Predicting Performance of Object Recognition

Michael Boshra, *Member*, *IEEE*, and Bir Bhanu, *Fellow*, *IEEE*

Abstract—We present a method for predicting fundamental performance of object recognition. We assume that both scene data and model objects are represented by 2D point features and a data/model match is evaluated using a vote-based criterion. The proposed method considers data distortion factors such as uncertainty, occlusion, and clutter, in addition to model similarity. This is unlike previous approaches, which consider only a subset of these factors. Performance is predicted in two stages. In the first stage, the similarity between every pair of model objects is captured by comparing their structures as a function of the relative transformation between them. In the second stage, the similarity information is used along with statistical models of the data-distortion factors to determine an upper bound on the probability of recognition error. This bound is directly used to determine a lower bound on the probability of correct recognition. The validity of the method is experimentally demonstrated using real synthetic aperture radar (SAR) data.

Index Terms—Bounds on recognition performance, model-based real-world object recognition, modeling data distortion, performance validation, synthetic aperture radar images, theory of performance prediction.

1 Introduction

Model-based object recognition is concerned with identifying and localizing instances of model objects in scene data. It involves extracting features from the scene data and searching for a good match between the extracted features and those of the model objects. Fundamentally, performance of recognition depends on the following factors. 1) *Distortion of the data features:* data distortion is a result of sensor noise, presence of clutter, occlusion, and imperfections of feature-extraction algorithms. 2) *Characteristics of the model objects:* this includes factors such as object articulations and the degree of similarity between objects. Intuitively, recognition becomes harder as the degree of model similarity increases and vice versa. 3) *Matching criterion:* the most popular criterion is based on the number of consistent data/model feature pairs (votes), or a weighted version of it. More complicated criteria, which are mainly statistical, have been proposed. Some of them account for unmatched features (e.g., [14]), while others capture the spatial relationships between matched/unmatched features (e.g., [3], [6], [7]).

Object recognition performance is typically determined experimentally. That is, a set of scene images is passed to the recognition system and the results obtained are used to determine the recognition performance (e.g., Monte Carlo experiments). The scene images can be either real, or synthetic with artificial distortion introduced to them. Experimental performance evaluation has two major limitations. *First,* it does not provide an insight into the relationship between recognition performance and various data and model factors that affect it. In other words, the experimental approach can tell us *what* performance to expect for a given set of model objects at specific data distortion rates. However, it does not tell us *why* this is the expected performance. Such an insight is important for designing better recognition systems, as it can provide fundamental answers to questions such as: 1) What are the performance limits? 2) How much distortion can be tolerated without degrading performance? 3) Is a given matching criterion sufficient to achieve desired levels of performance? 4) How many model objects can be accommodated without substantially degrading performance? Without such a fundamental understanding of the relationship between performance and the factors affecting it, design of object recognition systems will remain an art rather than a science. *Second*, the performance obtained experimentally depends on the actual implementation of the recognition system, which can be based on approaches such as alignment [8], [12], hypothesis accumulation [18], [22], or tree search [9], [11]. Note that even though these techniques might use the same matching criterion, they might still produce different performance results depending on how "closely" they follow the chosen criterion. For example, alignment-based systems use a looser notion of feature consistency in order to achieve polynomial-time complexity. Another example, performance of some hypothesis-accumulation-based systems depends on the resolution of the object/pose accumulator.

A major goal of our research is the development of a formal framework for predicting fundamental performance of object recognition. In this paper, we address the prediction problem in the context of the following object recognition task. 1) Both scene data and model objects are represented by locations of 2D point features that are discretized at some resolution. 2) An instance of a model object in the scene data is assumed to be obtained by applying some 2D transformation to the object (e.g., translation, rigid, affine). In our experiments, however, we only consider 2D translations, due to the nature of the recognition system at hand. Notice that, since we are dealing with discretized point features, the space of

- *M. Boshra is with AuthenTec Inc., 709 S. Harbor City Blvd., Melbourne, FL 32901. E-mail: mboshra@authentec.com.*
- *B. Bhanu is with the Center for Research in Intelligent Systems, University of California, Riverside, California 92521. E-mail: bhanu@vislab.ucr.edu.*

Manuscript received 28 Sept. 1999; revised 31 May 2000; accepted 26 June 2000.
Recommended for acceptance by J.R. Beveridge.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 110676.

applicable transformations for a given object is discretized. 3) The matching criterion is vote-based. In particular, votes for a given object/pose hypothesis are the model features that are consistent with one or more data features.

We present a statistical method for formally predicting a lower bound on the *probability of correct recognition* (PCR), considering the following: 1) *Data Factors:* uncertainty (positional), occlusion (missing features), and clutter (spurious features) and 2) *Model Factor:* structural similarity (degree of structural "overlap" between pairs of model objects). It has been a challenge to model these factors in a single approach for performance prediction. The performance predicted by our method is fundamental, since it is obtained by analyzing the information provided by both scene data and model objects, *independent* of the particular algorithm used for vote-based recognition. The proposed method is validated using real synthetic aperture radar (SAR) data.

The remainder of the paper is organized as follows: The next section reviews related research and highlights our contributions. Section 3 presents an overview of the method. Section 4 describes the statistical modeling of the data distortion factors, while Section 5 discusses object-similarity issues. A lower bound on PCR is derived in Section 6. In Section 7, we demonstrate the validity of the proposed method by comparing actual PCR plots, as a function of data distortion, with predicted bounds. Finally, conclusions are drawn in Section 8.

2 Related Research

There have been several research efforts for analyzing the performance of feature-based object recognition. Most of these efforts address the problem of discriminating objects from clutter. A representative sample of these approaches can be outlined as follows: Grimson and Huttenlocher [10] used a statistical occupancy model for estimating the distribution of the fraction of consistent data/model feature pairs for an erroneous hypothesis. This distribution was obtained under the assumptions of bounded uncertainty and uniform clutter models. It was used to determine the minimum fraction of consistent feature pairs required to achieve a specific probability of false alarm. Sarachik [21] estimated probability distributions of weighted votes conditioned on hypothesis validity and invalidity. These distributions were obtained assuming a Gaussian uncertainty model and uniform clutter and occlusion models. The likelihood-ratio test was used to decide on whether to accept a given hypothesis. Alter and Grimson [1] calculated the likelihoods of observed data features conditioned on hypothesis validity and invalidity and used the likelihood-ratio test as a decision criterion. This work considered both bounded and Gaussian uncertainty models, in addition to uniform clutter and occlusion models. Lindenbaum [17] extended the modeling of clutter to include objects of known shapes as well as random features. This hybrid model, which considered the similarity between clutter and model objects, was used to determine bounds on the number of features required to reliably distinguish objects, at any poses, from clutter.

The approaches outlined above considered a subset of the object recognition problem, which is object/clutter discrimination. Very few efforts have addressed the problem of discriminating objects from other model objects, which is obviously essential for predicting performance of object recognition. It is complicated by the need to consider not only the similarity between model objects, but also the *interaction* between object similarity and data distortion. Lindenbaum [16] presented a probabilistic method for deriving bounds on the number of features required to achieve recognition with a certain degree of confidence. This method considered object similarity, bounded uncertainty and occlusion. Extreme-case analysis was used to model the interaction between similarity and occlusion, thus leading to the generation of relatively loose bounds. We note that the method presented in [17], outlined above, for object/clutter discrimination can be used to analyze object/object discrimination in the presence of uncertainty, similarity, and clutter.

A comparison between the approaches outlined above and ours is shown in Table 1. From this table, it can be observed that our approach is unique in that it simultaneously considers uncertainty, occlusion, clutter, and object similarity. Previous approaches considered only a subset of these factors. Compared to other object/object discrimination methods [16], [17], other unique aspects of this work are consideration of objects represented by point features and validation using real data.

3 Overview of the Method

The performance-prediction problem considered in this paper can be formally defined as follows: We are given:

1. A set of model objects, $\mathcal{DB} = \{\mathcal{M}_i\}$, where each object $\mathcal{M}_i$ is represented by a set of discretized 2D point features, $\mathcal{M}_i = \{F_{ik}\}$,
2. statistical models for data distortion, and
3. a class of applicable transformations, $\mathcal{T}$.

Our objective is to predict a lower bound on PCR, as a function of distortion levels, assuming a vote-based matching criterion (to be formally defined later). Recognition is considered successful if the hypothesized object is the same as the actual one and the error between the hypothesized pose and the actual one is lying within a subspace, $\mathcal{T}_{acc} \subset \mathcal{T}$.

Fig. 1 shows a block diagram of the proposed method. It can be described as follows:

- **Data-Distortion Models.** Data distortion factors are statistically modeled using uniform probability distribution functions (PDFs) as follows: 1) *Uncertainty:* The actual location of a feature corresponding to an extracted data feature is described using a uniform distribution. Notice that the uncertainty distribution is discrete, since we are dealing with discrete features. We assume that the distributions associated with the data features are independent. While this assumption might not be entirely accurate (e.g., among neighboring features), we argue that it is reasonable in most practical situations. 2) *Occlusion:* Let O be the number of occluded features. We

TABLE 1
A Comparison Between Various Performance Prediction Approaches (U, O, C, and S Denote Uncertainty, Occlusion, Clutter, and Similarity, Respectively)

Work	Discrimination	Data/Model Features	Transformation	Factors			
				U	O	C	S
Grimson and Huttenlocher [10]	object/clutter	2-D/2-D lines	rigid	X		X	
Sarachik [21]	object/clutter	2-D/2-D points	affine	X	X	X	
Alter and Grimson [1]	object/clutter	2-D/3-D points & lines	weak perspective	X	X	X	
Lindenbaum [17]	object/clutter	2-D/2-D boundary pts.	affine	X		X	X
Lindenbaum [16]	object/object	2-D/2-D boundary pts.	rigid	X	X		X
This work	object/object	2-D/2-D discretized pts.	general trans.	X	X	X	X

assume that each subset of model features of size O is equally likely to be occluded. This distribution is more suitable for modeling features that are missing due to inherent instability or imperfections of the feature extraction process. It is less suitable for modeling features that are missing due to obscuration by other objects, since it does not take into account the spatial correlation between occluded/unoccluded features. This type of spatial correlation can be captured by using Markov random fields [15], at the expense of significantly increasing the complexity of the statistical analysis. In Section 7, we outline a simpler alternative approach, which can consider "spatial" occlusion, without increasing the complexity of the analysis. 3) *Clutter:* Clutter features are assumed to be uniformly distributed within some region, of arbitrary shape, surrounding the object. This distribution is useful for modeling clutter features which do not appear in specific structural patterns. Modeling "structural" clutter requires considering its similarity with the model objects. While this topic is a subject of future research, we note that the concepts presented in this paper, particularly Sections 5 and 6, can be used as a basis for considering structural clutter in performance prediction.

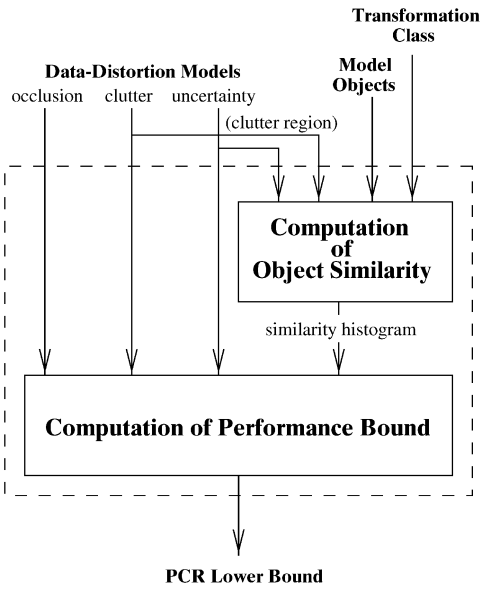

Fig. 1. Proposed performance-prediction method.

- **Computation of Object Similarity.** The objective of this stage is to compute the similarity between all pairs of model objects. Our measure of object similarity depends on the amount of data uncertainty. This agrees with the intuitive observation that as objects become more "blurred," it becomes more difficult to distinguish between them. This is equivalent to saying that they become more "similar." The similarity of an object, $\mathcal{M}_j$, to another one, $\mathcal{M}_i$, is defined as the number of votes that $\mathcal{M}_j$ would get given an "uncertain" instance of $\mathcal{M}_i$, as a function of the relative transformation between the two objects. Notice that the space of relative transformations is defined by the applicable transformation class $\mathcal{T}$. Obviously, the resulting function, which we call the *similarity function*, is probabilistic. Another important function is the *effective-size function*, which is simply the number of features of $\mathcal{M}_j$ that lie inside the clutter region surrounding $\mathcal{M}_i$, as a function of $\mathcal{T}$. The output of this stage is a *similarity histogram*, which accumulates information derived from the similarity and effective-size functions for all pairs of model objects.
- **Computation of Performance Bound.** The objective of this stage is to compute a lower bound on PCR. For each model object, an upper bound on the probability

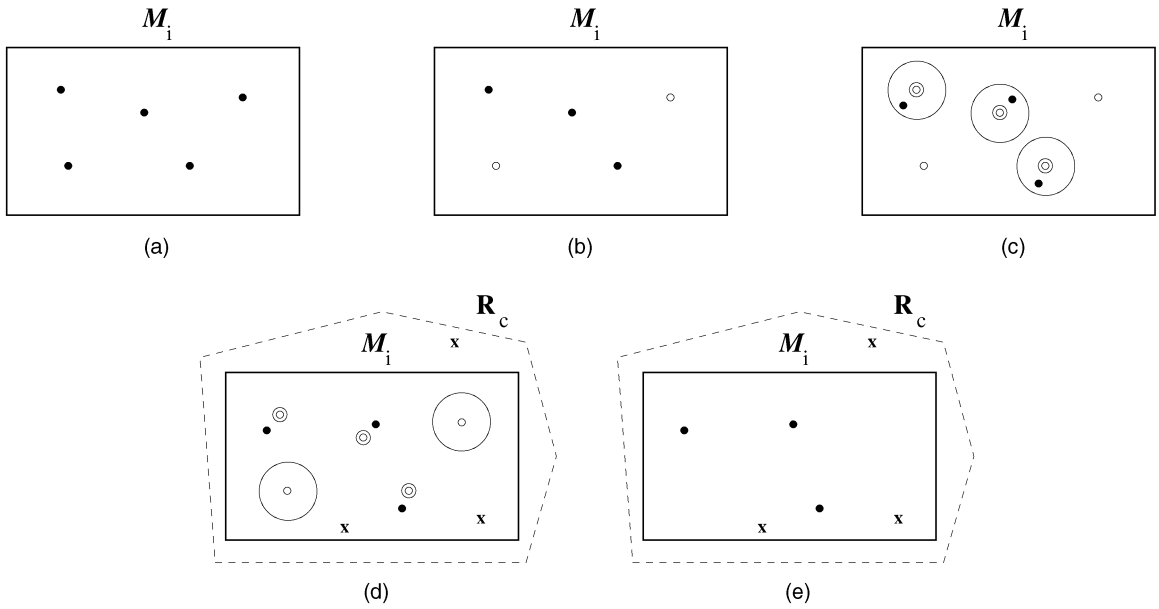

Fig. 2. An example of the distortion process: (a) Original object which consists of five features shown as dark points, (b) after occlusion ($O = 2$; the small circles represent original locations of occluded features), (c) after perturbation ($R_u(\cdot)$ = a circle centered at the feature's true location; the small double circles represent original locations of perturbed features), (d) after clutter ($C = 3$; the small crosses represent clutter features; notice that there are no clutter features inside the uncertainty regions of occluded features), and (e) object after distortion.

of recognition failure, given specific levels of data distortion, is computed. A recognition failure occurs, if the votes for *any* erroneous object/pose hypothesis reach or exceed those for the given "distorted" object. The probability of recognition failure is calculated by considering similarity and its interaction with distortion. The average of the computed upper bounds on the probability of recognition failure, for all model objects, directly leads to the desired lower bound on PCR.

4 Data-Distortion Models

In this section, we formally define the three sources of data distortion considered in this paper.

- **Uncertainty.** The effect of uncertainty is to perturb positions of object features according to some PDF. Since we are assuming that this PDF is uniform, it can be represented by an *uncertainty region*, $R_u(\cdot)$. An uncertain instance of a model object, $\mathcal{M}_i$, is obtained by randomly perturbing each feature in $\mathcal{M}_i$ within its uncertainty region. Formally, it can be represented as

$$\mathcal{D}_u(\mathcal{M}_i, R_u(\cdot)) = \{P_u(R_u(F_{ik})) : F_{ik} \in \mathcal{M}_i\},$$

  where $P_u(R)$ is a function that returns a randomly selected feature within region R.
- **Occlusion.** The effect of occlusion is to eliminate some object features. An occluded instance of object $\mathcal{M}_i$ can be represented as follows:

$$\mathcal{D}_o(\mathcal{M}_i, O) = \mathcal{M}_i - \mathcal{P}_o(\mathcal{M}_i, O),$$

  where $\mathcal{P}_o(\mathcal{M}_i, O)$ is a function that returns a randomly selected subset of O features from $\mathcal{M}_i$.
- **Clutter.** The effect of clutter is the addition of spurious features. We assume that these features are uniformly distributed within a *clutter region*, R_c, which can be of arbitrary shape (e.g., convex hull of object features, bounding box, etc.). An ambiguity occurs if a clutter feature happens to fall within the uncertainty region of an occluded one, since it can not be differentiated from the no-occlusion/no-clutter case. In our work, we assume the latter case, for simplicity.[1] Accordingly, we restrict clutter features to lie outside the uncertainty regions of occluded ones. Formally, a cluttered instance of $\mathcal{M}_i$ can be defined as follows:

$$\mathcal{D}_c(\mathcal{M}_i, C, R_c, R_x) = \mathcal{M}_i \cup \mathcal{P}_c(C, R_c - R_x),$$

  where $\mathcal{P}_c(C, R)$ is a function that returns C randomly generated features within region R and R_x is a region that is excluded from receiving clutter features. In our context, R_x is the union of the uncertainty regions associated with the *occluded* features. We refer to $R_c - R_x$, or simply R'_c, as the *effective clutter region*.
- **Combined Distortion.** In general, a distorted instance of object $\mathcal{M}_i$, $\widehat{\mathcal{M}_i}(R_u(\cdot), O, C, R_c)$, is obtained by occluding O features in $\mathcal{M}_i$, perturbing unoccluded ones within their uncertainty regions $R_u(\cdot)$ and then adding C clutter features within the effective clutter region R'_c. Formally,

$$\begin{aligned}&\widehat{\mathcal{M}_i}(R_u(\cdot), O, C, R_c)\\ &\quad = \mathcal{D}_c(\mathcal{D}_u(\mathcal{D}_o(\mathcal{M}_i, O), R_u(\cdot)), C, R_c, R_x),\end{aligned}$$

1. In other words, we do not distinguish between unoccluded object features and clutter features that fall within uncertainty regions of occluded object features. This simplification is valid in our case, since locations of both object and clutter features are described by similar (uniform) PDFs.

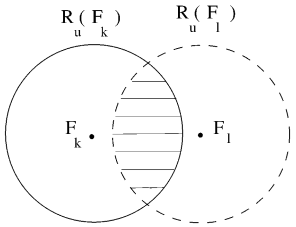

Fig. 3. An illustration of feature/feature similarity: the similarity between features F_k and F_l is proportional to the overlapping area of their uncertainty regions.

where $R_x = \cup_k R_u(F_{ik}), \forall F_{ik} \in (\mathcal{M}_i - \mathcal{D}_o(\mathcal{M}_i, O))$. An example of the distortion process is shown in Fig. 2.

5 Computation of Object Similarity

In this section, we formally define a similarity measure between model objects and present the method used to construct the similarity histogram.

5.1 Definition of Object Similarity

We introduce a sequence of definitions that lead to a quantitative measure of the similarity between a pair of model objects.

- **Feature Consistency.** A feature, F_k, is said to be consistent with another feature, F_l, if F_k can be interpreted as an uncertain measurement of F_l. It can be easily seen that the condition for consistency in our case is $F_k \in R_u(F_l)$.
- **Vote-Based Criterion.** Let $\mathcal{M}_i^{\hat{\tau}}$ be a *hypothesis* of object $\mathcal{M}_i$ at pose $\hat{\tau} \in \mathcal{T}$ relative to a distorted object, $\widehat{\mathcal{M}}$. The votes for $\mathcal{M}_i^{\hat{\tau}}$ given $\widehat{\mathcal{M}}$, $\mathrm{VOTES}(\mathcal{M}_i^{\hat{\tau}}; \widehat{\mathcal{M}})$, is the number of features in $\mathcal{M}_i$ that have one or more consistent features in $\widehat{\mathcal{M}}$. Formally,

$$\mathrm{VOTES}(\mathcal{M}_i^{\hat{\tau}}; \widehat{\mathcal{M}}) = \mid \{F_{ik}^{\hat{\tau}} : F_{ik}^{\hat{\tau}} \in \mathcal{M}_i^{\hat{\tau}} \text{ and } \exists \hat{F}_l \in \widehat{\mathcal{M}} \text{ s.t. } \hat{F}_l \in R_u(F_{ik}^{\hat{\tau}})\} \mid .$$

- **Feature/Feature Similarity.** The similarity between features F_k and F_l, $S_{ff}(F_k, F_l)$, is defined as the probability that an uncertain measurement of F_k is consistent with F_l. The degree of similarity is proportional to the extent of overlap between $R_u(F_k)$ and $R_u(F_l)$ (see Fig. 3). Specifically,

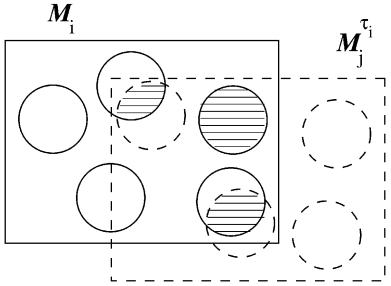

Fig. 4. Similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$. Notice that $S_j^{\tau_i} \in [1, 3]$, and $E(S_j^{\tau_i}) \approx 2$.

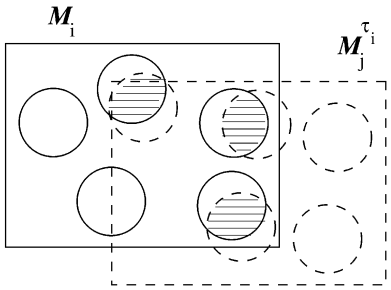

Fig. 5. Uniform model of similarity for the object/hypothesis pair shown in Fig. 4. Notice that the feature/feature similarity for each of the three similar feature pairs is $P_j^{\tau_i} \approx \frac{2}{3}$.

$$S_{ff}(F_k, F_l) = \frac{\mathrm{AREA}(R_u(F_k) \cap R_u(F_l))}{\mathrm{AREA}(R_u(F_k))},$$

where $\mathrm{AREA}(R)$ is the area of region R. We sometimes refer to feature pairs with overlapping/nonoverlapping uncertainty regions as *similar/dissimilar* pairs.

- **Object/Feature Similarity.** The similarity between a model object, $\mathcal{M}_i$, and a feature, F_l, $S_{of}(\mathcal{M}_i, F_l)$, is defined as the probability that an uncertain measurement of *any* feature in $\mathcal{M}_i$ (i.e., a feature in $\mathcal{D}_u(\mathcal{M}_i, R_u(\cdot))$, see Section 4) is consistent with F_l. Formally,

$$S_{of}(\mathcal{M}_i, F_l) = 1 - \prod_k (1 - S_{ff}(F_{ik}, F_l)).$$

- **Object/Hypothesis Similarity.** Let $\mathcal{M}_j^{\tau_i}$ be a hypothesis of object $\mathcal{M}_j$ at pose $\tau_i \in \mathcal{T}$ relative to $\mathcal{M}_i$. The similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$, $S_{oh}(\mathcal{M}_i, \mathcal{M}_j^{\tau_i})$ or simply $S_j^{\tau_i}$, can be defined as the number of votes for $\mathcal{M}_j^{\tau_i}$ given an uncertain instance of $\mathcal{M}_i$, $\mathcal{D}_u(\mathcal{M}_i, R_u(\cdot))$, i.e.,

$$S_j^{\tau_i} = \mathrm{VOTES}(\mathcal{M}_j^{\tau_i}; \mathcal{D}_u(\mathcal{M}_i, R_u(\cdot))).$$

Obviously, $S_j^{\tau_i}$ is a random variable. It can be easily shown that the bounds of $S_j^{\tau_i}$ are

$$\min(S_j^{\tau_i}) = \mid \{F_{jk}^{\tau_i} : S_{of}(\mathcal{M}_i, F_{jk}^{\tau_i}) = 1\} \mid$$
$$\max(S_j^{\tau_i}) = \mid \{F_{jk}^{\tau_i} : S_{of}(\mathcal{M}_i, F_{jk}^{\tau_i}) > 0\} \mid,$$

and its expected value is

$$E(S_j^{\tau_i}) \approx \sum_k S_{of}(\mathcal{M}_i, F_{jk}^{\tau_i}),$$

where $F_{jk}^{\tau_i} \in \mathcal{M}_j^{\tau_i}$. An example of object/hypothesis similarity is shown in Fig. 4.

- **Uniform Model of Object/Hypothesis Similarity.** In order to simplify the process of performance prediction, described in Section 6, we assume the following:

1. The uncertainty regions associated with the features of each of $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ are nonoverlapping.
2. There is a one-to-one correspondence between similar features in $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$.
3. The similarity between every pair of similar features is the average of object/feature similarity

Initialize global similarity histogram SH

for each model object $\mathcal{M}_i \in \mathcal{DB}$ **do**

  Initialize local similarity histogram for object $\mathcal{M}_i$, SH_i

  for each model object $\mathcal{M}_j \in \mathcal{DB}$ **do**

    for each $\tau_i \in \mathcal{T}$ such that $\mid \mathcal{M}_j^{\tau_i} \cap R_c \mid > 0$ **do**

      if $(i \neq j) \vee \neg(\tau_i \in \mathcal{T}_{acc})$ **then**

        Compute similarity parameters $(N_j^{\tau_i}, P_j^{\tau_i})$

        Increment $SH_i(|\mathcal{M}_i|, |\mathcal{M}_j^{\tau_i} \cap R_c|, N_j^{\tau_i}, \lfloor N_j^{\tau_i} P_j^{\tau_i} + \frac{1}{2} \rfloor)$ by 1

      end if

    end for

  end for

  Add SH_i to SH

end for

Fig. 6. Similarity-computation algorithm.

considering features of $\mathcal{M}_j^{\tau_i}$ that are similar to features in $\mathcal{M}_i$. This average is denoted by $P_j^{\tau_i}$.

The above assumptions result in a "uniform" model of the similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$. As an example, Fig. 5 illustrates the uniform model of similarity corresponding to the object/hypothesis pair shown in Fig. 4. Based on such a model, we can approximate the PDF of $S_j^{\tau_i}$ by the following binomial distribution:

$$P_{S_j^{\tau_i}}(s_j^{\tau_i}) = B_{S_j^{\tau_i}}(s_j^{\tau_i}; N_j^{\tau_i}, P_j^{\tau_i}),$$

where

$$P_X(x) = \Pr[X = x],$$

$$N_j^{\tau_i} = \max(S_j^{\tau_i}),$$

$$P_j^{\tau_i} = \frac{E(S_j^{\tau_i})}{N_j^{\tau_i}},$$

$$B_X(x; n, p) = K(n, x) p^x (1-p)^{n-x},$$

and $K(a, b) = \frac{a!}{(a-b)!\, b!}$.

5.2 Construction of the Similarity Histogram

As discussed in the previous section, the similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ is described using two parameters, $(N_j^{\tau_i}, P_j^{\tau_i})$, which define the associated binomial distribution. For our purpose of performance prediction, two additional parameters are added: 1) The size of $\mathcal{M}_i$, $|\mathcal{M}_i|$, and 2) the *effective size* of $\mathcal{M}_j^{\tau_i}$, which is simply the number of features that can contribute votes to $\mathcal{M}_j^{\tau_i}$. Potential vote-contributing features of $\mathcal{M}_j^{\tau_i}$ are those which lie inside the clutter region R_c, i.e., $\mathcal{M}_j^{\tau_i} \cap R_c$ and so the effective size of $\mathcal{M}_j^{\tau_i}$ is simply $|\mathcal{M}_j^{\tau_i} \cap R_c|$.[2] Similarity information is accumulated in a 4D histogram, whose dimensions correspond to $(|\mathcal{M}_i|, |\mathcal{M}_j^{\tau_i} \cap R_c|, N_j^{\tau_i}, N_j^{\tau_i} P_j^{\tau_i})$.

Fig. 6 shows the algorithm used to build the similarity histogram. For each model object $\mathcal{M}_i$, we calculate the similarity between $\mathcal{M}_i$ and all the *erroneous* hypotheses competing with it. These hypotheses are selected such that: 1) They have at least one feature inside the clutter region R_c, and 2) for those that belong to $\mathcal{M}_i$, the relative pose is outside the subspace of acceptable pose error $\mathcal{T}_{acc}$. The similarity information associated with $\mathcal{M}_i$ is accumulated in a local histogram, SH_i. These local histograms, for all $\mathcal{M}_i \in \mathcal{DB}$, are added to form a global similarity histogram, SH.

6 Computation of Performance Bound

In this section, we analyze the votes for $\mathcal{M}_i$ and a competing hypothesis $\mathcal{M}_j^{\tau_i}$, given a distorted instance of $\mathcal{M}_i$, and derive a lower bound on PCR.

6.1 Motivating Example

We start by presenting an example to illustrate the effects of data distortion and object similarity on the vote process. This example, which is illustrated in Fig. 7, can be described as follows:

- Initially, without any distortion, $\mathcal{M}_i$ gets a number of votes equal to its size, which is five. On the other

2. In the implementation, we have also included features of $\mathcal{M}_j^{\tau_i}$ outside R_c that are similar to features in $\mathcal{M}_i$. Obviously, these similar feature pairs are very close to the boundary of R_c.

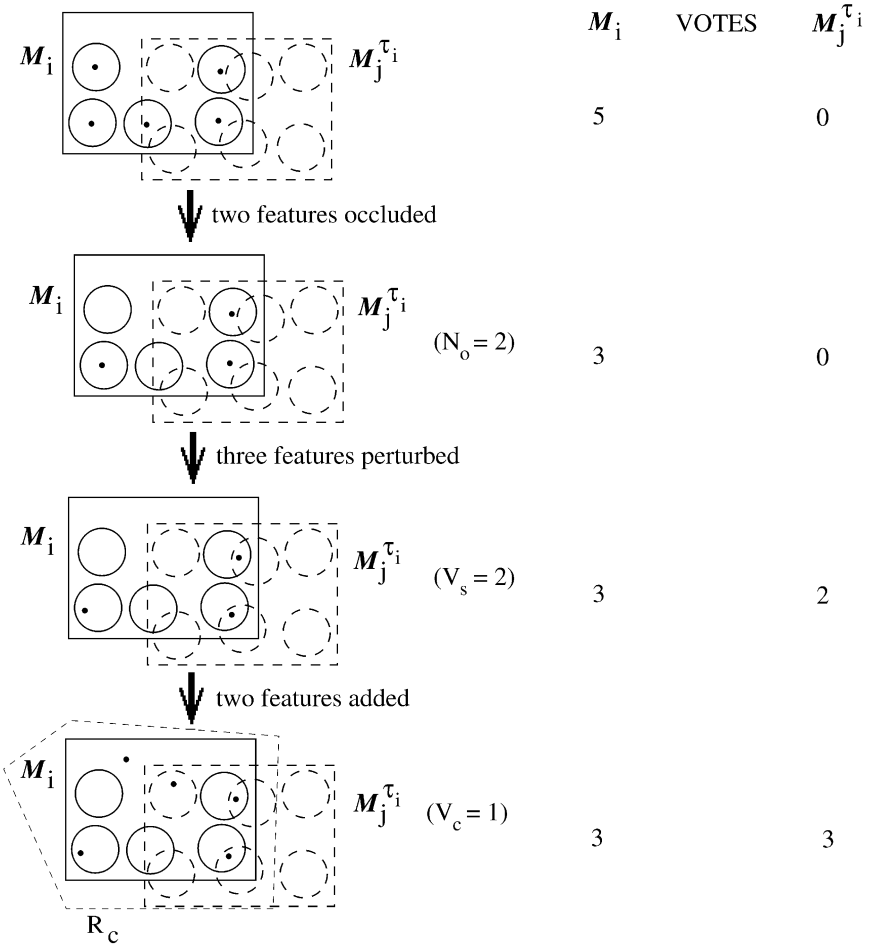

Fig. 7. An example of the vote process for $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$, as $\mathcal{M}_i$ gets distorted through occlusion, uncertainty, and clutter, respectively.

hand, $\mathcal{M}_j^{\tau_i}$ gets no votes, since there are no features of $\mathcal{M}_i$ inside the uncertainty regions of its features. Notice that we are considering the uniform model of similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$, which is described in Section 5.1.

- The first distortion step is to occlude two features in $\mathcal{M}_i$. Obviously, this reduces the number of votes for $\mathcal{M}_i$ to three and contributes no votes to $\mathcal{M}_j^{\tau_i}$. In addition, notice that occlusion reduces the number of similar feature pairs from three (which is $N_j^{\tau_i}$) to two.
- The second step in distortion is to perturb unoccluded features of $\mathcal{M}_i$ randomly within their uncertainty regions. Notice that, in Fig. 7, each of the two unoccluded similar features in $\mathcal{M}_i$ moves to the region overlapping with the uncertainty region of the corresponding feature in $\mathcal{M}_j^{\tau_i}$. This results in two votes for $\mathcal{M}_j^{\tau_i}$ and no change in the votes for $\mathcal{M}_i$.
- The third and final distortion step is to add two clutter features. It can be observed that one of them happens to fall within the uncertainty region of a feature in $\mathcal{M}_j^{\tau_i}$, thus increasing the number of votes for $\mathcal{M}_j^{\tau_i}$ from two to three.

The above distortion process results in a "tie" between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$, which is considered a failure in recognition.

The above example gives us a valuable insight into the vote process. *First*, the number of votes for $\mathcal{M}_i$, denoted by V_i, is simply the number of unoccluded features. That is, given a distorted instance of $\mathcal{M}_i$, $\widehat{\mathcal{M}_i}(R_u(\cdot), O, C, R_c)$ or simply $\widehat{\mathcal{M}_i}$, we have

$$V_i \quad = \quad |\,\mathcal{M}_i\,| - O. \tag{1}$$

Second, the number of votes for $\mathcal{M}_j^{\tau_i}$, which we denote by $V_j^{\tau_i}$, is a random variable. The votes for $\mathcal{M}_j^{\tau_i}$ come from two sources: 1) object $\mathcal{M}_i$, due to similarity and 2) clutter features, due to random coincidence. Accordingly, we can express $V_j^{\tau_i}$ as

$$V_j^{\tau_i} \quad = \quad V_s + V_c, \tag{2}$$

where V_s and V_c are random variables corresponding to the numbers of similarity and clutter votes for $\mathcal{M}_j^{\tau_i}$, respectively. *Third*, the number of similarity votes, V_s, is bounded by another random variable, N_o, which is the number of *unoccluded similar features* ($N_o \leq N_j^{\tau_i}$). Notice that in the previous example, we have $N_o = 2$, $V_s = 2$, and $V_c = 1$. In the next section, we determine the PDF of $V_j^{\tau_i}$ as a function of N_o, V_s, and V_c.

6.2 Probability Distribution of Hypothesis Votes

The PDF of $V_j^{\tau_i}$ can be determined as follows: Since V_s depends on N_o, we can express the PDF of V_s as

$$P_{V_s}(v_s) = \sum_{n_o} P_{N_o}(n_o) P_{V_s}(v_s; n_o), \quad (3)$$

where $P_{V_s}(v_s; n_o) = \Pr[V_s = v_s; N_o = n_o]$. From (2) and (3), we can express the PDF of $V_j^{\tau_i}$ as

$$P_{V_j^{\tau_i}}(v_j^{\tau_i}) = \sum_{n_o} P_{N_o}(n_o) \sum_{v_s} P_{V_s}(v_s; n_o) P_{V_c}(v_j^{\tau_i} - v_s; n_o, v_s), \quad (4)$$

where $P_{V_c}(v_c; n_o, v_s) = \Pr[V_c = v_c; N_o = n_o, V_s = v_s]$.

The PDF of N_o and the conditional PDFs of V_s and V_c are determined based on the statistical data-distortion models defined in Section 4 and the uniform model of similarity described in Section 5.1.

- **PDF of N_o.** Assuming uniform occlusion and similarity models, the process of occluding O features in $\mathcal{M}_i$ can be viewed as picking O balls from an urn containing $N_j^{\tau_i}$ white balls and $(\mid \mathcal{M}_i \mid -N_j^{\tau_i})$ black balls without replacement, where the white and black balls correspond to similar and dissimilar features in $\mathcal{M}_i$, respectively. Accordingly, we can describe N_o by the following hypergeometric distribution:

$$P_{N_o}(n_o) = H_{N_o}(N_j^{\tau_i} - n_o; O, N_j^{\tau_i}, \mid \mathcal{M}_i \mid -N_j^{\tau_i}), \quad (5)$$

where

$$H_X(x; n, a, b) = \frac{K(a, x) K(b, n - x)}{K(a + b, n)}.$$

Note that

$$n_o \in [\max(0, N_j^{\tau_i} - O), \min(N_j^{\tau_i}, \mid \mathcal{M}_i \mid -O)].$$

- **Conditional PDF of V_s.** Based on the assumptions of uniform models for uncertainty and similarity, it can be easily shown that the conditional PDF of V_s is represented by the following binomial distribution:

$$P_{V_s}(v_s; n_o) = B_{V_s}(v_s; n_o, P_j^{\tau_i}).$$

Notice that $v_s \in [0, n_o]$, if $P_j^{\tau_i} < 1$, and $v_s = n_o$, if $P_j^{\tau_i} = 1$.

- **Conditional PDF of V_c.** Determining the conditional PDF of V_c is considerably more involved than the first two PDFs. The effective clutter region R_c' can be split into two subregions, R_{V_c}' and $R_c' - R_{V_c}'$, such that a clutter feature falling within the first (second) subregion will (will not) result in a vote for $\mathcal{M}_j^{\tau_i}$. Region R_{V_c}' is the union of the uncertainty regions associated with features in $\mathcal{M}_j^{\tau_i} \cap R_c$ that are *not* already contributing similarity votes. Due to our modeling of clutter, explained in Section 4, features in $\mathcal{M}_j^{\tau_i} \cap R_c$ that are similar to *occluded* features in $\mathcal{M}_i$ effectively have "truncated" uncertainty regions. An example of R_{V_c}' is shown in Fig. 8. Under the assumption of uniform similarity, it can be shown that: 1) The area of a truncated uncertainty region is $\mathrm{AREA}(R_u(\cdot))(1 - P_j^{\tau_i})$ and 2) the numbers of features in $\mathcal{M}_j^{\tau_i} \cap R_c$ with truncated and full uncertainty regions are $n_t = N_j^{\tau_i} - n_o$, and $n_f = \mid \mathcal{M}_j^{\tau_i} \cap R_c \mid - v_s - n_t$ respectively. Accordingly, the conditional PDF of V_c can be approximated by the following binomial distribution:

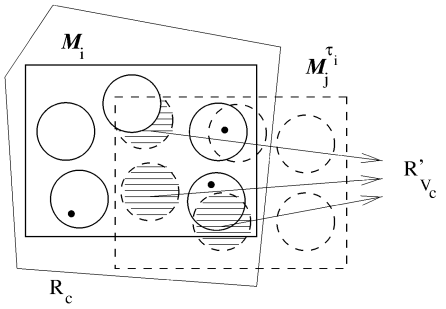

Fig. 8. An illustration of clutter vote region R_{V_c}'.

$$P_{V_c}(v_c; n_o, v_s) \approx B_{V_c}\left(v_c; C, \frac{\mathrm{AREA}(R_{V_c}')}{\mathrm{AREA}(R_c')}\right), \quad (6)$$

where

$$\mathrm{AREA}(R_{V_c}') = \mathrm{AREA}(R_u(\cdot))(n_f + (1 - P_j^{\tau_i}) n_t),$$

and

$$\mathrm{AREA}(R_c') = \mathrm{AREA}(R_c) - \mathrm{AREA}(R_u(\cdot)) O.$$

It is obvious that the lower bound of v_c is 0, while its upper bound is either $\min(n_f + n_t, C)$ or $\min(n_f, C)$, depending on whether $P_j^{\tau_i} < 1$, or $P_j^{\tau_i} = 1$, respectively.

6.3 Lower Bound on PCR

Let $\mathcal{N}_i$ be the set of erroneous object/pose hypotheses associated with $\mathcal{M}_i$. That is,

$$\mathcal{N}_i = \{\mathcal{M}_j^{\tau_i} : \mathcal{M}_j \in \mathcal{DB}, \text{and } \tau_i \in \mathcal{T} \text{ s.t. } \mid \mathcal{M}_j^{\tau_i} \cap R_c \mid > 0\} - \{\mathcal{M}_i^{\tau_i} : \tau_i \in \mathcal{T}_{acc}\}.$$

The probability of misinterpreting $\widehat{\mathcal{M}_i}$, as any hypothesis in $\mathcal{N}_i$, can be expressed as

$$\Pr[\mathcal{N}_i; \widehat{\mathcal{M}_i}] = \Pr[\exists\ \mathcal{M}_j^{\tau_i} \in \mathcal{N}_i \ \text{ s.t. } V_j^{\tau_i} \geq V_i]. \quad (7)$$

From (1) and (4), we can determine the probability that the votes for $\mathcal{M}_j^{\tau_i}$ reach or exceed those for $\mathcal{M}_i$:

$$\Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}_i}] \quad = \quad \sum_{v_j^{\tau_i} \geq \mid \mathcal{M}_i \mid - O} P_{V_j^{\tau_i}}(v_j^{\tau_i}). \quad (8)$$

The probability of recognition failure, $\Pr[\mathcal{N}_i; \widehat{\mathcal{M}_i}]$, can be bounded as follows (refer to (7)):

$$\Pr[\mathcal{N}_i; \widehat{\mathcal{M}_i}] < \sum_{\mathcal{M}_j^{\tau_i} \in \mathcal{N}_i} \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}_i}].$$

The above inequality can be directly used to determine the following lower bound on the probability of correctly recognizing $\mathcal{M}_i$, given $\widehat{\mathcal{M}_i}$:

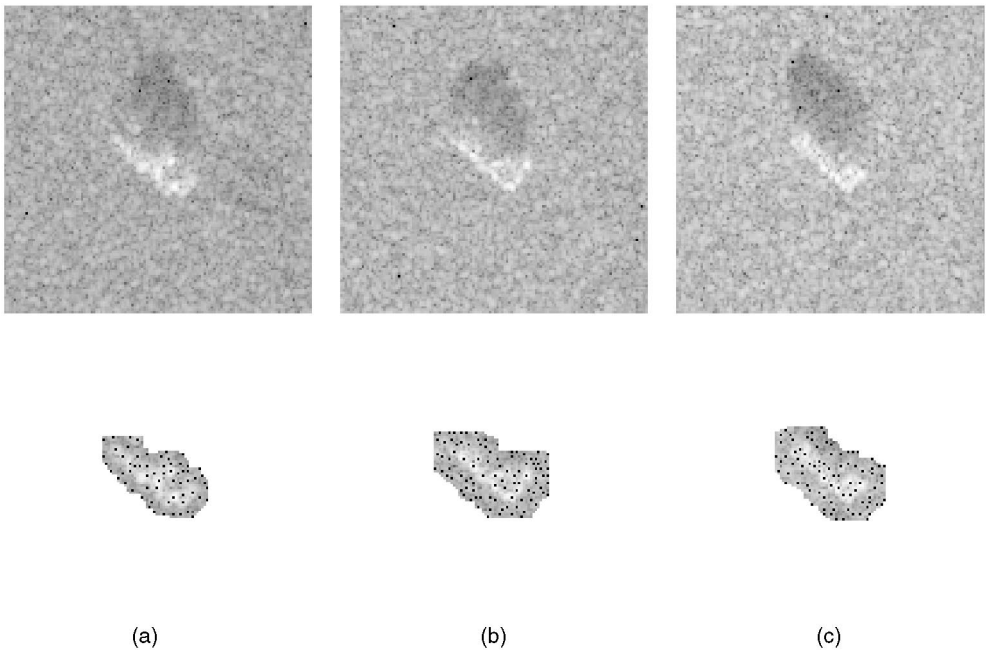

Fig. 9. Examples of SAR images at depression angle 17° and azimuth angle 132°. (a) T72, (b) BMP2, and (c) BTR70.

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}_i}] > 1 - \sum_{\mathcal{M}_j^{\tau_i} \in \mathcal{N}_i} \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}_i}]. \quad (9)$$

From the discussion in the previous section, it can be observed that $V_j^{\tau_i}$ and, in turn, $\Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}_i}]$ depend on only four object-dependent parameters: the size of $\mathcal{M}_i$, the effective size of $\mathcal{M}_j^{\tau_i}$, and the similarity parameters $(N_j^{\tau_i}, P_j^{\tau_i})$. Let

$$W(a,b,c,d) = \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}_i}],$$

such that $a = \mid \mathcal{M}_i \mid$ and $b = \mid \mathcal{M}_j^{\tau_i} \cap R_c \mid$, $c = N_j^{\tau_i}$ and $d = \lfloor N_j^{\tau_i} P_j^{\tau_i} + \frac{1}{2} \rfloor$. We can rewrite (9) as:

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}_i}] > 1 - \sum_a \sum_b \sum_c \sum_d SH_i(a,b,c,d) W(a,b,c,d). \quad (10)$$

Summing (10) for all model objects $\mathcal{M}_i \in \mathcal{DB}$ and taking the average, we obtain the following lower bound on average PCR for model set $\mathcal{DB}$, which is the main result of this paper:

$$\mathrm{PCR}(\mathcal{DB}) > 1 - \frac{1}{\mid \mathcal{DB} \mid} \sum_a \sum_b \sum_c \sum_d SH(a,b,c,d) W(a,b,c,d). \quad (11)$$

7 Experimental Validation

In this section, we validate the proposed performance-prediction method by comparing actual recognition performance with predicted lower bounds.

7.1 Recognition Task

Validation is presented in the context of a target recognition task using SAR images. These images are obtained from the MSTAR public data domain [20]. In such a task, each model target is represented by a number of SAR views, which sample its signature at a variety of azimuth angles, at a specific depression angle. For our purposes, we treat each view as an independent "object."

Recognition is performed by comparing a data view with the set of model views. The features used for recognition correspond to locations of peaks in a SAR view, which are commonly called *scattering centers*. These peaks are obtained by comparing the value of each pixel with its eight neighbors. Each view, model or data, is represented by the 30 peaks that are strongest in magnitude of radar returns. Notice that since we are considering a fixed number of peaks, the numbers of occluded and clutter features (O and C) in a view are always the same. The space of applicable transformations $\mathcal{T}$ is discrete 2D translation in the image plane [4], [13]. Only the exact location of the object is assumed to be acceptable; i.e., $\mathcal{T}_{acc} = \{\mathbf{0}\}$.

7.2 Model and Test Sets

We consider a model database consisting of three military targets: T72, BMP2, and BTR70. These targets are represented by 231, 233, and 233 available views, respectively, at depression angle 17°. Examples of these views along with associated peaks are shown in Fig. 9. For the test data, we have selected seven test sets. These sets can be classified into two groups, depending on the nature of the distortion in the data.

1. **Synthetic-Distortion Group.** This group, sets $\mathcal{TD}_1$ through $\mathcal{TD}_5$, is obtained by introducing synthetic distortion to the model data. This is done according to the distortion process described in Section 4, with the additional constraint that features can not be eight-neighbors, in order to simulate the peak-extraction process. Each test set corresponds to specific uncertainty and clutter regions and a variety of occlusion/clutter (O/C) values. It is constructed by distorting each model view four times, for every O/C value. Accordingly, the size of a test set is $4\times \mid \mathcal{DB} \mid \times N_{oc}$, where N_{oc} is the number of selected

TABLE 2
Description of the Test Sets Used in the Experiments

Set	Distortion	Uncertainty Region	Clutter Region		Occluded/ Clutter Features	Size
			Shape	Factor		
$\mathcal{TD}_1$	synthetic	0-neighbor	convex hull	1	$18, 19, \cdots, 27$	$4 \times 697 \times 10$
$\mathcal{TD}_2$	synthetic	4-neighbor	convex hull	1	$9, 10, \cdots, 20$	$4 \times 697 \times 12$
$\mathcal{TD}_3$	synthetic	8-neighbor	convex hull	1	$0, 1, \cdots, 15$	$4 \times 697 \times 16$
$\mathcal{TD}_4$	synthetic	4-neighbor	convex hull	2	$9, 10, \cdots, 24$	$4 \times 697 \times 16$
$\mathcal{TD}_5$	synthetic	4-neighbor	convex hull	3	$9, 10, \cdots, 24$	$4 \times 697 \times 16$
$\mathcal{TD}_6$	Δ configuration	4-neighbor	convex hull	1	estimated	464
$\mathcal{TD}_7$	Δ depression angle	4-neighbor	convex hull	1	estimated	581

O/C values and $\mid \mathcal{DB} \mid = 697$ in our case. These sets are described in Table 2. In this table, note that the clutter region is a dilation of the convex hull of model features, i.e., a scaling of the convex hull by some factor assuming its center as the reference point. Furthermore, notice that $R_u(\cdot)$ for $\mathcal{TD}_1$ is a zero-neighbor region, which implies that there is no positional uncertainty in the features.

2. **Real-Distortion Group.** This group, sets $\mathcal{TD}_6$ and $\mathcal{TD}_7$, corresponds to variants of the model set. Set $\mathcal{TD}_6$ is obtained by changing target configurations [4]. A configuration of a target can be changed in several ways such as changing the number of fuel barrels, using different flash lights, etc. Set $\mathcal{TD}_7$ is obtained by changing the depression angle from 17° to 15°. It is well-known that these changes can significantly alter the view structure [20]. The distortion parameters $R_u(\cdot)$ and R_c are empirically chosen as $R_u(\cdot) =$ four-neighbor region, and $R_c =$ convex hull of view features. The O/C value for each test view is estimated through finding the best match with model views within a difference of $\pm 3^\circ$ azimuth angles. If no views exist within this range, then the test view is matched with the model view that is nearest in azimuth. The real sets are also summarized in Table 2. Notice that since both test and model views are real, the notion of occluded and clutter features is relative not absolute. In other words, a true feature in a test view that is missing in the corresponding model view is considered as a clutter feature. Similarly, a spurious feature in a model view that does not correspond to any feature in the corresponding test view is considered as an occluded model feature. It is obvious that the overall recognition performance can be improved through learning the true model features, a subject which is beyond the scope of this paper.

7.3 Results

In this section, we compare actual performance, determined experimentally, with predicted lower bounds. The performance metric we use is PCR as a function of O/C rate, for given uncertainty and clutter regions. Actual PCR plots are determined using a recognition system that examines *all* the relevant subspace of the 2D translation space, for each pair of test and model views. This relevant subspace is the bounding box on the translations that lead to at least one pair of consistent test and model features. Accordingly, the performance of this system is *optimal*, for the given vote-based criterion. The predicted PCR plot is obtained as described in Section 6 with two modifications. *First*, notice that the shape of R_c in the test sets is view dependent. Accordingly, we substitute $\mathrm{AREA}(R_c)$ by the average of clutter-region areas corresponding to views of the model set.[3] *Second*, the above mentioned peak-extraction process imposes the constraint that features can not be eight-neighbors of each other. Accordingly, the conditional clutter-vote PDF, defined in (6), needs to be modified in order to consider such a constraint. Unfortunately, this significantly increases the complexity of estimating the clutter-vote PDF. In the Appendix, we present an approximate method for estimating such a PDF.

We start by analyzing the results obtained using the synthetic-distortion group. Fig. 10 shows actual and predicted PCR plots for three uncertainty regions, where the clutter region is assumed to be the convex hull of view features (test sets $\mathcal{TD}_1$, $\mathcal{TD}_2$, and $\mathcal{TD}_3$). Fig. 11 shows actual and predicted PCR plots for three different clutter regions, assuming a four-neighbor uncertainty region (test sets $\mathcal{TD}_2$, $\mathcal{TD}_4$, and $\mathcal{TD}_5$). From both of these figures, we observe that the proposed method consistently predicts reasonably tight bounds on PCR performance. In particular, we observe the following:

3. This average is found to be 532, $2,043$, and $4,532$ pixels for convex-hull regions at scale factors of 1, 2, and 3, respectively.

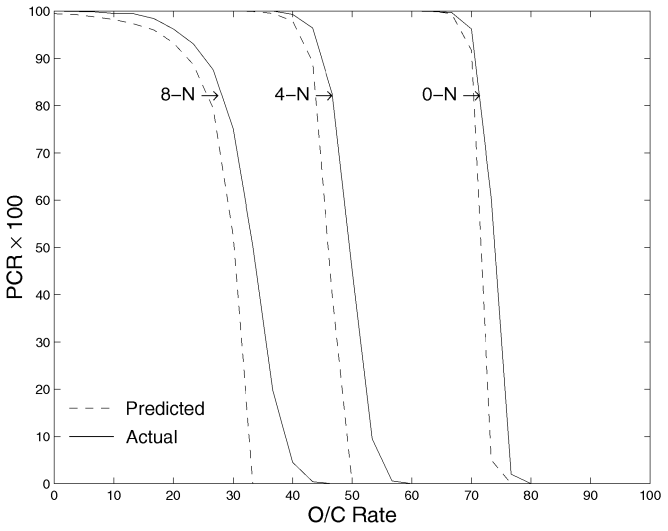

Fig. 10. Actual and predicted PCR plots for a variety of uncertainty regions.

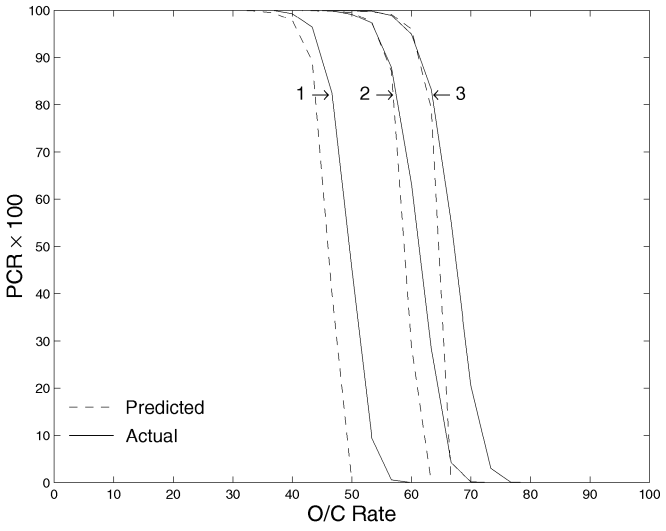

Fig. 11. Actual and predicted PCR plots assuming a convex-hull clutter region at a variety of scale factors (the numbers associated with the plots represent the corresponding scale factor of the clutter region).

- The breakpoint in performance is consistently predicted with high accuracy. In all cases, the predicted plot breaks either exactly with the actual one, or very slightly before it.
- The predicted PCR is very close to the actual one along the knee section of the PCR plot (the section between the first two breaks in the plot). It can be observed that the difference between the two values is consistently less than 3 percent.
- The predicted and actual PCR plots tend to diverge beyond the knee section. This can be explained as follows: At low distortion levels, the probability that several hypotheses simultaneously get more votes than the true one is negligible compared to that of a single hypothesis (we assume here that the model data set does not consist of identical or very similar views). Accordingly, in such a case, the lower bound defined in (11) is an estimate of the actual performance. However, as distortion increases, the probability that several erroneous hypotheses simultaneously get more votes than the true one becomes nonnegligible and so (11) becomes a strict lower bound on performance.
- The results shown in Fig. 10 confirm the intuition that performance is inversely proportional to the size of the uncertainty region. Furthermore, the results in Fig. 11 show that performance is directly proportional to the size of the clutter region. This is expected since, given a fixed number of clutter features, the increase in the size of the clutter region corresponds to a decrease in their density.

Now, we analyze the results obtained using the real-distortion group. Figs. 12a and 12b show actual and predicted PCR plots for real test sets $\mathcal{TD}_6$ and $\mathcal{TD}_7$, respectively. From these figures, we observe that the proposed method does a reasonable job of predicting performance. For example, it predicts the breakpoint accurately in both cases. However, we also observe that the predicted PCR bound is slightly overoptimistic along the knee section of the plot. This is due to differences between the actual distortion models and the uniform models assumed by our method. One of the important differences is that the assumed occlusion model does not capture the spatial correlation among occluded features. More accurate prediction can be obtained by learning the distortion models. This topic has received attention in the literature, mainly on learning distributions of positional uncertainty (e.g., [19]). This is due to its critical importance in designing object recognition systems [2].

A goal of our research is to develop statistical techniques for learning uncertainty, occlusion, and clutter models. In this work, we present an initial method for learning the occlusion model. The straightforward approach is to learn the spatially-correlated occlusion model and then use it to determine the PDF of unoccluded similar features (i.e., the PDF of N_o, see Section 6.2). The difficulty of this approach is that it involves learning not only the spatially-correlated occlusion model, but also the spatial correlation between similar feature pairs. Fortunately, there is a substantially simpler technique in our context (performance prediction of vote-based object recognition): Instead of learning the spatially-correlated occlusion and similarity models and then inferring the PDF of N_o, we can directly learn the PDF of N_o. This process is performed as follows: 1) Half of the test set, the odd-numbered views, is selected for learning. 2) For each selected test view, $\widehat{\mathcal{M}_i}$, we find the best matching model view, $\mathcal{M}_i$, as described in Section 7.2. From this match, we construct $\overline{\mathcal{M}}_i$, which is view $\mathcal{M}_i$ after occlusion (without either uncertainty or clutter). 3) For each erroneous hypothesis $\mathcal{M}_j^{\tau_i} \in \mathcal{N}_i$, we calculate the number of similar feature pairs between $\mathcal{M}_j^{\tau_i}$ and each of $\mathcal{M}_i$ and $\overline{\mathcal{M}}_i$, $N_j^{\tau_i}$ and $\bar{N}_j^{\tau_i}$, respectively. 4) The values of $\bar{N}_j^{\tau_i}$ are used to estimate the PDF of N_o, as a function of $\mid \mathcal{M}_i \mid$, $O = \mid \mathcal{M}_i \mid - \mid \overline{\mathcal{M}}_i \mid$, and $N_j^{\tau_i}$ (refer to Section 6.2). Using the learned PDF of N_o instead of the one defined in (5), we obtain the bounds shown in Figs. 13a and 13b, for test sets $\mathcal{TD}_6$ and $\mathcal{TD}_7$, respectively.[4]

4. In the very rare case of missing learned PDFs, the one defined in (5) is used.

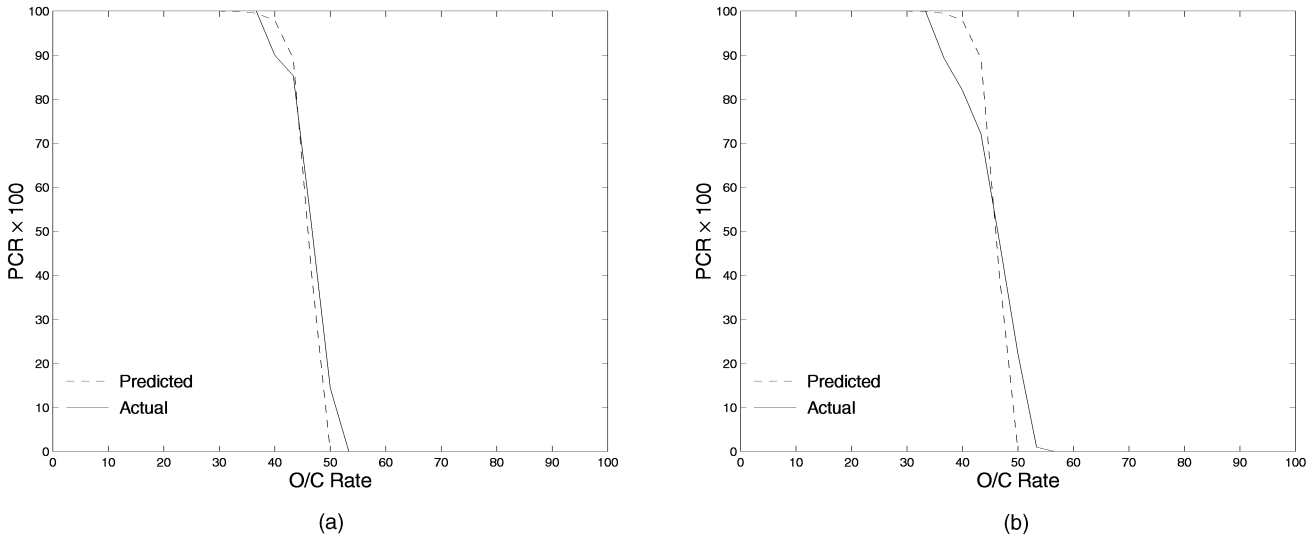

Fig. 12. Actual and predicted PCR plots for real-distortion sets: (a) configuration differences ($\mathcal{TD}_6$) and (b) depression angle change ($\mathcal{TD}_7$).

From these figures, we observe that the predicted bounds become considerably more accurate, compared to those obtained by assuming uniform distortion models (Fig. 12). In particular, the discrepancy almost vanished for set $\mathcal{TD}_6$ and decreased by about 50 percent for set $\mathcal{TD}_7$, where the extent of discrepancy is measured by the area between actual and predicted plots at the knee section. Learning the other distortion models, including the shape of the clutter region, is expected to eliminate the remaining slight overestimation of performance.

8 Conclusions

Most previous efforts for predicting object recognition performance focused on the problem of object/clutter discrimination. This paper extends these efforts by addressing the problem of object/object discrimination. The method presented in this paper considers data distortion factors such as uncertainty, occlusion, and clutter, in addition to object similarity. This is in contrast to the very few relevant approaches in this area, which consider only a subset of these factors. A target recognition task involving real SAR data and point features is used for validating the proposed method. Validation is done by comparing predicted PCR plots with ones determined experimentally, for a number of test sets involving both synthetic and real distortion. For all the test sets, the results show that the proposed method consistently predicts reasonably tight bounds on actual performance. However, for the real-distortion sets, the predicted bounds slightly overestimate performance around the knee section of the plot because of differences between assumed and actual distortion models. As an initial step towards learning the distortion models, we have presented a method for learning the PDF of the random variable that encodes the combined effect of similarity and occlusion on the vote process. We have shown that the use of the learned PDF leads to more accurate performance predictions. To the best of our knowledge, this work is the first which validated an object-recognition prediction method using real data. In addition to learning distortion models, future research work includes: 1) further validation of the method using additional real data and 2) prediction of other performance metrics such as an upper bound on PCR (a possible approach that is specific to SAR target recognition is described in [5]) and the confusion matrix.

Finally, we discuss the generalization aspects of the proposed method. This method has been presented in the context of a 2D/2D recognition problem involving discretized point features and a translation space. This context is selected, simply because it reflected the recognition task at hand. The proposed method, however, can be generalized to other recognition tasks as follows: *First,* consideration of arbitrary transformation spaces is straightforward. The only change that needs to be made is in the selection of the set of erroneous hypotheses (see Section 6.3). *Second,* handling of nondiscretized features is slightly more involved as the set of erroneous hypotheses in this case will be of infinite size. This set can be made finite by sampling the transformation space and appropriately "dilating" the uncertainty regions associated with the sampled hypotheses to account for the nonsampled ones (e.g., see [16]). *Third,* feature attributes (e.g., peak magnitude, etc.) can be modeled by including them in the uncertainty and clutter regions. In this case, the dimensionality of these regions will increase from 2 to $2+n$, where n is the number of attributes. Other than that, the method remains conceptually the same. *Fourth,* consideration of the 3D/3D problem will, as in the previous case, involve increasing the dimensionality of the uncertainty and clutter regions from 2 to 3. In conclusion, we believe that the work presented in this paper lays a theoretical/conceptual foundation that is necessary to solve the problem of predicting object recognition performance.

Appendix

The feature adjacency constraint can be represented by a *separation region*, $R_s(\cdot)$. For example, if features can not occupy the same location, then $R_s(\cdot)$ is simply a 1×1 window centered at the feature's location. Another example, if the

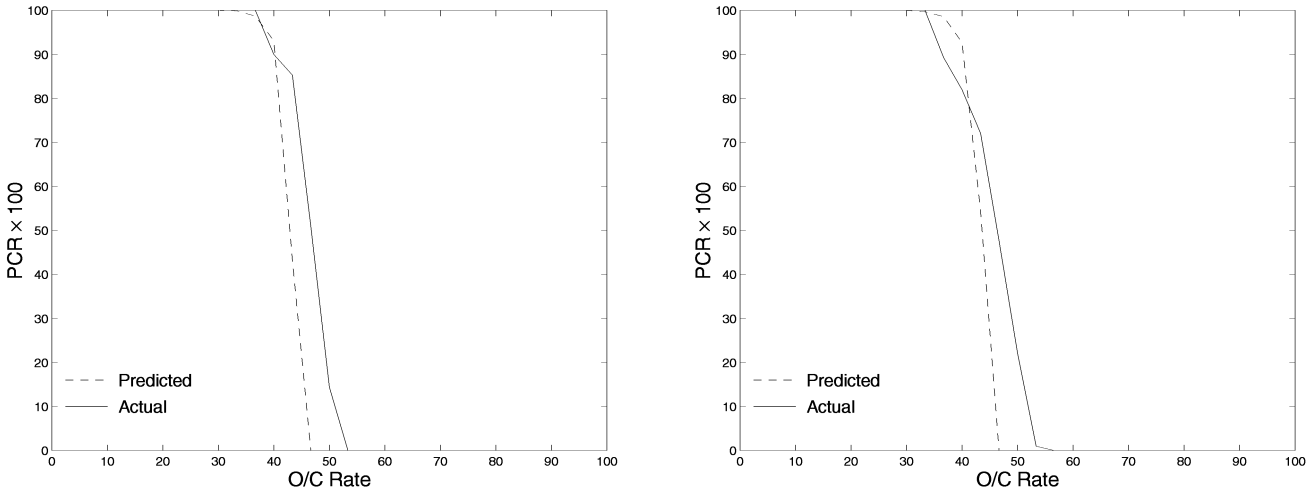

Fig. 13. Actual and predicted PCR plots for real-distortion sets after learning the PDF of N_o: (a) configuration differences ($\mathcal{TD}_6$), (b) depression angle change ($\mathcal{TD}_7$).

point features used correspond to image peaks (as in our experiments, see Section 7), then no two features can be eight neighbors. In such a case, $R_s(\cdot)$ is a 3×3 window centered at the feature's location.

The conditional PDF of V_c can be approximated as follows:

- The feature-adjacency constraint results in smaller effective clutter and clutter-vote regions, which we denote by R''_c and R''_{V_c}, respectively. The area of R''_c can be expressed as

$$\mathrm{AREA}(R''_c) \approx \mathrm{AREA}(R'_c) - \mathrm{AREA}(R_s(\cdot))(\mid \mathcal{M}_i \mid -O).$$

  Regarding the clutter-vote region, R''_{V_c}, consider the subset of features in $\mathcal{M}_i$ that are similar to those in $\mathcal{M}_j^{\tau_i}$, but are lying *outside* the overlapping regions (see Fig. 8). Typically, the separation region associated with one of the features in $\mathcal{M}_i$ covers a section of the uncertainty region associated with the corresponding similar feature in $\mathcal{M}_j^{\tau_i}$. The area of that covered section, E_d, is approximately estimated as follows: 1) The shapes of the uncertainty and separation regions are modeled as squares, whose areas are the same as the original respective regions. 2) For a pair of similar features, the relationship between their square uncertainty regions is obtained by moving one region relative to the other along one of the sides of the reference square region, until the area of the overlapping region corresponds to the similarity between the two features. 3) The expected location of the feature of $\mathcal{M}_i$, given that it lies outside the overlapping region, is estimated. 4) The area of the covered section, E_d, is then easily calculated. The area of R''_{V_c} is estimated as follows:

$$\mathrm{AREA}(R''_{V_c}) = \mathrm{AREA}(R'_{V_c}) - (n_o - v_s)E_d.$$

- From the above discussion and that in Section 6.2, potential vote-contributing features of hypothesis $\mathcal{M}_j^{\tau_i}$ are effectively associated with uncertainty regions of different sizes. One way of simplifying the computations is to fix the number of features, and assume that they are associated with identical uncertainty regions, the area of each is $\mathrm{AREA}(R''_{V_c})/(n_t + n_f)$. However, as we experimentally found out, this can lead to slightly overoptimistic results (if there is a significant difference between the areas of full and truncated uncertainty regions). A more accurate approximation can be obtained by fixing the uncertainty-region area, and assuming that we have an "effective" number of potential vote-contributing features, n_e, where $n_e = \lfloor \mathrm{AREA}(R''_{V_c})/\mathrm{AREA}(R_u(\cdot)) + 0.5 \rfloor$. In such a case, the average area of an uncertainty region, E_u, is $\mathrm{AREA}(R''_{V_c})/n_e$, which is approximately the same as $\mathrm{AREA}(R_u(\cdot))$.
- Next, we estimate the *effective* area of a separation region, E_s, which is simply the ratio of the union of the separation regions associated with the C clutter features to C. If the density of the clutter features $C/\mathrm{AREA}(R''_c)$ is low, then E_s is close to $\mathrm{AREA}(R_s(\cdot))$. However, for high densities, which is typical in our data, E_s can be considerably smaller than $\mathrm{AREA}(R_s(\cdot))$, due to the overlapping between the separation regions. In our work, we have $E_s = \min(\mathrm{AREA}(R_s(\cdot)), \mathrm{AREA}(R''_c)/C)$.
- The conditional PDF of V_c can be approximated as

$$P_{V_c}(v_c; n_o, v_s) \approx K(C, v_c) \frac{L(\mathrm{AREA}(R''_{V_c}), E_u, v_c) L(\mathrm{AREA}(R''_c - R''_{V_c}) - v_c(\max(0, E_s - E_u)), E_s, C - v_c)}{L(\mathrm{AREA}(R''_c), E_s, C)},$$

  where $L(X, Y, n) = \prod_{i=0}^{n-1}(X - iY)$.

ACKNOWLEDGMENTS

This work was supported in part by DARPA/AFOSR grant F49620-97-1-0184. The contents and information do not

reflect positions or policies of the US government. This research was conducted while Michael Boshra was with the Center for Research in Intelligent Systems, University of California, Riverside.

REFERENCES

[1] T.D. Alter and W.E.L. Grimson, "Verifying Model-Based Alignments in the Presence of Uncertainty," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 344-349, 1997.
[2] T.D. Alter and D.W. Jacobs, "Uncertainty Propagation in Model-Based Recognition," *Int'l J. of Computer Vision,* vol. 27, no. 2, pp. 127-159, 1998.
[3] A. Amir and M. Lindenbaum, "Grouping-Based Nonadditive Verification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 2, pp. 186-192, Feb. 1998.
[4] B. Bhanu and G. Jones III, "Recognizing MSTAR Target Variants and Articulations," *Proc. SPIE Conf. Algorithms for Synthetic Aperture Radar Imagery VI,* vol. 3,721, pp. 507-519, 1999.
[5] M. Boshra and B. Bhanu, "Bounding SAR ATR Performance Based on Model Similarity," *Proc. SPIE Conf. Algorithms for Synthetic Aperture Radar Imagery VI,* vol. 3,721, pp. 716-729, 1999.
[6] Y. Boykov and D.P. Huttenlocher, "A New Bayesian Framework for Object Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 517-523, June 1999.
[7] T.M. Breuel, "Higher-Order Statistics in Object Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 707-708, 1993.
[8] M. Dhome, M. Richetin, J. Lapreste, and G. Rives, "Determination of the Attitude of 3D Objects from a Single Perspective View," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 11, no. 12, pp. 1,265-1,278, Dec. 1989.
[9] O.D. Faugeras and M. Hebert, "The Representation, Recognition, and Locating of 3D Objects," *Int'l J. of Robotic Res.,* vol. 5, no. 3, pp. 27-52, 1986.
[10] W.E.L. Grimson and D.P. Huttenlocher, "On the Verification of Hypothesized Matches in Model-Based Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 12, pp. 1,201-1,213, Dec. 1991.
[11] W.E.L. Grimson and T. Lozano-Perez, "Localizing Overlapping Parts by Searching the Interpretation Tree," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, no. 4, pp. 469-482, July. 1987.
[12] D.P. Huttenlocher and S. Ullman, "Recognizing Solid Objects by Alignment with an Image," *Int'l J. of Computer Vision,* vol. 5, no. 2, pp. 195-212, 1990.
[13] G. Jones III and B. Bhanu, "Recognition of Articulated and Occluded Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 7, pp. 603-613, July 1999.
[14] W.M. Wells III, "Statistical Approaches to Feature-Based Object Recognition," *Int'l J. of Computer Vision,* vol. 21, no. 1, pp. 63-98, 1997.
[15] S.Z. Li, *Markov Random Field Modeling in Computer Vision.* New York: Springer-Verlag, 1995.
[16] M. Lindenbaum, "Bounds on Shape Recognition Performance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 7, pp. 666-680, July 1995.
[17] M. Lindenbaum, "An Integrated Model for Evaluating the Amount of Data Required for Reliable Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 11, pp. 1,251-1,264, Nov. 1997.
[18] J. Mao, P.J. Flynn, and A.K. Jain, "Integration of Multiple Feature Groups and Multiple Views into a 3D Object Recognition System," *Computer Vision and Image Understanding,* vol. 62, no. 3, pp. 309-325, 1995.
[19] A.R. Pope and D.G. Lowe, "Learning Appearance Models for Object Recognition," *Proc. Int'l Workshop Object Representation for Computer Vision,* pp. 201-219, 1996.
[20] T. Ross, S. Worrell, V. Velten, J. Mossing, and M. Bryant, "Standard SAR ATR Evaluation Experiments Using the MSTAR Public Release Data Set," *Proc. SPIE Conf. Algorithms for Synthetic Aperture Radar Imagery V,* vol. 3,370, pp. 566-573, 1998.
[21] K.B. Sarachik, "The Effect of Gaussian Error in Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 4, pp. 289-301, Apr. 1997.
[22] G. Stockman, "Object Recognition and Localization via Pose Clustering," *Computer Vision, Graphics, and Image Processing,* vol. 40, no. 3, pp. 361-387, 1987.

Michael Boshra received the BSc and MSc degrees in computer science from the University of Alexandria, Egypt, in 1988 and 1992, respectively, and the PhD degree in computing science from the University of Alberta, Canada, in 1997. He is currently a technical staff member at AuthenTec Inc., working on the development of algorithms for fingerprint identification. Prior to joining AuthenTec, he was a postdoctoral fellow at the Center for Research in Intelligent Systems (CRIS) at the University of California, Riverside from 1997 to 1999. From 1989 to 1992, he was a research assistant at the National Research Center, Giza, Egypt. His current research interests include fingerprint identification, object recognition, performance prediction, and multidimensional indexing. He is a member of the IEEE and ACM.

Bir Bhanu (S'72-M'82-SM'87-F'96) received the SM and EE degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, the PhD degree in electrical engineering from the Image Processing Institute, University of Southern California, Los Angeles, and the MBA degree from the University of California, Irvine. He also received the BS degree (with honors) in electronics engineering from the Institute of Technology, BHU, Varanasi, India, and the ME degree (with distinction) in electronics engineering from Birla Institute of Technology and Science, Pilani, India. Currently, Dr. Bhanu is the director of the Center for Research in Intelligent Systems (CRIS) at the University of California, Riverside where he has been a professor and director of Visualization and Intelligent Systems Laboratory (VISLab) since 1991. Previously, he was a senior Honeywell fellow at Honeywell Inc. in Minneapolis, Minnesota. He has been on the faculty of the Department of Computer Science at the University of Utah, Salt Lake City, and has worked at Ford Aerospace and Communications Corporation, in California, INRIA-France and IBM San Jose Research Laboratory, in California. He has been the principal investigator of various programs for DARPA, NASA, NSF, AFOSR, ARO and other agencies and industries in the areas of learning and vision, image understanding, pattern recognition, target recognition, navigation, image databases, and machine vision applications. He is the coauthor of books on *Computational Learning for Adaptive Computer Vision* (forthcoming), *Genetic Learning for Adaptive Image Segmentation,* (Kluwer 1994), and *Qualitative Motion Understanding,* (Kluwer 1992). He has received two outstanding paper awards from the Pattern Recognition Society and has received industrial awards for technical excellence, outstanding contributions, and team efforts. He has been the guest editor of several IEEE transactions and other journals and on the editorial board of various journals. He holds 10 US and international patents and more than 200 reviewed technical publications in the areas of his interest. He has been general chair for IEEE Workshops on Applications of Computer Vision, chair for the DARPA Image Understanding Workshop, general chair for the IEEE Conference on Computer Vision and Pattern Recognition, and program chair for the IEEE Workshops on Computer Vision Beyond the Visible Spectrum. Dr. Bhanu is a fellow of IEEE, AAAS, and IAPR, and a member of ACM, AAAI, and SPIE.