Advances in Pattern Recognition

Advances in Pattern Recognition is a series of books which brings together current developments in all areas of this multi-disciplinary topic. It covers both theoretical and applied aspects of pattern recognition, and provides texts for students and senior researchers.

Springer also publishes a related journal, **Pattern Analysis and Applications.** For more details see: http://link.springer.de

The book series and journal are both edited by Professor Sameer Singh of Exeter University, UK.

Also in this series:

Principles of Visual Information Retrieval
Michael S. Lew (Ed.)
1-85233-381-2

Statistical and Neural Classifiers: An Integrated Approach to Design
Šarūnas Raudys
1-85233-297-2

Advanced Algorithmic Approaches to Medical Image Segmentation
Jasjit Suri, Kamaledin Setarehdan and Sameer Singh (Eds)
1-85233-389-8

NETLAB: Algorithms for Pattern Recognition
Ian T. Nabney
1-85233-440-1

Object Recognition: Fundamentals and Case Studies
M. Bennamoun and G.J. Mamic
1-85233-398-7

Bir Bhanu and Ioannis Pavlidis (Eds)

Computer Vision Beyond the Visible Spectrum

With 156 Figures

Springer

Bir Bhanu, PhD, Fellow IEEE, AAAS, IAPR, Senior Honeywell Fellow (Ex.)
Center for Research in Intelligent Systems, University of California at Riverside, CA, USA

Ioannis Pavlidis, PhD
Department of Computer Science, University of Houston, TX, USA

Series editor
Professor Sameer Singh, PhD
Department of Computer Science, University of Exeter, Exeter, EX4 4PT, UK

British Library Cataloguing in Publication Data
Computer vision beyond the visible spectrum. — (Advances in pattern recognition)
1. Computer vision 2. Pattern recognition systems
I. Bhanu, Bir II. Pavlidis, Ioannis
006.3′7
ISBN 1852336048

Library of Congress Cataloging-in-Publication Data
CIP data available.

Advances in Pattern Recognition ISSN 1617-7916
ISBN 1-85233-604-8 Springer London Berlin Heidelberg
Springer is a part of Springer Science+Business Media
springeronline.com

Printed and bound in the United States of America
34/3830-543210 Printed on acid-free paper SPIN 10868565

Preface

Traditionally, computer vision has focused on the visible band for a variety of reasons. The visible band sensors are cheap and easily available. They are also sensitive in the same electromagnetic band as the human eye, which makes the produced data more interesting from the psychophysiology point of view. In fact, computer vision was pre-occupied for a long time with the problem of understanding and imitating the human visual system. Recently, this obsession subsided and computer vision research focused more on solving particular application problems with or without the help of the human visual paradigm. A case in point is the significant progress achieved in object tracking.

It so happens that many imaging applications cannot be addressed in the visible band. For example, visible sensors cannot see in the dark; thus, they are not very useful in military applications. Visible radiation cannot penetrate the human body and, therefore, cannot be a viable medical imaging modality. Other electromagnetic bands and sensor modalities have been identified and developed over the years that can solve all these problems, which are beyond the reach of the visible spectrum. Initially, it was primarily phenomenological and sensory work that was taking place. Later came algorithmic work, and with that computer vision beyond the visible spectrum was born.

In this book, we explore the state-of-the-art in *Computer Vision Beyond the Visible Spectrum (CVBVS)* research. The book is composed of nine chapters which are organized around three application axes:

1. Military applications with an emphasis on object detection, tracking, and recognition.
2. Biometric applications with an emphasis on face recognition.
3. Medical applications with an emphasis on image analysis and visualization.

Although the chapters describe research, they are not written as typical research papers. They have a tutorial flavor appropriate for a book.

The book opens with the military applications since they represent the birthplace of CVBVS. All the major modalities used in military applications

are represented in the first five chapters. These include SAR (Synthetic Aperture Radar), laser radar, hyperspectral, and infrared. The first five chapters also address fundamental issues with regard to object detection, tracking, and recognition, sometimes in more than one modality. This allows comparative evaluation of these important computational imaging questions across the electromagnetic spectrum.

In Chapter 1, Boshra and Bhanu et al. describe a theoretical framework for predicting the performance of object (target) recognition methods. The issue of identifying military targets in imagery is of great importance in military affairs. For years, target recognition was based purely on heuristics, and as a result performance was brittle. Boshra and Bhanu's work is representative of a more rigorous methodological approach, which promises to transform target recognition from art to science.

In Chapter 2, Bhanu and Jones unveil specific methods for improving the performance of an SAR target recognition system. SAR is probably the most successful imaging modality for military applications, because of its all-weather capability. Bhanu and Jones' methods conform to the model-based framework and involve incorporation of additional features, exploitation of a priori knowledge, and integration of multiple recognizers.

In Chapter 3, Arnold et al. present target recognition methods in a different modality, namely, three-dimensional laser radar. Three-dimensional laser radars measure the geometric shape of targets. The main approach described in this chapter is quite appealing because it bypasses detection and segmentation processes.

In Chapter 4, Kwon et al. deal with target recognition in the context of hyperspectral imagery. The basic premise of hyperspectral target recognition is that the spectral signatures of target materials are measurably different than background materials. Therefore, it is assumed that each relevant material, characterized by its own distinctive spectral reflectance or emission, can be identified among a group of materials based on spectral analysis of the hyperspectral data. Kwon et al. use independent component analysis (ICA) to generate a target spectral template. ICA is a method well-suited to the modular character of hyperspectral imagery.

In Chapter 5, Vaswani et al. close the sequence of military application papers by presenting a method for object detection and compression in infrared imagery. The proposed solution is guided by the limitations of the target platform, which is an infrared camera with on-board chip. The object detection method is computationally efficient, to deal with the limited processing power of the on-board chip. It is also paired with a compression scheme to facilitate data transmission.

Chapter 6 deals with biometrics and signals a transition from the military to civilian security applications. Wolff et al. present a face recognition approach based on infrared imaging. Infrared has advantages over visible imaging for face recognition, especially in the presence of variable lighting conditions.

Wolff et al. provide quantitative support for this argument by unveiling a system that performs comparative evaluation.

Chapter 7 opens the medical applications part of the book. It refers to cardiovascular image analysis of magnetic resonance imagery (MRI). While SAR is probably the most successful modality for military applications, one could make the case that MRI is the most successful modality for medical applications. Initially, MRI was treated much like x-rays. A radiologist, without any machine assistance, was interpreting the raw imagery. Increasingly, however, computer vision methods aid in this interpretation. In this chapter, Sonka et al. present techniques for 3D segmentation and quantitative assessment of left and right cardiac ventricles, arterial and venous trees, and arterial plaques.

In Chapter 8, Fenster et al. present segmentation and visualization techniques in another very important medical imaging modality, that is, ultrasound. Specifically, the authors describe methods to reconstruct ultrasound information into 3D images to facilitate interactive viewing. They also describe automated and semi-automated segmentation methods to quantify organ and pathology volume for monitoring disease.

In Chapter 9, Berry et al. introduce some very interesting image analysis work on a novel medical imaging modality, namely, terahertz pulsed imaging. Vis-a-vis the more established MRI and ultrasound modalities, terahertz pulsed imaging is the "new kid on the block". Berry et al. propose Fourier transforms and wavelets to analyze spectroscopic information of materials. They actually demonstrate that these methods perform as well as traditional analysis methods for material properties and predict a number of biomedical applications that stand to benefit form this technology.

The book can be used for instruction in graduate seminars or as a reference for the independent researcher. Although CVBVS is a broad and fast moving field, the balanced selection of key theoretical and practical issues represented in the chapters of the book will maintain their relevance for some time. It is our sincere hope that the book will serve as a springboard for the individual researcher who is interested in CVBVS research.

A number of people have contributed in our effort and we are deeply grateful to all of them. These certainly include the authors of the individual chapters and the reviewers who patiently went through three review cycles. We are especially grateful to Pradeep Buddharaju who handled most of the last minute editing and thanks to whom the book assumed its finished form.

Houston, Texas — *Ioannis Pavlidis*
Riverside, California — *Bir Bhanu*
USA
January 2004

Contents

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.

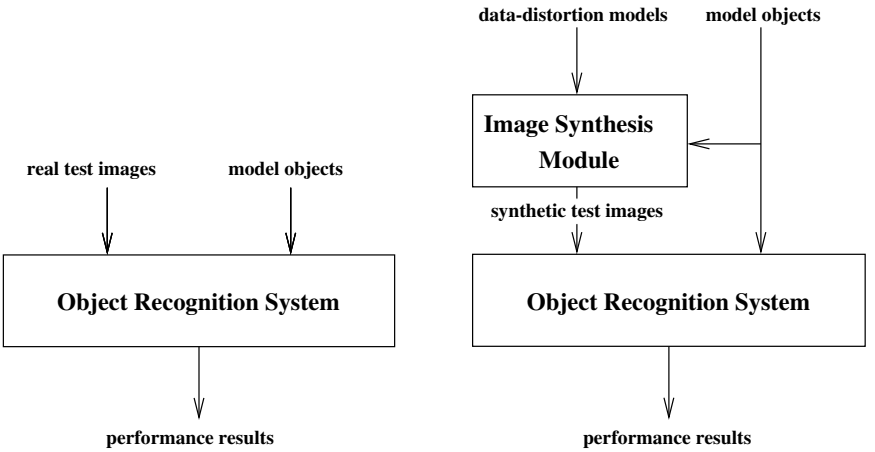


Figure 1.1. Empirical approaches for estimation of object recognition performance: (a) using real data, (b) using synthetic data.

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

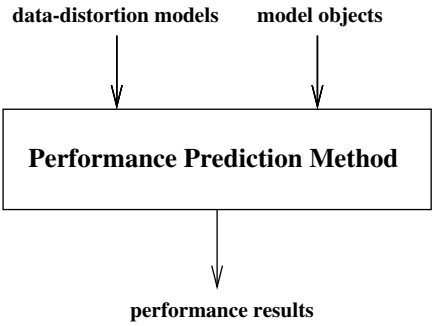


Figure 1.2. Formal estimation of object recognition performance.

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

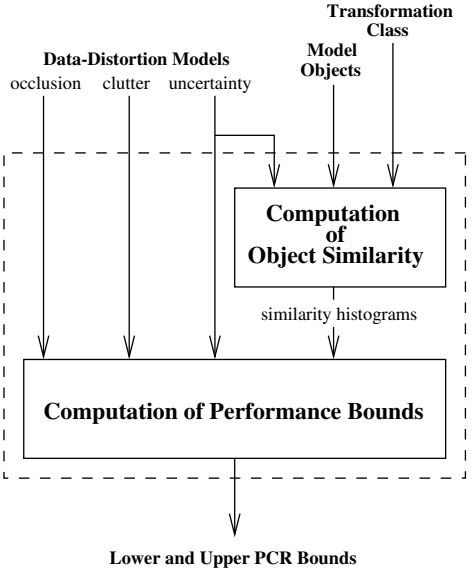


Figure 1.3. Block diagram of performance–prediction method.

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.

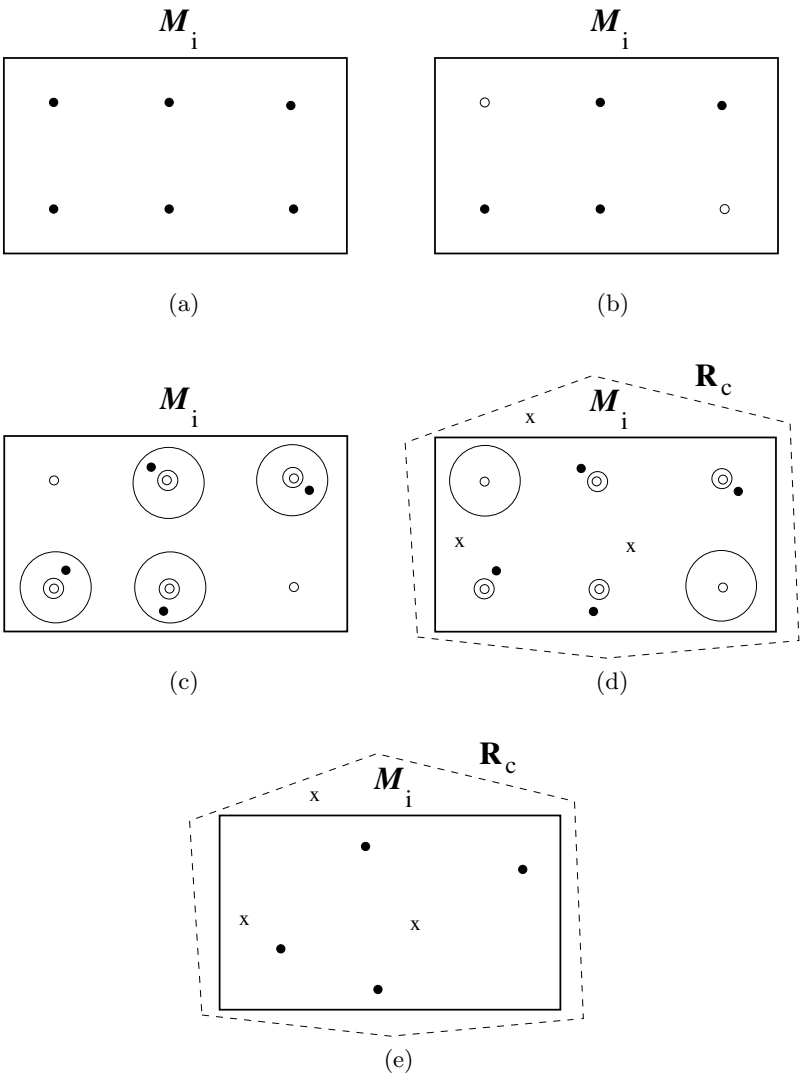


Figure 1.4. An illustration of the different stages of the distortion process: (a) original object consisting of six features (dark circles), (b) after occlusion ($O = 2$; small circles represent occluded features), (c) after perturbation ($R_u(\cdot)$ = a circle; small double circles represent locations 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

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

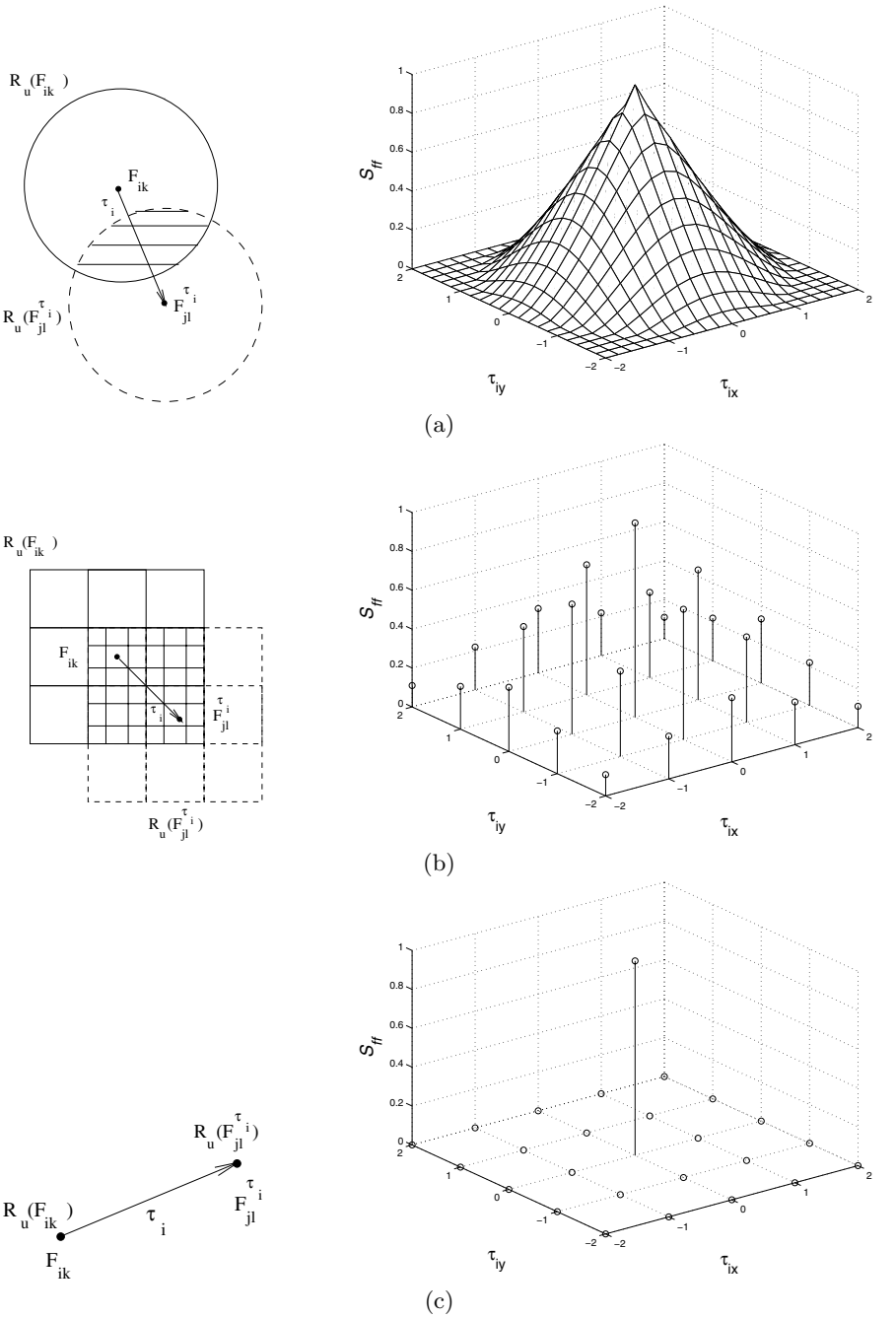


Figure 1.5. : An illustration of feature/feature similarity for a variety of consistency regions assuming $\mathcal{T}$ is the space of 2D translations: (a) a circle of unit radius, (b) a discrete eight-neighbor region, (c) a point region (implies absence of positional uncertainty). Note that the components 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.

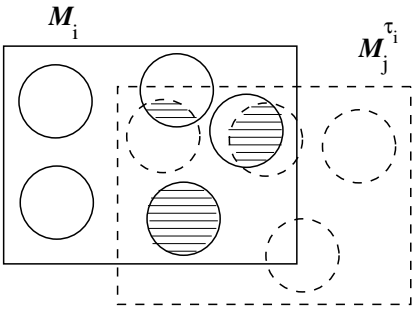


Figure 1.6. An illustration of object/hypothesis similarity. Notice that there are three similar feature pairs with feature/feature similarity values of approximately $\frac{1}{3}$, $\frac{2}{3}$ and 1. Accordingly, we have $S_j^{\tau_i} \in [1, 3]$, and $E(S_j^{\tau_i}) \approx 2$.

• **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 $P_X(x) = \Pr[X = x]$, $B_X(x; n, p) = K(n, x)p^x(1-p)^{n-x}$, $K(a, b) = \frac{a!}{(a-b)!\ b!}$,

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

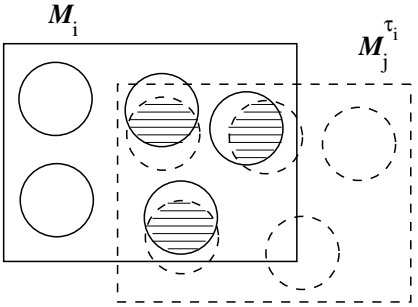


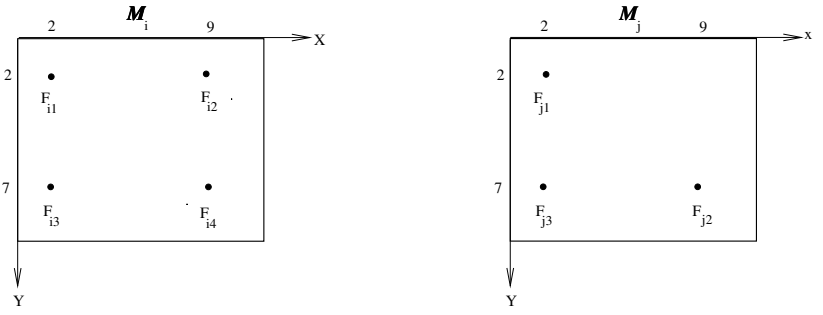
Figure 1.7. Uniform similarity model for object/hypothesis pair shown in Figure 1.6. Notice that similar feature pairs have constant feature/feature similarity, $P_j^{\tau_i} \approx \frac{2}{3}$ and $N_j^{\tau_i} = 3$.

• **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.

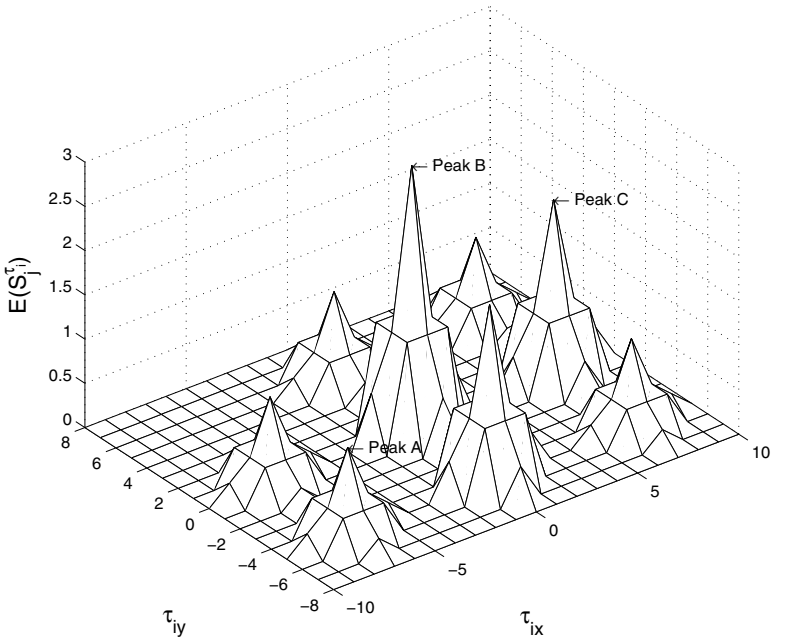
1.5.2 Construction of Similarity Histograms

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 $\mathcal{M}_i$.

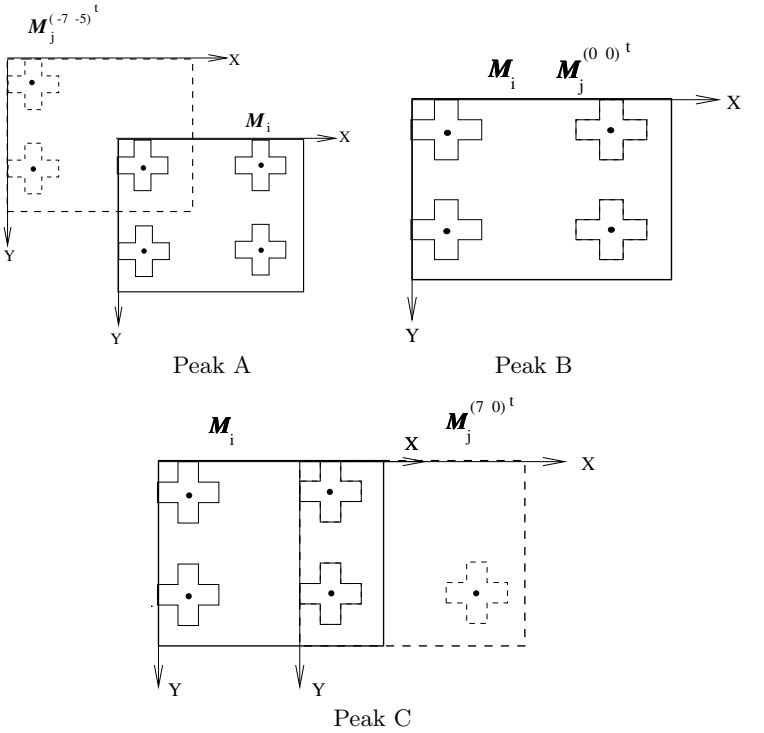


Figure 1.9. Three hypotheses corresponding to peaks A, B and C shown in Fig. 1.8(b), assuming four-neighbor consistency region.

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 $\mathcal{M}_j^{\tau_i}$.

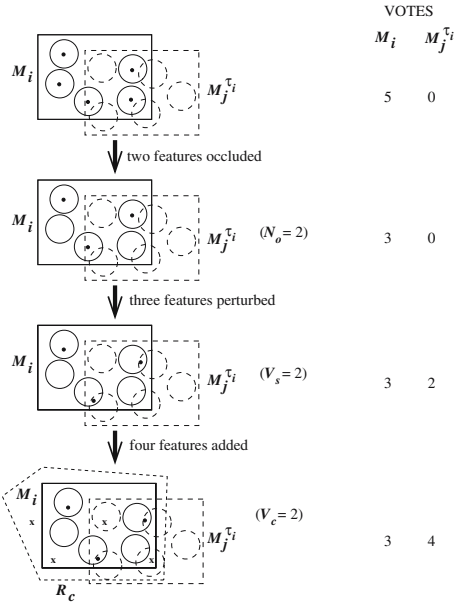


Figure 1.11. An example showing the vote process for object $\mathcal{M}_i$ and erroneous hypothesis $\mathcal{M}_j^{\tau_i}$, as $\mathcal{M}_i$ gets distorted.

- 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 $\mathcal{M}_i$.

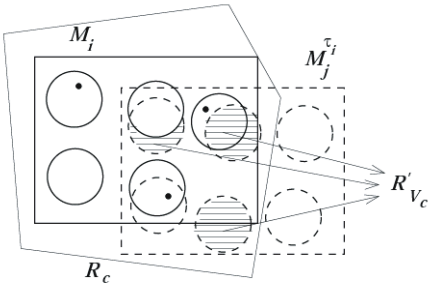


Figure 1.12. An illustration of the clutter vote region R'_{V_c}, assuming the uniform model of similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^{\tau_i}$.

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