Chapter 1

A Theoretical Framework for Predicting Performance of Object Recognition

Michael Boshra[1] and Bir Bhanu[2]

[1] Center for Research in Intelligent Systems, University of California, Riverside, California 92521, michael@cris.ucr.edu
[2] Center for Research in Intelligent Systems, University of California, Riverside, California 92521, bhanu@cris.ucr.edu

Summary. The ability to predict the fundamental performance of model-based object recognition is essential for transforming the object recognition field from an art to a science, and to speed up the design process for recognition systems. In this chapter, we address the performance–prediction problem in the context of a common recognition task, where both model objects and scene data are represented by locations of 2D point features. The criterion used for estimating matching quality is based on the number of consistent data/model feature pairs, which we refer to as "votes." We present a theoretical framework for prediction of lower and upper bounds on the probability of correctly recognizing model objects from scene data. The proposed framework considers data distortion factors such as uncertainty (noise in feature locations), occlusion (missing features), and clutter (spurious features). In addition, it considers structural similarity between model objects. The framework consists of two stages. In the first stage, we calculate a measure of the structural similarity between every pair of objects in the model set. This measure is a function of the relative transformation between the model objects. In the second stage, the model similarity information is used along with statistical models of the data distortion factors to determine bounds on the probability of correct recognition. The proposed framework is compared with relevant research efforts. Its validity is demonstrated using real synthetic aperture radar (SAR) data from the MSTAR public domain, which are obtained under a variety of depression angles and object configurations.

1.1 Introduction

Model-based object recognition has been an active area of research for over two decades (e.g., see surveys [1, 2, 3, 4]). It is concerned with finding instances of known model objects in scene data. This process involves extracting features from the scene data, and comparing them with those of the model objects using some matching criterion. Performance of the recognition process depends on the amount of distortion in the data features. Data distortion

can be classified into three types: (1) *uncertainty:* noise in feature locations and other feature attributes; (2) *occlusion:* missing features of data object of interest; and (3) *clutter:* spurious data features which do not belong to the data object of interest. In addition to data distortion, recognition performance depends on the degree of structural similarity between model objects. The often-overlooked similarity factor can have a profound impact on performance. Intuitively, the difficulty of recognizing a specific object is proportional to the degree of its similarity with the rest of the objects in the model set.

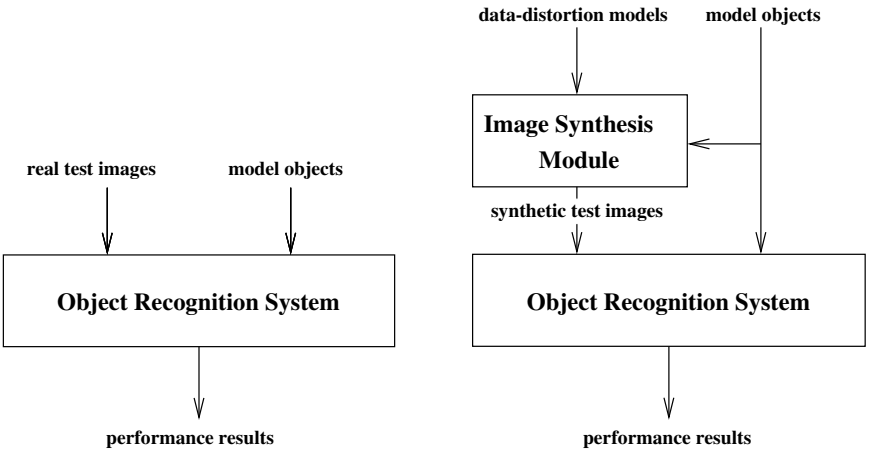


Figure 1.1. Empirical approaches for estimation of object recognition performance: (a) using real data, (b) using synthetic data.

Performance of object recognition is typically estimated empirically. This is done by passing a set of scene images containing known model objects to a recognition system, and then analyzing the output of the system. The set of scene images can be either real [5, 6, 7], or synthetic with artificial distortion introduced to them [8, 9, 10]. Both scenarios are illustrated in Figures 1.1(a) and 1.1(b), respectively. Empirical performance evaluation has a number of limitations:

1. It does not provide an understanding of the relationship between object recognition performance and the various data and model factors that affect it. In other words, the empirical approach can provide an answer to the question of *what* performance to expect, for a given set of model objects and specific data distortion rates. However, it does not explain *why* this is the expected performance. Such an understanding is critical for designing better object recognition systems, as it can provide fundamental answers to questions such as: (a) When does performance break down as a function of the amount of data distortion? (b) What are the

performance limits when using a specific sensor? (c) Is a given feature-selection scheme sufficient for achieving desired levels of performance? (d) What is the largest size of a model set that can be accommodated without significantly degrading performance? Fundamental understanding of the relationship between performance and the factors affecting it is essential for the advancement of the field of object recognition from an art to a science.

2. The performance estimated empirically is dependent upon the actual implementation of the object recognition system. This implementation can be based on recognition approaches such as alignment [11, 12], hypothesis accumulation [13, 14], or tree search [15, 16]. Note that the performance obtained using these approaches can be different, even if they use similar matching criteria. For example, systems that use a vote accumulator (Hough space) will generate different performance estimates depending on the resolution of the accumulator. Another example, alignment-based systems, achieve polynomial-time complexity by using a "looser" notion of data/model feature consistency.
3. Empirical evaluation requires the presence of an actual object recognition system. Obviously, this can considerably slow down the design process.

In this chapter, we address the performance–prediction problem in the context of a typical object recognition task. It can be described as follows. (1) Both model objects and scene data are represented by discretized locations of 2D point features. (2) A data object is assumed to be obtained by applying a 2D transformation to the corresponding model object. Notice that the space of possible 2D transformations is naturally discretized, since we are dealing with discrete 2D point features. (3) The data/model matching quality is estimated using a vote-based criterion. In particular, the quality of a given match hypothesis is estimated by counting the number of consistent data/model feature pairs, which we refer to as "votes."

We present a statistical method for formally predicting lower and upper bounds on the *probability of correct recognition* (PCR) for the task outlined above. The proposed method considers data distortion factors such as uncertainty, occlusion, and clutter, in addition to model similarity. Integrating these data and model factors in a single approach has been a challenging problem. The performance predicted is fundamental in the sense that it is obtained by analyzing the information provided by both the data and model features, *independent* of the particular vote-based matching algorithm. A schematic diagram of the prediction method is shown in Figure 1.2. It can be contrasted with the diagrams of the empirical approaches shown in Figure 1.1. The validity of the proposed method is demonstrated using real synthetic aperture radar (SAR) data from the MSTAR public domain. This data set is obtained under a variety of depression angles and object configurations.

The remainder of this chapter is organized as follows. The next section reviews related research efforts, and highlights our contributions. Section 1.3

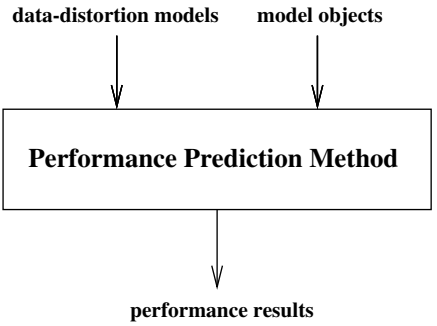


Figure 1.2. Formal estimation of object recognition performance.

presents an overview of the proposed method. Sections 1.4 and 1.5 describe the statistical modeling of the data distortion factors, and the object similarity, respectively. Derivation of lower and upper bounds on PCR is presented in Section 1.6. The validity of those bounds is demonstrated in Section 1.7, by comparing actual PCR plots, as a function of data distortion, with predicted lower and upper bounds. Finally, conclusions and directions for future research are presented in Section 1.8.

1.2 Relevant Research

Several research efforts have addressed the problem of analyzing performance of feature-based object recognition. Most of these efforts focus on the problem of discriminating objects from clutter. We present here a representative sample of these efforts. Grimson and Huttenlocher [17] presented a statistical method for estimating the probability distribution of the fraction of consistent data/model feature pairs for an erroneous hypothesis. They derived such a distribution using a statistical occupancy model (Bose–Einstein model), assuming bounded feature uncertainty and uniform clutter models. This distribution was used to determine the minimum fraction of consistent feature pairs required to achieve a desired probability of false alarm. Sarachik [18] studied the problem of predicting the receiver operating characteristic (ROC) curve for a specific recognition algorithm. The ROC curve described the relationship between the probability of correct recognition and that of a false alarm. The chosen algorithm used a weighted voting criterion based on Gaussian feature uncertainty. A statistical analysis was presented to determine the probability distributions of the weighted votes for both valid and invalid hypotheses, assuming uniform occlusion and clutter models. These distributions were used along with the likelihood-ratio test to predict the ROC curve. Alter and Grimson [8] used statistical knowledge about sources of data distortion to design a recognition criterion, based on the likelihood-ratio test. The likelihoods of observed data-feature set, conditioned on hypothesis validity and

invalidity, were calculated by assuming both bounded and Gaussian feature uncertainty models, in addition to uniform occlusion and clutter models. Lindenbaum [19] extended the modeling of clutter to include background objects of known shape, in addition to uniformly distributed random features. This hybrid model was incorporated into a statistical analysis to predict the number of features needed to guarantee recognition of a data object at a given confidence level. The analysis considered bounded feature uncertainty, as well as structural similarity between a given data object and background ones. Irving et al. [20] derived a theoretical bound on the ROC curve of an object detection task. The generalized likelihood-ratio test was used to discriminate between model objects at various poses and random clutter. The likelihoods of both clutter and model objects were modeled using 2D Poisson processes. Modeling object likelihood as a Poisson process was based on the assumption of independence of object views at discretized poses. This work considered bounded feature uncertainty, and uniform occlusion and clutter models.

The problem of discriminating objects from other model objects has received considerably less attention than object/clutter discrimination. This problem is obviously central to integrated performance prediction of object recognition. It requires consideration of not only data distortion but also object similarity. In addition, it requires consideration of the *interaction* between object similarity and data distortion. Lindenbaum [21] presented a probabilistic analysis for predicting lower and upper bounds on the number of data features required to achieve a certain confidence level in object localization or recognition. It explicitly considered the similarity between different model objects, as well as the self-similarity between a model object and an instance of itself at a different relative pose. The data distortion factors considered were bounded uncertainty and occlusion. The analysis considered extreme cases in modeling the interaction between occlusion and similarity, thus resulting in the generation of relatively loose bounds. We note that the analysis presented in [19], outlined above, can be used in the context of object/object discrimination considering uncertainty and clutter, as well as object similarity. Grenander et al. [22] addressed the problem of predicting fundamental error in object pose estimation. In their work, objects were represented by templates at the pixel level. A minimum mean-square-error estimator, the Hilbert–Schmidt estimator, was used to estimate object pose in the presence of pixel uncertainty. Performance of object/object discrimination was determined partially empirically through the synthesis of distorted templates of one object, and then using the likelihood-ratio test, based on the Hilbert–Schmidt estimator, as a recognition criterion (refer to Figure 1.1(b)).

The methods outlined above are summarized and compared with our method in Table 1.1. This table highlights the main contribution of our work, namely the integration of uncertainty, occlusion, clutter, and similarity factors in a single approach for performance prediction. As shown in the table, previous methods considered only a subset of these factors. It can also be seen that our method is unique among other object/object discrimination methods

Table 1.1. Comparison between performance-analysis methods (U, O, C, and S denote uncertainty, occlusion, clutter, and similarity, respectively).

Work	Discrimination	Data/Model Features	Transform.	Factors			
				U	O	C	S
Grimson et al. [17]	object/clutter	2D/2D lines	rigid	X		X	
Sarachik [18]	object/clutter	2D/2D points	affine	X	X	X	
Alter and Grimson [8]	object/clutter	2D/3D points & lines	weak persp.	X	X	X	
Lindenbaum [19]	object/clutter	2D/2D boundary pts.	affine	X		X	X
Irving et al. [20]	object/clutter	2D/2D points	2D transl.	X	X	X	
Lindenbaum [21]	object/object	2D/2D boundary pts.	rigid	X	X		X
Grenander et al. [22]	object/object	pixel-level templates	rotation	X			
This work	object/object	2D/2D discretized pts.	2D transl.	X	X	X	X

in that it considers point features. Another unique aspect of this work is not just the new theory but also the validation using real data. We note that parts of this work have appeared in [23, 24, 25].

1.3 Overview

In this section, we present an overview of the proposed performance–prediction method. Our problem can be formally defined as follows. We are given the following:

1. A set of model objects, $\mathcal{MD} = \{\mathcal{M}_i\}$, where each object $\mathcal{M}_i$ is represented by discretized locations of 2D point features, $\mathcal{M}_i = \{F_{ik}\}$.
2. Statistical data distortion models.
3. A class of data/model transformations, $\mathcal{T}$.

Our objective is to predict lower and upper bounds on PCR as a function of data distortion. We consider recognition to be successful only if the selected hypothesized object is the actual one, and the difference between the hypothesized pose and the actual one is small. The pose error can be represented by the relative pose of the hypothesized object with respect to the actual one. It is considered acceptable if it lies within a subspace, $\mathcal{T}_{acc} \subset \mathcal{T}$. We assume in this work that $\mathcal{T}_{acc} = \{\mathbf{0}\}$, i.e., only exact object location is acceptable.

A block diagram of the proposed method is shown in Figure 1.3. The main elements in this diagram can be described as follows:

• **Data-Distortion Models:** The data distortion factors are statistically modeled using uniform probability distribution functions (PDFs).

1. *Uncertainty:* The location of the data feature corresponding to a model feature is described by a uniform distribution. Notice that the uncertainty PDFs are discrete, since the feature locations considered in this work are discretized. We further assume that the PDFs associated with different

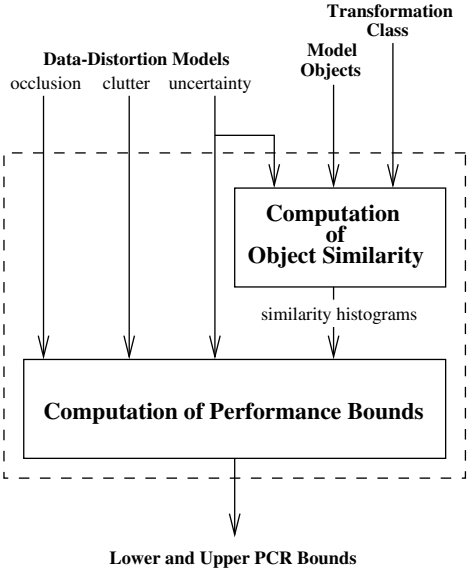


Figure 1.3. Block diagram of performance–prediction method.

model features are independent. We argue that such independence assumption is reasonable in most practical applications.

2. *Occlusion:* We assume that every subset of model features is equally likely to be occluded as any other subset of the same size. This assumption is more appropriate for modeling features that are missing due to inherent instability or imperfections of feature extraction. It is less suitable for modeling features that are missing due to being occluded by other objects, since it does not consider the spatial-correlation aspect among occluded/unoccluded features. Spatial correlation can be captured by using Markov random fields [9, 26, 27], at the expense of significantly increasing the complexity of the analysis. In Section 1.7, we outline a simpler approach that can implicitly consider the spatial-correlation factor, without increasing the analysis complexity.
3. *Clutter:* We assume that clutter features are uniformly distributed within a region surrounding the object. This distribution is useful for modeling random clutter, which does not have specific spatial structure. Modeling "structural" clutter requires analyzing its similarity with model objects. We note that the similarity-modeling concepts presented in this work can be used in modeling of structural clutter. This topic is a subject for future research.

• **Computation of Object Similarity:** The purpose of this stage is to compute the structural similarity information among all pairs of model objects. Our definition of object similarity depends on the amount of feature uncer-

tainty. In particular, the similarity between two model objects is directly proportional to feature uncertainty. This agrees with the intuitive observation that as different objects become more "blurred," it becomes more difficult to differentiate between them, which is, in a sense, equivalent to saying that they become more "similar." The similarity between a model object, $\mathcal{M}_i$, and another one, $\mathcal{M}_j$, is defined as the number of votes for $\mathcal{M}_j$ given an *uncertain* instance of $\mathcal{M}_i$, i.e., an instance of $\mathcal{M}_i$ that is obtained by randomly perturbing its features. Accordingly, the number of votes for $\mathcal{M}_j$ is a random variable. The chosen definition of similarity depends on the relative transformation between $\mathcal{M}_i$ and $\mathcal{M}_j$, defined by transformation class $\mathcal{T}$. Accordingly, the similarity between $\mathcal{M}_i$ and $\mathcal{M}_j$ can be viewed as a probabilistic function, which we call the similarity function. The similarity information is encoded in two histograms, which we call all-similarity and peak-similarity histograms, to be described in Section 1.5.2. These histograms are used for predicting lower and upper PCR bounds, respectively.

• **Computation of Performance Bounds:** The objective of this stage is to compute PCR bounds. The computation is based on estimating the PDF of the votes for a specific erroneous object/pose hypothesis, given a "distorted" instance of a given model object. The estimation process takes into account the structural similarity between the model object and the erroneous hypothesis. The vote PDF is used to determine the probability of a recognition failure, which occurs if the erroneous hypothesis gets same or more votes than the distorted object. This information is integrated for potential erroneous hypotheses to determine the PCR bounds.

1.4 Data-Distortion Models

We formally model the effects of the three distortion factors considered in this work on a "perfect" model object. This modeling is used to determine the vote PDF in Sect. 1.6.2.

• **Uncertainty:** The effect of the uncertainty factor is to perturb locations of model features according to some PDF. Since this PDF is assumed to be uniform, it can be represented by a region. Let $F_{ik} \in \mathcal{M}_i$ be a model feature, and $\widehat{F}_{ik}$ be a distorted instance of it. Define $R_u(F_{ik})$ to be the *consistency region* associated with F_{ik}. Such region bounds the possible locations of $\widehat{F}_{ik}$ given F_{ik}, i.e., $\widehat{F}_{ik} \in R_u(F_{ik})$. Likewise, we can say that $F_{ik} \in \bar{R}_u(\widehat{F}_{ik})$, where $\bar{R}_u(\cdot)$ is the reflection of $R_u(\cdot)$ about the origin. Practically, $R_u(\cdot)$ is the same as $\bar{R}_u(\cdot)$, because $R_u(\cdot)$ is symmetric about the origin (e.g., circle, square). Accordingly, we assume in this work for simplicity that $\bar{R}_u(\cdot) = R_u(\cdot)$. An uncertain instance of $\mathcal{M}_i$ can be obtained by uniformly perturbing each of its features within corresponding consistency region. This can be formally 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 feature selected randomly within region R.

• **Occlusion:** The effect of occlusion is the elimination of some model features. An occluded instance of $\mathcal{M}_i$ can be formally defined as

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

where $\mathcal{P}_o(\mathcal{M}_i, O)$ is a function that returns a subset of O features selected randomly from $\mathcal{M}_i$. For a fixed O, all subsets generated by $\mathcal{D}_o(\mathcal{M}_i, O)$ are equally likely to occur, since we are assuming uniform occlusion.

• **Clutter:** The effect of clutter on a model object is the addition of spurious features to it. They are assumed to be uniformly distributed within a *clutter region*, R_c, surrounding the model object. This region can have an arbitrary shape (e.g., bounding box of model features, convex hull, etc). A cluttered instance of $\mathcal{M}_i$ can be defined as

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

where $\mathcal{P}_c(C, R)$ is a function that returns C features selected randomly within region R, and R_x is a region that clutter features are excluded from falling into. The reason for including R_x is explained below.

• **Combined Distortion:** Consideration of the combined effects of uncertainty, occlusion, and clutter on a model object raises an ambiguous situation. It takes place when a model feature gets occluded and then a spurious feature falls within its consistency region. The ambiguity arises from the fact that this situation can not be distinguished from the no-occlusion/no-clutter case. In order to simplify the analysis, we assume the latter case. This can be modeled by restricting the clutter features to lie outside region R_x, defined as the union of the consistency regions of occluded features. We refer to $R_c - R_x$, or simply R'_c, as the effective clutter region. A distorted instance of $\mathcal{M}_i$, $\widehat{\mathcal{M}}_i(R_u(\cdot), O, C, R_c)$, can be obtained by first occluding O features of $\mathcal{M}_i$, perturbing unoccluded ones within their consistency regions $R_u(\cdot)$, and then randomly adding C clutter features within the effective clutter region R'_c. This can be represented formally as:

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

where $R_x = \cup_k R_u(F_{ik})$, $\forall F_{ik} \in (\mathcal{M}_i - \mathcal{D}_o(\mathcal{M}_i, O))$. Figure 1.4 shows an example of the distortion process.

1.5 Computation of Object Similarity

In this section, we formally define a measure of the structural similarity between model objects, and outline the method used to construct the similarity histograms.

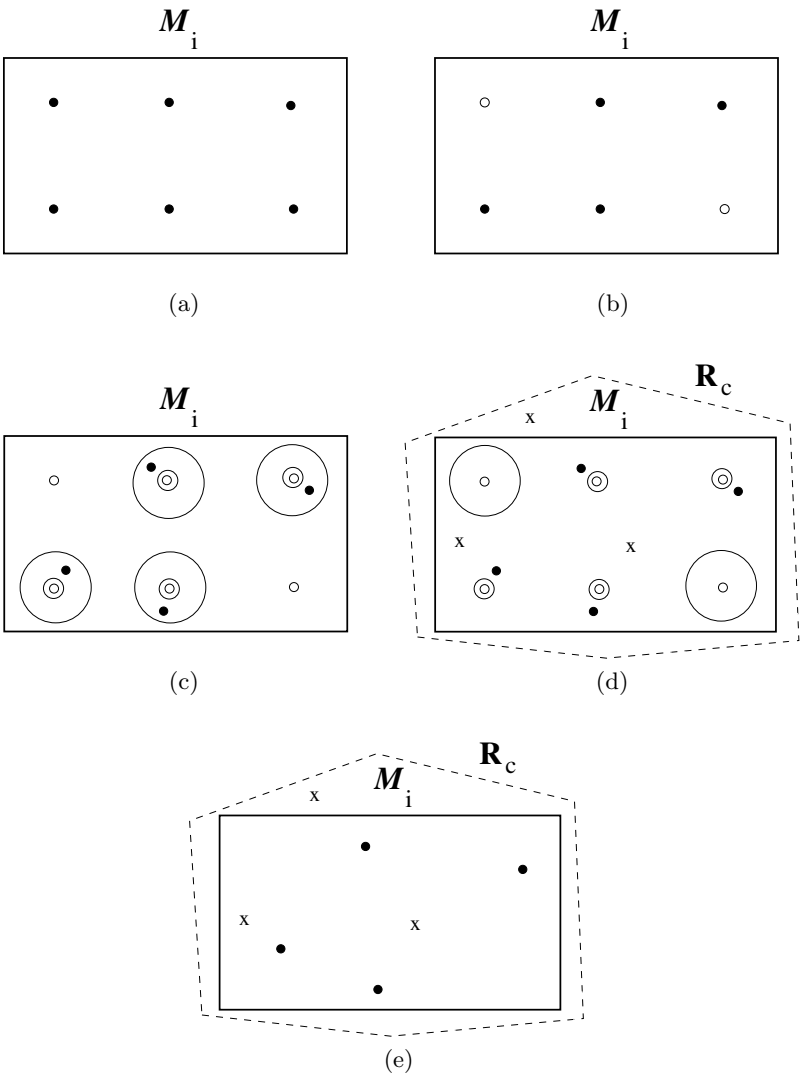


Figure 1.4. An illustration of the different stages of the distortion process: (a) original object consisting of six features (dark circles), (b) after occlusion ($O = 2$; small circles represent occluded features), (c) after perturbation ($R_u(\cdot)$ = a circle; small double circles represent locations of features before perturbation), (d) after clutter ($C = 3$; small crosses represent clutter features; notice absence of clutter features inside consistency regions of the two occluded features), (e) distorted object.

1.5.1 Definition of Object Similarity

We introduce a number of definitions that lead to a definition of the similarity between a pair of model objects.

• **Vote-Based Criterion:** Let $\mathcal{M}_i^{\hat{\tau}} = \{F_{ik}^{\hat{\tau}}\}$ be object $\mathcal{M}_i$ at pose $\hat{\tau} \in \mathcal{T}$ with respect to a data object, $\widehat{\mathcal{M}}$. We refer to $\mathcal{M}_i^{\hat{\tau}}$ as a *hypothesis* of object $\mathcal{M}_i$ at location $\hat{\tau}$. The votes for $\mathcal{M}_i^{\hat{\tau}}$, given $\widehat{\mathcal{M}}$, is the number of features in $\mathcal{M}_i^{\hat{\tau}}$ that are "consistent" with at least a data feature in $\widehat{\mathcal{M}}$. A model feature, $F_{ik}^{\hat{\tau}} \in \mathcal{M}_i^{\hat{\tau}}$, is said to be consistent with a data feature, $\widehat{F}_l \in \widehat{\mathcal{M}}$, if $\widehat{F}_l$ falls within the consistency region of $F_{ik}^{\hat{\tau}}$, i.e., $\widehat{F}_l \in R_u(F_{ik}^{\hat{\tau}})$. Accordingly, we can formally define the votes for $\mathcal{M}_i^{\hat{\tau}}$ given $\widehat{\mathcal{M}}$ as follows:

$$\text{VOTES}(\mathcal{M}_i^{\hat{\tau}}; \widehat{\mathcal{M}}) = |\{F_{ik}^{\hat{\tau}} : F_{ik}^{\hat{\tau}} \in \mathcal{M}_i^{\hat{\tau}} \text{ and } \exists \hat{F}_l \in \widehat{\mathcal{M}} \text{ s.t. } \hat{F}_l \in R_u(F_{ik}^{\hat{\tau}})\}| \quad (1.1)$$

• **Feature/Feature Similarity:** Let us assume that we are given a pair of model features, $F_{ik} \in \mathcal{M}_i$ and $F_{jl}^{\tau_i} \in \mathcal{M}_j^{\tau_i}$, where $\mathcal{M}_j^{\tau_i}$ is a hypothesis of object $\mathcal{M}_j$ at location $\tau_i \in \mathcal{T}$ with respect to object $\mathcal{M}_i$. The similarity between F_{ik} and $F_{jl}^{\tau_i}$, denoted by $S_{ff}(F_{ik}, F_{jl}^{\tau_i})$, is defined as the probability that an uncertain measurement of F_{ik} is consistent with $F_{jl}^{\tau_i}$. Formally, we have

$$S_{ff}(F_{ik}, F_{jl}^{\tau_i}) = \frac{\text{AREA}(R(F_{ik}) \cap R(F_{jl}^{\tau_i}))}{\text{AREA}(R(F_{ik}))},$$

where $\text{AREA}(R)$ is area of region R. Obviously, $S_{ff}(F_{ik}, F_{jl}^{\tau_i})$ lies in the range $[0, 1]$. It is proportional to the extent of overlap between the consistency regions of F_{ik} and $F_{jl}^{\tau_i}$ ($R(F_{ik})$ and $R(F_{jl}^{\tau_i})$). Figure 1.5 illustrates $S_{ff}(F_{ik}, F_{jl}^{\tau_i})$ as a function of τ_i, for a sample of three consistency regions. In some cases, we refer to feature pairs with overlapping/nonoverlapping consistency regions as *similar/dissimilar* feature pairs, respectively.

• **Object/Feature Similarity:** We define the similarity between an object, $\mathcal{M}_i$, and a feature, $F_{jl}^{\tau_i} \in \mathcal{M}_j^{\tau_i}$, as the probability that $F_{jl}^{\tau_i}$ is consistent with an uncertain measurement of *any* feature in $\mathcal{M}_i$. We can formally define object/feature similarity, denoted by $S_{of}(\mathcal{M}_i, F_{jl}^{\tau_i})$, as

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

• **Object/Hypothesis Similarity:** Let us denote the similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ as $S_{oh}(\mathcal{M}_i, \mathcal{M}_j^{\tau_i})$ or simply $S_j^{\tau_i}$. We define $S_j^{\tau_i}$ as the number of votes for hypothesis $\mathcal{M}_j^{\tau_i}$, given an uncertain instance of $\mathcal{M}_i$, which is $\mathcal{D}_u(\mathcal{M}_i, R_u(\cdot))$ (refer to Section 1.4). Formally,

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

It is obvious that $S_j^{\tau_i}$ is a random variable. The minimum value of $S_j^{\tau_i}$ is the number of coincident feature pairs of $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$. It can be expressed as

$$\min(S_j^{\tau_i}) = |\{F_{jk}^{\tau_i} : S_{of}(\mathcal{M}_i, F_{jk}^{\tau_i}) = 1\}|.$$

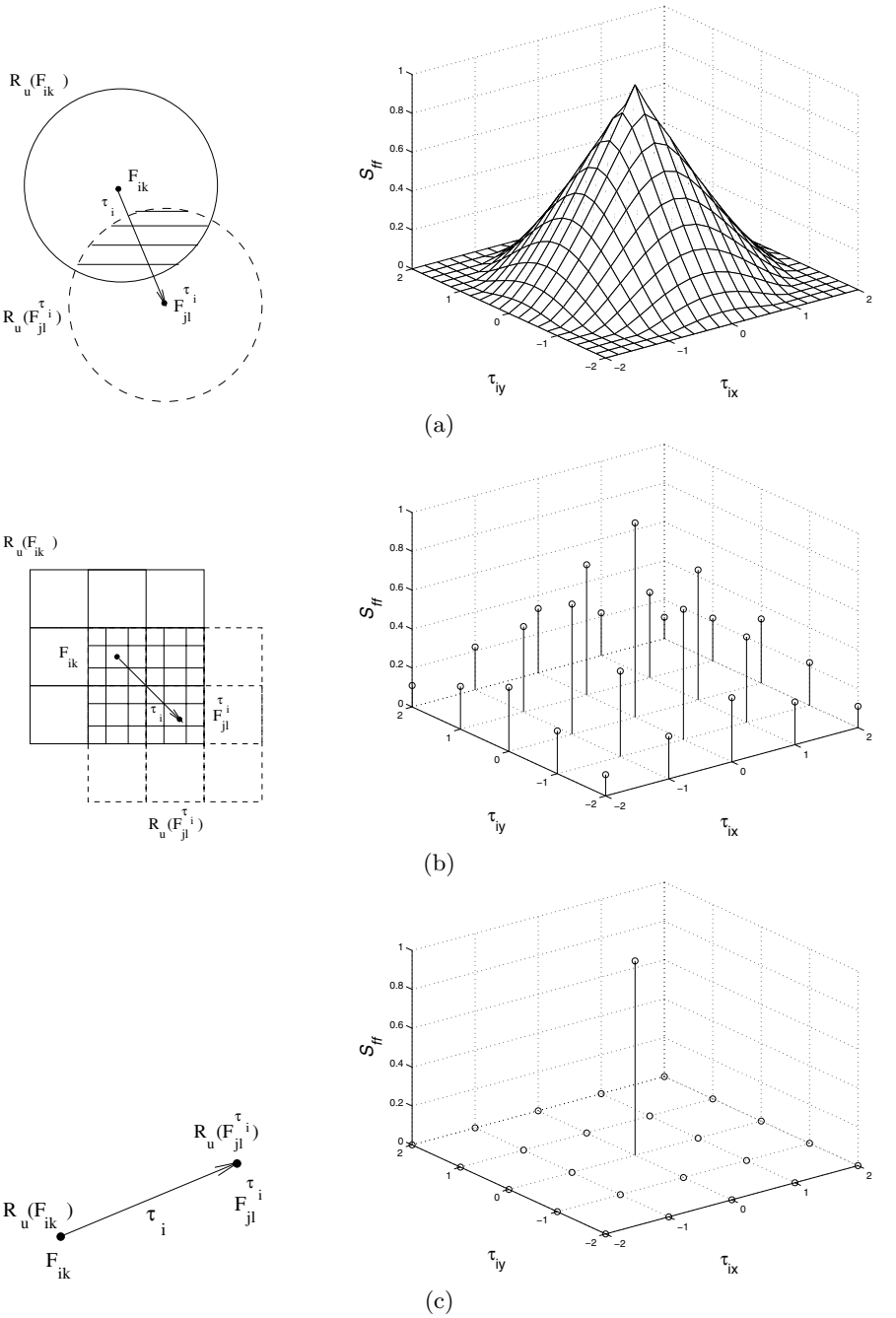


Figure 1.5. : An illustration of feature/feature similarity for a variety of consistency regions assuming $\mathcal{T}$ is the space of 2D translations: (a) a circle of unit radius, (b) a discrete eight-neighbor region, (c) a point region (implies absence of positional uncertainty). Note that the components of τ_i along the x- and y-axes are represented by τ_{ix}, and τ_{iy}, respectively. We assume here for simplicity that $F_{ik} = F_{jl}^{\tau_i}$ when $\tau_i = \mathbf{0}$.

On the other hand, the maximum value of $S_j^{\tau_i}$ is the number of features of $\mathcal{M}_j^{\tau_i}$ that are similar to features in $\mathcal{M}_i$ (i.e., whose consistency regions overlap with consistency regions of features in $\mathcal{M}_i$). Thus,

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

The expected value of $S_j^{\tau_i}$ can be approximated as

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

where $F_{jk}^{\tau_i} \in \mathcal{M}_j^{\tau_i}$. Figure 1.6 shows an example of object/hypothesis similarity.

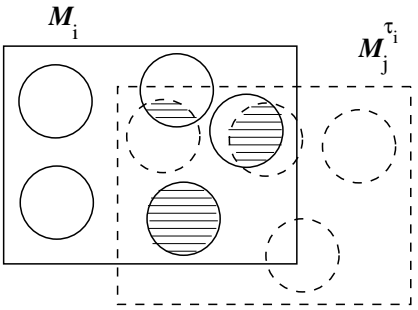


Figure 1.6. An illustration of object/hypothesis similarity. Notice that there are three similar feature pairs with feature/feature similarity values of approximately $\frac{1}{3}$, $\frac{2}{3}$ and 1. Accordingly, we have $S_j^{\tau_i} \in [1, 3]$, and $E(S_j^{\tau_i}) \approx 2$.

• **Uniform Model of Object/Hypothesis Similarity:** In order to make the prediction of PCR bounds mathematically tractable, we make the following reasonable assumptions about $\mathcal{M}_i$, $\mathcal{M}_j^{\tau_i}$ and the structure of their similar feature pairs:

1. The consistency regions of the features that belong to each of $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ are not overlapping.
2. The correspondence between similar features in $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ is bijective (one-to-one).
3. The feature/feature similarity between every pair of similar features is a constant value, $P_j^{\tau_i}$. It is the average object/feature similarity of the features in $\mathcal{M}_j^{\tau_i}$ that are similar to features in $\mathcal{M}_i$.

The above assumptions result in a "uniform" view of the structural similarity between object $\mathcal{M}_i$ and hypothesis $\mathcal{M}_j^{\tau_i}$. As an illustration, Figure 1.7 shows the uniform model corresponding to the object/hypothesis pair shown in Figure 1.6. The uniform similarity model leads to the approximation of 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 $P_X(x) = \Pr[X = x]$, $B_X(x; n, p) = K(n, x)p^x(1-p)^{n-x}$, $K(a, b) = \frac{a!}{(a-b)!\ b!}$,

$$N_j^{\tau_i} = \max(S_j^{\tau_i}), \text{ and}$$
$$P_j^{\tau_i} = \frac{E(S_j^{\tau_i})}{N_j^{\tau_i}}.$$

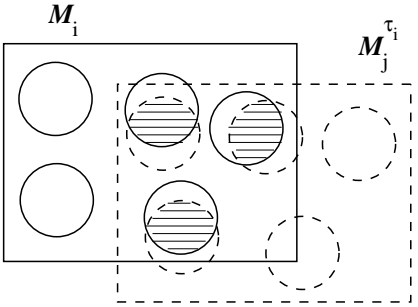


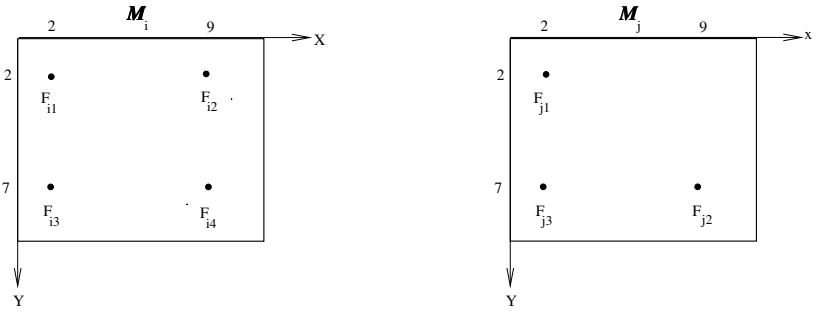
Figure 1.7. Uniform similarity model for object/hypothesis pair shown in Figure 1.6. Notice that similar feature pairs have constant feature/feature similarity, $P_j^{\tau_i} \approx \frac{2}{3}$ and $N_j^{\tau_i} = 3$.

• **Object/Object Similarity:** The similarity between a pair of objects, $\mathcal{M}_i$ and $\mathcal{M}_j$, is defined as the object/hypothesis similarity $S_j^{\tau_i}$, for all $\tau_i \in \mathcal{T}$. Thus, object/object similarity can be viewed as a probabilistic function. As an illustration, Figure 1.8(a) shows a pair of simple model objects. The corresponding expected-similarity function, $E(S_j^{\tau_i})$, is shown in Figure 1.8(b). Note that peaks in the expected-similarity function correspond to object hypotheses that have a higher degree of similarity with $\mathcal{M}_i$ than neighboring ones. A sample of these hypotheses, referred to as *peak hypotheses*, is shown in Figure 1.9. In our work, peak hypotheses are used for predicting an upper bound on PCR.

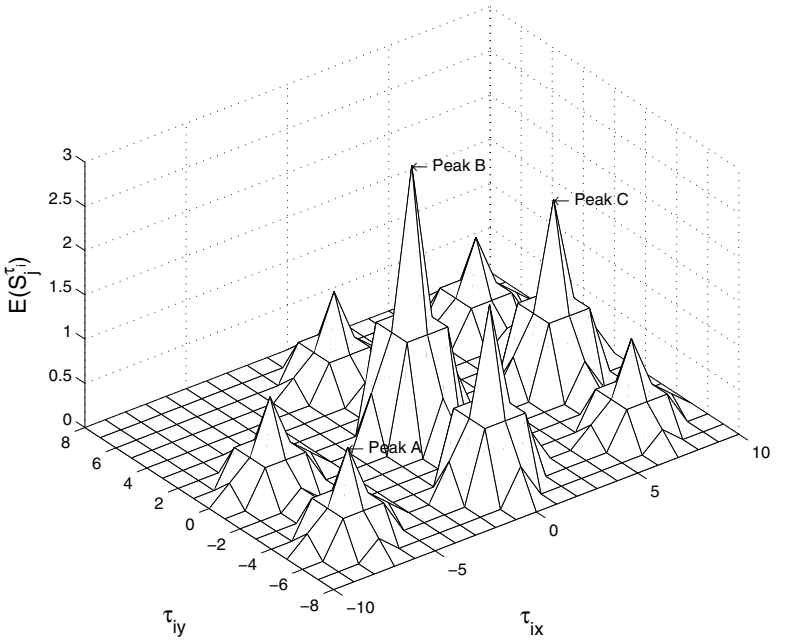
1.5.2 Construction of Similarity Histograms

As discussed in the previous section, we describe the object/hypothesis similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ by two parameters, $(N_j^{\tau_i}, P_j^{\tau_i})$. For our purpose of performance prediction, we add two more parameters:

1. The size of $\mathcal{M}_i$, $|\mathcal{M}_i|$.
2. The *effective size* of $\mathcal{M}_j^{\tau_i}$, $|\mathcal{M}_j^{\tau_i} \cap R_c|$, which is the number of features of $\mathcal{M}_j^{\tau_i}$ that lie inside the clutter region R_c.



(a)



(b)

Figure 1.8. An illustration of object/object similarity: (a) model objects $\mathcal{M}_i$ and $\mathcal{M}_j$, (b) corresponding expected-similarity function, $E(S_j^{\tau_i})$, assuming four-neighbor consistency region, and 2D translation space.

Thus, we encode the information of object/hypothesis similarity using tuple $(\mid \mathcal{M}_i \mid, \mid \mathcal{M}_j^{\tau_i} \cap R_c \mid, N_j^{\tau_i}, N_j^{\tau_i} P_j^{\tau_i})$. Accordingly, the similarity information is accumulated in 4D histograms [3].

Two similarity histograms are needed in our work, one for storing similarity information corresponding to all erroneous hypotheses, and the other

[3] When calculating the effective size of $\mathcal{M}_j^{\tau_i}$, we have also included features of $\mathcal{M}_j^{\tau_i}$ that lie outside R_c but are similar to features in $\mathcal{M}_i$.

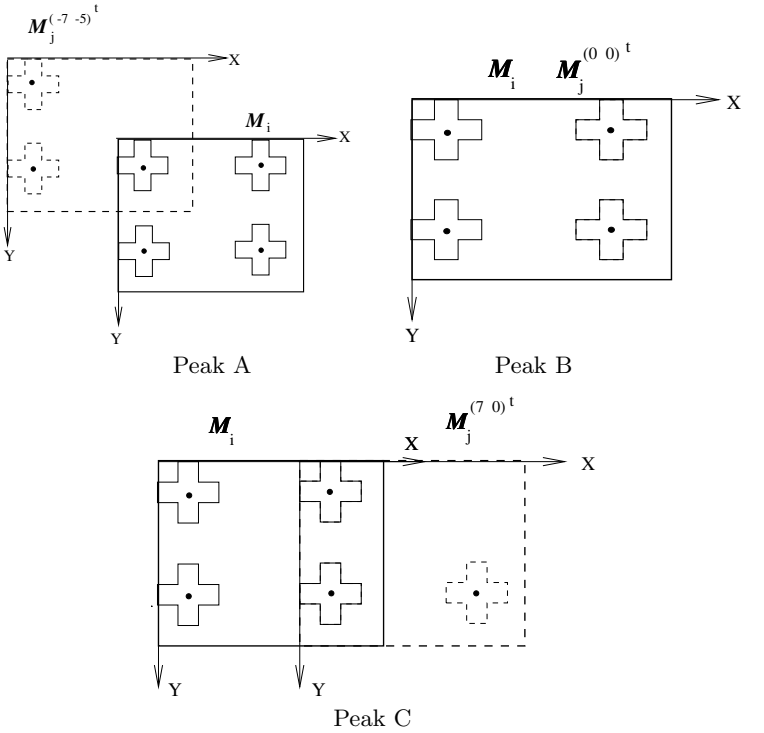


Figure 1.9. Three hypotheses corresponding to peaks A, B and C shown in Fig. 1.8(b), assuming four-neighbor consistency region.

for storing information corresponding to peak hypotheses only. They are referred to as all- and peak-similarity histograms, respectively. The algorithm used to construct these histograms is outlined in Figure 1.10. It calculates the similarity between every model object $\mathcal{M}_i$, and all the erroneous hypotheses competing with it. The erroneous hypotheses are selected to satisfy the following two criteria:

1. Each has at least one feature inside clutter region R_c.
2. For hypotheses that belong to $\mathcal{M}_i$, the relative pose, τ_i, lies outside $\mathcal{T}_{acc}$, defined in Section 1.3.

The similarity information associated with $\mathcal{M}_i$ is accumulated in local all- and peak-similarity histograms, ASH_i and PSH_i, respectively. These histograms, for all $\mathcal{M}_i \in \mathcal{MD}$, are then added to form global similarity histograms, ASH and PSH, respectively.

Initialize global similarity histograms ASH and PSH
for each model object $\mathcal{M}_i \in \mathcal{MD}$ **do**
  Initialize local similarity histograms for $\mathcal{M}_i$, ASH_i and PSH_i
  for each model object $\mathcal{M}_j \in \mathcal{MD}$ **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 $ASH_i(|\mathcal{M}_i|, |\mathcal{M}_j^{\tau_i} \cap R_c|, N_j^{\tau_i}, \lfloor N_j^{\tau_i} P_j^{\tau_i} + \frac{1}{2} \rfloor)$ by 1
        if $\mathcal{M}_j^{\tau_i}$ is a peak hypothesis **then**
          Increment $PSH_i(|\mathcal{M}_i|, |\mathcal{M}_j^{\tau_i} \cap R_c|, N_j^{\tau_i}, \lfloor N_j^{\tau_i} P_j^{\tau_i} + \frac{1}{2} \rfloor)$ by 1
        end if
      end if
    end for
  end for
  Add ASH_i to ASH
  Add PSH_i to PSH
end for

Figure 1.10. Similarity-computation algorithm.

1.6 Computation of Performance Bounds

In this section, we derive the PDF of votes for an erroneous hypothesis, and use this PDF for predicting lower and upper bounds on PCR.

1.6.1 Motivating Example

We start by presenting an example to illustrate the combined effects of data distortion and object similarity on the vote process. This example, illustrated in Figure 1.11, assumes the uniform model of similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$, which is defined in Section 1.5.1. It can be described as follows:

- Prior to being distorted, $\mathcal{M}_i$ has five votes, since it consists of five features. On the other hand, $\mathcal{M}_j^{\tau_i}$ does not have any features of $\mathcal{M}_i$ within the consistency regions of its features. Accordingly, it gets no votes.
- The first distortion step involves occlusion of two features in $\mathcal{M}_i$. Obviously, this reduces the number of votes for $\mathcal{M}_i$ from five to three. At this point, $\mathcal{M}_j^{\tau_i}$ still does not get any votes. Notice that the number of similar feature pairs between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$ decreases from three (which is $N_j^{\tau_i}$; refer to Section 1.5.1) to two.
- The second step involves randomly perturbing the three unoccluded features in $\mathcal{M}_i$ within their consistency regions. This keeps the number of votes for $\mathcal{M}_i$ at three. On the other hand, observe that both of the two unoccluded similar features of $\mathcal{M}_i$ move to the regions that overlap with the consistency regions of their corresponding similar features in $\mathcal{M}_j^{\tau_i}$. This contributes two votes to $\mathcal{M}_j^{\tau_i}$.

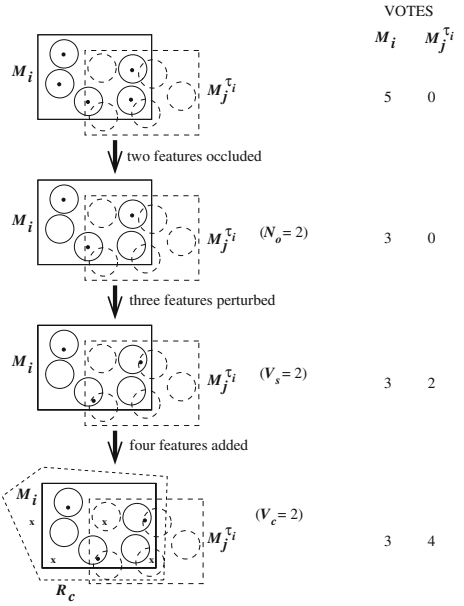


Figure 1.11. An example showing the vote process for object $\mathcal{M}_i$ and erroneous hypothesis $\mathcal{M}_j^{\tau_i}$, as $\mathcal{M}_i$ gets distorted.

- In the final distortion step, four clutter features are randomly added within clutter region R_c. Two of these features happen to fall within the consistency regions of two new features of $\mathcal{M}_j^{\tau_i}$. This contributes two extra votes for $\mathcal{M}_j^{\tau_i}$, thus bringing its total number of votes to four. The number of votes for $\mathcal{M}_i$ stays the same (recall from Section 1.4 that clutter features are excluded from falling into consistency regions of occluded ones).

The above example shows how data distortion and model similarity can result in a recognition failure by reducing the number of votes for the correct hypothesis, and increasing them for an erroneous one. It also provides us with the following valuable insight into the distribution of votes for both correct and incorrect hypotheses, as a function of data distortion:

- The number of votes for $\mathcal{M}_i$, denoted by V_i, is simply the number of unoccluded features of $\mathcal{M}_i$. That is, for 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

$$V_i = \mid \mathcal{M}_i \mid -O. \tag{1.2}$$

- The number of votes for $\mathcal{M}_j^{\tau_i}$, denoted by $V_j^{\tau_i}$, comes from two different sources: (1) object $\mathcal{M}_i$, due to structural similarity (second distortion step in Figure 1.11), and (2) clutter features, due to random coincidence (third distortion step in Figure 1.11). Thus, $V_j^{\tau_i}$ is a random variable that can be expressed as follows:

$$V_j^{\tau_i} = V_s + V_c, \tag{1.3}$$

  where V_s and V_c are random variables that represent *similarity* and *clutter* votes for $\mathcal{M}_j^{\tau_i}$, respectively.
- The number of similarity votes, V_s, is bounded by the number of similar feature pairs that remain unoccluded, which we denote by N_o (obviously, $N_o \leq N_j^{\tau_i}$).

In the above example, it can be seen that $V_s = 2$, $V_c = 2$, and $N_o = 2$. In the next section, we use these three random variables to determine the PDF of $V_j^{\tau_i}$.

1.6.2 Probability Distribution of Hypothesis Votes

In order to determine the PDF of $V_j^{\tau_i}$, we need to determine the PDFs of V_s (number of similarity votes), V_c (number of clutter votes), and N_o (number of unoccluded similar features). In the previous section, we have seen that V_s depends on N_o. Accordingly, we can express the PDF of V_s as

$$P_{V_s}(v_s) = \sum_{n_o} P_{V_s}(v_s; n_o) P_{N_o}(n_o), \tag{1.4}$$

where $P_{V_s}(v_s; n_o) = \Pr[V_s = v_s; N_o = n_o]$. From (1.3) and (1.4), we can represent the PDF of $V_j^{\tau_i}$ as

$$P_{V_j^{\tau_i}}(v_j^{\tau_i}) = \sum_{n_o} P_{V_j^{\tau_i}}(v_j^{\tau_i}; n_o) P_{N_o}(n_o), \tag{1.5}$$

where

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

and $P_{V_c}(v_c; n_o, v_s) = \Pr[V_c = v_c; N_o = n_o, V_s = v_s]$. We estimate the PDF of N_o and the conditional PDFs of V_s and V_c based on the uniform models of data distortion and structural similarity, presented in Sections 1.4 and 1.5.1, respectively.

• **PDF of N_o:** The process of occluding O features in $\mathcal{M}_i$ can be viewed as picking O balls from an urn, which contains $N_j^{\tau_i}$ white balls and ($\mid \mathcal{M}_i \mid - N_j^{\tau_i}$) black balls, with no replacement. In our case, the white (black) balls represent features in $\mathcal{M}_i$ that are similar (dissimilar) to features in $\mathcal{M}_j^{\tau_i}$. Based on the

uniform occlusion and similarity models, the PDF of N_o can be described by the following hypergeometric distribution,

$$P_{N_o}(n_o) = H_{N_o}(N_j^{\tau_i} - n_o; O, N_j^{\tau_i}, \mid \mathcal{M}_i \mid -N_j^{\tau_i}), \tag{1.6}$$

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:** It can be easily shown that the conditional PDF of V_s is represented by the following distribution!binomial binomial distribution:

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

This distribution is obtained based on the assumptions of uniform uncertainty and similarity models. Notice that $P_j^{\tau_i} < 1$ implies $v_s \in [0, n_o]$, while $P_j^{\tau_i} = 1$ implies $v_s = n_o$.

• **Conditional PDF of V_c:** The estimation of the PDF of V_c is considerably more involved than those of N_o and V_s. It can be outlined as follows. Let $R'_{V_c} \subset R'_c$ be the largest region such that a clutter feature falling within it will contribute a vote for $\mathcal{M}_j^{\tau_i}$. Region R'_{V_c} is the union of the consistency regions of features in $\mathcal{M}_j^{\tau_i} \cap R_c$ that do not have any features of $\mathcal{M}_i$ within their consistency regions. They are basically all the features of $\mathcal{M}_j^{\tau_i} \cap R_c$ minus those that have similar features of $\mathcal{M}_i$ within their consistency regions. A slight complexity arises from our clutter modeling explained in Section 1.4: features in $\mathcal{M}_j^{\tau_i} \cap R_c$ that are similar to *occluded* features in $\mathcal{M}_i$ are effectively associated with "truncated" consistency regions. Figure 1.12 shows an example of R'_{V_c} . Based on the assumption of uniform similarity, we can show the following:

1. The area of a truncated consistency region is $\text{AREA}(R_u(\cdot))(1 - P_j^{\tau_i})$.
2. The numbers of potential vote-contributing features with truncated and full consistency 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.

Splitting the effective clutter region R'_c into two subregions, R'_{V_c} and $R'_c - R'_{V_c}$, we can approximate the conditional PDF of V_c by the following binomial distribution,

$$P_{V_c}(v_c; n_o, v_s) \approx B_{V_c}\left(v_c; C, \frac{\text{AREA}(R'_{V_c})}{\text{AREA}(R'_c)}\right), \tag{1.7}$$

where

$$\text{AREA}(R'_{V_c}) = \text{AREA}(R_u(\cdot))(n_f + (1 - P_j^{\tau_i})n_t), \text{ and}$$
$$\text{AREA}(R'_c) = \text{AREA}(R_c) - O \times \text{AREA}(R_u(\cdot)).$$

The lower bound of v_c is 0, while the upper bound is either $\min(n_f + n_t, C)$ if $P_j^{\tau_i} < 1$, or $\min(n_f, C)$ if $P_j^{\tau_i} = 1$[4].

[4] The area of R'_c is calculated by assuming, for simplicity, that clutter region R_c totally covers the consistency regions of the features of $\mathcal{M}_i$.

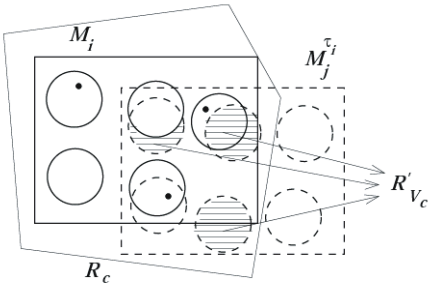


Figure 1.12. An illustration of the clutter vote region R'_{V_c}, assuming the uniform model of similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$.

1.6.3 Lower Bound on PCR

Let $\mathcal{H}_i$ be the set of erroneous object/pose hypotheses corresponding to $\mathcal{M}_i$. It can be defined as

$$\mathcal{H}_i=\{\mathcal{M}_j^{\tau_i} : \mathcal{M}_j \in \mathcal{MD} \text{ and } \tau_i \in \mathcal{T} \text{ s.t. } |\mathcal{M}_j^{\tau_i} \cap R_c| > 0\} - \{\mathcal{M}_i^{\tau_i} : \tau_i \in \mathcal{T}_{acc}\}.$$

We can express the probability of misinterpreting a distorted instance of $\mathcal{M}_i$, $\widehat{\mathcal{M}}_i$, as any hypothesis in $\mathcal{H}_i$, as

$$\Pr[\mathcal{H}_i; \widehat{\mathcal{M}}_i] = \Pr[\exists\ \mathcal{M}_j^{\tau_i} \in \mathcal{H}_i \text{ s.t. } V_j^{\tau_i} \geq V_i]. \tag{1.8}$$

The probability that $\mathcal{M}_j^{\tau_i}$ "beats" $\mathcal{M}_i$ (i.e., $\mathcal{M}_j^{\tau_i}$ reaches or exceeds votes for $\mathcal{M}_i$) can be obtained from (1.2) and (1.5):

$$\Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}}_i] = \sum_{v_j^{\tau_i} \geq |\mathcal{M}_i| - O} P_{V_j^{\tau_i}}(v_j^{\tau_i}). \tag{1.9}$$

From (1.8) and (1.9), we obtain the following upper bound on the probability of recognition failure:

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

The above inequality directly leads to the following lower bound on PCR:

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] > 1 - \sum_{\mathcal{M}_j^{\tau_i} \in \mathcal{H}_i} \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}}_i]. \tag{1.10}$$

From the derivation of the vote PDF discussed in the previous section, we can observe 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: size of $\mathcal{M}_i$, effective size of $\mathcal{M}_j^{\tau_i}$, and the two similarity parameters $(N_j^{\tau_i}, P_j^{\tau_i})$. Define

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

such that $a = \mid \mathcal{M}_i \mid$, $b = \mid \mathcal{M}_j^{\tau_i} \cap R_c \mid$, $c = N_j^{\tau_i}$ and $d = \lfloor N_j^{\tau_i} P_j^{\tau_i} + \frac{1}{2} \rfloor$. Equation (1.10) can be rewritten as

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] > 1 - \sum_a \sum_b \sum_c \sum_d ASH_i(a,b,c,d) W(a,b,c,d). \quad (1.11)$$

Taking the average of (1.11) over all model objects in the model set $\mathcal{MD}$, we obtain the following lower bound on average PCR for $\mathcal{MD}$:

$$\mathrm{PCR}(\mathcal{MD}) > 1 - \frac{1}{|\mathcal{MD}|} \sum_a \sum_b \sum_c \sum_d ASH(a,b,c,d) W(a,b,c,d), \quad (1.12)$$

where

$$ASH(a,b,c,d) = \sum_i ASH_i(a,b,c,d).$$

1.6.4 Upper Bound on PCR

In this section, we present three possible upper bounds on PCR. These bounds differ from each other in the degree of their tightness, and their reliance on assumptions.

The first upper bound can be obtained by observing that recognition fails if *any* hypothesis in $\mathcal{H}_i$ beats $\mathcal{M}_i$. The probability that this event takes place for a given hypothesis, $\mathcal{M}_j^{\tau_i}$, is $\Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}}_i]$, which is defined in (1.9). The maximum of these probabilities among all hypotheses in $\mathcal{H}_i$ forms a lower bound on the probability of recognition failure. This directly leads us to the following upper bound on PCR:

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] < 1 - \max_{\tau_i, j} \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}}_i], \quad (1.13)$$

where $\mathcal{M}_j^{\tau_i} \in \mathcal{H}_i$. Obviously, we do not expect this bound to be tight, since it involves only a *single* erroneous hypothesis.

One possible approach for obtaining a bound that is tighter than the one in (1.13) is to consider a subset of the hypotheses in $\mathcal{H}_i$, and make the assumption that the vote PDF's for these hypotheses are independent. Obviously, the vote-independence assumption is not reasonable among adjacent hypotheses due to the overlap of their consistency regions. We propose to consider hypotheses that correspond to *peaks* in the expected similarity function (refer to Section 1.5.1). The rationale behind such choice can be outlined as follows:

- Peak hypotheses tend to occur at random locations, which makes the vote-independence assumption among them more reasonable than when applied to adjacent ones.

- Peak hypotheses have a higher degree of similarity with given model object $\mathcal{M}_i$, than neighboring ones. Accordingly, we can say that a distorted instance of $\mathcal{M}_i$ is more likely to be misinterpreted as a peak hypotheses than an off-peak neighbor.

Let $\mathcal{H}_{pi}$ be the set of peak hypotheses associated with model object $\mathcal{M}_i$. That is,

$$\mathcal{H}_{pi} = \{\mathcal{M}_j^{\tau_i} : \mathcal{M}_j^{\tau_i} \in \mathcal{H}_i, \text{ and } \tau_i \text{ is a peak in } E(S_j^{\tau_i})\}.$$

Based on the vote-independence assumption, we can obtain the following upper bound on PCR:

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] < \min_j \prod_{\tau_i} (1 - \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}}_i]), \tag{1.14}$$

where $\mathcal{M}_j^{\tau_i} \in \mathcal{H}_{pi}$. This upper bound is tighter than the one defined in (1.13), since it considers a representative subset of model hypotheses that belong to a single object, instead of just a single hypothesis as in (1.13).

In order to obtain a bound that is tighter than the one in (1.14), we need to consider hypotheses in $\mathcal{H}_i$ that belong to *all* model objects, not just a single object as in (1.14). One possible way of achieving this goal is to make the assumption that the vote PDFs for peak hypotheses that belong to *different* model objects are independent. This assumption leads us to the following upper bound,

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] < \prod_{j,\tau_i} (1 - \Pr[\mathcal{M}_j^{\tau_i}; \widehat{\mathcal{M}}_i]), \tag{1.15}$$

where $\mathcal{M}_j^{\tau_i} \in \mathcal{H}_{pi}$. We note that the above assumption can be invalid in some extreme cases, such as when two of the model objects are identical or very similar. However, we argue that it is reasonable in many practical scenarios. For example, let us consider the domain of object recognition using SAR data, which is the main motivation behind our efforts (see the experimental results in Section 1.7). It is well known that the features extracted from a SAR image are extremely sensitive to the geometry of the object shape. Accordingly, even in the presence of similar objects, we can expect the corresponding feature sets to be considerably different [10].

In our experiments, we use the tightest bound defined in (1.15). This bound can be rewritten as

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] < \prod_a \prod_b \prod_c \prod_d (1 - W(a,b,c,d))^{PSH_i(a,b,c,d)}. \tag{1.16}$$

Taking the geometric mean of (1.16) for all model objects in $\mathcal{MD}$, we obtain an estimate of the upper bound of PCR corresponding to $\mathcal{MD}$,

$$\mathrm{PCR}(\mathcal{MD}) < \prod_a \prod_b \prod_c \prod_d (1 - W(a,b,c,d))^{\frac{PSH(a,b,c,d)}{|\mathcal{MD}|}}, \tag{1.17}$$

where

$$PSH(a,b,c,d) = \sum_i PSH_i(a,b,c,d).$$

1.7 Experimental Validation

In this section, we validate the proposed prediction method by comparing predicted bounds on performance with actual ones determined empirically.

1.7.1 Recognition Task

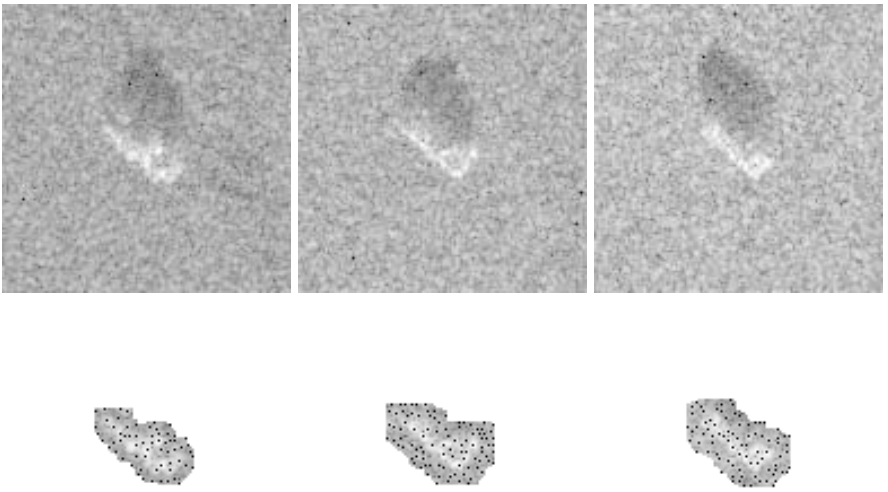
The recognition task that we consider in this work involves recognition of objects using SAR data [10]. We use real data from the MSTAR public domain [28]. The model set consists of a number of objects, which are military targets. Each object is represented by a number of SAR views that sample its signature at several azimuth angles and a specific depression angle. We consider each view as an independent 2D "object." Each model or data view is represented by locations of scattering centers, which are image peaks. These scattering centers correspond to eight-neighbor peaks. The features of each object are the 30 peaks that are strongest in magnitude of radar returns. Recognition involves comparing peaks extracted from data and model views assuming that the space of applicable transformations, $\mathcal{T}$, is discrete 2D translations in the image plane [5, 10].

1.7.2 Model and Test Sets

Our model set consists of three objects, which are T72, BMP2, and BTR70. The numbers of views for these objects are 231, 233, and 233, respectively, and so the total number of views is 697. All of these views are obtained at depression angle 17°. Figure 1.13 shows a representative sample of a single view for each object, along with associated scattering centers.

The test data consists of seven sets, $\mathcal{TD}_1$ through $\mathcal{TD}_7$. They are classified into two groups depending on whether the data distortion in the set is synthetic or real.

1. **Synthetic-Distortion Group:** This group consists of sets $\mathcal{TD}_1$ through $\mathcal{TD}_5$. It is obtained by synthetically distorting the model set, following the distortion process outlined in Section 1.4. We have added the constraint that no features are eight-neighbors, in order to simulate the process of peak extraction. The distortion in each set is characterized by specific consistency and clutter regions, and a number of occlusion/clutter (O/C) values. Notice that the numbers of occluded and clutter features are always the same in our case, since we are considering a fixed number of features



(a) T72 (b) BMP2 (c) BTR70

Figure 1.13. Examples of SAR images and associated scattering centers at depression angle 17°, and azimuth angle 132°.

for both model and data views. The test subset corresponding to a specific O/C value consists of four randomly generated distorted instances for each model view. Accordingly, the size of each test subset is $4 \mid \mathcal{MD} \mid$, and the total size of the test set is $4n_{oc} \mid \mathcal{MD} \mid$, where n_{oc} is the number of O/C values considered in the set. The upper section of Table 1.2 describes the five synthetically distorted test sets. Note that the fourth column in this table consists of two subcolumns: the first one describes the basic shape of the clutter region, while the second describes the scale of that shape. In particular, the clutter region is obtained by scaling the basic clutter region by some factor. Furthermore, notice that the consistency region $R_u(\cdot)$ for test set $\mathcal{TD}_1$ is the zero-neighbor region, which implies the absence of any positional uncertainty.

2. **Real-Distortion Group:** This group consists of two sets, $\mathcal{TD}_6$ and $\mathcal{TD}_7$. These sets correspond to variants of the model set involving real distortion. The first set, $\mathcal{TD}_6$, is obtained by changing the configurations of the model objects [5]. Examples of changing the object configuration are using different flash lights, changing numbers of barrels, etc. The second set, $\mathcal{TD}_7$, is obtained by changing the depression angle from 17° to 15°. Note that due to the nature of the SAR imaging process, such a small angle change can result in a significant variation in the object signature [28]. For both sets, the distortion parameters are determined as follows. The

Table 1.2. Description of the test sets used in the experiments.

Set	Distortion	Consistency Region	Clutter Region Shape	Clutter Region Factor	Occluded/ Clutter Features	Size
$\mathcal{TD}_1$	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 config.	4-neighbor	convex hull	1	estimated	464
$\mathcal{TD}_7$	Δ dep. angle	4-neighbor	convex hull	1	estimated	581

consistency and clutter regions, $R_u(\cdot)$ and R_c, are empirically chosen to be four-neighbor region and convex hull of view features, respectively. The O/C value is estimated for each test view through finding the best matching model view within a difference of $\pm 3^\circ$ azimuth angles, and counting the number of unmatched features. In the case of the absence of a model view within $\pm 3^\circ$ azimuth angles, the test view is matched with the available model one that is closest in azimuth. The lower section of Table 1.2 summarizes the two real sets.

Note that our models are defined from real data that may in fact contain some distortions. Thus, our notion of occlusion and clutter is relative, not absolute. That is, a spurious feature in the model view that does not match any feature in the test view is considered as a true model feature that is occluded. Furthermore, a true feature in the test view that is missing in the model view is considered as a clutter feature. Obviously, recognition performance would be improved through learning the true model features, or by giving a variety of weights to the model features depending on the probability that they correspond to true features. This topic is beyond the scope of this work.

1.7.3 Results

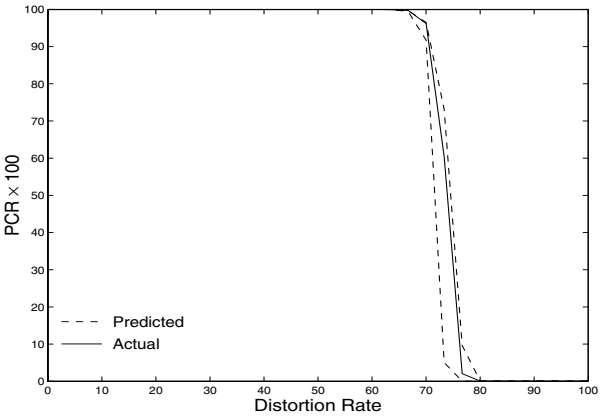
Our selected performance measure is PCR as a function of data distortion, in particular, the occlusion/clutter rate (O/C). As mentioned, the consistency and clutter regions are assumed to be fixed for each test set. The empirical performance is determined using an object recognition system, which uses the vote-based criterion defined in (1.1). The recognition system examines *all* the relevant translations between a given test view, and each model one. The translations examined are defined by the bounding box on the translations that lead to at least a single match between a test feature and a model one. Accordingly, the performance determined by the recognition system is *optimal*, assuming the given vote-based criterion. The predicted bounds are obtained as described in Section 1.6 with two minor modifications. The first one involves the clutter region. Notice that in our experiments, the exact shape of the

clutter region is not fixed, but differs from one view to another. Accordingly, AREA(R_c) is replaced by the average of clutter-region areas for all model views. Recall that AREA(R_c) is used when calculating the vote PDF of an erroneous hypothesis (refer to Section 1.6.2). The second modification involves the conditional PDF of clutter votes, defined in (1.7). Notice that in our recognition task, image peaks cannot be eight-neighbors. This fact needs to be considered in the estimation of the clutter vote PDF, in order to obtain more accurate performance bounds. An approximate method for estimating such a PDF is presented in the Appendix.

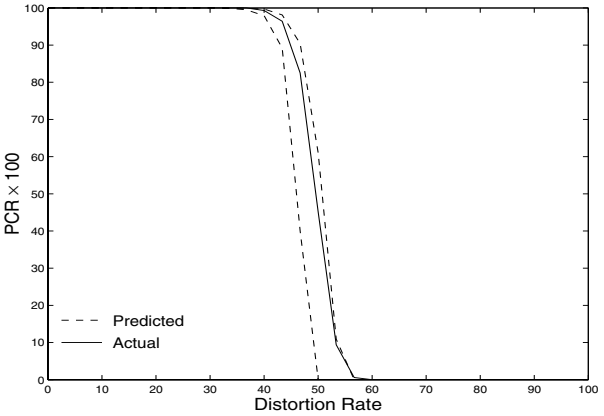
We first analyze the results involving the synthetic-distortion test group. Figures 1.14 and 1.15 show actual and predicted performance plots for the test sets $\mathcal{TD}_1$ through $\mathcal{TD}_5$. From these figures, we observe the following:

- In all cases, our method consistently succeeds in predicting reasonably tight bounds on actual performance.
- The predicted lower bounds consistently predict the actual breakpoint in performance with high accuracy. In all cases, the breakpoint in the lower-bound plot either coincides with the actual one, or occurs very slightly before it.
- The prediction method confirms the intuitive observation that performance degrades when the size of the consistency region increases. This can be observed by comparing Figures 1.14(a)–(c). The method also confirms the intuitive observation that performance improves when the density of clutter features decreases, which takes place when the size of the clutter region increases. Again, this can be observed by comparing Figures 1.14(b), 1.15(a), and 1.15(b).
- The predicted lower bound is consistently very close to the actual PCR plot along the knee section, defined as the section between the first two breakpoints in the plot. In all cases, the difference between the lower bound and corresponding PCR in the knee section is less than 3%. Beyond the knee section, the lower bound diverges. This observation can be explained as follows. Let $\mathcal{H}_i(D) \subset \mathcal{H}_i$ be set of erroneous hypotheses that simultaneously beat true hypothesis at distortion level D, where $D = O = C$. At relatively low levels of distortion, the probability that more than one erroneous hypotheses will *simultaneously* beat the true one is negligible compared to that of only a single erroneous hypothesis [5]. Accordingly, in case of recognition failure at low distortion levels, $\mathcal{H}_i(D)$ has only a single element in most cases. This makes the lower bound defined in (1.10) an estimate of the actual performance. However, as distortion increases, the above assertion ceases to be valid, which makes the lower bound in (1.10) a strict one.
- The predicted upper bound becomes less tight beyond the breakpoint of performance, but then becomes very tight towards the end of the plot. This can be explained as follows. It is easy to see that the size of $\mathcal{H}_i(D)$ is

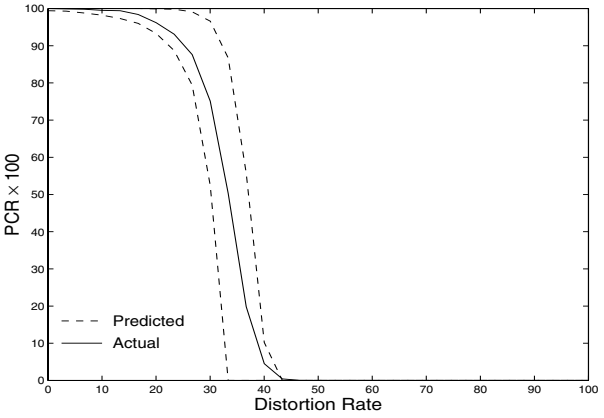
[5] This is assuming we do not have identical or very similar model objects.



(a)



(b)



(c)

Figure 1.14. Actual and predicted PCR plots for synthetic-distortion sets involving different consistency regions: (a) $\mathcal{TD}_1$, (b) $\mathcal{TD}_2$, (c) $\mathcal{TD}_3$.

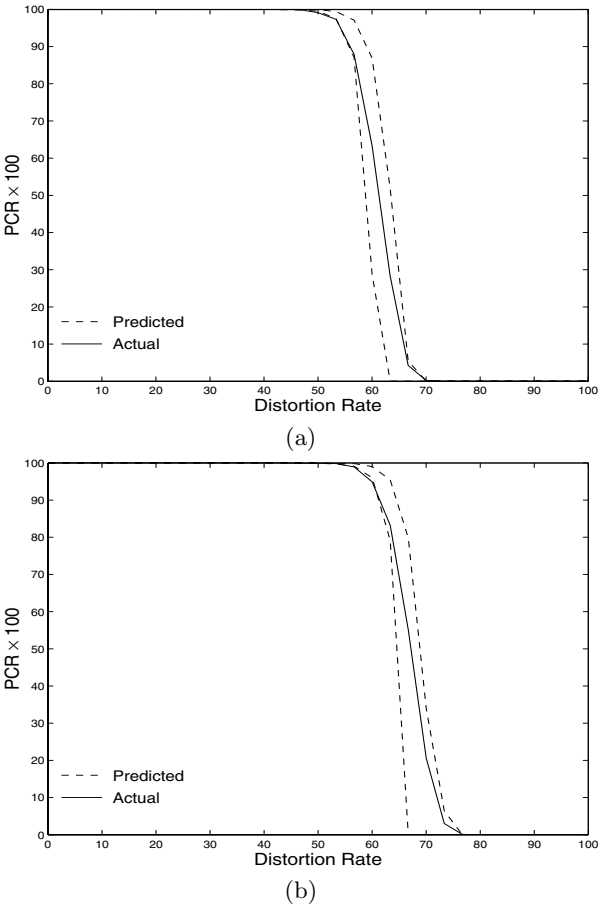


Figure 1.15. Actual and predicted PCR plots for synthetic-distortion sets involving different clutter regions: (a) $\mathcal{TD}_4$, (b) $\mathcal{TD}_5$.

proportional to D. At low distortion levels, the size of $\mathcal{H}_i(D)$ is small, and the probability that it has no peak hypothesis is generally high. Accordingly, the upper bound defined in (1.15) becomes a less tight one (recall that this upper bound represents the probability that no peak hypothesis beats the true one). Note that we can quantitatively measure the tightness of the upper bound using the ratio

$$F = \frac{1 - \text{predicted\_upper\_bound}}{1 - \text{actual\_PCR}},$$

where $F \in [0, 1]$. As distortion increases, the size of $\mathcal{H}_i(D)$ increases as well, and the probability it has no peak hypothesis decreases. This increases the tightness of the upper bound, until it eventually coincides with the actual PCR, at which case F becomes equal to one.

We now turn to analyzing the real-distortion group. Actual and predicted PCR plots for real-distortion test sets $\mathcal{TD}_6$ and $\mathcal{TD}_7$ are shown in Figures 1.16(a) and 1.16(b), respectively. These figures show that our prediction method succeeds in predicting reasonably accurate bounds on actual performance. For example, the breakpoint in performance is predicted accurately in both cases. However, we also observe that the predicted bounds are over-optimistic in the knee section of the plot. The reason is obviously the existence of some differences between the actual distortion models and the assumed uniform ones. One of the important differences is that the assumed uniform occlusion model does not account for the spatial correlation among occluded/unoccluded features.

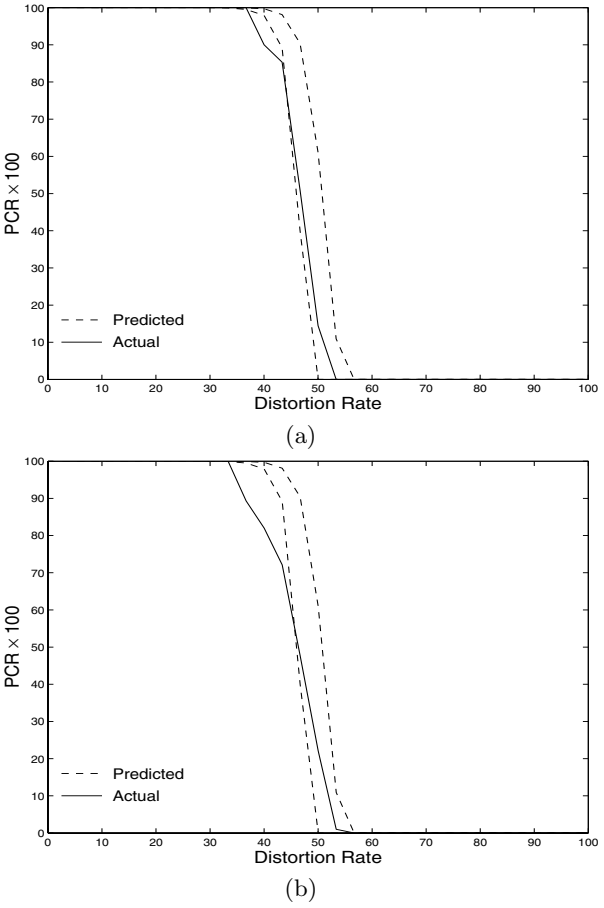
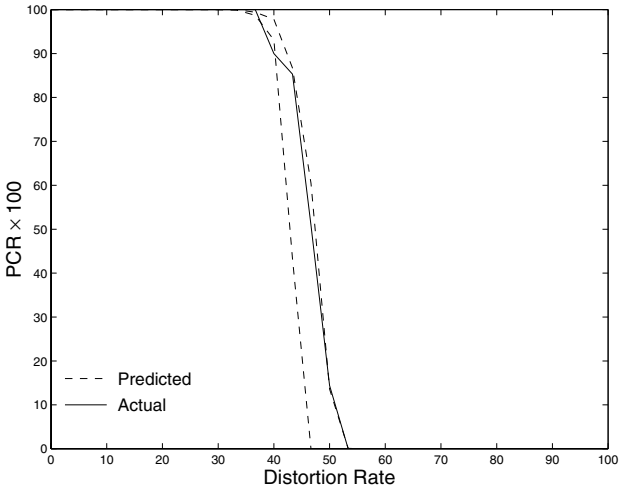


Figure 1.16. Actual and predicted PCR plots for real-distortion group: (a) $\mathcal{TD}_6$, (b) $\mathcal{TD}_7$.

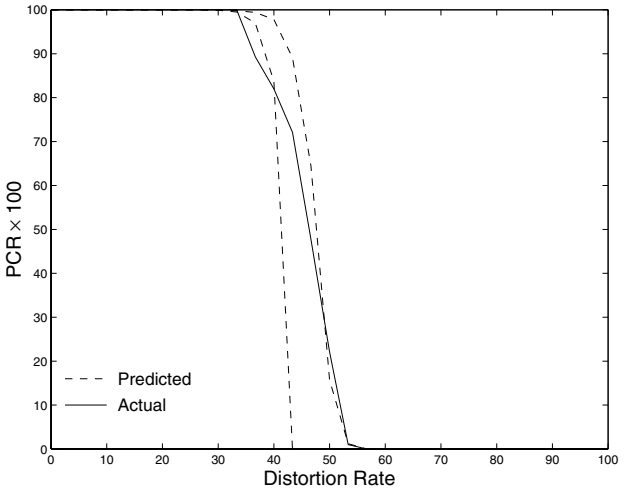
A goal of our research is to develop statistical techniques for learning uncertainty, occlusion, and clutter models. The topic of learning distortion models has received some attention in the literature, particularly in learning distributions of feature positional uncertainty [29]. We present here an initial method for learning the statistical model of occlusion, for the purpose of performance prediction. A possible direct approach is to learn the spatially correlating occlusion model from training data (e.g., using Markov random fields [26]), and then use it to determine the prediction bounds. In particular, the occlusion model can be used to determine the PDF of the number of unoccluded similar features (PDF of N_o), which is subsequently used to determine the PDF of votes for an erroneous hypothesis (see Section 1.6.2). The difficulty with this approach is that it involves not only a spatially correlating occlusion model, but also a spatially correlating similarity model, one that considers the spatial correlation among similar/dissimilar feature pairs. Fortunately, in the context of our problem, there is a significantly simpler approach: Instead of learning the spatially correlating occlusion and similarity models, and then estimating their combined effect on the PDF of N_o, we directly learn the PDF of N_o. The learning process can be outlined as follows:

1. We use half of the test set for the learning process (odd-numbered views).
2. For each selected test view, $\widehat{\mathcal{M}}_i$, we search for the best matching model view, $\mathcal{M}_i$, as described earlier when estimating the distortion rate for a test view (see Section 1.7.2). From this match, we eliminate unmatched features of $\mathcal{M}_i$. The resulting view, $\overline{\mathcal{M}}_i$, corresponds to model view $\mathcal{M}_i$ after occlusion, but without either uncertainty or clutter.
3. For each erroneous hypothesis corresponding to $\mathcal{M}_i$, $\mathcal{M}_j^{\tau_i} \in \mathcal{H}_i$, we calculate the number of similar feature pairs with each of $\mathcal{M}_i$ and $\overline{\mathcal{M}}_i$, denoted by $N_j^{\tau_i}$ and $\bar{N}_j^{\tau_i}$, respectively.
4. For each tuple $(\mid \mathcal{M}_i \mid, O = \mid \mathcal{M}_i \mid - \mid \overline{\mathcal{M}}_i \mid, N_j^{\tau_i})$, we histogram corresponding values of $\bar{N}_j^{\tau_i}$ to estimate the conditional PDF of N_o.

We have replaced the PDF of N_o defined in (1.6) by the learned PDF. The resulting new predicted bounds are shown in Figures 1.17(a) and 1.17(b) for test sets $\mathcal{TD}_6$ and $\mathcal{TD}_7$, respectively. Comparing these figures to the corresponding ones in Figure 1.16, we observe that the predicted bounds obtained using the learned PDF of N_o are considerably more accurate than the ones obtained by assuming the uniform distortion models. In particular, the extent of over-optimism between actual and predicted plots almost disappears for set $\mathcal{TD}_6$, and decreases by about 65% for set $\mathcal{TD}_7$. This is assuming that the extent of over-optimism is measured by the area between the actual plot and the predicted lower-bound plot at the knee section. More accurate prediction can be obtained through comprehensive learning of the PDFs of all the distortion parameters involved.



(a)



(b)

Figure 1.17. Actual and predicted PCR plots for real-distortion group using learned PDF of N_o: (a) $\mathcal{TD}_6$, (b) $\mathcal{TD}_7$.

1.8 Conclusions

Object recognition performance is typically evaluated empirically. A major limitation of empirical evaluation is that it does not give us an insight into the recognition process. Such an insight is fundamental for transforming the field of object recognition from an art to a science. Most efforts for formal analysis of object recognition performance focus on the problem of object/clutter discrimination. This work extends those efforts by also considering the prob-

lem of object/object discrimination. An integrated approach is presented for predicting lower and upper bounds on recognition performance. Such an approach simultaneously considers data distortion factors such as uncertainty, occlusion, and clutter, in addition to model similarity. This is in contrast to the other few object/object discrimination approaches, which consider only a subset of these factors. The method is validated in the context of a recognition task involving real SAR data, and point features for matching. Validation is done by comparing predicted lower and upper PCR bounds with actual PCR plots determined experimentally. The selected test sets involve both synthetic and real distortion. In all cases, the results demonstrate that the prediction method consistently succeeds in predicting reasonably tight bounds on actual performance. The predicted bounds, however, are slightly over-optimistic in the case of test sets involving real distortion. More accurate bounds can be obtained through comprehensive learning of all the distortion models involved in the recognition process. As an initial step towards achieving this goal, we have presented a method for learning the statistical distribution of the parameter that encodes combined effect of occlusion and similarity on performance. The impact of using such a learned distribution on the accuracy of the predicted bounds has been demonstrated. To the best of our knowledge, this work is the first that validates a prediction method for object recognition using real data.

We finally discuss possible ways to extend the work presented here. The recognition task considered in this work involves discretized 2D point features for both data and model objects, and a 2D translation space. The proposed method, however, can be extended in a number of ways. We explore some of them below:

- Consideration of more general data/model transformations (e.g., rigid, affine) is straightforward. It would mainly involve extending the set of erroneous hypotheses for a model object, to reflect the extra degrees of freedom (see Sect. 1.6.3).
- Replacement of discretized feature sets by non discretized sets is a more involved process. This is because, in such a case, the erroneous-hypothesis set for a model object would be of infinite size. One possible approach is sampling the erroneous hypotheses in the transformation space, and then "dilating" the consistency regions associated with model features to account for non-sampled hypotheses (e.g., [21]). We note that consistency-region dilation would be needed only when computing lower bounds on performance.
- It is also possible to incorporate feature attributes [5], and/or increase dimensionality of feature locations. Both cases would mainly involve increasing the dimensionality of consistency and clutter regions to be equal to the sum of location dimensionality and number of attributes. Otherwise, the method remains basically the same.

- Another interesting possibility is to describe the positional uncertainty by a Gaussian distribution, instead of a uniform one. This would involve replacing the uniform vote-based criterion by a weighted one, which depends on the distance between corresponding data and model features, as well as the standard deviation of the Gaussian distribution [18]. This would require making modifications to both the measure of object/hypothesis similarity, and the method used to calculate the vote distributions for both true and erroneous hypotheses. The prediction method, however, remains conceptually the same.
- The performance theory presented here can be integrated with adaptive algorithms to optimize object recognition performance in practical scenarios (e.g., with respect to a desired trade-off between probability of false alarm and probability of correct recognition [30]).

In conclusion, the work presented here is believed to lay a theoretical/conceptual foundation that is important for developing a general theory for performance prediction of object recognition.

Appendix

We present an approximate method for estimating the conditional PDF of clutter votes V_c, considering the constraint that no two features can be eight neighbors (imposed by the feature extraction process). We can represent the feature-adjacency constraint by a separation region, $R_s(\cdot)$. In our case, $R_s(\cdot)$ is a 3×3 window centered at the feature's location. The method can be outlined as follows:

- It can be seen that the feature-adjacency constraint reduces effective clutter and clutter-vote regions, R'_c and R'_{V_c}, to smaller ones, R''_c and R''_{V_c}, respectively (see Figure 1.18). We can express the area of R''_c as

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

  For clutter-vote region R''_{V_c}, let us first consider features in $\mathcal{M}_i$ that are similar to those in $\mathcal{M}_j^{\tau_i}$, but are lying *outside* the overlapping regions (e.g., upper right feature of $\mathcal{M}_i$ in Figure 1.18). Generally, the separation region corresponding to one of these features partially covers the consistency region corresponding to the similar feature in $\mathcal{M}_j^{\tau_i}$. We can approximate the area of the covered section, denoted by E_d, as follows. (1) Both the consistency and separation regions are modeled as squares, having the same areas as their original respective regions. (2) The relationship between square consistency regions corresponding to similar feature pairs is modeled by movement of one square with respect to the other parallel to one of the edges, until the area of the overlapping region is the same as the original one (which corresponds to feature/feature similarity). (3) Assuming the

square modeling of consistency regions, we can easily estimate expected location of a feature of $\mathcal{M}_i$, assuming that it is outside the overlapping region. Given the expected location, we can easily estimate the area of the covered section, E_d. We can then estimate the area of R''_{V_c} as

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

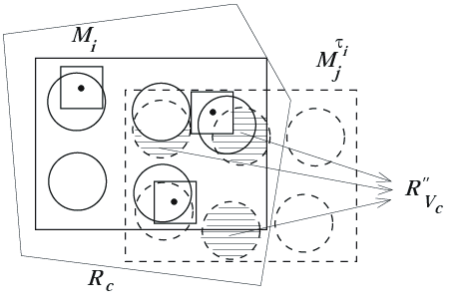


Figure 1.18. An illustration of clutter vote region R''_{V_c}, in the presence of separation regions, shown as squares.

- The above discussion, as well as that in Section 1.6.2, implies that features of $\mathcal{M}_j^{\tau_i}$ that can result in votes for $\mathcal{M}_j^{\tau_i}$ are effectively associated with consistency regions of various sizes. In order to simplify the calculations, we assume that we have an "effective" number of these features, n_e, such that each feature is associated with a full consistency region. It can be easily shown that $n_e = \lfloor \text{AREA}(R''_{V_c})/\text{AREA}(R_u(\cdot)) + 0.5 \rfloor$. In such a case, the effective area of the consistency region, E_u, is $\text{AREA}(R''_{V_c})/n_e$, which is approximately the same as $\text{AREA}(R_u(\cdot))$.
- Next, we need to estimate the average *effective* area of a separation region, E_s. This area is simply the ratio between the area corresponding to the union of the separation regions of the C clutter features, to C. If the clutter features are spread apart from each other, then E_s is close to $\text{AREA}(R_s(\cdot))$. However, if their density is high, which is typical in our case, then there can be significant overlap between their separation regions, thus making E_s considerably smaller than $\text{AREA}(R_s(\cdot))$. In our work, we have

$$E_s \approx \min(\text{AREA}(R_s(\cdot)), \text{AREA}(R''_c)/C). \tag{1.18}$$

- Now, we are finally at a position to approximate the conditional PDF of V_c:

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

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

Acknowledgments

This work was supported in part by AFOSR grant F49620-02-1-0315. The contents and information do not reflect positions or policies of the U.S. Government.

References

[1] Arman, F., Aggarwal, J.: Model-based object recognition in dense-range images: A review. ACM Comput. Surveys **25** (1993) 5–43
[2] Besl, P., Jain, R.: Three-dimensional object recognition. ACM Comput. Surveys **17** (1985) 75–145
[3] Chin, R., Dyer, C.: Model-based recognition in robot vision. ACM Comput. Surveys **18** (1986) 67–108
[4] Suetens, P., Fua, P., Hanson, A.: Computational strategies for object recognition. ACM Comput. Surveys **24** (1992) 5–61
[5] Bhanu, B., III, G.J.: Recognizing target variants and articulations in synthetic aperture radar images. Optical Engineering **39** (2000) 712–723
[6] Boshra, M., Zhang, H.: A constraint-satisfaction approach for 3-D object recognition by integrating 2-D and 3-D data. Comput. Vision Image Understand. **73** (1999) 200–214
[7] Wells, W.: Statistical approaches to feature-based object recognition. Int. J. of Computer Vision **21** (1997) 63–98
[8] Alter, T., Grimson, W.: Verifying model-based alignments in the presence of uncertainty. In: Proc. IEEE Conf. Comput. Vision and Patt. Recogn., San Juan, Puerto Rico (1997) 344–349
[9] Boykov, Y., Huttenlocher, D.: A new Bayesian framework for object recognition. In: Proc. IEEE Conf. Comput. Vision and Patt. Recogn. Volume 2., Fort Collins, Colorado (1999) 517–523
[10] III, G.J., Bhanu, B.: Recognition of articulated and occluded objects. IEEE Transactions on Pattern Anal. and Mach. Intell. **21** (1999) 603–613
[11] Dhome, M., Richetin, M., Lapreste, J., Rives, G.: Determination of the attitude of 3-D objects from a single perspective view. IEEE Transactions on Pattern Anal. and Mach. Intell. **11** (1989) 1265–1278
[12] Huttenlocher, D., Ullman, S.: Recognizing solid objects by alignment with an image. Int. J. of Computer Vision **5** (1990) 195–212
[13] Dhome, M., Kasvand, T.: Polyhedra recognition by hypothesis accumulation. IEEE Trans. on Pattern Anal. and Mach. Intell. **9** (1987) 429–439
[14] Stockman, G.: Object recognition and localization via pose clustering. Comput. Vision Graphics Image Process. **40** (1987) 361–387

[15] Faugeras, O., Hebert, M.: The representation, recognition and locating of 3-D objects. Int. J. of Robotic Res. **5** (1986) 27–52
[16] Grimson, W., Lozano-Perez, T.: Localizing overlapping parts by searching the interpretation tree. IEEE Trans. on Pattern Anal. and Mach. Intell. **9** (1987) 469–482
[17] Grimson, W., Huttenlocher, D.: On the verification of hypothesized matches in model-based recognition. IEEE Transactions on Pattern Anal. and Mach. Intell. **13** (1991) 1201–1213
[18] Sarachik, K.: The effect of Gaussian error in object recognition. IEEE Transactions on Pattern Anal. and Mach. Intell. **19** (1997) 289–301
[19] Lindenbaum, M.: An integrated model for evaluating the amount of data required for reliable recognition. IEEE Transactions on Pattern Anal. and Mach. Intell. **19** (1997) 1251–1264
[20] Irving, W., Washburn, R., Grimson, W.: Bounding performance of peak-based target detectors. In: Proc. SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery IV. Volume 3070. (1997) 245–257
[21] Lindenbaum, M.: Bounds on shape recognition performance. IEEE Transactions on Pattern Anal. and Mach. Intell. **17** (1995) 665–680
[22] Grenander, U., Miller, M., Srivastava, A.: Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR. IEEE Transactions on Pattern Anal. and Mach. Intell. **20** (1998) 790–802
[23] Boshra, M., Bhanu, B.: Predicting performance of object recognition. IEEE Trans. on Pattern Anal. and Mach. Intell. **22** (2000) 956–969
[24] Boshra, M., Bhanu, B.: Validation of SAR ATR performance prediction using learned distortion models. In: Proc. SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery VII. Volume 4053., Orlando, Florida (2000) 558–566
[25] Boshra, M., Bhanu, B.: Predicting an upper bound on SAR ATR performance. IEEE Trans. on Aerospace and Electronic Syst. **37** (2001) 876–888
[26] Li, S.: Markov Random Field Modeling in Image Analysis. Springer-Verlag, New York (2001)
[27] Ying, Z., Castanon, D.: Feature-based object recognition using statistical occlusion models with one-to-one correspondence. In: Proc. Int. Conf. on Comput. Vision. Volume 1., Vancouver, Canada (2001) 621–627
[28] Ross, T., Worrell, S., Velten, V., Mossing, J., Bryant, M.: Standard SAR ATR evaluation experiments using the MSTAR public release data set. In: Proc. SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery V. Volume 3370., Orlando, Florida (1998) 566–573
[29] Pope, A., Lowe, D.: Probabilistic models of appearance for 3-D object recognition. Int. J. of Computer Vision **40** (2000) 149–167
[30] Bhanu, B., Lin, Y., Jones, G., Peng, J.: Adaptive target recognition. Machine Vision and Applications **11** (2000) 289–299