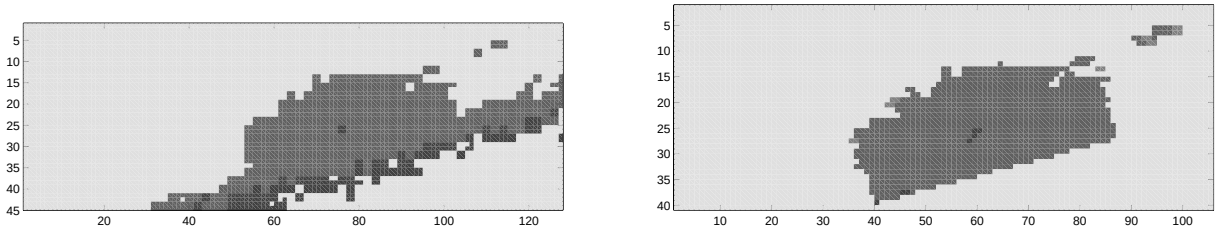

Figure 5. Left panel: target/background plane intersection. Right panel: final target segmentation.

nearest neighbor whose range value is lower by more than a pre-determined threshold. Once edge curves have been obtained from this procedure, they are partitioned into fragments via a "march-down-the-curve" method. Starting from an edge pixel, the edge curve is traced upwards and downwards and segmented into fragments of a certain size. These fragments are used to construct features using a 2-D Point Feature model, in which 2-D real-space centroid coordinates are assigned to each fragment. Each image feature is thus represented by a 2-D vector,

$$\mathbf{Y}_i = \begin{bmatrix} x_i & y_i \end{bmatrix}^T, \quad \text{for } 1 \leq i \leq n, \tag{7}$$

where n is the total number of image features. The location of the model features depends on the pose, denoted by β. The model features are represented by real matrices containing information about the 2-D real-space coordinates of the model feature point. This particular form of representation is selected because it admits to a linear projection model which facilitates the solution of the estimation problem. Each model feature projected by β is represented by,

$$\hat{\mathbf{Y}}_j(\beta) = \mathbf{M}_j\beta, \quad \text{for } 1 \leq j \leq m, \tag{8}$$

where $\mathbf{M}_j$ is a matrix whose size depends on the choice of the pose vector and m is the total number of model features.

The extracted feature points from the segmented range image and the rendered 3-D CAD model image of an M60 tank are shown as points on the detected edge curves in the left and right panels, respectively, of Fig. 6.

5.2. Alignment and Scoring

The objective of alignment and scoring is to determine the optimal pairings between the model and image features, i.e., to find (and score) the alignment between the model and the image the yields the greatest metrical consistency among the paired features. This constraint can be expressed mathematically using different measures. In this work, we employ a statistical approach known as posterior marginal pose estimation.[10]

5.2.1. Image Feature Statistical Model

Let $\mathbf{Y}_i$ denote the ith image feature, $1 \leq i \leq n$, and $\hat{\mathbf{Y}}_j(\beta)$ denote the jth model feature, $1 \leq j \leq m$, projected into the image plane at pose β. The ith image feature, $\mathbf{Y}_i$, is modeled as a continuous random variable whose conditional probability density function given the pose vector β is given by,

$$p(\,\mathbf{Y}_i \mid \beta\,) = \sum_{j=1}^{m} \frac{1 - \Pr(B)}{m} \frac{\exp\left[-(\mathbf{Y}_i - \hat{\mathbf{Y}}_j)^T \Lambda_{ij}^{-1} (\mathbf{Y}_i - \hat{\mathbf{Y}}_j)/2\right]}{(2\pi)^{p/2}|\Lambda_{ij}|^{1/2}} + \frac{\Pr(B)}{A_p}. \tag{9}$$

In Eq. 9: $\Pr(B)$ is the probability that the image feature belongs to the background; A_p denotes the volume of the feature-space uncertainty region; and Λ_{ij} is the covariance matrix for $\mathbf{Y}_i$, when $\mathbf{Y}_i$ corresponds to $\hat{\mathbf{Y}}_j$ at the given pose. An image feature matched to a model feature is assumed to be normally distributed about the predicted position in the image given by the projection of the model feature with a specific pose. Features matched to the background are assumed to be uniformly distributed over the feature-space uncertainty region.

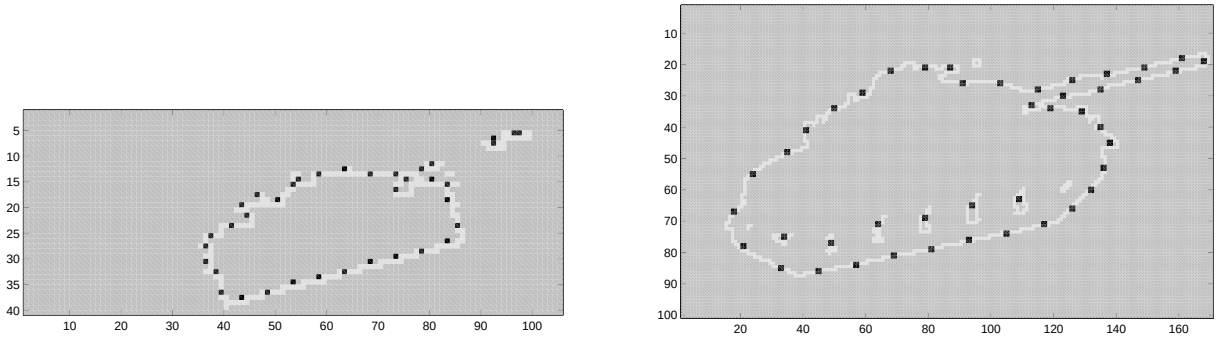

Figure 6. Left panel: feature points from the segmented range image of an M60 tank. Right panel: feature points from the rendered 3-D CAD model image of an M60 tank.

Assuming that the image features are conditionally independent, the joint conditional probability density of the image features, $\mathcal{Y} = \{ \mathbf{Y}_i : 1 \leq i \leq n \}$, given the pose vector β is given by,

$$p(\mathcal{Y} \mid \beta) = \prod_{i=1}^{n} p(\mathbf{Y}_i \mid \beta). \tag{10}$$

5.2.2. Posterior Marginal Pose Estimation

Maximum aposteriori probability (MAP) estimation is used to find the pose estimate by maximizing the posterior probability density of the pose, β, given the observed image features $\mathcal{Y} = \{ \mathbf{Y}_i : 1 \leq i \leq n \}$,

$$\hat{\beta}(\mathcal{Y}) = \text{argmax}_{\beta} \left[p(\beta \mid \mathcal{Y}) \right]. \tag{11}$$

Using Bayes' rule and assuming that there is essentially uniform prior distribution for pose allows us to reduce Eq. 11 to the following maximum-likelihood form:

$$\hat{\beta}_{\text{ML}}(\mathcal{Y}) = \text{argmax}_{\beta} \left[p(\mathcal{Y} \mid \beta) \right]. \tag{12}$$

The zero gradient condition, which is the necessary condition for an extremum, then turns Eq. 12 into a nonlinear weighted least-squares problem,

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}(\hat{\beta}_{\text{ML}}) \mathbf{M}_j^T (\mathbf{Y}_i - \mathbf{M}_j \hat{\beta}_{\text{ML}}) = \mathbf{0}, \tag{13}$$

where

$$w_{ij}(\beta) \equiv \Pr(i\text{th image feature} \rightarrow j\text{th model feature} \mid \text{pose} = \beta). \tag{14}$$

The EM algorithm can be used to solve this problem producing a sequence of estimates whose associated likelihood sequence is monotonically increasing.[10]

It is important to note that the computation involved in PMPE formulation of object recognition shares the same mechanism with that of ML range imaging problem. Both methods use the EM algorithm to find statistically optimal estimates of the variables of interest.

Because the EM algorithm is a local optimization method, it needs a good initial value to converge to the global maximum. In this work, we assume that a good initial pose is provided and we are interested in refining the value of the pose vector by a local search in the pose space.

5.2.3. Statistical Scoring using PMPE

The final recognition decision is made by comparing the quality of the final alignment between the image and each of the object models in the model library. Suppose the model library contains K possible object models,

$$\mathcal{M}_k = \{ \mathbf{M}_{j_k}, 1 \leq j_k \leq m_k \}, \text{ for } 1 \leq k \leq K. \tag{15}$$

PMPE is used to find the optimal pose estimate for each of the models. Let $\hat{\beta}_k$ denote the PMPE pose estimate for the kth model,

$$\hat{\beta}_k = \text{argmax}_{\beta} \left[p(\mathcal{Y} \mid \beta, \mathcal{M}_k) \right]. \tag{16}$$

The logarithm of the likelihood function (scaled properly) is used to score the degree of match between the image features and the projected model features.[15] We declare the object model with the highest score to be the target in the input image.

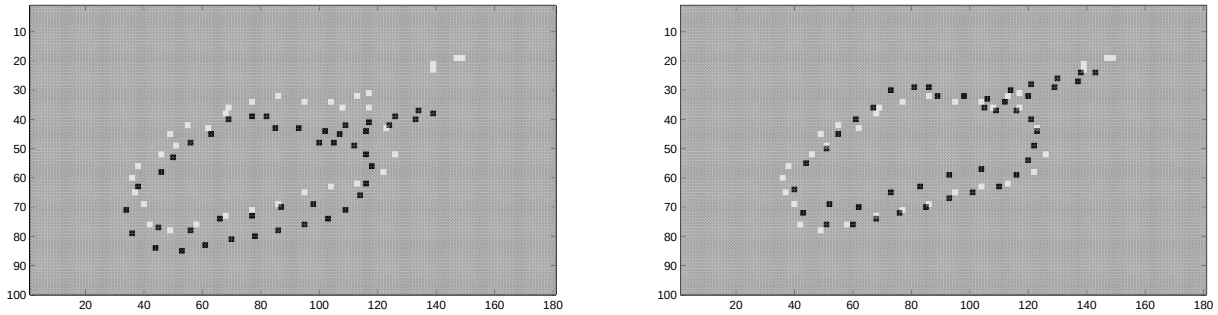

Figure 7. Left panel: initial alignment of the M60 model features with the M60 image features. Right panel: final alignment of the M60 model features with the M60 image features.

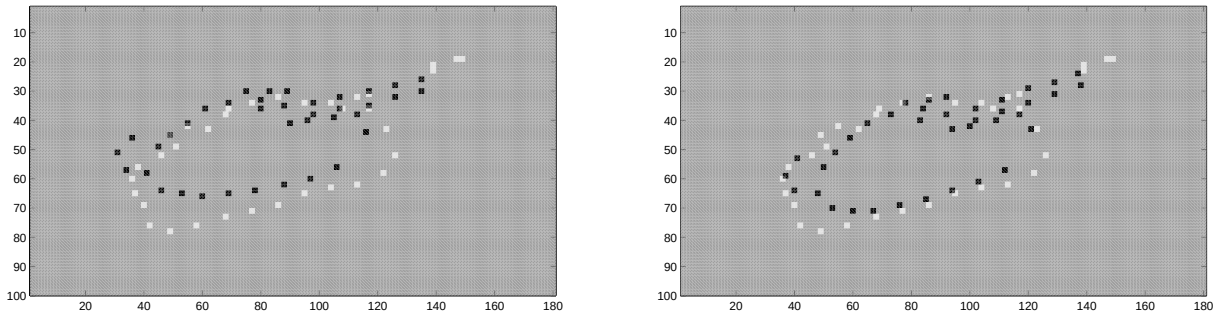

Figure 8. Left panel: initial alignment of the T80 model features with the M60 image features. Right panel: final alignment of the T80 model features with the M60 image features.

6. PERFORMANCE RESULTS

In this section, we present performance results pertaining to the sample image, shown in the right panel of Fig. 3, which contains the M60 tank as the target. Three representative models have been selected from the model library to perform matching with the image of interest: an M60 tank model, a T80 tank model, and a Ford GPA Jeep model. Other results can be found in Ref. 14.

We start with the alignment of the image with the correct model. One scheme for assessing the quality of alignment is to display the image and the model features together. The nature of the alignnment is readily apparent from these images. Note that in all figures in which alignment between two sets of features are displayed, the white dots represent the model features whereas the black dots represent the model features. In all cases and the pixel size is $0.1\,\mathrm{m} \times 0.1\,\mathrm{m}$. Initially, the object model pose was manually adjusted until it appeared to be close to correct. Then pose refinement via PMPE was used to pull the object into the "correct pose". The initial and final alignments can be seen in Fig. 7. The left panel shows the initial alignment between two sets of features and the right panel illustrates the alignment of image features with the model features projected onto the image with the pose estimated by the PMPE algorithm. As seen, there is a good match between the object in the image and the correct model, i.e., the boundaries of the object in the image are in close agreement with the outline of the model points.

The initial and the final alignments corresponding to the T80 tank model are shown in Fig. 8. The final alignment results for the M60 tank model and the T80 tank model show that there is a better match between the image data and the M60 tank model than between the image data and the T80 tank model. (As will be seen in the scores given later in Table 1, this means that the data is more supportive of the claim that an M60 tank, rather than a T80 tank, was present.) It is easy to understand the behavior of the pose estimate. The pose is estimated such that maximum number of matches is achieved between the features. Therefore the algorithm tries to match regions of the image and the model where there is a cluster of features, such as the gun-barrel region.

The last matching example we will show is with the Jeep model. The initial and final alignments are presented in Fig. 9. This figure is a clear manifestation of an important characteristic of recognition system: a model whose physical size is considerably different from the true object is forced to have a low alignment score.

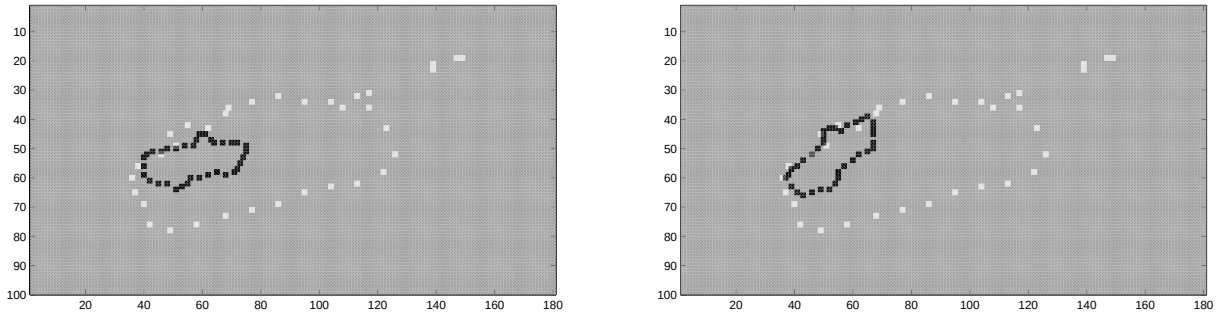

Figure 9. Left panel: initial alignment of the GPA jeep model features with the M60 image features. Right panel: final alignment of the GPA jeep model features with the M60 image features.

Object Model	Initial Score	Final Score	Number of EM Iterations	Number of Model Features
M60 Tank Model	95.06	154.53	39	44
T80 Tank Model	98.49	133.94	19	38
GPA Jeep Model	37.38	60.66	12	35

Table 1. Initial and final scores, number of EM iterations, and number of mode features for our fitting of the M60, T80, and GPA Jeep CAD models to the features derived from our preprocessed/segmented/feature-extracted laser radar range image of an M60 tank. The final scores indicate that the image data is more supportive of the claim that an M60 tank was present, rather than a T80 tank or a GPA jeep.

The scores corresponding to the final alignments shown in Figs. 7–9 are given in Table 1. This table also includes the initial alignment scores, thus quantifying the effectiveness of PMPE refinement. Comparing these scores leads to our final recognition decision that the target in the input image is an M60 tank.

7. MULTIPLE TRIALS

The single-trial experiment conducted in the previous section was able to reach a correct decision about the identity of the target in the image. A single trial, however, is not sufficient to demonstrate the robustness of the recognition algorithm. To be more confident about the behavior of the scores, we performed a multiple-trial matching experiment for the correct model, the M60 tank model, and one of the incorrect models, the T80 tank model, using simulated data randomly generated from the M60 range truth. Only a small number of trials could be performed due to large run-time requirement of this experiment. The resulting scores obtained from sets of statistically independent, identically distributed trials for both of the models are given in Fig. 10 in the form of a scatter diagram. The results show that the algorithm correctly chooses the M60 model in every case, thus giving us a reasonable level of confidence that the algorithm is working well.

8. CONCLUSIONS

In this work, we built an end-to-end system that uses low resolution, noisy laser radar range images as its input and provides recognition decisions at its output. We presented performance results for our system using laser radar data from MIT Lincoln Laboratory IRAR data release together with 3-D CAD models which account for the possible military targets that could have been present in this data.

Despite the poor resolution in the input image, our system acquired enough information to arrive at a correct decision about the identity of the target. As expected, the results imply that mismatch is most likely to be between targets of similar sizes. The use of actual dimensions in the alignment process thus greatly reduces the probability of misclassification precluding the possibility of confusing targets which, although similarly shaped, have very different dimensions.

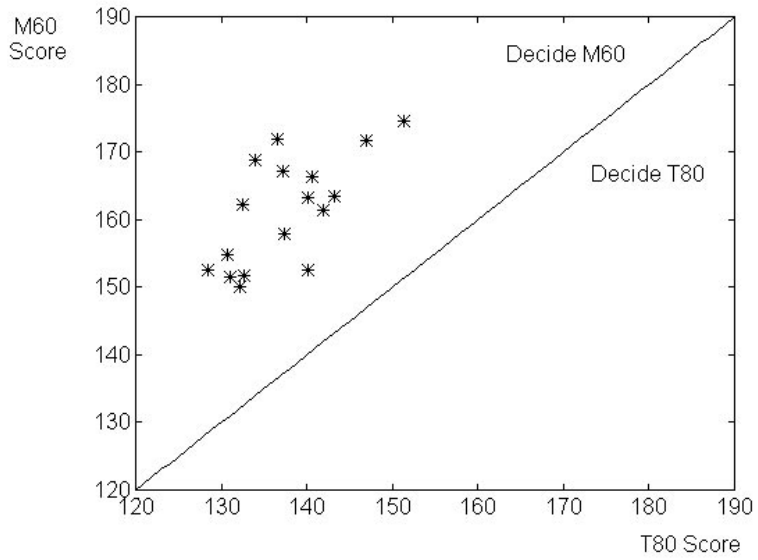

Figure 10. Scoring results for a 17-trial experiment comparing M60 and T80 tank models with statistically independent, identically distributed M60 range images.

REFERENCES

1. R.C. Harney and R.J. Hull, "Compact infrared radar technology," Proc. SPIE **227**, 162–170 (1980).
2. A.B. Gschwendtner, R.C. Harney, and R.J. Hull, "Coherent IR radar technology," in *Optical and Laser Remote Sensing*, D.K. Killinger and A. Mooradian, Eds., Springer-Verlag, Berlin, 1983.
3. J.W. Goodman, "Statistical properties of laser speckle," in *Laser Speckle and Related Phenomena*, J.C. Dainty, Ed., Springer-Verlag, Berlin, 1975.
4. R.H. Kingston, *Detection of Optical and Infrared Radiation*, Springer-Verlag, Berlin, 1978, Chap. 3.
5. J.H. Shapiro, R.W. Reinhold, and D. Park, "Performance analyses for peak detecting laser radars," Proc. SPIE **663**, 38–56 (1986).
6. M.B. Mark and J.H. Shapiro, "Multipixel, multidimensional laser radar system performance," Proc. SPIE **783**, 109–122 (1987).
7. S.M. Hannon and J.H. Shapiro,"Laser radar target detection with a multipixel joint range-intensity processor," Proc. SPIE **999**, 162–175 (1988).
8. S.M. Hannon and J.H. Shapiro, "Active-passive detection of multipixel targets," Proc. SPIE, **1222**, 2–23 (1990).
9. D.R. Greer, I. Fung, and J.H. Shapiro, "Maximum-likelihood multiresolution laser radar range imaging," IEEE Trans. on Image Process. **6**, 36–46(1997).
10. W.M. Wells, III, "Statistical approaches to feature-based object recognition," International J. Comput. Vision **21**, 63–98 (1997).
11. J.K. Bounds, "The Infrared Airborne Radar Sensor Suite," Research Laboratory of Electronics Tech. Report 610, MIT, December 1996.
12. IRAR data release, http://cis.jhu.edu/mit_cis/laserradar/IRAR/IRARmain.html
13. T.J. Green, Jr. and J.H. Shapiro, "Maximum-likelihood laser radar range profiling with the expectation-maximization algorithm," Opt. Eng. **31**, 2343–2354 (1992).
14. A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. **39**, 1–38 (1977).
15. A.E. Koksal, "Using multiresolution range-profiled real imagery in a statistical object recognition system," S.M. Thesis, Dept. of ElectEng. and Comput. Sci., MIT, May 1998.