A BAYESIAN FRAMEWORK FOR CONSIDERING PROBABILITY DISTRIBUTIONS OF IMAGE SEGMENTS AND SEGMENTATIONS

BY

STEVEN MICHAEL LAVALLE

B.S., University of Illinois at Urbana-Champaign, 1990

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1993

Urbana, Illinois

ABSTRACT

We have seen many probabilistic approaches to image segmentation. Traditionally the general paradigm has been to formulate statistical image models and design an algorithm which iteratively approaches an optimal segmentation. The resulting segmentation is often utilized by a more abstract process, such as a model-based recognition system. Although this segmentation is optimal (or near-optimal) with respect to the chosen models, the problem is generally considered underconstrained. Consequently, the segmentation may not contain the best homogeneous regions needed by the abstract process (i.e., a recognition system cannot exert complex model-based influences directly on the selection of an optimal segmentation).

We depart from this paradigm and develop a framework for probabilistically maintaining sets of alternative homogeneous regions, and segmentations. Depending on the image size and complexity, and on the application, a probability distribution can be constructed over segmentations of the entire image, or a distribution over partial segmentations can be formed. We develop an efficient representation scheme, and a probabilistic mechanism for applying Bayesian, model-based evidence to guide the construction of the representation.

Since probability distributions over the space of alternatives are of primary importance to us (rather than some decision criteria necessary to select a single segmentation), we have developed a general Bayesian formalism for determining the posterior probability that the union of regions is homogeneous, given that the individual regions are homogeneous. This method does not rely on estimation, and properly treats the issues involved when sample sets are small and estimation performance degrades. We apply the general formulation to the implicit polynomial surface model, on range images. This model involves three high-dimensional manifold integrations over a density function on the parameter space; therefore, we present a novel Monte Carlo-based computation scheme for efficient evaluation of the integrals. Several experimental results are presented using planar and quadric models on real and synthetic range data, using a Gaussian noise model.

DEDICATION

In memory of my mom, Rose M. LaValle (1943-1992)

ACKNOWLEDGMENTS

I would first like to thank my advisor, Professor Seth Hutchinson, for providing an environment which encourages the development of new ideas. He has always been open and patient with discussions of different research ideas, whether or not they had been carefully refined. The quality of this work was greatly improved through his many helpful comments and suggestions.

I would also like to thank my comrades in our Vision/Robotics research group for endless hours of fun (including good research discussions), especially Andrés Castaño, Armando Fox, Ken Moroney, Sandeep Pandya, and Steve Sullivan. They have provided much entertainment during those late hours in the lab.

I especially thank Becky Anderson for reading an earlier draft of this thesis, and providing numerous helpful comments.

I thank Pat Flynn and the MSU Pattern Recognition and Image Processing Lab for providing us with range images.

Lastly (and mostly), I thank my wife, Tammy, for being patient and loving during those endless hours spent on this work.

This work was sponsored by NSF under grant #IRI-9110270.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Image segmentation has been a topic of active research for over two decades. The segmentation problem is typically defined as the low-level vision task of extracting a set of homogeneous regions (called segments) from an image, often for the purpose of higher-level processing. Although considerable research effort has yielded a number of approaches to the problem, segmentation remains a difficult problem in its general formulation. As Horn has pointed out, one of the primary difficulties in evaluating a segmentation method is the lack of a clear definition of the "correct" segmentation [1]. This definition usually depends on the intended application of the segmentation result. Szeliski argues that low-level image models often underconstrain the solution, and advocates the use of uncertainty estimation [2]. This type of difficulty in segmentation provides much of the impetus of our work: to represent multiple segments and segmentations, probabilistically, in a Bayesian framework.

For a conceptual understanding of segmentation, it is useful to consider the formal description given by Horowitz and Pavlidis [3]. Intuitively, one wants the segments in a segmentation to be homogeneous or uniform in some sense. Abstractly, this can be formalized with a *homogeneity predicate* (or *uniformity predicate*). Consider an image, D, as a set of elements, with each element containing some information. It is assumed that we are given some homogeneity predicate, H, which applies to subsets of D. If R

is some subset of D, then predicate, H, returns $true$ if the set is homogeneous or $false$ otherwise.[1]

Examples of simple homogeneity predicates on intensity images, taken from [4], are

- $H(R) = true$ if and only if all elements of R have the same intensity value,
- $H(R) = true$ if and only if the intensity does not differ by more than some fixed amount for any two elements of R.

These are merely provided as intuitive examples. In Chapter 4, we consider a large class of models that induce uniformity predicates, and in Chapter 5 we consider $H(R) = true$ if the points in R were observed from the same object surface patch in the presence of noise.

A segmentation is defined for some H and D as a collection of disjoint nonempty subsets $X_1, \ldots, X_n$ such that

1. $\bigcup X_i = D$
2. X_i is connected (optional and must be consistently defined)
3. $H(X_i) = true \ \ \forall X_i$
4. $H(X_i \cup X_j) = false \ \ \forall X_i \neq X_j$.

We consider a segment to be a maximal region such that $H(R) = true$.[2] Pavlidis points out that it is possible to define uniformity predicates that do not lead to a unique segmentations [4]; however, we will not consider this class of predicates.

The set of all partitions of D into connected regions is typically large, and due to noise and other uncertainty, evaluation of a uniformity predicate may not be directly

[1] This is primarily a conceptual formulation, and in practice a logical predicate is usually not specified in this form.

[2] In Chapter 3, we will introduce alternative, but equivalent definitions of segments and segmentations.

possible. This has lead to numerous approaches to image segmentation. Section 1.1 provides a brief, general survey of segmentation work and related approaches. Section 1.2 introduces the motivation and goals for our Bayesian, segmentation sample space framework. Section 1.3 concludes this chapter with an outline this thesis.

1.1 General Approaches to Segmentation

The type of image often determines the segmentation models that are appropriate. Images are usually rectangular arrays, with elements consisting of intensity values or 3D coordinates of points. The intensity values could be obtained through a standard camera-imaging system, which could yield a set of gray-scale values, or could encode color information with, for instance, RGB values. Segmentations are obtained from other types of intensity images, such as X-ray radiographs for manufacturing inspection [5], or Electron-Microscopic Autoradiography (EMA) for identifying structures in biological tissue [6]. Images consisting of 3D coordinates are referred to as range images, and the information is often obtained through some form of active sensing, such as triangulation with a camera and laser light source. The information could also be obtained through binocular (or trinocular) stereo matching [7],[8],[9] or a motion sequence [10]. Ikeuchi and Kanade give a detailed description of a variety of imaging sensors [11].

Most segmentation approaches can be considered *edge-based* or *region-based* (and in many cases both). Edge-based approaches consider information at the boundaries of two adjacent regions. When a region contains an edge the homogeneity predicate is $false$. Edge-based methods are characterized by their computational efficiency [12], and a good overview of edge-based methods is given by Horn [1].

Region-based approaches use information that is common to the points within regions. Region-based methods usually begin with some set of small regions, and attempt to construct larger regions by iteratively merging regions with similar features [13],[14].

One example of a region-based method is the split-and-merge procedure [4]. With this approach, the image is recursively split into regions that are believed to be homogeneous with high confidence. Regions with similar features are merged until some termination criterion is met. Other region-based approaches are mentioned in Section 5.1. There has been recent interest in the integration of edge-based and region-based methods [15],[16],[17],[18]. Such methods gain a significant increase in performance by the combination of both types of evidence.

Aside from the general strategy (e.g., edge-based vs. region-based), segmentation approaches differ by the types of models of homogeneity that are considered. Constancy of intensity or color is one of the more straightforward models. Segmentation is often performed by thresholding in the intensity/color space. A description of a variety of thresholding techniques appears in [19].

The facet model defines another type of homogeneity [20]. In general, a set of points under this model is homogeneous if the points can be well-approximated by some smooth surface such as a plane, quadric, or spline surface. Polynomial models are particularly popular for range image segmentation, and several approaches are discussed in Section 5.1. We use the implicit polynomial surface model to demonstrate our Bayesian framework in Chapters 5 and 6.

Models of texture have also been used extensively for segmentation. A texture is often considered as a local shape or pattern repeated throughout a region of the image. The segmentation goal is to partition the image into components having the same texture. Many approaches to texture involve the Markov random field (MRF) model, which is naturally suited to model local dependencies that arise with texture models [21],[22],[23]. Other models employ the use of probabilistic relaxation [24],[25], Gabor transforms [26],[27], Voronoi tesselations [28], and autoregressive random fields [29]. The MRF and proba-

bilistic relaxation approaches will be discussed in further detail in Sections 2.2 and 2.3, respectively.

Since there are numerous ways to partition an image, concerns about computational efficiency have inspired a number of computational strategies for producing an optimal or near-optimal segmentation. Three general strategies that appear are: iterative optimization [30] (and other MRF approaches), pyramidal/multiresolution approaches [31],[32],[33], [34],[35], and region-merging [4],[14].

Image restoration and pixel-labeling are problems closely related to image segmentation. With restoration, an observed image is considered to be the result of a transformation that is applied to an original image. The transformation contains components such as noise and blurring, and the goal is to determine a best estimate of the original image from the observed image. For an intensity image (the standard application) the goal is to select an assignment for each pixel from a set of possible intensities. We can view each of these assignments as a *pixel-labeling*, and can also build appropriate models for labeling pixels with qualitative information such as "edge" or "concave region" [36]. Labelings induce a partition (with not necessarily connected regions) of the image, in which each element of the partition is the set of all pixels receiving a certain labeling. Using qualitative labels (or line-processes, discussed in Section 2.2) one can construct a partition into connected segments, which is a segmentation in the sense previously defined. Hence, advancement in the area of image restoration often applies to image segmentation. For instance, Geman and Geman stimulated widespread use of Markov random field models for segmentation [30]. A comprehensive survey of image restoration methods can be found in [37].

Finally, due to inherent noise and uncertainty in images, probabilistic schemes have played a significant role in the development of segmentation methods. Existing general approaches are: statistical clustering, Markov random field/energy minization, and

probabilistic relaxation. These approaches are closely related to our work, and consequently Chapter 2 is devoted to their discussion.

1.2 Our Motivation and Objectives

In this section, we present our set of objectives for this work. Immediately following each objective is an explanation of our motivation for that objective. Some additional motivation is provided in Section 2.4 after other probabilistic approaches have been described in detail. We consider each of following objectives to be desirable characteristics for a segmentation approach; consequently, each guided the development of our framework. In Section 9.1 we return to these objectives, and discuss how they have been achieved through the work presented in this thesis.

To develop a system capable of representing any number of alternative segments or segmentations, and their corresponding probabilities.

Since the space of alternatives is often underconstrained using low-level models [2], one approach is to introduce more constraints through the use of higher-level models, for instance, at the recognition level. For this to occur, it is unreasonable to select a single, apparently best, segmentation to send to the higher-level process. The single segmentation has been formed by making all of the decisions using low-level models, and all other information is lost. For the higher-level models to participate in the segmentation process, it seems useful to at least give some set of alternative segmentations. Additional evidence can then begin to be applied by the higher-level process to constrain the space of segmentations, eventually resulting in a unique solution.

Rather than simply representing a set of alternatives, consider also obtaining probabilities for each of the alternatives. The probabilities give much more information than

is present in the set of alternatives alone. For instance, if the leading segmentation obtains ten times the probability of its leading competitor, then the confidence in the segmentation should be high. If the top ten segmentations have approximately the same probability, some other process may have to be performed to further constrain the solution. When a Bayesian formalism is used to generate the probabilities, there is a natural way to combine evidence from multiple models, and this will be seen in Section 4.6.

To derive the framework completely from underlying statistical models, which can be experimentally determined.

A segmentation approach in which necessary parameters can be experimentally determined is favorable to one with parameters that are set arbitrarily. This is particularly true when performance is extremely sensitive to the parameter settings. If the user of the segmentation algorithm is forced to guess at meaningless parameter settings for each image application, the algorithm is difficult to use. If the parameters can be experimentally determined from a set of representative scenes and/or a fixed imaging sensor using some estimation procedure, the approach is advantageous. This trend in segmentation has generated careful study of sensor models (as in [11]), and some abandonment of earlier, primarily heuristic approaches (as in [38]).

To develop a framework capable of handling complex statistical image models (not necessarily those restricted to local dependencies).

We state the handling of complex statistical image models as an objective since we would like to directly model statistical dependencies between fairly distant elements in an image. Such dependency occurs, for example, with a polynomial surface patch model. All of the points on a surface patch are related in a way that is not easily

expressed with very local dependencies. The MRF models, for instance, directly specify very local dependencies between image elements.[3] These concepts are further discussed in Chapter 2.

To build a system capable of estimating the amount of information present in the image under a particular statistical image model.

For a typical application, it is useful to know the degree to which a particular image model is providing information regarding the segmentation. With a probability distribution over segments and segmentations available, a formal measure of information content can be directly quantified. One natural measure is the information entropy, which is a function of a probability distribution. A clear discussion of the characteristics of an entropy measure is provided in [39]. Several alternative, entropy-like functions can be found in [40],[41].

An entropy measure can be used, for example, to select between different models, or to decide to combine several models synergistically. Szeliski argues that a measure of uncertainty can be used to guide search, indicate when more sensing is required, and integrate new information [2]. We presently do not employ entropy measures to guide our algorithms, but in general we can estimate the amount of information present by applying an entropy measure directly to the probability distributions of segments and segmentations.

To develop a framework that applies to the most general images and models possible.

[3]These models can, in general, encode very global dependencies, but the resulting computations quickly become intractable.

It is desirable to have a segmentation approach that applies to the most general images and models possible. Consequently, in our formulations, we attempt to add only a minimal number of constraints to the framework, as the discussion progresses. For instance, this is the motivation for working with general probability spaces in Chapter 4, as opposed to some particular image model. The restrictions we make are based either on necessity or to add clarity to the presentation.

There has been a substantial interest in recent years in unifying segmentation approaches into a common framework [42]. We believe our framework is not yet ready to satisfy this broad goal; however, we do believe that appropriate extensions may be possible to incorporate a wider variety of segmentation models and applications.

To develop a framework that readily supports extensions to incorporate higher-level models.

The Bayesian formalism provides a natural way to combine evidence from several models. In general, a Bayesian approach begins with some prior distribution and some evidence, and yields a posterior distribution. A multiple model approach treats the posterior distribution from one model as the prior distribution for the next model. The second posterior distribution reflects the application of both models. This concept can be applied directly to segment and segmentation distributions, and also to region pairs, as discussed in Section 4.6.

To explain the implications of computationally based simplifications and, when possible, to allow monotonic improvement of accuracy by increasing computation.

When simplifications are made for the purpose of computational efficiency, the resulting performance degradation should be characterized. We should also be able to improve

performance, recovering from some of the degradation caused by simplifications, by accepting more computational cost. Model simplifications are described in this thesis, and with the methods discussed in Chapters 6 and 7, performance can be increased with an increase in computational cost.

To avoid prior specification or estimation of the number of segments in order to perform segmentation.

Often, it is undesirable to specify the number of segments that will be present in some unknown image before performing segmentation. In many approaches to segmentation, the segments are treated as classes, and pattern recognition techniques (such as clustering) are used to define these classes, yielding a favorable segmentation. With these methods, knowledge of the number of classes (i.e., segments) in the image is usually necessary to guide the generation of the segmentation, although a statistically based termination criterion is used in [43].

To design efficient algorithms that are capable of handling large images with realistic probability models.

As a natural goal, we seek algorithms that can cope with images of practical complexity. Enumerating *all* possible segmentation alternatives for a real image is an infeasible task. Therefore it is important to represent only the segmentations having sufficiently large probabilities. Also, we will need an efficient representation of the space of alternative segmentations. Our representation allows us to represent large portions of the space succinctly, and to develop efficient algorithms to construct this representation for a given image.

To design algorithms and representations that allow for straightforward parallelization.

With the current widespread interest in parallel architectures, the issue of parallelizability should be taken seriously when designing segmentation algorithms. When possible, we have developed procedures that are amenable to parallel processing. Specifically, the computational methods introduced in Chapters 6 and 7 are parallelizable in a straightforward manner.

1.3 Organization of the Thesis

Figure 1.1 gives a sketch of the dependencies among chapters in this thesis. Entire chapters are not needed to continue in some cases, and often only minor references are made to previous chapters without a direct dependency indicated. In this section we will provide a brief outline of the material that follows, and indicate the crucial portions of the chapters.

In Chapter 2 the major probabilistic approaches to segmentation, which are more closely related to this work than the general approaches mentioned in Section 1.1, are introduced and discussed. The purpose is to provide a context for our work and a point of comparison. Section 2.1 gives an overview of clustering methods, and presents an example of a segmentation application from Silverman and Cooper [43]. Section 2.2 discusses Markov Random Field (MRF) models (particularly, much of the work by Geman and Geman [30]), and their application to segmentation. Relaxation labeling and related methods are briefly discussed in Section 2.3. Finally, Section 2.4 discusses some of the similarities of the methods, and views them in light of our objectives introduced in Section 1.2.

Chapters 3 and 4 provide the theoretical foundation of our general approach. Chapter 3 deals primarily with the representations and operations needed to construct and maintain

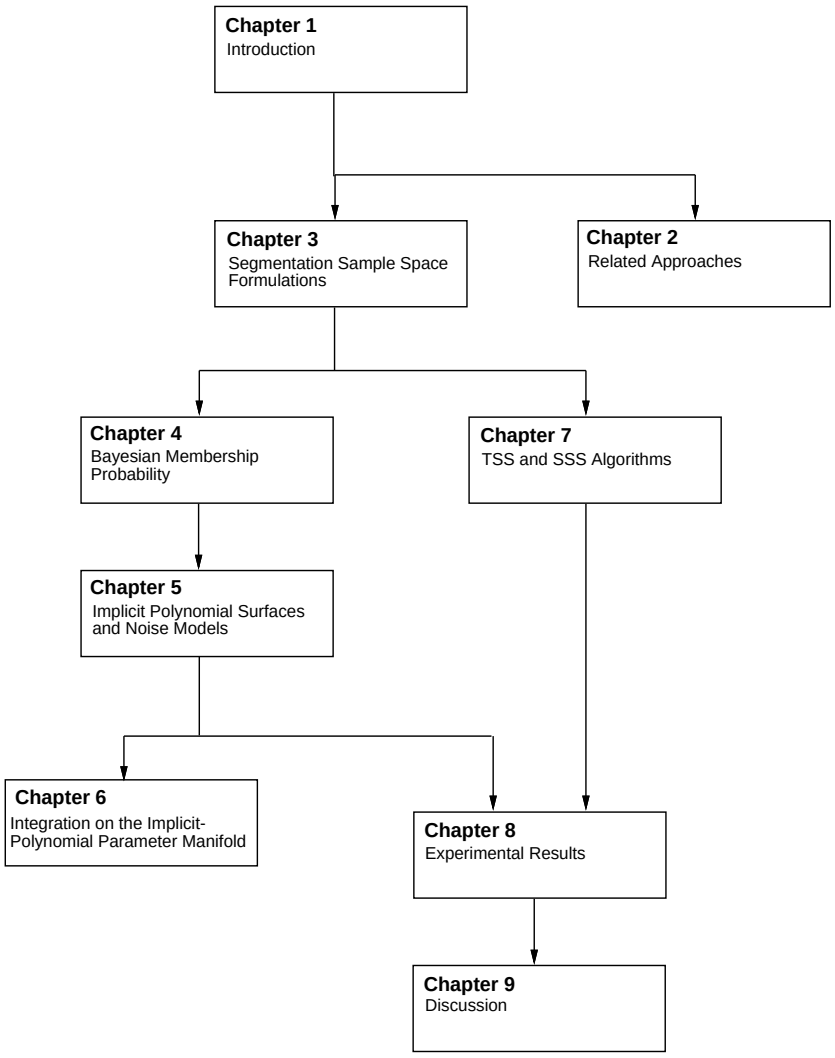


Figure 1.1 Dependencies between chapters in this thesis.

a probability distribution on spaces of segments or segmentations. Chapter 4 gives a general Bayesian model for determining the probability assignments, which applies to a key component of the theory in Chapter 3.

Section 3.1 gives the technical definitions of regions, segments and segmentations. These definitions are important to understand and are slightly different than the usual intuitive notions of these concepts. Sections 3.2 and 3.3 formally define the probability spaces needed for the discussion of distributions of segments and segmentations. An operation called *refinement* is introduced which allows the incremental construction of a

segment distribution representation. A simple example is presented in Section 3.3 illustrating the concepts. Model-based probability assignments are needed for the refinement operation, and the general issues are addressed in Section 3.4. Section 3.5 describes the relationship between segment space distributions and segmentation space distributions, and how a segmentation space representation is constructed. Finally, Section 3.6 discusses the relationships between different types of prior distributions on the spaces of segments and segmentations.

Chapter 4 introduces the membership probability model, which is used at the refinement step. The general statistical model that we begin with has appeared in a similar form in work by Geman [44], and also Szeliski [2], in the context of MRF models. The first three sections of the chapter are the most important. Section 4.1 discusses the general perspective and motivation for Chapter 4. Section 4.2 provides the basic, general definitions of the probability spaces and random variables that will be used in the chapter. Section 4.3 derives the Bayesian membership probability for a useful class of models. Sections 4.4-4.7 describe other models, extensions, and simplifications of the material presented in Section 4.3. Section 4.8 provides an example illustrating the key concepts in the chapter, and Section 4.9 briefly presents a more complex instance of the Bayesian membership probability model, in an application to an MRF texture model.

Chapter 5 demonstrates the utility of the Bayesian membership model by applying it to an important class of models used in range image segmentation: the implicit polynomial surfaces. These are generally difficult to utilize since parameter estimation is often difficult, and closed-form, point-to-surface distance expressions do not exist. An introduction to implicit surfaces, including general background and motivation, is given in Section 5.1. Section 5.2 presents an example motivating the use of our membership model in this context and compares it to an estimation-based approach to segmentation. The application of the membership model from Chapter 4 is described in detail in

Sections 5.3 and 5.4. Specific expressions for planar and quadric surface cases are given in Section 5.5.

Chapter 6 discusses the computational implications of determining the membership probability under the implicit surface model. In general, the computation requires an integration to be performed over a half-hypersphere manifold. The parameterization equations for this manifold are presented in Section 6.1. Closed-form expressions for hypersphere area, required in the computation, are presented in Section 6.2. Section 6.3 describes the Monte-Carlo technique used to obtain numerical estimates of the integrals. The convergence rate of this integration method is slow for many cases, and we present a technique which greatly reduces the number of iterations required to compute the integrals in Section 6.4. A simplified method for computing the membership probability, applicable in some cases, is presented in Section 6.5. The chapter concludes with some additional implementation issues in Section 6.6.

Chapter 7 specifies the algorithms used to construct representations of distributions of segments and segmentations, and may be considered as a continuation of the discussion presented in Chapter 3. In Chapter 3, the representations and operations are specified, and in Chapter 7, the algorithms and decisions used in our experiments are presented. These algorithms are primarily concerned with the efficiency issues involved in building the distribution representation. Section 7.1 presents algorithms for generating a representation and probability distribution over a set of segments. Section 7.2 presents algorithms for generating a single segmentation, or a probability distribution over a set of segmentations.

Experimental results using planar and quadric surface models with range images are given in Chapter 8. The experimentation can be divided into three major categories:

1. Membership probability results for varying region types

2. Obtaining a probability distribution on a space of segments

3. Obtaining a probability distribution on a space of segmentations

The experimental objectives and conclusions are introduced in Section 8.1. The first category, presented in Section 8.2, demonstrates the concepts introduced in Chapters 4-6. Specifically, the effects of increased noise, location of points in space, and small sample sizes on membership probability are discussed. Sections 8.3 and 8.4, representing the second and third categories, demonstrate the algorithms of Chapter 7, and also the theory presented in Chapter 3. Probability distributions over segments and segmentations are obtained by repeatedly using membership probability computations. Since it is important to control the experiments in a way that allows us to declare the underlying model, vary the noise, and make other deviations for the purpose of observation, some of the images presented are synthetic. We present many results for real range images, in which the data set is obtained through active triangulation.

Section 9.1 of Chapter 9 recalls the objectives introduced in Section 1.2, and summarizes how they have been addressed in the thesis. Section 9.2 presents several directions for future research. A few concluding remarks are given in Section 9.3.

CHAPTER 2

RELATED APPROACHES

The purpose of this chapter is to present other probabilistically based methods of image segmentation to provide a context and point of comparison to our work. These approaches can be organized into several categories, depending on the general method and models used. Section 2.1 describes the statistical clustering approach. The clustering methods are taken from general pattern recognition, in which features are extracted from the image and classification is made in the feature space, yielding an image segmentation. Section 2.2 discusses the MRF-based approach to segmentation, in which a spatially dependent, stochastic model is used, and the segmentation goal is formulated as an energy minimization problem. Relaxation labeling, discussed in Section 2.3, also models spatial dependency and often uses a deterministic, iterative algorithm to yield a locally optimal segmentation. This categorization is not completely rigid since several researchers combine approaches. Finally, in Section 2.4, our segmentation goals are discussed in comparison to the previous probabilistic segmentation work. Specifically, we point out that all of these methods involve model parameter estimation and iterative convergence toward a single, goal segmentation.

2.1 Clustering Methods

2.1.1 General description

Many of the clustering methods are derived from general pattern recognition research. The problem in pattern recognition occurs when one wants to design a classifier, and the class labels of the training set are unknown. This is known as the *unsupervised learning* problem. It may also be the case that the number of classes is not known, although most clustering algorithms require this number. It is, however, assumed that some functional form of the class-conditional densities is known (e.g., that the densities are multivariate Gaussian).

By considering segments as classes, clustering can be applied to the segmentation problem. For texture models it may be the case that a labeled training set is available from which texture parameter values can be obtained; however, for most applications, this is not the case. The clustering approach is naturally suited to the polynomial surface model, in which each class represents a single polynomial patch in the image. In this case it is difficult to estimate in advance the number of surface patches one expects to appear in the image.

Since the class densities are unknown and the training set is not labeled, the clustering approach is not inherently probabilistic; however, we consider it in this categorization since clustering often incorporates parameter estimation and statistical decisions. It is also common for clustering to be used as a single stage in a probabilistic segmentation approach, which could also, for instance, incorporate MRF methods (see [33],[45]).

A general introduction to clustering can be found in [46],[47], and only a brief introduction is provided here. Four basic components involved in most clustering algorithms are:

1. Define a feature metric space.

2. Determine feature values that correspond to pixels or regions.

3. Iteratively group pixels or regions with close features in the metric space.

4. Terminate based on some stopping criterion (if the number of classes is unknown).

The first step corresponds to choosing some image features to use for clustering. Features could correspond to single pixels, or could be extracted from regions in the image. For instance, these features could be texture parameters, coefficients of a polynomial surface that approximates some region of the image, or intensity values. It is expected that pixels or regions with close feature values (in which a distance metric on the feature space is used to measure closeness) are similar and should belong to the same segment. Consequently, locally clustered feature values can be compared and grouped. One of the useful aspects of the clustering approach is its natural ability to combine evidence from multiple sources by concatenating feature vectors [48].

The second step may involve some straightforward transformation of the data into feature space, as in the color space used in [49], or may use some region-based parameter estimation. For example, Silverman and Cooper use maximum likelihood estimation to obtain surface parameter vector estimates for clustering [43].

The decisions involved in the third step depend on the particular clustering algorithm chosen. Among these are K-means clustering [50],[51] and agglomerative clustering [46],[43], which is discussed in the next section. Some algorithms also allow partitioning of the clusters [47].

The fourth step is necessary when the number of classes is not known. Without this knowledge, it could conceivably be correct to leave each data point in a class by itself, or to group them all together. Some decision criterion is therefore necessary to determine the number of resulting classes. This decision criterion is often based on some

assumptions about the general form of the class-conditional densities (e.g., multivariate Gaussian).

Clustering has been applied to a variety of image types and models. Silverman and Cooper [43] segment intensity images into regions in which the intensities can be well-approximated by planar or quadric surfaces. Bell combines clustering with a Monte-Carlo approach to segment radiograph images, to determine manufacturing defects [52].

As stated previously, most clustering algorithms require specification of the number of classes. Some recent work has been done specifically addressing the problem of determining the number of classes, known as *cluster validation*, in the context of image segmentation applications. Zhang and Modestino give a discussion of cluster validation that applies Akaike's information criterion to deciding the number of classes a priori [51]. Jolion et al. propose a robust clustering algorithm which provides reliable performance without specifying the number of clusters [48].

Histogram thresholding is related to clustering, and a general survey of thresholding techniques can be found in [53]. An intensity image can be transformed into a histogram space, and the goal is to use thresholding on groups of pixels or regions, represented in the histogram, which are closely concentrated. This is equivalent to deciding on class boundaries, in the pattern recognition context. Since necessary statistical models are difficult to obtain, the thresholding methods are often heuristic [13].

2.1.2 A clustering example

As an example of the clustering approach, applied to segmentation of intensity images with explicit planar and quadric polynomial models, we will discuss some details of the work of Silverman and Cooper [43]. Later, in Chapter 5, we will apply our model to the implicit polynomial facet model for range images; therefore, this clustering application is useful for comparison.

Silverman and Cooper applied the agglomerative hierarchical clustering algorithm [46] to the problem of intensity image segmentation. The assumption is made that the image intensities can be represented by piecewise quadric (or planar) surfaces with additive white noise. Certain texture models can be used instead, but these will not be treated here.

Initially, the image is partitioned into a rectangular array of square regions, called *blocks*. By iteratively grouping blocks to create larger regions, called *clusters*, an image segmentation is obtained in which the intensities in each agglomerate region are well-approximated by a single quadric (or planar) equation.

We will describe some of the mathematics next, with some of the notation altered from [43] to conform to the notation that we will introduce in subsequent chapters. Let $D[i,j]$ denote the intensity value at the i^{th} row and j^{th} column of a rectangular image, and let D denote the entire image. A quadric intensity function is represented as

$$h(\mathbf{u};i,j) = u_0 + u_1 i + u_2 j + u_3 ij + u_4 i^2 + u_5 j^2. \tag{2.1}$$

Let the k^{th} block be denoted by R_k, which is simply a set of intensity values at the indexed locations. For any block (or later any set of blocks), a parameter $\hat{\mathbf{u}}_\mathbf{k}$ can be chosen to yield a best approximation to the set of intensity values in R_k in the least-squares sense, as described in [46].

The noise model consists of Gaussian iid noise with known variance σ_k for block R_k. The joint probability of all of the intensity values, given the polynomial coefficients $\mathbf{u}_\mathbf{k}$ for every block R_k, is used as a global criterion to choose clusters for merging. This joint probability density function (pdf) for the data set is a product of all of the Gaussian densities,[1]

$$p(D|R_k, \mathbf{u}_\mathbf{k}, \sigma_k\ \forall k) = \prod_{k=1}^{K} \prod_{D[i,j]\in R_k} \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left\{-\frac{1}{2}\left[\frac{D[i,j] - h(\mathbf{u}_\mathbf{k};i,j)}{\sigma_k}\right]^2\right\}. \tag{2.2}$$

[1]In the condition, we are given all region data, region parameters, and noise variances.

1. Initialize blocks R_k and determine parameter estimates $\hat{\mathbf{u}}_\mathbf{k}$.
2. Out of all adjacent cluster pairs select two clusters, e.g., R_{m_1} and R_{m_2}, which have the closest parameter values.
3. Replace R_{m_1} and R_{m_2} with $R_{m_1} \cup R_{m_2}$, and compute $\hat{\mathbf{u}}_{\mathbf{m_1},\mathbf{m_2}}$.
4. If termination criterion met, then terminate.
5. Go to 2.

Figure 2.1 The agglomerative clustering algorithm.

The inside product can be multiplied out to obtain

$$p(D|R_k, \mathbf{u_k}, \sigma_k \; \forall k) = \prod_{k=1}^{K} \frac{1}{(2\pi\sigma_k^2)^{\frac{|R_k|)}{2}}} exp\left\{ -\frac{1}{2} \sum_{D[i,j] \in R_k} \left[\frac{D[i,j] - h(\mathbf{u_k}; i, j)}{\sigma_k} \right]^2 \right\}, \quad (2.3)$$

in which $|R_k|$ represents the number of points in R_k. For any cluster R_k, we can obtain an estimate of $\mathbf{u_k}$, denoted by $\hat{\mathbf{u}}_\mathbf{k}$. Also, for any pair of clusters R_r and R_s, we can form a new cluster $R_{rs} = R_r \cup R_s$, and obtain a corresponding parameter estimate $\hat{\mathbf{u}}_\mathbf{rs}$. The agglomerative clustering algorithm is described in Figure 2.1.

The expression (2.3) is used to decide the most likely pair to merge. A likelihood ratio is formed for each pair, using parameter estimates. This ratio compares the likelihood (2.2) with the pair of regions merged to the likelihood without the pair merged. The pair of regions that maximizes this ratio is selected for merging. Explicitly, the ratio is

$$\frac{p(R_{rs}|\hat{\mathbf{u}}_\mathbf{rs}) \prod_{k \neq r,s} p(R_k|\hat{\mathbf{u}}_\mathbf{k})}{\prod_k p(R_k|\hat{\mathbf{u}}_\mathbf{k})} = \frac{p(R_{rs}|\hat{\mathbf{u}}_\mathbf{rs})}{p(R_r|\hat{\mathbf{u}}_\mathbf{r}) p(R_s|\hat{\mathbf{u}}_\mathbf{s})}. \quad (2.4)$$

The densities of the form $p(R_k|\hat{\mathbf{u}}_\mathbf{k})$ represent the joint pdf of R_k, given the parameter estimate and observed intensities, taken directly from (2.3).

We omit some details here, but the maximum likelihood criterion above can be expressed as a minimization of a quadratic distance measure,

$$(\hat{\mathbf{u}}_\mathbf{r} - \hat{\mathbf{u}}_\mathbf{rs})^t M_r (\hat{\mathbf{u}}_\mathbf{r} - \hat{\mathbf{u}}_\mathbf{rs}) + (\hat{\mathbf{u}}_\mathbf{s} - \hat{\mathbf{u}}_\mathbf{rs})^t M_s (\hat{\mathbf{u}}_\mathbf{s} - \hat{\mathbf{u}}_\mathbf{rs}). \quad (2.5)$$

The matrices M_r and M_s are computed directly from the region data in R_r and R_s. Each term in the sum is similar to a Mahalanobis distance metric in the coefficient space [46]. The distance function is minimized as the difference between $\hat{\mathbf{u}}_{\mathbf{rs}}$ and $\hat{\mathbf{u}}_{\mathbf{r}}$ and the difference between $\hat{\mathbf{u}}_{\mathbf{rs}}$ and $\hat{\mathbf{u}}_{\mathbf{s}}$ become small. In other words, if the polynomial coefficients are approximately equal, the clusters should be merged.

The clustering iterations continue until a stopping criterion is met that asserts that even the best possible merge is likely to be incorrect. The condition is formulated as a Bayesian, two-hypothesis test, based on the parameter likelihoods.

2.2 MRF/Gibbs Methods

2.2.1 General introduction

The Markov random field (MRF) approach models the image as a lattice of random variables, with each variable having explicit dependency on some local neighborhood consisting of other random variables. One of the earliest instances of this type of system is the Ising model [54], used in statistical physics to explain ferromagnetism. The general model that is most often used in computer vision was introduced in a seminal paper by Geman and Geman [30], in the context of image restoration. A primary appeal of the approach is the fact that any MRF formulation (which applies to a variety of image models) can be expressed as an energy minimization problem, in which parallelism can be exploited. The primary difficulties with the approach are the computational complexity of the optimization, and the problem of MRF parameter estimation [55]. An extension of the MRF approach, which encompasses many continuous field and deterministic segmentation and boundary detection methods, can be found in [42],[56]. For a discussion of uncertainty estimation with MRF models, see [2].

The MRF approach is applied primarily to intensity images, and the statistical models capture local dependency relationships between pixels. The approach has been applied to modeling noise processes and texture [21],[22],[23], color-constancy [57], blurring [30],[58], boundary modeling [59],[58], and locally dependent nonlinear image transformations [30]. The approach is less appropriate for modeling more globally dependent models such as piecewise polynomial surfaces [2].

2.2.2 Model definitions and the Hammersley-Clifford Theorem

Some details of the approach will now be presented. For more complete introductions, we refer the reader to [60],[30],[61]. A set of *sites* is denoted by $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$. Each site has a discrete-valued random variable associated with it, denoted by X_s for site s. The dependencies between the sites are modeled by a *neighborhood system* $\mathcal{G}$. Each site, s, contains a set of sites G_s in $\mathcal{G}$ with the restrictions that

1. $s \notin G_s$,

2. $s \in G_r$ if and only if $r \in G_s$.

Each G_s represents a set of local neighbors for the site s. The restrictions define the neighborhoods in an intuitive way, and the the pair $\{\mathcal{S}, \mathcal{G}\}$ together define an undirected graph; the sites are the nodes, and each neighborhood gives a set of arcs which are drawn between s and the neighboring sites in G_s. A subset $C \subseteq S$ is called a *clique* if each site in C is a neighbor of every other site in C.

Typically, two coupled random fields are used to model an intensity image: the *intensity process*, F, and the *line process*, L. Together, these constitute the sites: $S = F \cup L$. The intensity process represents the "true" intensity values of the image pixels. The line process represents boundaries between the individual pixels. For any two adjacent pixels in the image (using standard four-neighbors), there is a site representing a line segment

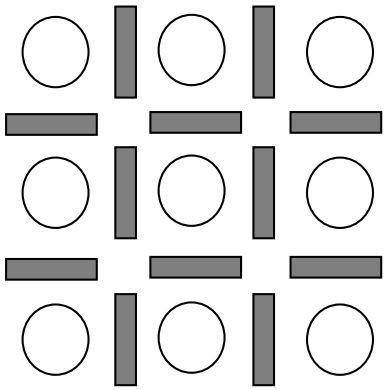


Figure 2.2 Depiction of the coupled MRF processes. Circles represent the intensity process, and shaded bars represent the line process.

between them. Figure 2.2 depicts the two interleaved processes with circles for the intensity process and shaded bars for the line process. Geiger and Girosi discuss the difficulties with bias induced from vertical and horizontal lines [55]. Other types of information have been represented with MRF variables. For example, Chou and Brown represented region and edge type labelings [36], and Modestino and Zhang consider high-level, qualitative labelings [62].

Let each random variable, X_s, take on values from $\Lambda = \{0, 1, \ldots, L-1\}$ for some integer L. For the coupled process, the line and intensity sites will take on values from different sets; however, this distinction is merely notational and does not hinder the following discussion. This gives rise to a sample space on the sites with each outcome represented by assigning values from Λ to each of the sites. The set of all outcomes is

$$\Omega = \{x = (x_{s_1}, \ldots, x_{s_N}) : x_{s_i} \in \Lambda, 1 \leq i \leq N\}. \tag{2.6}$$

The event corresponding to a single outcome is denoted by $\{X = x\}$.

Two conditions are necessary for the system of random variables X to be an MRF

$$P(X = x) > 0 \quad \forall x \in \Omega \tag{2.7}$$

and

$$P(X_s = x_s | X_r = x_r \; \forall r \neq s) = P(X_s = x_s | X_r = x_r \; \forall r \in G_s). \tag{2.8}$$

The first condition simply asserts that every outcome must have nonzero probability. The second condition states that the probability of values at a site s depends only on the site variables that are neighbors of s. The other sites in the image have no direct influence on the probability distribution of X_s. There is an implicit global dependency throughout the sites since the probabilities of the neighboring sites depend on *their* neighbors. When considering the outcomes x, two sites that are connected in $\{\mathcal{S}, \mathcal{G}\}$ have dependent probabilities through transitivity of the neighborhood dependencies. This is a natural extension of the concept of a Markov chain, which has a linear dependency, to a general, graph-structured dependency. It is often assumed that the system is noncausal, and the entire image is presented at once. Goutsias and Mendel have considered the use of causal random field models for segmentation [63].

Consider a probability mapping π on Ω. It conveniently turns out that the two MRF conditions above impose a strong restriction on the form of valid probability distributions on Ω. The following theorem is obtained, which is clearly explained and proved in [60]:

Theorem 1 *(Hammersley-Clifford Theorem) A set of random variables X with the neighborhood graph $\{\mathcal{S}, \mathcal{G}\}$ is an MRF if and only if there exists a function V such that*

$$\pi(x) = \frac{e^{-\frac{1}{T}U(x)}}{Z} \quad \forall x \in \Omega, \tag{2.9}$$

in which T and Z are constants, and

$$U(x) = \sum_{C \in \mathbf{C}} V_c(x). \tag{2.10}$$

In (2.9), $\mathbf{C}$ represents the set of all cliques in $\{\mathcal{S}, \mathcal{G}\}$, and T denotes "temperature," representing the degree of peaking in the distribution. As T approaches infinity, the distribution $\pi(x)$ approaches uniformity. The normalizing constant, Z, is often called

the *partition* function. The exponential function given by (2.9) and (2.10) is called a Gibbs distribution. The function U represents the *energy function*. Note that smaller values for the energy function correspond to larger probabilities $\pi(x)$. Each function V_c is called a *potential*.

The potential functions are specified as part of the particular model. There is one function for each clique, meaning that some potential function exists for each fully connected grouping of sites. Hence, there are potentials for singletons, pairs, triplets, and so forth. A larger potential value penalizes outcomes of certain combinations of variables, while smaller potential encourages certain combinations. Determining these potential functions (the estimation problem) can often be difficult, depending on the application.

The goal is to find the labeling, x, that minimizes the posterior energy function, given a prior energy function, and an observation of the degraded image. The general approach can be summarized as follows:

1. Define a prior MRF and potentials.

2. Determine the form of the posterior probability map from models and the image observation.

3. Find the best image estimate through optimization of an energy function.

2.2.3 An MRF example

We present an example which illustrates the combination of priors and the degradation model into a posterior MRF, discussed in [44]. Consider a set of sites representing the intensity pixels in a rectangular array image (i.e., $L = \emptyset$ and $S = F$). The sites represent binary intensity values, being zero or one. Define the neighborhood set of a pixel to be

the four-neighbor set (above, below, left, right neighboring pixels).[2] Let y denote the observed intensity site values that comprise the given image.

Using this neighborhood model, the only cliques in $\{\mathcal{S}, \mathcal{G}\}$ are singletons and pairs. Let singletons have zero potential and let all pairs have identical potential. Using a constant $\beta > 0$, a consistent potential function can be defined as

$$U(x) = -\beta \sum_{[s,t] \in \mathbf{C}} x_s x_t. \tag{2.11}$$

Larger values of β will encourage larger regions of constant intensity since the energy for cliques with 1-1 intensity values will be lower than for mixed, 0-1 and 1-0 values. Using this prior energy function, the prior probability is

$$\pi(x) = \frac{1}{Z} exp\left\{ \beta \sum_{[s,t] \in \mathbf{C}} x_s x_t \right\}. \tag{2.12}$$

Assume that the only contributing factor to the degradation of the image is iid Gaussian, additive noise with zero mean and variance σ^2. The conditional probability of the observed image, given the true image is

$$\pi(y|x) = \left(\frac{1}{2\pi\sigma^2} \right)^{|S|/2} exp\left\{ -\frac{1}{2\sigma^2} \sum_{s \in S} (y_s - x_s)^2 \right\}. \tag{2.13}$$

This form is obtained from expanding a product of Gaussians representing the joint site density.

Bayes' rule is used to obtain the posterior probability

$$\pi(x|y) = \frac{\pi(y|x)\pi(x)}{\pi(y)}. \tag{2.14}$$

The denominator, $\pi(y)$, is very difficult to compute, but is usually not directly needed, and therefore treated as a normalization factor. The posterior becomes

$$\pi(x|y) = \frac{1}{Z_p} exp\left\{ \beta \sum_{[s,t]} x_s x_t - \frac{1}{2\sigma^2} \sum_{s \in S} (y_s - x_s)^2 \right\}. \tag{2.15}$$

[2]This is known as the Ising model.

Above, Z_p represents the combined constant from Z and $\pi(y)$. The argument of the *exp* function is again an energy function, and by Theorem 1 the posterior distribution is also an MRF. This is an example of a posterior energy which is minimized to obtain a good labeling.

2.2.4 Energy optimization and parameter estimation

In recent years here has been considerable interest in improving MRF energy optimization algorithms. Geman and Geman used a simulated annealing approach (also called stochastic relaxation) to determine the maximum a posteriori estimate (MAP) of the image [30]. The temperature, T, is controlled in a manner that guarantees convergence to the optimal energy state, but the rate of convergence can be slow in practice. Other techniques have been developed which yield performance tradeoffs. Besag [64] proposed the iterative conditional modes (ICM) method as a feasible alternative to stochastic relaxation. Marroquin et al. gave an approach, called maximizer of posterior marginals (MPM), which defines a segmentation error metric, minimized to yield the best labeling. Empirical comparison of these three approaches was done in [65], concluding that for many cases, ICM was the most efficient, robust and produced the most reasonable segmentations.

Derin and Elliot derive a recursive formulation of the posterior energy function and propose the dynamic programming formalism to determine the MAP estimate [22]. Due to high computational complexity, an iterative, suboptimal approximation is used, which sequentially processes strips in the image. Cohen and Cooper give a parallel, hierarchical algorithm for optimizing the energy function in a Gaussian MRF [21].

An interesting approach to the minimization is taken by Chou and Brown [36]. The site values are arranged in a hierarchy, in which various nodes represent subsets of the set of labelings that can be assigned. Using this representation, a highest confidence first

(HCF) method is developed for efficient energy minimization with least commitment under inaccurate models.

The issues involved in MRF parameter estimation have also been carefully considered. In all of the MRF algorithms, some model parameter estimation must be performed to obtain the energy function. Often, methods perform parameter estimation "off-line" as a preprocessing stage to segmentation [22]. Cohen and Cooper discuss the problem of adaptively estimating parameters during segmentation in the context of texture models [21]. Subrahmonia et al. propose an iterative scheme which performs global optimization of the energy function and parameter estimation for 3-D surfaces through a single performance functional [66]. An alternative adaptive estimation/optimization scheme has been proposed by Lakshmanan and Derin [45]. Manjunath and Chellappa [23] argue that for texture models, estimation on small windows and simple, nearest-neighbor clustering can be used as a starting point for the energy optimization to yield results comparable to the adaptive scheme [45].

2.3 Relaxation Labeling

The relaxation labeling approach (also called probabilistic relaxation) has many similarities to the MRF approach, but has been less popular in recent times. One of the early appearances of relaxation labeling is in work by Rosenfeld, Hummel, and Zucker [67]. We leave a complete introduction to the subject to Peleg [68], and Rosenfeld and Kak [69], and instead mention some of its features and issues.

Like the MRF approach, a probability distribution for each site over a set of labels is considered. A *compatibility* measure is used to model the interaction between pairs of sites, when determining probability assignments. One begins with a prior distribution of labelings. Through an assignment rule based on compatibility, one iteratively improves estimates of the "true" probabilities. After several iterations, the probability assignments

stabilize. For most versions of relaxation labeling, the probabilities are merely syntactic (in the sense that they obey the axioms of probability) and are not derived from a statistically valid image model [30],[69].

Peleg provides a rigorous probability updating rule, which explicitly defines quantities (as individual labeling probabilities and pairwise, joint probabilities) needed for building the models. Some heuristic estimation is often required to determine these probabilities. For this estimation in the context of shape analysis, see [70].

It has been difficult to characterize the quality of the obtained labeling. Faugeras and Berthod, however, defined measures of consistency and ambiguity in the distribution of labelings, and gave an iterative optimization approach to obtaining a good labeling [71]. Some extensions of the relaxation labeling paradigm are presented in [72]. A more recent discussion of relaxation labeling, applied to the problem of supervised and unsupervised texture classification can be found in [24],[25].

Don and Fu propose a segmentation method, similar to relaxation labeling, using a stochastic grammar [73]. Compatibility measures for pairs of image features are derived from stochastic rules in the grammar. An iterative optimization procedure is derived, with proven convergence, that yields a favorable segmentation. Some experimental results were obtained with intensity images, and hueristically assigned probability rules.

2.4 Comparative Discussion

Using the methods discussed, we can further address some of the objectives that were stated in Section 1.2. Most probabilistic approaches involve three steps:

1. Parameter estimation

2. Iterative decision making

3. Classification

These are similar to the steps indicated by Jeong and Lapsa: parameter determination, decision, and classification [74].

For clustering, parameter estimation is performed on regions or pixels to obtain values in the feature space. The clustering algorithm involves iterations of decision-based criteria and often a final termination decision. The classification is given by the resulting groupings of features in the feature/parameter space.

With the MRF model, parameter estimation is needed to determine the energy function. Some form of iterative optimization is generally used to determine the posterior distribution on the space of image labelings. A classification is induced by considering each possible label as defining a class, and the resulting site values as defining the class assignments.

With relaxation labeling, estimation (if it is performed) determines the compatibility measures. An iterative optimization is performed to yield a stable label assignment. Again, if the labels are considered as classes, a classification has been performed.

Our framework does not fit into these three steps. With our approach, parameter estimation can be completely avoided (however, the density in the parameter space is still utilized). We will introduce iterative methods, but they will not be working toward a single goal segmentation or labeling, but rather toward maintaining a distribution of segments or segmentations. Uncertain decisions can be left to some higher-level process.

In the remainder of the section, we evaluate the methods presented in this chapter in the context defined by our objectives in Section 1.2.

None of the methods that we have seen have attempted to represent a distribution over a space of segments or segmentations. The MRF approach does, however, explicitly define the space of all labelings, on which the optimization is performed. Dubes et al. have noticed through experimental observation that the optimal energy in an MRF formulation does not necessarily correspond to the "correct" segmentation [65]. Szeliski has

emphasized the importance of being able to estimate the amount of uncertainty present in a segmentation [2]. Due to the underconstrained nature of segmentation, we consider it appropriate to represent distributions of alternative segments and segmentations.

Methods that allow for statistical parameter estimation, such as MLE, have been preferred recently over those having many arbitrary parameters. This is one of the difficulties with many relaxation labeling applications, causing research to focus toward MRF-based methods and clustering. If there is no sound method to determine segmentation parameters reliably, the method becomes difficult to use. As problems of greater complexity are considered, the number of parameters tends to increase, making this issue increasingly important. Hence, we consider deriving all parameters from statistical models as one of our objectives.

Recall that the MRF approach is most naturally suited for locally dependent models (although a wide variety of them). We are interested in the application of a Bayesian model to polynomial surfaces, and hence the MRF formalism is not suitable. However, clustering is often used for polynomial surface models. Also, Jeong and Lapsa combine a polynomial model with the MRF model in a general decision criterion, used for segmentation [74].

Since the MRF approach is Bayesian, there is a natural extension to the case of multiple models, since it corresponds to repeated application of Bayes's rule [44]. Multiple models can also be incorporated into the clustering approach since feature spaces from different models can be concatenated. Hence, we also consider this as an objective, and in Chapter 4, we demonstrate how our Bayesian model can accomplish this.

Increased performance at the expense of increased computation can be beneficial if the tradeoff can be sufficiently characterized. This is the case for some MRF optimization algorithms. For instance, by adjusting the temperature, T, in the simulated annealing procedure, the convergence rate is increased at the expense of failing to guarantee an

optimal solution. Hence, one is allowed to trade off computation for accuracy, depending on the application. We also consider this as a useful attribute for a segmentation approach, and develop algorithms that behave in this manner.

The number of classes in the MRF approach is supplied as the number of labelings for a site. This does not directly relate, however, to the number of segments in a segmentation determined by the line sites. For this case it is a matter of interpretation as to whether or not the number of classes is given. For clustering, we have discussed the difficulty involved in determining the number of classes a priori, or formulating a termination criterion for ending the merging. This is the reason that clustering algorithms tend to oversegment large patches [75]. For texture models, it may be reasonable to expect some number of classes; however, for polynomial patch models, it does not seem reasonable to expect to know the number of patches in the image a priori. Hence we present a method which does not require the number of classes to be given.

Parallelizability has been a feature making many segmentation approaches popular. For instance, the local-dependency structure of the MRF allows for straightforward parallelization of many of the optimization algorithms [30],[2]. Don and Fu have also emphasized parallelism in their stochastic grammar optimization approach to segmentation [73].

CHAPTER 3

SEGMENTATION SAMPLE SPACE FORMULATIONS

As mentioned in Chapter 1, one of the keys to our approach is the ability to efficiently generate and represent the interesting portions of the sample space of segmentations. This chapter defines the general segmentation concepts needed for our approach, and gives the explicit mechanisms for constructing the sample space representation. Generating the representation will require probability assignments to be made at various stages, and in the next chapter we will present a Bayesian model which is used to exert evidence-based probabilities on the space of segmentations. This chapter will define precisely where this Bayesian model is applied, while leaving its details to Chapter 4. Although the method of constructing the sample space representation will be given in this chapter, the issue of making computationally based decisions will be deferred until Chapter 7, when algorithm details are discussed. Many of the concepts presented in this chapter also appear in [76].

While much of the past segmentation research has focused on obtaining a single, best segmentation with respect to image models, our approach will require carefully chosen representations of subsets of the space of segmentations. We will consider a segmentation here as a partition of the image into connected segments, and not as a set of possible labelings for image pixels as in many MRF methods. With a probabilistic model of segmentation, the favorability of a partition is influenced by the underlying

statistical model of the homogeneity. With the approaches discussed in Chapter 2, the goal is to iteratively converge to a single partition, optimizing (or attempting to optimize) some statistical criterion. The following sections organize the segmentation problem into a sound, probabilistic framework in which entire sets of partitions are implicitly represented. Evidence induced by the models of homogeneity is explicitly applied to these sets of partitions to determine posterior Bayesian probabilities on the space of all partitions. In this way, interesting portions of the space of all partitions can be represented, along with the corresponding probabilities.

Two different sample spaces, which describe these sets of segmentations, will be introduced. The segmentation sample space (SSS) describes the probability distribution over all possible segmentations. The segment sample space (TSS) is used to maintain a probability distribution over a space of single segments in the image. During the process of constructing a representation of the SSS, different TSSs are generated, with each one contributing to the representation of the SSS. The two spaces are fundamentally related through a set mapping.

In the next section, some of the basic definitions needed to discuss segmentations will be provided. In Section 3.2 the notion of a sample space of segmentations is introduced. Section 3.3 describes the segment sample space, which is used to apply evidence locally and construct representations of the space of segmentations. Following this, Section 3.4 introduces the general issues involved in applying model-based evidence to the construction of the segmentation space representation. This topic will be discussed in much further detail in Chapter 4. Section 3.5 gives the fundamental relationship between the segment sample space and the space of all segmentations. The final section presents the issue of obtaining prior distributions on the space of segmentations, which is usually of interest in a Bayesian approach. The definitions from this chapter are presented in Table 3.5.

3.1 Regions, Segments and Segmentations

The definitions and assumptions that follow treat the segmentation problem as that of finding a partition of an image which consists of a set of points, between which some adjacency relation is defined. Only those partitions that lead to connected groups of points are considered. The structure of these groupings will be precisely defined, leading to a sample space of segmentations in the probabilistic sense.

Before considering probabilistic formulations, it is necessary to consistently define familiar segmentation ideas. The input to the segmentation system is a set D of points. Associated with each element $D[i,j]$ is the representation of the point, which may be an intensity value and/or a set of coordinates in $\Re^3$, or may be other image information. In Chapter 5, when polynomial models are introduced, $D[i,j]$ will represent a point in $\Re^3$, with coordinates $\mathbf{x} = [x_1\ x_2\ x_3]$. The theory in Chapter 4 will not, however, require this restriction.

Since the elements of D are arranged in a matrix, adjacencies can be considered in the usual way. A given point $D[i,j]$ will have a set of points called *neighbors* to which it is *adjacent*. Using standard four-neighbors, this set is: $D[i-1,j]$, $D[i+1,j]$, $D[i,j-1]$, $D[i,j+1]$ (see Figure 3.1). One could also consider eight-neighbors by considering diagonally related points as adjacent. In general, for the theory to come, any adjacency relation may be used for a given set of data points.

We will introduce three terms that are used extensively in this thesis: regions, segments, and segmentations. Usually, a segmentation is considered as a partition of the image D; however, in our work, we will need to introduce some additional structure when defining a segmentation. The definitions are first introduced, along with some interpretations. In Section 3.3 an example is presented which illustrates the definitions introduced here, and subsequent concepts.

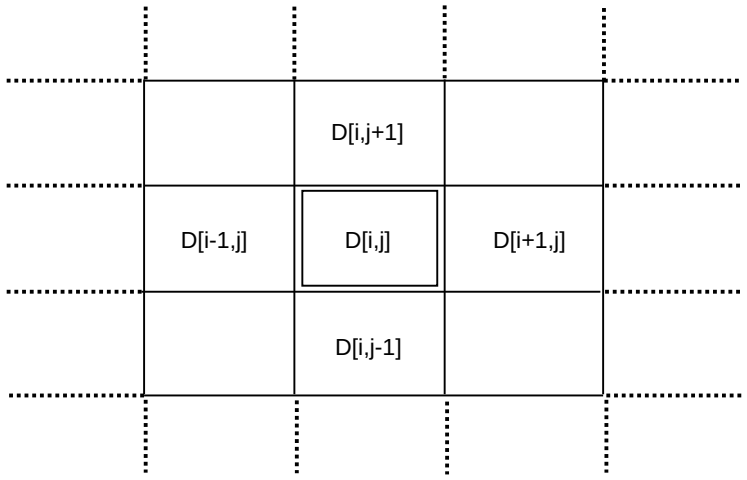


Figure 3.1 Simple, 4-neighbor adjacency for point $D[i,j]$ in the image D.

A *region*, R, is some connected subset of D.[1] By *connected* we mean that for any $D[i_1, j_1]$, $D[i_2, j_2] \in R$, there exists some sequence of elements of R, $\langle D[k_m, l_m] \rangle_{m=1,\ldots,n}$, in which each $D[k_{m-1}, l_{m-1}]$ is adjacent to $D[k_m, l_m]$, and $D[k_1, l_1] = D[i_1, j_1]$, $D[k_n, l_n] = D[i_2, j_2]$. This is the standard definition of a connected region for image segmentation. Two regions, R_1 and R_2, will be called *adjacent* if there exist some $D[i_1, j_1] \in R_1$ and $D[i_2, j_2] \in R_2$ that are adjacent.

For a given problem, we work with a pairwise-disjoint set of regions, $\mathcal{R}$, in which every element of D is contained in some region. We will soon describe how these regions are obtained, but first we continue with the related definitions.

A *segment*, T, is a connected set of regions (e.g., $S = \{R_1, R_2, R_3\}$ is a segment consisting of three regions). A set of regions is *connected* if their union is connected, in the sense as defined above.

A *segmentation*, S, denotes a set of segments that forms a partition of $\mathcal{R}$. Note that a segmentation implicitly defines a partition of D.

[1] Note the table of definitions appearing at the end of this chapter.

Note that given a segmentation, S, and two adjacent segments[2] $T_1, T_2 \in S$, a new segmentation, S', can be formed by replacing T_1 and T_2 with $T_1 \cup T_2$, and keeping all other segments fixed. These are the familiar notions of region merging and segmentation, expressed in our set-theoretic terms. With these definitions, however, region merging does not really occur. Rather, segments are merged by taking the union of two segments to yield a larger collection of regions (as opposed to taking the set union of the regions to create a larger region).

It is often profitable to begin with some initial partition of the image into small regions, and construct new segmentations by combining regions. This the standard approach taken in the region merging paradigm [69]. One justification for this is the savings in complexity achieved by considering this smaller set of possible segmentations. Another reason, which will become more evident when polynomial models are introduced in Chapter 5, is that often some minimal number of points is required in a region before the statistical models can be effectively employed. The initial segmentation represents the starting point in a region-merging algorithm. For instance, Silverman and Cooper begin with an initial image of *blocks*, which corresponds to an initial partition of the image into a grid of square regions [43]. Blocks are merged to yield *clusters*, which correspond to segments.

Let Π denote the set of all segmentations that could be constructed from the regions, $\mathcal{R}$. At one extreme, Π includes the partition induced by the original regions. At another extreme, Π contains the partition corresponding to combining all regions into one segment. The implication of starting with $\mathcal{R}$ is that there are many *image* partitions that are not considered.

Segmentations could, of course, be constructed by *partitioning* regions in $\mathcal{R}$. However, when the models are formulated in later chapters, it will be assumed that each region

[2]We mean that the union of the segments is connected.

is homogeneous.[3] If some region is not known to be homogeneous, we might want to consider splitting it into possibly homogeneous components. In a sense, $\mathcal{R}$ defines the granularity of the segmentation. The granularity can be made arbitrarily fine, up to the limiting case in which all elements of $\mathcal{R}$ are singleton subsets of D. Hence, using $\mathcal{R}$ as a basis does not impose any limitations on the space of segmentations that we can consider.

In practice, $\mathcal{R}$ can be constructed using a segmentation algorithm that allows one to select some level of confidence required for points to be grouped. This confidence level can be set in such a way that homogeneity can be assumed to hold with a very high degree of confidence over the regions that are constructed. In Chapter 8, we present experimental results on range data, with $\mathcal{R}$ obtained by recursively splitting regions that cannot be approximated by a single plane or that contain a Canny edge. The idea is to make extremely conservative choices about which points to keep together as regions, since combinations of regions that are uncertain will be considered at the higher level, probabilistically. This is essentially the approach taken in split-and-merge algorithms [77],[4]. A region-splitting algorithm that provides regions in a split-and-merge approach should also be sufficient for providing $\mathcal{R}$ in our context. In general, the initial regions used for other region-merging approaches are also sufficient. The set $\mathcal{R}$ represents the building blocks from which segmentation events are constructed in this setting (as it is also the building blocks for region-merging approaches).

3.2 The Segmentation Sample Space

In this section we consider Π in a probability space, by considering subsets of Π and their associated probabilities. Hence a probability, $P(\{S\})$, is associated with each possible $S \in \Pi$. In real applications, the size of Π is quite large. Therefore, as will

[3] In practice, one cannot usually guarantee this, and we expect a natural performance degradation when nonhomogeneous regions are presented.

be seen shortly, we have developed a technique in which much of the sample space Π is represented implicitly by accessing only selected subsets of Π. This will allow tremendous savings over representing every element $S \in \Pi$ and its corresponding probability.

The Segmentation Sample Space (SSS) is represented by the probability triple $(\Pi, \mathcal{A}, P)$. Recall that Π is the set of all segmentations that can be formed using $\mathcal{R}$. In the triple, $\mathcal{A}$ represents the set of all subsets of Π (i.e., the power set of Π), and P denotes a probability mapping, defined on $\mathcal{A}$. The SSS adheres to the standard probability axioms:

$$0 \leq P(A) \leq 1, \ \forall A \in \mathcal{A}, \tag{3.1}$$

$$P(\Pi) = 1, \tag{3.2}$$

and

$$A_1 \cap A_2 = \emptyset \rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2), \ \forall A_1, A_2 \in \mathcal{A}. \tag{3.3}$$

3.3 The Segment Sample Space

A segment sample space (TSS) essentially describes all ways to construct segments that contain some specified region. Its general purpose is to specify locally how to apply evidence, ultimately determining a probability mapping on Π. This concept will now be defined in detail. The structure and operations on this space will be introduced independently of the previous discussion about the SSS. When the SSS and the TSS are related in Section 3.5, both will be considered simultaneously.

For a particular region $R_i \in \mathcal{R}$, Θ_i is the set of all possible segments that contain R_i. One such element of Θ_i is the singleton $\{R_i\}$. Another might contain R_i and several adjacent regions. If D is connected, then the entire set $\mathcal{R}$ also belongs to Θ_i. By this definition, for any region $R_i \in \mathcal{R}$, there is a corresponding set of segments Θ_i.

A TSS is defined by the triple $(\Theta_i, \mathcal{B}_i, P)$. As defined before, Θ_i is the set of all segments that include R_i, $\mathcal{B}_i$ is the set of all subsets of Θ_i, and P is the probability mapping on $\mathcal{B}_i$.[4]

The next step is to define a method for implicitly representing elements of $\mathcal{B}_i$. For real image applications, the number of segments in Θ_i will typically be extremely large; the set $\mathcal{B}_i$ is exponentially larger. Any scheme that requires enumeration of either of these sets would be severely hindered by combinatorial explosion. It will be necessary to select certain elements of $\mathcal{B}_i$ in an organized manner, when evidence induced by the models is applied, and a representation of a TSS is built. This provides motivation for the definitions and representations that follow, which will be used to apply Bayesian evidence efficiently, affecting the probability distribution on Θ_i.

3.3.1 Compact representation of TSS events

To represent elements of $\mathcal{B}_i$, we introduce two set definitions and a function. Specifically, we can define an element $B \in \mathcal{B}_i$ by describing a set of regions which must be included in every element of B, and a set of regions which must be *not* be included in any element of B. These two sets essentially contain all of the information that is common to the segments in B. The *inclusion set*, I, is a set of regions which always includes R_i. The *exclusion set*, E, is also a set of regions, each of which is required to be adjacent to some region in I. Finally, $I \cap E = \emptyset$.

We define a function $\tau(I, E)$, which maps to some $B \in \mathcal{B}_i$, as

$$\tau(I, E) = \{T \in \Theta_i : I \subseteq T, E \cap T = \emptyset\}. \tag{3.4}$$

This definition precisely represents what was previously described in words: $\tau(I, E)$ gives the set of all segments that include all of the regions in I, and exclude all regions in E.

[4]We have a different probability map for each possible TSS, and should technically use P_i. We refrain from this to make some of the later notation simpler.

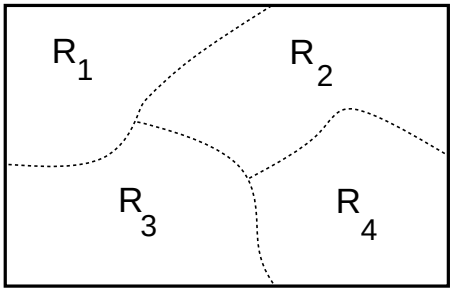


Figure 3.2 A simple image composed of only four regions is provided as an example.

Proposition 1 *The mapping defined by τ is well-defined and onto $\mathcal{B}_i$.*

Proof *Well-defined:* Note that $\tau(I, E)$ is a maximal subset with respect to $\subseteq$. Suppose τ could map to two different events, B_1 and B_2. Then it must also map to $B_1 \bigcup B_2$ by (3.4). If $B_1 \neq B_2$, neither can satisfy the maximality condition since $B_1 \subset B_1 \bigcup B_2$ and $B_2 \subset B_1 \cup B_2$. This leads to a contradiction; hence τ is well-defined.

Onto: Suppose there exists some $B \in \mathcal{B}_i$ which is not the image of τ for any I or E sets. Let $I = \bigcap_{S \in B} S$. Let E be the set of all regions that are not in $\cup_{S \in B} S$, and are adjacent to some region in I. The set $\tau(I, E)$ is precisely the original event B, assumed not to be the image of $\tau(I, E)$, which is a contradiction. □

The proposition implies that every event $B \in \mathcal{B}_i$ has a well-defined representation in terms of I and E sets.

Before continuing further, an example will be presented to illustrate the concepts. Figure 3.2 shows a hypothetical image D consisting of four regions: R_1, R_2, R_3 and R_4. Region R_1 is selected as the fixed region; Θ_1 is then the set of all segments that include R_1,

$$\begin{aligned} \Theta_1 \;=\; & \{\{R_1\}, \{R_1, R_2\}, \{R_1, R_3\}, \{R_1, R_2, R_3\}, \\ & \{R_1, R_2, R_4\}, \{R_1, R_3, R_4\}, \{R_1, R_2, R_3, R_4\}\}. \end{aligned} \tag{3.5}$$

Table 3.1 Segment space events are represented for the four-region example. The columns under I and E denote the include and exclude sets, respectively. The column under $\tau(I, E)$ gives the corresponding event in $\mathcal{B}_1$.

h	I	E	$\tau(I, E)$
0	$\{R_1\}$	$\{\}$	Θ_1
1	$\{R_1\}$	$\{R_2\}$	$\{\{R_1\}, \{R_1, R_3\}, \{R_1, R_3, R_4\}\}$
2	$\{R_1, R_2\}$	$\{\}$	$\{\{R_1, R_2\}, \{R_1, R_2, R_4\}, \{R_1, R_2, R_3\}, \{R_1, R_2, R_3, R_4\}\}$
3	$\{R_1\}$	$\{R_2, R_3\}$	$\{\{R_1\}\}$
4	$\{R_1, R_3\}$	$\{R_2\}$	$\{\{R_1, R_3\}, \{R_1, R_3, R_4\}\}$
5	$\{R_1, R_2\}$	$\{R_3\}$	$\{\{R_1, R_2\}, \{R_1, R_2, R_4\}\}$
6	$\{R_1, R_2, R_3\}$	$\{\}$	$\{\{R_1, R_2, R_3\}, \{R_1, R_2, R_3, R_4\}\}$
7	$\{R_1, R_3\}$	$\{R_2, R_4\}$	$\{\{R_1, R_3\}\}$
8	$\{R_1, R_3, R_4\}$	$\{R_2\}$	$\{\{R_1, R_3, R_4\}\}$
9	$\{R_1, R_2\}$	$\{R_3, R_4\}$	$\{\{R_1, R_2\}\}$
10	$\{R_1, R_2, R_4\}$	$\{R_3\}$	$\{\{R_1, R_2, R_4\}\}$
11	$\{R_1, R_2, R_3\}$	$\{R_4\}$	$\{\{R_1, R_2, R_3\}\}$
12	$\{R_1, R_2, R_3, R_4\}$	$\{\}$	$\{\{R_1, R_2, R_3, R_4\}\}$

Note that $\{R_1, R_4\} \notin \Theta$ since R_1 is not adjacent to R_4. For larger images, the number of segments is typically a small fraction of the total number of possible (connected or not) region groupings. This is because the excluded regions tend to disconnect the included regions from the rest of the image. This happens in the example when R_4 is cut off from R_1 by excluding R_2 and R_3.

Table 3.1 shows the results of applying τ for some events in $\mathcal{B}_1$. Each entry in the rightmost column is a set of segments, with each segment specified by a list of its constituent regions. In rows 3 and 7 to 12 of the table, $\tau(I, E)$ represents singleton elements of $\mathcal{B}_1$. These cases are termed *ground events*. As will be seen shortly, ground events play an important role in our approach. It is important to note that $\tau(I, E)$ provides a compact representation for potentially large subsets of Θ_i. For example, in row 2, four elements of Θ_1 are implicitly represented by $I = \{R_1, R_2\}$ and $E = \emptyset$. The savings in representation become greater as the number of regions in $\mathcal{R}$ increases.

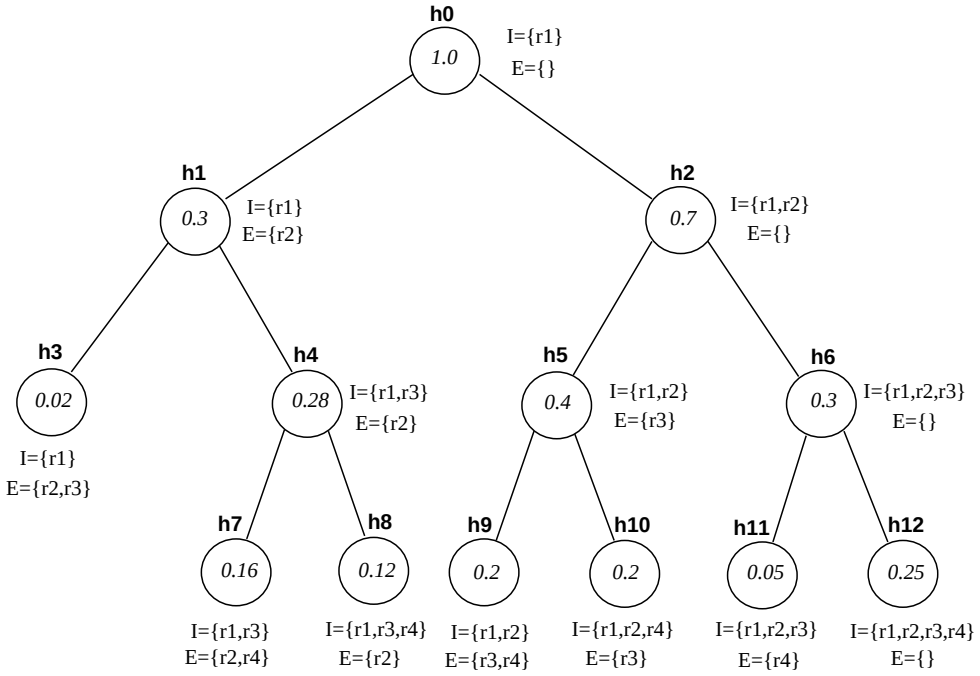


Figure 3.3 This figure shows a tree of events on Θ_i corresponding to four regions. The I and E sets are indicated for each hypothesis. Each node is also assigned a consistent probability.

The rows in Table 3.1 can be organized into a binary tree structure, as illustrated in Figure 3.3. Such a tree represents a hierarchy of events in Θ_1. Each node, with its I and E sets given, represents one of the rows in Table 3.1. Node h_0 in the tree represents the event Θ_1 (taken as an element of $\mathcal{B}_1$). The two child nodes of h_0, namely, h_1 and h_2, represent disjoint events whose union is represented by h_0. For any two nodes in the tree that share the same parent, the two nodes together represent disjoint events whose union is the event represented by the parent. The ground events in $\mathcal{B}_1$ appear as leaves of the tree, since these events represent single segments. This tree structure conceptually represents the events that will be used to build a representation of the TSS. For illustration, a consistent probability has been assigned to each node in Figure 3.3.

We note that there is a close correspondence between this tree and classification and regression trees [78],[79],[80]. Classification and regression trees are used to represent a sample space efficiently, often for the purpose of pattern recognition. They differ, however, since these methods are concerned with the efficiency of the entire tree structure, since the size and depth of the classification tree affects the classification cost. In our framework, the tree is used only for conceptual purposes.

3.3.2 Constructing an approximate TSS representation

In this section, we are concerned with constructing a hierarchy of events, such as the one in Figure 3.3. To accomplish this it will be necessary to consider a sequence of partitions of Θ_i. Each partition in the sequence will represent one step in the construction of the tree. We will describe a cover, which is an approximate TSS representation, and a refinement, which is an operation that creates closer approximations.

We define a *cover*, C, of the TSS to be a set of pairwise-disjoint events in $\mathcal{B}_i$, which forms a partition of Θ_i.[5] For a tree such as the one shown in Figure 3.3, a cover can also be considered as a set of nodes whose corresponding events form a partition of Θ_i. For example (see Figure 3.4), a cover can be constructed from the tree using nodes h_1, h_6, h_9 and h_{10}, or from all of the ground nodes (the leaves of the tree). Also, h_0 alone creates a cover containing only one event, Θ_1, in the partition.

A cover can be considered as an approximate representation of Θ_i and P on the TSS. A cover C_1 is considered to be a better approximation than C_2, if C_1 can be obtained by partitioning some of the elements of C_2.[6] Consider a cover $C = \{B_1, B_2, \ldots, B_k\}$ with specified probabilities on each of the events, $\{P(B_1), P(B_2), \ldots, P(B_k)\}$. If C is the set of all ground events in $\mathcal{B}_i$, then an exact representation of the TSS is obtained;

[5] Note that this does not completely coincide with the definition of a cover in analysis, which does not require pairwise disjointness.

[6] By this concept of approximation, a partial ordering (not a linear ordering) is imposed on the set of all covers.

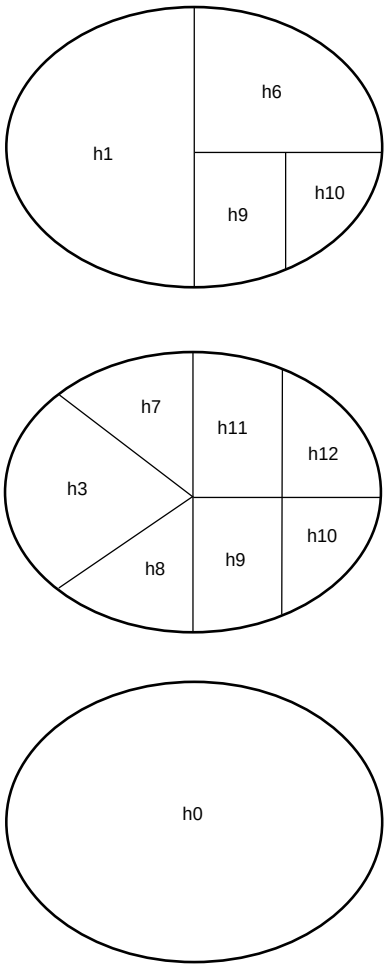


Figure 3.4 Several covers of Θ_1 obtained by taking nodes from the example tree.

all of the elements of Θ_i are explicitly represented in C, and the probability is given for each of them. Since the ground events are mutually exclusive, the probability for an arbitrary event $B \in \mathcal{B}_i$ is obtained by summing the probabilities $P(\{T\})$ for each $T \in B_i$. Hence if C is the set of all ground events in $\mathcal{B}_i$, then the entire probability map is fully determined. Suppose that $C = \{\Theta_i\}$. We know that $P(\Theta_i) = 1$; however, the probabilities of the other events in $\mathcal{B}_i$ cannot be directly determined. This corresponds to the poorest approximation possible, since no information is actually present about the TSS.

In the example appearing in Table 3.1 and Figure 3.3, we have enumerated all ground events of $\mathcal{B}_i$, and seven other events. In practice, it is intractable to construct an entire tree. It is reasonable, however, to expand and represent only the portion of the tree that corresponds to events with significant probability value $P(B)$. The precise details of constructing this tree efficiently in practice are discussed in Chapter 7. For now, we will present the operations for constructing the tree, as necessary preparation and motivation for the forthcoming chapters. The goal is to obtain a cover that explicitly represents ground events with high probability, and represents events with low probability as larger subsets of Θ_i. A cover with few events forms a compact representation of the TSS since each event is represented by I and E sets. This representation is useful since it can explicitly represent the most favorable segments (as ground events) and their corresponding probabilities. The represented ground events will be used later to build an SSS representation, and this extension will be discussed in Section 3.5.

Our approach to constructing a representation of the TSS is to start with an initial cover, and iteratively construct new covers, each of which gives a closer TSS approximation. In terms of classification and regression trees, this corresponds to the notion of impurity reduction through partitioning [79]. The goal in the classification and regression tree setting is to select finer partitions of the sample space to optimally reduce the expected loss due to approximate representation. In our framework, we will also be reducing the expected loss, but with an interest in obtaining a representation of ground events that have highest probability.

The *refinement mapping*, ρ, is used to perform a single iteration in building the representation of the TSS. It takes a cover, C, an event, $B_\rho \in C$, and a region, R_ρ, and yields a cover that represents a better approximation, $C' = \rho(C, B_\rho, R_\rho)$. The region R_ρ, termed the *refinement region*, is adjacent to some region in I_ρ and $R_\rho \notin I_\rho \cup E_\rho$, in which $B_\rho = \tau(I_\rho, E_\rho)$. The B_ρ is termed the *refinement event*. The cover, C', is termed the

Table 3.2 Enumeration of segments for the example image

Name	Segment
T_1	$\{R_1\}$
T_2	$\{R_1, R_2\}$
T_3	$\{R_1, R_3\}$
T_4	$\{R_1, R_2, R_3\}$
T_5	$\{R_1, R_2, R_4\}$
T_6	$\{R_1, R_3, R_4\}$
T_7	$\{R_1, R_2, R_3, R_4\}$

Table 3.3 Sequence of refinements to generate example tree

Cover	R_ρ	B_ρ	Partition of Θ_1
C_0	-	-	$\{\{T_1, T_2, T_3, T_3, T_4, T_5, T_6, T_7\}\}$
C_1	R_2	Θ_i	$\{\{T_2, T_4, T_5, T_7\}, \{T_1, T_3, T_6\}\}$
C_2	R_3	$\{T_1, T_3, T_6\}$	$\{\{T_2, T_4, T_5, T_7\}, \{T_1\}, \{T_3, T_6\}\}$
C_3	R_3	$\{T_2, T_4, T_5, T_7\}$	$\{\{T_2, T_5\}, \{T_4, T_7\}, \{T_1\}, \{T_3, T_6\}\}$
C_4	R_4	$\{T_3, T_6\}$	$\{\{T_2, T_5\}, \{T_4, T_7\}, \{T_1\}, \{T_3\}, \{T_6\}\}$
C_5	R_4	$\{T_2, T_5\}$	$\{\{T_2\}, \{T_5\}, \{T_4, T_7\}, \{T_1\}, \{T_3\}, \{T_6\}\}$
C_6	R_4	$\{T_4, T_7\}$	$\{\{T_2\}, \{T_5\}, \{T_4\}, \{T_7\}, \{T_1\}, \{T_3\}, \{T_6\}\}$

refined cover with respect to C. The refinement mapping is formally defined as

$$\rho(C, B_\rho, R_\rho) = (C \ - \ B_\rho) \cup \{\tau(I_\rho \cup \{R_\rho\}, E_\rho), \tau(I_\rho, E_\rho \cup \{R_\rho\})\}. \tag{3.6}$$

The only difference between C and C' is the replacement of B_ρ by $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and $\tau(I_\rho, E_\rho \cup \{R_\rho\})$. These two new events will be termed *refined events*. The refinement event B_ρ has been partitioned by using the refinement region R_ρ. All segments in B_ρ that include R_ρ are in $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$. The remaining elements of B_ρ (all those that exclude R_ρ) are in $\tau(I_\rho, E_\rho \cup \{R_\rho\})$.

By repeated applications of the refinement mapping, a tree can be constructed similar to the one in Figure 3.3. Table 3.3 depicts one way of constructing this tree from a sequence of refinement mappings. For convenience, Table 3.2 names all of the segments of Θ_i. Initially, we begin with C_0. This corresponds to the starting node, h_0, in the tree. Cover C_1 is constructed by using the mapping $C_1 = \rho(C_0, \Theta_1, R_1)$. In the tree, this corresponds to the creation of nodes h_1 and h_2. The cover C_2 is constructed using R_3 and event $\{T_1, T_3, T_6\}$, corresponding to the addition of nodes h_3 and h_4. The iterations are continued until cover C_6. No further refinements can be performed after C_6 since every element in the cover is a ground event.

Recall that, in practice, a tree is not constructed to the point at which all ground events are represented. The refinement mapping will not be applied to events that have very low probability. We will be interested in the ground events that have been represented, but the vast majority of all ground events will be represented only implicitly as members of some event in the cover. Each of these events is a subset of Θ_i, which can represent many segments that have very low probability and are therefore uninteresting. The details involved in making these representation choices, as well as choosing the refinement event and the refinement region, will be discussed in Chapter 7.

3.4 Models for Refined-Event Probability Assignments

Although constructing covers by refinement has been specified structurally, no attention has yet been given to determining the probability assignments to the events that are created by refinement. One primary issue must be considered: we are not given a complete representation of P on the TSS. This would require one probability assignment for every ground event. If this was given, then the probability of some other event B is simply the summation over all of the ground events that are subsets of B.

Recall that each refinement removes one event in a cover of Θ_i and replaces it with two disjoint events whose union is the original event. The basic strategy in building a TSS representation is to determine probability assignments of the new events when this step is performed. This requires deciding how to divide the probability of the original event between the two new events.

There are two basic mechanisms that exert influence on this probability assignment. As with any Bayesian framework, there is some prior distribution on the sample space. Also, after the application of evidence, some posterior distribution is obtained. Before constructing the tree, a prior distribution will be defined implicitly on the TSS. Model-based evidence will be used, along with the prior distribution, to determine probability assignments at the refinement step. These issues will be discussed in the remainder of this chapter.

Using the refinement mapping just described, successive partitions are constructed from Θ_i as prescribed by (3.6). Recall that in this operation, after selecting B_ρ and R_ρ, we partition $B_\rho = \tau(I_\rho, E_\rho)$ into $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and $\tau(I_\rho, E_\rho \cup \{R_\rho\})$. For probabilistic consistency, it is necessary to have

$$P(\tau(I_\rho, E_\rho)) = P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) + P(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{3.7}$$

It is assumed inductively that $P(B_\rho)$ is known, and that the two probabilities on the right side of (3.7) must be determined. Before the first refinement is performed, $B_\rho = \Theta_i$, and $P(\Theta_i) = 1$, reflecting the starting condition. At each iteration we will have $P(\tau(I_\rho, E_\rho))$ and need to determine probability assignments on the right side of (3.7) while making use of priors and model-based evidence. This will determine the probability assignments at each refinment, and hence the probabilities for all of the events in a cover.

If it were reasonable to enumerate all of Θ_i, then it would be possible to apply evidence directly to the ground events. This would yield a complete representation of the probability map on the TSS. To avoid this enumeration the probability assignments are made in this incremental form. When considering the efficiency of resulting covers, the assigned proability will dramatically affect the choice of B_ρ and R_ρ for a new refinement. This issue is of an algorithmic nature and is discussed in Chapter 7.

It has been assumed in (3.7) that the probability of B_ρ is never altered by the refinement operation. In general, it could be the case that evidence about R_ρ could cause $P(B_\rho)$ to increase or decrease. Although we have not yet introduced specific models of homogeneity, it is important to note that in the most general setting a model could be considered that causes $P(B_\rho)$ to change after R_ρ is considered. This is precisely the issue that arises with taxonomic hierarchies, analyzed by Pearl [81]. An efficient method of propagating evidence-based, posterior probabilities throughout a hierarchy of events is presented in Pearl's work, but the construction of the hierarchy by the refinement mapping is not considered. Models that cause $P(B_\rho)$ to change are much more difficult to analyze in our context.

An empirical investigation of this more difficult probability model, applied to our TSS representation framework, can be found in [82]. Pearl's more general, Bayesian networks have also been applied to computer vision problems. Agosta, and Binford et al. have considered them for model-based object recognition applications [83],[84]. Sarkar

and Boyer have proposed Bayesian networks for a hierarchical organization of perceptual features [85].

3.4.1 Simplifications and membership probability

The purpose of this section is to reduce the probability assignment of (3.7) into a form that we can directly utilize. The resulting expressions concisely represent the dependencies on R_i and R_ρ. The membership probability is defined, and computation using general image models is presented in Chapter 4.

An alternative way to represent the probability terms on the right side in (3.7) is by

$$P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) = P(\tau(I_\rho \cup \{R_\rho\}, E_\rho) \,|\, B_\rho)\; P(B_\rho) \tag{3.8}$$

and

$$P(\tau(I_\rho, E_\rho \cup \{R_\rho\})) = P(\tau(I_\rho, E_\rho \cup \{R_\rho\}) \,|\, B_\rho)\; P(B_\rho). \tag{3.9}$$

This can be seen by observing that since $\tau(I_\rho \cup \{R_\rho\}, E_\rho) \subseteq B_\rho$, we have

$$P(\tau(I_\rho \cup \{R_\rho\}, E_\rho), B_\rho) = P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)). \tag{3.10}$$

Similarly, for $\tau(I_\rho, E_\rho \cup \{R_\rho\}) \subseteq B_\rho$, we have

$$P(\tau(I_\rho, E_\rho \cup \{R_\rho\}), B_\rho) = P(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{3.11}$$

Since it is assumed that $P(B_\rho)$ is known, the conditionals of (3.8) and (3.9) are all that need to be determined. Recall that $B_\rho = \tau(I_\rho, E_\rho)$. The conditionals become

$$P_I \equiv P(\tau(I_\rho \cup \{R_\rho\}, E_\rho) \,|\, \tau(I_\rho, E_\rho)) \tag{3.12}$$

and

$$P_E \equiv P(\tau(I_\rho, E_\rho \cup \{R_\rho\}) \,|\, \tau(I_\rho, E_\rho)). \tag{3.13}$$

The probabilities, P_I and P_E, are called *membership probabilities* and are very important for the remainder of this thesis. We consider P_I as the probability that R_ρ is a

member of the maximal homogeneous segment that contains R_i, and P_E as the probability that R_ρ is a not a member of the maximal homogeneous segment that contains R_i. Note that $P_I + P_E = 1$ since the two events together represent a partition of $\tau(I_\rho, E_\rho)$, which is given. This implies that only one of these probability assignments is necessary to determine the probabilities to assign to the new events created by a refinement. The expressions above can be simplified further by making use of the following proposition:

Proposition 2 *If I, I_1, I_2 are include sets with $I = I_1 \cup I_2$, and E, E_1, E_2 are exclude sets with $E = E_1 \cup E_2$ then $\tau(I, E) = \tau(I_1, E_1) \cap \tau(I_2, E_2)$.*

Proof We have $\tau(I, E) = \tau(I_1 \cup I_2, E_1 \cup E_2)$. By definition

$$\tau(I_1 \cup I_2, E_1 \cup E_2) = \{T \in \Theta_i : (I_1 \cup I_2) \subseteq T, (E_1 \cup E_2) \cap T = \emptyset\}. \tag{3.14}$$

This can equivalently be expressed as

$$\{T \in \Theta_i : I_1 \subseteq T, I_2 \subseteq T, E_1 \cap T = \emptyset, E_2 \cap T = \emptyset\}. \tag{3.15}$$

This is the same as

$$\{T \in \Theta_i : I_1 \subseteq T\} \cap \{T \in \Theta_i : I_2 \subseteq T\} \cap \{T \in \Theta_i : E_1 \cap T = \emptyset\} \cap \{T \in \Theta_i : E_2 \cap T = \emptyset\}. \tag{3.16}$$

But this is equivalent to

$$\{T \in \Theta_i : I_1 \subseteq T, E_1 \cap T = \emptyset\} \cap \{T \in \Theta_i : I_2 \subseteq T, E_2 \cap T = \emptyset\}, \tag{3.17}$$

which is simply $\tau(I_1, E_1) \cap \tau(I_2, E_2)$. □

By using Proposition 2 with $I_1 = I_\rho$, $I_2 = \{R_i, R_\rho\}$, $E_1 = E_\rho$, $E_2 = \emptyset$, (3.12) can be rewritten as

$$P_I = P(\tau(\{R_i, R_\rho\}, \emptyset), \tau(I_\rho, E_\rho) \,|\, \tau(I_\rho, E_\rho)). \tag{3.18}$$

Also, (3.13) can be rewritten, by using the proposition with $I_1 = I_\rho$, $I_2 = \{R_i\}$, $E_1 = E_\rho$, $E_2 = \{R_\rho\}$, as

$$P_E = P(\tau(\{R_i\}, \{R_\rho\}), \tau(I_\rho, E_\rho) \,|\, \tau(I_\rho, E_\rho)). \tag{3.19}$$

These can both be simplified by making use of the fact that for any two events B_1 and B_2, $P(B_1, B_2|B_2) = P(B_1|B_2)$. By using this directly, the two can be reduced to

$$P_I = P(\tau(\{R_i, R_\rho\}, \emptyset) \,|\, \tau(I_\rho, E_\rho)) \tag{3.20}$$

and

$$P_E = P(\tau(\{R_i\}, \{R_\rho\}) \,|\, \tau(I_\rho, E_\rho)). \tag{3.21}$$

The expressions (3.20) and (3.21) are expressed in a form explicitly indicating the importance of adding R_ρ to I_ρ or E_ρ. This is the fundamental distinction between the event B_ρ and the two refined events. It is natural to expect that the probability due to evidence will depend directly on the new region that has been brought into consideration, and this has been precisely represented by these expressions.

3.4.2 IE-independent and IE-dependent models

In general, for determining P_I (or P_E), a significant issue must be considered: the notion of IE-independence. In the next chapter, we will derive expressions for computing the membership probability that use information obtained from each region. The probability, P_I, from (3.20) depends on R_ρ, I_ρ and E_ρ. If a model uses information from all of these regions, it is termed *IE-dependent*. Recall that R_i is contained in every TSS segment. If a model uses information only from R_p and $R_i \in I_\rho$, then it is termed *IE-independent*, since the membership probability is independent of the regions in I and E (except R_i). Explicitly, IE-independence can be expressed for P_I as

$$P(\tau(\{R_i, R_\rho\}, \emptyset) \,|\, \tau(I_\rho, E_\rho)) = P(\tau(\{R_i, R_\rho\}, \emptyset)), \tag{3.22}$$

and for P_E as

$$P(\tau(\{R_i\}, \{R_\rho\}) \,|\, \tau(I_\rho, E_\rho)) = P(\tau(\{R_i\}, \{R_\rho\})). \tag{3.23}$$

The choice between these models depends on the application. In Chapter 8, the differences when using polynomial surface models are examined. The IE-dependent model

is, of course, the more general model since it does not require an additional assumption; however, the computations that it will tend to produce are more costly. Since the IE-independent model uses only R_ρ and R_i, only one membership probability computation is performed for each potential refinement region. If I_ρ and E_ρ are also considered, then a membership probability computation must be performed for each refinement.

Depending on the model, however, it may not always be possible to introduce the IE-independence assumption. Since the membership probabilities with this assumption depend on the relationship between R_ρ and R_i, there must be a significant amount of information available R_i. Typically, this will imply that better utilization of model evidence will be possible when R_i contains more data points. If this implication is not intuitively clear, it should become apparent when surface models are introduced in Chapter 5.

3.4.3 Posterior evidence-based probabilities

In the next chapter, a general statistical formalism will be developed which will be used to determine P_I and P_E for both IE-independent models and IE-dependent models. By using this formalism, the probability assignments can be made at each step of the refinement, allowing Θ_i to be partitioned to yield better approximations to the TSS. The models assume some prior value of P_I and P_E and yield an evidence-based, posterior probability.

For the IE-independent model, and a generic piece of evidence, e, the posterior probability of P_I is obtained through Bayes' rule as

$$P(\tau(\{R_i, R_\rho\}, \emptyset) \mid e) = \frac{P(e \mid \tau(\{R_i, R_\rho\}, \emptyset))\ P(\tau(\{R_i, R_\rho\}, \emptyset))}{P(e)}, \tag{3.24}$$

and the posterior for P_E is

$$P(\tau(\{R_i, R_\rho\}, \emptyset)^C \mid e) = \frac{P(e \mid \tau(\{R_i, R_\rho\}, \emptyset)^C)\ P(\tau(\{R_i, R_\rho\}, \emptyset)^C)}{P(e)}. \tag{3.25}$$

We have $\tau(\{R_i, R_\rho\}, \emptyset)^C = \Theta_i - \tau(\{R_i, R_\rho\}, \emptyset)$, which is the event complementary to $\tau(\{R_i, R_\rho\}, \emptyset)$. Note that $\tau(\{R_i, R_\rho\}, \emptyset)$ and $\tau(\{R_i\}, \{R_\rho\})$ form a partition of Θ_i, hence $\tau(\{R_i, R_\rho\}, \emptyset)^C = \tau(\{R_i\}, \{R_\rho\})$.

Similar expressions can be written for the IE-dependent case:

$$\begin{aligned} & P(\tau(\{R_i, R_\rho\}, \emptyset) \mid \tau(I_\rho, E_\rho), e) \\ & = \frac{P(e \mid \tau(\{R_i, R_\rho\}, \emptyset), \tau(I_\rho, E_\rho)) P(\tau(\{R_i, R_\rho\}, \emptyset), \tau(I_\rho, E_\rho))}{P(e, \tau(I_\rho, E_\rho))} \end{aligned} \tag{3.26}$$

and

$$\begin{aligned} & P(\tau(\{R_i, R_\rho\}, \emptyset)^C \mid \tau(I_\rho, E_\rho), e) \\ & = \frac{P(e \mid \tau(\{R_i, R_\rho\}, \emptyset)^C, \tau(I_\rho, E_\rho))\ P(\tau(\{R_i, R_\rho\}, \emptyset)^C, \tau(I_\rho, E_\rho))}{P(e, \tau(I_\rho, E_\rho))}. \end{aligned} \tag{3.27}$$

By making use of Proposition 2, the IE-dependent probabilities can be reduced to

$$P(\tau(\{R_i, R_\rho\}, \emptyset) \mid \tau(I_\rho, E_\rho), e) = \frac{P(e \mid \tau(I_\rho \cup \{R_\rho\}, E_\rho)) P(\tau(I_\rho \cup \{R_\rho\}, E_\rho))}{P(e, \tau(I_\rho, E_\rho))} \tag{3.28}$$

and

$$P(\tau(\{R_i, R_\rho\}, \emptyset)^C \mid \tau(I_\rho, E_\rho), e) = \frac{P(e \mid \tau(I_\rho, E_\rho \cup \{R_\rho\})) P(\tau(I_\rho, E_\rho \cup \{R_\rho\}))}{P(e, \tau(I_\rho, E_\rho))}. \tag{3.29}$$

The particular interpretation of e will be discussed in the next chapter.

3.5 Segment Sample Space to Segmentation Sample Space Relationships

Now that the TSS has been defined, its relationship to the SSS will be discussed. This relationship is interesting since the SSS is a space of segmentations, representing all alternatives we wish to consider, and a TSS describes a distribution over segments that represent a portion of each of the segmentations. It will turn out that the TSS

can be used as a building block with which representations of events on the SSS can be constructed. It is necessary to relate the distribution over Θ_i to the distribution on the SSS since traditionally one is interested in full image segmentations. Also, it is interesting to investigate the distribution arising from the composition of a few neighboring segments, since often only a few segments form a useful constraint for recognition [86].

It is assumed here that we have methods to compute the probabilities described in the previous section. Chapters 4-6 present a detailed discussion of this computation.

3.5.1 The segment-to-segmentation mapping

The mapping from a TSS to the SSS is determined by specifying a function from the set of events on a TSS to the events on the SSS. Formally, this mapping is expressed as $f_i : \mathcal{B}_i \rightarrow \mathcal{A}$. From this function, the relationship between the probability maps can also be determined. To define f_i, first consider a ground event on a TSS, denoted by $\{T\}$.[7] We define f_i on this ground event as

$$f_i(\{T\}) = \{S \in \Pi : T \in S\}. \tag{3.30}$$

The event $f_i(\{T\}) \in \mathcal{A}$ is the set of all segmentations that include the segment T. Since every $T \in \Theta_i$ contains R_i and segments in a segmentation are disjoint, we have

$$f_i(\{T_1\}) \cap f_i(\{T_2\}) = \emptyset \ \forall T_1, T_2 \in \Theta_i, T_1 \neq T_2. \tag{3.31}$$

In other words, no single segmentation can contain two distinct segments that belong to the same Θ_i. By using this fact and (3.30), we can define the mapping for a general event, $B \in \mathcal{B}_i$

$$f_i(B) = \bigcup_{T \in B} f(\{T\}) = \bigoplus_{T \in B} \{S \in \Pi : T \in S\}. \tag{3.32}$$

Above, $\oplus$ represents the orthogonal sum, or union of disjoint sets.

[7] We use $\{T\}$ instead of T since the ground event is a singleton subset of Θ_i.

By applying f_i to each ground event in Θ_i, we obtain a set of events which forms a partition of Π, with each set in the partition corresponding to a segment from the TSS. Hence, there is a correspondence between segments in a TSS and the events in the SSS obtained from f_i. The relationship between Θ_i and Π is similar to the notion of refinement used by Shafer [87]. In fact, the mapping f_i is very similar to what Shafer terms a *refining mapping*.

The next step is to define the relationship between the probability maps on the two spaces. To be very precise in this section and in Section 3.5.2, we will use P_Π to denote the probability map on the SSS, and P_Θ to denote the probability map on a TSS. The goal is to make the probability maps coincide on events that are equivalent through f_i. Explicitly, the probability assigned to a ground event in the TSS is assigned directly to the corresponding event on the SSS. This is

$$P_\Pi(f_i(\{T\})) = P_\Theta(\{T\}). \tag{3.33}$$

For an arbitrary event $B \in \mathcal{B}_i$ we use (3.32) to obtain

$$P_\Pi(f_i(B)) = \sum_{T \in B} P_\Theta(\{T\}) = P_\Theta(B). \tag{3.34}$$

This is true since the ground events are disjoint.

If we are given some probability map defined on $\mathcal{B}$, we have constraints only for the corresponding map on $\mathcal{A}$. This is due to the fact that a single segment in Θ_i maps to a subset of Π. This implies that a TSS can be considered as an approximate representation of the SSS. It will therefore be required only that probabilities of ground events on the SSS sum up to the probability of the corresponding ground event on the TSS. Also, it is assumed as in Section 3.4 that the probability assignment at the TSS is not altered by evidence from some later refinement.

3.5.2 Considering SSS representations using TSS representations

It will now be shown that the probability relationship, (3.34), along with the method of constructing TSS approximations, can be used to build SSS approximations that are better than the closest approximation that could be obtained by some Θ_i. The technique involves using TSSs repeatedly for different initial regions and organizing the results into a hierarchy of partitions on Π (very similar in appearance to the hierarchy constructed on Θ_i).

The process begins by choosing some initial region R_{i_1}, and building a representation of the ground events on Θ_{i_1} (or at least a representation of some of them). Denote this set of ground events by X. Take some event $\{T\} \in X$, and denote the subset of Π given by $f_1(\{T\})$ by A_{i_1}. Consider defining a new image, D', from D by taking

$$D' = D - \bigcup_{R \in T} R. \tag{3.35}$$

The image, D', corresponds to the removal of T from the original image, D. An initial region, R_{i_2}, is selected as a subset of D', and the corresponding TSS, Θ_{i_2}, can be constructed. Ground events in this TSS give single segments in D'. With Θ_{i_2} the event A_{i_1} is given from D; therefore, Θ_{i_2} gives a distribution over some segment pairs in the image D. Each pair represents T and some segment from Θ_{i_2}.

An example of this concept is described in terms of the example in Figures 3.2 and 3.3. Figure 3.5 shows the extended tree arising from constructing several TSSs in the manner just described. To keep the diagram from being cluttered, the I and E sets have not been indicated. Instead, the region, R_ρ, introduced at each refinement, is indicated. A left child indicates adding R_ρ to E. The right indicates adding R_ρ to I. Note that the upper portion of this tree coincides with Figure 3.3. In this figure, however, each square node represents the root node of the tree for a new TSS. At the root of the tree, R_1 was chosen, as before, to construct Θ_1. The nodes h_3, h_7, h_9, and h_{19} mark the initiation of

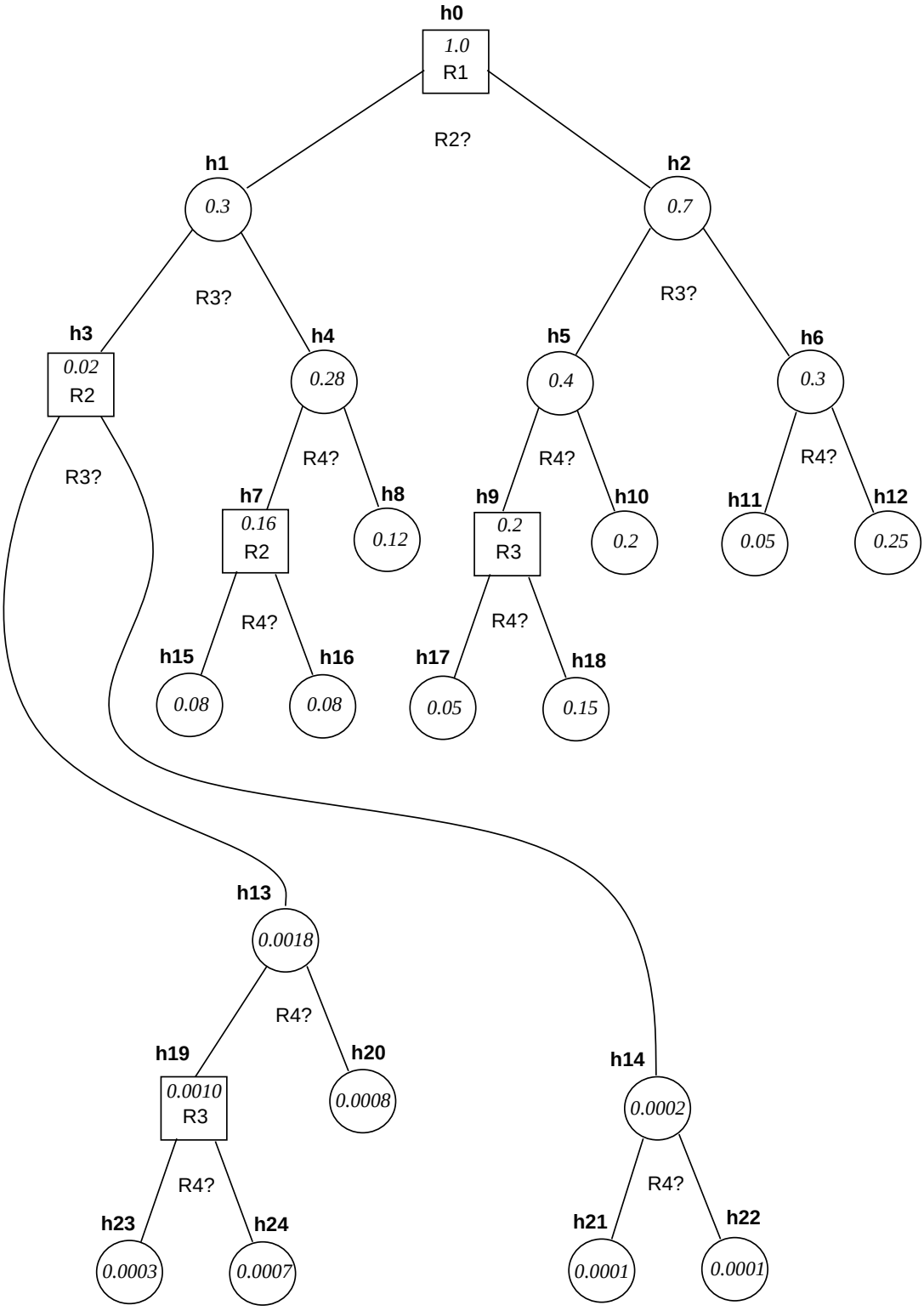


Figure 3.5 Full segmentation tree represented for example.

Table 3.4 The SSS ground events represented for four-region example

h	Segmentation	P(S)
8	$\{\{R_1, R_3, R_4\}, \{R_2\}\}$	0.12
10	$\{\{R_1, R_2, R_4\}, \{R_3\}\}$	0.2
11	$\{\{R_1, R_2, R_3\}, \{R_4\}\}$	0.05
12	$\{\{R_1, R_2, R_3, R_4\}\}$	0.25
15	$\{\{R_1, R_3\}, \{R_2, R_4\}\}$	0.08
16	$\{\{R_1, R_3\}, \{R_2\}, \{R_4\}\}$	0.08
17	$\{\{R_1, R_2\}, \{R_3\}, \{R_4\}\}$	0.05
18	$\{\{R_1, R_2\}, \{R_3, R_4\}\}$	0.05
20	$\{\{R_1\}, \{R_2, R_4\}, \{R_3\}\}$	0.0008
21	$\{\{R_1\}, \{R_2, R_3\}, \{R_4\}\}$	0.0001
22	$\{\{R_1\}, \{R_2, R_3, R_4\}\}$	0.0001
23	$\{\{R_1\}, \{R_2\}, \{R_3\}, \{R_4\}\}$	0.0003
24	$\{\{R_1\}, \{R_2\}, \{R_3, R_4\}\}$	0.0007

TSSs started by choosing some new initial region. The new starting region is indicated in the box. The new images with regions removed are shown in Figure 3.6. From left to right, they are: the initial image, the new image at h_3 (corresponding to the removal R_1), the new image at h_9 and h_{19}, and the new image at h_7. For this example, 5 TSSs were constructed, with all 13 possible segmentations represented. Table 3.4 indicates the segmentation represented at each of the terminating nodes, with its corresponding probability.

Figure 3.7 gives a schematic representation of the TSSs that are constructed. The SSS ground events generated by each TSS construction are indicated. By looking at this figure it becomes evident that the problem of finding SSS ground events is a kind of search problem. In practice T and R_{i_n} must be chosen at each TSS step, in an attempt to represent portions of the SSS that correspond to higher probabilities.

This again is an example representing the entire space; here the SSS is represented in its entirety since this is possible for such a simple image. In practice, this space will

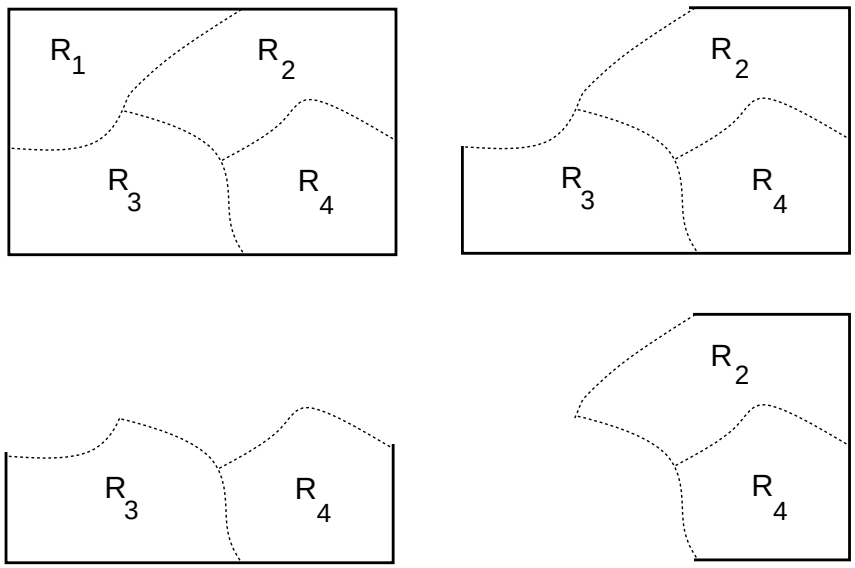


Figure 3.6 Partial images corresponding to various TSS constructions.

not be fully represented (indeed, not one of the TSSs needs to be fully constructed). This raises issues of efficiently constructing interesting portions of the SSS. This will be discussed at length in Chapter 7.

In Figure 3.5, probabilities have been assigned in a manner consistent with (3.34). Consider, as before, Θ_{i_1} and Θ_{i_2}, with $T_1 \in \Theta_{i_1}$. Segment T_1 is removed when considering some new region R_{i_2} used to construct Θ_{i_2}. Consider some segment $T_2 \in \Theta_{i_2}$. The joint probability P_Π of T_1 and T_2 is given by

$$P_\Pi(f_1(\{T_1\}) \cap f_2(\{T_2\})) = P_\Pi(f_2(\{T_2\})|f_1(\{T_1\}))P_\Pi(f_1(\{T_1\})). \tag{3.36}$$

We need to relate the probabilities on the right side of (3.36) to the TSS probabilities, $P_{\Theta_{i_1}}(\{T_1\})$ and $P_{\Theta_{i_2}}(\{T_2\})$. We assume that the probability maps on different TSSs are independent. This implies that there is no statistical dependency between different segments in the same segmentation. One approach to considering segment interdependence is discussed in Section 9.2.3. With the independence assumption and (3.34), and

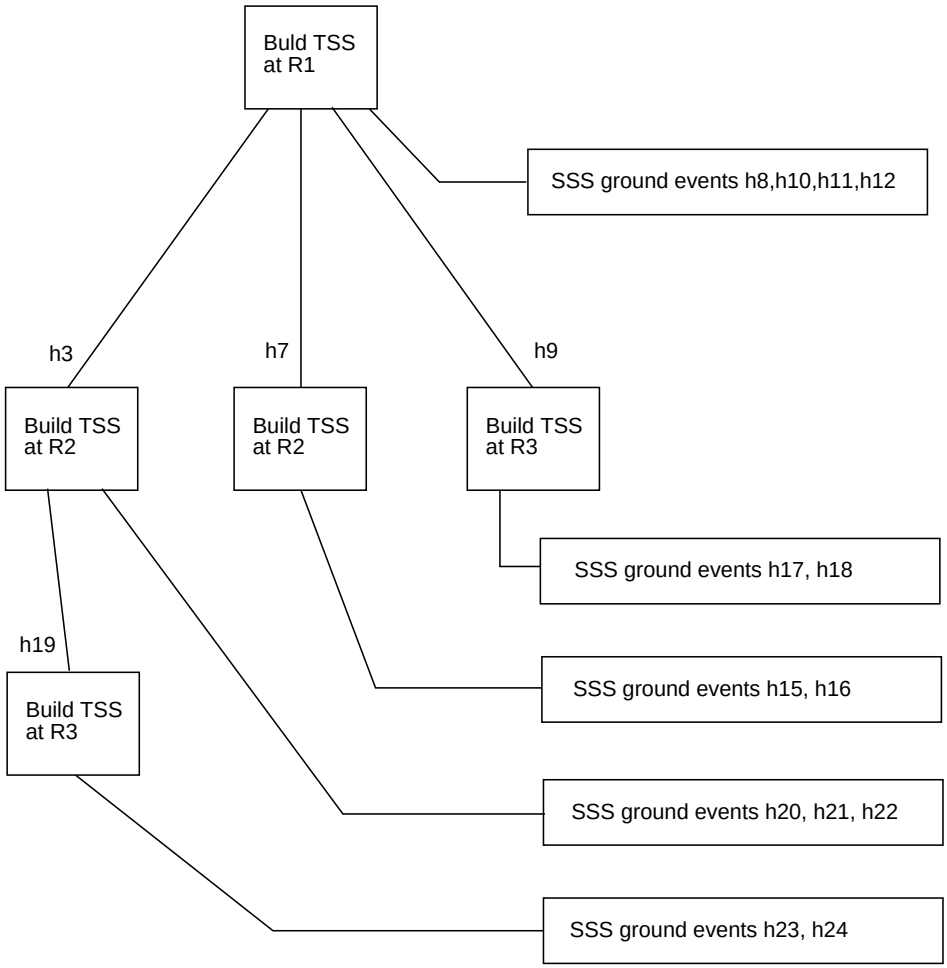


Figure 3.7 Schematic representation of the various TSSs that are constructed in the example image.

if $T_1 \cup T_2 = \emptyset$, we have

$$P_{\Pi}(f_2(\{T_2\})|f_1(\{T_1\}) = P_{\Theta}(\{T_2\}), \tag{3.37}$$

using the probability map on Θ_{i_2}. Also we have

$$P_{\Pi}(f_1(\{T_1\})) = P_{\Theta}(\{T_1\}), \tag{3.38}$$

using the probability map on Θ_{i_1}.

Together (3.37) and (3.38) give the probability of the event on Π corresponding to the set of all segmentations that contain both T_1 and T_2. In general (without the independence assumption), an expression analogous to (3.36) is applied to n segments and TSS spaces by

$$P_\Pi(f_1(\{T_1\}), f_2(\{T_2\}), \ldots, f_n(\{T_n\}))$$

$$= P_\Pi(f_n(\{T_n\})|f_{n-1}(\{T_{n-1}\}), \ldots, f_1(\{T_1\})) \cdots P_\Pi(f_2(\{T_2\}))|f_1(\{T_1\})) \; P_\Pi(f_1(\{T_1\})). \tag{3.39}$$

By using this assignment, the probabilities corresponding to ground events on the SSS can ultimately be obtained using individual TSS probabilities.

Note that, in practice, one may not even care about constructing ground events on the SSS. One contribution of this thesis is that it allows the consideration of events that represent partial segmentations, rather than representing explicit segmentations. Depending on the application, it may be necessary only to construct trees as in Figure 3.7 to some size. For instance, for model-based object recognition using objects composed of polynomial surface patches, three segments may be sufficient to determine the position of an object. Faugeras and Hebert provide conditions for which this holds [86]. By constructing the tree to represent only three-segment groups, an appropriate distribution is generated. Each three-segment group corresponds to a set of segmentations; however, the rest of the image (i.e., the regions that do not appear in any I or E sets) may not be interesting for the application.

3.6 Obtaining Noninformative Priors

Any Bayesian approach must include some consideration of a prior distribution. In this section we consider different distributions on TSS and SSS probability spaces, prior to the application of model-based evidence. The general goal is to reflect some kind

of uniformity, due to the lack of information that affects the probability distribution. It might be the case that one wants to introduce some bias through the priors, but this discussion will primarily be concerned with trying to eliminate unwanted bias to yield uniformity. The problem here is that, in the beginning, a representation of the sample space does not even exist. Even after refinements have been performed, the full representation does not exist. The goal is to try to be as "uniform" as possible, while not completely representing the space. If a prior with large deviations from some desired uniformity is selected and very little relevant information is available, the posterior could be heavily biased. If the priors do not reflect the desired uniformity, misleading results could be obtained.

This same problem occurs with Bayesian parameter estimation [88]. When the samples are few, the dependency on priors is strong. As the number of points increases, the dependency on priors decreases. In the limiting case with full information, the prior distribution essentially does not influence the resulting posterior density at all. Similarly, the resulting prior issues in our setting will become critical only if, in the particular application, the model-based evidence is weak. This can occur, for example, with a noisy image in which the regions contain very few points. The issues of *how* noisy and *how* few depend on the model. For polynomial surface models these issues will be discussed in Chapter 5.

There will, in general, be three alternative notions of uniformity that pertain to our formalism. The first kind, termed *segmentation uniformity*, is simply the condition that

$$P(\{S_1\}) = P(\{S_2\}) \ \ \forall S_1, S_2 \in \Pi \tag{3.40}$$

or

$$P(\{S\}) = \frac{1}{|\Pi|} \ \ \forall S \in \Pi, \tag{3.41}$$

in which $|\Pi|$ is the number of possible segmentations. This appears to be the most natural definition of uniformity. The problem with guaranteeing such an initial condition is that it

requires enumerating Π before being able to determine the prior. The methods that have been discussed are aimed at avoiding this enumeration. Hence, segmentation uniformity is difficult to explicitly use; however, it serves as a reference for comparing other types of uniformity.

An alternative uniformity, which will be called *segment uniformity*, is that for any space Θ_i, generated with the procedure described in Section 3.5.2, we have

$$P(\{T\}) = \frac{1}{|\Theta_i|} \quad \forall T \in \Theta_i. \tag{3.42}$$

This states that each of the possible segments in a TSS has equal prior probability. This also seems natural; however, we must consider that for this to be equivalent to the segmentation uniformity, $f(\{T\})$, for various $T \in \Theta_i$, must contain the same number of elements. This is implied by the probability constraint (3.34). It will certainly not be true in general that the sizes of different $f(\{T\})$ are equal. For example, if $T = \mathcal{R}$ (contains all regions), there is only one element in $f(\{T\})$, because there is only one segmentation possible: all points grouped into a single region. But if $T = \{R_i\}$, for some $R_i \in \mathcal{R}$, there will be numerous segmentations that contain the segment $\{R_i\}$. Thus, with respect to the segmentation uniformity, the segment uniformity can be considered as a kind of bias.

The third and final type of uniformity that we consider is *membership uniformity*, which is based on the prior probability needed at each step of the refinement. Consider a binary sample space with two outcomes: $\{\tau(\{R_i, R_\rho\}, \emptyset), \tau(\{R_i, R_\rho\}, \emptyset)^C\}$. Recall that membership refers to the specific distinction between whether or not R_ρ is a member of the maximal homogeneous segment that includes R_i. When the IE-independent model is used, this is precisely the sample space representing the refinement. If we assume that each of these two outcomes has probability of $1/2$ as a prior, then we have membership uniformity. Under the IE-independent model, membership uniformity is achieved using the prior probability assignment of $1/2$ for each refinement. Under the IE-dependent

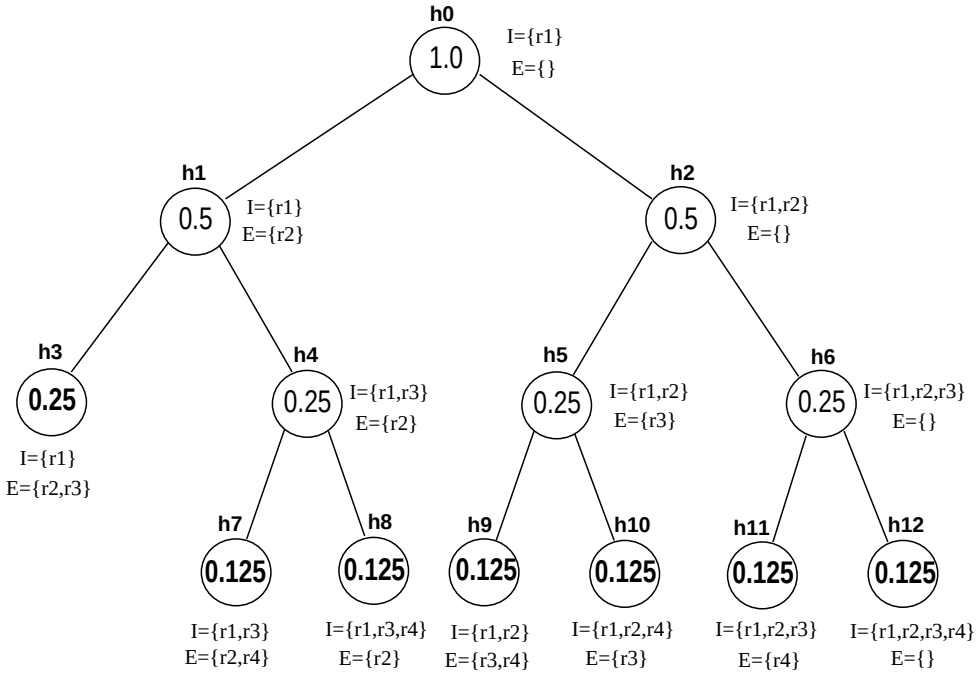


Figure 3.8 Probability distribution induced by membership probability of 0.5

model, a uniform prior can also be used with the interpretation that although the evidence is IE-dependent, the prior probability does not depend on knowledge about the other regions in I_ρ or E_ρ.

A natural question to ask is, "What does membership uniformity imply about the segment uniformity?" To see the relationship of membership uniformity to segment uniformity consider Figure 3.8. As long as the tree is height balanced, the two prior definitions are equivalent; however, this is not true in general. With this example, we see that h_3 obtains twice as much prior probability as the other ground events. With segment uniformity this would not be the case. In general, the membership-uniformity prior induces probabilities on the segment space which are off by a factor of two for every level in the tree that one ground event differs from the others.

Note that the depth in the tree directly corresponds to the total number of elements in the I and E sets of the node. Hence, ground events that are not deep in the tree tend to have fewer regions. Thus, membership uniformity is actually a bias toward segments with fewer regions. By changing membership probability, to $P(\tau(\{R_i, R_\rho\}, \emptyset)) = 3/4$ and $P(\tau(\{R_i, R_\rho\}, \emptyset)^C) = 1/4$ for instance, we can obtain a counterbias toward including more regions. Hence, at each refinement, the prior probability that R_ρ should be added to I_ρ is significantly higher than the probability that R_ρ should be added to E_ρ.

Experiments indicate that when evidence is strong, the bias due to priors is readily overcome. We have also observed that membership uniformity is usually closer to segmentation uniformity than it is to segment uniformity. This is due to the fact that segments with fewer regions (given higher prior probability) tend to cause more TSSs to be constructed than larger segments. The SSS probabilities are obtained from these individual TSSs using (3.39). If the number of segments n is larger, the prior SSS probability will tend to be smaller, compensating for the small-segment bias with respect to segment uniformity.

Table 3.5 Notation used in Chapter 3, with elements sorted by the order of their introduction.

Symbol	Definition	Section
D	The set of image elements	3.1
R	A connected subset of D	3.1
$\mathcal{R}$	The set of all regions	3.1
T	A segment (connected set of regions) from D	3.1
S	A segmentation of D (or partition of $\mathcal{R}$)	3.1
Π	Set of all segmentations that can be generated from $\mathcal{R}$	3.1
SSS	Segmentation Sample Space $(\Pi, \mathcal{A}, P)$	3.2
$\mathcal{A}$	Set of all subsets of Π (i.e., the events on the SSS)	3.2
A	An event on the SSS	3.2
TSS	Segment Sample Space $(\Theta_i, \mathcal{B}_i, P)$	3.3
R_i	An initial region for starting a TSS	3.3
Θ_i	The set of all segments that include R_i	3.3
$\mathcal{B}_i$	Set of all subsets of Θ_i (i.e., the events on a TSS)	3.3
B	An event on a TSS	3.3
I	Inclusion set, containing regions	3.3.1
E	Exclusion set, containing regions	3.3.1
$\tau(I, E)$	$\{T \in \Theta_i : I \subseteq T, E \cap T = \emptyset\}$	3.3.1
h	A node in the TSS tree, corresponding to a TSS event	3.3.1
C	A cover (a partition of Θ_i)	3.3.2
$\rho(C, B_\rho, R_\rho)$	The refinement mapping	3.3.2
R_ρ	A region chosen for the refinement	3.3.2
B_ρ	An event chosen for the refinement	3.3.2
P_I	Probability of including R_ρ	3.4.1
P_E	Probability of excluding R_ρ	3.4.1
e	A general variable denoting evidence	3.4.3

CHAPTER 4

BAYESIAN MEMBERSHIP PROBABILITY

4.1 General Introduction

This chapter presents a general statistical formalism, used to make probability assignments at a refinement step when constructing the TSS representation discussed in Chapter 3. Aside from the context developed in the previous chapter, this model can also be used as a Bayesian region-merging decision function. The membership probability refers to the posterior probability that the segment formed by adding the refinement region to I is homogeneous, when the original segment is given to be homogeneous. In other words, this is the probability that the refinement region is a member of the maximally homogeneous segment.

There are two interdependent[1] sets of random variables which are used in our statistical context. These represent the *parameter space* and the *observation space*. These spaces encompass a wide class of image models, and we will first describe some of their basic interpretations.

The parameter space directly captures the notion of homogeneity: every region has a parameter value (a point in the parameter space) associated with it, which is unknown to

[1]By two *interdependent* sets of random variables, we mean that a pair of random variables, formed by taking one from each set, are generally dependent.

the observer.[2] Two regions are defined to be homogeneous if they share the same parameter value. By using the homogeneity predicate, H, from Chapter 1, we can conceptually express the type of homogeneity considered in this chapter:

- $H(R) = true$ if and only if every subset of R has the same underlying parameter value associated with it.

The precise meaning of this underlying parameter value will be developed in this and subsequent sections.

As an example the parameter space could represent the coefficients of a polynomial surface describing image intensities, as in the model used by Silverman and Cooper [43]. If the parameter value is known for each region, an ideal segment is a maximal set of connected regions sharing the same parameter value. It is assumed that the observer (receiving information only from the image) does not know the associated parameter value for any of the regions. If it is possible to determine the parameter values for each of the regions, then the ideal segmentation can be trivially determined.

The observation space defines statistics that are functions of the image elements, and contain information about a region's parameter value. Although the parameter values are not known in general, a statistical model can be introduced which directly uses two probability density functions (pdf's) to determine the membership probability. These pdf's represent the prior model and the degradation model. The prior model is represented by a density on the parameter space (usually uniform), before any observations have been made. The degradation model is represented by a conditional density on the observation space, for each given parameter value. Intuitively, this density characterizes the observations we would expect to make, if the parameter value is given. In Chapter 5, the degradation model will represent the effects of Gaussian noise.

[2]The observer refers to the machine, which receives only the image data.

Each of these concepts has been used in similar contexts for image segmentation. In fact we have borrowed the terms *prior model* and *degradation model* from Geman [44]. In that context similar models are used for a Bayesian formalization of the MRF approach. Szeliski also defines a Bayesian model for MRFs, and terms what we call the degradation model, the *sensor model* [2].

Traditionally, a segmentation is determined by performing parameter estimation, and then grouping regions (or pixels) with similar parameter values into segments. It is important to note, however, that parameter estimation is not directly part of the segmentation goal. As discussed in Chapter 1, segmentation requires only a connected partition of the image. In contrast to traditional segmentation work, we directly use the densities on the observation and parameter spaces, rather than deriving the results from parameter estimates based on observation.

It may appear that we would want to perform parameter estimation for different regions and compare the resulting parameters to determine the probability of homogeneity. We have chosen *not* to perform estimation and are most concerned with the cases in which estimation is unreliable. These cases tend to yield the greatest amount of uncertainty when attempting to assess homogeneity, and are therefore the most interesting when building a probability distribution over varying segments. In Chapter 5, a fairly complex model (implicit polynomial surfaces with Gaussian noise) is introduced which uses the results discussed in this chapter, and a comparison is made in Section 5.2 with the estimation model for planar surfaces with noise.

Recall from Section 3.3.1 that an event on the TSS is represented by $\tau(I_\rho, E_\rho)$, in which I_ρ and E_ρ are the include and exclude sets. At each level of refinement some region, R_ρ, is selected and used to partition $\tau(I_\rho, E_\rho)$ into two refined events, $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and $\tau(I_\rho, E_\rho \cup \{R_\rho\})$. Recall from Section 3.4 that the goal is to determine the membership probability P_I (or P_E). Two models were introduced which yielded

different membership probability expressions: the general IE-dependent model, and the more specialized IE-independent model. The section concluded with some Bayesian expressions for the membership probability, given some evidence e. For P_I, the IE-independent model gave

$$P_I = P(\tau(\{R_i, R_\rho\}, \emptyset)|e). \tag{4.1}$$

Under the IE-dependent model, and some evidence e, the membership probability was

$$P_I = P(\tau(\{R_i, R_\rho\}, \emptyset)\ |\tau(I_\rho, E_\rho), e). \tag{4.2}$$

The event $\tau(\{R_i, R_\rho\}, \emptyset)$ will be used extensively throughout this chapter, and hence we will simplify its notation by letting $\tau_{i\rho} \equiv \tau(\{R_i, R_\rho\}, \emptyset)$. By using this notational convenience, the evidence-based membership probability (3.24) for the IE-independent model becomes

$$P_I = P(\tau_{i\rho}\ |e) = \frac{P(e\ |\tau_{i\rho})P(\tau_{i\rho})}{P(e)}. \tag{4.3}$$

For the IE-dependent model (3.28) the equivalent expression is

$$P_I = P(\tau_{i\rho}\ |\tau(I_\rho, E_\rho), e) = \frac{P(e\ |\tau_{i\rho}, \tau(I_\rho, E_\rho))P(\tau_{i\rho}, \tau(I_\rho, E_\rho))}{P(e, \tau(I_\rho, E_\rho))}. \tag{4.4}$$

Since $P_E = 1 - P_I$, it is sufficient to consider only P_I.

In the remaining sections we explicitly describe the model of evidence e, and how it applies to the membership probability. In the next section, the formal definitions of the random variables and densities are introduced. The IE-independent membership probability is derived in a general setting in Section 4.3. The most important expressions are (4.23) and (4.43). In Section 4.4, the IE-dependent model expressions are derived. The concepts are similar to those for the independent model, but the notation is more difficult. The reader may wish to browse through the derivations for the IE-dependent case, and make note of the resulting expressions (4.51) and (4.81). In some cases, it may be that estimation is reliable for one of the regions (e.g., R_i), and unreliable for

the other (R_ρ). In this case, some simplifications can be made, and the corresponding expressions are derived in Section 4.5. A natural extension of the membership probability to multiple-independent parameter/observation spaces is provided in Section 4.6. Again, the derivations are similar to those of the single IE-independent model, with the main result appearing in (4.102). In Section, 4.7 some expressions are briefly described, which correspond to discrete parameter and/or observation space versions of some of the earlier expressions. An illustrative example of the IE-independent model, along with some interpretations of the resulting membership probabilities, is presented in Section 4.8. Section 4.9 briefly presents a more complex instance of the Bayesian membership probability model, in an application to an MRF texture model.

4.2 Bayesian-Model Definitions

This section defines the observation space, parameter space, degradation model, and prior model. To be as general as possible, we begin with general probability space and random variable concepts, as found in [89]. The definitions presented here are critical for understanding the remaining portion of this chapter, although an understanding of the measure-theoretic notions is not necessary.

Recall that for each element, $D[i,j]$, of the image there is a vector of data that may contain 3D position, intensity, color, or other information, and let the dimension of this vector be l. For the n^{th} component of the vector at $D[i,j]$ we have a *probability triple*, $(\Omega_n, \mathcal{F}_n, P_n[i,j])$. In the probability triple, Ω_n represents the set of all possible outcomes of the n^{th} component. In practice we could, for instance, take Ω_n to be $\Re$ or some finite set of labels. We use $\mathcal{F}_n$ to denote a σ-algebra of measurable subsets of Ω_n, representing the *events*. If $\Omega_n = \Re$, then $\mathcal{F}_n$ could be the set of Borel-measurable sets on $\Re$. If Ω_n is finite, then $\mathcal{F}_n$ could be the power set of Ω_n.[3]

[3]This is the situation that occurred for the TSS and SSS definitions in Chapter 3.

The third element of the triple, $P_n[i,j]$, is the *probability mapping*, represented abstractly as $P_n[i,j] : \mathcal{F}_n \rightarrow \Re[0,1]$. This function assigns real-valued probabilities to each of the events in $\mathcal{F}_n$. We use the index $[i,j]$ to denote the fact that each image element can have a different probability map, but we assume that the n^{th} component has the same events for all of the image elements, and consequently we do not write $\mathcal{F}_n[i,j]$ or $\Omega_n[i,j]$. As will be observed shortly, we will not be required to directly know or estimate these probability mappings in order to determine the membership probability.

To simplify the discussion, we will reduce each of these probability triples to random variables, denoted by $X_n[i,j]$. In general, to define a random variable, we need a σ-measurable function, $h : \Omega_n \rightarrow \Re$. This corresponds to assigning real values to each of the outcomes, as done in elementary probability theory. If $\Omega_n = \Re$, then h can be taken as an identity map. We let $X_k[i,j]$ take on values denoted by $x_n[i,j]$, which are the images of elements of Ω_n under h. A given image D yields for each i and j:

$$D[i,j] = \{X_1[i,j] = x_1[i,j], X_2[i,j] = x_2[i,j], \ldots, X_l[i,j] = x_l[i,j]\}. \tag{4.5}$$

We next define the *parameter space* at some region, R_k, with a probability triple, $(\mathcal{U}, \mathcal{F}_\mathcal{U}, P_{\mathcal{U}_k})$. As before, $\mathcal{U}$ is a set of outcomes, $\mathcal{F}_\mathcal{U}$ is the set of events, and $P_{\mathcal{U}_k}$ is a probability mapping, unique to R_k. For simplicity, we will represent the parameter space as a finite vector of random variables, although, in general, this is not required. Denote the parameter random variables for some region R_k by $\mathbf{U_k} = [U_k^1\ U_k^2\ \ldots\ U_k^r]$. A vector value that can be taken on by $\mathbf{U_k}$ is denoted with $\mathbf{u_k}$.

The observation space is more difficult to define than the parameter space. We could consider representing the observation space as the set of all $X_n[i,j]$ in D. This implies that the observed image is used to directly determine the membership probabilities. We will develop a more general observation space in which transformations can be performed on the image variables to yield the observations, as opposed to directly using the image variables. This will be useful in applications in which it is impractical to consider large

numbers of random variables, and the information can be condensed into one or a few variables.

For each region R_k in the image, let W_k denote the set of all random variables at all of the points in R_k:

$$W_k = \bigcup_{D[i,j] \in R_k} \{X_1[i,j], \ldots, X_l[i,j]\}. \tag{4.6}$$

We will in general allow the observations to be obtained through functions of the variables in W_k and $\mathbf{U_k}$. This dependency on the parameter space will be explained shortly. In practice the observations may depend only on W_k.

We consider a set, Ψ_k, of measurable functions on the product space formed by taking the Cartesian products of the sample spaces for the elements of W_k, and $\mathcal{U}_k$. We could, for example, use only one function, representing the mean of the data variables. In Chapter 5, we will use a function to represent the sum-of-squares displacements of the points in a region to a surface determined by the parameter value. We will also introduce a model which applies the identity map to each of the variables in W_k. The general set of functions can be represented as

$$\Psi_k = [\psi_k^1(W_k, \mathbf{U_k}), \psi_k^2(W_k, \mathbf{U_k}), \ldots, \psi_k^d(W_k, \mathbf{U_k})]. \tag{4.7}$$

Each element of Ψ_k is a measurable function of random variables, and hence is a random variable. Therefore, we can use random variable notation to represent the observation space: $\mathbf{Y_k} = [Y_k^1\ Y_k^2 \ldots Y_k^d]$, in which superscripts are used for indexing. Hence we have $Y_k^n = \psi_k^1(W_k, \mathbf{U_k})$. Let $\mathbf{y_k}$ denote a vector of values that $\mathbf{Y_k}$ may take on. The vector $\mathbf{Y_k}$ will represent the *observation space* of region R_k.

In the discussion that follows, the random variables previously defined will be treated as continuous random variables, having corresponding probability density functions (pdf's). The results presented in the next four sections apply to general probability spaces equipped with Lebesgue integration; however, for simplicity, expressions will appear as

in standard probability texts. Some equivalent expressions for discrete random variables are given in Section 4.7.

The pdf's associated with the random variables $\mathbf{Y_k}$ and $\mathbf{U_k}$ will be utilized extensively. Let $p(\mathbf{y_k})$ be the joint pdf of $\mathbf{Y_k}$, and let $p(\mathbf{u_k})$ be the joint pdf of $\mathbf{U_k}$. Let $p(\mathbf{y_k}, \mathbf{u_k})$ denote the combined joint pdf. It is assumed that in general $p(\mathbf{y_k})$ and $p(\mathbf{u_k})$ are interdependent sets of random variables.[4] The combined joint pdf can be represented as

$$p(\mathbf{y_k}, \mathbf{u_k}) = p(\mathbf{y_k} \,|\mathbf{u_k}) p(\mathbf{u_k}). \tag{4.8}$$

The marginal pdf for $\mathbf{y_k}$ is then given by

$$p(\mathbf{y_k}) = \int p(\mathbf{y_k}|\mathbf{u_k}) p(\mathbf{u_k}) \mathbf{du_k}. \tag{4.9}$$

The pdf's appearing in the integrand above represent the basic premises from which the membership probability will be constructed.

The joint pdf in (4.8) consists of two components which correspond to the two densities previously indicated: $p(\mathbf{y_k}|\mathbf{u_k})$ is the *degradation model*, and $p(\mathbf{u_k})$ is the *prior model*.[5] Each of these components is specified as part of the Bayesian model needed to determine the membership probability. The prior model has been represented by a density on the parameter space, before any observations are made (this is typically taken to be a uniform density). The degradation model has been represented by a conditional density on the observation $\mathbf{y_k}$, given some parameter value $\mathbf{u_k}$. Given a model of degradation that pertains to the $X_n[i,j]$, the degradation model can be obtained though transformation of random variables.

Since the degradation model is given, if the parameter value of some region, e.g., R_1, is also given, the density on the observation space represented by $\mathbf{Y_1}$ is independent of

[4] By interdependent it is meant that the observation space variables depend on the parameter space variables.

[5] It will often be assumed that the prior parameter space pdf does not depend on a particular region, and will be denoted by $p(\mathbf{u})$.

any observation variables at other regions. Explicitly this means that for any two regions, such as R_1 and R_2,

$$p(\mathbf{y_1}|\mathbf{u_1}, \mathbf{y_2}) = p(\mathbf{y_1}|\mathbf{u_1}). \tag{4.10}$$

Above, $\mathbf{y_1}$ and $\mathbf{y_2}$ represent the observations from R_1 and R_2, respectively. Hence the degradation model densities are sufficient to determine the marginal density of $\mathbf{Y_k}$, given by (4.9), regardless of observations from other regions.

There is an important connection with the models introduced here and Bayesian parameter estimation. The random variables from some region R_k, namely, $X_1[i,j] \ldots X_l[i,j]$, can be interpreted as a training set. The variables $\mathbf{Y_k}$ can be interpreted as a set of *statistics* (i.e., functions of the training set). With Bayesian estimation, a *sufficient statistic* is a random variable that essentially represents all of the information present in the training set that pertains to the parameter [46],[88]. Explicitly using the notation presented, $\mathbf{y_k}$ is a sufficient statistic if

$$p(\mathbf{u_k}|\mathbf{y_k}, \mathbf{w_k}) = p(\mathbf{u_k}|\mathbf{y_k}). \tag{4.11}$$

Here, $\mathbf{w_k}$ represents an observed value for the set of random variables in W_k. This means that the observation of $\mathbf{y_k}$ contains as much information pertaining to the parameter density as all of $X_n[i,j]$ in region R_k.

The issues that apply in estimation theory also apply here. Recall that we do not perform parameter estimation, and only densities on this parameter space are considered. A sufficient statistic in our context means that all information available to determine the posterior density on the parameter space is present in the observation space.

The sufficient statistic issue will become important when the IE-dependent model is developed in Section 4.4. It will not be assumed in general that any of the functions determining $\mathbf{y_k}$ are required to be sufficient statistics. Section 5.3 presents an observation space that is not a sufficient statistic, and also one that is trivially a sufficient statistic.

4.3 *IE*-Independent Membership Probability

In this section we use the definitions of the previous section to derive an expression for the membership probability, given observations from R_ρ and R_i under the *IE*-independent model. We begin with the definitions of the parameter space, observation space, degradation model, prior model, and the refined event probabilities from Section 3.4. The result is an expression requiring three integrations on the parameter space, given by (4.23) and (4.43).

Recall that with the *IE*-independent model, only information obtained from R_ρ, the refinement region, and R_i, the initial region for Θ_i, is relevant to determining the membership probability. The vectors $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$ represent the observation spaces of R_ρ and R_i, respectively. In other words, the random variables $\mathbf{Y}_\rho$ correspond to applying the Ψ_ρ functions to the variables, $X_1[i,j], \ldots, X_l[i,j]$, which are taken from points belonging to R_ρ. Similarly, $\mathbf{Y_i}$ is obtained from R_i.

Note that $\tau_{i\rho}$ and $\tau_{i\rho}^C$ are two important events that affect the interdependence of $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$. These explicitly are

$$\tau_{i\rho} = \{T \in \Theta_i : R_\rho \in T\} \tag{4.12}$$

and

$$\tau_{i\rho}^C = \{T \in \Theta_i : R_\rho \notin T\}. \tag{4.13}$$

These events are depicted in Figures 4.1 and 4.2. The dashed-line for T indicates that it can vary over all of the segments in $\tau_{i\rho}$ (or $\tau_{i\rho}^C$). The difference between these two events can be regarded as a difference in the assumption about the boundary of T. Since it is always true that $R_i \in T$, the event $\tau_{i\rho}$ implies that the data elements in R_ρ and R_i share the same (but unknown) parameter vector. The event $\tau_{i\rho}^C$ implies the opposite, that R_ρ and R_i have distinct (unknown) parameter values.

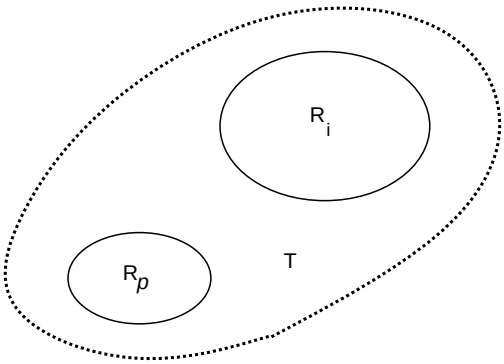


Figure 4.1 An element, T, of the event $\tau_{i\rho}$.

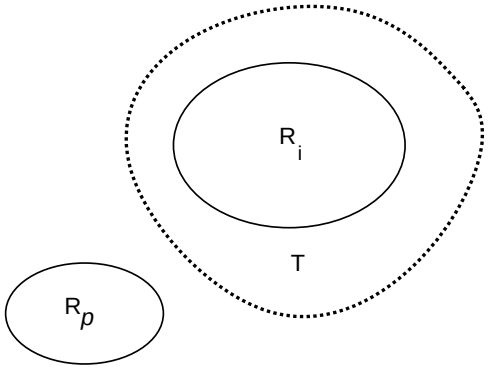


Figure 4.2 An element, T, of the event $\tau_{i\rho}^{C}$.

The parameter space captures an important dependency between observations made from R_ρ and observations made from R_i. The two events $\tau_{i\rho}$ and $\tau_{i\rho}^{C}$ will induce different conditional densities on the observation space of R_ρ. This difference between the two will ultimately lead to a formulation of the Bayesian posterior membership probability.

The functions Ψ_ρ and Ψ_i, applied to the image elements yield observations of the random variables $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$. These serve as the evidence used to determine the membership probability, which is represented as

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i}). \tag{4.14}$$

There is a natural dependency between $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$ when the event $\tau_{i\rho}$ is given, since $R_\rho \cup R_i$ hypothesized to be homogeneous (since each region belongs to the inclusion set). We can apply Bayes' rule on $\tau_{i\rho}$ and $\mathbf{y}_\rho$, while leaving $\mathbf{y_i}$ as a condition:[6]

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i}) = \frac{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})P(\tau_{i\rho}|\mathbf{y_i})}{p(\mathbf{y}_\rho|\mathbf{y_i})}. \tag{4.15}$$

The denominator can be expanded to yield

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i}) = \frac{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})P(\tau_{i\rho}|\mathbf{y_i})}{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})P(\tau_{i\rho}|\mathbf{y_i}) + p(\mathbf{y}_\rho|\tau_{i\rho}^C, \mathbf{y_i})P(\tau_{i\rho}^C|\mathbf{y_i})}. \tag{4.16}$$

Note that the numerator and the first term of the denominator are identical. The only difference between the two terms in the denominator is the condition $\tau_{i\rho}$ or $\tau_{i\rho}^C$. The denominator is the standard normalizing factor from Bayes' rule, over the binary sample space, $\{\tau_{i\rho}, \tau_{i\rho}^C\}$.

Before proceeding further, the prior membership probability, $P(\tau_{i\rho})$, must be specified, corresponding to the probability before values for $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$ are observed. Recall that Section 3.6 discussed the issue of membership priors, stating that in the presence of no evidence some membership probability must be assigned. In this chapter we will denote the prior membership probability as P_0.

We note that the two probabilities in the denominator of (4.16) can be expressed in terms P_0. We first note that $\tau_{i\rho}$ is marginally independent of $\mathbf{Y_i}$, but $\tau_{i\rho}$ is conditionally dependent on $\mathbf{Y_i}$ given the observation $\mathbf{y}_\rho$, because, in the absence of any information about the region R_ρ (which would be contained in the observed $\mathbf{y}_\rho$), the membership probability will be unaffected when given the observation from R_i. In other words, before $\mathbf{y}_\rho$ has been observed, the observation of $\mathbf{Y_i}$ has no effect on the membership probability. In (4.16) we can therefore set

$$P(\tau_{i\rho}|\mathbf{y_i}) = P(\tau_{i\rho}) = P_0. \tag{4.17}$$

[6]The mixture of discrete events and densities in Bayes' rule does not present a problem. See for instance Stark and Woods [90].

Similarly for the negation,

$$P(\tau_{i\rho}^{C}|\mathbf{y_i}) = P(\tau_{i\rho}^{C}) = 1 - P_0. \tag{4.18}$$

By substituting these into (4.16), the expression becomes

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i}) = \frac{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})P_0}{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})P_0 + p(\mathbf{y}_\rho|\tau_{i\rho}^{C}, \mathbf{y_i})[1 - P_0]}. \tag{4.19}$$

Using the fact that $a/(a+b)$ is equal to $(1 + b/a)^{-1}$ for $0 < a, b < \infty$, the resulting expression is

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i}) = \left[1 + \frac{(1 - P_0)\ p(\mathbf{y}_\rho|\tau_{i\rho}^{C}, \mathbf{y_i})}{P_0\ p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})}\right]^{-1}. \tag{4.20}$$

The expression above can be decomposed into a function of two ratios; one depends only on the prior membership probability, and the other depends on the events $\tau_{i\rho}$ and $\tau_{i\rho}^{C}$ and the observation variables. Let λ_0 denote the ratio based only on priors, and let $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ denote the ratio representing the effects of the observation. With (4.20) these become

$$\lambda_0 \equiv \frac{1 - P_0}{P_0} \tag{4.21}$$

and

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) \equiv \frac{p(\mathbf{y}_\rho|\tau_{i\rho}^{C}, \mathbf{y_i})}{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_i})}. \tag{4.22}$$

By using these definitions, the membership probability expression becomes

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i}) = \frac{1}{1 + \lambda_0\ \lambda_1(\mathbf{y}_\rho, \mathbf{y_i})}. \tag{4.23}$$

The λ_0 and $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ ratios represent an interesting decomposition of the factors contributing to the membership probability. The range of values of these ratios is restricted to $0 \leq \lambda_0 < \infty$ and $0 \leq \lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) < \infty$. When either takes on the value of 1, the ratio essentially does not bias the posterior membership probability. For example, first consider what happens when $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = 1$. The posterior membership probability remains the same as the prior membership probability. If the prior is 1/3, then λ_0 is

2, and the resulting posterior membership probability is 1/3. As λ_0 approaches 0, the posterior membership probability approaches 1. Alternatively, as λ_0 approaches infinity, the posterior membership probability approaches 0. Also, $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ behaves in a similar manner. Consider the case in which λ_0 is 1. This corresponds to a membership prior of 1/2, representing membership uniformity as discussed in Section 3.6. The evidence will guide the ratio $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$; if the event $\tau_{i\rho}^C$ causes a larger pdf value to be obtained for the observation $\mathbf{y}_\rho$ than the event $\tau_{i\rho}$, then by (4.22) the ratio $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ will be greater than one. This causes a decrease in the membership posterior which is expected since the event $\tau_{i\rho}^C$ is more favorable. Analogous behavior occurs when the event $\tau_{i\rho}$ is favorable over $\tau_{i\rho}^C$. This decomposition of the membership probability into contributing factors corresponds to our intuitive expectations.

The next goal is to express the ratio $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ in terms of the given model densities $p(\mathbf{u_k})$ and $p(\mathbf{y_k}|\mathbf{u_k})$. Together these two will be combined in (4.23) to give the resulting membership posterior probability. The ratio λ_0 is simply defined using the membership prior in (4.21), leaving the rest of this section to determining $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$.

Consider (4.9), conditioned on some event A. The expression becomes

$$p(\mathbf{y}|A) = \int p(\mathbf{y}|\mathbf{u}, A)p(\mathbf{u}|A)\mathbf{du}. \tag{4.24}$$

We can use (4.24) with $A = \{\tau_{i\rho}^C, \mathbf{y_i}\}$ in the numerator and $A = \{\tau_{i\rho}, \mathbf{y_i}\}$ in the denominator of (4.22), and the expression (4.22) is expanded into

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho, \tau_{i\rho}^C, \mathbf{y_i})p(\mathbf{u}_\rho|\tau_{i\rho}^C, \mathbf{y_i})\mathbf{du}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho, \tau_{i\rho}, \mathbf{y_i})p(\mathbf{u}_\rho|\tau_{i\rho}, \mathbf{y_i})\mathbf{du}_\rho}. \tag{4.25}$$

The integrand in the numerator of (4.25) represents the joint density of $\mathbf{Y}_\rho$ and $\mathbf{U}_\rho$ under the condition $\{\tau_{i\rho}^C, \mathbf{y_i}\}$. The integral is taken over the parameter space, yielding a marginal density (with respect to $\mathbf{U}_\rho$) of $\mathbf{Y}_\rho$. Similarly, the integrand in the denominator of (4.25) represents the joint density under the condition $\{\tau_{i\rho}, \mathbf{y_i}\}$, which is integrated to give another marginal density on $\mathbf{Y}_\rho$.

The terms $p(\mathbf{y}_\rho|\mathbf{u}_\rho, \tau_{i\rho}, \mathbf{y_i})$ and $p(\mathbf{y}_\rho|\mathbf{u}_\rho, \tau_{i\rho}^C, \mathbf{y_i})$, appearing in (4.25), can be reduced. Recall that $p(\mathbf{y}_\rho|\mathbf{u}_\rho)$ is given. By using (4.10), we observe that $\mathbf{y_i}$ can be removed from the conditions. Since there is no longer a dependency on information from R_ρ, the events $\tau_{i\rho}$ and $\tau_{i\rho}^C$ can also be removed. These statements are equivalent to asserting that nothing is learned about the degradation model when observations are made from other regions having the same given parameter value. Essentially, all other information is irrelevant once the parameter value for $\mathbf{u}_\rho$ is given. These observations result in

$$p(\mathbf{y}_\rho|\mathbf{u}_\rho, \tau_{i\rho}, \mathbf{y_i}) = p(\mathbf{y}_\rho|\mathbf{u}_\rho) \tag{4.26}$$

and

$$p(\mathbf{y}_\rho|\mathbf{u}_\rho, \tau_{i\rho}^C, \mathbf{y_i}) = p(\mathbf{y}_\rho|\mathbf{u}_\rho). \tag{4.27}$$

Consequently, $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho|\tau_{i\rho}^C, \mathbf{y_i})\mathbf{du}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho|\tau_{i\rho}, \mathbf{y_i})\mathbf{du}_\rho}. \tag{4.28}$$

Using Bayes' rule, under the condition $\tau_{i\rho}$, the following replacement can be made in the denominator of (4.25):

$$p(\mathbf{u}_\rho|\tau_{i\rho}, \mathbf{y_i})\mathbf{du}_\rho = p(\mathbf{u}_\rho|\mathbf{y_i}, \tau_{i\rho})\mathbf{du}_\rho = \frac{p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})p(\mathbf{u}_\rho|\tau_{i\rho})\mathbf{du}_\rho}{p(\mathbf{y_i}|\tau_{i\rho})}. \tag{4.29}$$

Similarly, the term in the numerator can be replaced using Bayes' rule under the condition $\tau_{i\rho}^C$

$$p(\mathbf{u}_\rho|\tau_{i\rho}^C, \mathbf{y_i})\mathbf{du}_\rho = \frac{p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C)p(\mathbf{u}_\rho|\tau_{i\rho}^C)\mathbf{du}_\rho}{p(\mathbf{y_i}|\tau_{i\rho}^C)}. \tag{4.30}$$

The only information obtained from R_ρ and R_i comes from the observation of the random variables $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$. Hence, if the condition is $\tau_{i\rho}$, without $\mathbf{y_i}$ or $\mathbf{y}_\rho$, no evidence is present to influence the density of $\mathbf{U}_\rho$. This similarly holds for the $\tau_{i\rho}^C$ case, allowing some reductions:

$$p(\mathbf{u}_\rho|\tau_{i\rho}) = p(\mathbf{u}_\rho) \tag{4.31}$$

and

$$p(\mathbf{u}_\rho|\tau_{i\rho}^C) = p(\mathbf{u}_\rho). \tag{4.32}$$

Here $p(\mathbf{u}_\rho)$ denotes the noninformative prior density on the parameter space for R_ρ when no relevant information is present.

For similar reasons some reductions can be made to the denominators of (4.29) and (4.30). Since no information has been obtained from R_ρ (via $\mathbf{Y}_\rho$), $\mathbf{Y_i}$ does not depend on $\tau_{i\rho}$. The event $\tau_{i\rho}$ only hypothesizes that R_i and R_ρ have the same underlying parameter value, but provides no information about R_ρ. This observation yields

$$p(\mathbf{y_i}|\tau_{i\rho}) = p(\mathbf{y_i}) \tag{4.33}$$

and

$$p(\mathbf{y_i}|\tau_{i\rho}^C) = p(\mathbf{y_i}). \tag{4.34}$$

When the reductions of (4.31), (4.32), (4.33), and (4.34) are substituted back into (4.29) and (4.30), the expressions become

$$p(\mathbf{u}_\rho|\mathbf{y_i}, \tau_{i\rho})\mathbf{du}_\rho = \frac{p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})p(\mathbf{u}_\rho)\mathbf{du}_\rho}{p(\mathbf{y_i})} \tag{4.35}$$

and

$$p(\mathbf{u}_\rho|\mathbf{y_i}, \tau_{i\rho}^C)\mathbf{du}_\rho = \frac{p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{p(\mathbf{y_i})}. \tag{4.36}$$

These Bayesian expansions may be substituted back into (4.25) to yield

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)\dfrac{p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C)p(\mathbf{u}_\rho)}{p(\mathbf{y_i})}\mathbf{du}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)\dfrac{p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})p(\mathbf{u}_\rho)}{p(\mathbf{y_i})}\mathbf{du}_\rho}. \tag{4.37}$$

Note that $p(\mathbf{y_i})$ does not depend on the variables of integration in either the numerator or denominator above, hence, they can be removed, and cancelled to yield

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})p(\mathbf{u}_\rho)\mathbf{du}_\rho}. \tag{4.38}$$

This form clearly displays the influence of the membership events $\tau_{i\rho}$ and $\tau_{i\rho}^C$. Both of the integrals are taken over the parameter space of R_ρ. The distinction between the numerator and denominator is, as expected, the condition $\tau_{i\rho}^C$ or $\tau_{i\rho}$. Two densities still remain that were not given as part of the model. These are $p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C)$ and $p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})$. It will turn out that the denominator term can be expressed in terms of the given model densities; however, the numerator term, in this general setting would require further specification of the model. It is unclear how the observation variables $\mathbf{Y_i}$ should be affected when the parameter value is given for some other region that is not even related R_i (due to the given event $\tau_{i\rho}^C$). It seems natural to make the assumption that knowing the parameter value for one region is independent from the observation variables from some other regions, given to belong to some other segment.

If a model is introduced in which parameter densities of adjacent segments (which are assumed to have distinct parameter values) are dependent, this is the natural place to model it (i.e., by defining $p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C)$ in (4.38)). For instance, if the parameter space represents surface normals to a plane, there may be knowledge that all surfaces corresponding to segments in the scene meet at right angles. An assumption such as this would cause some expectation about the parameter densities of neighboring segments if the parameter value for some segment is known. From this point onward we preclude models of this type, since they require higher-level image models than we wish to consider at this stage.

Using the independence assumption just discussed, we obtain

$$p(\mathbf{y_i}) = p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho}^C). \tag{4.39}$$

Using this, (4.37) becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{y_i})p(\mathbf{u}_\rho)\mathbf{d}\mathbf{u}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})p(\mathbf{u}_\rho)\mathbf{d}\mathbf{u}_\rho}. \tag{4.40}$$

Again, $p(\mathbf{y_i})$ can be removed from the numerator integrand to obtain

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{p(\mathbf{y_i}) \int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{y_i}|\mathbf{u}_\rho, \tau_{i\rho})p(\mathbf{u}_\rho)\mathbf{du}_\rho}. \tag{4.41}$$

By using (4.9), $p(\mathbf{y_i})$ can be expressed as

$$p(\mathbf{y_i}) = \int p(\mathbf{y_i}|\mathbf{u_i})p(\mathbf{u_i})\mathbf{du_i}. \tag{4.42}$$

The condition $\{\mathbf{u}_\rho, \tau_{i\rho}\}$ in the denominator of (4.40) can be rewritten. This condition means that R_ρ and R_i have the same underlying parameter value, and the parameter value for R_ρ is given. Hence it follows that, given $\tau_{i\rho}$, we have $\mathbf{u_i} = \mathbf{u}_\rho$. If the prior model is assumed not to be region dependent, then we can write $p(\mathbf{u}_\rho) = p(\mathbf{u_i}) = p(\mathbf{u})$. This yields

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho\right]\left[\int p(\mathbf{y_i}|\mathbf{u_i})p(\mathbf{u_i})\mathbf{du}_i\right]}{\int p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})\mathbf{du}}. \tag{4.43}$$

This expression, in conjunction with λ_0, will give the Bayesian membership proability $P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i})$ using (4.23).

An interesting interpretation can be derived by treating the pdf's in the integrals as random variables. Define a random variable Z_k as a measurable function on the parameter space $\mathbf{u}$. Define this function to be precisely the value of the conditional pdf $p(\mathbf{y_k}|\mathbf{u_k})$. Hence, Z_k is the random variable corresponding the marginal density $p(\mathbf{y_k}|\mathbf{u_k})$. The components of (4.43) can be written as expectations of Z_ρ and Z_i by

$$\int p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{u})\mathbf{du} = \int Zp(\mathbf{u})\mathbf{du} = E[Z_\rho], \tag{4.44}$$

$$\int p(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})\mathbf{du} = E[Z_i], \tag{4.45}$$

and

$$\int p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})\mathbf{du} = E[Z_\rho\ Z_i]. \tag{4.46}$$

By using these definitions, (4.43) becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{E[Z_\rho]\ E[Z_i]}{E[Z_\rho\ Z_i]}. \tag{4.47}$$

The definition of independence for Z_n and Z_i is

$$E[Z_\rho]\ E[Z_i] = E[Z_\rho\ Z_i]. \tag{4.48}$$

When this occurs, $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = 1$, corresponding to the case of no information. Recall that when $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = 1$, the priors, represented by λ_0, completely determine the resulting membership probability. This is precisely what we would expect; if the observations made from R_ρ and R_i are independent of the assumption that $\tau_{i\rho}$ or $\tau_{i\rho}^C$, then it is natural to expect that the priors will completely determine the probability. This result indicates that (4.43) is similar to correlation.

4.4 IE-Dependent Membership Probability

Recall that with the IE-dependent model, the probability of homogeneity is expressed as

$$P_I = P(\tau_{i\rho}\ |\tau(I_\rho, E_\rho), e). \tag{4.49}$$

This expression is similar to that for the IE-independent model, with the only difference being the condition $\tau(I_\rho, E_\rho)$. We would therefore expect some similarities between the derivation of this membership probability and that for the IE-independent model, and, indeed, this is the case.

Due to the added complexity of the IE-dependent model, some new issues arise. To begin with, we will retain the assumption that the excluded regions E_ρ do not affect the observation variables. In the most general model, these regions may be considered.

With the IE-independent model, only R_i and R_ρ were used to influence the parameter space density. With the IE-dependent model, all of the regions that belong to I_ρ will be

used. It may be possible that the models are formulated in such a way that is possible to create an observation variable obtained from the region given by

$$R_I \equiv \bigcup_{R_k \in I_\rho} R_k. \tag{4.50}$$

In other words there are $\mathbf{Y_I}$ and $\mathbf{U_I}$ for R_I, with a given conditional pdf $p(\mathbf{y_I}|\mathbf{u_I})$.

Alternatively, we may treat each region in I_ρ individually, with each region having its own set of observation variables. The observations from each region in I_ρ, together with $\mathbf{y}_\rho$ can be used to determine the posterior membership probability.

These two alternatives (creating one large region or considering individual observation spaces), in general, produce different results. We describe the situation in which the two coincide by again considering the notion of a sufficient statistic, which was introduced in Section 4.2. The two alternatives coincide when both $\mathbf{Y_I}$ and the observation variables from each of the regions in I_ρ are sufficient statistics. When this occurs, the same information regarding the parameter density will be obtained whether observations are made from each region individually, or one observation is made from the union of the regions. Since each individual region observation is equivalent to using all of the points in I_ρ directly, the resulting parameter densities will be the same for both, yielding identical results for either approach. In Section 5.4.1.2, this difference is shown for particular observation spaces and the implicit polynomial surface model.

For the case in which one single region R_I is constructed, the IE- independent model may be applied directly, with R_i being replaced with R_I. This results in the evidence-based ratio,

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_I}) = \frac{\left[\int p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{u})\mathbf{du}\right]\left[\int p(\mathbf{y_I}|\mathbf{u})p(\mathbf{u})\mathbf{du}\right]}{\int p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y_I}|\mathbf{u})p(\mathbf{u})\mathbf{du}}. \tag{4.51}$$

The situation is more interesting when individual observations are considered from each of the regions in I_ρ. Recall that with the IE-independent model, only information obtained from R_ρ and R_i is relevant to determining the membership probability. For the

IE-dependent case, all of the regions in I_ρ become relevant. Let $I_\rho = \{R_1, \ldots, R_m\}$ be the region names for the included regions. The random variables obtained from these regions using $\Psi_1, \ldots, \Psi_m$ are $\mathbf{Y_1}, \ldots, \mathbf{Y_m}$. It is assumed again that all information to be considered is represented by these variables. All of these observations, together, are the evidence that is used to determine the membership probability. This is represented as

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}, \tau(I_\rho, E_\rho)). \tag{4.52}$$

To simplify some of the notation, it will be assumed that the condition $\tau(I_\rho, E_\rho)$ is always given. Hence, from this point onward, the assumption will be utilized, but not expressed in the conditionals.

Figures 4.3 and 4.4 depict the membership events for the dependent model. For $\tau_{i\rho}$ in this model, it is assumed that I_ρ (which includes R_i) and R_ρ all belong to any segment, T, in $\tau_{i\rho} \cap \tau(I_\rho, E_\rho)$. It is also assumed that none of the regions in E_ρ belong to $T \in \tau_{i\rho} \cap \tau(I_\rho, E_\rho)$. The second figure is similar to the first, with the difference being the condition $\tau_{i\rho}^C$ replacing $\tau_{i\rho}$.

We can again apply Bayes' rule to expand the membership probability, giving

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) P(\tau_{i\rho}|\mathbf{y_1}, \ldots, \mathbf{y_m})}{p(\mathbf{y}_\rho|\mathbf{y_1}, \ldots, \mathbf{y_m})}. \tag{4.53}$$

Expanding the denominator as before yields

$$P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m})$$

$$= \frac{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) P(\tau_{i\rho}|\mathbf{y_1}, \ldots, \mathbf{y_m})}{p(\mathbf{y}_\rho|\tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) P(\tau_{i\rho}|\mathbf{y_1}, \ldots, \mathbf{y_m}) + p(\mathbf{y}_\rho|\tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m}) P(\tau_{i\rho}^C|\mathbf{y_1}, \ldots, \mathbf{y_m})}. \tag{4.54}$$

Once again, the numerator and the first term of the denominator are identical. The only difference between the two terms in the denominator is the event $\tau_{i\rho}$ or $\tau_{i\rho}^C$.

It is again assumed that the membership prior is given. In (4.53) we can set

$$P(\tau_{i\rho}|\mathbf{y_1}, \ldots, \mathbf{y_m}) = P(\tau_{i\rho}) \equiv P_0, \tag{4.55}$$

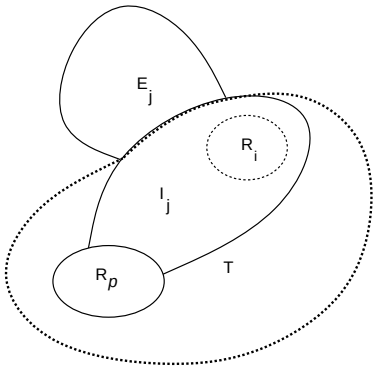


Figure 4.3 An element, T, of the event $\tau_{i\rho}$, given $\tau(I_\rho, E_\rho)$.

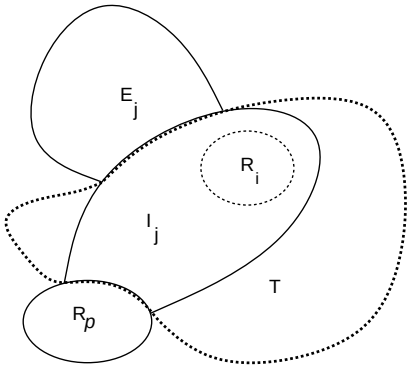


Figure 4.4 An element, T, of the event $\tau_{i\rho}^C$, given $\tau(I_\rho, E_\rho)$.

since no observation has been made from R_ρ. Similarly for the negation

$$P(\tau_{i\rho}^C|\mathbf{y_1},\ldots,\mathbf{y_m}) = P(\tau_{i\rho}^C) = 1 - P_0. \tag{4.56}$$

Substituting these into (4.53), the expression becomes

$$P(\tau_{i\rho}|\mathbf{y}_\rho,\mathbf{y_1},\ldots,\mathbf{y_m}) = \frac{p(\mathbf{y}_\rho|\tau_{i\rho},\mathbf{y_1},\ldots,\mathbf{y_m})P_0}{p(\mathbf{y}_\rho|\tau_{i\rho},\mathbf{y_1},\ldots,\mathbf{y_m})P_0 + p(\mathbf{y}_\rho|\tau_{i\rho}^C,\mathbf{y_1},\ldots,\mathbf{y_m})[1-P_0]}. \tag{4.57}$$

Again, using the fact that $a/(a+b)$ is equal to $(1+b/a)^{-1}$ for $0 < a, b < \infty$, the resulting expression is

$$P(\tau_{i\rho}|\mathbf{y}_\rho,\mathbf{y_1},\ldots,\mathbf{y_m}) = \left[1 + \frac{(1-P_0)\; p(\mathbf{y}_\rho|\tau_{i\rho}^C,\mathbf{y_1},\ldots,\mathbf{y_m})}{P_0\; p(\mathbf{y}_\rho|\tau_{i\rho},\mathbf{y_1},\ldots,\mathbf{y_m})}\right]^{-1}. \tag{4.58}$$

Define λ_0 as before, representing a ratio based entirely on priors (4.21). Define $\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m})$ as the ratio representing the effects of the model-based evidence. This results in

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) \equiv \frac{p(\mathbf{y}_\rho | \tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m})}{p(\mathbf{y}_\rho | \tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m})}, \tag{4.59}$$

and consequently

$$P(\tau_{i\rho} | \mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{1}{1 + \lambda_0 \ \lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m})}. \tag{4.60}$$

The λ_0 and $\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m})$ ratios represent the same decomposition of the factors contributing to the membership probability. By using (4.24) with $A = \{\tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m}\}$ in the numerator and $A = \{\tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}\}$ in the denominator of (4.59), the expression (4.59) is expanded into

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{\int p(\mathbf{y}_\rho | \mathbf{u}_\rho, \tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m}) p(\mathbf{u}_\rho | \tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m}) \mathbf{du}_\rho}{\int p(\mathbf{y}_\rho | \mathbf{u}_\rho, \tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) p(\mathbf{u}_\rho | \tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) \mathbf{du}_\rho}. \tag{4.61}$$

The integrands in (4.61) above represent the joint density $p(\mathbf{y}_\rho, \mathbf{u}_\rho)$ corresponding to region R_ρ, under the two different conditions. The $\mathbf{u}_\rho$ is integrated out to give the marginal density of $\mathbf{Y}_\rho$ under the remaining conditions.

The terms $p(\mathbf{y}_\rho | \mathbf{u}_\rho, \tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m})$ and $p(\mathbf{y}_\rho | \mathbf{u}_\rho, \tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m})$ can be simplified. Since the joint pdf $p(\mathbf{y}_\rho, \mathbf{u}_\rho)$ is essentially given, $\mathbf{Y}_\rho$ is independent of $\tau_{i\rho}$ and $\mathbf{y_1}, \ldots, \mathbf{y_m}$ when $\mathbf{u}_\rho$ is given. Also, $\mathbf{Y}_\rho$ is independent of $\tau_{i\rho}^C$ and $\mathbf{y_1}, \ldots, \mathbf{y_m}$ when $\mathbf{u}_\rho$ is given. The simplification is

$$p(\mathbf{y}_\rho | \mathbf{u}_\rho, \tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) = p(\mathbf{y}_\rho | \mathbf{u}_\rho) \tag{4.62}$$

and

$$p(\mathbf{y}_\rho | \mathbf{u}_\rho, \tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m}) = p(\mathbf{y}_\rho | \mathbf{u}_\rho). \tag{4.63}$$

Then, $\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m})$ becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{\int p(\mathbf{y}_\rho | \mathbf{u}_\rho) p(\mathbf{u}_\rho | \tau_{i\rho}^C, \mathbf{y_1}, \ldots, \mathbf{y_m}) \mathbf{du}_\rho}{\int p(\mathbf{y}_\rho | \mathbf{u}_\rho) p(\mathbf{u}_\rho | \tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m}) \mathbf{du}_\rho}. \tag{4.64}$$

Using Bayes' rule, under the condition $\tau_{i\rho}$, the following replacement can be made in the denominator of (4.61):

$$p(\mathbf{u}_\rho|\mathbf{y_1},\ldots,\mathbf{y_m},\tau_{i\rho})\mathbf{du}_\rho = \frac{p(\mathbf{y_1},\ldots,\mathbf{y_m}|\mathbf{u}_\rho,\tau_{i\rho})p(\mathbf{u}_\rho|\tau_{i\rho})\mathbf{du}_\rho}{p(\mathbf{y_1},\ldots,\mathbf{y_m}|\tau_{i\rho})}. \tag{4.65}$$

Similarly, the term in the numerator can be replaced using Bayes' rule under the condition $\tau_{i\rho}^C$, giving

$$p(\mathbf{u}_\rho|\mathbf{y_1},\ldots,\mathbf{y_m},\tau_{i\rho}^C)\mathbf{du}_\rho = \frac{p(\mathbf{y_1},\ldots,\mathbf{y_m}|\mathbf{u}_\rho,\tau_{i\rho}^C)p(\mathbf{u}_\rho|\tau_{i\rho}^C)\mathbf{du}_\rho}{p(\mathbf{y_1},\ldots,\mathbf{y_m}|\tau_{i\rho}^C)}. \tag{4.66}$$

We can use the observations from (4.31) and (4.32) to simplify the parameter space densities. For similar reasons as for (4.33) and (4.34), another simplification can be made:

$$p(\mathbf{y_1},\ldots,\mathbf{y_m}|\tau_{i\rho}) = p(\mathbf{y_1},\ldots,\mathbf{y_m}) \tag{4.67}$$

and

$$p(\mathbf{y_1},\ldots,\mathbf{y_m}|\tau_{i\rho}) = p(\mathbf{y_1},\ldots,\mathbf{y_m}). \tag{4.68}$$

When these expressions are substituted back into (4.65) and (4.66) we obtain

$$p(\mathbf{u}_\rho|\tau_{i\rho},\mathbf{y_1},\ldots,\mathbf{y_m})\mathbf{du}_\rho = \frac{p(\mathbf{y_1},\ldots,\mathbf{y_m}|\mathbf{u}_\rho,\tau_{i\rho})p(\mathbf{u}_\rho)\mathbf{du}_\rho}{p(\mathbf{y_1},\ldots,\mathbf{y_m})} \tag{4.69}$$

and

$$p(\mathbf{u}_\rho|\tau_{i\rho}^C,\mathbf{y_1},\ldots,\mathbf{y_m})\mathbf{du}_\rho = \frac{p(\mathbf{y_1},\ldots,\mathbf{y_m}|\mathbf{u}_\rho,\tau_{i\rho}^C)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{p(\mathbf{y_1},\ldots,\mathbf{y_m})}. \tag{4.70}$$

Up until now the derivation has been, for the most part, the same as that for the independent case. Here is where a fundamental difference enters. Since $\mathbf{u}_\rho$ is given along with $\tau_{i\rho}$ and $\tau(I_\rho,E_\rho)$, the parameter values are given for each of the regions in I_ρ. This means that the common parameter value is given for all of the regions in I_ρ and R_ρ, hence the $\mathbf{y_k}$ are conditionally independent. Explicitly this is

$$p(\mathbf{y_k}|\mathbf{u}_\rho,\tau_{i\rho},\mathbf{y}_h) = p(\mathbf{y_k}|\mathbf{u}_\rho,\tau_{i\rho}) \;\; \forall k,h \text{ with } k \neq h. \tag{4.71}$$

From this, we have

$$p(\mathbf{y_1}, \ldots, \mathbf{y_m}|\mathbf{u}_\rho, \tau_{i\rho}) = \prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u}_\rho, \tau_{i\rho}). \tag{4.72}$$

This allows (4.69) to be rewritten as

$$p(\mathbf{u}_\rho|\tau_{i\rho}, \mathbf{y_1}, \ldots, \mathbf{y_m})\mathbf{du}_\rho = \frac{\left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u}_\rho, \tau_{i\rho})\right] p(\mathbf{u}_\rho)\mathbf{du}_\rho}{p(\mathbf{y_1}, \ldots, \mathbf{y_m})}. \tag{4.73}$$

The Bayesian expansions, (4.69) and (4.70), may be substituted back into (4.61) to yield

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)\dfrac{p(\mathbf{y_1}, \ldots, \mathbf{y_m}|\mathbf{u}_\rho, \tau_{i\rho}^C)p(\mathbf{u}_\rho)}{p(\mathbf{y_1}, \ldots, \mathbf{y_m})}\mathbf{du}_\rho}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)\dfrac{\left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u}_\rho, \tau_{i\rho})\right] p(\mathbf{u}_\rho)}{p(\mathbf{y_1}, \ldots, \mathbf{y_m})}\mathbf{du}_\rho}. \tag{4.74}$$

The assumption is made here, as before, that knowledge of the parameter value for an excluded region does not affect the observation random variable. In this context, the assumption becomes

$$p(\mathbf{y_1}, \ldots, \mathbf{y_m}) = p(\mathbf{y_1}, \ldots, \mathbf{y_m}|\mathbf{u}_\rho, \tau_{i\rho}^C). \tag{4.75}$$

Then (4.74) becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{\int \left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u}_\rho, \tau_{i\rho})\right] \dfrac{p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)}{p(\mathbf{y_1}, \ldots, \mathbf{y_m})}\mathbf{du}_\rho}. \tag{4.76}$$

We can remove $p(\mathbf{y_1}, \ldots, \mathbf{y_m})$ from the denominator integral (since it does not depend the parameter space) to obtain

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{p(\mathbf{y_1}, \ldots, \mathbf{y_m})\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{\int \left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u}_\rho, \tau_{i\rho})\right] p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}. \tag{4.77}$$

The condition $\{\mathbf{u}_\rho, \tau_{i\rho}\}$ will again be rewritten. Recall that in order to simplify notation, the condition $\tau(I_\rho, E_\rho)$ was dropped from (4.52). Thus $\tau(I_\rho, E_\rho)$ is an implicit condition in (4.77). Specifically, the condition $\{\mathbf{u}_\rho, \tau_{i\rho}\}$ can be rewritten as $\{\mathbf{u}_\rho, \tau_{i\rho}, \tau(I_\rho, E_\rho)\}$.

Hence the condition means that R_ρ and all of the regions in I_ρ have the same underlying parameter value, and the parameter value for R_ρ is given. Therefore, it follows that the parameter value for any $R_k \in I_\rho$, $\mathbf{u_k} = \mathbf{u}_\rho$. The condition can be replaced, and the expression becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{p(\mathbf{y_1}, \ldots, \mathbf{y_m}) \int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{\int \left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u}_\rho)\right] p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}. \tag{4.78}$$

The parameter value, $\mathbf{u}$, common to all of I_ρ, can be used to express $p(\mathbf{y_1}, \ldots, \mathbf{y_m})$ as

$$p(\mathbf{y_1}, \ldots, \mathbf{y_m}) = \int p(\mathbf{y_1}, \ldots, \mathbf{y_m}|\mathbf{u})p(\mathbf{u})\mathbf{du}. \tag{4.79}$$

Given that $R_1, \ldots, R_m$ all belong to the same segment, they all share the same underlying parameter value. Hence when the value $\mathbf{u}$ is given, $\mathbf{Y_1}, \ldots, \mathbf{Y_m}$ become conditionally independent. The independence allows

$$p(\mathbf{y_1}, \ldots, \mathbf{y_m}) = \int \left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u})\right] p(\mathbf{u})\mathbf{du}. \tag{4.80}$$

Now this is expressed in terms of the conditional and prior densities that are given with the model. This can be substituted back into (4.78), resulting in

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}) = \frac{\left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}\right] \left\{\int \left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u})\right] p(\mathbf{u})\mathbf{du}\right\}}{\int \left[\prod_{k=1}^{m} p(\mathbf{y_k}|\mathbf{u})\right] p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{u})\mathbf{du}}. \tag{4.81}$$

This form is intuitively pleasing since it is nearly the same as (4.43) for IE-independent membership probability. The distinction is that the product of pdf's over different regions which appear here in the integrals replaces the pdf corresponding only to region R_i for the independent model.

4.5 Dirac Delta-Function Approximation and Its Applicability

Suppose that the parameter value for R_i can be reliably estimated, and that the parameter value for R_ρ cannot. We show that some simplifications can be made to the membership probability expression (4.43), reducing computation. This is often a reasonble assumption, since it could be the case that there are several large regions in the image for which estimation is very reliable, and most of the uncertainty is due to smaller regions, which produce poor parameter estimates. If larger regions contribute more information about the parameter value and are used as initial regions (i.e., TSSs are built from them), then reliable estimation may be possible. Note, however, that in this general context it is difficult to directly relate the size of a region (or other properties) to the quality of estimates; these ideas are application dependent.

If the parameter value for R_i is known, then the IE-independent vs. IE-dependent distinction is no longer necessary. In either case, the observations corresponding to other regions in I_ρ have no effect on the parameter since their common parameter value is given.

We can consider estimating $\mathbf{u_i}$ by maximizing the observation likelihood

$$\hat{\mathbf{u}}_\mathbf{i} = \underset{\mathbf{u_i}}{argmax}\ p(\mathbf{y_i}|\mathbf{u_i}), \tag{4.82}$$

in which $\mathbf{y_i}$ is the observation from R_i, as discussed previously. This is the standard maximum likelihood estimator (MLE) of $\mathbf{u_i}$. By stating that the estimate is reliable, we mean that the density $p(\mathbf{y_i}|\mathbf{u_i})$ is sharply peaked. For the purpose of membership probability computation, the peaking of this density will be represented as a multiple of a Dirac delta function,

$$\delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i}) = \begin{cases} \infty & \text{if } \mathbf{u_i} = \hat{\mathbf{u}}_\mathbf{i} \\ 0 & \text{otherwise} \end{cases} \tag{4.83}$$

in which

$$\int_{\mathbf{U_i}} \delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i}) \mathbf{du_i} = 1. \tag{4.84}$$

Although the density $p(\mathbf{y_i}|\mathbf{u_i})$ is peaked, we need to allow for some unknown scaling constant, s. This must be allowed since $p(\mathbf{y_i}|\mathbf{u_i})$ is not normalized over $\mathbf{u_i}$ (i.e., the integral over the parameter space of the degradation density is not necessarily one). Hence we represent the density as

$$p(\mathbf{y_i}|\mathbf{u_i}) \approx s\delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i}) = \begin{cases} \infty & \text{if } \mathbf{u_i} = \hat{\mathbf{u}}_\mathbf{i} \\ 0 & \text{otherwise} \end{cases}. \tag{4.85}$$

Next, consider substituting (4.85) into (4.43). This yields

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho\right]\left[\int s\ \delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i})p(\mathbf{u_i})\mathbf{du_i}\right]}{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)\ s\ \delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i})p(\mathbf{u}_\rho)\mathbf{du}_\rho}. \tag{4.86}$$

The delta function representation simplifes two of the integrals into

$$\int s\ \delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i})p(\mathbf{u_i})\mathbf{du_i} = s\ p(\hat{\mathbf{u}}_\mathbf{i}) \tag{4.87}$$

and

$$\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)\ s\ \delta(\mathbf{u_i}, \hat{\mathbf{u}}_\mathbf{i})p(\mathbf{u}_\rho)\mathbf{du}_\rho = s\ p(\mathbf{y}_\rho|\hat{\mathbf{u}}_\mathbf{i})p(\hat{\mathbf{u}}_\mathbf{i}). \tag{4.88}$$

These have been rewritten using the standard substitution that occurs when using the delta function in an integral [90]. Using the simplifications we have

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{s\ p(\hat{\mathbf{u}}_\mathbf{i})}{s\ p(\mathbf{y}_\rho|\hat{\mathbf{u}}_\mathbf{i})p(\hat{\mathbf{u}}_\mathbf{i})}\left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho\right] = \frac{\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho}{p(\mathbf{y}_\rho|\hat{\mathbf{u}}_\mathbf{i})}. \tag{4.89}$$

The resulting simplified form has an interesting interpretation. The average value over $\mathbf{u}_\rho$ of the observation density (the numerator) is compared to the value of the density value at $\hat{\mathbf{u}}_\mathbf{i}$ (the denominator). As the known parameter value causes the observation to appear much more likely than average, the ratio approaches zero, and the membership probability will approach one. In the other case, as the known parameter value causes the observation to appear much less likely than average, then the membership probability approaches zero.

4.6 Multiple Independent Models

Since a Bayesian model is used to determine membership proability, there is a natural extension to the case of multiple, independent models of evidence. We begin the discussion with the prior membership probability. When the evidence-based probability is determined from the first model, the resulting probability can be treated as a membership prior for the next evidence model. This is a natural benefit of using this Bayesian approach, as opposed to formulating some decision criterion for each model. Hence, one can combine evidence from multiple sources.

For simplicity, the multiple model case will be derived only for the IE-independent case. The results carry over to the IE-dependent model in a straightforward manner. Consider having some set of models, with each one specified as in Section 4.2. Take some region R_q. There are m models, each with its own parameter space: $\mathbf{U_q^1}, \ldots, \mathbf{U_q^m}$. The superscripts denote different parameter spaces, not exponents. Also, consider sets of observation variables $\mathbf{Y_q^1}, \ldots, \mathbf{Y_q^m}$, with each set $\mathbf{y^k}$ corresponding to relevant observations about the parameter space $\mathbf{U^k}$. We require that these observation variables, for different parameter spaces, are taken from different $X_l[i,j]$. In other words, no single image variable $X_l[i,j]$ is used for two or more observation variable sets.

The functions Ψ_ρ and Ψ_i from each of the m models, applied to the image elements, yield observations of the random variables $\mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m}$ and $\mathbf{y_i^1}, \ldots, \mathbf{y_i^m}$. This will be all of the evidence used to determine the membership probability, and the posterior membership probability is represented as

$$P(\tau_{i\rho} | \mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m}, \mathbf{y_i^1}, \ldots, \mathbf{y_i^m}). \tag{4.90}$$

We apply Bayes rule as for the independent model to obtain

$$P(\tau_{i\rho} | \mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m}, \mathbf{y_i^1}, \ldots, \mathbf{y_i^m}) = \frac{p(\mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m} | \tau_{i\rho}, \mathbf{y_i^1}, \ldots, \mathbf{y_i^m}) P(\tau_{i\rho} | \mathbf{y_i^1}, \ldots, \mathbf{y_i^m})}{p(\mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m} | \mathbf{y_i^1}, \ldots, \mathbf{y_i^m})}. \tag{4.91}$$

The denominator can be expanded to yield

$$P(\tau_{i\rho}|\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m) = \frac{p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho},\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)P(\tau_{i\rho}|\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)}{p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho},\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)P(\tau_{i\rho}|\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)+p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho}^C,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)P(\tau_{i\rho}^C|\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)}. \tag{4.92}$$

Again the numerator and the first term of the denominator are identical. The only difference between the two terms in the denominator is the event $\tau_{i\rho}$ or $\tau_{i\rho}^C$.

The membership prior must be specified. In this expression it is the probability of membership, before any of the models are applied. The usual definitions result:

$$P(\tau_{i\rho}|\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m) = P(\tau_{i\rho}) \equiv P_0 \tag{4.93}$$

and

$$P(\tau_{i\rho}^C|\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m) = P(\tau_{i\rho}^C) = 1 - P_0. \tag{4.94}$$

By substituting these into (4.91), we obtain

$$P(\tau_{i\rho}|\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m) = \frac{p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho},\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)P_0}{p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho},\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)P_0 + p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho}^C,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)[1-P_0]}. \tag{4.95}$$

Once again, this can be rearranged to

$$P(\tau_{i\rho}|\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m) = \left[1 + \frac{(1-P_0)\; p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho}^C,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)}{P_0\; p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho},\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)}\right]^{-1}. \tag{4.96}$$

Define λ_0 to represent the prior ratio, (4.21). Define $\lambda(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)$ as the ratio representing the effects of all of the model-based evidence. This results in

$$\lambda(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m) \equiv \frac{p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho}^C,\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)}{p(\mathbf{y}_\rho^1,\ldots,\mathbf{y}_\rho^m|\tau_{i\rho},\mathbf{y}_i^1,\ldots,\mathbf{y}_i^m)}. \tag{4.97}$$

We haved assumed that the parametric models are independent. This means that all observations from one model are independent of the observations in another. From this

we have

$$p(\mathbf{y}_\rho^{\mathbf{1}}, \ldots, \mathbf{y}_\rho^{\mathbf{m}} | \tau_{i\rho}, \mathbf{y}_\mathbf{i}^{\mathbf{1}}, \ldots, \mathbf{y}_\mathbf{i}^{\mathbf{m}}) = \prod_{k=1}^{m} p(\mathbf{y}_\rho^k | \tau_{i\rho}, \mathbf{y}_i^k) \tag{4.98}$$

and

$$p(\mathbf{y}_\rho^{\mathbf{1}}, \ldots, \mathbf{y}_\rho^{\mathbf{m}} | \tau_{i\rho}^C, \mathbf{y}_\mathbf{i}^{\mathbf{1}}, \ldots, \mathbf{y}_\mathbf{i}^{\mathbf{m}}) = \prod_{k=1}^{m} p(\mathbf{y}_\rho^k | \tau_{i\rho}^C, \mathbf{y}_i^k). \tag{4.99}$$

Using (4.98) and (4.99) in (4.97), the result is

$$\lambda(\mathbf{y}_\rho^{\mathbf{1}}, \ldots, \mathbf{y}_\rho^{\mathbf{m}}, \mathbf{y}_\mathbf{i}^{\mathbf{1}}, \ldots, \mathbf{y}_\mathbf{i}^{\mathbf{m}}) = \frac{\prod_{k=1}^{m} p(\mathbf{y}_\rho^k | \tau_{i\rho}^C, \mathbf{y}_i^k)}{\prod_{k=1}^{m} p(\mathbf{y}_\rho^k | \tau_{i\rho}, \mathbf{y}_i^k)} = \prod_{k=1}^{m} \frac{p(\mathbf{y}_\rho^k | \tau_{i\rho}^C, \mathbf{y}_i^k)}{p(\mathbf{y}_\rho^k | \tau_{i\rho}, \mathbf{y}_i^k)}. \tag{4.100}$$

Define each element of the product as

$$\lambda_k(\mathbf{y}_\rho^k, \mathbf{y}_i^k) \equiv \frac{p(\mathbf{y}_\rho^k | \tau_{i\rho}^C, \mathbf{y}_i^k)}{p(\mathbf{y}_\rho^k | \tau_{i\rho}, \mathbf{y}_i^k)}. \tag{4.101}$$

Note that each $\lambda_k(\mathbf{y}_\rho^k, \mathbf{y}_i^k)$ is equivalent to the definition for $\lambda_1(\mathbf{y}_\rho, \mathbf{y}_\mathbf{i})$ for the single IE-independent model case. Hence (4.43) can be used to compute the ratios corresponding to each of the independent models.

By using these definitions, the membership probability expression becomes

$$P(\tau_{i\rho} | \mathbf{y}_\rho, \mathbf{y}_\rho^{\mathbf{1}}, \ldots, \mathbf{y}_\rho^{\mathbf{m}}, \mathbf{y}_\mathbf{i}^{\mathbf{1}}, \ldots, \mathbf{y}_\mathbf{i}^{\mathbf{m}}) = \frac{1}{1 + \lambda_0 \prod_{k=1}^{m} \lambda_k(\mathbf{y}_\rho^k, \mathbf{y}_i^k)}. \tag{4.102}$$

For the case in which $m = 1$ this specializes to be (4.23) from the single IE-independent model.

Again, this framework has provided a decomposition of the evidence into components. The prior component λ_0 behaves as before, affecting membership probability. Each of the $\lambda_k(\mathbf{y}_\rho^k, \mathbf{y}_i^k)$ independently contributes to the region membership probability in the same manner as for the single, IE-independent model.

4.7 Discrete Random Variable Cases

In the discussion so far, both $\mathbf{Y_k}$ and $\mathbf{U_k}$ were introduced as mappings on a continuous probability space. For the sake of completeness, the corresponding expressions when either or both are discrete-valued sets of random variables are presented. There are no fundamental differences presented by these cases, hence no further derivations are explicated. The discrete cases are merely notational variants of the previously derived expressions. The single, IE-independent model cases are given; the IE-dependent and multiple model distinctions can be easily extrapolated from these expressions.

If the observation variables $\mathbf{Y_k}$ are discrete valued, and the $\mathbf{U_k}$ are continuous, then the evidence-based ratio is

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int P(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{u})\mathbf{du}\right]\left[\int P(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})\mathbf{du}\right]}{\int P(\mathbf{y}_\rho|\mathbf{u})P(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})\mathbf{du}}. \tag{4.103}$$

Above, $P(\mathbf{y_k}|\mathbf{u})$ simply replaces $p(\mathbf{y_k}|\mathbf{u})$.

If the parameter space $\mathbf{U_k}$ is discrete valued, and the $\mathbf{Y_k}$ are continuous, then we have

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\sum_U p(\mathbf{y}_\rho|\mathbf{u})P(\mathbf{u})\right]\left[\sum_U p(\mathbf{y_i}|\mathbf{u})P(\mathbf{u})\right]}{\sum_U p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y_i}|\mathbf{u})P(\mathbf{u})}. \tag{4.104}$$

If both are discrete valued, the result is

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\sum_U P(\mathbf{y}_\rho|\mathbf{u})P(\mathbf{u})\right]\left[\sum_U P(\mathbf{y_i}|\mathbf{u})P(\mathbf{u})\right]}{\sum_U P(\mathbf{y}_\rho|\mathbf{u})P(\mathbf{y_i}|\mathbf{u})P(\mathbf{u})}. \tag{4.105}$$

4.8 An Illustrative Example

An example of the region membership probability model will now be discussed in detail. The discussion is provided to illustrate some of the concepts presented in this chapter. To keep things simple, the example involves the single, IE-independent model.

Take the parameter space to be the closed unit interval [0,1]. This could, for example, represent normalized intensity values in an intensity image (neglecting the discreteization). Let U denote the single parameter random variable, and $u \in [0,1]$ its value. Take Y to be a single, real-valued observation variable. Let each image element $D[i,j]$ contain only one variable $X[i,j]$, that also takes on real values. Consider some region R_k. Let N_k denote the number of points in R_k. Define the random variable Y_k by

$$Y_k = \frac{1}{N_k} \sum_{D[i,j] \in R_k} X[i,j]. \tag{4.106}$$

Hence, by this definition, Y_k is the mean of the $X[i,j]$ variables comprising R_k. We must next specify $p(u)$ and $p(y|u)$. We define the prior parameter space density to be uniform on $[0,1]$, hence we have $p(u) \equiv 1$. The density $p(y|u)$ represents the noise model. It will be assumed that all of the points in a region are observations of the same real value, under an additive noise model. In other words, without noise, each $X[i,j]$ in a region should take on the same value. This value, without noise, will also be on the unit interval. The parameter space represents this value, which is unknown to the observer because of noise. If this is considered as an intensity image application, all pixels in a region would have the same intensity value, without noise.

The pdf of the data variable $X[i,j]$, taken from region R_k, given the value u_k, will be distributed as $X[i,j] \sim n(u, \sigma^2)$. Explicitly, the density is

$$p(x[i,j] \mid u_k) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{(x[i,j]-u_k)^2}{-2\sigma^2}\right]. \tag{4.107}$$

The density of Y_k, given u_k will also be normally distributed, as $Y_k \sim n(u, \sigma^2/N_k)$, since this is the mean of the data elements. The specification of $p(y_k|u_k)$ is thus

$$p(y_k \mid u_k) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N_k}}} exp\left[\frac{N_k(y_k-u_k)^2}{-2\sigma^2}\right]. \tag{4.108}$$

The complete model has now been specified in accordance with the required densities presented in this chapter. In this context, the evidence-based ratio (4.43) becomes

$$\lambda_1(y_\rho, y_i) = \frac{\left[\int_0^1 p(y_\rho|u)p(u)du\right]\left[\int_0^1 p(y_i|u)p(u)du\right]}{\int_0^1 p(y_\rho|u)p(y_i|u)p(u)du}. \tag{4.109}$$

By using the density definitions just given, we obtain

$$\lambda_1(y_\rho, y_i) = \frac{\left[\int_0^1 \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N_\rho}}} exp\left[\frac{N_\rho(y_\rho - u)^2}{-2\sigma^2}\right] du\right]\left[\int_0^1 \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N_i}}} exp\left[\frac{N_i(y_i - u)^2}{-2\sigma^2}\right] du\right]}{\int_0^1 \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N_\rho}}} exp\left[\frac{N_\rho(y_\rho - u)^2}{-2\sigma^2}\right] \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N_i}}} exp\left[\frac{N_i(y_i - u)^2}{-2\sigma^2}\right] du}. \tag{4.110}$$

The term in front of each of the exp functions does not depend on u, and hence can be removed from each of the integrals, and cancelled:

$$\lambda_1(y_\rho, y_i) = \frac{\left[\int_0^1 exp\left[\frac{N_\rho(y_\rho - u)^2}{-2\sigma^2}\right] du\right]\left[\int_0^1 exp\left[\frac{N_i(y_i - u)^2}{-2\sigma^2}\right] du\right]}{\int_0^1 exp\left[\frac{N_\rho(y_\rho - u)^2}{-2\sigma^2}\right] exp\left[\frac{N_i(y_i - u)^2}{-2\sigma^2}\right] du}. \tag{4.111}$$

Recall that by using (4.111), along with the prior ratio λ_0, the membership probability can be determined with (4.23). We are now prepared to evaluate the membership probability for several cases. The results presented were determined from numerical integration of (4.111).

Consider an example in which $\sigma^2 = 0.02$. Also, assume that $\tau_{i\rho}$ is true (unknown to the observer), hence R_ρ and R_i have the same parameter value u, which is unknown. Take the underlying parameter vector to be $u = 1/3$. The densities of the image elements $X[i,j]$ are identical, and shown in Figure 4.5. Assume also that the membership priors are given by $P(\tau_{i\rho}) = P(\tau_{i\rho}^C) = 1/2$. Consider the case in which $Y_i = 1/3$. Given this model, it is most likely that the observation of y_ρ will be close to Y_i. Figure 4.6 shows the membership proability plotted against different values for y_ρ when $N_\rho = N_i = 1$. This

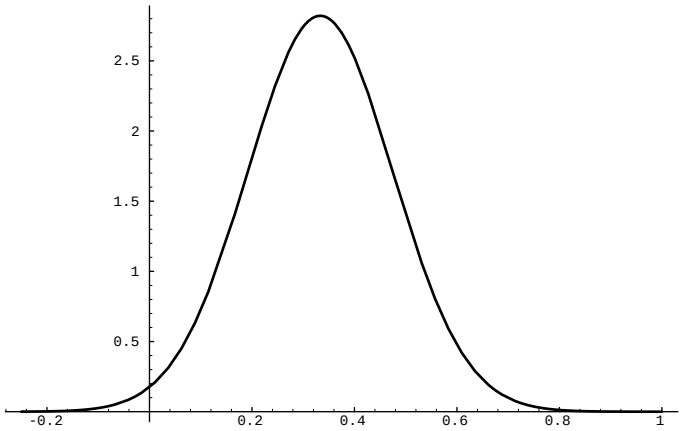


Figure 4.5 The conditional density of $X[i,j]$ when $u = 1/3$ is given.

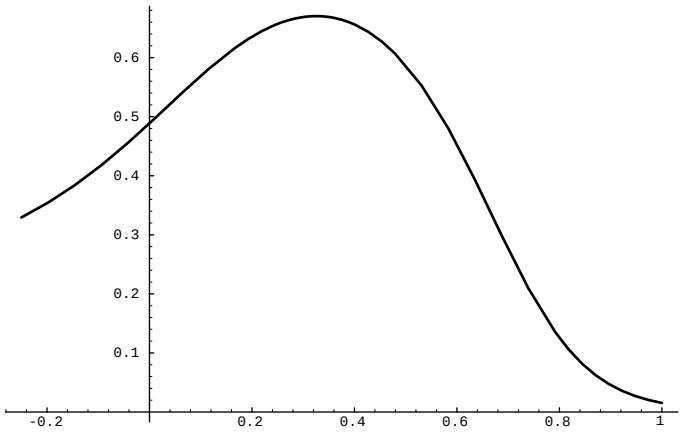


Figure 4.6 Membership probability for different y_ρ when $P(\tau_{i\rho}) = 0.5$, $u = 1/3$, $y_i = 1/3$, $N_\rho = 1$, $N_i = 1$, and $\sigma^2 = 0.02$.

peaks at $y_\rho = 1/3$ since this represents the correct parameter value. Note that for most of the range of observations, the membership probability is between 0.4 and 0.6. With no information available, the priors should cause the membership probability to be 0.5. Since there is only one sample for each region, there is very little information, and this is reflected by fact that most of the probability is close to 0.5, the prior value. Suppose the number of samples in the regions is increased from 1 to 5. Figure 4.7 shows the resulting membership probability. When $y_\rho = 1/3$, the membership probability is higher than it was for the 1-sample case. Also the graph is more peaked around the correct parameter value. This results since more information is available when 5 points are used for each region. When the number of points in each region is 50, the distribution is even more peaked (see Figure 4.8). Note that the maximum membership value (occurring now at $y_\rho = 1/3$) is now well above 0.9. This again, corresponding to the increase in information due to more samples.

Consider changing the membership prior (leaving the rest of the experiment fixed) by letting $P(\tau_{i\rho}) = 0.9$ (consequently $P(\tau_{i\rho}^C) = 0.1$). Consider again, the cases of 1 sample and 5 samples. Figure 4.9 shows the resulting membership probabilities for the one-point case. Most of the probability is close to the prior probability of membership, as expected. In Figure 4.10, it can be seen that this distribution becomes peaked in the presence of more evidence, tending to override the strong prior bias.

Next, consider the situation in which R_ρ and R_i have distinct parameter values, which implies that $\tau_{i\rho}^C$ is true. The same graphs will be obtained, since it is never known to the observer whether the two share the same parameter value. Only the *difference* in observations of y_ρ and y_k will affect the membership probability. If these regions truly do not have the same parameter value, then it is much more likely to observe largely differing values for Y_ρ and Y_i. Assume, for instance that $y_i = 1/3$ and $y_\rho = 2/3$ for 5-point region example. The point on the graph in Figure 4.7 at $y_\rho = 2/3$ gives us the

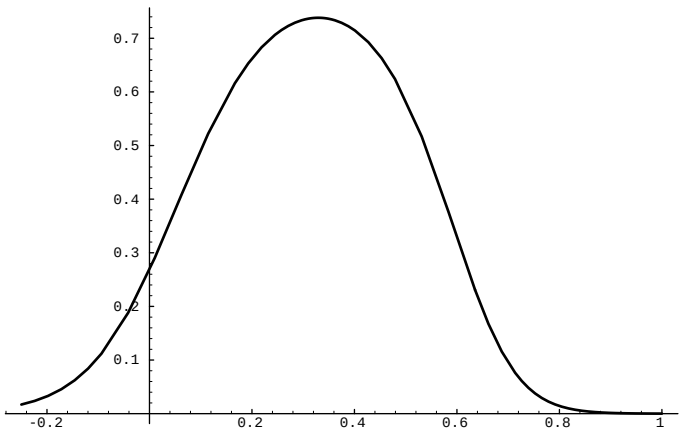


Figure 4.7 Membership probability for different y_ρ when $P(\tau_{i\rho}) = 0.5$, $u = 1/3$, $y_i = 1/3$, $N_\rho = 5$, $N_i = 5$, and $\sigma^2 = 0.02$.

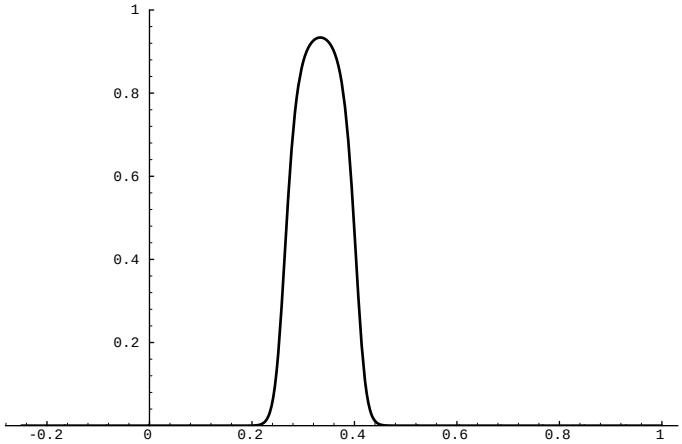


Figure 4.8 Membership probability for different y_ρ when $P(\tau_{i\rho}) = 0.5$, $u = 1/3$, $y_i = 1/3$, $N_\rho = 50$, $N_i = 50$, and $\sigma^2 = 0.02$.

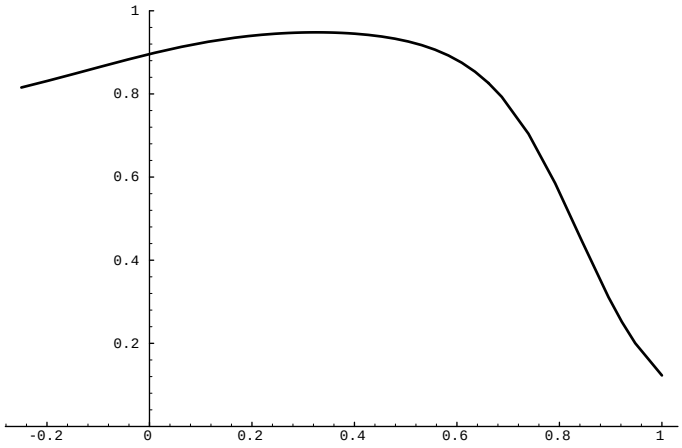


Figure 4.9 Membership probability for different y_ρ when $P(\tau_{i\rho}) = 0.9$, $u = 1/3$, $y_i = 1/3$, $N_\rho = 1$, $N_i = 1$, and $\sigma^2 = 0.02$.

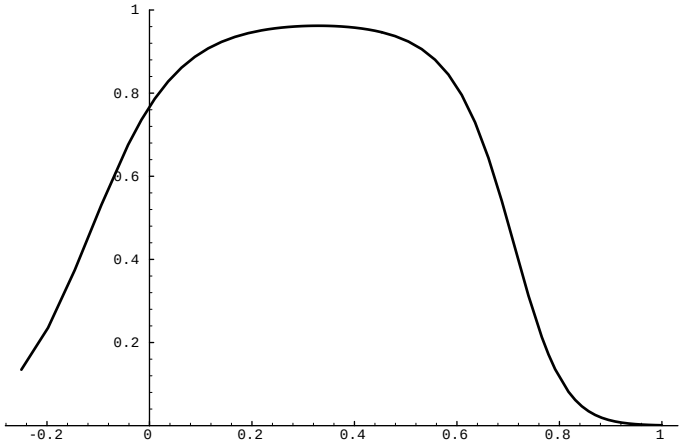


Figure 4.10 Membership probability for different y_ρ when $P(\tau_{i\rho}) = 0.9$, $u = 1/3$, $y_i = 1/3$, $N_\rho = 5$, $N_i = 5$, and $\sigma^2 = 0.02$.

resulting membership probability (which is about 0.2). From this observation, all of the membership probability graphs are the same, since only y_ρ and y_i are used as information.

4.9 Expressions for a Texture Model on Intensity Data

In this section we briefly describe how the four components of the region merging probability model are applied to a standard class of texture models, previously considered in [21],[43]. An image element, $D[i,j]$ represents a single intensity value, $X[i,j]$. The observation space, $\mathbf{Y_k}$ is defined as the vector of points $X[i,j]$ in some region R_k (this corresponds to applying identity maps to the data). We have an N-dimensional parameter space, $\mathbf{U_k}$ representing the texture parameters. The mean, μ_k, in R_k is represented by U_1, and U_2 represents the variance, σ_k^2. The remaining $N-3$ parameters are the interaction parameters, usually denoted with β. In a first-order MRF, for example, there are four parameters corresponding to interactions of $X[i,j]$ with $X[i+1,j]$, $X[i-1,j]$, $X[i,j+1]$, $X[i,j-1]$. The expressions we present pertain to any general order of MRF interactions, and the image element of the l^{th} parameter interaction is denoted by $T_l(x)$. Recall Section 2.2, which described the general MRF approach.

We define the prior model by assigning a uniform density to the parameter space. Let $\mathbf{u'_k}$ denote a vector of minimum parameter values, and $\mathbf{u''_k}$ denote a vector of maximum parameter values. The prior density is

$$p(\mathbf{u_k}) = \prod_{i=1}^{N} \frac{1}{u_i'' - u_i'}. \tag{4.112}$$

The degradation density for R_k is:[7]

$$p(\mathbf{y_k}|\mathbf{u_k}) = \prod_{x \in R_k} \frac{1}{\sqrt{2\pi u_2^2}} exp\left\{ \frac{-1}{2u_2^2} \left[x - u_1 - \sum_{l=3}^{N} u_k(T_l(x) - u_1) \right]^2 \right\}. \tag{4.113}$$

[7]This joint density uses an efficiency-based independence assumption, also used in [21, 43].

By expanding the product, we obtain

$$p(\mathbf{y_k}|\mathbf{u_k}) = (2\pi\mathbf{u}_2^2)^{\frac{|R_k|}{2}} exp\left\{\frac{-1}{2u_2^2}\sum_{x\in R_k}\left[x - u_1 - \sum_{l=3}^{N} u_l(T_l(x) - u_1)\right]^2\right\}. \tag{4.114}$$

Finally, (4.43) can be specialized to obtain

$$\frac{\left[\int_{\mathbf{u}'_\mathbf{k}}^{\mathbf{u}''_\mathbf{k}} e^{\left\{\frac{-1}{2u_2^2}\sum_{x\in R_1}\left[x-u_1-\sum_{l=3}^{N} u_l(T_l(x)-u_1)\right]^2\right\}}\mathbf{du_k}\right]\left[\int_{\mathbf{u}'_\mathbf{k}}^{\mathbf{u}''_\mathbf{k}} e^{\left\{\frac{-1}{2u_2^2}\sum_{x\in R_2}\left[x-u_1-\sum_{l=3}^{N} u_l(T_l(x)-u_1)\right]^2\right\}}\mathbf{du_k}\right]}{\prod_{i=1}^{N}(u''_i - u'_i)\int_{\mathbf{u}'_\mathbf{k}}^{\mathbf{u}''_\mathbf{k}} e^{\left\{\frac{-1}{2u_2^2}\sum_{x\in R_1\cup R_2}\left[x-u_1-\sum_{l=3}^{N} u_l(T_l(x)-u_1)\right]^2\right\}}\mathbf{du_k}}. \tag{4.115}$$

CHAPTER 5

IMPLICIT POLYNOMIAL SURFACES AND NOISE MODELS

In this chapter we apply the Bayesian membership probability model of Chapter 4 to the case in which the parameter space represents a set of implicit polynomial surfaces. We develop specifications of the parameter and observation spaces, and the two required probability densities: the prior model, $p(\mathbf{u})$, and the degradation model, $p(\mathbf{y}|\mathbf{u})$. Each parameter value $\mathbf{u}$, in this context, represents a vector of coefficients which uniquely identifies a polynomial surface. Each observation component represents a function of the signed-distances of the observed data points in a region to a polynomial surface.

The model developed in this chapter is summarized as follows:

Parameter Space Coefficient vector representing implicit polynomial equations

Observation Space Functions of displacements of observed points from a given surface

Degradation Model Densities of functions of displacements from a given surface

Prior Model A noninformative uniform distribution over a space of polynomial surfaces

We later use this formulation for the cases of planar and quadric surfaces. The parameter space arising from these models provides a challenging integration problem when computing the membership probability (due to high dimension of the parameter space), and this

issue is treated in Chapter 6. Our experiments were conducted using the formulations in this chapter, and the results are presented in Chapter 8.

Section 5.1 discusses the use of implicit surface models for segmentation, and related work. A motivational example is presented in Section 5.2, displaying some of the features of the Bayesian membership model in comparison an estimation-based approach. Section 5.3 describes the four components needed to use the Bayesian membership model. Section 5.3.1 defines the implicit polynomial surface model, and the choices that arise in determining its parameter space. Two alternative parameter spaces may be used (the *nonhomogeneous* parameter space and the *homogeneous* parameter space), and the implications of each are treated. Section 5.3.2 defines the observation space for the implicit polynomial model. In Sections 5.3.3 and 5.3.4, the degradation and prior models are defined. Section 5.4 presents simplifications and resulting membership probability expressions. Section 5.5 gives the parameter space and observation space formulations for the case of planar and quadric models, as a specialization of the results in the previous sections.

5.1 Surface Models in Image Segmentation

Image segmentation approaches using the surface patch model are generally considered region-based, since the homogeneity predicate holds when the union of two regions can be represented by a single polynomial surface. Besl and Jain use bivariate polynomials of variable order for segmentation and select the best model by analyzing fitting-error signs and the mean-square error [91]. Leonardis et al. also use bivariate polynomials of variable order and select an appropriate image description through a cost/benefit objective function to obtain a segmentation [92]. Silverman and Cooper use explicit quadric and planar equations to model surfaces patches in intensity images for clustering-based

segmentation [43]. Faugeras and Hebert perform segmentation on range images using implicit

As a demonstration of our approach, we have selected the implicit polynomial surface model for the following reasons, each of which will be discussed:

- A broad class of images/objects can be represented using this model.
- Implicit surfaces have gained increasing popularity both for segmentation and model-based object recognition.
- The model is sufficiently challenging to demonstrate the applicability of our proposed framework.

Implicit polynomial surface models have been used extensively in computer vision, with particular use for range-image segmentation, object modeling, and model-based object recognition. An *implicit equation* is an equation of the form $f(\mathbf{x}) = 0$, in which $\mathbf{x}$ represents a point in the space (for our purposes, $\Re^3$). We are concerned with the case in which $f(\mathbf{x})$ is a polynomial, and the *implicit surface* is the solution set, $\{\mathbf{x} : f(\mathbf{x}) = 0\}$, to the implicit equation.

One reason for the use of implicit surfaces is the range of expressive power, compared to parametric equations (e.g., of the form $x_3 = f(x_1, x_2)$). For instance, objects such as spheres, ellipsoids, and tori cannot be represented by a single explicit equation. Also, any rotation or translation can be applied to an implicit surface, with relative ease. With an explicit model, some rotations will cause f to no longer be a function. These features have made implicit surfaces attractive for recognition and object models, particularly with range data, in which the image directly corresponds to points in the scene (as opposed to intensities).

For object modeling by implicit polynomial surfaces, we paraphrase a description presented by Taubin [93]. An *object*, $\mathcal{O}$, is a collection of surface patches,

$$\mathcal{O} = \bigcup_{j=1}^{q} \mathcal{O}_j, \tag{5.1}$$

in which each surface patch is connected, and a subset of an implicit surface,

$$\mathcal{O}_j \subseteq \{\mathbf{x} : f_j(\mathbf{x}) = 0\}. \tag{5.2}$$

This is the ideal model, and due to noise, the observed surface patches are only approximated by subsets of implicit surfaces.

An image segmentation can also be considered in this form by defining the segments to be connected sets in the image, that correspond to implicit surface patches. Hence a segmentation is considered as a set of subsets of implicit surfaces. In terms of the homogeneity predicate we have

- $H(R_1 \cup R_2) = true$ if and only if $\exists j$ such that the points in $R_1 \cup R_2$ belong to $\mathcal{O}_j$.

Bolle and Cooper have modeled objects appearing in ranges images with patches of planes, spheres, and cylinders for position estimation [94]. Faugeras and Hebert have used implicit quadric and planar models for object modeling, segmentation, and recognition [86]. Taubin and Cooper have developed an efficient estimation procedure for implicit polynomial curves and surfaces of arbitrary order, with application to object recognition [93],[95]. Several other applications of surface models to recognition can be found in [96],[97]. A survey of 3-D surface models and parameter estimation is presented in [98].

For demonstration of our proposed framework, this model is particularly challenging for several reasons. First, the dimension of the parameter space is high when considering our interest in evaluating integrals on the parameter space, appearing in (4.43). For the planar case, there are 3 or 4 parameters, and for quadrics (second-degree equations) there are 9 or 10 parameters. For higher-degree models, the number of parameters increases

considerably. Second, we will need to work extensively with a function of distance of points to a given implicit surface. For implicit models, a closed-form expression does not exist, and we use a distance approximation that has been used for implicit surface estimation [95]. Finally, the surface patch model involves statistical dependencies between data points that are mutually distant in the image, which we stated as a goal for our approach.

5.2 Comparison to Estimation and Decisions

Before continuing with a detailed presentation, we introduce a simple example which illustrates the purpose of the membership model in this context. In particular, we indicate a primary difficulty with using an estimation-based approach, and show when the Bayesian membership model produces a reasonable result. This example is provided to motivate the concepts to be introduced in this chapter, and consequently, some of the details are omitted, to be clarified in the coming sections.

Our example involves a planar model with range data (represented by x_1, x_2 and x_3 coordinates). The parameter space for a region, R_k, corresponds to the set of possible unit normals to a plane (in the positive x_3 direction), represented by $\mathbf{u_k} = [u_1\ u_2\ u_3]^T$. In the parameter space, u_j is component of the unit normal in the x_j direction. The parameter space represents the unit hemisphere in $\Re^3$.

We are given two regions R_1 and R_2 which consist of 26 and 34 points, respectively. The points lie at contiguous integer coordinates in the x_1-x_2 plane, and the x_3 value is sampled from a Gaussian distribution with zero mean, and unit variance. The projection of the points in R_1 and R_2, into the x_1-x_2 plane, is depicted in Figure 5.1. The correct parameter value (unknown to the observer) is $\mathbf{u_1} = \mathbf{u_2} = [0\ \ 0\ \ 1]^T$. This corresponds to the x_1-x_2 plane, where the points would lie if there were no noise component.

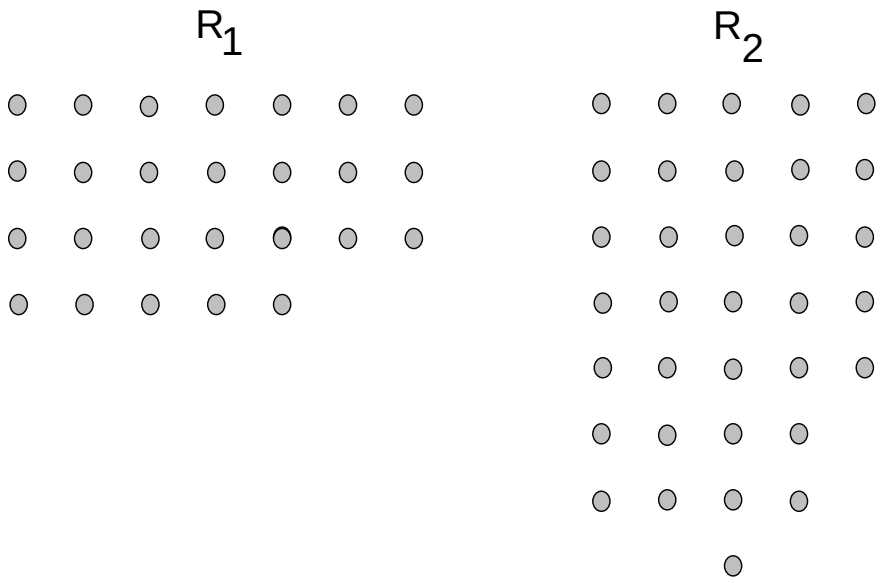


Figure 5.1 The relative x_1-x_2 locations of the points in R_1 and R_2.

Take R_2 as the initial region, and consider the membership probability corresponding to the event $\tau(\{R_1, R_2\}, \emptyset)$. Alternatively, this can be considered as the probability that $R_1 \cup R_2$ is homogeneous. The correct parameter value (unknown to the observer) is

$$\mathbf{u_1} = \mathbf{u_2} = [0 \;\; 0 \;\; 1]^T. \tag{5.3}$$

This corresponds to the x_1-x_2 plane, where the points would lie if there were no noise component.

Figures 5.2 and 5.3 represent the densities of the observed data sets in R_1 and R_2, given the parameter value (and using the optimal choice for scalar offset in the plane equation). The details of these densities are yet to be discussed; however, the plot is obtained by extending a radius from the center of the sphere, representing the likelihood of each surface normal by a longer or shorter radius. The radius function in the plots is given by

$$c_1 + c_2\, p(\mathbf{y_k}|\mathbf{u_k}), \tag{5.4}$$

with c_1 and c_2 chosen to make the make the function visible. Observe that the densities tend to be aligned in different directions for R_1 and R_2. This result is due to the spatial

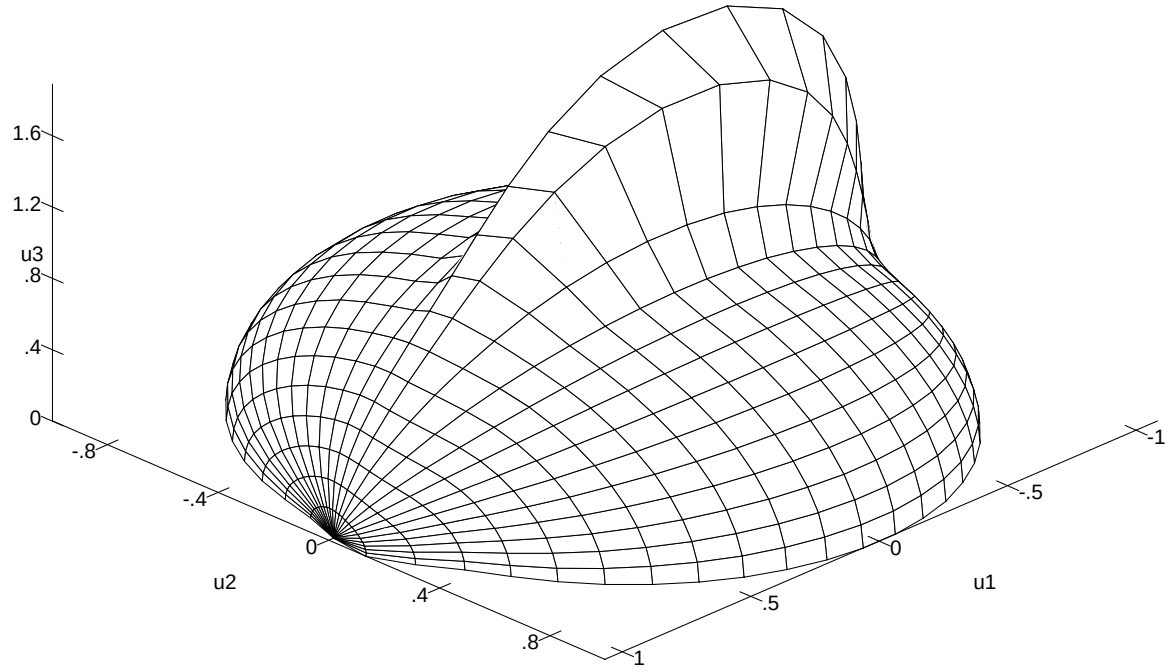


Figure 5.2 The conditional observation space density for R_1

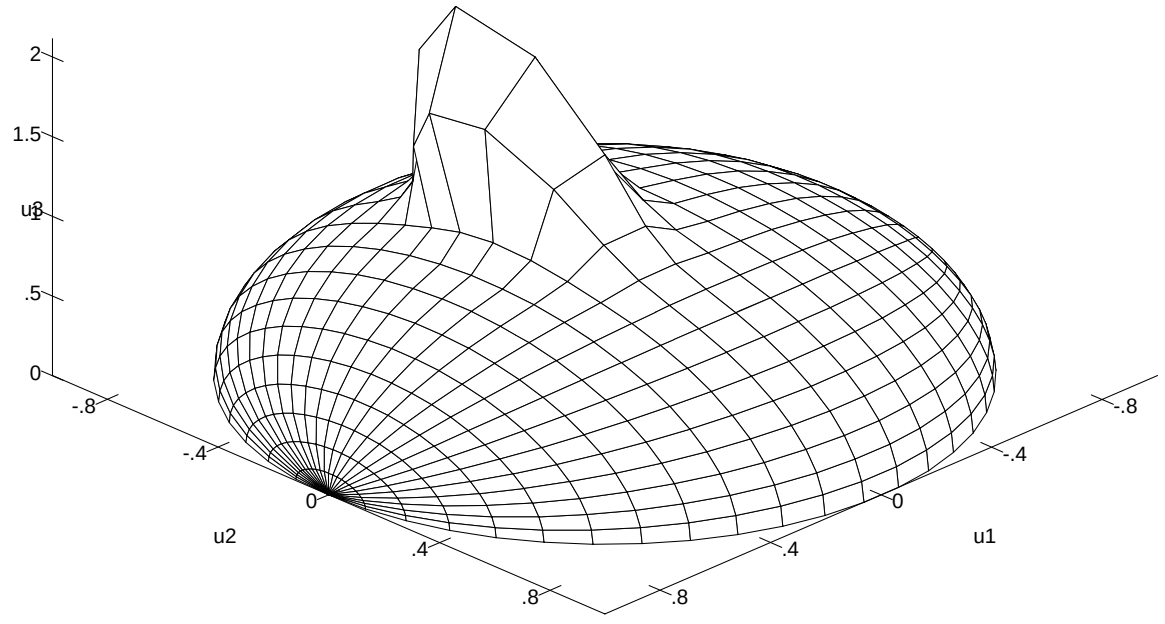


Figure 5.3 The conditional observation space density for R_2

relationship of the points in the regions. The regions are narrow along one direction, and most of the density aligns in an orthogonal direction.

Suppose we wish to determine the membership probability based on the distance between parameter estimates for R_1 and R_2. Or consider the simpler task of formulating a decision criterion, based on the distance, which determines whether or not R_1 and R_2 have the same underlying parameter value. We can use maximum likelihood estimation to obtain optimal parameter estimates $\hat{\mathbf{u}}_\mathbf{1}$ and $\hat{\mathbf{u}}_\mathbf{2}$. This involves selecting values that minimize the sum-of-squares error over the points. Using a standard least-squares minimization technique, the optimal values for our example are

$$\hat{\mathbf{u}}_\mathbf{1} = [-0.258305 \quad -0.025461 \quad 0.965728]^T \tag{5.5}$$

and

$$\hat{\mathbf{u}}_\mathbf{2} = [0.0953739 \quad -0.147327 \quad 0.984479]^T. \tag{5.6}$$

To compute distance, one natural choice is the Euclidean distance between the estimates in the parameter space. For any two parameter vectors $\mathbf{u}$ and $\mathbf{v}$ we define the Euclidean metric m by

$$m(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}. \tag{5.7}$$

For our example, the distance is computed to be

$$m(\hat{\mathbf{u}}_\mathbf{1}, \hat{\mathbf{u}}_\mathbf{2}) = 0.3325\ . \tag{5.8}$$

Since the parameter space represents surface normals to the plane, this is a measure of the distance in orientation of the estimated planes. There is a 21.588 deg angle between these surface normals, reflecting a substantial difference.

A distance-based criterion could, in terms of the homogeneity predicate, be

$$H(R_1 \cup R_2) = true \text{ if } m(\hat{\mathbf{u}}_\mathbf{1}, \hat{\mathbf{u}}_\mathbf{2}) < \epsilon \tag{5.9}$$

for some small $\epsilon > 0$. In most applcations, it would seem that ϵ should be selected small enough to cause the uniformity predicate to evaluate to $false$ if the distance is as large as 0.3325 (i.e., the planes differ in orientation by 21.588 deg). But in this case, we have observed that the distance can be this large, even though R_1 and R_2 have the same underlying parameter value.

Suppose we were to remedy this situation by choosing ϵ large enough to declare $H(R_1 \cup R_2) = true$ for our example, such as $\epsilon = 0.40$. This would cause another problem. Suppose we observe two square regions R_3 and R_4 with $|R_3| = 400$ and $|R_4| = 400$ with integer spacing, and their correct parameter vectors are the same as the estimates for R_1 and R_2:

$$\mathbf{u_3} = [-0.258305 \quad -0.025461 \quad 0.965728]^T \tag{5.10}$$

and

$$\mathbf{u_4} = [0.0953739 \quad -0.147327 \quad 0.984479]^T. \tag{5.11}$$

We would incorrectly declare $H(R_3 \cup R_4) = true$, if the correct estimates are obtained. The difficulty is that the metric space loses some of the information in the parameter and observation spaces that is relevant to determine homogeneity. In this instance, we observed that both the number of points and the position of points in space can affect the decision criterion. We will observe that the Bayesian membership model naturally encodes this information in the prior and degradation model densities.

Our problem is actually more difficult than simply formulating a decision. Since decision making is avoided in our approach, we need to compute the probability of homogeneity for the union of two regions, as opposed to declaring their union homogeneous vs. not homogeneous. For this particular example, using the methods presented in this chapter, we obtain the following intuitively satisfying membership probabilities:

$$P(\tau(\{R_1, R_2\}, \emptyset)|\mathbf{y_1}, \mathbf{y_2}) = 0.767 \tag{5.12}$$

and

$$P(\tau(\{R_3, R_4\}, \emptyset)|\mathbf{y_3}, \mathbf{y_4}) \approx 0.000\,. \tag{5.13}$$

These are computed with a prior membership probability of 0.5. The details involved in modeling and performing this computation are discused in this chapter and in Chapter 6.

5.3 The Membership Components for Implicit Polynomial Surfaces

For the models introduced here, we will be concerned only with the case in which $D[i,j]$ represents a point in $\Re^3$, specified by $[x_1, x_2, x_3]$ coordinates. For simplicity of notation, we will denote some element of the image D by $\mathbf{x}$ instead of $D[i,j]$. Rather than using the i,j indices, we will index points that belong to some region R_k by $\mathbf{x} \in R_k$, assuming the set of regions, $\mathcal{R}$, has been given.

It is assumed that all points in a region R_k came from the same polynomial surface patch, with some noise occurring in the observation process. This is consistent with the assumption that regions are homogeneous, made in Chapter 3. The segmentation goal in this context is to determine maximal connected sets of regions that belong to the same polynomial surface. The problem is that the observer cannot directly obtain the surface parameters, due to noise.

5.3.1 The parameter space

We will now introduce the general, implicit polynomial model, applying to 3D surfaces. A more general formulation of implicit polynomial models, pertaining to curves and surfaces of arbitrary degree and dimension, has been developed by Taubin [93]. A polynomial can be considered as a linear combination of monomial basis functions. Let $\mathcal{W}$ denote a set of monomials in x_1,x_2, and x_3. The j^{th} element of $\mathcal{W}$ is represented by

$x_1^{a_j} x_2^{b_j} x_3^{c_j}$.[1] The integers a_j, b_j, and c_j represent the exponents of the variables. Other basis functions could be used in $\mathcal{W}$, but we avoid this generality for clarity.

There are two distinct ways in which the elements of $\mathcal{W}$ can be combined to define implicit polynomial surfaces. These determine the *nonhomogeneous parameter space* and the *homogeneous parameter space* (or simply the nonhomogeneous space and the homogeneous space). The difference arises when we consider the definition of the parameter space (the coefficients from the linear combination of basis functions).

Before giving the formal expressions we provide an intuitive explanation, in terms of the simple plane equation. Recall that a plane can be uniquely expressed by its unit normal and some real-valued constant. The nonhomogeneous case defines the parameter space as the space of all possible unit normals to a plane (in a canonical direction). The homogeneous case defines the parameter space obtained by adjoining both the space of all unit normals and the space of all scalar offsets. Higher-degree polynomial models have similar interpretations.

Let $\mathbf{u}$ denote a vector of dimension N, and there are N elements of $\mathcal{W}$. The two ways to construct implicit polynomial surfaces are: nonhomogeneous,

$$\phi(\cdot, \mathbf{u}, d) \equiv \sum_{w_j \in \mathcal{W}} u_j w_j = \sum_{j=1}^{N} u_j x_1^{a_j} x_2^{b_j} x_3^{c_j} + d = 0 \tag{5.14}$$

and homogeneous,

$$\phi(\cdot, \mathbf{u}) \equiv \sum_{w_j \in \mathcal{W}} u_j w_j = \sum_{j=1}^{N+1} u_j x_1^{a_j} x_2^{b_j} x_3^{c_j} = 0 \tag{5.15}$$

with

$$a_{N+1} = b_{N+1} = c_{N+1} = 0. \tag{5.16}$$

The $\cdot$'s used here indicate that we have an implicit function with $\mathbf{x}$ as a vector of variables. In later expressions, we will refer to $\phi(\mathbf{x}, \mathbf{u})$ and $\phi(\mathbf{x}, \mathbf{u}, d)$, which yield a nonzero value

[1] Since we are considering linear combinations of these functions, without loss of generality we can assume there are no real-valued coefficients in front of the monomials.

unless $\mathbf{x}$ is on the surface. The *degree* of the polynomial model is the maximum over j of $a_j + b_j + c_j$. The planar model is of degree one, and the quadric model is of degree two.

With the nonhomogeneous space, it is assumed that $1 \notin \mathcal{W}$. In other words, the scalar offset d is not treated as part of the parameter vector $\mathbf{u}$. Specification of $\mathbf{u}$ will identify an implicit surface up to the scalar offset. This identifies an uncountable number of distinct polynomial surfaces, and in practice, a unique surface is obtained by choosing an optimal value for d, given some u and some data points; the details of this optimization will be deferred until Section 5.4.2.

With the homogeneous space, it is assumed that $w_{N+1} = 1$. The scalar offset is thus incorporated into the parameter space, and the linear combination of basis functions uniquely identifies an implicit polynomial equation. We can convert a polynomial represented by the nonhomogeneous model into the homogeneous representation by letting $u_{k+1} = d$ and adding 1 to the set $\mathcal{W}$; similarly, we can transform in the other direction. The homogeneous/nonhomogeneous distinction may appear trivial at this point; however, we will be integrating over the parameter space to obtain membership probabilities, and the definition of the parameter space directly affects this.

No estimation will be required to compute the membership probability with the homogeneous model. We will need to integrate over the parameter space, and every parameter value determines a unique surface. However, for the nonhomogeneous model, the estimation of d is required. The fundamental difference between the nonhomogeneous and homogeneous spaces can be interpreted as whether d is treated as a parameter, with a value that can estimated, or as a random variable, with some density associated with it.

From our experiments, we have found the nonhomogeneous space to be more computationally efficient. The next chapter discusses our Monte Carlo-based integration method, which is used to compute the membership probabilities. It will be seen that the performance of the integration method degrades as the function on the parameter space be-

comes more peaked. It turns out that the homogeneous space tends to have more sharply peaked densities than the nonhomogeneous space, hence we have experimented primarily with the nonhomogeneous space. In this section, the nonhomogeneous/homogeneous distinctions are minor notational variants.

With the present formulation, there are redundant representations of the solution sets (i.e., there are many parameter vectors that describe the same surface in $\Re^3$, even when d is specified). As indicated, we will be considering probability densities on the parameter space $\mathbf{u}$, and be performing integration, as prescribed by (4.43). Hence, it is profitable to choose some restriction of the parameter space that facilitates the integration. It is also important that this restriction retain the full expressive power of the original parameter space. This constrained parameter space will be termed the *parameter manifold.*

For the nonhomogeneous space, observe that $\phi(\cdot, \mathbf{u}, d)$ and $m\phi(\cdot, \mathbf{u}, d)$ for any $m > 0$ represent the same zero set.[2] In other words, we have replaced $\mathbf{u}$ by $m\,\mathbf{u}$ and d by $m\,d$. Also, $\phi(\cdot, \mathbf{u}, d)$ and $-\phi(\cdot, \mathbf{u}, d)$ represent the same zero sets. The sign and scalar-constant multiplications can also be performed for the homogeneous space to yield identical solution sets.

One way to enforce uniqueness is to add the constraints $\|u\| = 1$ and $u_1 \geq 0$. Here, $\|\cdot\|$ is taken to be the standard Euclidean norm. The constraint $\|u\| = 1$ represents the equation of a unit hypersphere centered at the origin of the parameter space. This is a single polynomial constraint, and this represents an $(N-1)$-dimensional manifold in $\Re^N$. Combined with $u_1 \geq 0$, we obtain one-half of a hypersphere. We will denote this half-hypersphere by Σ^N, which is a subset of $\Re^N$.[3] These constraints do not restrict the expressive power of the parameter space, and the space has been reduced to a compact manifold. We can use the same constraint for the homogeneous space, yielding Σ^{N+1} by using the Euclidean norm on the $N+1$ parameters.

[2]This is simply scalar multiplication by m.

[3]This notation differs slightly from the usual hypersphere notation, S^{N-1}, which is a subset of $\Re^N$.

In general, other constraints could be used. The set of all parameter points that correspond to the same zero set form a line through the origin (in the parameter space).[4] The Σ^N restriction selects a single point on this line, at unit distance from the origin. The Euclidean norm was chosen because we have a parameterization of the resulting manifold that is used for integration, in Chapter 6. The l^1 or l^∞ norms (as opposed to the l^2, Euclidean norm) could alternatively be used, giving rise to hypercube parameter manifolds. We will later specify a prior density on this parameter manifold, and the choice of constraints will, in general, affect the actual distribution over the space of surfaces.

This definition of polynomial surfaces is fairly standard. When performing parameter estimation, the homogeneous/nonhomogeneous distinction is not important. This is because the parameter values are selected to minimize some performance functional, regardless of whether d is treated as a separate scalar or as one of the parameters. Since, in general, we consider the parameter space as a vector of random variables and define densities on them, the distinction becomes important.

5.3.2 The observation space

The next step is to define the observation space, $\mathbf{Y_k}$, for the implicit polynomial model. In this section, we will use the subscript $\mathbf{k}$ to refer to parameters/observations from some region R_k. Recall from Chapter 4 that $\mathbf{Y_k}$ is specified by defining a set of functions from the data in each region R_k. Also, recall that we do not work with a direct definition of $p(\mathbf{y_k})$, but rather we define $p(\mathbf{y_k}|\mathbf{u_k})$ and $p(\mathbf{u_k})$. The density $p(\mathbf{y_k})$ is obtained as a marginal density, through integration on the parameter space. This does not mean that a parameter estimate is necessary for computing the membership probability, but that the observation density depends on values of $\mathbf{u_k}$, and only a pdf on $\mathbf{U_k}$ is known.

[4]The nonhomogeneous parameter space is actually projective space of dimension $N-1$ [99].

The observations considered here are functions of the signed distances of the points $\mathbf{x} \in R_k$ from the surface determined by $\mathbf{u_k}$. These signed distances will be referred to as *displacements*. For the homogeneous parameter manifold, the parameter $\mathbf{u_k}$ uniquely determines a zero set; however, for the nonhomogeneous parameter manifold, we will consider the distances of the points from the surface determined by $\mathbf{u_k}$ and by selecting an optimal value for the scalar offset d_k. Assuming the given membership model is true (e.g., the objects truly are composed of quadric surface patches), the only reason for any displacement to be observed is sensor-based noise.

The functions Ψ_k from Chapter 4 will now be defined for this particular context. For the homogeneous case, define $\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))$ to be the displacement of the point $\mathbf{x}$ to the nearest point on the surface described by the zero set, $\{\mathbf{x} : \phi(\mathbf{x}, \mathbf{u_k}) = 0\}$. The function $\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))$ takes on negative values on one side of the surface and positive on the other. Equivalently, we will use $\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}, d_k))$ for the nonhomogeneous case.

There are a number of ways in which the displacements δ at each of the points in a region can be combined to yield an observation space. Recall that it is important to consider the effects on choices for $p(\mathbf{u_k})$ when defining the parameter space; it is also important to consider the effects on choices for $p(\mathbf{y_k}|\mathbf{u_k})$ when defining the observation space. For some choices of functions that define the observation space, it may be difficult or impossible to formulate a degradation model that can be used for computation.

Consider the following possible definitions for the observation space with the homogeneous case (similar expressions apply for the nonhomogeneous case by replacing $\phi(\cdot, \mathbf{u_k})$ with $\phi(\cdot, \mathbf{u_k}, d_k)$):

sum of squares (l^2 norm),[5]

$$y_k(R_k, \mathbf{u_k}) = \sum_{\mathbf{x} \in R_k} [\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))]^2, \tag{5.17}$$

[5]Note that we use y_k instead of $\mathbf{y_k}$ when the observation space is scalar.

median of squares,

$$y_k(R_k, \mathbf{u_k}) = \delta(\mathbf{x}_i, \phi(\cdot, \mathbf{u_k})) \tag{5.18}$$

in which

$$|\delta(\mathbf{x}_i, \phi(\cdot, \mathbf{u_k}))| = median|\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))|, \tag{5.19}$$

identity maps,

$$\mathbf{y_k}(R_k, \mathbf{u_k}) = [\delta(\mathbf{x}_1, \phi(\cdot, \mathbf{u_k}))\ \ \delta(\mathbf{x}_2, \phi(\cdot, \mathbf{u_k}))\ \ \cdots\ \ \delta(\mathbf{x}_{|R_k|}, \phi(\cdot, \mathbf{u_k}))], \tag{5.20}$$

and l^∞ **norm**,

$$y_k(R_k, \mathbf{u_k}) = \max_{\mathbf{x} \in R_k} |\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))|. \tag{5.21}$$

It will be seen in the next section, when degradation densities are presented, that the identity maps model can be alternatively expressed as

$$\mathbf{y_k}(R_k, \mathbf{u_k}) \equiv [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_{|R_k|}]. \tag{5.22}$$

This equivalence occurs within the degradation model, since the identity maps model is a function of the displacements.

These functions of the displacements are often used for polynomial parameter estimation. All of these are statistics in the estimation-theoretic sense. In other words, they could be used for estimation of $\mathbf{u_k}$, given the observations $\mathbf{y_k}$. Taubin has considered both the sum-of-squares and l^∞ norm observation spaces [93].

For our experiments we have considered the sum-of-squares and the identity-maps models. The other two models could be used for the observation space; however, some work must be done to efficiently compute the integrals that will result from using them. We included the models here merely to indicate the generality of the formalism.

Although we have defined the observation space with functions of the displacements, a closed-form expression for the displacement of a point from a polynomial surface does

not exist in general. We use a displacement estimate presented by Taubin and Cooper [95]:

$$\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k})) = \frac{\phi(\mathbf{x}, \mathbf{u_k})}{\|\nabla_x \phi(\cdot, \mathbf{u_k})\|}. \tag{5.23}$$

This is a good approximation for small displacements, and it will turn out that for our approach good approximations are required only for small displacements. Large displacement errors will not cause difficulty because of the approximately zero tail values of the chi-square and Gaussian pdfs. These will be discussed in the next section.

5.3.3 The degradation model

To define the degradation model, we first need to express the density corresponding to the displacement of an observed point from a given surface. Initially, we will give the expressions for the homogeneous case, and then mention the minor difference for the nonhomogeneous case. We use a simple probability model for range-scanning error, used and justified by Bolle and Cooper [94], and also used by Taubin [93]. The model asserts that the displacements of observed points from their true surfaces are a set of Gaussian independent identically distributed random variables with zero mean and some known variance σ^2. It is possible to estimate noise variance in an image before performing segmentation, and methods that accomplish this are discussed in [100],[93]. Extensions to segment-dependent variances and nonzero means could also be considered, but these are primarily notational variants. The degradation model utilized here was merely chosen as a representative of possible models that can be used. In practice, for different imaging systems, other models may be more appropriate. Ikeuchi and Kanade provide a detailed discussion of the modeling of a variety of range-imaging sensors [11].

The expressions are given for the homogeneous case, and again we replace $\phi(\cdot, \mathbf{u})$ with $\phi(\cdot, \mathbf{u}, d)$ for the nonhomogeneous case. Explicitly, the displacement density is

$$p(\delta|\mathbf{u}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{\frac{-\delta(\mathbf{x}, \phi(\cdot, \mathbf{u}))^2}{2\sigma^2}\right\}. \tag{5.24}$$

Since taking the sum-of-squares displacements of Gaussian variables yields the chi-square density, the degradation density using (5.17) is

$$p(y_k|\mathbf{u_k}) = \chi^2_{m_k}(y_k) = \frac{1}{2^{m_k/2}\Gamma(m_k/2)} y_k^{m_k/2-1} e^{-y_k/2}. \tag{5.25}$$

Here y_k is the sum-of-squares for a given region, R_k, and parameter value $\mathbf{u_k}$ (and d_k is given for the nonhomogeneous case). Also, $\Gamma(\cdot)$ is the standard gamma function and $m_k = |R_k|$ (the number of elements in R_k). Above, (5.25) represents the chi-square pdf with m_k degrees of freedom.

Next, consider using the identity-maps observation space (5.22). Each coordinate in the observation space represents the displacement of one point from a given surface. The observation space random variables, $\mathbf{Y_k}$ (corresponding to R_k), are conditionally independent, given the parameter value $\mathbf{u_k}$. Hence, the observation space density is

$$p(\mathbf{y_k}|\mathbf{u_k}) = \prod_{\mathbf{x}\in R_k} p(\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))|\mathbf{u_k}). \tag{5.26}$$

This results in a product of Gaussians,

$$p(\mathbf{y_k}|\mathbf{u_k}) = \prod_{\mathbf{x}\in R_k} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2}\left(\frac{\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))}{\sigma}\right)^2\right\}. \tag{5.27}$$

The product can be expanded, and the exponents summed to obtain

$$p(\mathbf{y_k}|\mathbf{u_k}) = \frac{1}{(2\pi\sigma^2)^{\frac{|R_k|}{2}}} exp\left\{-\frac{1}{2}\sum_{\mathbf{x}\in R_k}\left(\frac{\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))}{\sigma^2}\right)^2\right\}. \tag{5.28}$$

While the homogeneous parameter manifold uniquely identifies a surface with $\mathbf{u_k}$, the nonhomogeneous parameter manifold requires the extra specification of d_k. Although d_k is not treated as a random variable, for the nonhomogeneous space, we write it in the

condition of the degradation density. The expressions corresponding to (5.25) and (5.28) are

$$p(y_k|\mathbf{u_k}, d_k) = \frac{1}{2^{m_k/2}\Gamma(m_k/2)} y_k^{m_k/2-1} e^{-y_k/2} \tag{5.29}$$

and

$$p(\mathbf{y_k}|\mathbf{u_k}, d_k) = \frac{1}{(2\pi\sigma^2)^{\frac{|R_k|}{2}}} exp\left\{-\frac{1}{2}\sum_{\mathbf{x}\in R_k}\left(\frac{\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}, d_k))}{\sigma^2}\right)^2\right\}, \tag{5.30}$$

respectively.

5.3.4 The prior model

The prior model is the only piece of the membership model that remains to be specified. It is a specification of the density on the parameter manifold in the presence of no information. Since the parameter space has been restricted to a bounded set, we can define the prior pdf to have equal value everywhere on the parameter manifold. This captures the notion of uniformity due to the lack of information; however, it is important to note that our choice of parameter manifold affects the prior distribution on the space of surfaces. If other constraints were used on the parameter space, and we assumed a constant-valued pdf, the distribution would be somewhat different from the one we have selected here. Once some information is present, i.e., some observed data points, this distinction becomes less important.

Since the density over the parameter manifold must integrate to one, the uniform density is the inverse of the surface area of the half-hypersphere that defines the parameter manifold. In Section 6.2, some simple expressions are derived for this area.[6] The area of an $(N-1)$-dimensional parameter manifold is denoted here by A_N, and we obtain

$$p(\mathbf{u_k}) = \frac{1}{A_N} \quad \forall R_k \in \mathcal{R}. \tag{5.31}$$

Recall that N is the number of monomial basis functions.

[6] A 2-dimensional parameter manifold is a unit hemisphere, and the area is 2π.

5.4 Determining the Membership Expressions

Using the results of Section 5.3, the membership probability can be expressed explicitly. Recall from Chapter 4 that $P(\tau_{i\rho}|\mathbf{y}_\rho, \mathbf{y_i})$ was given by (4.23), which was a function of two ratios, $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ and λ_0. Recall that $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ is the only portion of the membership probability that depends on the observation space, parameter space, degradation model, or prior model. Using this ratio, the membership probability is directly computed.

5.4.1 General expressions

5.4.1.1 The IE-independent model case

For simplicity, the expressions in this section apply to the homogeneous parameter manifold only. Hence, every parameter value uniquely identifies a surface. In Section 5.4.2 we describe how d is eliminated when using the nonhomogeneous space.

We will obtain expressions for $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ by specializing (4.43). We do not have to distinguish $\mathbf{U}_\rho$ from $\mathbf{U_i}$ since both have the same prior distribution, hence they will both be referred to as $\mathbf{U}$. In general, with the implicit polynomial model, we are integrating over half of the unit hypersphere. These observations yield

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{u})d\mathbf{u}\right]\left[\int_{\Sigma^N} p(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})d\mathbf{u}\right]}{\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y_i}|\mathbf{u})p(\mathbf{u})d\mathbf{u}}. \tag{5.32}$$

We can use (5.31) directly to obtain

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\frac{1}{A_N}\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u})d\mathbf{u}\right]\left[\frac{1}{A_N}\int_{\Sigma^N} p(\mathbf{y_i}|\mathbf{u})d\mathbf{u}\right]}{\frac{1}{A_N}\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y_i}|\mathbf{u})d\mathbf{u}}. \tag{5.33}$$

The areas, A_N, are removed from the integrals since they do not depend on $\mathbf{u}$, and a pair of them will cancel.

We will now give the expressions using the sum-of-squares and identity maps observation spaces. First consider the sum-of-squares; by substitution of the degradation density (5.25) into (5.33), the expression for $\lambda_1(y_\rho, y_i)$ is

$$\frac{\left[\int_{\Sigma^N} \frac{1}{2^{m_\rho/2}\Gamma(m_\rho/2)} y_\rho^{m_\rho/2-1} e^{-y_\rho/2} d\mathbf{u}\right] \left[\int_{\Sigma^N} \frac{1}{2^{m_i/2}\Gamma(m_i/2)} y_i^{m_i/2-1} e^{-y_i/2} d\mathbf{u}\right]}{A_N \int_{\Sigma^N} \frac{1}{2^{m_\rho/2}\Gamma(m_\rho/2)} y_\rho^{m_\rho/2-1} e^{-y_\rho/2} \frac{1}{2^{m_i/2}\Gamma(m_i/2)} y_i^{m_i/2-1} e^{-y_i/2} d\mathbf{u}}. \tag{5.34}$$

The fractions that appear in front of the exponential functions depend only on the degrees of freedom, and not on $\mathbf{u}$. Hence, these can be removed from the integrands, and cancelled in the ratio to obtain

$$\lambda_1(y_\rho, y_i) = \frac{\left[\int_{\Sigma^N} y_\rho^{m_\rho/2-1} e^{-y_\rho/2} d\mathbf{u}\right] \left[\int_{\Sigma^N} y_i^{m_i/2-1} e^{-y_i/2} d\mathbf{u}\right]}{A_N \int_{\Sigma^N} y_\rho^{(m_\rho/2-1)} y_i^{(m_i/2-1)} e^{-(y_\rho+y_i)/2} d\mathbf{u}}. \tag{5.35}$$

Now consider the identity-maps observation space. In (5.28) the degradation density was expressed as a function of the sum-of-squares of displacements. For some region R_k, let z_k denote this function of the sum of squares:

$$z_k = \left\{ -\frac{1}{2} \sum_{\mathbf{x} \in R_k} \left(\frac{\mathbf{x} - \phi(\cdot, \mathbf{u})}{\sigma} \right)^2 \right\}. \tag{5.36}$$

The expression above is the sum of squares multiplied by $1/2\sigma^2$. Using this definition and (5.28) we have

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int_{\Sigma^N} \frac{1}{(2\pi\sigma^2)^{\frac{|R_\rho|}{2}}} e^{z_\rho} d\mathbf{u}\right] \left[\int_{\Sigma^N} \frac{1}{(2\pi\sigma^2)^{\frac{|R_i|}{2}}} e^{z_i} d\mathbf{u}\right]}{A_N \int_{\Sigma^N} \frac{1}{(2\pi\sigma^2)^{\frac{|R_\rho|}{2}}} e^{z_\rho} \frac{1}{(2\pi\sigma^2)^{\frac{|R_i|}{2}}} e^{z_i} d\mathbf{u}}. \tag{5.37}$$

The fractions in front of the exponents do not depend on $\mathbf{u}$, and can be removed from the integrals and cancelled to obtain

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int_{\Sigma^N} e^{z_\rho} d\mathbf{u}\right] \left[\int_{\Sigma^N} e^{z_i} d\mathbf{u}\right]}{A_N \int_{\Sigma^N} e^{(z_\rho + z_i)} d\mathbf{u}}. \tag{5.38}$$

5.4.1.2 The IE-dependent model case

For the IE-dependent model, similar expressions are obtained through direct substitution of the four membership components into (4.51) or (4.81). In order to use (4.51), the definitions must be extended, however, to R_I.

We next discuss the difference between using (4.51) or (4.81). First, consider the identity maps observation space. Since all of the data are directly used to obtain a joint density, the observation space is trivially a sufficient statistic; the sum-of-squares space is not.

It was stated in Section 4.4 that the two different formulations of the IE-dependent membership model are equivalent when $\mathbf{y}$ is a set of sufficient statistics. Indeed, this is true for our identity maps space. Consider two regions, R_{i_1} and R_{i_2}, which belong to I_ρ. They share the same parameter value, $\mathbf{u}$, since the implicition is made that $H(R_{i_1} \cup R_{i_2}) = true$. Since the observation variables are conditionally independent given $\mathbf{u}$, the joint degradation density is

$$p(\mathbf{y}_{i_1}, \mathbf{y}_{i_2}|\mathbf{u}) = \left[\frac{1}{(2\pi\sigma^2)^{\frac{|R_{i_1}|}{2}}} e^{z_{i_1}}\right]\left[\frac{1}{(2\pi\sigma^2)^{\frac{|R_{i_2}|}{2}}} e^{z_{i_2}}\right] = \frac{1}{(2\pi\sigma^2)^{\frac{|R_{i_1}|+|R_{i_2}|}{2}}} e^{z_{i_1}+z_{i_2}}. \quad (5.39)$$

Also, the degradation density for $R_i = R_{i_1} \cup R_{i_2}$ is

$$\frac{1}{(2\pi\sigma^2)^{\frac{|R_i|}{2}}} e^{z_i} = \frac{1}{(2\pi\sigma^2)^{\frac{|R_{i_1}|+|R_{i_2}|}{2}}} e^{z_{i_1}+z_{i_2}}. \quad (5.40)$$

These forms are equal, and the two approaches outlined in Section 4.4 coincide. For the sum-of-squares density, this equivalence does not occur.

Note that the integrands of (5.35) and (5.38) are all functions of the sum-of-squares displacements. This is a nice benefit of using the Gaussian noise model; however, in general, one can expect the sum-of-squares model to be simpler, since the observation space is one dimensional, as opposed to varying with the number of points. These integrals will be evaluated numerically using Monte-Carlo integration, which is discussed

in the next chapter. To perform the function evaluations for these two models efficiently, it will be necessary to construct an efficient representation of the sum-of-squares error for a given parameter value. This is the motivation for Section 5.4.3, but before this, Section 5.4.2 presents expressions for the nonhomogeneous membership probability.

5.4.2 Eliminating d_k for the nonhomogeneous space

In this section we consider the nonhomogeneous case, and the problem of eliminating d_k. The goal is to determine the optimal value for d_k, given some parameter value $\mathbf{u_k}$ in R_k. We should denote this as $d_k(\mathbf{u_k})$; however, to simplify the notation we will write only d_k. Since d_k is required, along with $\mathbf{u_k}$, to uniquely identify a surface, consider representing (5.32) as

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y_i}) = \frac{\left[\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u}, \hat{d}_\rho)p(\mathbf{u})d\mathbf{u}\right]\left[\int_{\Sigma^N} p(\mathbf{y_i}|\mathbf{u}, \hat{d}_i)p(\mathbf{u})d\mathbf{u}\right]}{\int_{\Sigma^N} p(\mathbf{y}_\rho|\hat{d}_{i\rho}, \mathbf{u})p(\mathbf{y_i}|\mathbf{u}, \hat{d}_{i\rho})p(\mathbf{u})d\mathbf{u}}, \tag{5.41}$$

in which $\hat{d}_\rho$ (or $\hat{d}_i$) represents values of d_ρ in region R_ρ (or d_i in R_i) given $\mathbf{u}_\rho$ (or $\mathbf{u_i}$). Also, $\hat{d}_{i\rho}$ represents the value for $d_{i\rho}$ in region[7] $R_\rho \cup R_i$. Using the sum-of-squares and identity-maps model expressions, (5.29) and (5.30), the expressions for $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ will appear identical to (5.35) and (5.38), with the implicit understanding that the value for d_k is given for each $\mathbf{u}$.

We need to derive a simple expression for the optimal value of d_k, given a region R_k and $\mathbf{u_k}$. We will see in Section 5.4.3 that the estimated sum-of-squares displacements in some region, R_k, can be written

$$\sum_{\mathbf{x}\in R_k} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k}, d_k))]^2 = \frac{\sum_{\mathbf{x}\in R_k} [\phi(\mathbf{x}, \mathbf{u_k}, d_k)]^2}{\frac{1}{|R_k|}\sum_{\mathbf{x}\in R_k} \|\nabla_x \phi(\cdot, \mathbf{u_k}, d_k)\|^2}. \tag{5.42}$$

[7] In general this may not be a region by our technical definition due to lack of connectivity, but this will not affect the values.

This displacement function depends on d_k, and the integrations in the membership probability are taken over $\mathbf{u_k}$. This gives an uncountable number of choices for the value of d_k, each yielding a different displacement.

We will select the optimal value for d_k, denoted by $\hat{d}_k$, which minimizes the sum-of-squares displacements for a given value of $\mathbf{u_k}$. This requires one corresponding $\hat{d}_k$ for every value $\mathbf{u_k}$. It turns out that we can obtain an expression for $\hat{d}_k$, and work it into a quadratic form which depends only on $\mathbf{u_k}$. This is done by selecting a value for d_k for which (5.42) is optimized. This is represented as

$$\sum_{\mathbf{x}\in R_k} [\hat{\delta}(x, \phi(\cdot, \mathbf{u_k}, \hat{d}_k))]^2 = \min_{d_k} \sum_{\mathbf{x}\in R_k} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k}, d_k))]^2. \tag{5.43}$$

A single value $\hat{d}_k$ is chosen over all of the points in R_k, since we can always find a value for $\hat{d}_k$ which reduces the displacement of a single point to zero.

The optimal choice for d_k occurs when

$$\frac{\partial}{\partial d_k} \left\{ \sum_{\mathbf{x}\in R_k} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k}, d_k))]^2 \right\} = 0. \tag{5.44}$$

Using (5.42) we have

$$\frac{\partial}{\partial d_k} \left\{ \frac{\sum_{\mathbf{x}\in R_k} [\phi(\mathbf{x}, \mathbf{u_K}, d_k)]^2}{\sum_{\mathbf{x}\in R_k} \|\nabla_x \phi(\cdot, \mathbf{u_k}, d_k)\|^2} \right\} = 0. \tag{5.45}$$

Note that $\nabla_x \phi(\cdot, \mathbf{u_k}, d_k)$ is *not* a function of d_k since there is no monomial in $\mathbf{x}$ appearing in front of d_k in (5.14). Consequently, it does not affect the gradient, and the denominator may be removed to obtain

$$\frac{\partial}{\partial d_k} \sum_{\mathbf{x}\in R_k} [\phi(\mathbf{x}, \mathbf{u_k}, d_k)]^2 = 0. \tag{5.46}$$

By a slight abuse of terminology, we will denote the j^{th} element of the parameter vector $\mathbf{u}_k$ by u_j. In other words, the reference to R_k will be understood, and not explicitly mentioned. Using the explicit definition for nonhomogeneous surfaces, (5.14), we have

$$\frac{\partial}{\partial d_k} \sum_{\mathbf{x}\in R_k} ((\sum_{j=1}^{N} u_j x_1^{a_j} x_2^{b_j} x_3^{c_j}) + d_k)^2 = \frac{\partial}{\partial d_k} \sum_{\mathbf{x}\in R_k} ((\sum_{j=1}^{N} u_j w_j) + d_k)^2 = 0. \tag{5.47}$$

In the second part of the equation we are denoting the monomial $x_1^{a_j} x_2^{b_j} x_3^{c_j}$ (evaluated at some $\mathbf{x} \in R_k$) by w_j, for notational convenience. Performing the differentiation we obtain

$$2 \sum_{\mathbf{x} \in R_k} ((\sum_{j=1}^{N} u_k w_k) + d_k) = 0. \tag{5.48}$$

Solving for d_k, we obtain

$$\hat{d}_k = \frac{-1}{|R_k|} \sum_{\mathbf{x} \in R_k} (\sum_{j=1}^{N} u_j w_j), \tag{5.49}$$

in which $|R_k|$ is taken to be the number of elements in R_k. Using $R_k = R_i$, $R_k = R_\rho$, and $R_k = R_{i\rho}$ we obtain the values $\hat{d}_i$, $\hat{d}_\rho$, and $\hat{d}_{i\rho}$, respectively.

For computational efficiency, the expression can be rewritten as

$$\hat{d}_k = \frac{-1}{|R_k|} (\sum_{j=1}^{N} u_j \sum_{\mathbf{x} \in R_k} w_j), \tag{5.50}$$

or, equivalently,

$$\hat{d}_k = -\sum_{j=1}^{N} u_j \bar{w}_j. \tag{5.51}$$

Above, each $\bar{w}_j$ is the mean of a single monomial over all points in R_k:

$$\bar{w}_j = \frac{1}{|R_k|} \sum_{j=1}^{N} w_j. \tag{5.52}$$

Hence the $\hat{d}_k$ are simple linear combinations of the means. These means do not depend on the parameter value, and they can be computed in advance for each region.

5.4.3 Efficient representation of the sum-of-squares displacements

We are required to perform several integrations on the parameter manifold in order to determine the membership probability. The observation space density apears in the integrands and hence will be converted into a computationally efficient form to facilitate the integrations.

When the integrals are evaluated numerically, it will become particularly important to minimize the computational cost of performing a single integrand evaluation for a given parameter value. The goal is to represent the sum-of-squares displacements in a way that does not involve a summation over all of the points in a region. This can be accomplished by writing the sum-of-squares as a ratio of two matrix quadratic forms. For each region, two matrices are stored, corresponding to the numerator and the denominator quadratic forms. These are computed during a preprocessing stage and are independent of any parameter vector. The vector of the quadratic form is the parameter vector. For each function evaluation a different parameter value is used, and simple matrix multiplications are performed to yield the sum-of-squares displacements. These resulting matrices will be symmetric, and in practice only the upper triangles are determined, and used to compute the sum-of-squares.

5.4.3.1 Sum of squares for the homogeneous space

First, we will treat the homogeneous space. Using the displacement estimate (5.23), the sum-of-squares expression is

$$\sum_{\mathbf{x}\in R_k} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))]^2 = \sum_{\mathbf{x}\in R_k} \frac{[\phi(\mathbf{x}, \mathbf{u_k})]^2}{\|\nabla_x \phi(\cdot, \mathbf{u_k})\|^2}. \tag{5.53}$$

We want to compute the sums independently from the parameter value, so that the function can be evaluated for new parameter vectors without recomputing expensive sums. Based entirely of the need for computational efficiency, we borrow a simplification used by Taubin and Cooper [95]. In their work, the simplification was performed to facilitate optimization for the purpose of parameter estimation of implicit surfaces. This simplification makes the assumption that the magnitude of the gradient remains fairly constant over the set of points, for a given parameter value. Using this, (5.53) can be rewritten with a numerator summation and a denominator summation. Since their

definition for the parameter space coincides with our homogeneous parameter manifold, this simplification is equally valid for our work.

For the homogeneous space, this simplification yields

$$\sum_{\mathbf{x}\in R_k} \frac{[\phi(\mathbf{x},\mathbf{u_k})]^2}{\|\nabla_x \phi(\cdot,\mathbf{u_k})\|^2} \approx \frac{\sum_{\mathbf{x}\in R_k} [\phi(\mathbf{x},\mathbf{u_k})]^2}{\frac{1}{|R_k|}\sum_{\mathbf{x}\in R_k} \|\nabla_x \phi(\cdot,\mathbf{u_k})\|^2}. \tag{5.54}$$

Recall that $\phi(\cdot,\mathbf{u_k})$ is linear in the parameters $\mathbf{u_k}$. The numerator above must be a quadratic function in the parameter value, since it is a linear function squared. The denominator is also quadratic, since the gradient yields a linear function of $\mathbf{u_k}$, and the magnitude squared yields a quadratic. The numerator and denominator represent sums of quadratics, and hence are in turn quadratic. From this, they can each be expressed as an $(N+1)\times(N+1)$ quadratic form. This gives

$$\sum_{\mathbf{x}\in R_k} [\hat{\delta}(\mathbf{x},\phi(\cdot,\mathbf{u_k}))]^2 \approx \frac{\mathbf{u_k}^T M_h \mathbf{u_k}}{\mathbf{u_k}^T Q_h \mathbf{u_k}}. \tag{5.55}$$

Both M_h and Q_h are symmetric nonnegative definite matrices. The parameter vector has been represented in column form, and $\mathbf{u_k}^T$ denotes the transpose of $\mathbf{u_k}$ (row form). The specific matrices are cumbersome to denote, and the specific planar and quadric cases are given in Section 5.5. Each element of M_h and Q_h are functions only of $\mathbf{x}\in R_k$, and not dependent on $\mathbf{u_k}$. From this, the matrices remain static as different parameter values are chosen.

5.4.3.2 Sum of squares for the nonhomogeneous space

Now we discuss the nonhomogeneous space. This space is more difficult since an optimal value for d_k must be determined for each value of $\mathbf{u_k}$. It will turn out that we can still obtain a quadratic form that incorporates the optimal value for d_k. The value for d_k is not explicitly computed, but represented implicitly in the quadratic form.

First, we argue that the nonhomogeneous denominator matrix, Q_n, is a submatrix of Q_h. Recall that $\nabla_x \phi(\cdot, \mathbf{u_k}, d_k)$ is *not* a function of d_k. Also, u_{N+1} does not appears in $\nabla_x \phi(\cdot, \mathbf{u_k})$. Consider the quadratic form obtained by expanding the denominator of (5.42), and denote the corresponding $N \times N$ quadratic form matrix by Q_n. Since neither d_k nor u_{N+1} appears in the gradient expressions, the quadric polynomials that represent the gradient for the two cases are identical. Therefore, Q_h contains zeros in its final row and column (these represent factors combined with u_{N+1}). We obtain Q_n by simply removing the final row and column of Q_h.

The only work left is to determine an $N \times N$ quadratic form with M_n for the numerator matrix. The estimate for d_k, given by (5.51), can be substituted into the numerator of (5.42) to yield

$$\sum_{\mathbf{x} \in R_k} [\phi(\mathbf{x}, \mathbf{u_k}, \hat{d}_k)]^2 = \sum_{\mathbf{x} \in R_k} (\sum_{j=1}^{N} u_j w_j) + \hat{d}_k)^2 = \sum_{\mathbf{x} \in R_k} (\sum_{j=1}^{N} u_j (w_j - \bar{w}_j))^2. \tag{5.56}$$

The result in (5.56) is quadratic in $\mathbf{u_k}$, and can be expanded into a quadratic form $\mathbf{u_k}^T M_n \mathbf{u_k}$. The matrix M_n is $N \times N$ nonnegative symmetric with the $(s,t)^{th}$ component given by

$$\sum_{\mathbf{x} \in R_k} (w_s - \bar{w}_s)(w_t - \bar{w}_t). \tag{5.57}$$

This matrix form evaluated on the points is a quadratic polynomial in $\mathbf{u_k}$ with d_k eliminated. The elimination brought the expression from linear to quadratic. This procedure for eliminating d_k and setting up the quadratic form has been used in [50] for the problem of linear estimation for planes.

These matrices produce an efficient representation of the sum-of-squares displacement, similar to (5.55),

$$\sum_{\mathbf{x} \in R_k} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))]^2 \approx \frac{\mathbf{u_k}^T M_n \mathbf{u_k}}{\mathbf{u_k}^T Q_n \mathbf{u_k}}. \tag{5.58}$$

In Chapter 6, these forms will be used to represent the membership-probability integrands.

5.5 Special Cases

In this section we present specializations of the models introduced to the cases of planes and quadrics. These are the cases that we examined in our experiments. The planar case is interesting because the sum-of-squares expression simplifies further, and because the nonhomogeneous parameter manifold can be easily visualized (since it is embedded in $\Re^3$). The quadric case raises some of the more interesting issues involved in using higher-degree models, and provides a concrete example of the implicit models for a more interesting case.

5.5.1 The planar model

The set of monomials for the planar model is simply $\mathcal{W} = \{x_1,\ x_2,\ x_3\}$. The non-homogeneous parameter manifold is $\mathbf{u_k} = [u_1\ u_2\ u_3]^T$, and the corresponding equation is

$$\phi(\cdot, \mathbf{u_k}, d_k) \equiv u_1 x_1 + u_2 x_2 + u_3 x_3 + d_k = 0. \tag{5.59}$$

The uniqueness constraints are $u_3 \geq 0$ and $\|\mathbf{u_k}\| = u_1^2 + u_2^2 + u_3^2 = 1$. These constraints yield the unit hemisphere, shown in Figure 5.4. For this case the parameter manifold corresponds to normals to the plane since

$$\nabla_x (u_1 x_1 + u_2 x_2 + u_3 x_3 + d_k) = [u_1\ u_2\ u_3]. \tag{5.60}$$

From (5.60), and since $\|\mathbf{u_k}\| = 1$, the displacement estimate (5.23) becomes exact:

$$\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}, d_k)) = \phi(\mathbf{x}, \mathbf{u_k}, d_k) = u_1 x_1 + u_2 x_2 + u_3 x_3 + d_k. \tag{5.61}$$

The sum-of-squares displacements (5.17) is

$$\sum_{\mathbf{x} \in R_k} (u_1 x_1 + u_2 x_2 + u_3 x_3 + d_k)^2. \tag{5.62}$$

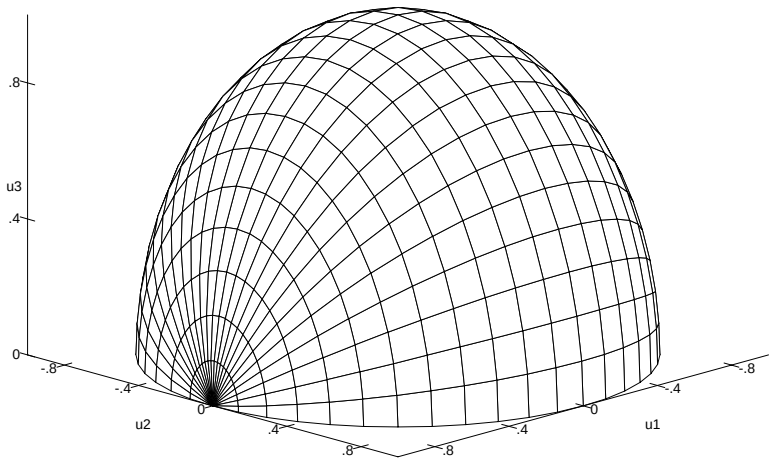


Figure 5.4 The nonhomogeneous parameter manifold $\mathbf{u_k}$ for the planar model is a unit hemisphere at the origin in $\Re^3$.

The resulting matrix M_n is

$$\begin{pmatrix} \sum_{\mathbf{x}\in R_k}(x_1-\bar{x}_1)^2 & \sum_{\mathbf{x}\in R_k}(x_1-\bar{x}_1)(x_2-\bar{x}_2) & \sum_{\mathbf{x}\in R_k}(x_1-\bar{x}_1)(x_3-\bar{x}_3) \\ \sum_{\mathbf{x}\in R_k}(x_2-\bar{x}_2)(x_1-\bar{x}_1) & \sum_{\mathbf{x}\in R_k}(x_2-\bar{x}_2)^2 & \sum_{\mathbf{x}\in R_k}(x_2-\bar{x}_2)(x_3-\bar{x}_3) \\ \sum_{\mathbf{x}\in R_k}(x_3-\bar{x}_3)(x_1-\bar{x}_1) & \sum_{\mathbf{x}\in R_k}(x_3-\bar{x}_3)(x_2-\bar{x}_2) & \sum_{\mathbf{x}\in R_k}(x_3-\bar{x}_3)^2 \end{pmatrix}, \quad (5.63)$$

in which

$$\bar{x}_i = \frac{1}{|R_k|}\sum_{\mathbf{x}\in R_k} x_i. \quad (5.64)$$

The homogeneous planar equation is

$$u_1x_1 + u_2x_2 + u_3x_3 + u_4 = 0, \quad (5.65)$$

and $\mathcal{W} = \{x_1, x_2, x_3, 1\}$. The uniqueness constraints are $u_3 \geq 0$ and $\|\mathbf{u_k}\| = u_1^2 + u_2^2 + u_3^2 + u_4^2 = 1$. The matrix, M_h, is

$$\begin{pmatrix} \sum_{\mathbf{x}\in R_k} x_1^2 & \sum_{\mathbf{x}\in R_k} x_1x_2 & \sum_{\mathbf{x}\in R_k} x_1x_3 & \sum_{\mathbf{x}\in R_k} x_1 \\ \sum_{\mathbf{x}\in R_k} x_2x_1 & \sum_{\mathbf{x}\in R_k} x_2^2 & \sum_{\mathbf{x}\in R_k} x_2x_3 & \sum_{\mathbf{x}\in R_k} x_2 \\ \sum_{\mathbf{x}\in R_k} x_3x_1 & \sum_{\mathbf{x}\in R_k} x_3x_2 & \sum_{\mathbf{x}\in R_k} x_3^2 & \sum_{\mathbf{x}\in R_k} x_3 \\ \sum_{\mathbf{x}\in R_k} x_1 & \sum_{\mathbf{x}\in R_k} x_2 & \sum_{\mathbf{x}\in R_k} x_3 & \sum_{\mathbf{x}\in R_k} 1 \end{pmatrix}. \tag{5.66}$$

The denominator matrix, Q_h, is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5.67}$$

The quadratic form in the denominator is simply $\mathbf{u_k}^T Q_h \mathbf{u_k} = u_1^2 + u_2^2 + u_3^2$.

5.5.2 The quadric model

For the quadric case, the monomials for the nonhomogeneous case are

$$\mathcal{W} = \{x_1^2,\ x_2^2,\ x_3^2,\ x_1x_2,\ x_1x_3,\ x_2x_3,\ x_1,\ x_2,\ x_3\}, \tag{5.68}$$

and the parameter manifold is given by $\mathbf{u_k} = [u_1\ \ u_2\ \ \ldots\ \ u_9]$. It is difficult to express M_n on a single page. It is similar in appearance to (5.63), although with dimensions 9×9. The denominator matrix Q_n is determined by expanding the gradient, and organizing

the terms into a matrix. This yields

$$Q_n = \sum_{\mathbf{x}\in R_k} \begin{pmatrix} 4x_1^2 & 0 & 0 & 2x_1x_2 & 2x_1x_3 & 0 & 2x_1 & 0 & 0 \\ 0 & 4x_2^2 & 0 & 2x_1x_2 & 0 & 2x_2x_3 & 0 & 2x_2 & 0 \\ 0 & 0 & 4x_3^2 & 0 & 2x_1x_3 & 2x_2x_3 & 0 & 0 & 2x_3 \\ 2x_1x_2 & 2x_1x_2 & 0 & x_1^2+x_2^2 & x_2x_3 & x_1x_3 & x_2 & x_1 & 0 \\ 2x_1x_3 & 0 & 2x_1x_3 & x_2x_3 & x_1^2+x_3^2 & x_1x_2 & x_3 & 0 & x_1 \\ 0 & 2x_2x_3 & 2x_2x_3 & x_1x_3 & x_1x_2 & x_2^2+x_3^2 & 0 & x_3 & x_2 \\ 2x_1 & 0 & 0 & x_2 & x_3 & 0 & 1 & 0 & 0 \\ 0 & 2x_2 & 0 & x_1 & 0 & x_3 & 0 & 1 & 0 \\ 0 & 0 & 2x_3 & 0 & x_1 & x_2 & 0 & 0 & 1 \end{pmatrix}. \tag{5.69}$$

The homogeneous quadric parameter space is 10-dimensional, with the addition of 1 to $\mathcal{W}$. The matrix Q_h is similar to Q_n, with one row and column of zeros added. The matrix M_h is computed by taking the sum over all $\mathbf{x}$ in R_k of

$$\begin{pmatrix} x_1^4 & x_1^2x_2^2 & x_1^2x_3^2 & x_1^3x_2 & x_1^3x_3 & x_1^2x_2x_3 & x_1^3 & x_1^2x_2 & x_1^2x_3 & \frac{1}{2}x_1^2 \\ x_1^2x_2^2 & x_2^4 & x_2^2x_3^2 & x_1x_2^3 & x_1x_2^2x_3 & x_2^2x_3 & x_1x_2^2 & x_2^3 & x_2^2x_3 & \frac{1}{2}x_2^2 \\ x_1^2x_3^2 & x_2^2x_3^2 & x_3^4 & x_1x_2x_3^2 & x_1x_3^3 & x_2x_3^2 & x_1x_3^2 & x_2x_3^2 & x_3^3 & \frac{1}{2}x_3^2 \\ x_1^3x_2 & x_1x_2^3 & x_1x_2x_3^2 & x_1^2x_2^2 & x_1^2x_2x_3 & x_1x_2^2x_3 & x_1^2x_2 & x_1x_2^2 & x_1x_2x_3 & \frac{1}{2}x_1x_2 \\ x_1^3x_3 & x_1x_2^2x_3 & x_1x_3^3 & x_1^2x_2x_3 & x_1^2x_3^2 & x_1x_2x_3^2 & x_1^2x_3 & x_1x_2x_3 & x_1x_3^2 & \frac{1}{2}x_1x_3 \\ x_1^2x_2x_3 & x_2^3x_3 & x_2x_3^3 & x_1x_2^2x_3 & x_1x_2x_3^2 & x_2^2x_3^3 & x_1x_2x_3 & x_2^2x_3 & x_2x_3^2 & \frac{1}{2}x_2x_3 \\ x_1^3 & x_2^2x_1 & x_3^2x_1 & x_1^2x_2 & x_1^2x_3 & x_1x_2x_3 & x_1^2 & x_1x_2 & x_1x_3 & \frac{1}{2}x_1 \\ x_1^2x_2 & x_2^3 & x_2x_3^2 & x_1x_2^2 & x_1x_2x_3 & x_2^2x_3 & x_1x_2 & x_2^2 & x_2x_3 & \frac{1}{2}x_2 \\ x_1^2x_3 & x_2^2x_3 & x_3^3 & x_1x_2x_3 & x_1x_3^2 & x_2x_3^2 & x_1x_3 & x_1x_3 & x_3^2 & \frac{1}{2}x_3 \\ \frac{1}{2}x_1^2 & \frac{1}{2}x_2^2 & \frac{1}{2}x_3^2 & \frac{1}{2}x_1x_2 & \frac{1}{2}x_1x_3 & \frac{1}{2}x_2x_3 & \frac{1}{2}x_1 & \frac{1}{2}x_2 & \frac{1}{2}x_3 & 1 \end{pmatrix}. \tag{5.70}$$

CHAPTER 6

INTEGRATION ON THE IMPLICIT-POLYNOMIAL PARAMETER MANIFOLD

This chapter is concerned with efficient evaluation of the integrals that appear in the membership probability expressions. In particular, we need to evaluate the integrals in the expression of $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$, using the implicit polynomial models (5.35) and (5.38). In general, we want to perform the integration

$$\int_{\Sigma^N} h(\mathbf{u}) d\mathbf{u} \tag{6.1}$$

in which $h(\mathbf{u})$ is some arbitrary function on the parameter manifold.

The region of integration is a half-hypersphere, and is typically high-dimensional (the dimension is two for the nonhomogeneous planar model, and eight for nonhomogeneous quadrics). Therefore, standard numerical techniques for integration are impractical. Using the method presented in this chapter, the integrals can be approximately evaluated, yielding the membership probability for the implicit polynomial model. Section 6.1 develops a parameterization which transforms the region of integration from the half-hypersphere to a unit hypercube, as shown in Figure 6.1. The half-hypersphere area normalization factors, which are needed to express $p(\mathbf{u})$, are derived in Section 6.2. Section 6.3 discusses the Monte-Carlo integration method that we use to evaluate the integrals. An alternative method, using polynomial approximation is discussed in Appendix A.2.

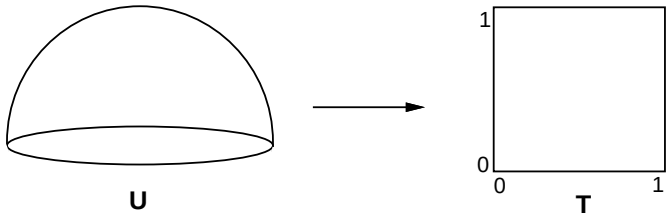


Figure 6.1 The parameter manifold is transformed into the unit hypercube for integration.

In Section 6.4, a technique is described which significantly improves the convergence of the Monte-Carlo integration. Section 6.5 describes how the Dirac delta-function approximation (discussed in Section 4.5) can be used to efficiently compute the membership probability in some cases. Finally, Section 6.6 discusses some implementation details.

6.1 Hypersphere Parameterization

As formulated in (6.1), the integral is technically an $(N-1)$-dimensional manifold integration in $\Re^N$. In this section we present an appropriate parameterization and corresponding transformation Jacobian, which will convert the integral into a standard volume integral over a hypercube region. This avoids the use of any calculus of differential forms or other intricate methods for evaluating the manifold integral. The reader should be able to follow the conversion that takes place in this section; however, a more detailed derivation of the transformation of the manifold integral into a volume integral is given in Appendix A.

The unit hypersphere, including its interior in $\Re^N$, may be parameterized as [101]

$$\begin{aligned}
u_1 &= g'_1(\mathbf{t}) \equiv t_N sin(t_1)sin(t_2)sin(t_3)\dots sin(t_{N-1}) \\
u_2 &= g'_2(\mathbf{t}) \equiv t_N sin(t_1)sin(t_2)sin(t_3)\dots cos(t_{N-1}) \\
&\vdots \\
u_{N-2} &= g'_{N-2}(\mathbf{t}) \equiv t_N sin(t_1)sin(t_2)cos(t_3) \\
u_{N-1} &= g'_{N-1}(\mathbf{t}) \equiv t_N sin(t_1)cos(t_2) \\
u_N &= g'_N(\mathbf{t}) \equiv t_N cos(t_1).
\end{aligned} \tag{6.2}$$

A similar parameterization can also be found in [102]. For the case of $N = 2$ or $N = 3$, this parameterization specializes to standard polar and spherical coordinates, respectively. On Σ^N we have $t_N = 1$, which is the case we will be interested in. For the full hypersphere we have $0 \leq t_j \leq \pi$ for $0 \leq j \leq N-2$ and $0 \leq t_{N-1} \leq 2\pi$. As discussed in the previous chapter, we are interested only in half of the hypersphere, Σ^N, and will consequently take $0 \leq t_{N-1} \leq \pi$.

The integrand of (6.1) is a function of $\mathbf{u}$, and the equations of (6.2) express each element of $\mathbf{u}$ as a function of $\mathbf{t}$. We take $\mathbf{t}$ as the N-dimensional vector of parameter variables, $[t_1 \ \dots \ t_N]$. From this the integrand will be represented as $h(g'_1(\mathbf{t}), \dots, g'_N(\mathbf{t}))$.

Recall from advanced calculus that when performing a curvilinear transformation of integrals, the Jacobian must be computed to relate a differential volume element in the initial space to the corresponding volume element in the transformed space [103]. The magnitude of the Jacobian appears in the transformed integrand (for instance, with standard spherical coordinates this factor is $t_3^2 sin(t_1)$).

The differential relationship is

$$du_1 \dots du_N = |J'_N| dt_1 \dots dt_N \tag{6.3}$$

and $|J'_N|$ represents the magnitude of the Jacobian. The Jacobian is a matrix of partial derivatives, corresponding to all pairings of elements from $\mathbf{u}$ and $\mathbf{t}$. This is represented

as

$$J'_N \equiv \frac{\partial(u_1, u_2, \ldots, u_N)}{\partial(t_1, t_2, \ldots, t_N)}. \tag{6.4}$$

For the parameterization (6.2), the magnitude of the Jacobian is [102]

$$|J'_N| = t_N^{N-1} sin^{N-2}(t_1) sin^{N-3}(t_2) \cdots sin^2(t_{N-3}) sin(t_{N-2}). \tag{6.5}$$

For our purposes, t_N is fixed at one; therefore, it can be eliminated to obtain

$$|J'_N| = sin^{N-2}(t_1) sin^{N-3}(t_2) \cdots sin^2(t_{N-3}) sin(t_{N-2}). \tag{6.6}$$

At this point, the region of integration has been converted to a hypercube of size π with respect to each axis (with t_N fixed at one). If we use the following slightly modified, nonstandard transformation, the region of integration becomes a unit hypercube. We choose this form to make the discussion in Section 6.3 simpler. The modified transformation is

$$\begin{aligned}
u_1 &= g_1(\mathbf{t}) \equiv t_N sin(\pi t_1) sin(\pi t_2) sin(\pi t_3) \ldots sin(\pi t_{N-1}) \\
u_2 &= g_2(\mathbf{t}) \equiv t_N sin(\pi t_1) sin(\pi t_2) sin(\pi t_3) \ldots cos(\pi t_{N-1}) \\
&\vdots \\
u_{N-2} &= g_{N-2}(\mathbf{t}) \equiv t_N sin(\pi t_1) sin(\pi t_2) cos(\pi t_3) \\
u_{N-1} &= g_{N-1}(\mathbf{t}) \equiv t_N sin(\pi t_1) cos(\pi t_2) \\
u_N &= g_N(\mathbf{t}) \equiv t_N cos(\pi t_1).
\end{aligned} \tag{6.7}$$

Relating to the previous transformation, we have $g_i(\mathbf{t}) = g'_i(\pi \mathbf{t})\ \forall i$. The new magnitude of the Jacobian, with $t_N = 1$, is

$$|J_N| = \pi^{N-1} sin^{N-2}(\pi t_1) sin^{N-3}(\pi t_2) \cdots sin^2(\pi t_{N-3}) sin(\pi t_{N-2}). \tag{6.8}$$

Finally, the transformed integral becomes

$$\int_0^1 \cdots \int_0^1 h(g_1(\mathbf{t}), \ldots, g_N(\mathbf{t}))\, |J_N|\, dt_1 dt_2 \ldots dt_{N-1}. \tag{6.9}$$

The volume integration technique of Section 6.3 applies directly to the integral in this form.

6.2 Expressions for the Area of Σ_N

Recall from Section 5.3.4 that we need the area of the half-hypersphere to determine prior model density (5.31). Using (6.9) with $h(\cdot) \equiv 1$, we obtain an expression for the half-hypersphere area. This requires computing

$$A_N = \int_0^1 \cdots \int_0^1 |J_N| \, dt_1 dt_2 \ldots dt_{N-1}. \tag{6.10}$$

Since the limits of integration are constant, and the integrand is a product of univariate functions, this multiple integral can be separated into a product of single integrals:

$$\prod_{k=1}^{N-2} \pi \int_0^1 sin^{N-k-1}(\pi t_k) dt_k. \tag{6.11}$$

By simple substitution and from standard integration tables we have

$$\pi \int_0^1 sin^r(\pi t) dt = \int_0^\pi sin^r(t) dt = \begin{cases} \pi & \text{if } r = 0 \\ \dfrac{2[2 \cdot 4 \cdots (r-1)]}{1 \cdot 3 \cdots r} & \text{if } r > 0 \text{ is odd} \\ \dfrac{\pi[1 \cdot 3 \cdots (r-1)]}{2 \cdot 4 \cdots r} & \text{if } r > 0 \text{ is even} \end{cases} . \tag{6.12}$$

The product in (6.11) can be expanded with (6.12) substituted, and, through cancellations, we obtain an explicit formula for the half-hypersphere area for $N \geq 3$,

$$A_N = \begin{cases} \dfrac{\pi^{(N/2)}}{(N/2-1)!} & \text{if } N \text{ is even} \\ 2^{(N-1)/2} \left[\dfrac{\pi^{(N-1)/2}}{(N-2)(N-4) \cdots 1} \right] & \text{if } N \text{ is odd} \end{cases} . \tag{6.13}$$

Recall that the prior density, $p(\mathbf{u})$ is $1/A_N$. The prior density value for four special instances mentioned in the previous chapter are presented in Table 6.1.

6.3 Monte Carlo Integration

The Monte-Carlo integration method iteratively approximates a definite integral by uniformly sampling from the domain of integration, and averaging the function values

Table 6.1 Specific prior model densities

Parameter Manifold	$p(\mathbf{u})$
nonhomogeneous planar	$1/2\pi$
homogeneous planar	$1/\pi^2$
nonhomogeneous quadric	$105/16\pi^4$
homogeneous quadric	$24/\pi^5$

at the samples. The integrand is treated as a random variable. A sampling/averaging scheme yields a parameter estimate of the mean, or expected value of the random variable. We use the most basic Monte Carlo method. More elaborate schemes with faster convergence rates are discussed in [104]; however, improvement in the convergence rate for these methods is possible only for low-dimensional cases ($N \leq 3$). For our purposes, since the dimension is high, the basic Monte Carlo integration methods is appropriate. For a complete introduction to Monte-Carlo integration, see [105].

For our purposes the basic Monte Carlo integration method is appropriate for the following reasons:

- The number of iterations required for a certain accuracy does *not* depend on the dimension of the domain of integration.

Most numerical integration techniques (such as quadrature) require some number of samples along each of the N axes. Consequently, the complexity of such methods with respect to dimension is usually high. The Monte-Carlo method is derived purely from properties of random variables and sampling, and hence does not depend on dimension. For our purposes the integration time will increase somewhat for higher degree models since, although the number of iterations is fixed, the amount of time required to evaluate the integrand increases.

- High accuracy is not an important concern, and computation efficiency can be increased (or decreased) by decreasing (increasing) the number of samples.

Since our approach represents distributions of alternative segments and segmentations, there is no great loss for some small error ($\approx 1\%$) in the membership probability computation. If the goal is to select a single, best segment or segmentation, small errors may be an issue; however, slightly incorrect membership probabilities will slightly change the ordering with respect to probability of the segments or segmentations in the resulting TSS representation. This effect can be seen in some of the experiments in Section 8.4.

- There are no computational dependencies between consecutive iterations.

Since the samples are taken independently, each of the function evaluations can be performed independently. From this, straightforward parallelization is possible by having any number of computation modules perform averages and by using recursive binary subdivision to propagate the final average.

In the derivation that follows, we treat the region of integration as a vector of random variables, denoted by $\mathbf{T}$. Take $h : \mathbf{T} \rightarrow \Re$ and $h \in L^2$.[1] The integral (6.9) is represented as

$$I(h) = \int_T h(\mathbf{t}) d\mathbf{t}. \tag{6.14}$$

Take a set of n independent samples $\mathbf{t_1}$, $\mathbf{t_2}$, $\cdots$, $\mathbf{t_n}$, drawn uniformly from the space $\mathbf{T}$. The n^{th} estimate of $I(h)$ is

$$I_n(h) = \frac{1}{n} \sum_{i=1}^{n} h(\mathbf{t_i}). \tag{6.15}$$

By the strong law of large numbers, $\|I_n(h) - I(h)\| \rightarrow 0$ as $n \rightarrow \infty$, with probability one. Consider the variance of the estimate, $\sigma_n^2 = E[I_n(h) - I(h)]^2$. From (6.15), we observe by linearity that $E[I_n(h)] = E[I(h)] = I(h)$. From this observation we obtain an expression for the variance of the estimate [106],

$$\sigma_n^2 = E[I_n^2] - 2E[I_n(h)]E[I(h)] + E[I(h)]^2 = \frac{1}{n}\{I(h^2) - [I(h)]^2\}. \tag{6.16}$$

This indicates that the error rate is reduced at a rate of $1/n$.

[1] By $h \in L^2$, we mean that $\int h^2 < \infty$.

6.4 Improving the Monte-Carlo Convergence Rate

Note that the integration accuracy depends on the difference between $I(f^2)$ and $[I(f)]^2$. In our context, this relationship affects the performance of the integrations on the conditional densities. We have determined through experimentation that this difference increases as the amount of information in the regions increases.[2] For instance, a region with more points will have greater integration error than a region with fewer points. Also, if the variance of the noise model is decreased, the amount of information increases. These changes cause the density on the parameter hypersphere to become more peaked, and the Monte Carlo convergence becomes too slow to be of practical use.

In this section, we identify a rectangular region in the domain of integration which contains nearly all of the points that significantly contribute to the integral. As the region information increases, the size of this rectangular region shrinks accordingly. The random sampling is performed only inside the rectangular box, and the number of samples required is significantly reduced (by thousands of times in many practical cases).

First we consider the case of evaluating an integral in the numerator of (5.33) for some region R_k with size n. We are evaluating the degradation density at various points throughout the integration. We can take some maximum value, $k_1 >> n\sigma^2$, such that sample points that yield a value for the sum-of-squares displacements greater than k_1 contribute relatively little to the integration, since the sum-of-squares and degradation densities asymptotically approach zero. We use the Cornish-Fischer approximation [107] to the chi-square cumulative distribution function to obtain a k_1 at the 99.9th percentile for some n. The left side of the equation below represents the set of all parameter values that yield sum-of-squares displacements less than k_1. Note that this is a subset of the

[2] We intuitively describe the concept of information, and could alternatively use the Shannon entropy on the parameter space as a formal measure.

right side:

$$\left\{\mathbf{u_k} : \frac{\mathbf{u_k}^T M \mathbf{u_k}}{\mathbf{u_k}^T Q \mathbf{u_k}} < k_1\right\} \subseteq \left\{\mathbf{u_k} : \frac{\mathbf{u_k}^T M \mathbf{u_k}}{max(\mathbf{u_k}^T Q \mathbf{u_k})} < k_1\right\}. \tag{6.17}$$

Since $\|\mathbf{u_k}\| = 1$, the maximum value in the denominator on the right above is given by the maximum eigenvalue of Q:

$$max(\mathbf{u_k}^T Q \mathbf{u_k}) = maxeig(Q) \equiv k_2. \tag{6.18}$$

Therefore, the right side of (6.19) is the equation of a solid ellipsoid, centered at $\mathbf{u_k} = \mathbf{0}$, which encloses all of the points in the parameter space that significantly contribute to the integration:

$$\left\{\mathbf{u_k} : \frac{\mathbf{u_k}^T M \mathbf{u_k}}{\mathbf{u_k}^T Q \mathbf{u_k}} < k_1\right\} \subseteq \left\{\mathbf{u_k} : \mathbf{u_k}^T M \mathbf{u_k} < k_1 k_2\right\}. \tag{6.19}$$

Let $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ denote the eigenvalues of M, in order of increasing magnitude. Also, let S denote the corresponding eigenvector matrix, which is a rotation matrix that aligns the ellipsoid with the coordinate axes (diagonalizing M). Take $\mathbf{u_k} = S\mathbf{v_k}$, and we obtain

$$\mathbf{u_k}^T M \mathbf{u_k} = (S\mathbf{v_k})^T M (S\mathbf{v_k}) = \mathbf{v_k}^T S^T M S \mathbf{v_k} = \mathbf{v_k}^T \Lambda \mathbf{v_k} \tag{6.20}$$

in which $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$. Using $\mathbf{v_k}$, the ellipsoid equation becomes

$$\sum_i \lambda_i v_i^2 < k_1 k_2. \tag{6.21}$$

The half-lengths of the principle ellipsoid axes are

$$b_i = \sqrt{\frac{k_1 k_2}{\lambda_i}}. \tag{6.22}$$

The rectangular subset of $\Re^N$ that has corners located at coordinates $\pm b_i$ encloses the ellipsoid, and the rectangular faces are tangent to the ellipsoid surface. We can apply the inverse transformation to the equations (6.2) to map the corners of the box into T. These form a rectangular subset, T', of T in which the corners have coordinates we denote by $1/2 \pm c_i$.[3]

[3]Some of rectangular faces in the parameter space may lie outside the unit hypersphere. When the first axis is found that is outside, the remaining c_i are set to their maximum value, 1/2.

The portion of parameter space that significantly contributes to the resulting integral lies within T', hence we need to draw samples uniformly only from T'. Using these results, an integral in the numerator of (5.33) can be computed by

$$\frac{1}{nF}\sum_{\mathbf{v_k}} f\left(\frac{\sum_i \lambda_i v_i^2}{\mathbf{v_k}^T S^T Q S \mathbf{v_k}}\right). \tag{6.23}$$

Above, F represents the ratio of the area of T' to T, and n is the number of samples, $\mathbf{v_k}$, that are used. The F also represents the factor by which the number of required samples is reduced.

To compute the denominator integral, we use the smaller rectangular region, T', (and corresponding rotation matrix S) of the two regions R_i and R_ρ. If that region is R_ρ, then the integral is computed by

$$\frac{1}{nF}\sum_{\mathbf{v_k}} f\left(\frac{\sum_i \lambda_i v_i^2}{\mathbf{v_k}^T S^T Q_\rho S \mathbf{v_k}}\right)\ f\left(\frac{\mathbf{v_k}^T S^T M_i S \mathbf{v_k}}{\mathbf{v_k}^T S^T Q_i S \mathbf{v_k}}\right) \tag{6.24}$$

in which the λ_i are the eigenvalues of M_1 and S is its eigenvector matrix.

The integral convergence is fastest using the nonhomogeneous parameter space, but we have also found it useful for performing integrations on the homogeneous parameter manifold. For nonhomogeneous planes, the rectangular box T' obtained from this method produces a very tight bound on the peak, yielding nearly constant performance, regardless of the amount of information in the region. For quadrics, we also obtain a significant reduction in the number of samples, allowing the practical computation of region merging probabilities.

6.5 Using the Dirac Delta Model to Compute the Membership Probability

Recall from Section 4.5 that if estimation can be reliably performed for one of the regions, a simplified version of the membership probability computation results. In this

section we discuss how this model can be used in some cases, particularly when the Monte-Carlo integration becomes inefficient.

As will be discussed in Chapter 7, it is best to select an initial region that contains the most information regarding the parameter manifold. In other words, it is best to select the region that is most likely to have a peaked parameter manifold distribution. In our experiments, we simply selected regions with the most data points. From the discussion in the previous section, this can, however, cause problems for the Monte-Carlo integration. For these cases, if estimation can be reliably performed, the delta model is appropriate. Using the nonhomogeneous parameter manifold, the goal is to estimate the parameter value for R_i, denoted by $\hat{\mathbf{u}}_\mathbf{i}$ (we can also compute estimates for d_i and d_ρ). Using (4.89), the expression for $\lambda_1(\mathbf{y}_\rho, \mathbf{y}_\mathbf{i})$ in the context of implicit polynomial surfaces becomes

$$\lambda_1(\mathbf{y}_\rho, \mathbf{y}_\mathbf{i}) = \frac{\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u}_\rho, \hat{d}_\rho(\mathbf{u}_\rho))p(\mathbf{u}_\rho)\mathbf{d}\mathbf{u}_\rho}{p(\mathbf{y}_\rho|\hat{\mathbf{u}}_\mathbf{i}, \hat{d}_i)}. \tag{6.25}$$

Hence, we need to obtain the estimate $\hat{\mathbf{u}}_\mathbf{i}$. For performing the estimate, we assume that the identity maps model is being used. The maximum likelihood estimate (MLE) of $\mathbf{u}_\mathbf{i}$ will be the value yielding the smallest sum-of-squares displacements

$$\sum_{\mathbf{x}\in R_i} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u}_\mathbf{i}, d_i))]^2 \approx \frac{\mathbf{u}_\mathbf{i}{}^T M_n \mathbf{u}_\mathbf{i}}{\mathbf{u}_\mathbf{i}{}^T Q_n \mathbf{u}_\mathbf{i}}. \tag{6.26}$$

For our experiments, we estimated $\mathbf{u}_\mathbf{i}$ by minimizing the numerator quadratic form. From linear algebra, it is known that the minimum of the quadratic form $\mathbf{u}_\mathbf{i}{}^T M_n \mathbf{u}_\mathbf{i}$ is attained by selecting $\hat{\mathbf{u}}_\mathbf{i}$ as the eigenvector corresponding to the minimum eigenvalue of M_n.

Parameter estimation is not the primary focus of our work, and for our experiments with the delta model, this estimation technique was simple and usually produced reasonable results. A better solution to this estimation problem has been used by Taubin [95]. In our context, his method corresponds to minimizing the homogeneous ratio to estimate

$\mathbf{u_i}$,

$$\sum_{\mathbf{x}\in R_i} [\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_i}))]^2 = \frac{\mathbf{u_i}^T M_h \mathbf{u_i}}{\mathbf{u_i}^T Q_h \mathbf{u_i}}. \tag{6.27}$$

This form is minimized by determining the *generalized eigenvectors* and *generalized eigenvalues* of the matrix equation

$$\mathbf{u_i}(M_h - \lambda Q_h) = 0. \tag{6.28}$$

The solution is given by the minimum eigenvector/eigenvalue pair.

Both of the methods discussed use the approximation given by (5.54). In general, one could alternatively apply a nonlinear optimization technique, such as Levenberg-Marquardt [108], directly to the minization of

$$\sum_{\mathbf{x}\in R_k} \frac{[\phi(\mathbf{x}, \mathbf{u_k})]^2}{\|\nabla_x \phi(\cdot, \mathbf{u_k})\|^2}. \tag{6.29}$$

This does not use the approximation in (5.54); however, this optimization will be more costly than the linear, eigenvector approaches.

6.6 Some Efficiency Issues

6.6.1 Efficient function evaluation

In this section, we provide the details necessary to efficiently evaluate the integrand at each iteration of the Monte-Carlo integration, using the nonhomogeneous model. We also use, however, the homogeneous matrices, M_h and Q_h. For each region in $\mathcal{R}$, we precompute and store the matrices M_n, M_h, Q_n, Q_h, discussed in Section 5.4.3.

First, consider one of the two integrals appearing in the numerator of (5.41). The operations in the j^{th} iteration of the integration procedure on R_k are

1. Obtain the parameter value $\mathbf{u_j}$.

2. Compute $\hat{y}_k$ from the nonhomogeneous quadratic form ratio, (5.58).

3. Use degradation density to compute $p(\hat{y}_k|\mathbf{u_j}, \hat{d}_k)$.

4. Multiply the Jacobian value, (6.8), to result of 3, and add to the total.

For the first step, the parameter value is determined from the given sample points using (6.7). For the second step, recall that the nonhomogeneous quadratic form, (5.58), implicitly encodes the optimal estimate $\hat{d}_k$. From this the observation value, $\hat{y}_k$, is directly obtained. Note that since the matrices are symmetric, only the upper triangles of M_n and Q_n are needed. Consequently, off-diagonal elements are multiplied by two. The third step involves substituting the values directly into the degradation density. Finally, the magnitude of the Jacobian at that sample point is determined and multiplied by the value obtained from the third step.

A single evaluation of the integrand in the denominator of (5.41) requires more work than the previous integrals. The operations in the j^{th} iteration are

1. Obtain the parameter value $\mathbf{u_j}$.

2. Compute optimal estimate $\hat{d}_{i\rho}$.

3. Compute $\hat{y}_i$ from homogeneous quadratic form ratio, (5.55).

4. Compute $\hat{y}_\rho$ from homogeneous quadratic form ratio, (5.55).

5. Use degradation density and estimates to compute $p(\hat{y}_i|\mathbf{u_i}, \hat{d}_{i\rho})$.

6. Use degradation density and estimates to compute $p(\hat{y}_\rho|\mathbf{u}_\rho, \hat{d}_{i\rho})$.

7. Multiply Jacobian and results from 5 and 6, and add to the total.

The first step is the same as before. In the second step, the optimal estimate for $\hat{d}_{i\rho}$ is directly computed directly from (5.51). The coordinate sums are stored for each region, and the sums for R_i and R_ρ are combined to yield the estimate.

Although we are using the nonhomogeneous model, the homogeneous quadratic form is used in the third and fourth steps. The estimate $\hat{d}_{i\rho}$ is treated as the $(N+1)^{th}$ parameter value, creating the vector

$$[u_1 \;\; u_2 \;\; \ldots \;\; u_N \;\; \hat{d}_{i\rho}]^T. \tag{6.30}$$

This vector is substituted directly into (5.55), using M_h and Q_h for each of R_i and R_ρ. Hence, the observation $\hat{y}_i$ and $\hat{y}_\rho$ values are obtained. Again, we store and work only with the upper triangles of the matrices. The observations are substituted into the degradation densities in the fifth and sixth steps. Finally, the two degradation density values are multiplied with the magnitude of the Jacobian to yield the result.

6.6.2 Multiple use of some integrals

Several further computational savings are possible for the IE-independent model. Two of the three integrals in the expression for $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$, (4.43), involve only a single region. Once these integrals are computed, we can typically expect the regions to appear in many other membership probability computations.

The initial-region integral,

$$\int_{\Sigma^N} p(\mathbf{y_i}|\mathbf{u}) d\mathbf{u}, \tag{6.31}$$

will be used for every membership probability computation performed in the construction of a TSS representation. It needs to be computed only once during the construction of the TSS representation.

Also, as we saw in the example in Chapter 3, one region is often used for refinement numerous times in a TSS. Hence

$$\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u}) d\mathbf{u} \tag{6.32}$$

can be stored and recalled many times.

If a full image segmentation is required, and the delta-function model is used at each refinement, all individual region integrals can be performed as a preprocessing stage. Recall that with this model the denominator integral of $\lambda_1(\mathbf{y}_\rho, \mathbf{y}_\mathbf{i})$,

$$\int_{\Sigma^N} p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{y}_\mathbf{i}|\mathbf{u})d\mathbf{u}, \tag{6.33}$$

reduces to a single function evaluation. Hence, the TSS and SSS contructions are performed very rapidly, since membership probabilities are quickly computed.

CHAPTER 7

TSS AND SSS ALGORITHMS

In Chapter 3, a framework was introduced which allows us to construct and represent distributions of segments and segmentations efficiently. In particular, the refinement operation is a single step used in the construction of the representations. Subsequently, Chapters 4-6 addressed the issue of assigning probabilities to refined events. The primary purpose of this chapter is to present efficient algorithms for building TSS and SSS representations. These algorithms require iterations of the refinement operation and corresponding membership probability computation. The algorithms presented here were used to obtain experimental results, which appear in Chapter 8.

Section 7.1 addresses the problem of generating a TSS representation. An algorithm is presented which returns a user-specified number, n, of ground events that have highest probability. A termination criterion is given, which, if met, guarantees that the top n segments were determined (i.e., there does not exist another segment with higher probability than one of the returned segments). Section 7.2 presents one algorithm to obtain a single, good segmentation and two algorithms for generating an SSS representation. A termination criterion that guarantees that the algorithm returns the best n segmentations is presented, but we expect that in many practical situations this criterion cannot be achieved with moderate computation. Hence, these algorithms can sometimes be considered as searches for several good segmentations. Section 7.2 concludes with

descriptions of modifications required to represent a distribution over a set of m-segment groupings.

7.1 Constructing TSS Representations

This section presents algorithms that explicitly represent a subset of the set of segments in Θ_i, in which each segment has a higher probability than the complementary set of segments. We consider the high-probability segments to be the interesting portion of the TSS. Given the large size of Θ_i that we expect in real applications (with dozens or hundreds of regions in $\mathcal{R}$), obtaining the segments that represent the interesting portion of the TSS may seem to be an unreasonable expectation. There are, however, some factors which enable us to approach this task. It is important to remember the difference between the size of Θ_i, and the number of segments that we find interesting. Suppose, for example, that the segments that have high probability contain only three of the four regions, although there are hundreds of regions in $\mathcal{R}$. In a situation such as this, there will be thousands of segments with very low probability in the TSS (for instance, the segment $\mathcal{R} \in \Theta_i$ if D is connected).

We want to determine a sequence of refinements that efficiently leads to a TSS representation. Specifically, we want to represent some number of ground events (corresponding to single segments) that have high probability. Note that each refinement increases the number of events in a cover by one (one event is removed, and replaced by two). Hence, restricting the total number of refinements and obtaining a compact representation of the TSS are closely related. Recall that the refinement event B_ρ and the refinement

region R_ρ completely determine a refinement mapping $\rho(C, B_\rho, R_\rho)$ to apply to a cover C. Hence, the following two issues pertain to the construction of a TSS representation:

- B_ρ should be selected to maximize the probability that it contains interesting ground events.
- R_ρ should be chosen to maximize the efficiency of the representation.

Recall that we represent some event $B \in \mathcal{B}_i$ using the sets I and E. We do not have direct access to specific segments contained in B, or even the number of segments in B. Determining these requires performing enough refinements to divide B into all of its ground events. We argue that if we are given a cover C, then the element $B \in C$ that has highest probability should be selected for refinement, since it is most likely to contain ground events with larger probabilities (assuming there is no information about the number of segments in each event).

The refinement region, R_ρ, which of course depends on the choice for B_ρ, is chosen in a way that uses the strongest evidence first. In other words, out of all of the choices for R_ρ, we seek the region that gives a membership probability closest to zero or one. The motivation for this choice is that in constructing a compact representation, it is beneficial to have the largest portion of the TSS possible (in terms of the number of segments) belonging to a set with low probability. This greatly reduces the number of segments that are still in consideration after the refinement. Considering the set of all valid choices for R_ρ, a membership probability close to zero will assign low probability to the event $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$. If the probability is sufficiently low, the new event will not be considered for subsequent refinement. The complementary event, $\tau(I_\rho, E_\rho \cup \{R_\rho\})$, will have high probability, and is likely to be selected for further refinement. An analogous situation occurs when the membership probability is close to one.

We need to define the set of possible alternatives for R_ρ for a single refinement step. Recall the constraints on I and E from Section 3.3. The sets I and E are mutually disjoint

sets of regions, I is connected, and each element in E is adjacent to some element in I. Our choice for R_ρ must create valid include and exclude sets: $I \cup \{R_\rho\}$ and $E \cup \{R_\rho\}$. We call the set of possible refinement regions for a given B_ρ, the *frontier*. This is defined and denoted as

$$\textbf{FRONTIER}(B_\rho) = \{R_k \in \mathcal{R} : \ R_k \notin I_\rho \cup E_\rho \text{ and } \{R_k\} \cup I_\rho \text{ is connected}\}. \tag{7.1}$$

For some set of regions, $\mathcal{R}_i$, as possible choices for refinement, we select $R_\rho \in \mathcal{R}_i$ by

$$\textbf{SELECT-}R_\rho(\mathcal{R}_i) = \underset{R_k \in \mathcal{R}_i}{argmin} \{min \ [\ P(\tau(I_\rho \cup \{R_k\}, E_\rho)|e), \ P(\tau(I_\rho, E_\rho \cup \{R_k\})|e) \ ]\} \, . \tag{7.2}$$

We use e to denote a generic piece of evidence as in Chapter 3. The expression selects the R_ρ that yields a membership probability closest to zero or one. We will often take $\mathcal{R}_i$ as the frontier set.

To present the algorithms formally, we will use some standard definitions, found in [109]. A *priority queue* is a finite set of elements with a linear ordering. An element, a, is inserted into a priority queue, X, (in a position that maintains the linear ordering) with $\textbf{INSERT}(X, a)$. The highest-priority element in X is removed and returned by $\textbf{EXTRACT-MAX}(X)$. The highest-priority element in X is referred to as $\textbf{MAX}(X)$ (without being removed). We will refer to the n^{th} element in the queue as $\textbf{NTH}(X, n)$. In practice, the priority queue can be efficiently implemented using heaps [109].

Recall that a cover C represents a set of TSS events that form a partition of Θ_i. This cover may contain ground events and nonground events. We divide the cover into two priority queues, C_g and C_h, such that

$$C = C_g \cup C_h \quad \text{and} \quad C_g \cap C_h = \emptyset, \ . \tag{7.3}$$

The queue C_g contains the ground events of C, and C_h contains the remaining, nonground events of C. We define the ordering on the queue to be $>$, applied to the probabilities of the events. Hence $\textbf{EXTRACT-MAX}(C_g)$ returns the ground event with highest probability.

The queue C_g contains the events that represent individual segments, which will eventually form the distribution of ground-event probabilities. The queue C_h represents the portion of the TSS left to be explored, and the first element of C_h is the unexplored event that has highest probability. Together, these two lists always determine a cover and hence an approximate representation of the TSS, as discussed in Chapter 3.

We will represent each element of C_h by its I and E sets. Hence, an event will be represented in the algorithms as $\{I, E\}$, implicitly representing the set of segments given by $\tau(I, E)$. For some event B_x in the algorithm descriptions that follow (here x represents some generic subscript), we will directly use its include and exclude sets, I_x and E_x. Thus, $B_x = \tau(I_x, E_x)$ will be implicitly represented by $\{I_x, E_x\}$.

We are interested in determining some number of ground events and their corresponding probabilities. One natural way to express this goal is to request a representation of n segments that have high probability.[1] The ideal situation would be to obtain the n segments with the highest probability out of all ground events in $\mathcal{B}_i$. The following proposition gives a termination criterion that achieves this:

Proposition 3 *If* $P(\textbf{NTH}(C_g, n)) > P(\textbf{MAX}(C_h))$, *then for any* $T \in \Theta_i$, *such that* $\{T\} \notin C_g$, $P(\{T\}) < P(\textbf{NTH}(C_g, n))$.

Proof Suppose to the contrary that $\exists T \in \Theta_i$ such that $\{T\} \notin C_g$, and $P(\{T\}) \geq P(\textbf{NTH}(C_g, n))$. Since $C_g \cup C_h = C$ (which covers Θ_i) and $\{T\} \notin C_g$, there is some $B \in C_h$ such that $\{T\} \in B$. This implies $P(\{T\}) \leq P(B)$. We also have $P(B) \leq P(\textbf{MAX}(C_h))$. This leads to a contradiction since:

$P(\textbf{MAX}(C_h)) < P(\textbf{NTH}(C_g, n)) \leq P(\{T\}) \leq P(B) \leq P(\textbf{MAX}(C_h))$. □

[1] Another reasonable request is to obtain some number of ground events, with total probability above some threshold (e.g. above 0.99).

7.1.1 An IE-independent algorithm which builds a TSS representation

We now present an algorithm for constructing a TSS representation, which exploits the computational savings of the IE-independent model. Recall that these savings are possible since under this model, the membership probability depends only on R_ρ, and not on the other regions in I_ρ or E_ρ. This implies that the membership probability (derived from $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$) can be precomputed and stored with each region in $\mathcal{R}$. When a region is selected for a refinement, the membership probability has already been determined. Hence for a TSS algorithm, using the IE-independent model, there will be fewer than $|\mathcal{R}|$ membership computations, since these represent all possible refinement regions, regardless of B_ρ. A similar approach is taken for agglomerative clustering, when the merging likelihood function is precomputed for all adjacent-region pairs in the image [43].

We could choose to determine the membership probabilities as needed, to avoid the possibility of unnecessary computations that correspond to low-probability events. However, making the selection for R_ρ with (7.2) requires knowing the membership probabilities for each region in the frontier set.

We compromise between the two extremes by incrementally computing the membership probabilities for the frontier set of each new refinement event, B_ρ. This way there are a sufficient number of membership probabilities available to use (7.2), and we will not be forced to compute the membership probabilities for all regions in $\mathcal{R} - \{R_i\}$.

The algorithm accepts as inputs: the regions, $\mathcal{R}$; the initial region, R_i; and the number of greatest-probability ground events to determine, n. We assume that the adjacency structure for $\mathcal{R}$ has been given. Note that we store the probabilities of any events that appear in C_h or C_g, but to keep the algorithm specification concise, these will not be explicitly represented.

```
GET-TOP-SEGMENTS (R, R_i, n)
1    NEW-MEM-PROBS(R_i, FRONTIER({{R_i}, ∅}))
2    C_g ← ∅;  C_h ← ∅
3    INSERT(C_h, {{R_i}, ∅})
4    repeat
5        B_ρ ←EXTRACT-MAX(C_h)
6        R_ρ ←SELECT-R_ρ (FRONTIER(B_ρ))
7        B_I ← {I_ρ ∪ {R_ρ}, E_ρ};  B_E ← {I_ρ, E_ρ ∪ {R_ρ}}
8        Compute and store P(B_I|e) and P(B_E|e)
9        if FRONTIER(B_I) ≠ ∅
10           then
11               INSERT(C_h, B_I)
12               NEW-MEM-PROBS(R_i, FRONTIER(B_I))
13           else
14               INSERT(C_g, B_I)
15       if FRONTIER(B_E) ≠ ∅
16           then
17               INSERT(C_h, B_E)
18           else
19               INSERT(C_g, B_E)
20   until [C_h = ∅] or [ |C_g| ≥ n and P(NTH(C_g, n)) > P(MAX(C_h))]
21   return C_g
```

Figure 7.1 An IE-independent algorithm returning the best n segments.

The formal algorithm description is presented in Figure 7.1. The function **NEW-MEM-PROBS** accepts as arguments, R_i, and some set of regions, e.g., $\mathcal{R}_i$. For any $R_k \in \mathcal{R}_i$ that is considered for the first time, the membership probability, $P(\tau(\{R_i, R_k\}, \emptyset)|e)$, is computed and stored. As more refinements are performed, the number of membership probabilities that are actually computed by **NEW-MEM-PROBS** decreases, since many regions are reconsidered.

Line 1 of GET-TOP-SEGMENTS determines the membership probabilities for the regions adjacent to R_i. These are sufficient for the selection of the first refinement region. Lines 2 and 3 initialize C_g and C_h. There are no ground events determined yet, hence $C_g = \emptyset$. We initialize C_h to contain an implicit representation of $\tau(\{R_i\}, \emptyset)$. The two priority queues together initially represent the entire TSS with a single event.

Lines 5 to 19 are repeated until the termination condition in Line 20 is met. Each iteration corresponds to a single refinment made to the cover $C_g \cup C_h$. The second part of the termination condition, $[\,|C_g| \geq n$ and $P(\textbf{NTH}(C_g, n)) > P(\textbf{MAX}(C_h))]$, uses Proposition 3 with the constraint that at least n ground events have been determined. Termination by this part of the condition guarantees that the n segments that have highest probability are represented. The $C_h = \emptyset$ portion of the condition is met when all ground events in Θ_i have been represented (i.e., no further expansion is possible). In practice, we can also terminate if $P(\textbf{FIRST}(C_h)) < \epsilon$ for some $\epsilon \approx 0$, or if $|C_g|$ exceeds some space limit (e.g., $20n$), without having met the criterion in Proposition 3. These final two conditions are merely heuristics and are used only to force an early termination, if desired.

Each iteration creates a new cover by removing an element from C_h, performing a refinement, and defining the next cover by adding the new events to C_h or C_g as appropriate. Line 5 removes the highest-probability element from C_h for refinement. Line 6 chooses the refinement region, R_ρ, by comparing the membership probabilities for elements of the frontier set, as prescribed by (7.2). Line 7 constructs the two refined events, B_I and B_E. In Line 8, the membership probability for the pair of regions, R_i and R_ρ, is used to compute and store $P(B_I|e)$ and $P(B_E|e)$, using (3.8) and (3.9).

The remaining portion of the iteration decides whether the new events, B_I and B_E, belong to C_h or C_g. Line 9 is equivalent to testing whether B_I is a TSS nonground event. If the condition is true, then Line 11 adds B_I to C_h. The addition of B_I to C_h may cause more membership probabilities to be evaluated (these are needed in advance for Line 6 since a new region, R_ρ, has been considered). If B_I is a ground event, then it is added to C_g in Line 14. Lines 15 to 19 proceed for B_E as Lines 9 to 14 do for B_I. Since B_E excludes R_ρ, no further membership probability computations are necessary (because the frontier has not been extended).

7.1.2 An IE-dependent algorithm which builds a TSS representation

With the IE-independent model, the membership probabilities are computed in batches, as the frontier is expanded. For the IE-dependent model, the membership probabilities must be computed at each refinement. The membership probabilities will, in general, be distinct for different refinements with the same R_ρ, since the probability now depends on I_ρ, which will be distinct for different iterations. Hence we do not perform the computation of membership probabilities for regions appearing in the frontier set, as with GET-TOP-SEGMENTS.

Also, we do not have the luxury of knowing in advance the membership probabilities needed to use (7.2). If the cost of computing the membership probability is not a concern, we could, of course, compute the membership probability for every R_k in the frontier set, and then use (7.2) to choose R_ρ.

In practice, we can make the refinement region selection based on the amount of information contained in a given region.[2] The region with the most information tends to yield a membership probability that is closer to zero or one. This consideration guided the selection of the refinment region for GET-TOP-SEGMENTS. One heuristic measure for this is the size of the region, which is what was used in our experiments. We could alternatively use the factor F from (6.23), which corresponds to the degree to which the conditional density over the parameter space is peaked. For the IE-dependent model, the modified version of GET-TOP-SEGMENTS is presented in Figure 7.2.

GET-TOP-SEGMENTS-DEP is similar to GET-TOP-SEGMENTS, with the differences being the location of membership probability computations, and **SELECT-**R_ρ selects the largest region. The function **MEM-PROB** computes and stores the membership prob-

[2]We intuitively describe the concept of information, and could alternatively use the Shannon entropy on the parameter space as a formal measure.

```
GET-TOP-SEGMENTS-DEP (R, R_i, n)
1     C_g ← ∅;  C_h ← ∅
2     INSERT(C_h, {{R_i}, ∅})
3     repeat
4         B_ρ ←EXTRACT-MAX(C_h)
5         R_ρ ←SELECT-R_ρ (FRONTIER(B_ρ))
6         MEM-PROB(B_ρ, R_ρ)
7         B_I ← {I_ρ ∪ {R_ρ}, E_ρ};  B_E ← {I_ρ, E_ρ ∪ {R_ρ}}
8         Compute and store P(B_I|e) and P(B_E|e)
9       if FRONTIER(B_I) ≠ ∅
10            then
11                INSERT(C_h, B_I)
12            else
13                INSERT(C_g, B_I)
14        if FRONTIER(B_E) ≠ ∅
15            then
16                INSERT(C_h, B_E)
17            else
18                INSERT(C_g, B_E)
19    until [C_h = ∅] or [ |C_g| ≥ n and P(NTH(C_g, n)) > P(MAX(C_h))]
20    return C_g
```

Figure 7.2 An IE-dependent algorithm returning the best n segments.

ability, which depends on the refinement event, B_ρ (using I_ρ), and the refinement region, R_ρ.

7.2 Constructing SSS Representations

The next step is to consider the distributions arising from combinations of different segments in the image. This leads to a distribution over the SSS ground events, each representing a full-image segmentation. Our goal is analogous to that of the previous section: we wish to consider covers of the SSS (which will be defined shortly), partitioned into ground and nonground SSS events, and to efficiently obtain a representation of the SSS ground events with the highest probability. The fundamental difference is that now

we are concerned with the concept of covers and ground events pertaining to the SSS, as opposed to the TSS.

Three algorithms will be presented. GREEDY-SEGMENTATION is the simplest algorithm, and generates a single segmentation by taking the best segments from a sequence of applications of GET-TOP-SEGMENTS (or GET-TOP-SEGMENTS-DEP) with $n = 1$. The other two algorithms generate a distribution over SSS ground events. BEAM-SEARCH-SEGMENTATIONS uses GET-TOP-SEGMENTS (or GET-TOP-SEGMENTS-DEP) as a single iteration and performs a search on the space of segmentations to obtain a set of segmentations that have high probability. GET-TOP-SEGMENTATIONS is similar to GET-TOP-SEGMENTS, except SSS refinements and covers are used as opposed to TSS refinements and covers.

In the algorithms that follow, it is necessary at various stages to choose a new initial region, R_i, to begin the construction of a new TSS. We can describe the ideal choice of an initial region using the concepts from Chapter 4, although in practice, some empirically based, heuristic choice of initial region may be used. We want to select the region that contains the most information about the parameter space. If there is enough information in some region that its parameter value is essentially known, then the delta model can be employed. That region would be an ideal choice, since the membership probability computations would be significantly reduced.

This choice of regions containing more information tends to cause more extreme values for $\lambda_1(\mathbf{y}_\rho, \mathbf{y_i})$ (tending toward zero or infinity), yielding a more compact TSS representation. If there is little information present in the initial region, then the resulting TSS probabilities are close to their prior values, and numerous competing TSS ground-events are obtained. This effect can be seen in our experiments in Section 8.3. In practice, we used the size of a region as an estimate of the amount of information present, and hence chose the largest regions as initial regions. If the goal is to obtain a distribution

```
GREEDY-SEGMENTATION (R)
1     S ← ∅
2     R_r ← ∅
3     repeat
4         R_i ←SELECT-INITIAL-REGION(R − R_r)
5         {B_g} ←GET-TOP-SEGMENTS(R − R_r, R_i, 1)
6         R_r ← R_r ∪ I_g
7         S ← S ∪ {I_g}
8     until R_r = R
9     return S
```

Figure 7.3 A greedy algorithm which returns one good segmentation.

over several adjacent segments, then the choice of initial regions may be restricted to regions that are adjacent to previously considered segments.

7.2.1 A greedy segmentation algorithm

If the goal is to obtain a single segmentation that has high probability, the following algorithm can be used. It is termed *greedy* since it picks the best segment from a TSS distribution at each iteration, in a sequence of iterations that produces a resulting segmentation. This can also be considered as a kind of *best-first* search, often appearing in AI algorithms [110]. The formal specification is given in Figure 7.3.

Line 1 initializes the variable, S, which will hold the segmentation. The variable $\mathcal{R}_r$ holds a list of regions that are removed from consideration at a given iteration, and is initialized to $\emptyset$ in Line 2. Lines 4 to 7 generate a segment to add to S. Line 4 selects the next initial region for a new TSS, removing the regions in $\mathcal{R}_r$ from consideration. In Line 5, GET-TOP-SEGMENTS (or GET-TOP-SEGMENTS-DEP) is used to obtain a single, best segment from the TSS constructed for R_i. The left side of the assignment is written as $\{B_\rho\}$ since we want to select the one ground event from the singleton set of events returned by GET-TOP-SEGMENTS. The variable $\mathcal{R}_r$ is extended in Line 6 so that the regions in the new segment will not be considered in subsequent iterations. Line

7 adds the new segment to the list of segments, which eventually becomes the full-image segmentation. Line 8 is the termination condition, which is met when the entire image is represented in S.

7.2.2 Algorithms for obtaining a distribution of segmentations

The next two algorithms build an approximate representation of the SSS. Just as the approximate TSS representation yielded a distribution over a set of segments, the approximate SSS representation yields a distribution over a set of segmentations. We first define an implicit representation of SSS events, which is analogous to the $\tau(I, E)$ representation of TSS events. We next define cover and refinement on the SSS, which are analogous to their TSS counterparts. Once these are introduced, we present the algorithms, BEAM-SEARCH-SEGMENTATIONS and GET-TOP-SEGMENTATIONS. It will be seen that BEAM-SEARCH-SEGMENTATIONS is a generalization of GREEDY-SEGMENT that considers b alternative segments from each TSS construction (GREEDY-SEGMENT uses *one* segment from each TSS construction). GET-TOP-SEGMENTATIONS uses SSS refinements and SSS covers to build an SSS representation in the same manner used in GET-TOP-SEGMENTS to build an TSS representation. A termination criterion analogous to the one given by Proposition 3 is used to guarantee that the best n segmentations in the SSS have been explicitly represented.

Recall from Section 3.5.2 that SSS representations can be built from multiple TSS representations. Given a TSS ground event, $\{T_1\}$, we can consider a second TSS by choosing an initial region not contained in T_1. The second TSS builds a set of segments that do not include any regions that are members of T_1. Each TSS ground event, e.g., $\{T_2\}$, in the second TSS corresponds to a pair of segments, $\{T_1, T_2\}$, in the original image. With respect to the SSS, the TSS ground event in the second TSS represents the set of all

segmentations containing both T_1 and T_2. From the TSS ground event $\{T_2\}$, a third TSS can be considered, and this process can be iterated until the entire image is represented.

We will argue that any SSS event that is constructed in the manner presented in Section 3.5.2 can be implicitly represented by a set of segments, F, an include set, I, and an exclude set E. The elements of F are the segments obtained from the ground events in the sequence of TSS constructions. The I and E sets are the include and exclude sets of the current TSS construction, which represents some TSS event, $\tau(I, E)$. In this context, we use $(\Theta_i, \mathcal{B}_i, P)$ to denote the current TSS.

Formally, we consider a function σ:

$$\sigma(F, I, E) = \{S \in \Pi : T \in S \ \forall T \in F\} \bigcap f_i(\tau(I, E)). \tag{7.4}$$

As defined in Section 3.5.2, f_i is the function that maps TSS events in $\mathcal{B}_i$ to their corresponding SSS events. In this context, the corresponding SSS event is

$$f_i(\tau(I, E)) = \bigcup_{T \in \tau(I,E)} \{S \in \Pi : T \in S\}. \tag{7.5}$$

Of the segmentations represented by (7.5), (7.4) represents only those that contain all of the segments in F. It is assumed that the segments in F were obtained from previous TSS constructions, and hence are fixed in the event $\sigma(F, I, E)$.

An *SSS cover*, $\mathbf{C}$, is a set of pairwise-disjoint events in $\mathcal{A}$ that form a partition of Π. An SSS cover may contain SSS ground events and/or nonground events. We consider (as was done with a TSS cover) dividing the SSS cover into two priority queues, $\mathbf{C}_g$ and $\mathbf{C}_h$, such that

$$\mathbf{C} = \mathbf{C}_g \cup \mathbf{C}_h \quad \text{and} \quad \mathbf{C}_g \cap \mathbf{C}_h = \emptyset. \tag{7.6}$$

The queue $\mathbf{C}_g$ contains the SSS ground events of $\mathbf{C}$. The queue $\mathbf{C}_h$ contains the remaining, nonground events of $\mathbf{C}$. Again, we define the queue ordering to be $>$, applied to the probabilities of the events.

In Section 3.3.2 we introduced the TSS refinement operation, and now we introduce the SSS refinement operation. An event, $A_\rho = \sigma(F_\rho, I_\rho, E_\rho)$, is selected as the *SSS refinement event.* A region, R_ρ, is selected as the *SSS refinement region.* We constrain R_ρ to be adjacent to some region in I_ρ and not in any of F_ρ, I_ρ, or E_ρ. An SSS refinement is obtained by applying the TSS refinement operation on the I_ρ and E_ρ sets in $\sigma(F_\rho, I_\rho, E_\rho)$. The SSS refinement operation applies to an SSS cover, $\mathbf{C}$, and yields a new cover, $\mathbf{C}'$, in which A_ρ is replaced by two *SSS refined events.*

Although SSS refinement applies the TSS refinement operation using the I_ρ and E_ρ sets of $\sigma(F_\rho, I_\rho, E_\rho)$, there are two cases that must be treated differently. We are concerned with the TSS event $\tau(I_\rho, E_\rho)$. If this event is not a TSS ground event, then an SSS refinement will be formed by performing a TSS refinement operation on $\tau(I_\rho, E_\rho)$. If $\tau(I_\rho, E_\rho)$ represents a TSS ground event, then a new initial region must be selected, which initiates the construction of a new TSS. These cases will now be described in further detail.

First, consider the case in which $A_\rho = \sigma(F_\rho, I_\rho, E_\rho))$, and $\tau(I_\rho, E_\rho)$ represents a nonground TSS event. The two SSS refined events are

$$A_I = \sigma(F_\rho, I_\rho \cup \{R_\rho\}, E_\rho) \tag{7.7}$$

and

$$A_E = \sigma(F_\rho, I_\rho, E_\rho \cup \{R_\rho\}). \tag{7.8}$$

We have the following proposition:

Proposition 4 *If $A_\rho = \sigma(F_\rho, I_\rho, E_\rho))$, and $\tau(I_\rho, E_\rho)$ represents a nonground TSS event, then A_I and A_E, given above, form a disjoint partition of A_ρ.*

Proof Using the definition of σ, (7.4), the intersection of A_I and A_E is

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap f_i(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) \bigcap f_i(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{7.9}$$

This set is empty since $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and $\tau(I_\rho, E_\rho \cup \{R_\rho\})$ represent TSS refinement events (which are known to be disjoint), and the application of f_i to each of them yields disjoint sets of segmentations (recall (3.32)).

Taking the union of A_I and A_E we obtain

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap \left[f_i(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) \bigcup f_i(\tau(I_\rho, E_\rho \cup \{R_\rho\})) \right]. \tag{7.10}$$

This is equivalent to

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap f_i(\tau(I_\rho, E_\rho)) = A_\rho \tag{7.11}$$

since

$$f_i(\tau(I_\rho, E_\rho)) = f_i(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) \cup f_i(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{7.12}$$

□

We now treat the case in which $\tau(I_\rho, E_\rho)$ is a TSS ground event. When this case occurs, we select a region, R_j, that is not in $\bar{F}_\rho \cup I_\rho$. The notation $\bar{F}$ is used to refer to the set of all regions that are contained in the segments of F. This region is used as the initial region for a new TSS. When presented with the SSS event $\sigma(F_\rho, I_\rho, E_\rho)$, the new event corresponding to the selection of R_ρ yields the same event, although with a slightly different notation. We state this equivalence as a proposition:

Proposition 5 *For an SSS ground event $\sigma(F_\rho, I_\rho, E_\rho)$ and some region $R_j \notin \bar{F}_\rho \cup I_\rho$, we have*

$$\sigma(F_\rho, I_\rho, E_\rho) = \sigma(F_\rho \cup \{I_\rho\}, \{R_j\}, \emptyset). \tag{7.13}$$

Proof In the first expression, F_ρ represents a set of segments, and I_ρ and E_ρ represent an additional segment (a TSS ground event). This additional segment can be identified by its set of regions, which is I_ρ. To obtain the second expression, we add the additional

segment, I_ρ, to F_ρ, and represent a new TSS with the include set, $\{R_j\}$, and exclude set, $\emptyset$.

This equivalence becomes clearer when the definition of σ is applied to left and right sides of (7.13) to obtain

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap f_i(\tau(I_\rho, E_\rho)) \tag{7.14}$$

and

$$\{S \in \Pi : T \in S \; \forall T \in (F_\rho \cup \{I_\rho\})\} \bigcap f_i(\tau(\{R_j\}, \emptyset)), \tag{7.15}$$

respectively. The first expression describes the set of all segmentations that contain the segments in F_ρ and contain the segment I_ρ (given by the ground event $\tau(I_\rho, E_\rho)$). The second expression describes the set of all segmentations that include F_ρ and I_ρ, and also contain some segment that contains R_j. The second condition is not restrictive since every segmentation must contain a segment that contains R_j (i.e., $f_i(\tau(\{R_j\}, \emptyset)) = \Pi$). Therefore, the two representations denote the same SSS event. □

If $\tau(I_\rho, E_\rho)$ is a TSS ground event, and some refinement region, R_ρ has been selected, then by Proposition 5 the SSS refined events are

$$A_I = \sigma(F_\rho \cup \{I_\rho\}, \{R_j, R_\rho\}, \emptyset) \tag{7.16}$$

and

$$A_E = \sigma(F_\rho \cup \{I_\rho\}, \{R_j\}, \{R_\rho\}), \tag{7.17}$$

in which R_j is the new initial region. Again, these represent a disjoint partition of A_ρ.

When $\tau(I_\rho, E_\rho)$ is a not a TSS ground event, the probabilities of the SSS refined events are computed by

$$P(\sigma(F_\rho, I_\rho \cup \{R_\rho\}, E_\rho)|e) = P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)|e) \prod_{T \in F_\rho} P(\{T\}) \tag{7.18}$$

and

$$P(\sigma(F_\rho, I_\rho, E_\rho \cup \{R_\rho\})|e) = P(\tau(I_\rho, E_\rho \cup \{R_\rho\})|e) \prod_{T \in F_\rho} P(\{T\}). \tag{7.19}$$

Each $P(\{T\})$ represents the probability of the ground event $\{T\}$ in its corresponding TSS. If $\tau(I_\rho, E_\rho)$ is a TSS ground event, then we use (3.39).

Using the concepts of SSS cover and SSS refinement, we can formulate an SSS analog of Proposition 3:

Proposition 6 *If* $P(\textsc{nth}(\mathbf{C}_g, n)) > P(\textsc{max}(\mathbf{C}_h))$, *then for any* $S \in \Pi$ *such that* $\{S\} \notin \mathbf{C}_g$, $P(\{S\}) < P(\textsc{nth}(\mathbf{C}_g, n))$.

Proof Use the same argument as the proof of Proposition 3, but with the consideration of SSS refinements and SSS covers. □

7.2.2.1 Using a beam search

The next algorithm generates a distribution of segmentations by repreatedly applying GET-TOP-SEGMENTS for different initial regions. We refer to the following algorithm as a beam-search because of its similarities to the classical AI-search strategy by that name [110]. The inputs are $\mathcal{R}$, the regions; n, the desired number of segmentations; and b, the beam size. The beam size represents the number of TSS ground events considered for any given TSS. When $b = 1$, this algorithm reduces to GREEDY-SEGMENTATION.

Its formal specification is presented in Figure 7.4. The function **SELECT-INITIAL-REGION** chooses an initial region from the region set, for a new TSS. In our experiments, this was chosen by region size.

Lines 1 to 3 perform some initializations. Line 1 chooses the initial region for the first TSS. The variable $\mathbf{C}_h$ is initialized to represent all of Π with a single event, and $\mathbf{C}_g$ is initialized to the empty set.

Each iteration in Lines 5 to 14 corresponds to one construction of a TSS representation, using GET-TOP-SEGMENTS or GET-TOP-SEGMENTS-DEP. Line 16 includes

BEAM-SEARCH-SEGMENTATIONS $(\mathcal{R},\ n,\ b)$

```
1    R_i ←SELECT-INITIAL-REGION(R)
2    C_g ← ∅;  C_h ← ∅
3    INSERT(C_h, {∅, {R_i}, ∅})
4    repeat
5        A_ρ ←EXTRACT-MAX(C_h)
6        R_i ← SELECT-INITIAL-REGION(R − F̄_ρ)
7        A_ρ ← {F_ρ ∪ {I_ρ}, {R_i}, ∅}
8        G ← GET-TOP-SEGMENTS(R − F̄_ρ, R_i, b)
9        for B_g ∈ G  do
10           if SSS-GROUND({F_ρ, I_g, E_g})
11               then
12                   INSERT(C_g, {F_ρ, I_g, E_g})
13               else
14                   INSERT(C_h, {F_ρ, I_g, E_g})
15   until [C_h = ∅] or [ |C_g| ≥ n and P(NTH(C_g, n)) > P(MAX(C_h))]
16   return C_g
```

Figure 7.4 An algorithm which performs a beam-search on the space of segmentations.

the termination condition from Proposition 6; however, in this instance the condition implies that no better n segmentations would be found by allowing the algorithm to execute until $\mathbf{C}_h$ is empty. It does not guarantee that the top n segmentations were determined because many segments are not considered due to the beam width. For instance, if $b = 2$, there could exist a favorable segmentation, S, in which some $T \in S$ is ranked third in its TSS. Although $P(\{S\})$ might be high, the segmentation would never be represented since only the top two segments from each TSS construction are used to form segmentations.

Line 5 removes the SSS event that has highest probability from the queue. Line 6 selects a valid initial region, and Line 7 converts the SSS refinement event into its equivalent representation. In Line 8, b top segments are computed, with the regions in $\bar{F}_\rho$ removed from $\mathcal{R}$.

Lines 10 to 14 create a new set of SSS refined events from the resulting best segments. If an SSS refined event is an SSS ground event, then it is added to $\mathbf{C}_g$. Otherwise, it is added to $\mathbf{C}_h$, and may be utilized at a later iteration.

7.2.2.2 Obtaining the top n segmentations

We next present an algorithm analogous to GET-TOP-SEGMENTS to find the best segmentations. Since the SSS refinement operation is more complicated than the TSS refinement, there are extra conditions for determining whether for an SSS event, $\sigma(F_\rho, I_\rho, E_\rho)$, is SSS ground, whether $\tau(I_\rho, E_\rho)$ represents a TSS ground event, or neither. The formal specification is given in Figure 7.5. Again, we use the notation $\bar{F}$ to refer to the set of all regions that are contained in the segments of F.

Since this algorithm is very similar to GET-TOP-SEGMENTS, only the major differences are indicated. Lines 1 to 4 perform the usual initialization. Lines 6 to 28 represent one SSS refinement applied to the cover. Line 29 includes the termination criterion, Proposition 6.

The primary difference in the conditions in Lines 10 to 28, when compared to the analogous portion of GET-TOP-SEGMENTS, is the treatment of the special case in which include and exclude sets represent a TSS ground event, and a new initial region must be selected. This occurs in Lines 18-19 and Lines 27-28.

7.2.2.3 Obtaining a distribution of m-segment events

In this section we consider some modifications to the previous algorithms to generate distributions of groups of some fixed number of adjacent segments. One requirement we impose is that initial regions must be selected from the set of regions adjacent to at least one of the regions in F. This guarantees obtaining a group of connected segments in each element of $\mathbf{C}_g$. Also, the user provides the first initial region, although this can be optional.

To perform a beam-search, the modifications to BEAM-SEARCH-SEGMENTATIONS are indicated in Figure 7.6. Line 1 is removed since R_i is given initially. The added condition in Line 10 determines whether an m-segment event has been determined. The

```
GET-TOP-SEGMENTATIONS (R, n)
1     R_i ←SELECT-INITIAL-REGION(R)
2     NEW-MEM-PROBS(R_i, FRONTIER({{R_i}, ∅}))
3     C_g ← ∅;  C_h ← ∅
4     INSERT(C_h, {∅, {R_i}, ∅})
5     repeat
6         A_ρ ←EXTRACT-MAX(C_h)
7         R_ρ ← SELECT-R_ρ(FRONTIER({I_ρ, E_ρ}) − F̄_ρ)
8         A_I ← {F_ρ, I_ρ ∪ {R_ρ}, E_ρ};  A_E ← {F_ρ, I_ρ, E_ρ ∪ {R_ρ}}
9         Compute and store P(A_I|e) and P(A_E|e) using (7.18) and (7.19)
10        if FRONTIER({I_I, E_I}) − F̄ ≠ ∅
11            then
12                INSERT(C_h, A_I)
13                NEW-MEM-PROBS(R_i, SELECT-R_ρ(FRONTIER({I_ρ, E_ρ}) − F̄_ρ)
14            else
15                if SSS-GROUND(A_I)
16                    INSERT(C_g, A_I)
17                else
18                    R_i ←SELECT-INITIAL-REGION(R − F̄_ρ)
19                    INSERT(C_h, {F_ρ ∪ I_I}, {R_i}, ∅})
20        if FRONTIER({I_E, E_E}) − F̄ ≠ ∅
21            then
22                INSERT(C_h, A_E)
23            else
24                if SSS-GROUND(A_E)
25                    INSERT(C_g, A_E)
26                else
27                    R_i ←SELECT-INITIAL-REGION(R − F̄_ρ)
28                    INSERT(C_h, {F_ρ ∪ {I_ρ}, {R_i}, ∅})
29    until [C_h = ∅] or [ |C_g| ≥ n and P(NTH(C_g, n)) > P(MAX(C_h))]
30    return C_g
```

Figure 7.5 An algorithm which returns the best n segmentations.

BEAM-SEARCH-M-SEGMENT-EVENTS $(\mathcal{R}, R_i,\ n,\ b,\ m)$

1 $R_i \leftarrow$ **SELECT-INITIAL-REGION**$(\mathcal{R})$

⋮

10 **if** **SSS-GROUND**$(A_\rho \cup \{T\})$ or **SEGMENTS**$(B_g) = m$

⋮

15 **return** $\mathbf{C}_g$

Figure 7.6 A beam-search algorithm which returns n m-segment groupings.

GET-TOP-M-SEGMENT-EVENTS $(\mathcal{R}, R_i,\ n,\ m)$

1 $R_i \leftarrow$ **SELECT-**$R_\rho(\mathcal{R})$

⋮

15 **if** **SSS-GROUND**(A_I) or **SEGMENTS**$(A_I) = m$

⋮

24 **if** **SSS-GROUND**(A_E) or **SEGMENTS**$(A_E) = m$

⋮

30 **return** $\mathbf{C}_g$

Figure 7.7 An algorithm which returns the n best m-segment groupings.

function **SEGMENT** returns the number of segments represented by B_g. If B_g represents m segments, then it is treated as an SSS ground event (although in truth it is not), and added to $\mathbf{C}_g$.

Similarly, modifications that can be made to GET-TOP-SEGMENTATIONS are given in Figure 7.7.

CHAPTER 8

EXPERIMENTAL RESULTS

8.1 Our Experimentation Strategy

In this chapter we present some experiments which illustrate the framework proposed in Chapters 3 to 7. Section 8.2 provides tabular comparisons of the membership probability with the planar model by varying the amount of noise, the number of points, the distance between the regions, the observation space, and the use of the delta model. Section 8.3 presents TSS representations on real and synthetic range images using GET-TOP-SEGMENTS. Section 8.4 presents some SSS representations, obtained using GET-TOP-SEGMENTATIONS and BEAM-SEARCH-SEGMENTATIONS, and also some segmentations obtained using GREEDY-SEGMENTATION.

From the results in Section 8.2 we conclude that the membership probabilities correspond to our intuitive expectations. The most general relationship to be observed from tables is: As the amount of information in the regions, e.g., R_1 and R_2, increases, the membership probability tends to one if $H(R_1 \cup R_2) = \mathit{true}$, and tends to zero if $H(R_1 \cup R_2) = \mathit{false}$. We also have concluded that both the identity-maps, (5.22), and sum-of-squares, (5.17), spaces are reasonable choices for the observation space, and the identity-maps space performs particularly well in the presence of little information. We have also noticed that the membership probability decreases as that distance between R_1

and R_2 increases, and that the delta model (discussed in Sections 4.5 and 6.5) is a reasonable alternative for adjacent regions when there is a sufficient amount of information in one of the regions.

From the results in Section 8.3 and 8.4 we conclude that the methods of this thesis are quite successful for practical-sized problems, on real and synthetic range data. On the synthetic data we created large regions, and the points were placed with varying noise levels. The larger regions make the differences between the segments (and segmentations) clearer in the figures. However, to demonstrate the methods on more challenging region sets and image data, we have presented several results in which the regions were generated automatically (through the fusion of a recursive fitting technique and Canny edge detection) on real range data. When very small regions are considered in these cases, many of the resulting segments (and segmentations) that have high probability are very similar.

The specific membership probability components used in this chapter were developed in Chapter 5. The regions and images used in the experiments are composed from range data in $\Re^3$. We consider the nonhomogeneous parameter manifolds from the planar and quadric models, discussed in Section 5.3.1. We use both the sum-of-squares and identity-maps observation spaces, as presented in Section 5.3.2. The Gaussian iid degradation model, presented in Section 5.3.3, is used to model the noise in the range images. We take the prior density on the parameter space to be uniform and consequently use the manifold area expressions from Section 6.2.

In the synthetic images, the data points are placed in the image along surfaces represented in the parameter space and points are randomly displaced by sampling from the Gaussian displacement density. We have two primary motivations for using synthetic data:

1. Synthetic data allows controlled experiments to be performed in which any number of model parameters can be changed. Regions and images are constructed which conform to the model parameters. When varying a parameter such as the noise variance the effects on membership probability, and on TSS and SSS distributions can be carefully studied. With a given real image the parameters are fixed due to the characteristics of the imaging process.

2. We can judge the performance of the membership model when statistical image models are correct. We sometimes wish to determine whether the general framework is appropriate for image segmentation, without taking into account some unknown differences between a hypothesized image model and one that is more experimentally correct for a particular range-scanner setup.

8.2 Membership Probabilities

This section demonstrates the membership probability by performing several series of computations of the membership probabilities for region pairs under the planar model. For each computation, different settings for the model variables are chosen, and the resulting probabilities appear in tables. The prior probability remains fixed at $1/2$ throughout all of the experiments. Also, all regions considered in the experiments are square, with integer distances between adjacent points.

We present several figures which illustrate the conditional density on the parameter space for several regions. These are plotted using the technique discussed in Section 5.2. Figures 8.1.(a)-(c) show the conditional observation density values for three four-point regions. The (x_1, x_2) coordinates of the points are $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$. The x_3 values are determined by sampling from the Gaussian distribution with zero mean. The points in the first two regions were generated with $\sigma^2 = 1.0$, corresponding to a high

degree of noise. The points for the region corresponding to Figure 8.1 were generated with $\sigma^2 = 0.01$, corresponding to little noise. Figure 8.1.(d) represents the conditional density values for a region with 100 points and $\sigma^2 = 1.0$.

When we refer to regions which contain little information, we are generally referring to regions of the type in Figures 8.1.(a) and (b), in which the density is not peaked. These are the cases that would give an estimation-based approach the most difficulty because it is not at all clear what value should be selected for the parameter (e.g., using MLE). Figures 8.1.(c) and (d), on the other hand, correspond to regions which contain much information, since either the variance is small, or there are many points in the region.

For two regions, R_1 and R_2, the following items were varied in the experiments:

- $R_1 \cup R_2$ is homogeneous vs. $R_1 \cup R_2$ is not homogeneous

By $R_1 \cup R_2$ is homogeneous, we mean that before the noise is appled, the points in R_1 and R_2 lie in the same plane. For the case in which $R_1 \cup R_2$ is not homogeneous, we rotate one of the regions by 30 deg before applying noise. Naturally, we conclude that in the presence of a significant amount of information, the probability tends toward one when $R_1 \cup R_2$ is homogeneous, and toward zero when $R_1 \cup R_2$ is not homoegeous.

- The number of points in R_1 and R_2

From the experiments we infer that as the number of points in R_1 and R_2 increases, the membership probability approaches zero or one. When the regions are small (e.g., less than 10 points), the membership probabilities tend to be closer to the prior membership probability. This coincides with our intuition that smaller regions contain less information and hence more ambiguity.

- The noise variance, σ^2, in R_1 and R_2

We also infer that the increase in variance causes the same type of convergence as the increase in region sizes. As the noise level increases the membership probability tends to

be closer to the prior membership probability. This is also expected intuitively since the larger amount of noise will cause more ambiguity.

- The distance between R_1 and R_2

A greater distance between the regions tends to cause a smaller membership probability, particularly when little information is present in the regions. This is expected, and we provide an intuitive explanation of the results. First, note that the distance between the regions does not affect the numerator integrals of (5.33), since each regions depends on the regions individually. Of the space of possible planes (which actually is the domain of integration), consider how many can closely approximate two small, adjacent square regions. As these regions are moved farther apart, there are fewer planes that can closely approximate both regions. This situation occurs in the denominator integral of (5.33), causing the membership probability to decrease as the regions are distanced.

We have observed that this difference in probabilities is much greater for quadrics. Recall that with the IE-independent model, only R_i and R_ρ are considered, and are not required to be adjacent. With the IE-dependent model, all of the regions in I_ρ are considered, and R_ρ is adjacent to at least one of them. For this reason, we have used the IE-dependent model for the quadric experiments presented in this chapter.

- The identity-maps observation space vs. the sum-of-squares observation space

We have observed that both of these models produce reasonable results for the membership probability. For smaller regions that have high variance, it seems that the identity-maps model produces results that most closely correspond to our intuitive expectations. This is due to the fact that the identity-maps model contains more information regarding the parameter space.

- Parameter manifold integrations vs. the delta function approximation

As discussed in Sections 4.5 and 6.5, the delta function approximation is appropriate when there is sufficient information regarding the parameter space in one of the regions. We have determined that indeed the delta function approximation succeeds when 1) the variance is low, 2) the regions contain few points, and 3) the regions are adjacent. The distance between regions is important since slight errors in the parameter estimate cause the surface to be displaced. When two regions are far apart, and a slightly inaccurate estimate is obtained for R_1, is it likely that the resulting surface has moved considerably away from R_2

We present six tables which illustrate the observations that we have made. The first three tables represent cases in which $R_1 \cup R_2$ is homogeneous, and the remaining three represent cases in which $R_1 \cup R_2$ is not homogeneous. The numbers in the Dist column correspond to the distances between the regions, as multiples of the width of R_2. Each table entry is obtained by averaging several experiments, to make the trends clearer. Tables 8.1 and 8.4 provide membership probabilities when both regions contain only four points. Tables 8.2 and 8.5 provide membership probabilities for medium-sized regions (49 points). Finally, Tables 8.3 and 8.6 provide membership probabilities for large regions (100 to 400 points).

8.3 TSS Representations

In this section we present experimental results which show TSS representations on four real range images and one synthetic image. Each TSS representation was obtained by running GET-TOP-SEGMENTS with $n = 20$, and the termination criterion given by Proposition 3 was met for every TSS experiment. We used a prior membership probability of 0.5 for the planar model, and 0.99 for the quadric model. A higher prior membership probability is required for the quadric case, otherwise the posterior membership probabilities are relatively low. This is due to the fact that even when the union

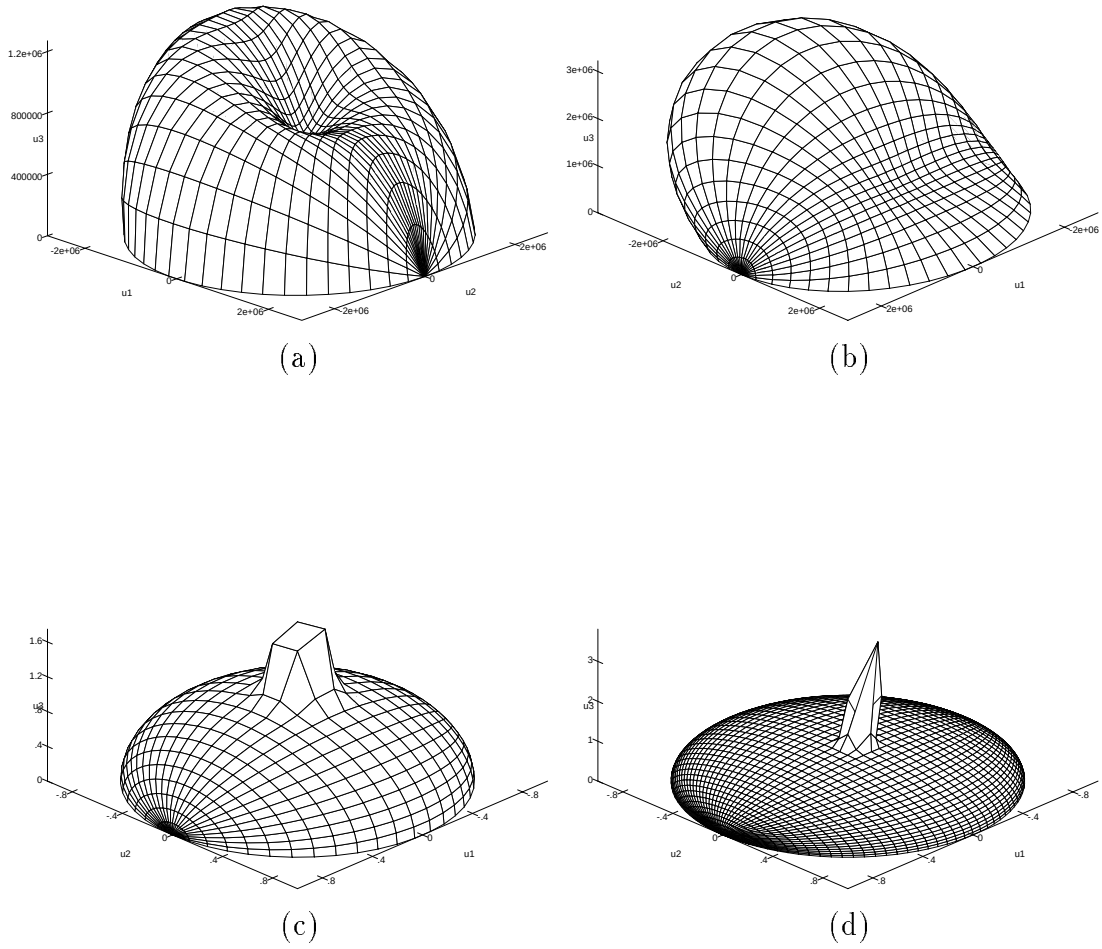


Figure 8.1 (a) The conditional observation space density for a four-point region with $\sigma^2 = 1.0$; (b) the density for another four-point region with $\sigma^2 = 1.0$; (c) the density for a four-point region with $\sigma^2 = 0.01$; (d) the density for a 100-point region with $\sigma^2 = 1.0$.

Table 8.1 Probabilities when R_1 and R_2 are small, and $R_1 \cup R_2$ is homogeneous.

Points		Noise	Dist	Integration		Delta Model	
$\|R_1\|$	$\|R_2\|$	σ^2	Δ	Identity	Chi-square	Identity	Chi-square
4	4	2.0	0	0.7794	0.9401	0.5459	0.9743
4	4	2.0	1	0.6301	0.9130	0.1958	0.5662
4	4	2.0	5	0.3620	0.7706	0.0728	0.1272
4	4	1.0	0	0.7138	0.9055	0.3032	0.8557
4	4	1.0	1	0.5523	0.8232	0.1355	0.3279
4	4	1.0	5	0.3061	0.6060	0.0851	0.1179
4	4	0.1	0	0.7257	0.8467	0.3801	0.5901
4	4	0.1	1	0.6221	0.7520	0.0988	0.3590
4	4	0.1	5	0.4238	0.5186	0.0079	0.0391
4	4	0.01	0	0.9730	0.9890	0.5220	0.8175
4	4	0.01	1	0.9337	0.9790	0.2415	0.5018
4	4	0.01	5	0.8551	0.9418	0.0972	0.1201
4	4	0.001	0	0.9704	0.9982	0.6178	0.6821
4	4	0.001	1	0.9582	0.9966	0.3769	0.4825
4	4	0.001	5	0.9397	0.9890	0.0761	0.0873
4	4	0.0001	0	0.9769	0.9997	0.5385	0.7421
4	4	0.0001	1	0.9755	0.9995	0.4029	0.4888
4	4	0.0001	5	0.9517	0.9987	0.1541	0.2793

Table 8.2 Probabilities when R_1 is medium-sized, and $R_1 \cup R_2$ is homogeneous.

Points		Noise	Dist	Integration		Delta Model	
$\|R_1\|$	$\|R_2\|$	σ^2	Δ	Identity	Chi-square	Identity	Chi-square
49	4	2.0	0	0.4820	0.9032	0.0326	0.6900
49	4	2.0	5	0.3826	0.7992	0.0236	0.2128
49	4	1.0	0	0.8018	0.9755	0.3801	0.9505
49	4	1.0	5	0.7156	0.9453	0.1490	0.4979
49	4	0.1	0	0.9326	0.9982	0.6765	0.9199
49	4	0.1	5	0.9343	0.9961	0.2784	0.7583
49	4	0.01	0	0.9969	0.9998	0.6233	0.9674
49	4	0.01	5	0.9958	0.9996	0.2438	0.6895
49	4	0.001	0	0.9993	1.0000	0.7137	0.9959
49	4	0.001	5	0.9993	1.0000	0.2524	0.7605
49	9	2.0	0	0.6268	0.9023	0.3290	0.5578
49	9	2.0	5	0.4811	0.7717	0.0975	0.2229
49	9	1.0	0	0.6808	0.9714	0.1786	0.8316
49	9	1.0	5	0.5048	0.9280	0.0408	0.3008
49	9	0.1	0	0.9687	0.9984	0.6354	0.8981
49	9	0.1	5	0.9295	0.9956	0.2853	0.6198
49	9	0.01	0	0.9993	0.9999	0.8678	1.0000
49	9	0.01	5	0.9979	0.9996	0.3714	0.7957
49	9	0.001	0	0.9997	1.0000	0.6367	0.9225
49	9	0.001	5	0.9973	1.0000	0.3074	0.5827
49	16	2.0	0	0.6993	0.9193	0.2991	0.6976
49	16	2.0	5	0.4762	0.7684	0.0824	0.2346
49	16	1.0	0	0.7641	0.9705	0.4397	0.7873
49	16	1.0	5	0.6309	0.9016	0.0756	0.4024
49	16	0.1	0	0.9205	0.9983	0.6797	0.9136
49	16	0.1	5	0.8959	0.9944	0.1674	0.6504
49	16	0.01	0	0.9968	0.9998	0.7635	0.9199
49	16	0.01	5	0.9863	0.9994	0.2910	0.5527
49	16	0.001	0	0.9600	1.0000	0.6323	0.9200
49	16	0.001	5	0.9586	0.9999	0.2510	0.4865
49	49	2.0	0	0.6466	0.9171	0.2366	0.7479
49	49	2.0	5	0.2993	0.6586	0.0708	0.2221
49	49	1.0	0	0.7199	0.9713	0.4284	0.7344
49	49	1.0	5	0.5681	0.8798	0.0800	0.0890
49	49	0.1	0	0.9473	0.9985	0.4419	0.9031
49	49	0.1	5	0.8419	0.9917	0.0412	0.2325
49	49	0.01	0	0.9967	0.9998	0.4430	0.9534
49	49	0.01	5	0.9847	0.9991	0.0440	0.0801
49	49	0.001	0	0.9998	1.0000	0.7331	1.0000
49	49	0.001	5	0.9940	0.9999	0.1430	0.3212

Table 8.3 Probabilities when R_1 is large, and $R_1 \cup R_2$ is homogeneous.

Points		Noise	Dist	Integration		Delta Model	
$\|R_1\|$	$\|R_2\|$	σ^2	Δ	Identity	Chi-square	Identity	Chi-square
100	4	2.0	0	0.7953	0.9854	0.3281	0.9140
100	4	2.0	5	0.7566	0.9718	0.1695	0.5701
100	4	1.0	0	0.9089	0.9929	0.3946	0.9410
100	4	1.0	5	0.8674	0.9870	0.2979	0.5628
100	4	0.1	0	0.9977	0.9994	0.7410	1.0000
100	4	0.1	5	0.9900	0.9989	0.4115	0.9135
100	4	0.01	0	0.9977	0.9999	0.6878	0.9999
100	4	0.01	5	0.9960	0.9999	0.4531	0.8637
100	4	0.001	0	0.9997	1.0000	0.8769	0.9600
100	4	0.001	5	0.9992	1.0000	0.5184	0.8807
100	16	2.0	0	0.7443	0.9844	0.2337	0.8650
100	16	2.0	5	0.6125	0.9552	0.0403	0.3798
100	16	1.0	0	0.9399	0.9943	0.6439	0.9601
100	16	1.0	5	0.9104	0.9839	0.1786	0.4550
100	16	0.1	0	0.9831	0.9994	0.7443	0.9600
100	16	0.1	5	0.9508	0.9984	0.2766	0.7859
100	16	0.01	0	0.9997	0.9999	0.6787	1.0000
100	16	0.01	5	0.9787	0.9999	0.2556	0.6651
100	16	0.001	0	0.9999	1.0000	0.7799	0.9990
100	16	0.001	5	0.9996	1.0000	0.3094	0.7409
100	100	2.0	0	0.7258	0.9853	0.3896	0.9002
100	100	2.0	5	0.5516	0.9178	0.0000	0.1360
100	100	1.0	0	0.9720	0.9940	0.4428	0.9157
100	100	1.0	5	0.6553	0.9603	0.0400	0.1585
100	100	0.1	0	0.9877	0.9996	0.6116	0.9897
100	100	0.1	5	0.8053	0.9976	0.0485	0.2847
100	100	0.01	0	0.9979	1.0000	0.6596	1.0000
100	100	0.01	5	0.8863	0.9997	0.2801	0.4273
100	100	0.001	0	0.9949	1.0000	0.6763	1.0000
100	100	0.001	5	0.9582	1.0000	0.1215	0.4010
400	4	2.0	0	0.9874	0.9988	0.6415	1.0000
400	4	2.0	5	0.9120	0.9983	0.3779	0.9999
400	16	2.0	0	0.9107	0.9987	0.6249	0.9999
400	16	2.0	5	0.8274	0.9975	0.3833	0.9060
400	100	2.0	0	0.9051	0.9988	0.3635	0.9795
400	100	2.0	5	0.5940	0.9953	0.1025	0.4855
400	400	2.0	0	0.9350	0.9987	0.7346	0.9999
400	400	2.0	5	0.6244	0.9931	0.0314	0.4395

Table 8.4 Probabilities when R_1 and R_2 are small and $R_1 \cup R_2$ is not homogeneous.

Points		Noise	Dist	Integration		Delta Model	
$\|R_1\|$	$\|R_2\|$	σ^2	Δ	Identity	Chi-square	Identity	Chi-square
4	4	2.0	0	0.7887	0.9373	0.7261	0.9761
4	4	2.0	1	0.6507	0.9100	0.4299	0.7442
4	4	2.0	5	0.3834	0.7664	0.2166	0.3659
4	4	1.0	0	0.7219	0.9127	0.6113	0.9161
4	4	1.0	1	0.5620	0.8418	0.2466	0.5829
4	4	1.0	5	0.2985	0.6412	0.0941	0.1544
4	4	0.1	0	0.5160	0.7842	0.2508	0.5196
4	4	0.1	1	0.4042	0.6448	0.1916	0.3528
4	4	0.1	5	0.2585	0.3870	0.1126	0.1910
4	4	0.01	0	0.0689	0.3840	0.0521	0.2302
4	4	0.01	1	0.0173	0.1142	0.0307	0.1486
4	4	0.01	5	0.0000	0.0135	0.0000	0.0021
4	4	0.001	0	0.0000	0.0000	0.0000	0.0000
4	4	0.001	1	0.0000	0.0000	0.0000	0.0000
4	4	0.001	5	0.0000	0.0000	0.0000	0.0000
4	4	0.0001	0	0.0000	0.0000	0.0000	0.0000
4	4	0.0001	1	0.0000	0.0000	0.0000	0.0000
4	4	0.0001	5	0.0000	0.0000	0.0000	0.0000

Table 8.5 Probabilities when R_1 is medium-sized, and $R_1 \cup R_2$ is not homogeneous.

Points		Noise	Dist	Integration		Delta Model	
$\|R_1\|$	$\|R_2\|$	σ^2	Δ	Identity	Chi-square	Identity	Chi-square
49	4	2.0	0	0.3592	0.8601	0.2505	0.4128
49	4	2.0	5	0.1021	0.7306	0.0023	0.1563
49	4	1.0	0	0.0191	0.7993	0.0014	0.1108
49	4	1.0	5	0.0000	0.4352	0.0000	0.2220
49	4	0.1	0	0.0000	0.0000	0.0000	0.0000
49	4	0.1	5	0.0000	0.0000	0.0000	0.0000
49	4	0.01	0	0.0000	0.0000	0.0000	0.0000
49	4	0.01	5	0.0000	0.0000	0.0000	0.0000
49	4	0.001	0	0.0000	0.0000	0.0000	0.0000
49	4	0.001	5	0.0000	0.0000	0.0000	0.0000
49	9	2.0	0	0.1713	0.8744	0.1011	0.4883
49	9	2.0	5	0.0890	0.5832	0.0380	0.2277
49	9	1.0	0	0.0015	0.8020	0.0000	0.1452
49	9	1.0	5	0.0001	0.3045	0.0000	0.1496
49	9	0.1	0	0.0000	0.0000	0.0000	0.0000
49	9	0.1	5	0.0000	0.0000	0.0000	0.0000
49	9	0.01	0	0.0000	0.0000	0.0000	0.0000
49	9	0.01	5	0.0000	0.0000	0.0000	0.0000
49	9	0.001	0	0.0000	0.0000	0.0000	0.0000
49	9	0.001	5	0.0000	0.0000	0.0000	0.0000
49	16	2.0	0	0.1708	0.8594	0.0042	0.5231
49	16	2.0	5	0.0641	0.5112	0.0347	0.3243
49	16	1.0	0	0.0215	0.8352	0.0000	0.0985
49	16	1.0	5	0.0001	0.3337	0.0000	0.1949
49	16	0.1	0	0.0000	0.0000	0.0000	0.0000
49	16	0.1	5	0.0000	0.0000	0.0000	0.0000
49	16	0.01	0	0.0000	0.0000	0.0000	0.0000
49	16	0.01	5	0.0000	0.0000	0.0000	0.0000
49	16	0.001	0	0.0000	0.0000	0.0000	0.0000
49	16	0.001	5	0.0000	0.0000	0.0000	0.0000
49	49	2.0	0	0.0788	0.8130	0.0334	0.3157
49	49	2.0	5	0.0007	0.3989	0.0001	0.0397
49	49	1.0	0	0.0000	0.8218	0.0000	0.0402
49	49	1.0	5	0.0000	0.1328	0.0000	0.0409
49	49	0.1	0	0.0000	0.0000	0.0000	0.0000
49	49	0.1	5	0.0000	0.0000	0.0000	0.0000
49	49	0.01	0	0.0000	0.0000	0.0000	0.0000
49	49	0.01	5	0.0000	0.0000	0.0000	0.0000
49	49	0.001	0	0.0000	0.0000	0.0000	0.0000
49	49	0.001	5	0.0000	0.0000	0.0000	0.0000

Table 8.6 Probabilities when R_1 is large, and $R_1 \cup R_2$ is not homogeneous.

Points		Noise	Dist	Integration		Delta Model	
$\|R_1\|$	$\|R_2\|$	σ^2	Δ	Identity	Chi-square	Identity	Chi-square
100	4	2.0	0	0.0000	0.6754	0.0000	0.0000
100	4	2.0	5	0.0000	0.2514	0.0000	0.0353
100	4	1.0	0	0.0000	0.0483	0.0000	0.0000
100	4	1.0	5	0.0000	0.0003	0.0000	0.0000
100	4	0.1	0	0.0000	0.0000	0.0000	0.0000
100	4	0.1	5	0.0000	0.0000	0.0000	0.0000
100	4	0.01	0	0.0000	0.0000	0.0000	0.0000
100	4	0.01	5	0.0000	0.0000	0.0000	0.0000
100	4	0.001	0	0.0000	0.0000	0.0000	0.0000
100	4	0.001	5	0.0000	0.0000	0.0000	0.0000
100	16	2.0	0	0.0000	0.6534	0.0000	0.0062
100	16	2.0	5	0.0000	0.0975	0.0000	0.0790
100	16	1.0	0	0.0000	0.0999	0.0000	0.0000
100	16	1.0	5	0.0000	0.0000	0.0000	0.0000
100	16	0.1	0	0.0000	0.0000	0.0000	0.0000
100	16	0.1	5	0.0000	0.0000	0.0000	0.0000
100	16	0.01	0	0.0000	0.0000	0.0000	0.0000
100	16	0.01	5	0.0000	0.0000	0.0000	0.0000
100	16	0.001	0	0.0000	0.0000	0.0000	0.0000
100	16	0.001	5	0.0000	0.0000	0.0000	0.0000
100	100	2.0	0	0.0000	0.6867	0.0000	0.0022
100	100	2.0	5	0.0000	0.0689	0.0000	0.0000
100	100	1.0	0	0.0000	0.0533	0.0000	0.0000
100	100	1.0	5	0.0000	0.0000	0.0000	0.0000
100	100	0.1	0	0.0000	0.0000	0.0000	0.0000
100	100	0.1	5	0.0000	0.0000	0.0000	0.0000
100	100	0.01	0	0.0000	0.0000	0.0000	0.0000
100	100	0.01	5	0.0000	0.0000	0.0000	0.0000
100	100	0.001	0	0.0000	0.0000	0.0000	0.0000
100	100	0.001	5	0.0000	0.0000	0.0000	0.0000
400	4	2.0	0	0.0000	0.0000	0.0000	0.0000
400	4	2.0	5	0.0000	0.0000	0.0000	0.0000
400	100	2.0	0	0.0000	0.0000	0.0000	0.0000
400	100	2.0	5	0.0000	0.0000	0.0000	0.0000
400	400	2.0	0	0.0000	0.0000	0.0000	0.0000
400	400	2.0	5	0.0000	0.0000	0.0000	0.0000

of two regions is homogeneous, there are many quadric surfaces which well-approximate one region and not the other, causing $\lambda_1(\mathbf{y}_\rho, \mathbf{y}_\mathbf{i})$ to be low.

The results obtained on the real range data were completely automated. Without regard to a degradation model that may be more experimentally correct for the given range-scanning system, we directly used the degradation model presented in Section 5.3.3. The initial regions were obtained by a fusion of label maps generated by three processes:

1. Regions obtained from a recursive splitting procedure, which attempts to closely fit planes to region data

2. A set of edges obtained from running the Canny edge detector [12] on the corresponding intensity image of the range data

3. The boundaries between valid and invalid data points.

The three of these, together, produce a set of small regions in which nearly all regions are homogeneous. There are inevitably, however, several regions which are not homogeneous. From the experiments, we have observed that these regions tend to be left alone (i.e., they are not included in a TSS representation, and treated as one-region segments in an SSS representation).

The synthetic image uses the planar model, and noise is introduced by sampling from the displacement density, (5.24). Hence, the degradation model is statistically valid for this image. In addition, for the synthetic image, the set of regions was constructed by hand, with each region being truly homogeneous. This synthetic example provides an opportunity to observe the behavior of the representations under correct model assumptions, and also to vary the some of the model parameters for comparison.

The synthetic image consists of 10000 data points (100 x 100). Figure 8.2.(a) shows the data set before noise is applied. When the points are projected into the x_1-x_2 plane, there is integer spacing between adjacent points. There is one four-sided pyramid in the

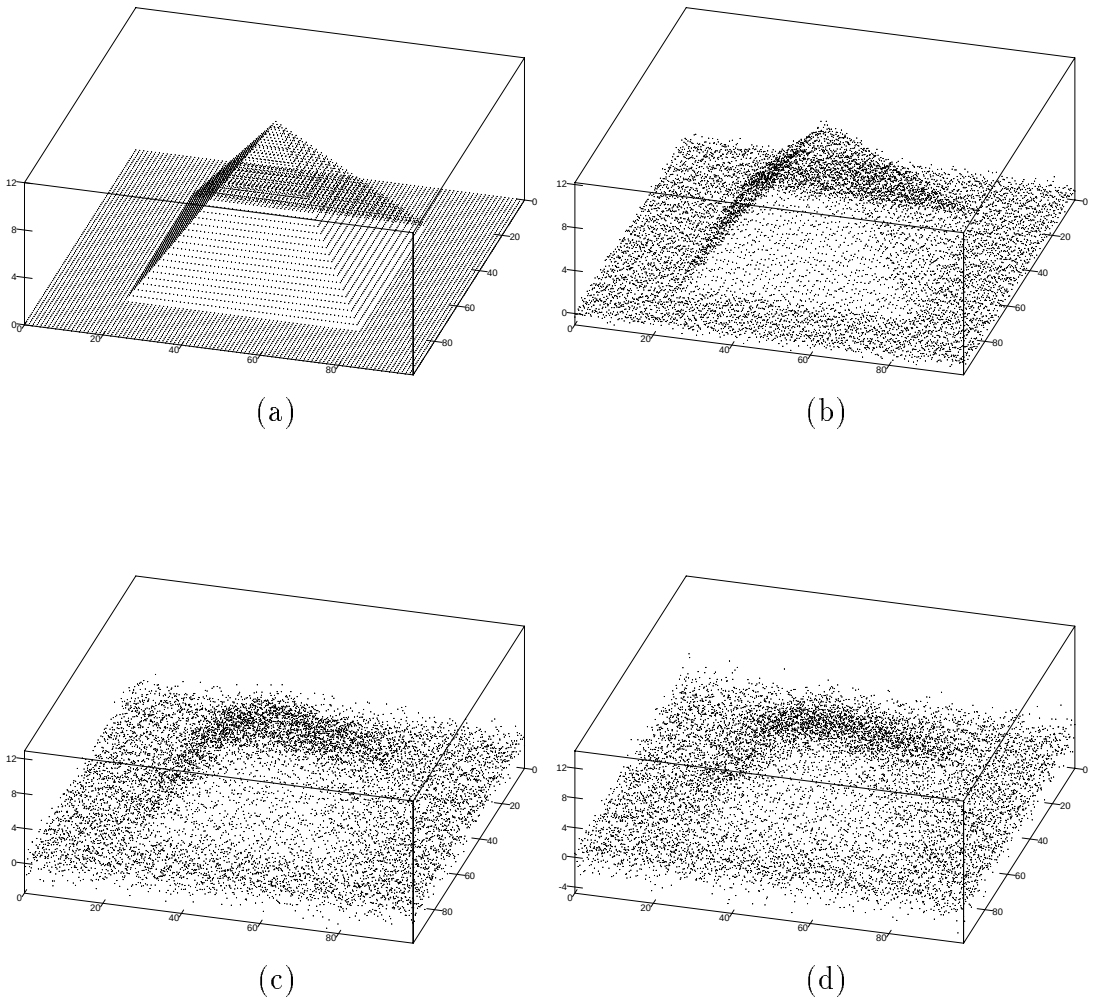


Figure 8.2 (a) The range data without noise; (b) the data with $\sigma^2 = 0.1$; (c) the data with $\sigma^2 = 1.0$; (d) the data with $\sigma^2 = 2.0$

image, with a plane in the background. Note that the height of the pyramid is distorted in the figure: the x_1 and x_2 coordinates range from 0 to 100, while the height of the pyramid, given by maximum value of x_3, is only 12. This makes the problem more challenging than the figure may suggest.

Three different noise levels were used for the experiments presented for the synthetic image. Figure 8.2.(b) shows the range image after applying noise with $\sigma^2 = 0.1$. Figure 8.2.(c) shows the range image after applying noise with $\sigma^2 = 1.0$. Figure 8.2.(d) shows the range image after applying noise with $\sigma^2 = 2.0$.

Figures 8.4-8.9 show results with $\sigma^2 = 1.0$ on the set of regions shown in Figure 8.3. These experiments use the sum-of-squares observation space (from Section 5.3.2), and the IE-independent model (from Section 3.4.2). These results correspond to a fairly high degree of noise, hence the segments that we would judge to be correct do not necessarily obtain the highest probability. In Figure 8.4, a TSS representation of Θ_{14} is obtained, and the correct segment is ranked fourth. We next show that the particular choice of initial region is not important (assuming the alternative regions roughly contain the same amount of information). Using an alternative initial region (inside the same segment), we obtain a TSS representation of Θ_{13}, shown in Figure 8.5. The use of R_{13} instead of R_{14} produces a similar result. Representations of Θ_{20}, Θ_{36}, Θ_{42}, and Θ_{73} are shown in Figures 8.6, 8.7, 8.8, and 8.9, respectively. Due to the noise level and use of the IE-independent model, the correct segment in Θ_{73} is not represented in the top twenty ground events. Note that the probabilities in this case are close to zero, and the probability distribution over the alternatives is fairly uniform.

Figures 8.10-8.15 show results obtained under the same conditions as the previous six experiments, but with with a lower noise level: $\sigma^2 = 0.1$. These results show the correct segment obtained with probability above 0.99 for every case.

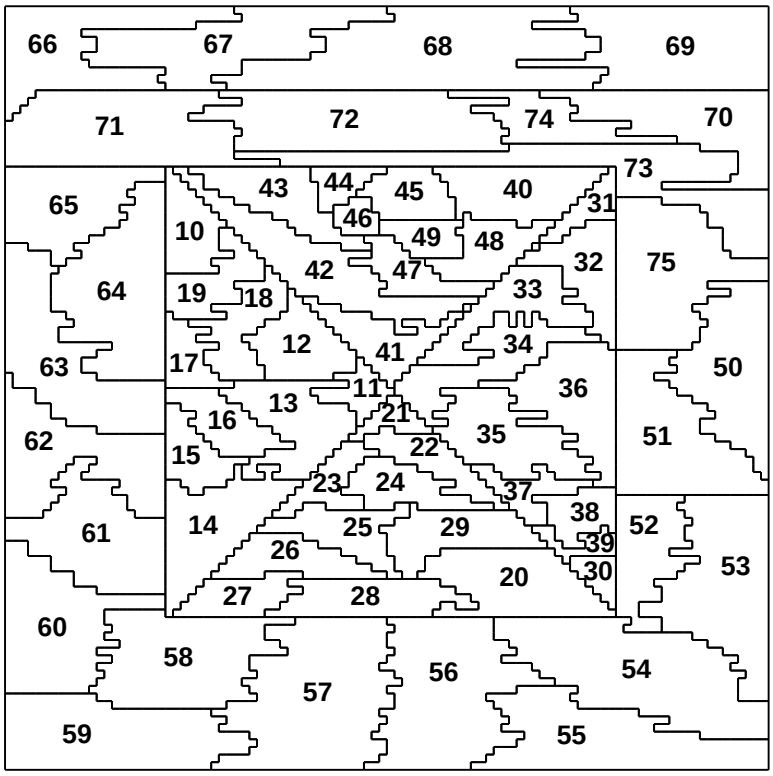


Figure 8.3 The set of regions, $\mathcal{R}$, that is presented to the TSS and SSS algorithms. This image shows the region boundaries projected into the x_1-x_2 plane, and the regions are labeled with integers for reference.

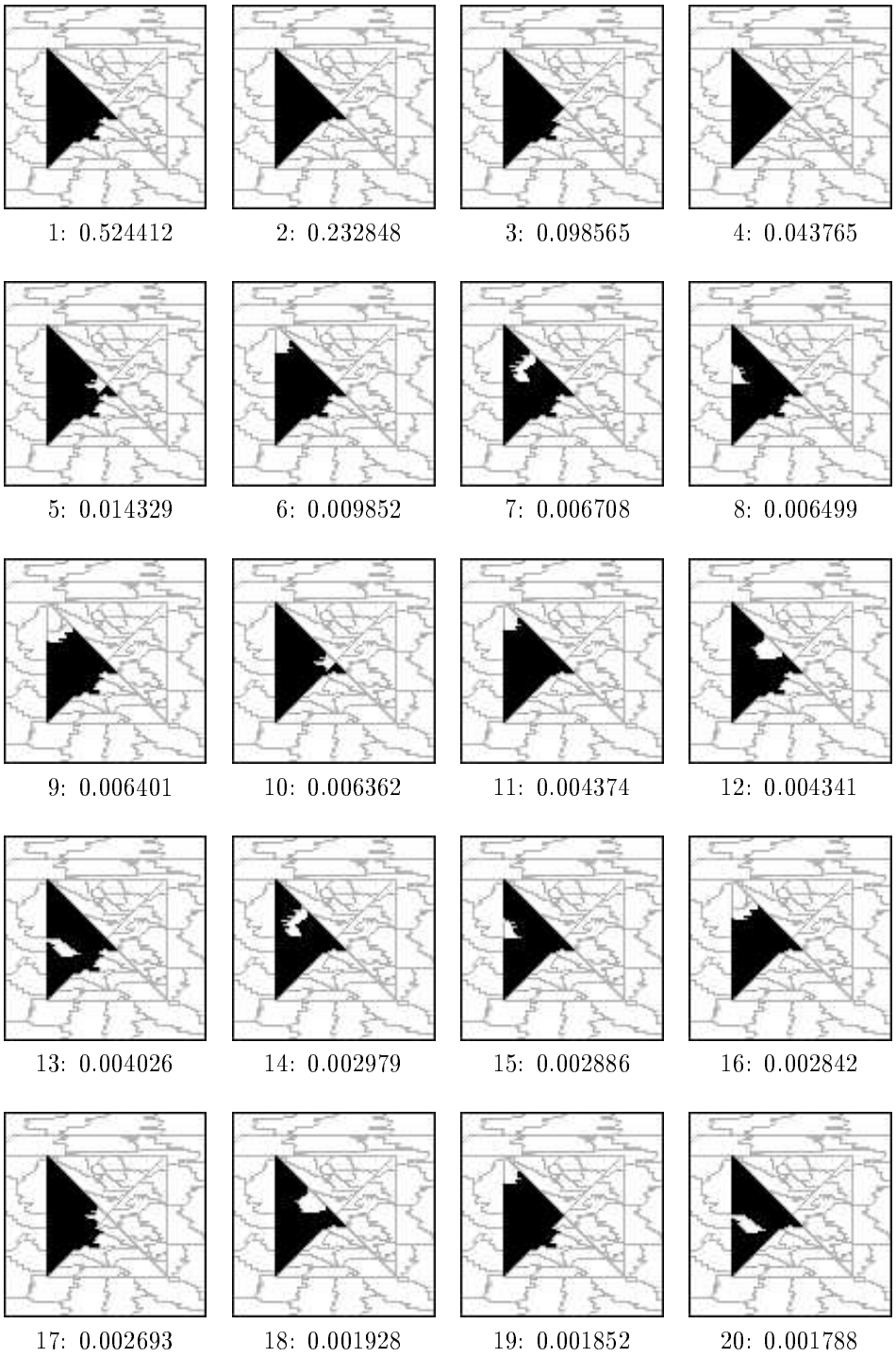


Figure 8.4 Twenty segments that have highest probability in Θ_{14}. There were 228 events in the final cover, with 25 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

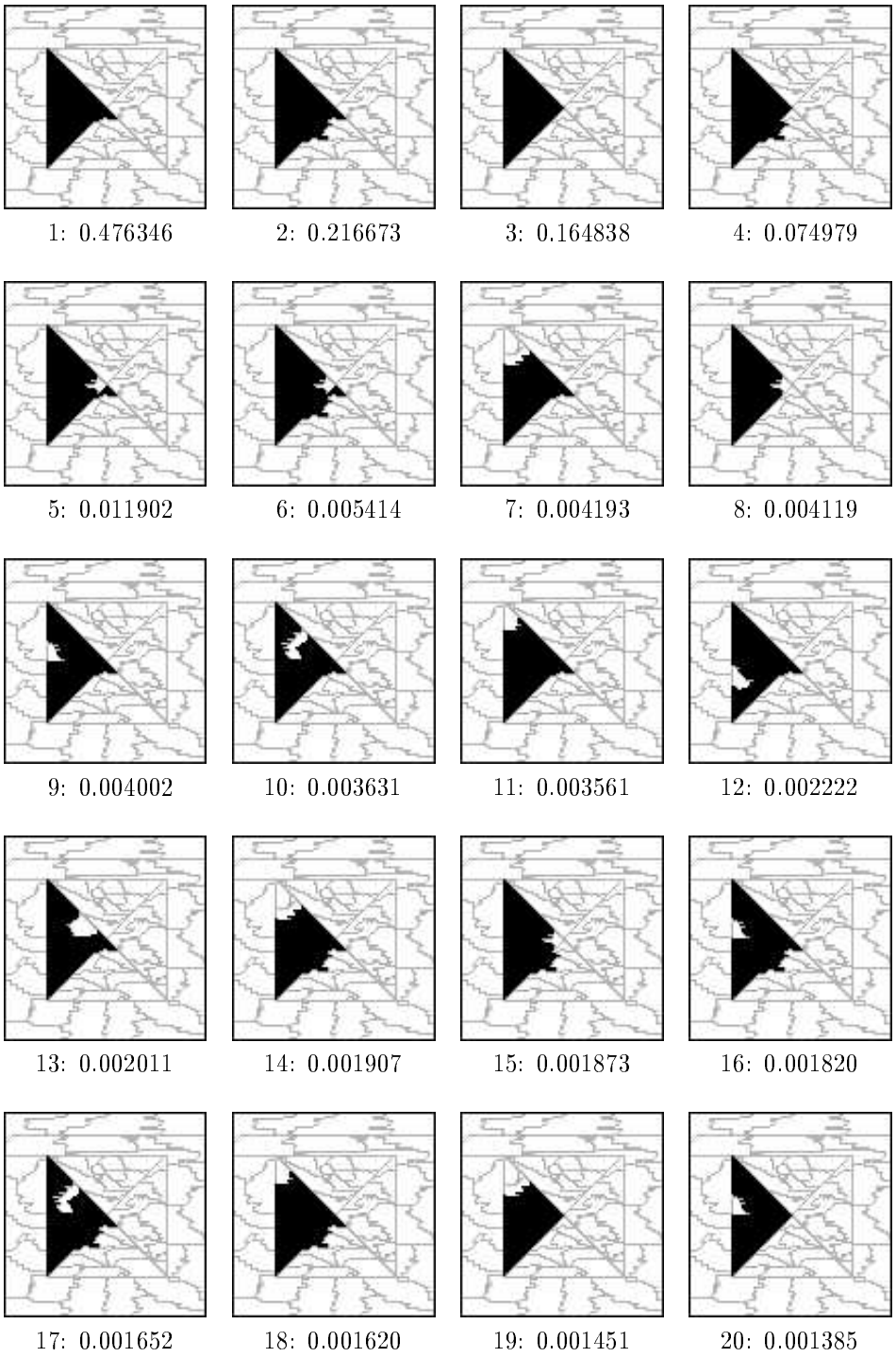


Figure 8.5 Twenty segments that have highest probability in Θ_{13}. There were 222 events in the final cover, with 29 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

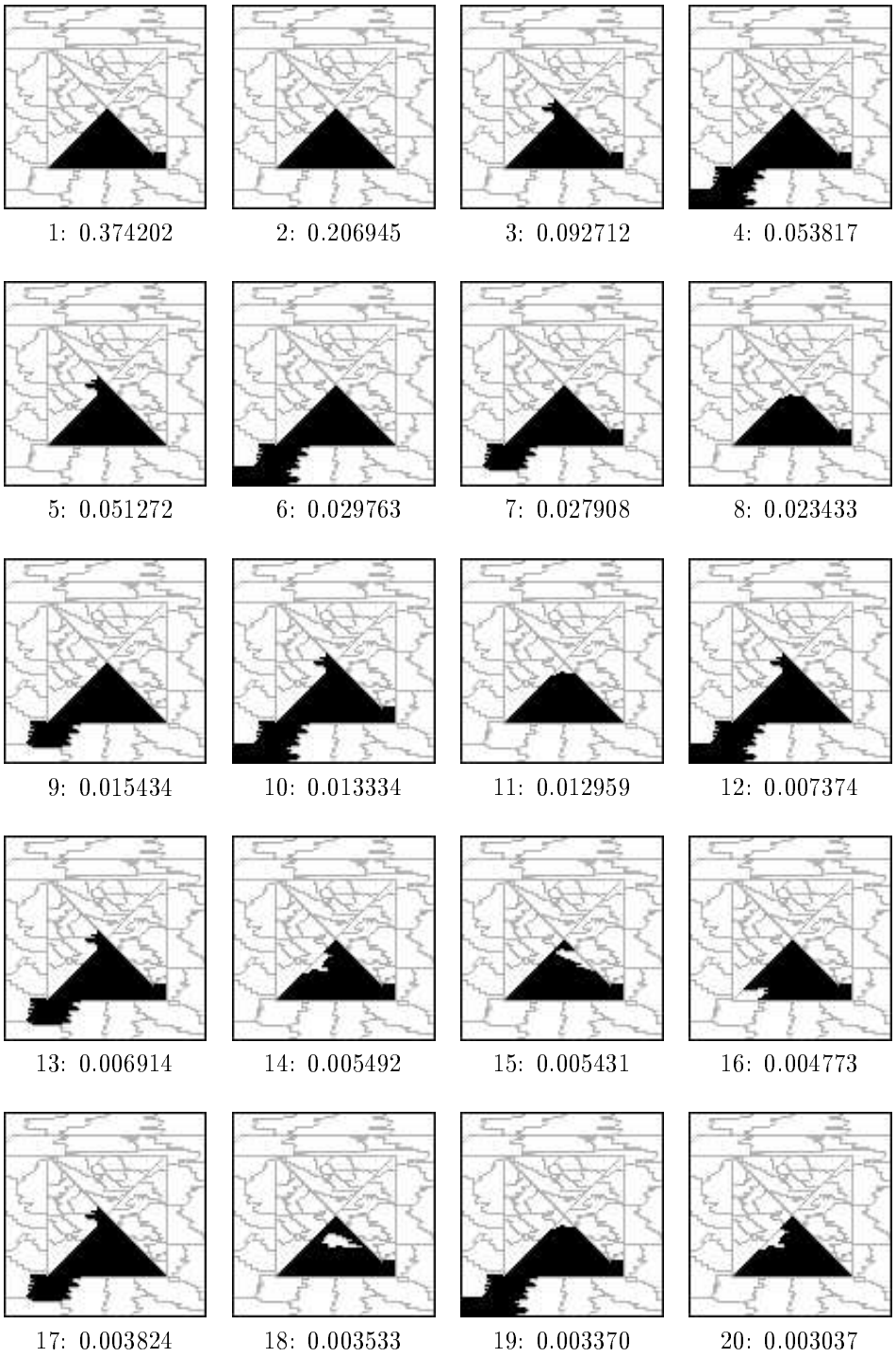


Figure 8.6 Twenty segments that have highest probability in Θ_{20}. There were 173 events in the final cover, with 27 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

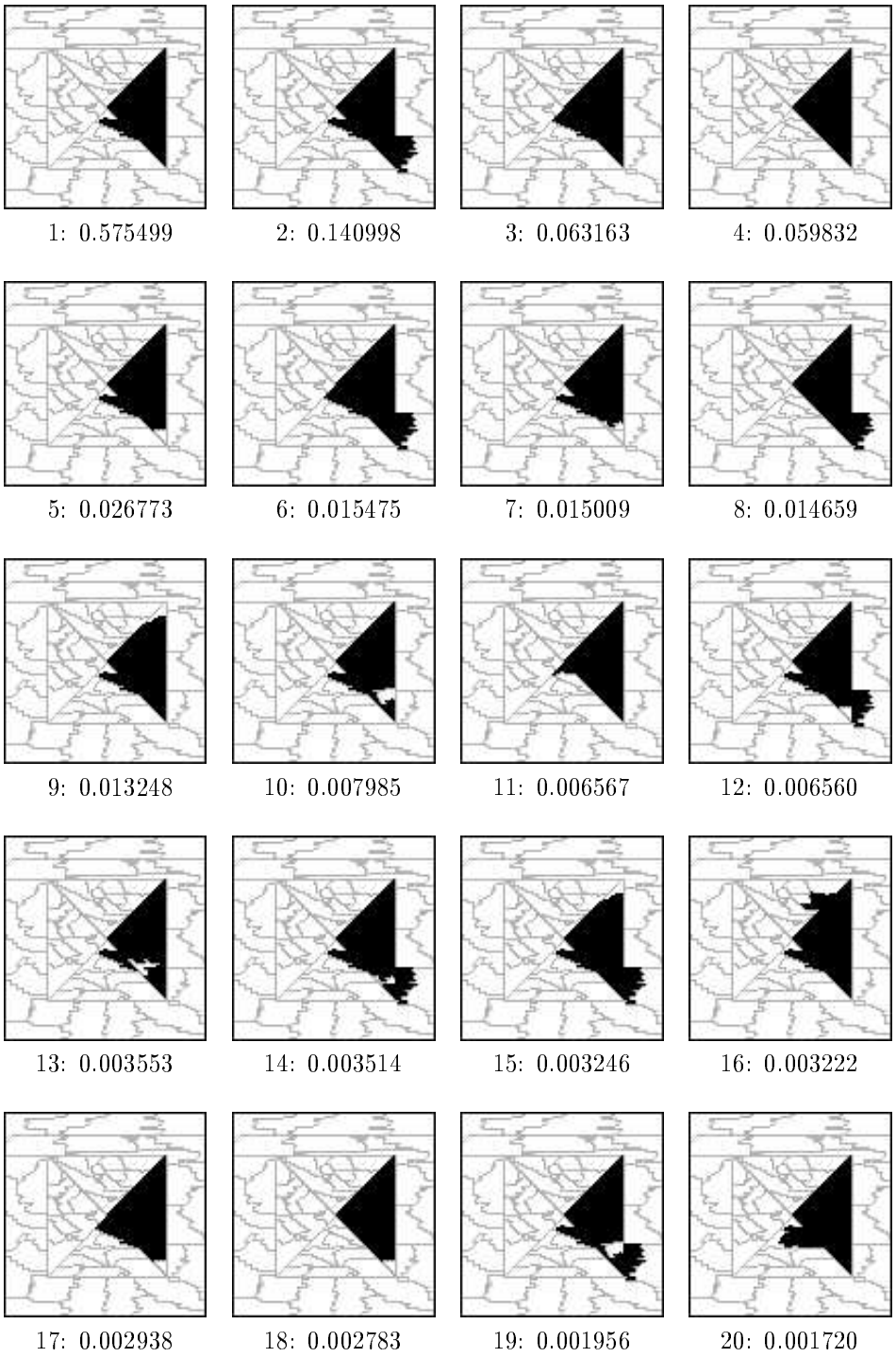


Figure 8.7 Twenty segments that have highest probability in Θ_{36}. There were 312 events in the final cover, with 20 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

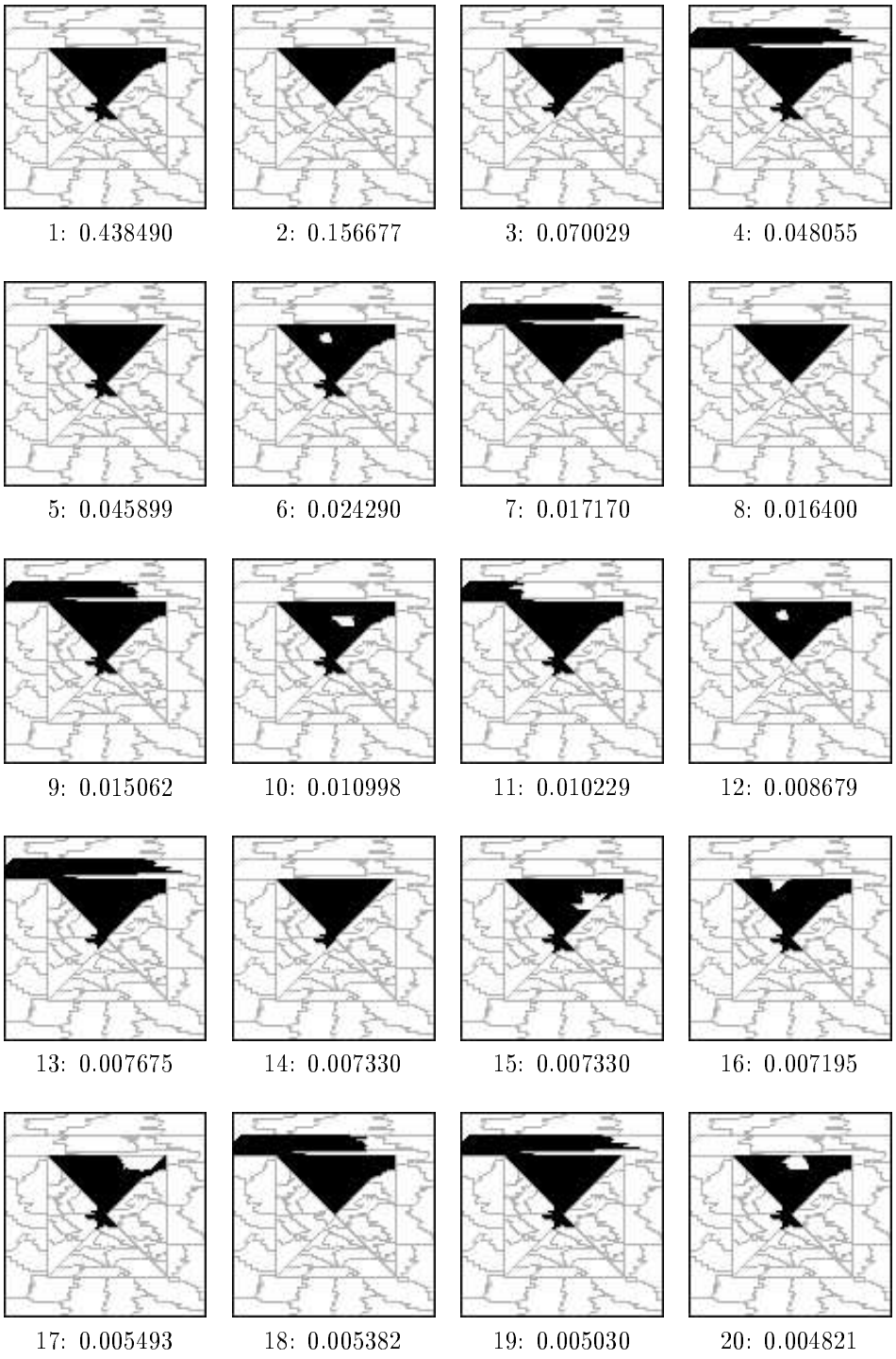


Figure 8.8 Twenty segments that have highest probability in Θ_{42}. There were 216 events in the final cover, with 36 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

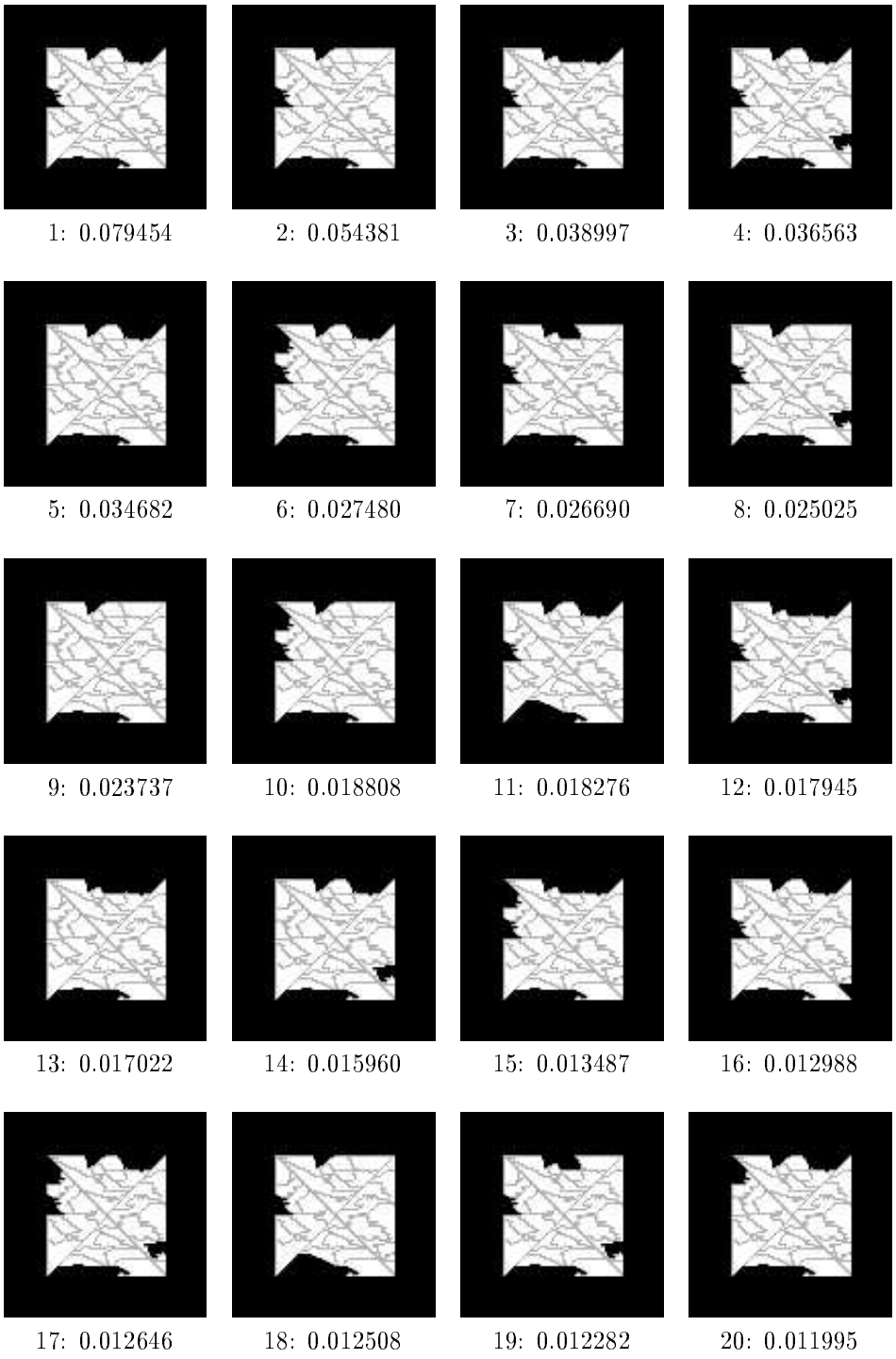


Figure 8.9 Twenty segments that have highest probability in Θ_{73}. There were 362 events in the final cover, with 22 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

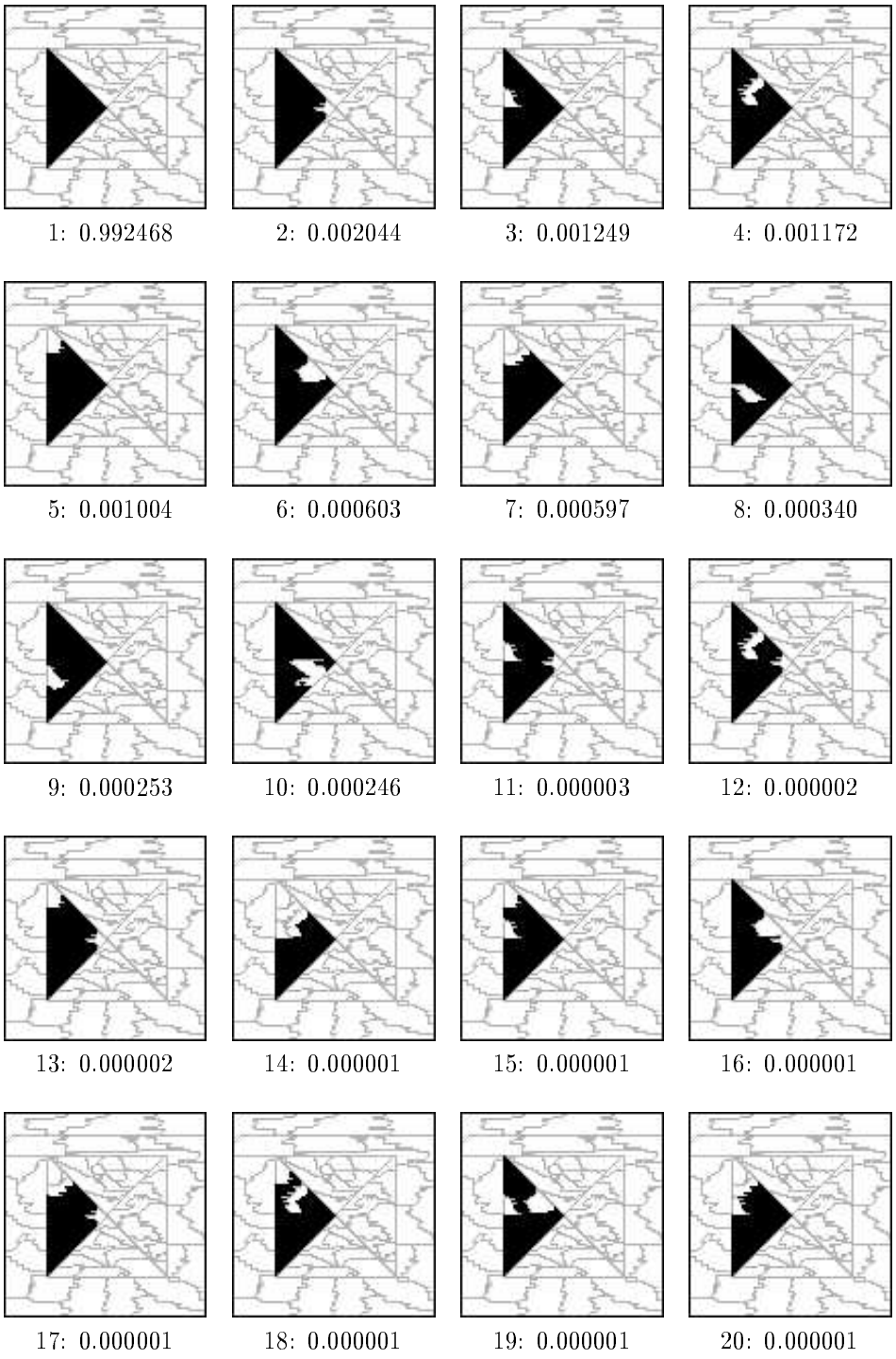


Figure 8.10 Twenty segments that have highest probability in Θ_{14}. There were 169 events in the final cover, with 27 ground events. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

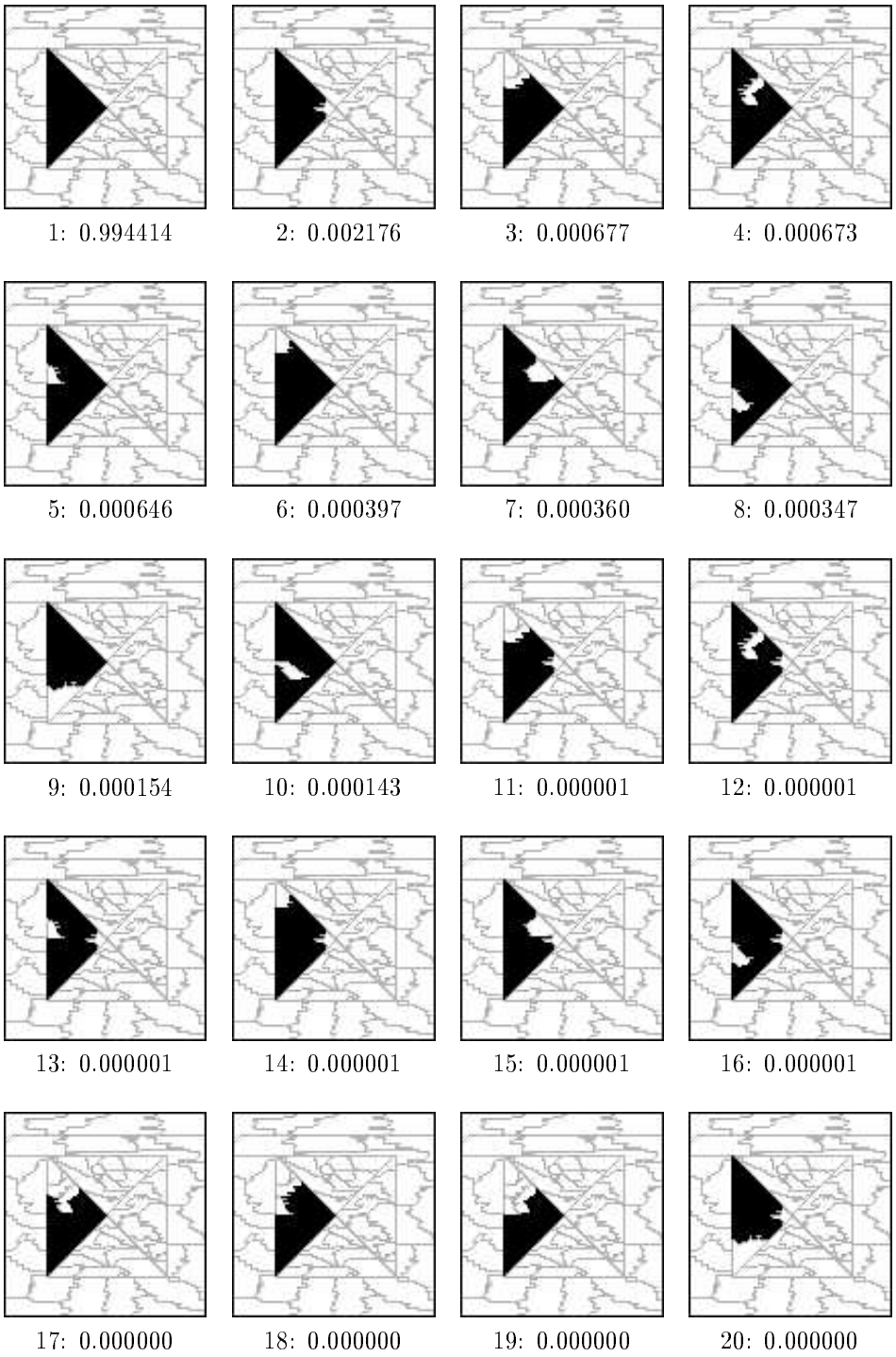


Figure 8.11 Twenty segments that have highest probability in Θ_{13}. There were 198 events in the final cover, with 28 ground events. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

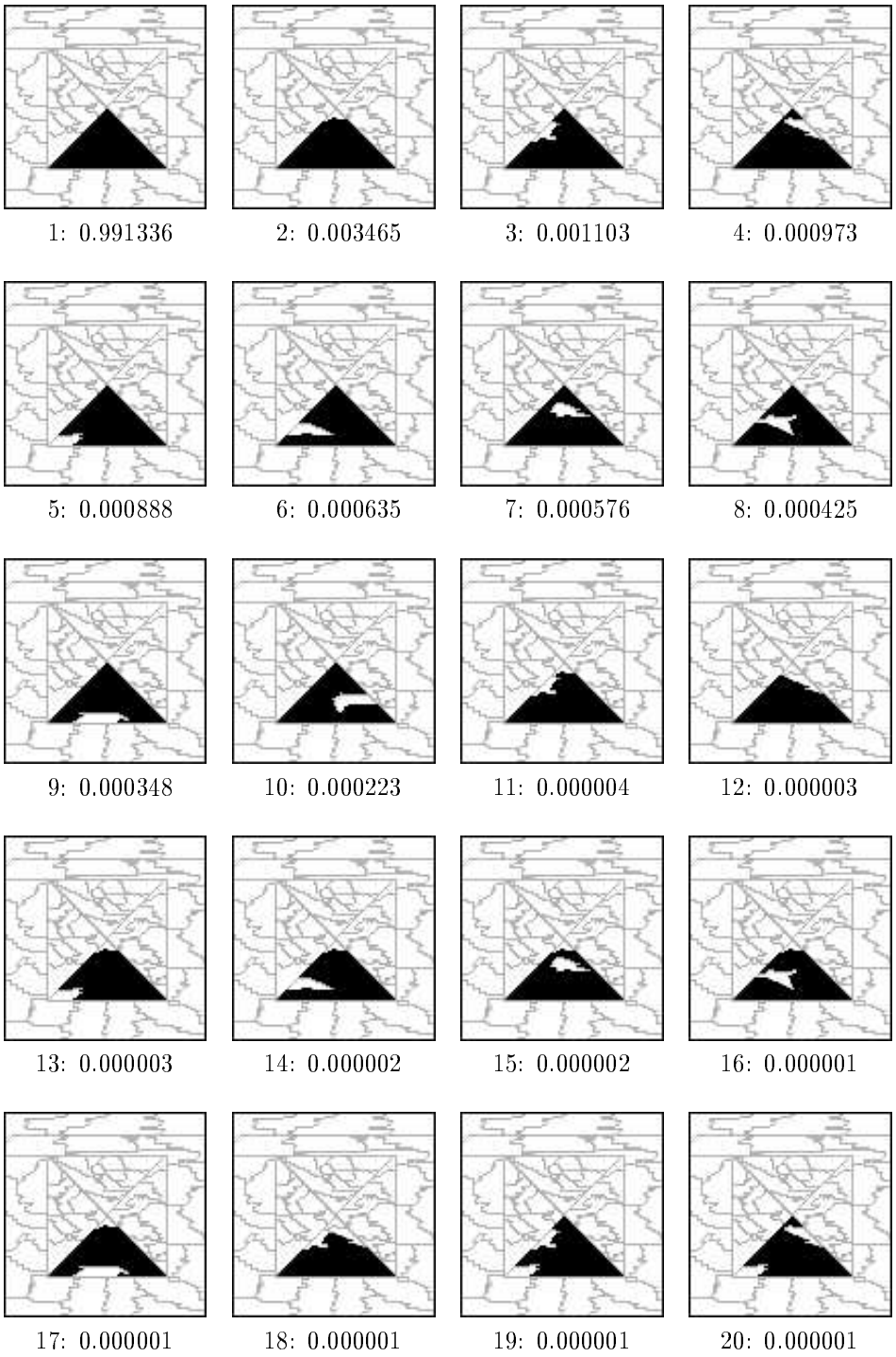


Figure 8.12 Twenty segments that have highest probability in Θ_{20}. There were 123 events in the final cover, with 25 ground events. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

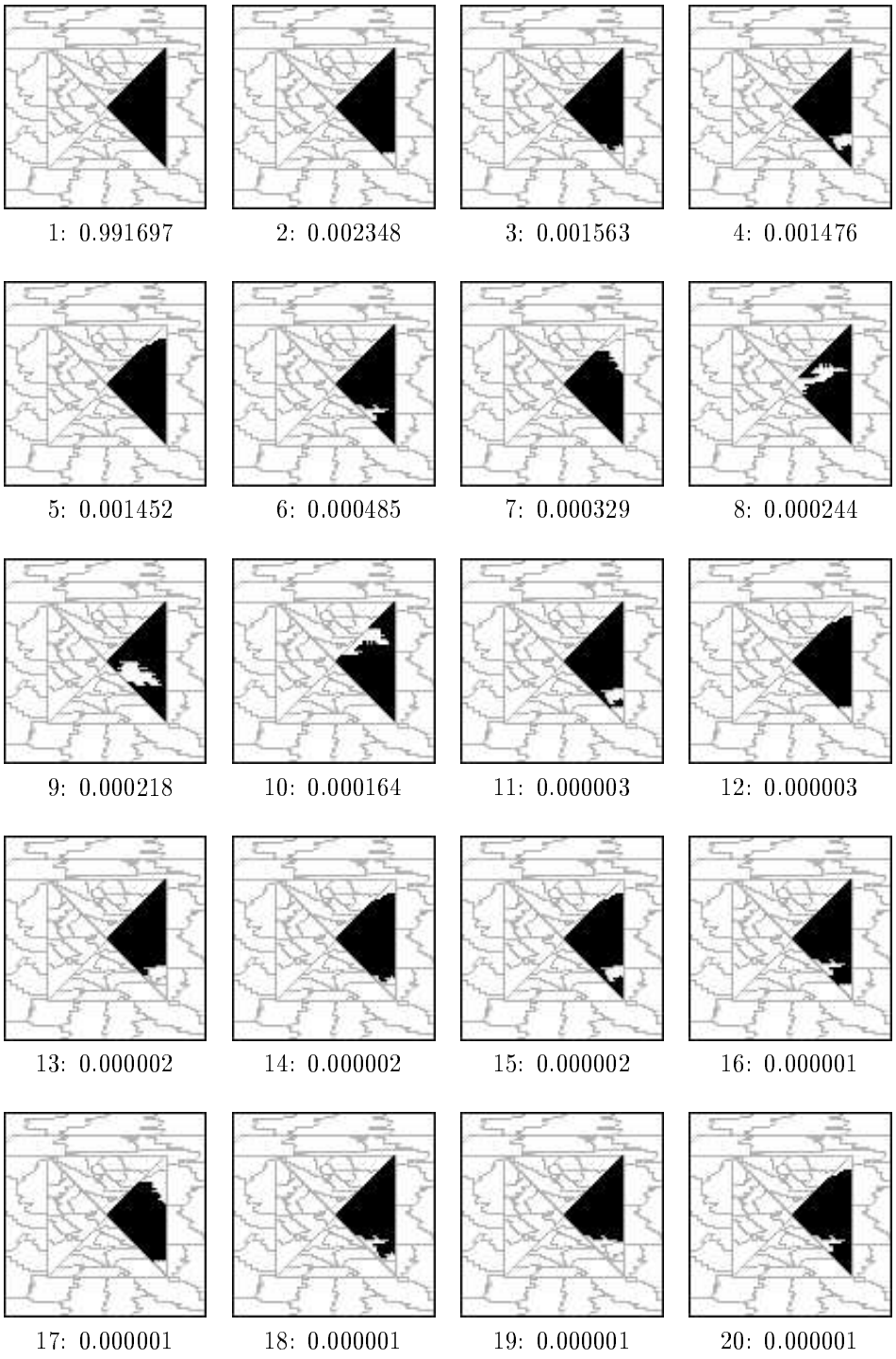


Figure 8.13 Twenty segments that have highest probability in Θ_{36}. There were 234 events in the final cover, with 22 ground events. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

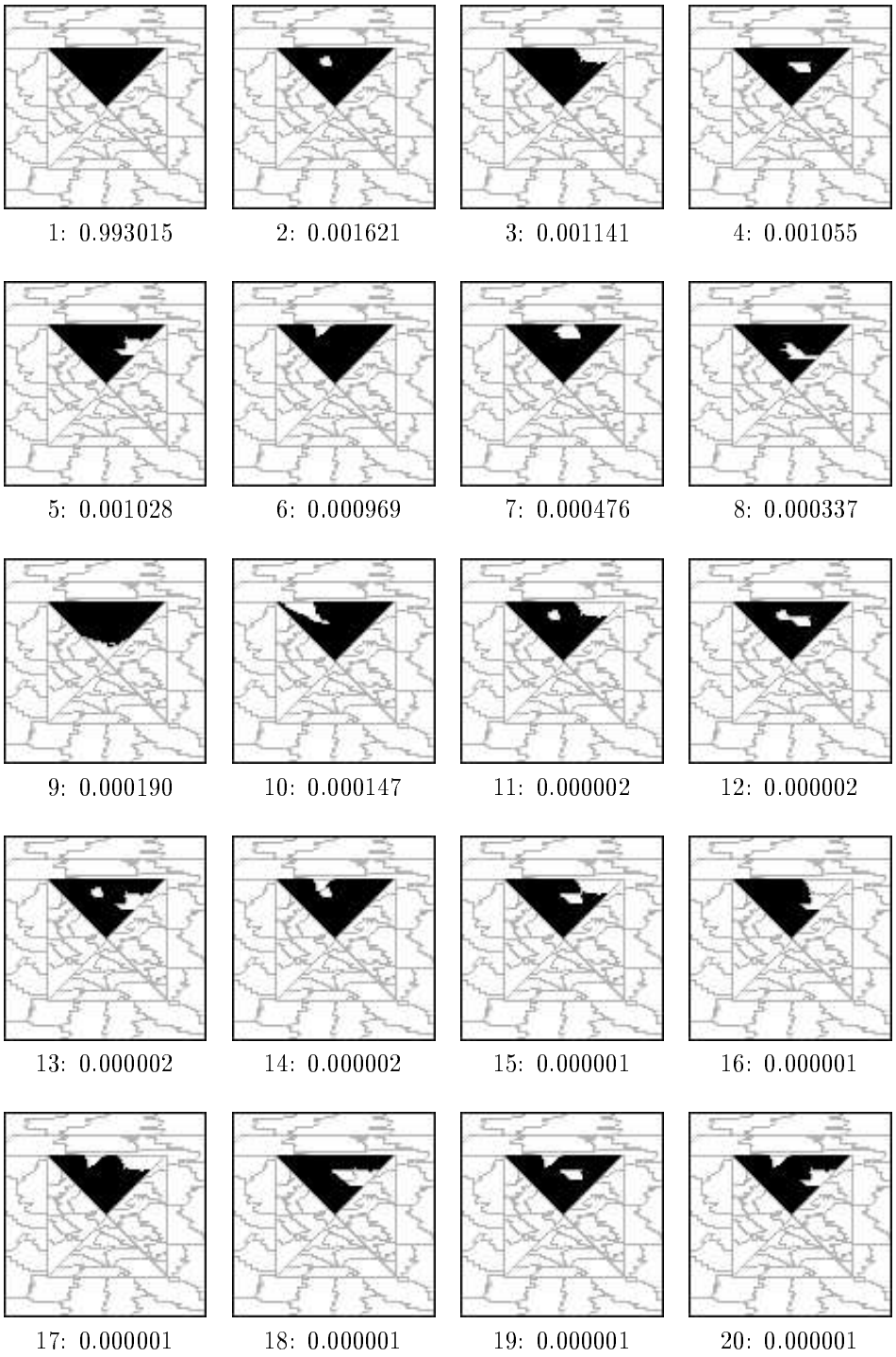


Figure 8.14 Twenty segments that have highest probability in Θ_{42}. There were 103 events in the final cover, with 26 ground events. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

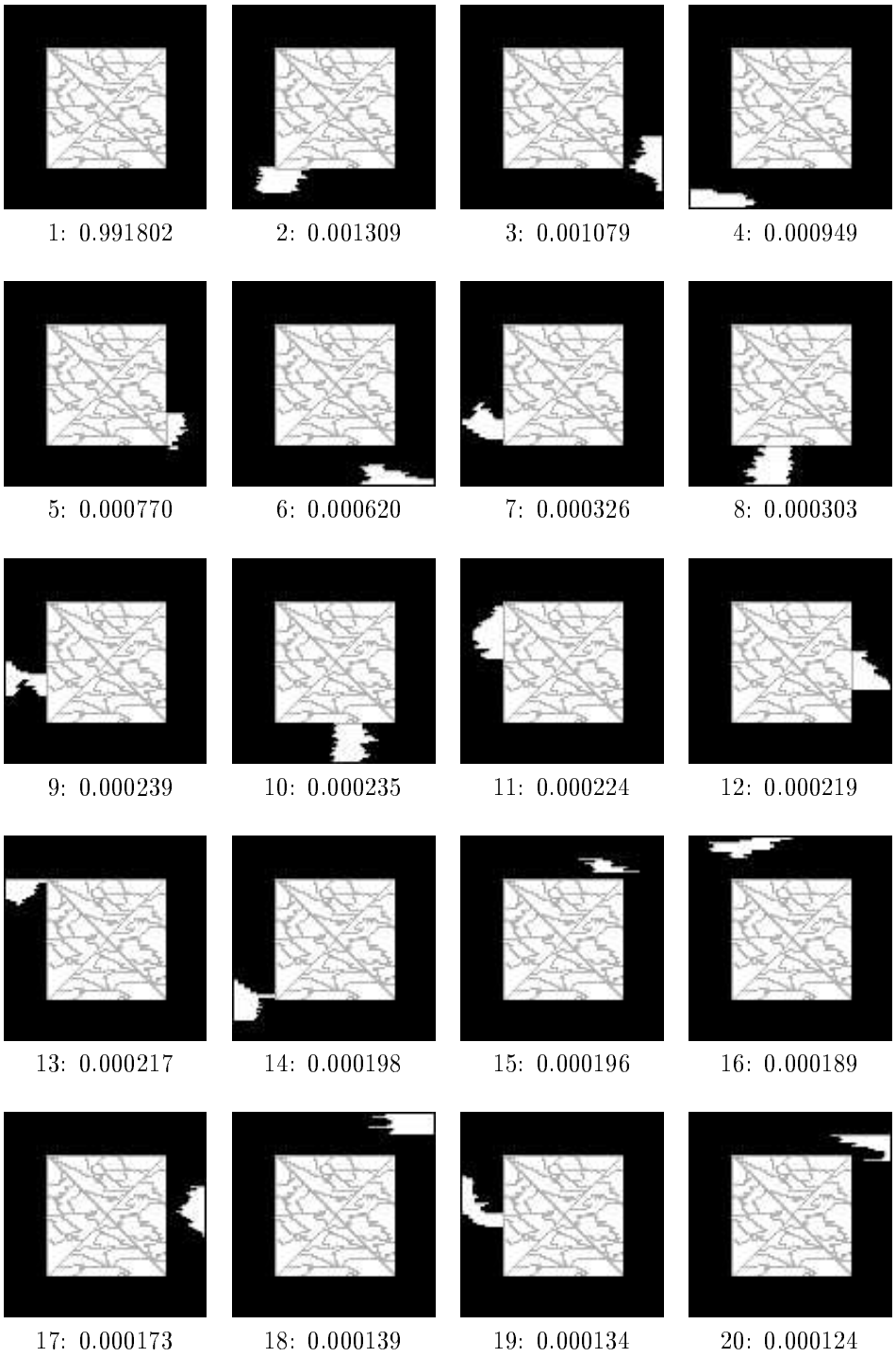


Figure 8.15 Twenty segments that have highest probability in Θ_{73}. There were 550 events in the final cover, with 29 ground events. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

We next present present three examples which use the IE-dependent model (keeping the sum-of-squares observation space). Figure 8.16 shows the TSS representation obtained for Θ_{73}. In this case, the correct segments is ranked second, which is significantly improved over the case in Figure 8.9. Figures 8.17 and 8.18 show two cases with an extremely high noise level, $\sigma^2 = 2.0$.

The final three TSS experiments on the synthetic image demonstrate the alternative use the identity-maps observation space. There are three representations of Θ_{14} shown in Figures 8.19-8.21, under different conditions.

We next turn our attention to the real range images. The first two use the planar model, and the other two use the quadric model. For each of these images we present an artificial intensity rendering of the data set, and a map of the region set, $\mathcal{R}$. The sum-of-squares observation space is used in each of these experiments.

We first consider the image and regions given by Figure 8.22. Three different initial regions, which are used in the experiments, are indicated in Figure 8.23. Figures 8.24-8.26 show the resulting TSS representations for each of the initial regions, using the IE-independent model. Since there are many small regions presented, the favorable segments in the TSS appear very similar. Figures 8.27-8.29 show some TSS representations, determined when smaller regions are removed. In these cases, the differences between the alternative segments are clearer. Figures 8.30-8.32 show the results when alternatively using the IE-dependent model.

We next consider the image shown in Figure 8.33, and an initial region, shown in Figure 8.34. A result using the IE-independent model is shown in Figure 8.35, and a result using the IE-dependent model is shown in Figure 8.36.

We finally consider two quadric images, shown in Figures 8.37 and 8.38, with initial regions shown in Figures 8.39.(a) and 8.39.(b). The initial regions shown in Figure 8.37.(c) were obtained by starting with around 900 tiny regions (using the method from

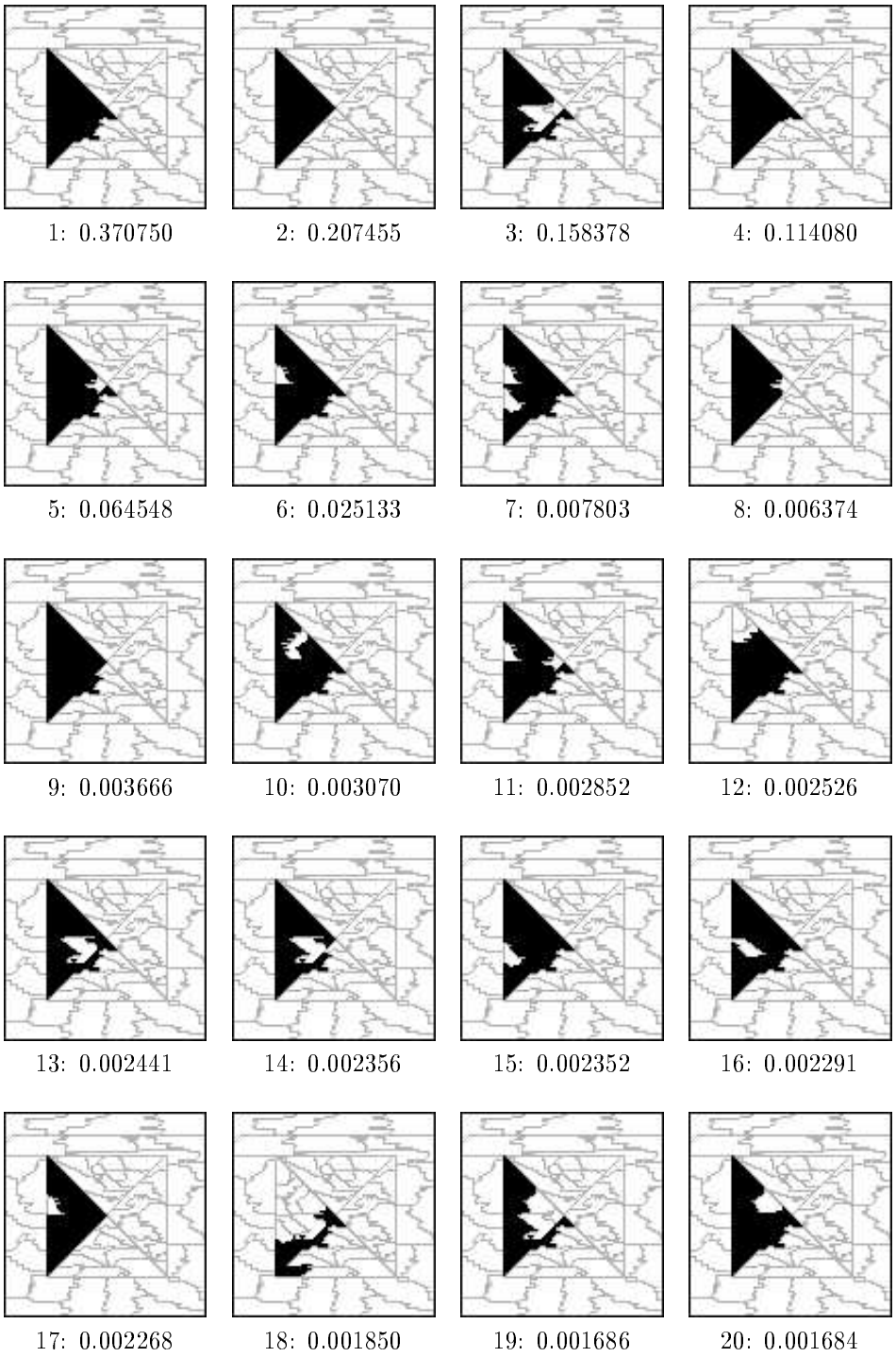


Figure 8.16 Twenty segments that have highest probability in Θ_{14}. There were 230 events in the final cover, with 29 ground events. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-dependent model are in use.

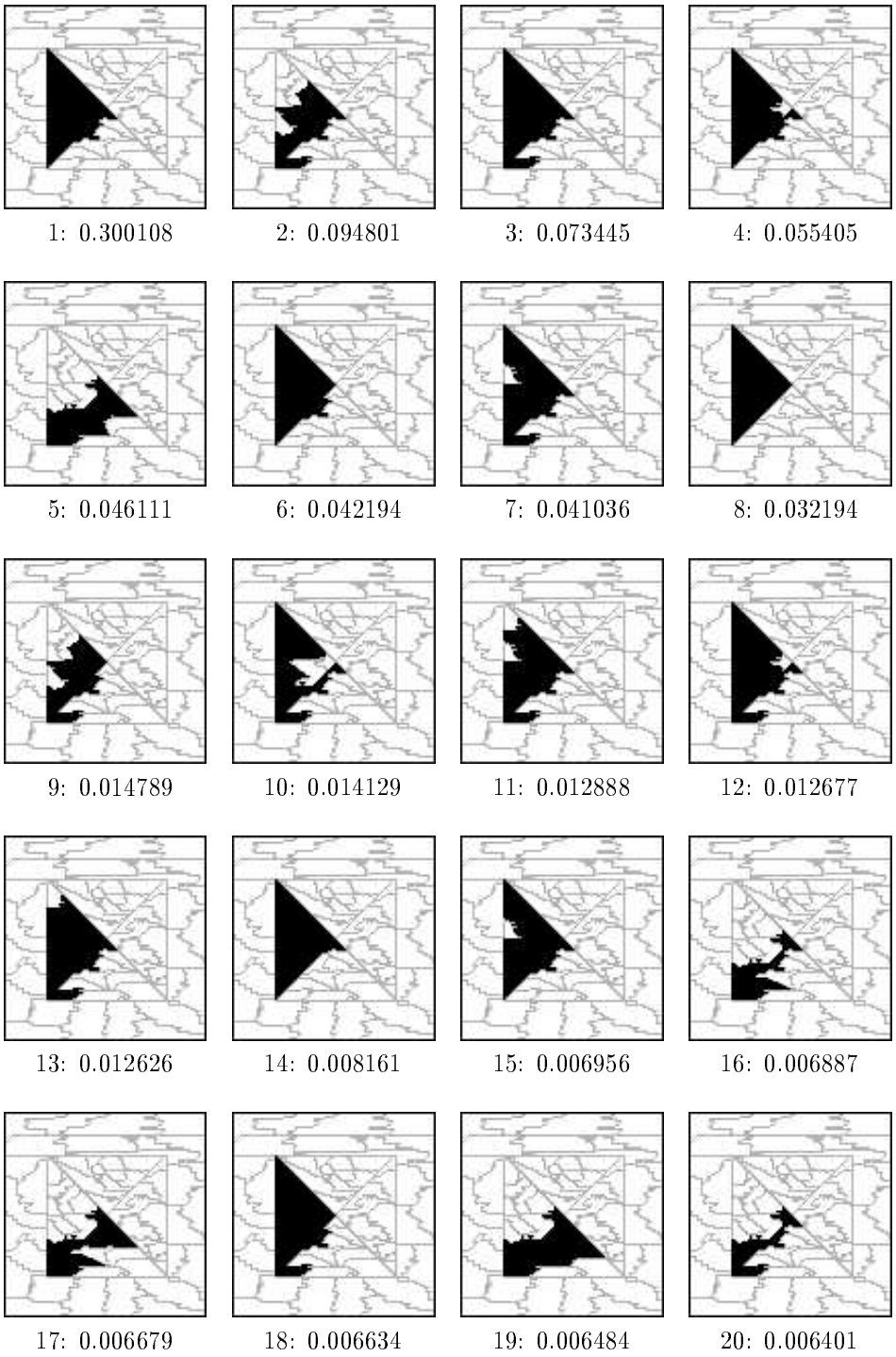


Figure 8.17 Twenty segments that have highest probability in Θ_{14}. There were 285 events in the final cover, with 29 ground events. In this experiment, $\sigma^2 = 2.0$, and the sum-of-squares observation space and IE-dependent model are in use.

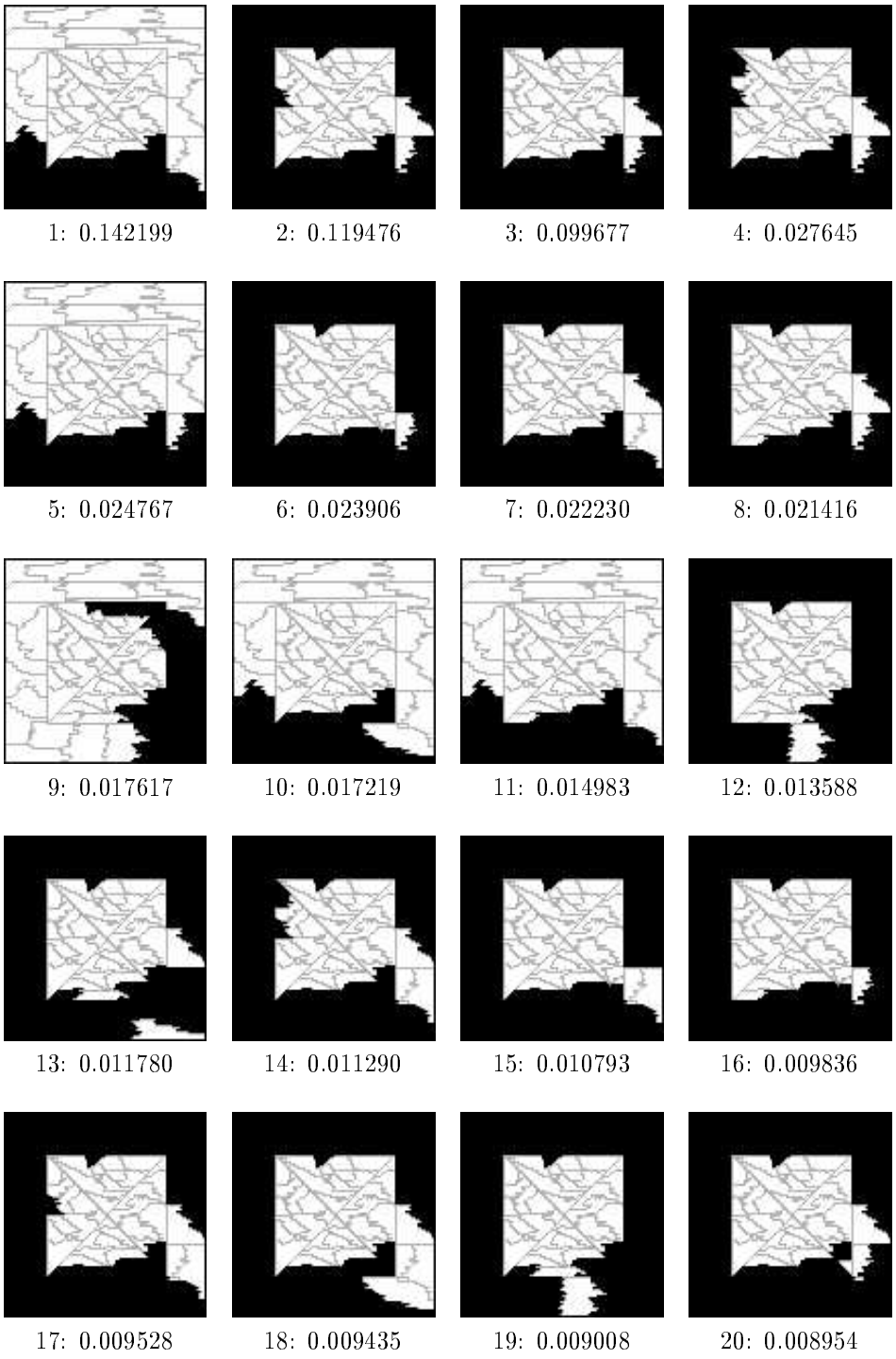


Figure 8.18 Twenty segments that have highest probability in Θ_{20}. There were 627 events in the final cover, with 41 ground events. In this experiment, $\sigma^2 = 2.0$, and the sum-of-squares observation space and IE-dependent model are in use.

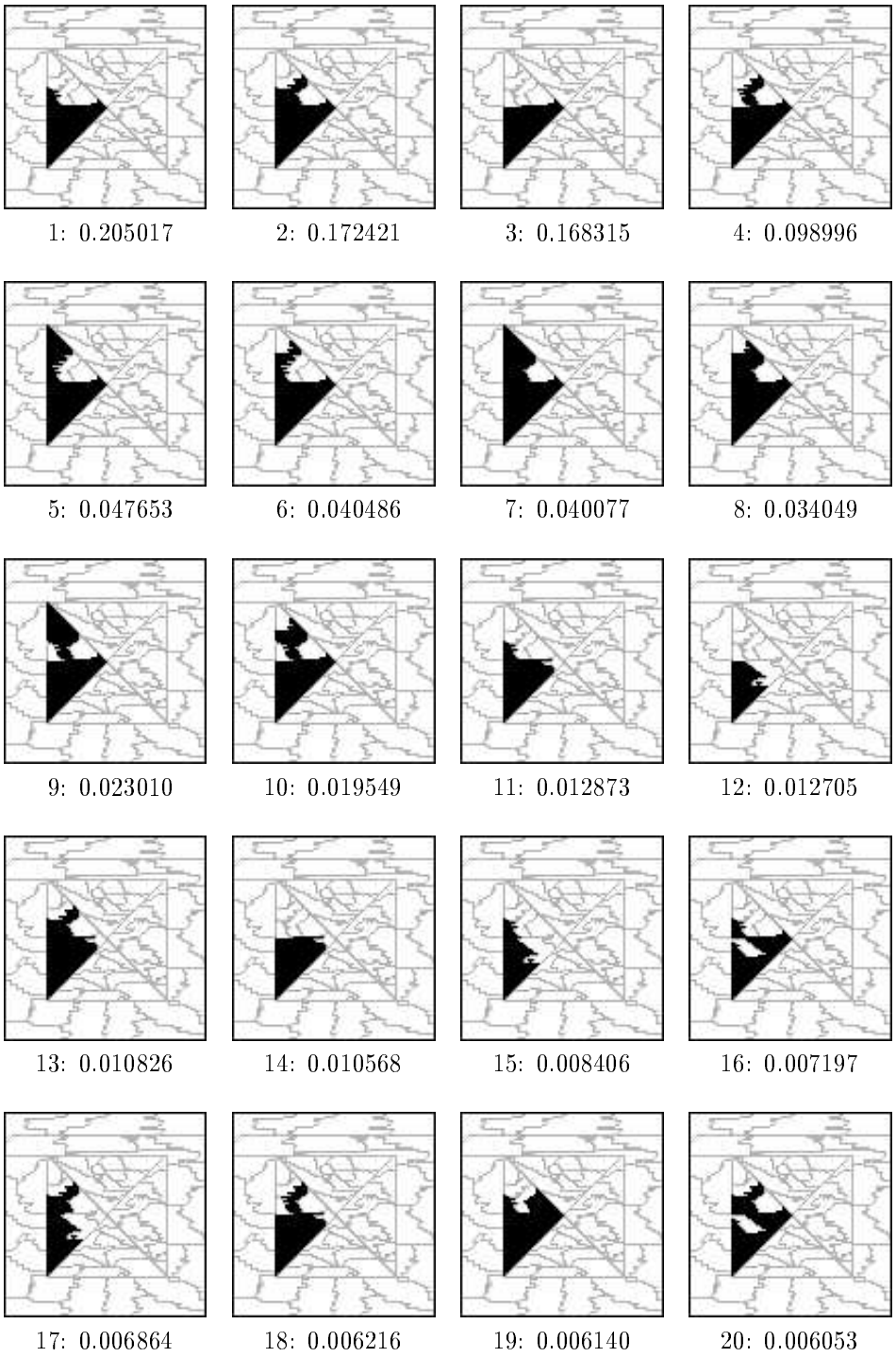


Figure 8.19 Twenty segments that have highest probability in Θ_{14}. There were 144 events in the final cover, with 25 ground events. In this experiment, $\sigma^2 = 1.0$, and the identity-maps observation space and IE-independent model are in use.

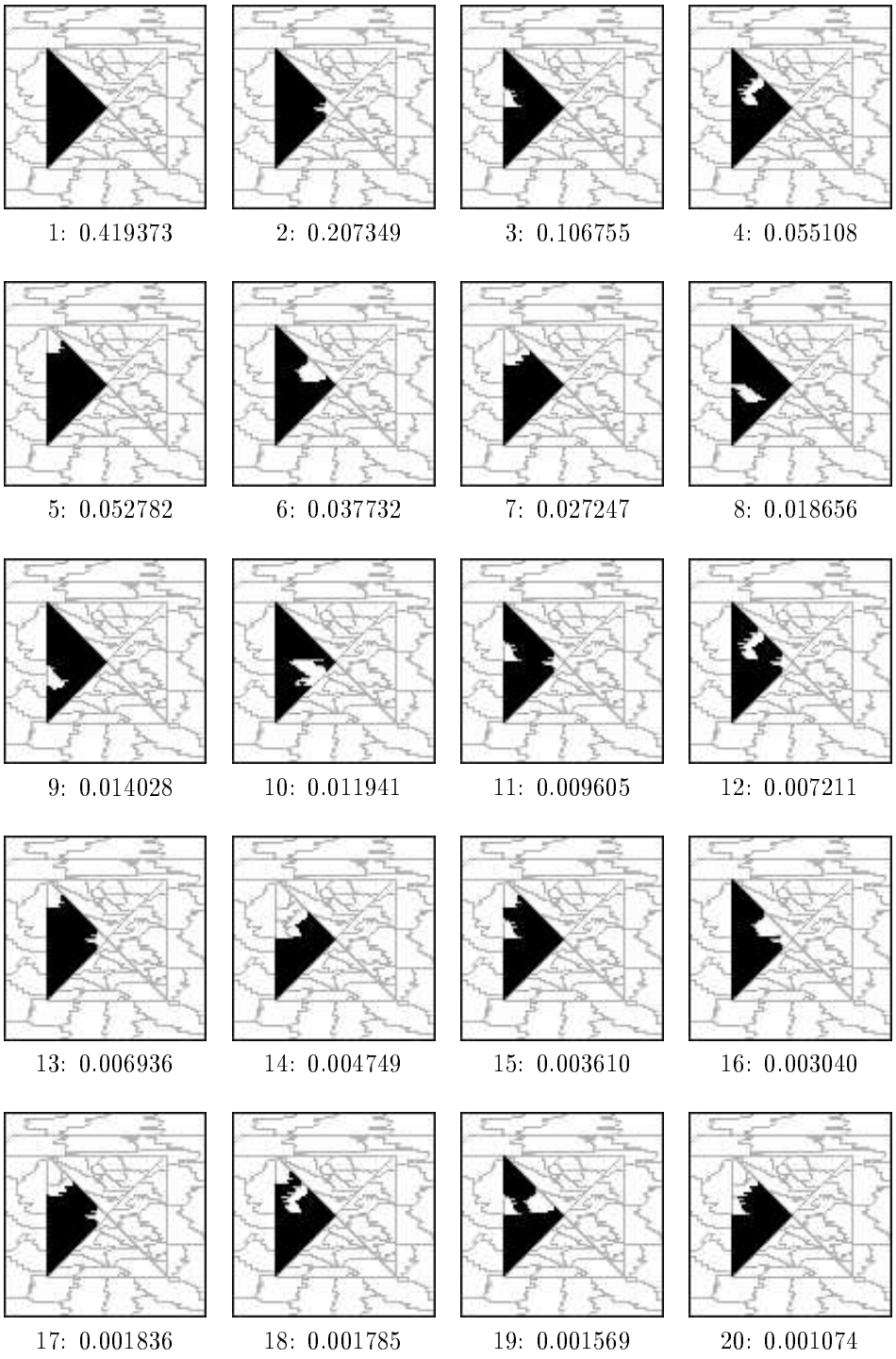


Figure 8.20 Twenty segments that have highest probability in Θ_{14}. There were 166 events in the final cover, with 25 ground events. In this experiment, $\sigma^2 = 0.1$, and the identity-maps observation space and IE-independent model are in use.

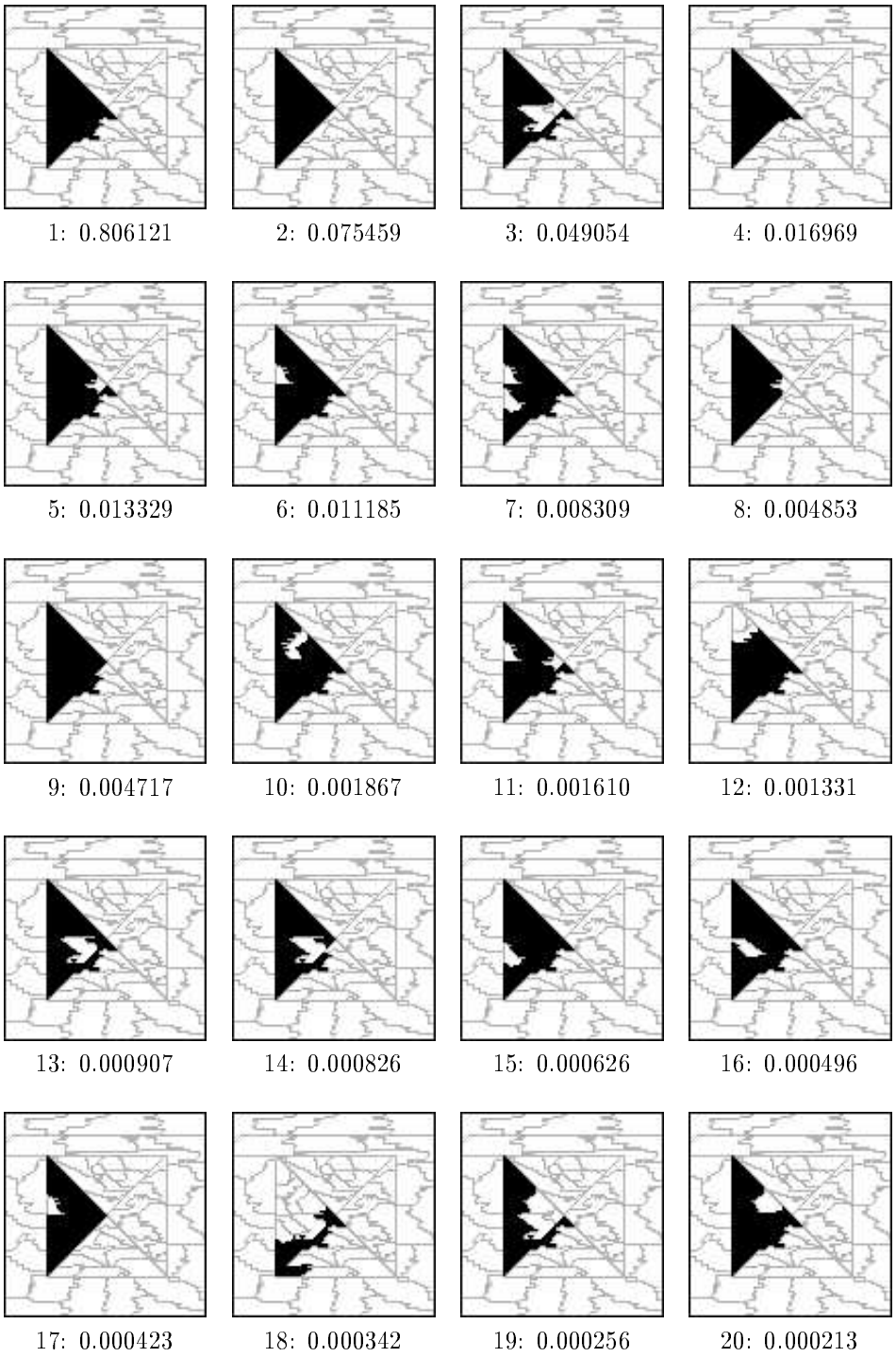
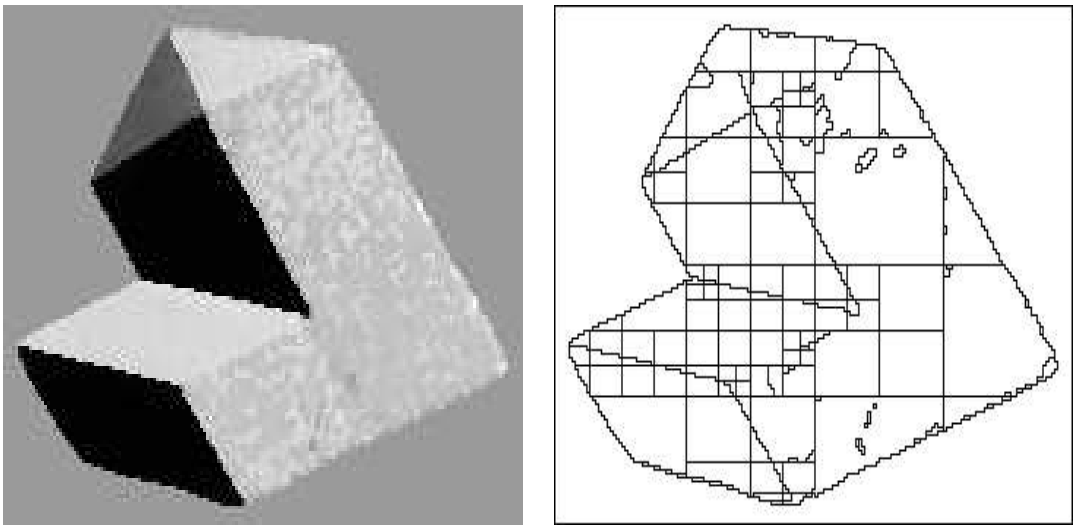
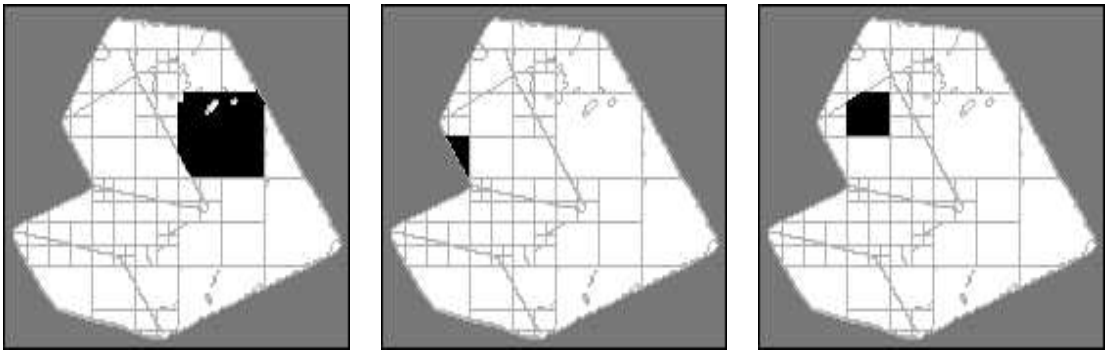


Figure 8.21 Twenty segments that have highest probability in Θ_{14}. There were 143 events in the final cover, with 25 ground events. In this experiment, $\sigma^2 = 1.0$, and the identity-maps observation space and IE-dependent model are in use.



(a) (b)

Figure 8.22 (a) A rendering of the data set, (b) the set of regions, $\mathcal{R}$.



(a) (b) (c)

Figure 8.23 Initial regions that were selected for TSS experiments: (a) R_{264} (b) R_{224}, (c) R_{131}.

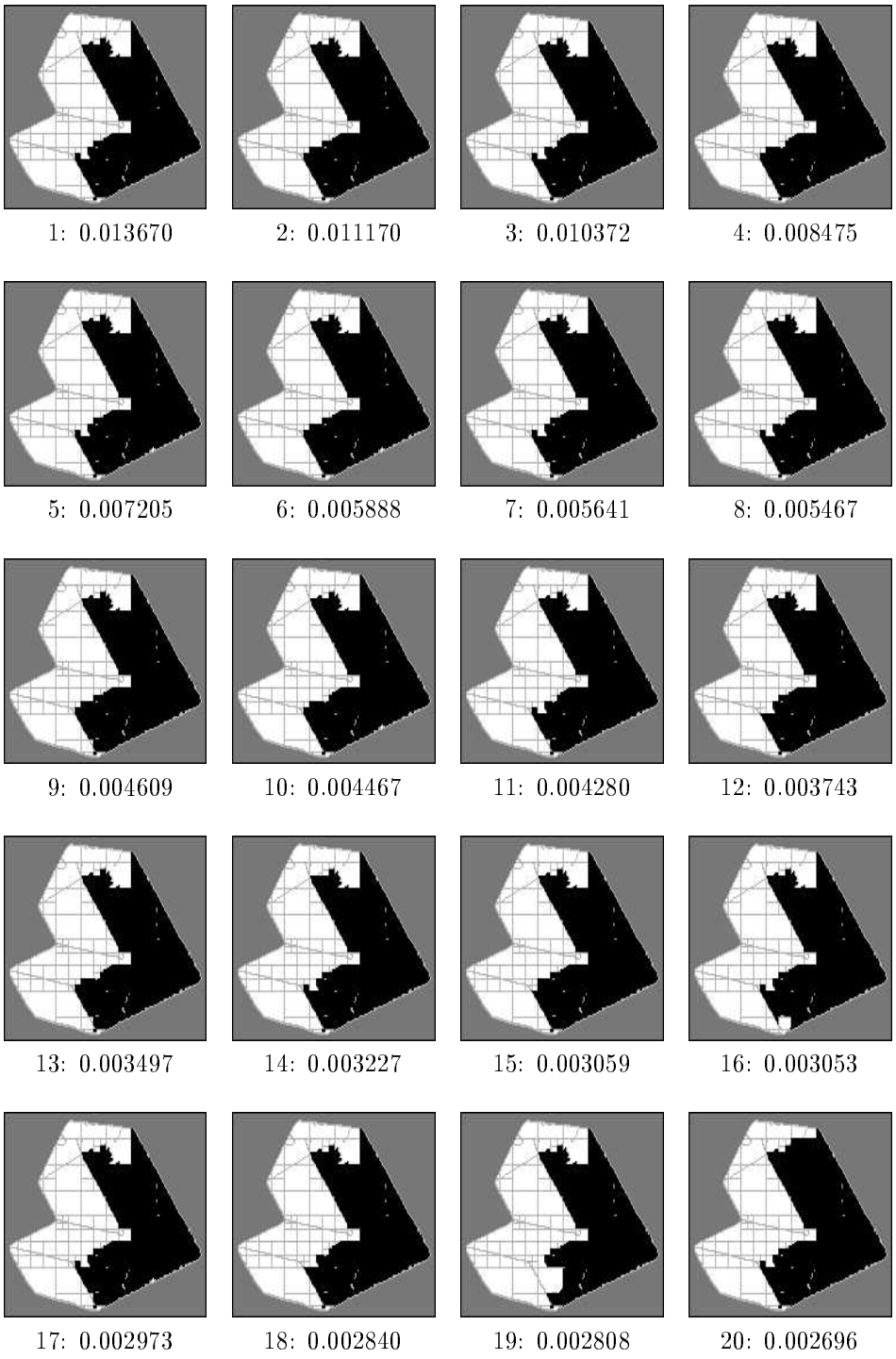


Figure 8.24 Twenty segments that have highest probability in Θ_{264}. There were 2294 events in the final cover, with 20 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use.

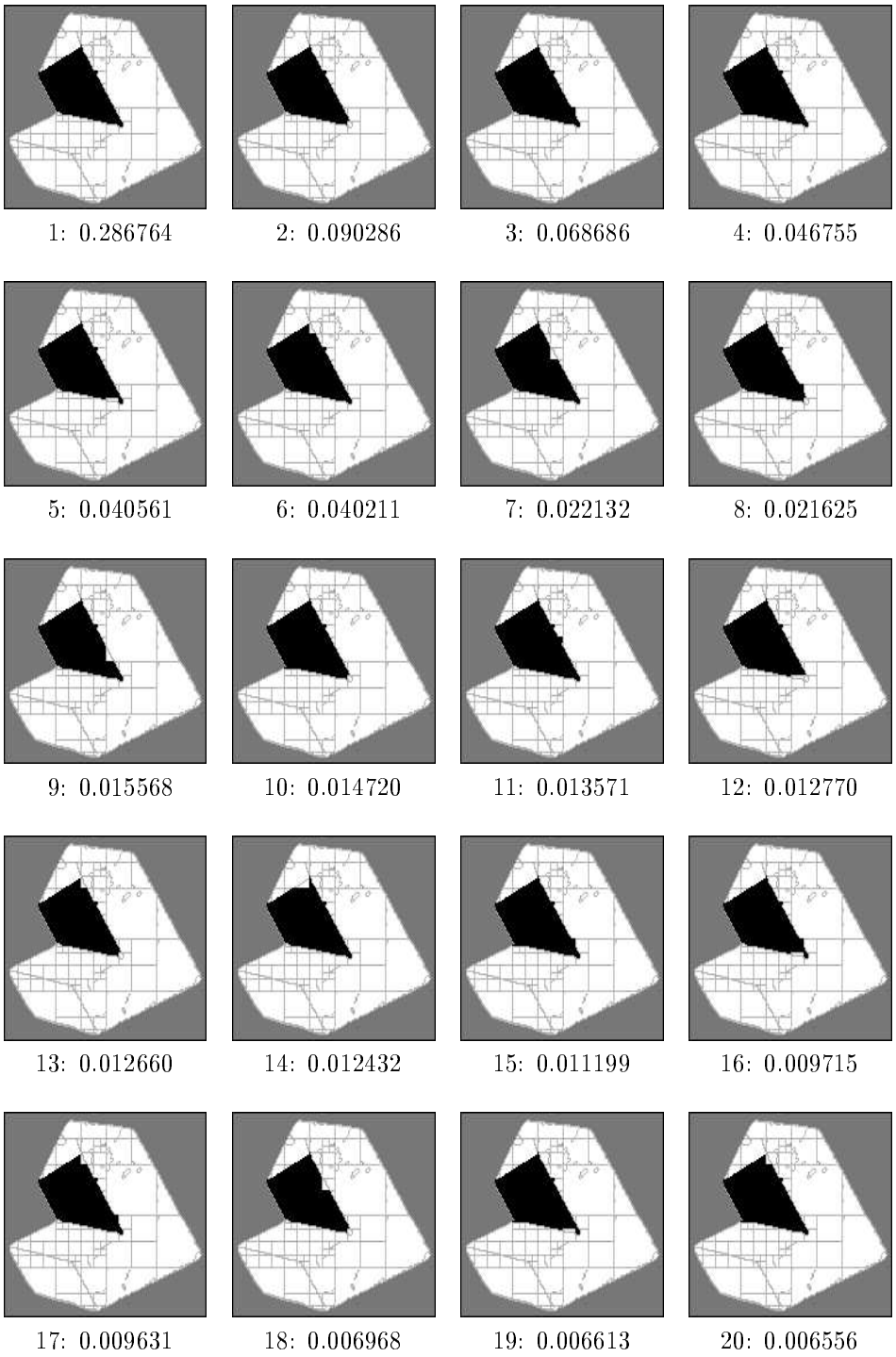


Figure 8.25 Twenty segments that have highest probability in Θ_{224}. There were 405 events in the final cover, with 29 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use.

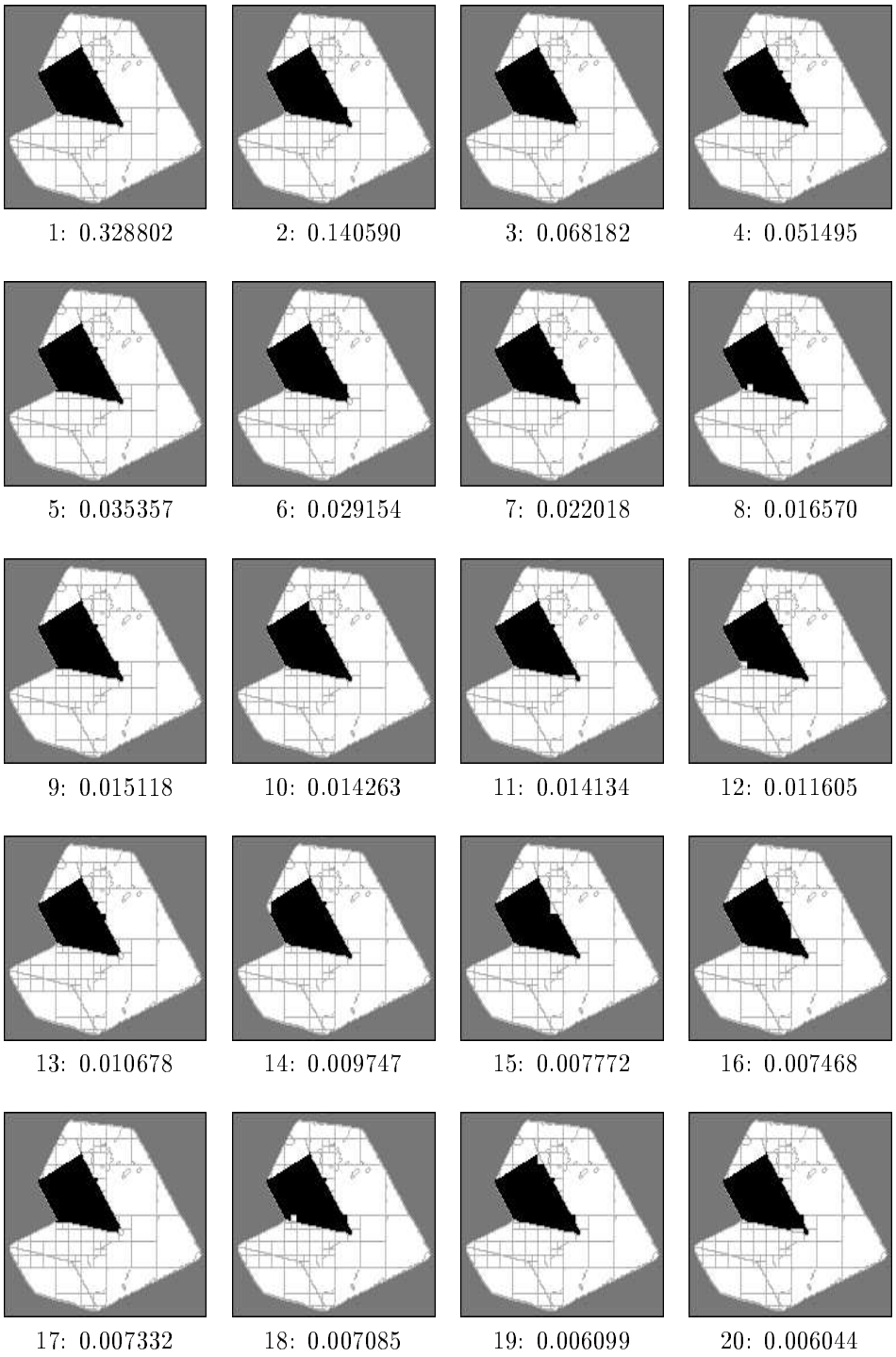


Figure 8.26 Twenty segments that have highest probability in Θ_{131}. There were 483 events in the final cover, with 31 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use.

Figure 8.27 Twenty segments that have highest probability in Θ_{264}. There were 361 events in the final cover, with 23 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use. Regions, R_k, such that $|R_k| < 24$ have been removed.

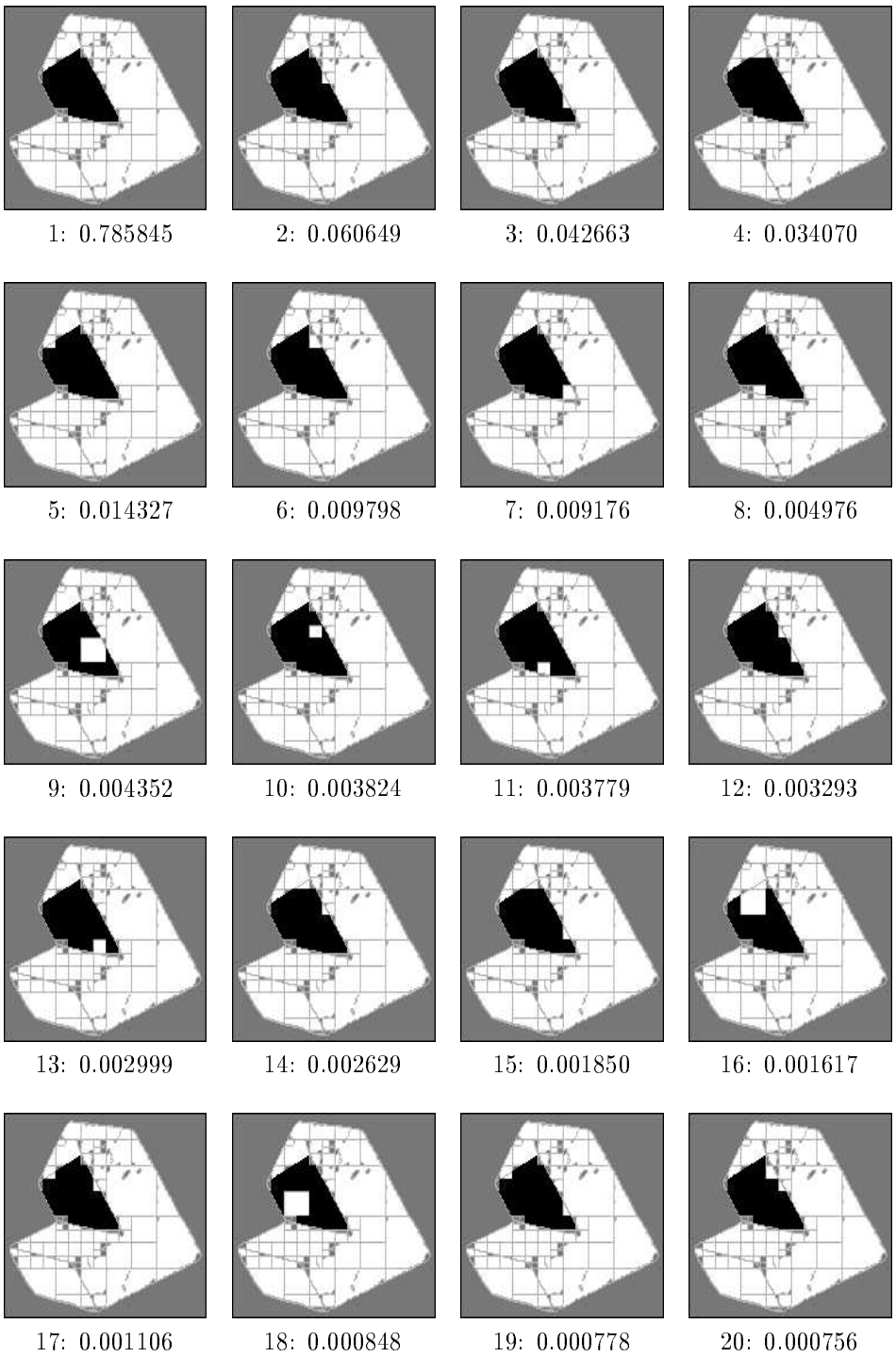


Figure 8.28 Twenty segments that have highest probability in Θ_{224}. There were 255 events in the final cover, with 28 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use. Regions, R_k, such that $|R_k| < 24$ have been removed.

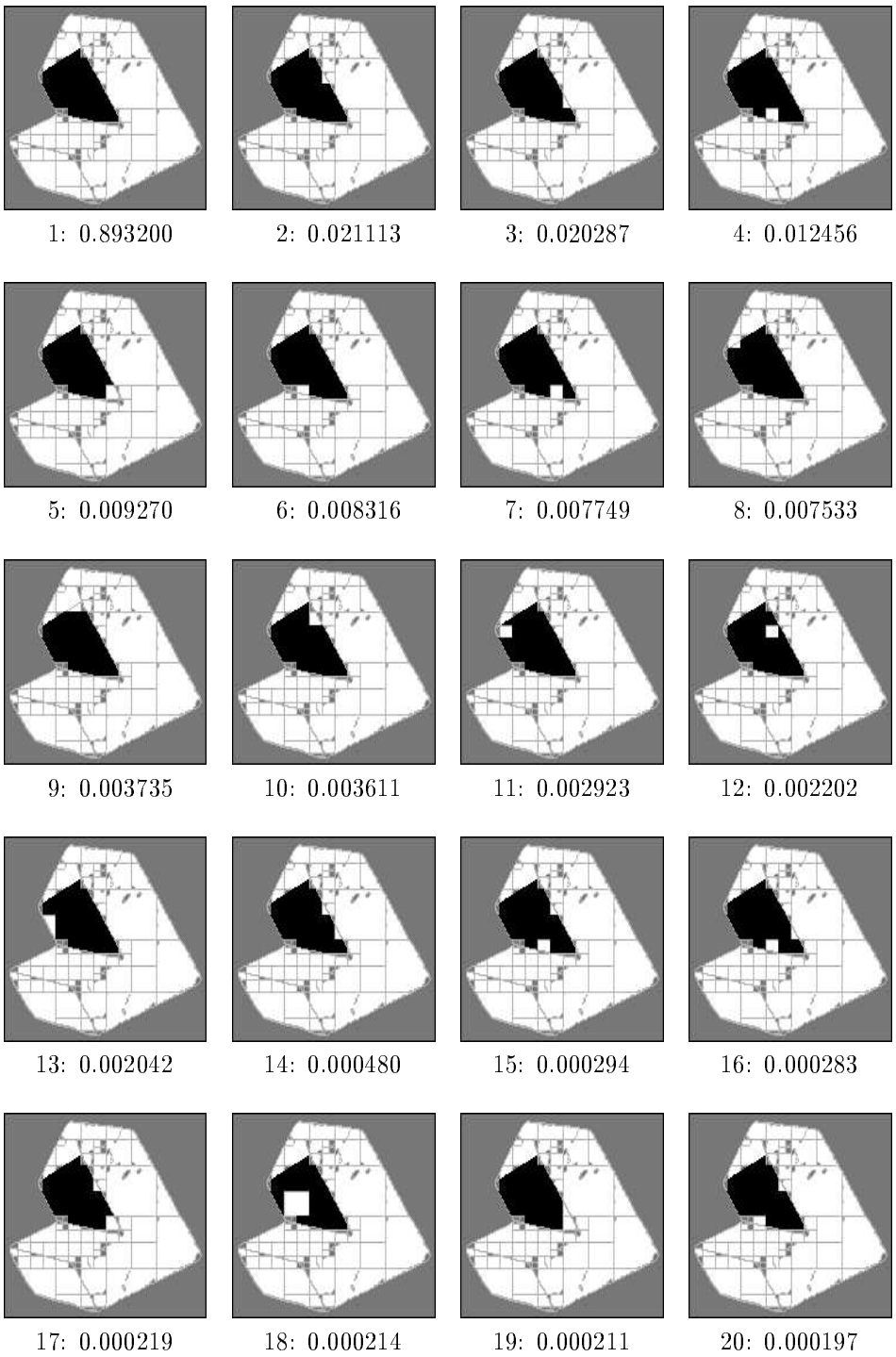


Figure 8.29 Twenty segments that have highest probability in Θ_{131}. There were 205 events in the final cover, with 27 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use. Regions, R_k, such that $|R_k| < 24$ have been removed.

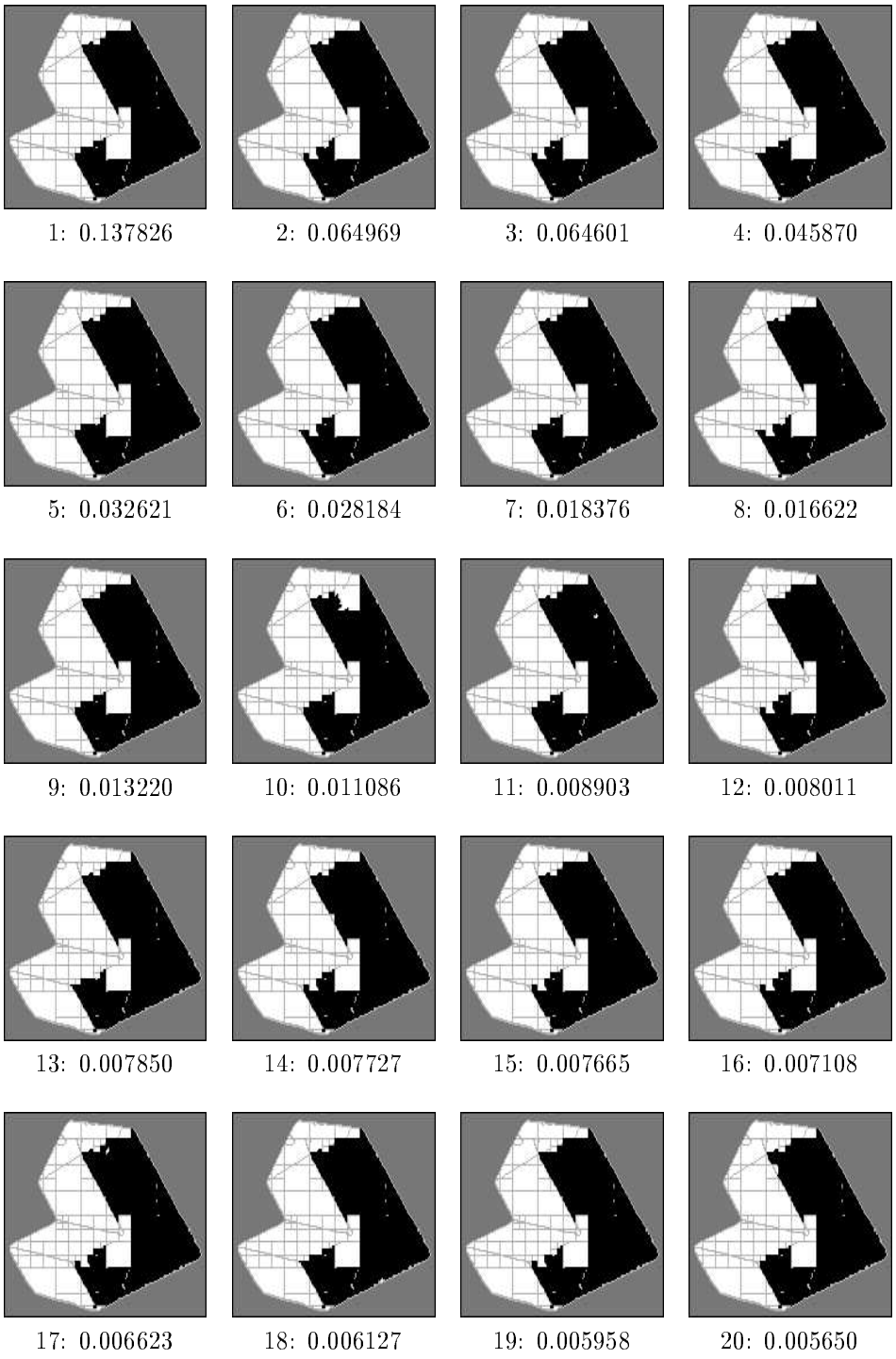


Figure 8.30 Twenty segments that have highest probability in Θ_{264}. There were 927 events in the final cover, with 38 ground events. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.

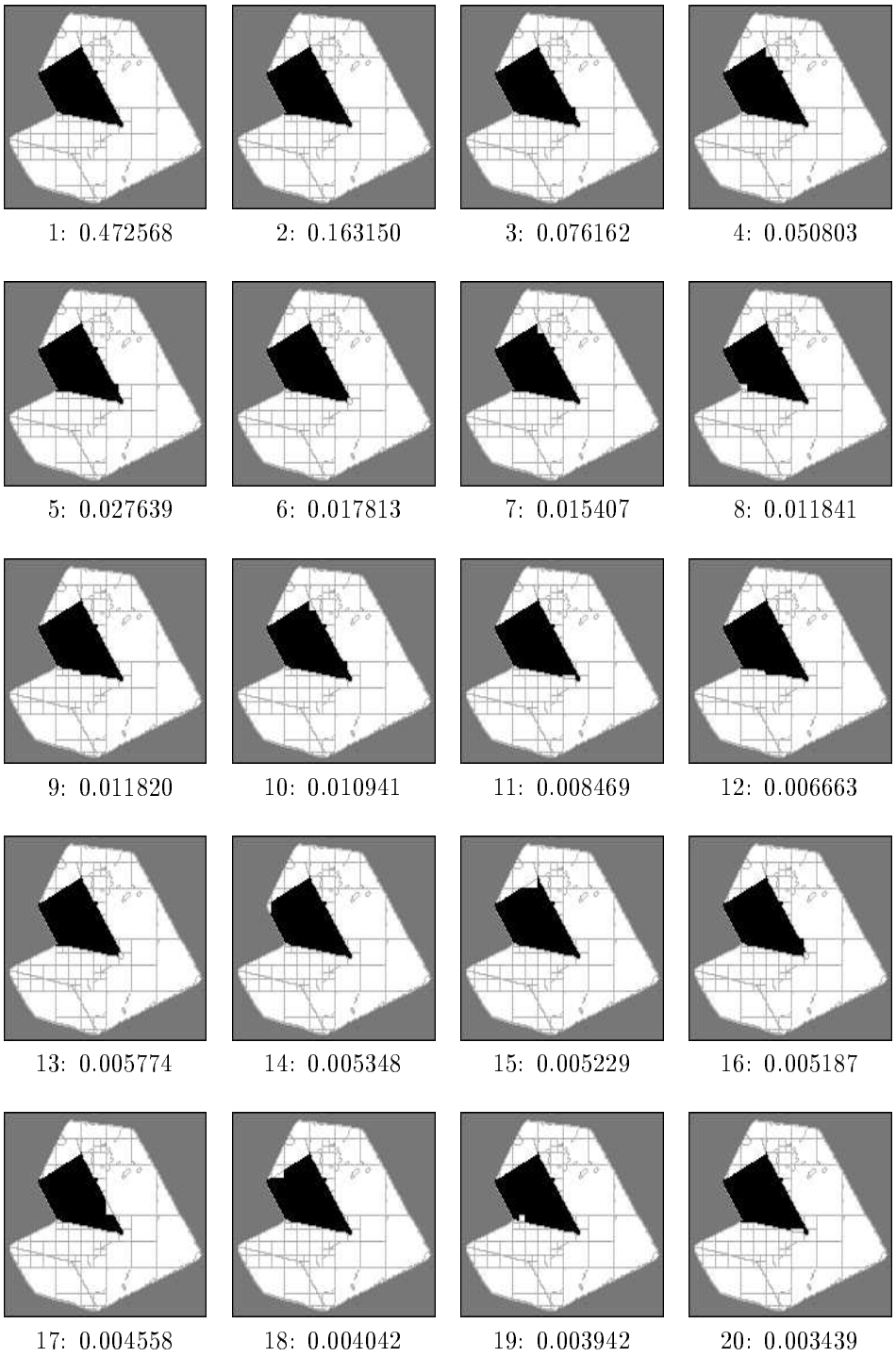


Figure 8.31 Twenty segments that have highest probability in Θ_{224}. There were 475 events in the final cover, with 34 ground events. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.

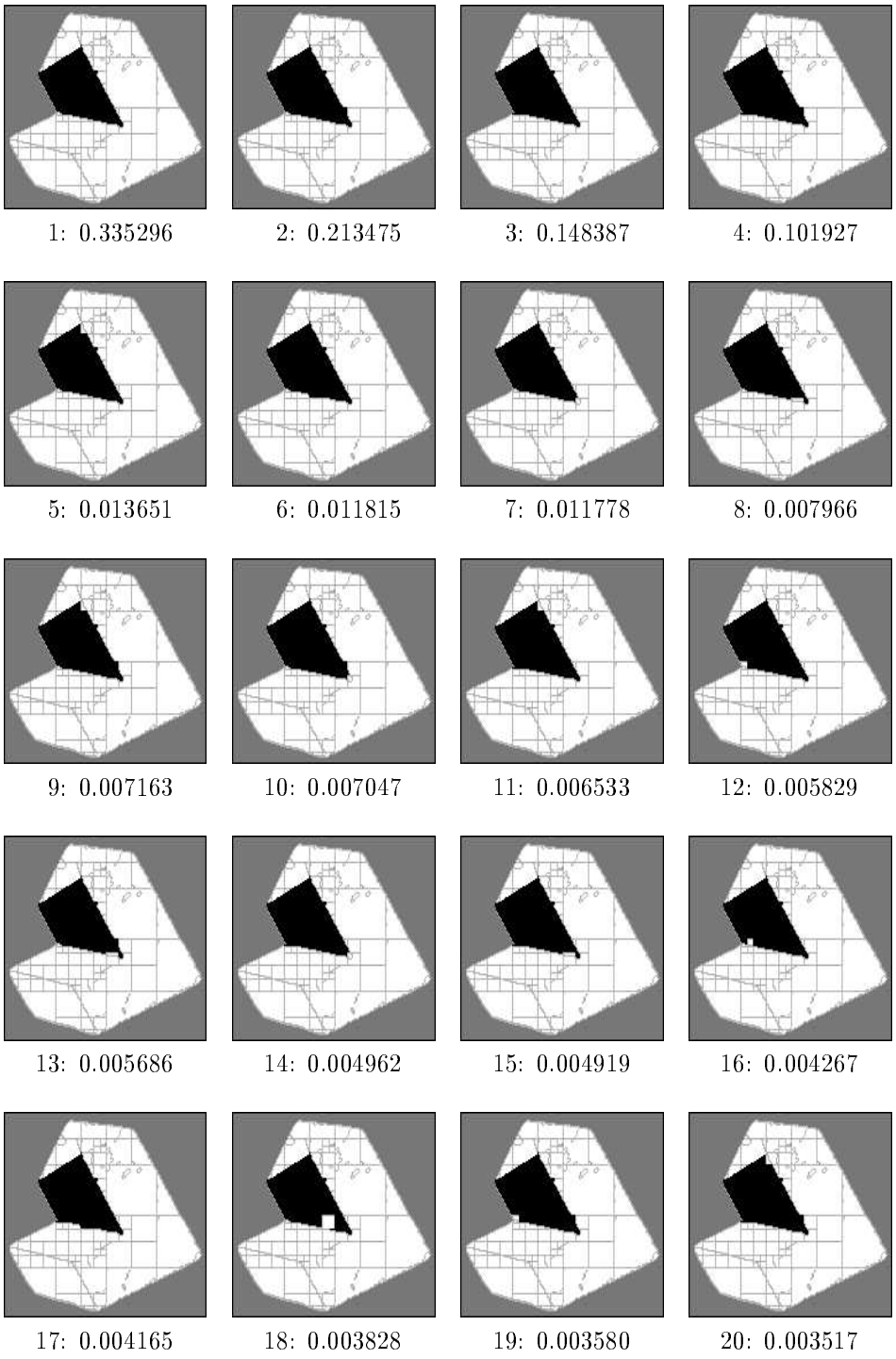
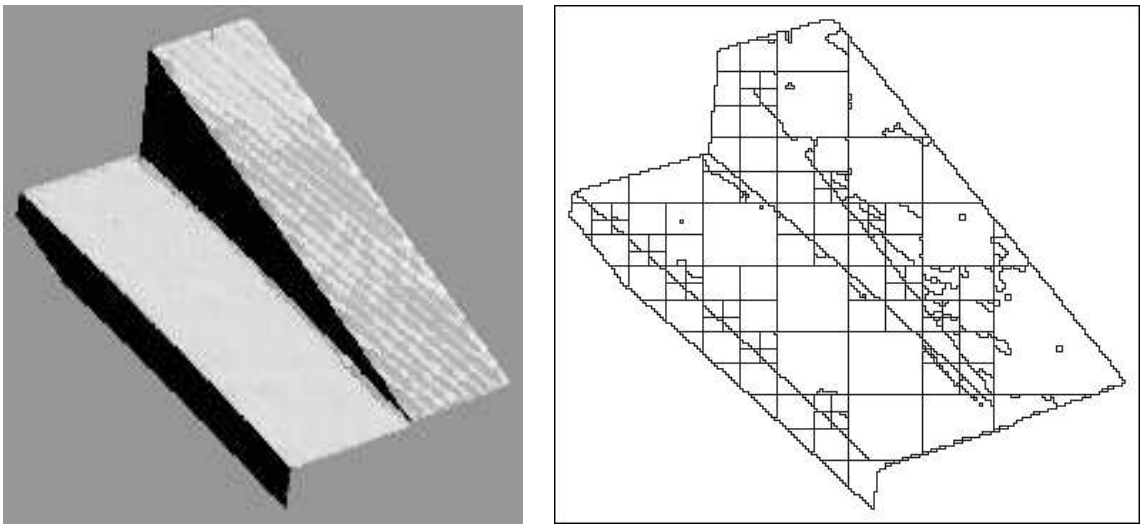


Figure 8.32 Twenty segments that have highest probability in Θ_{131}. There were 537 events in the final cover, with 41 ground events. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.



(a) (b)

Figure 8.33 (a) The rendering of the data set, (b) the set of regions, $\mathcal{R}$

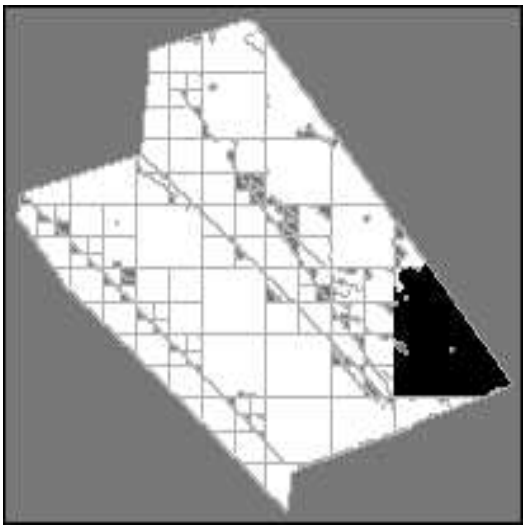


Figure 8.34 The initial region, R_{568}

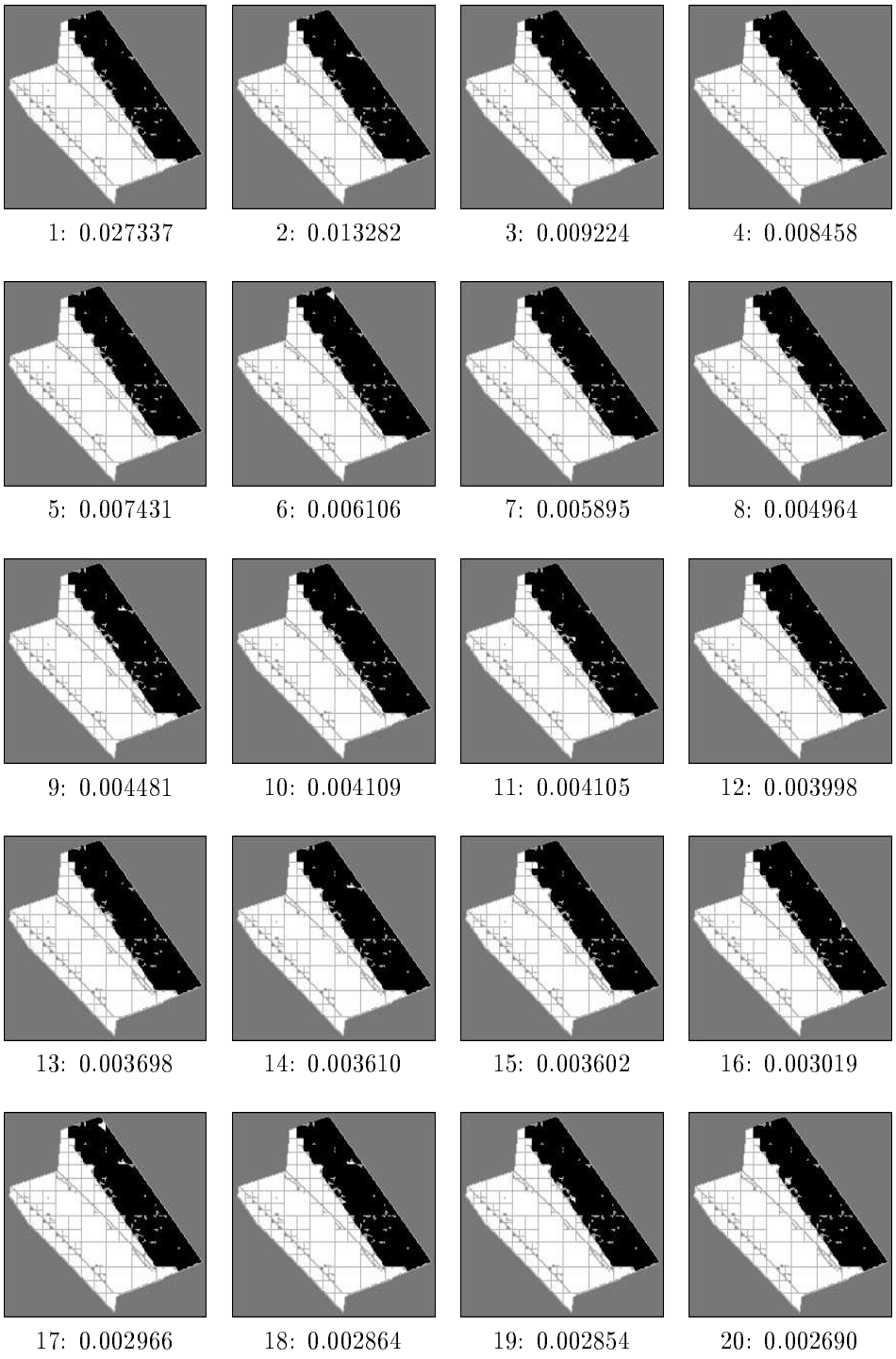


Figure 8.35 Twenty segments that have highest probability in Θ_{568}. There were 1638 events in the final cover, with 29 ground events. In this experiment, the sum-of-squares observation space and IE-independent model are in use.

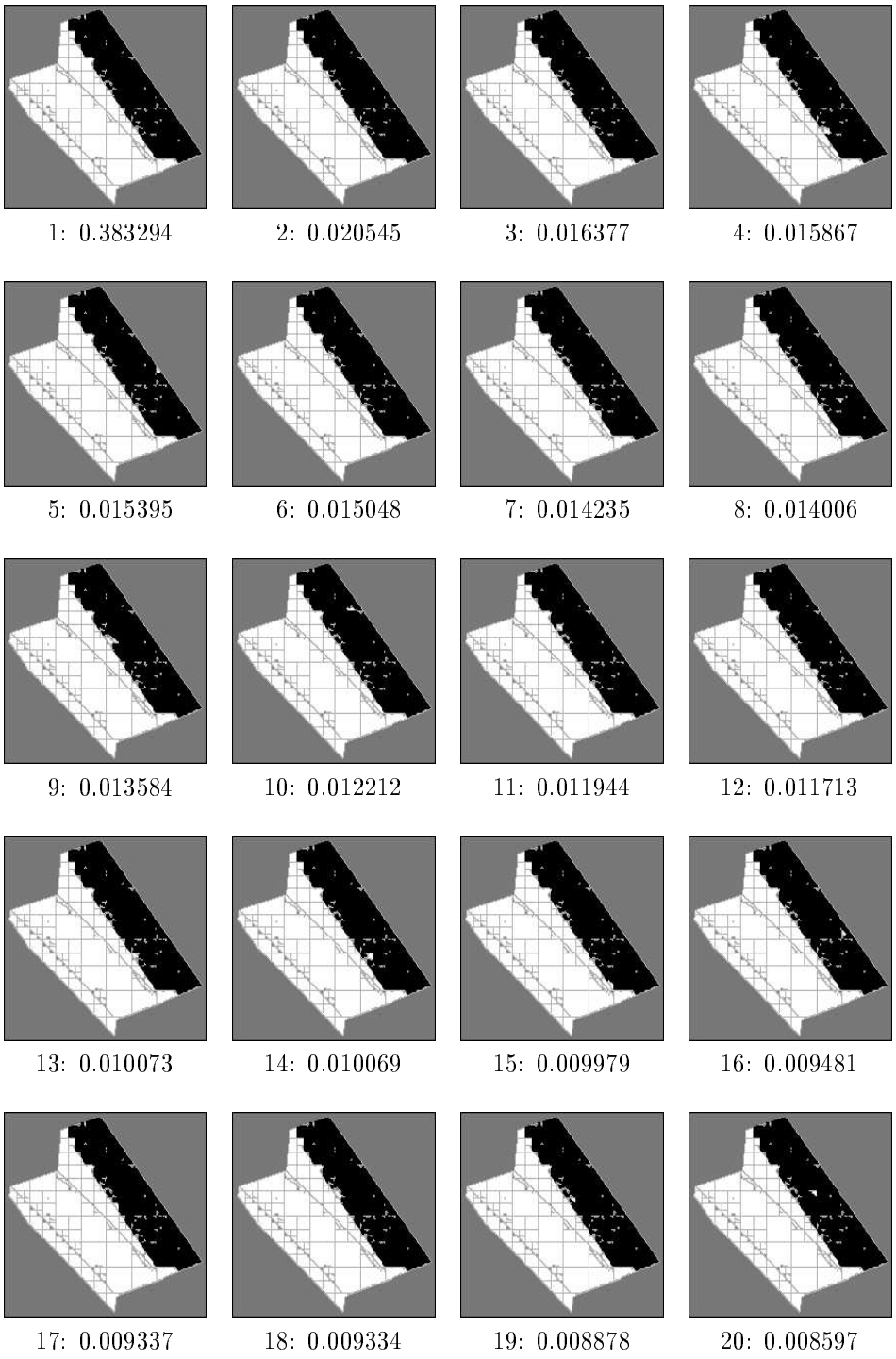
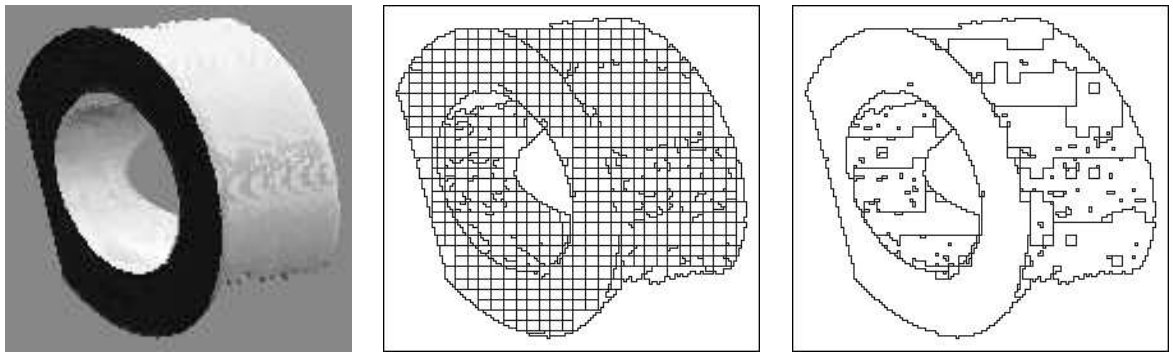
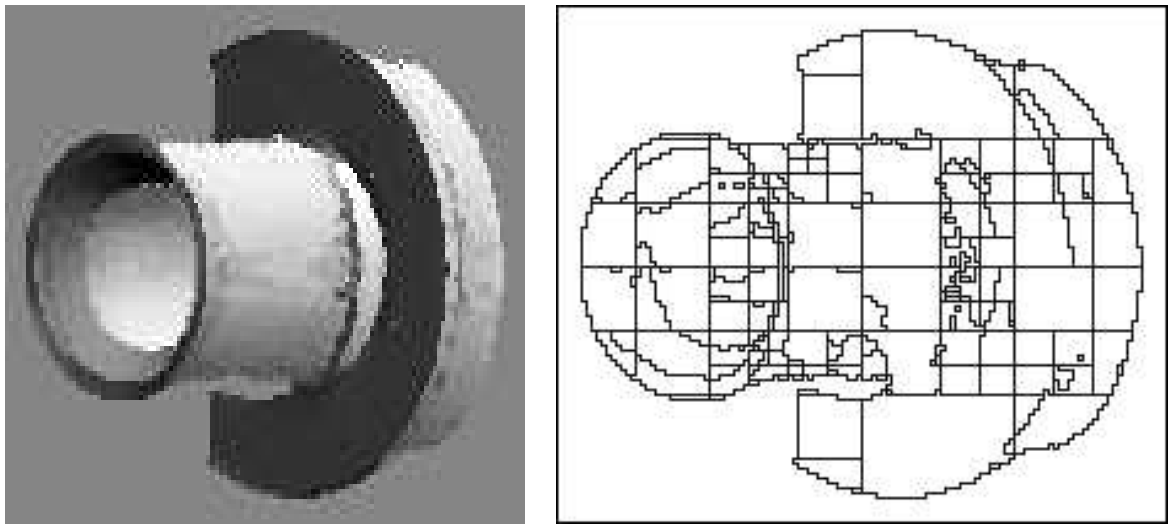


Figure 8.36 Twenty segments that have highest probability in Θ_{568}. There were 949 events in the final cover, with 35 ground events. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.



(a) (b) (c)

Figure 8.37 (a) The rendering of the data set; (b) the set of regions used in a preprocessing step; (c) the set of regions, $\mathcal{R}$



(a) (b)

Figure 8.38 (a) The rendering of the data set; (b) the set of regions, $\mathcal{R}$

(a) (b)

Figure 8.39 Initial regions that were selected for the quadric TSS experiments: (a) R_{783}, (b) R_{181}

this section), shown in Figure 8.37.(b), and iteratively grouping regions that have very high membership probability under the planar model, resulting in the regions shown in Figure 8.37.(c). The quadric experiments use the IE-independent model, since the membership probabilities become extremely small for quadrics when the regions are distant. The results are shown in Figures 8.40 and 8.41. The leading segment in Figure 8.40 is the one we would judge to be correct, and other reasonable alternatives are given. In Figure 8.41, many regions are not homogeneous, and the resulting segments crudely represent the cylindrical surface.

8.4 SSS Representations and Segmentations

In this section we present experimental results which show SSS representations and individual segmentations on the images that were used in Section 8.3. Each of the algorithms presented in Section 7.2 is demonstrated here. Some SSS representations were obtained by running GET-TOP-SEGMENTATIONS with $n = 20$. Others were obtained from BEAM-SEARCH-SEGMENTATIONS. The termination criterion given by Proposition 6 was met for every SSS experiment. Finally, some segmentations obtained using GREEDY- SEGMENTATION are presented. All of the experiments in this section use the sum-of-squares observation space.

Figure 8.42 shows an SSS representation for the synthetic image, with $\sigma^2 = 0.1$ and the IE-independent model. This result was obtained using GET-BEST-SEGMENTATIONS. Since the noise level is low, the correct segmentation receives first rank and has a high probability. When the noise level is increased to $\sigma^2 = 1.0$, some reasonable segmentations are obtained, as shown in Figure 8.43. The correct segmentation, however, does not appear in the top twenty.

The next four results are obtained using the IE-dependent model on the synthetic image. Figure 8.44 shows the SSS representation using GET-BEST-SEGMENTATIONS

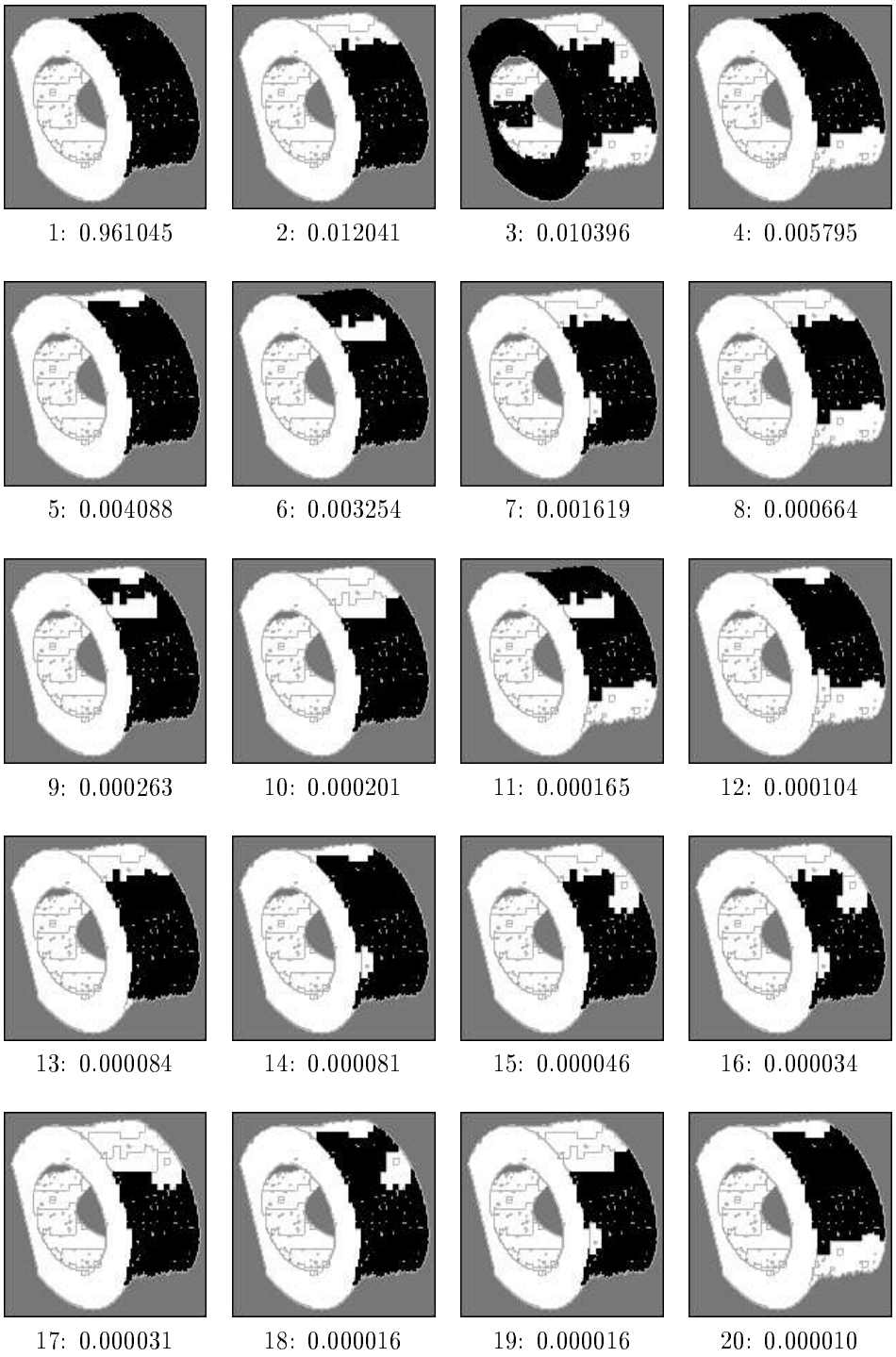


Figure 8.40 Twenty segments that have highest probability in Θ_{783}. There were 150 events in the final cover, with 37 ground events. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.

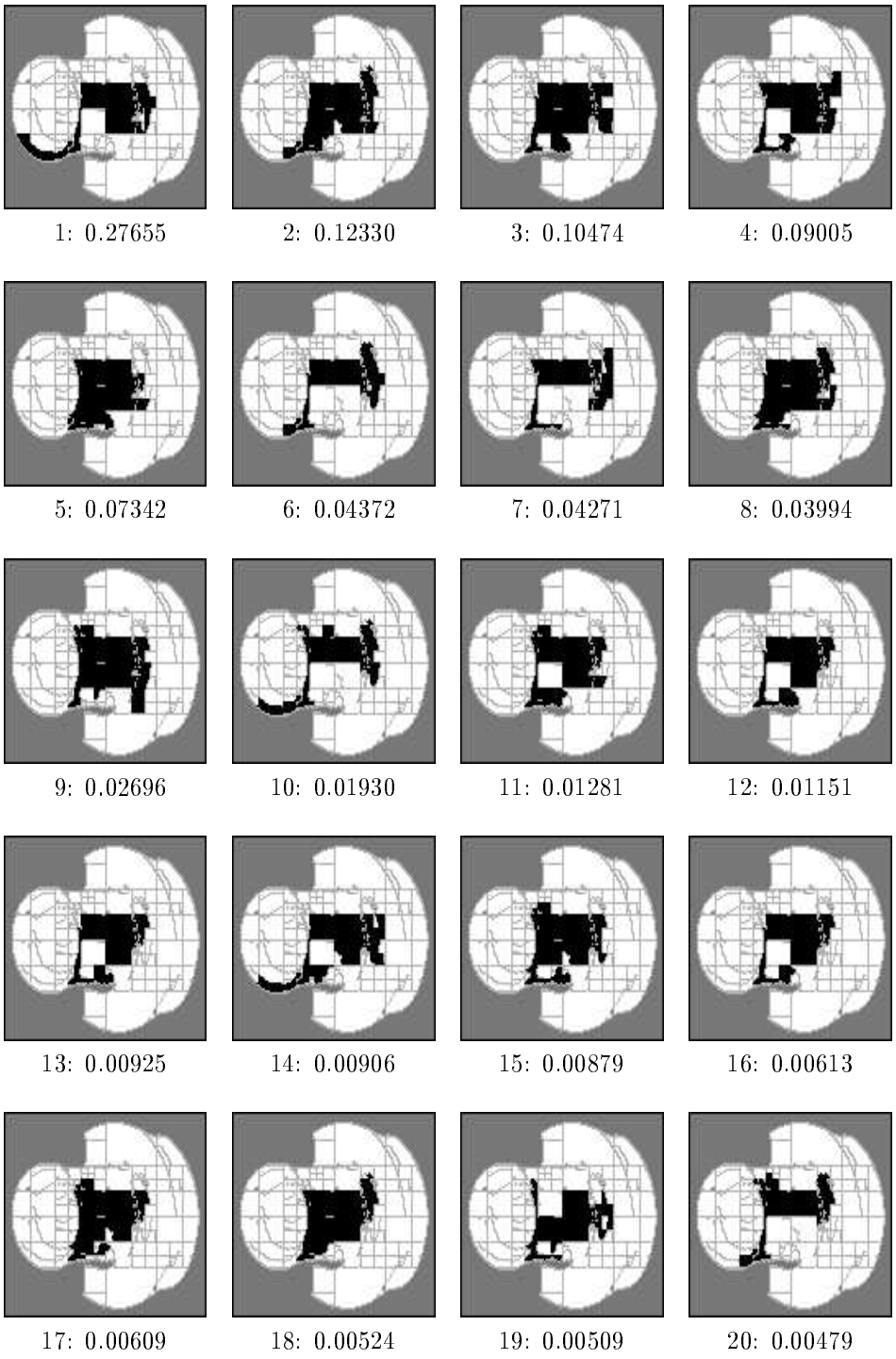


Figure 8.41 Twenty segments that have highest probability in Θ_{181}. There were 716 events in the final cover, with 28 ground events. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.

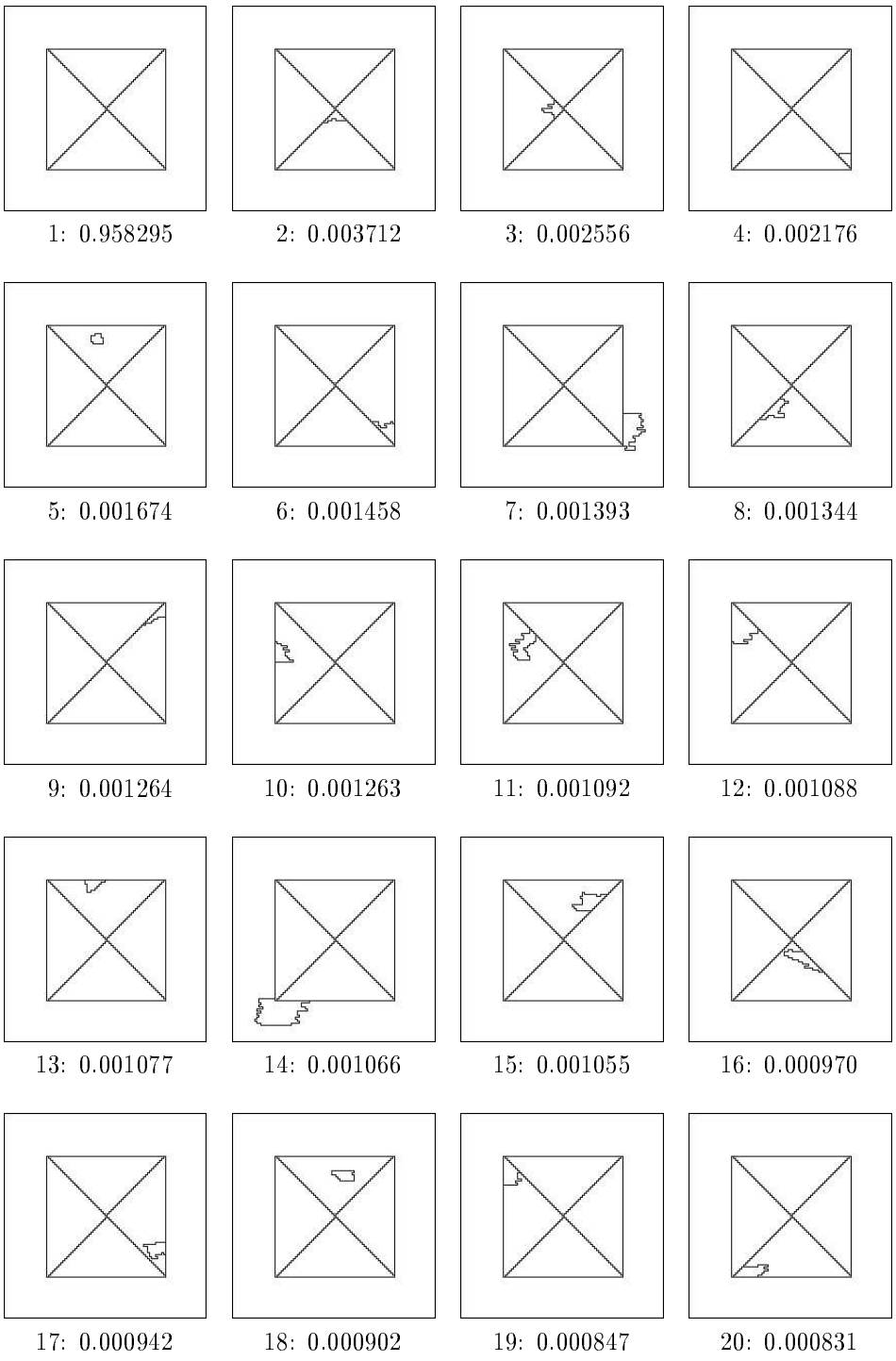


Figure 8.42 Twenty segmentations that have highest probability in Π. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-independent model are in use.

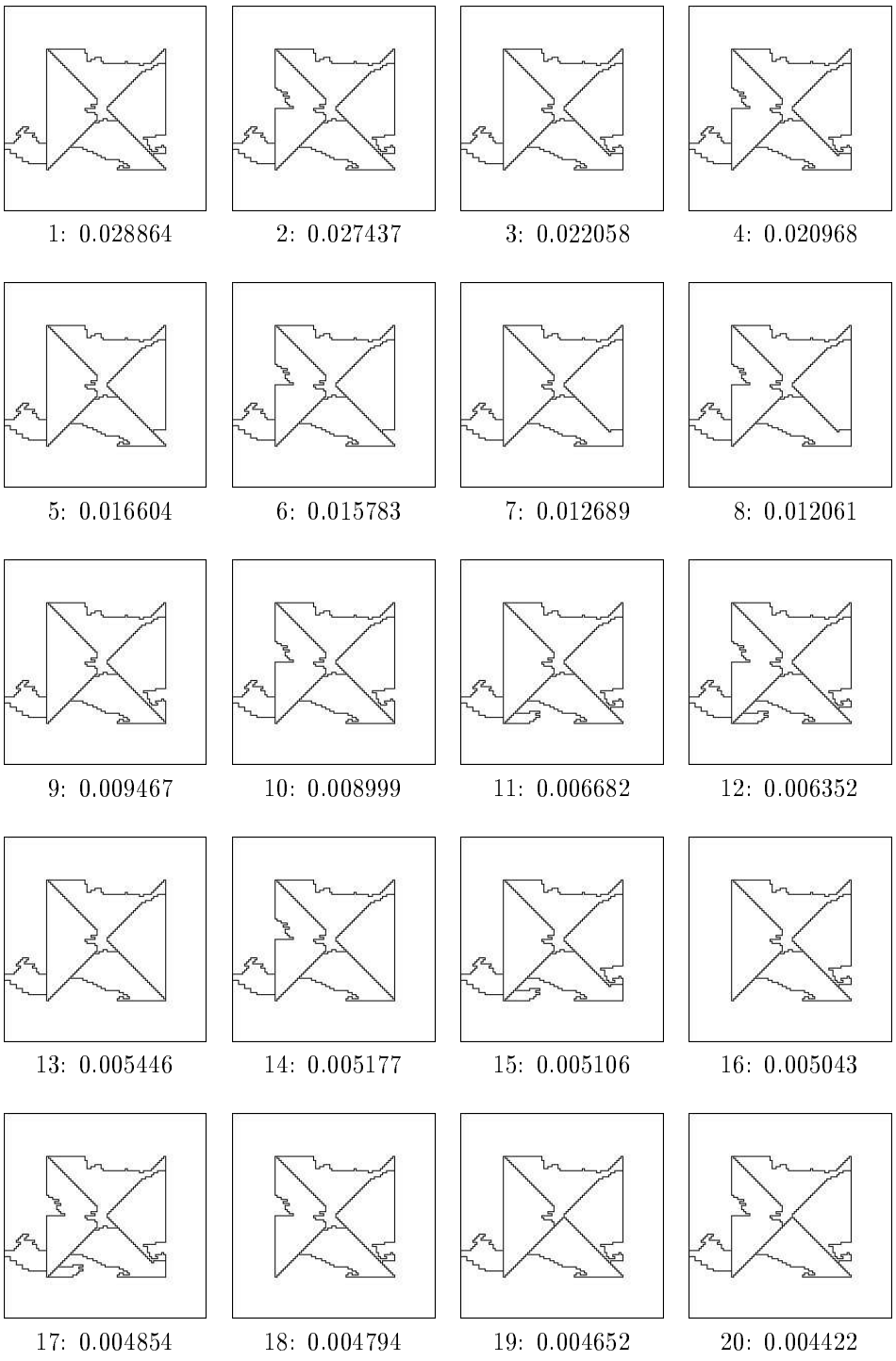


Figure 8.43 Twenty segmentations that have highest probability in Π. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-independent model are in use.

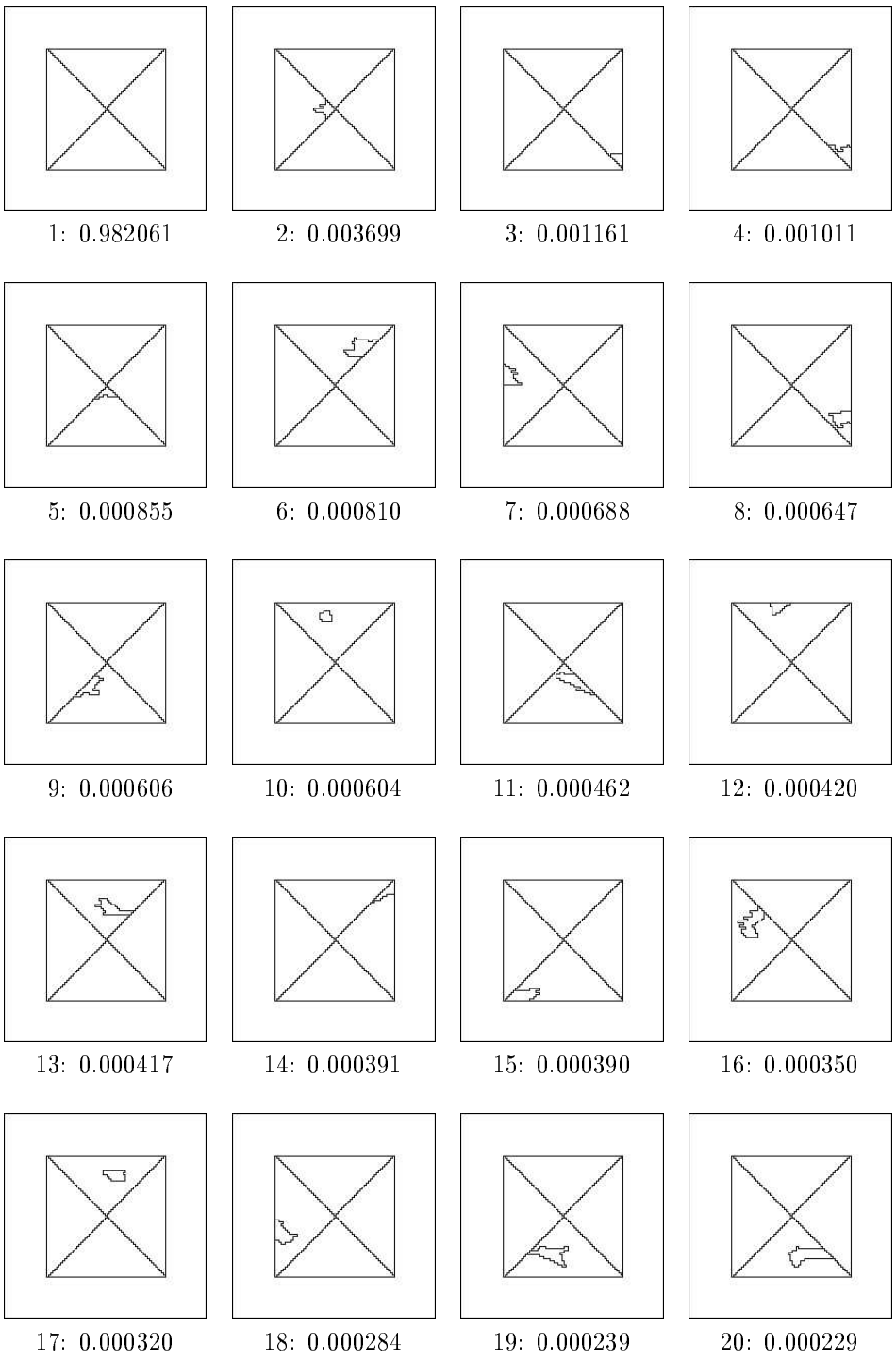


Figure 8.44 Twenty segmentations that have highest probability in Π. In this experiment, $\sigma^2 = 0.1$, and the sum-of-squares observation space and IE-dependent model are in use.

with $\sigma^2 = 0.1$. Again, the correct segmentation is decisively obtained with first rank. With $\sigma^2 = 1.0$, a significantly improved result from Figure 8.43 is obtained, shown in Figure 8.45. The best segmentation is very close to the correct segmentation, and the correct segmentation is ranked 14^{th}. Figures 8.46 and 8.47 show the results from applying BEAM-SEARCH-SEGMENTATIONS (with $\sigma^2 = 1.0$) with $b = 10$ and $b = 3$, respectively. Notice that when $b = 3$, the correct segmentation is no longer represented in the top twenty.

The probabilities of a given segmentation should, in theory, be identical for the past three experiments. There are minor differences in the probabilities, however, due to the crude approximation that was chosen for the Monte-Carlo integration. The inaccuracies are compounded when the membership-probabilities are multiplied to yield the segmentation probability, causing some of the segmentations to slightly change their rank in the final representation.

We next present several SSS results obtained on some of the real range images. Each of these is obtained using BEAM-SEARCH-SEGMENTATIONS with $b = 5$. Figures 8.48 and 8.49 show distributions of segmentations, using the IE-independent and IE-dependent models, respectively. Figure 8.50 shows another SSS result, using the IE-dependent model. Due to the size of the figures, and the number of small regions, the segmentations in these results appear quite similar. There are, however, minor variations in the segmentations, and each is a reasonable alternative.

We conclude this section with some segmentations obtained using the algorithm that obtains single segmentations, GREEDY-SEGMENTATION. Figure 8.51 shows some segmentations obtained for the synthetic image. Some segmentations for the real range data are shown in 8.52.

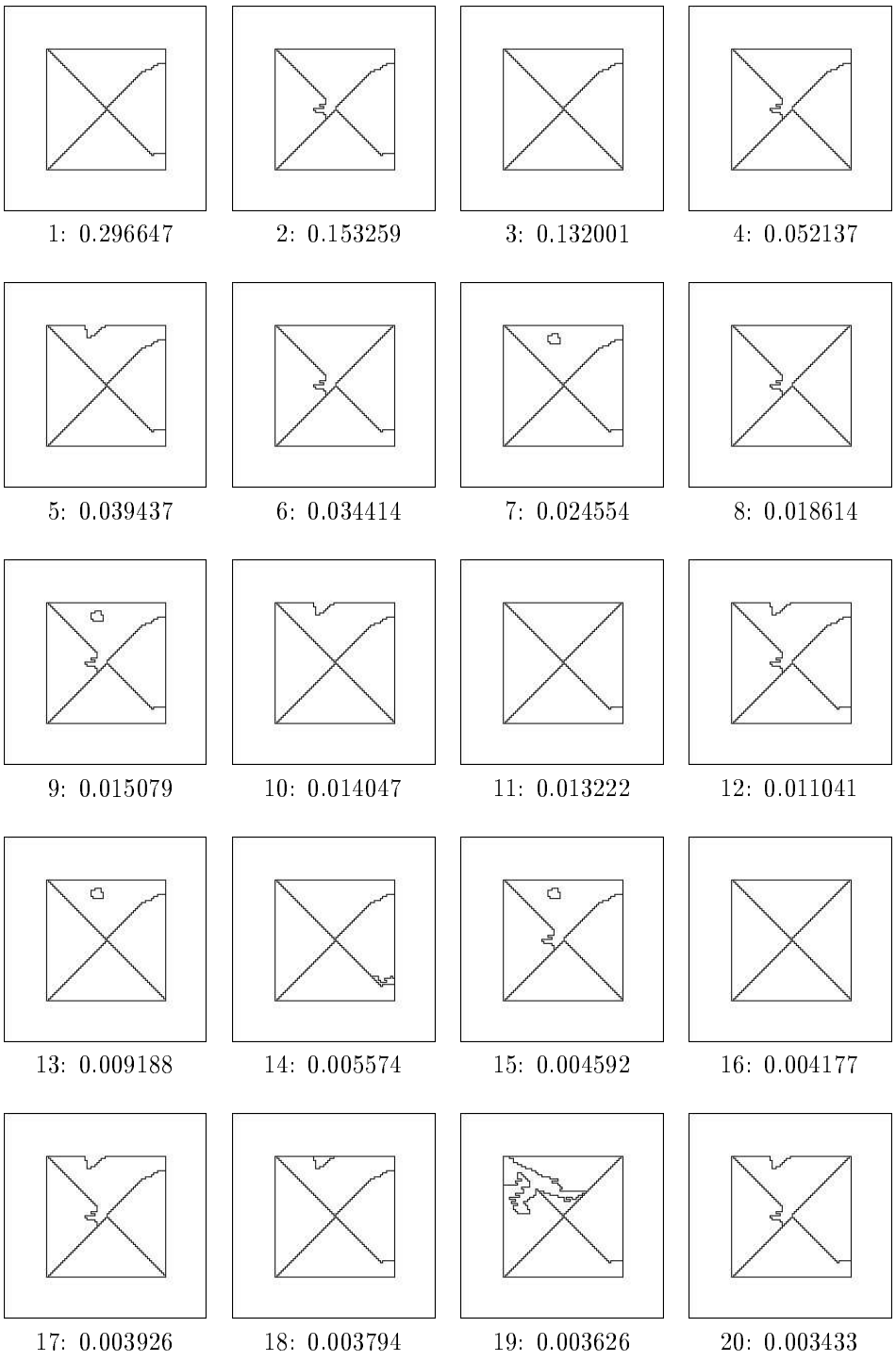


Figure 8.45 Twenty segmentations that have highest probability in Π. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-dependent model are in use.

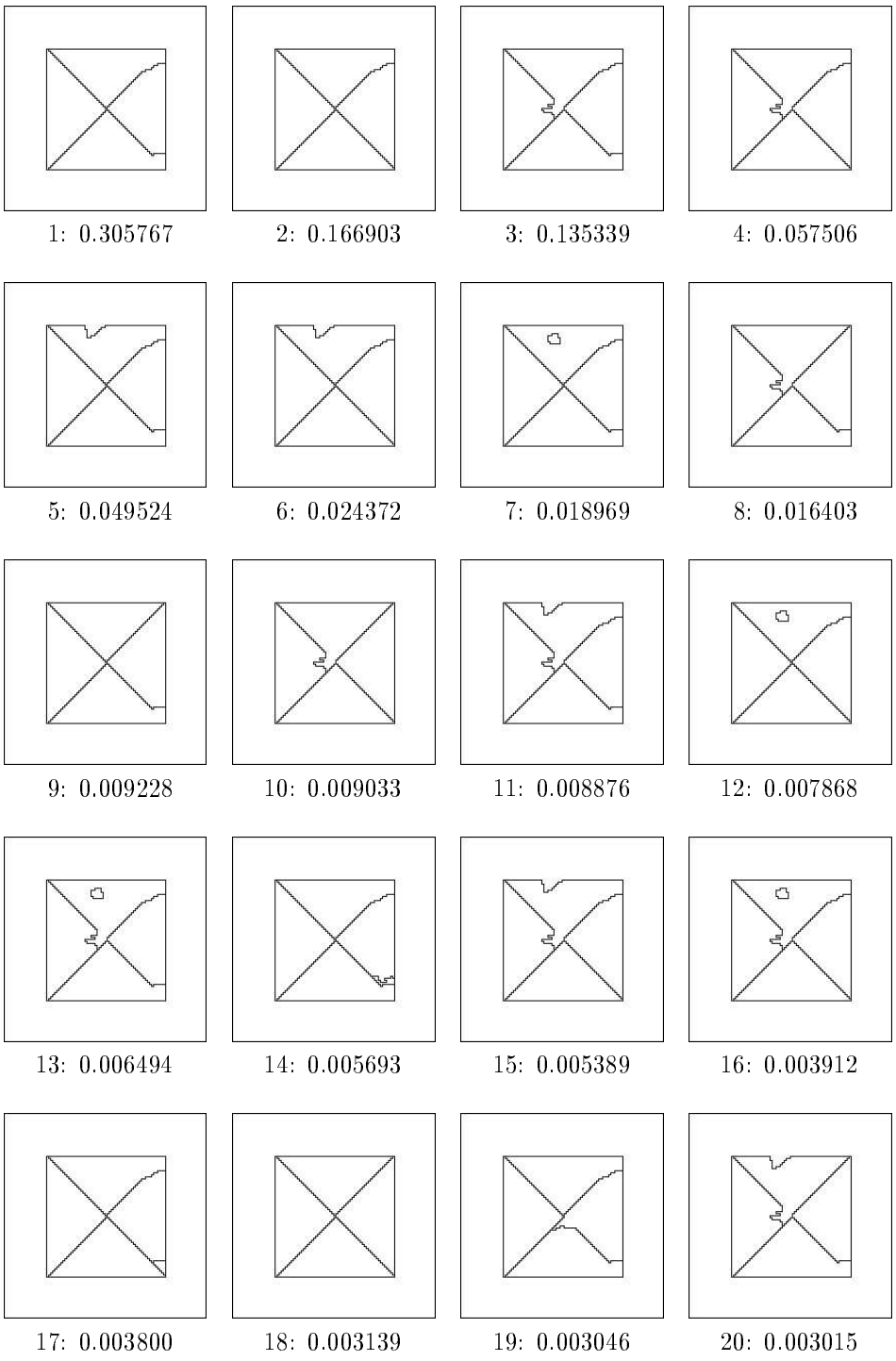


Figure 8.46 Twenty segmentations obtained from a beam-search with $b = 10$. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-dependent model are in use.

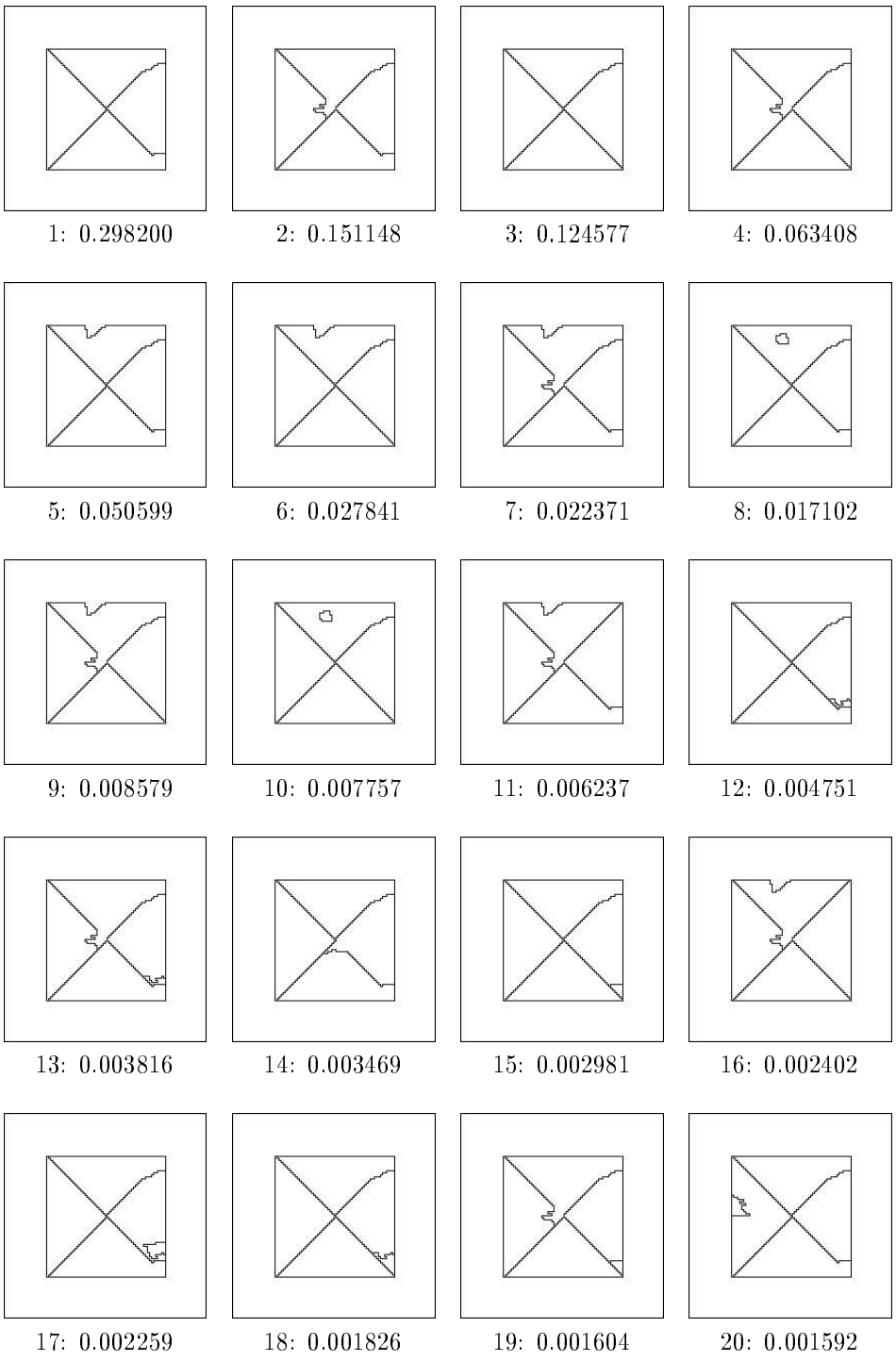


Figure 8.47 Twenty segmentations obtained from a beam-search with $b = 3$. In this experiment, $\sigma^2 = 1.0$, and the sum-of-squares observation space and IE-dependent model are in use.

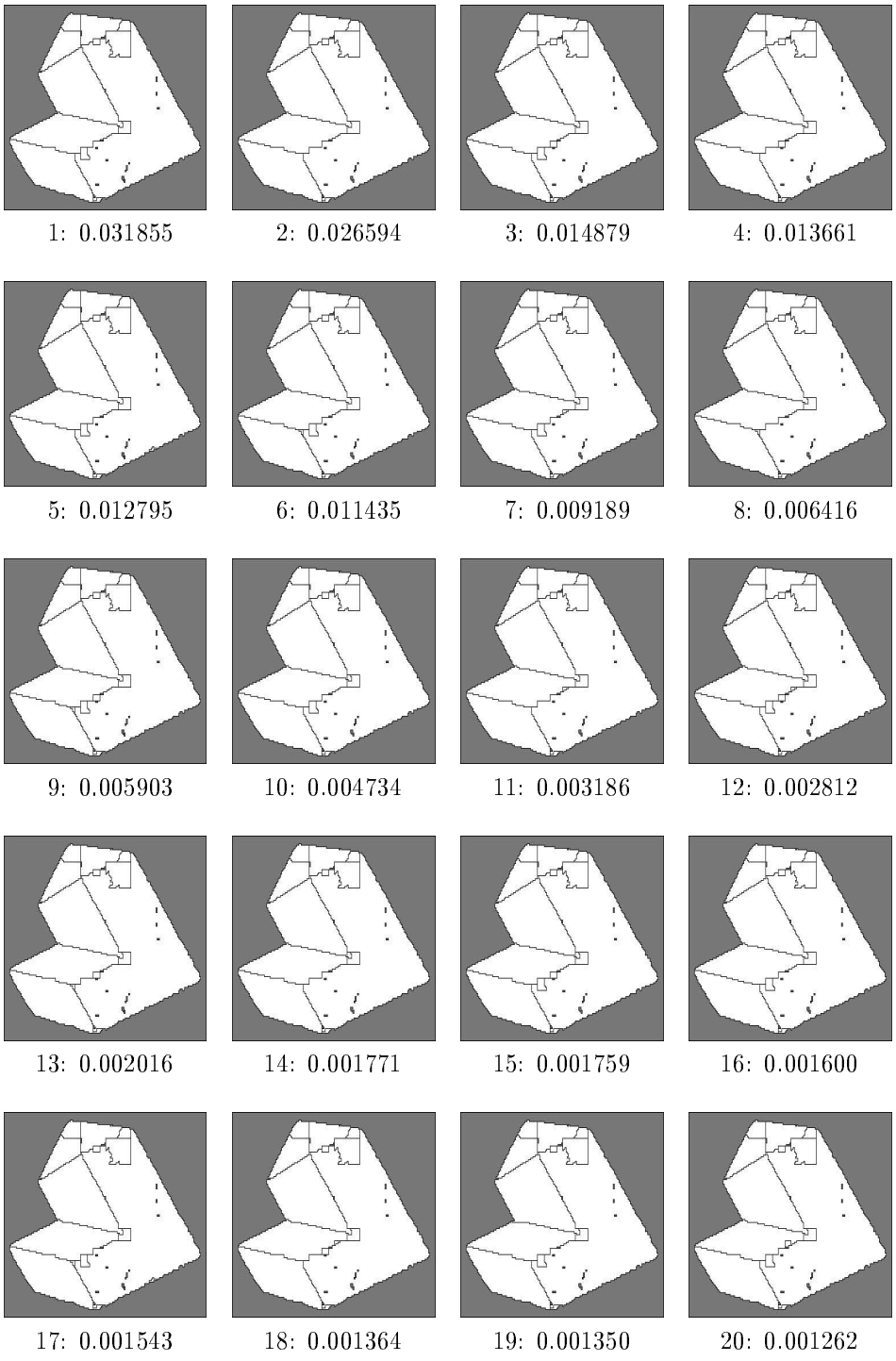


Figure 8.48 Twenty segmentations obtained from a beam-search with $b = 5$. In this experiment, the sum-of-squares observation space and IE-independent model are in use.

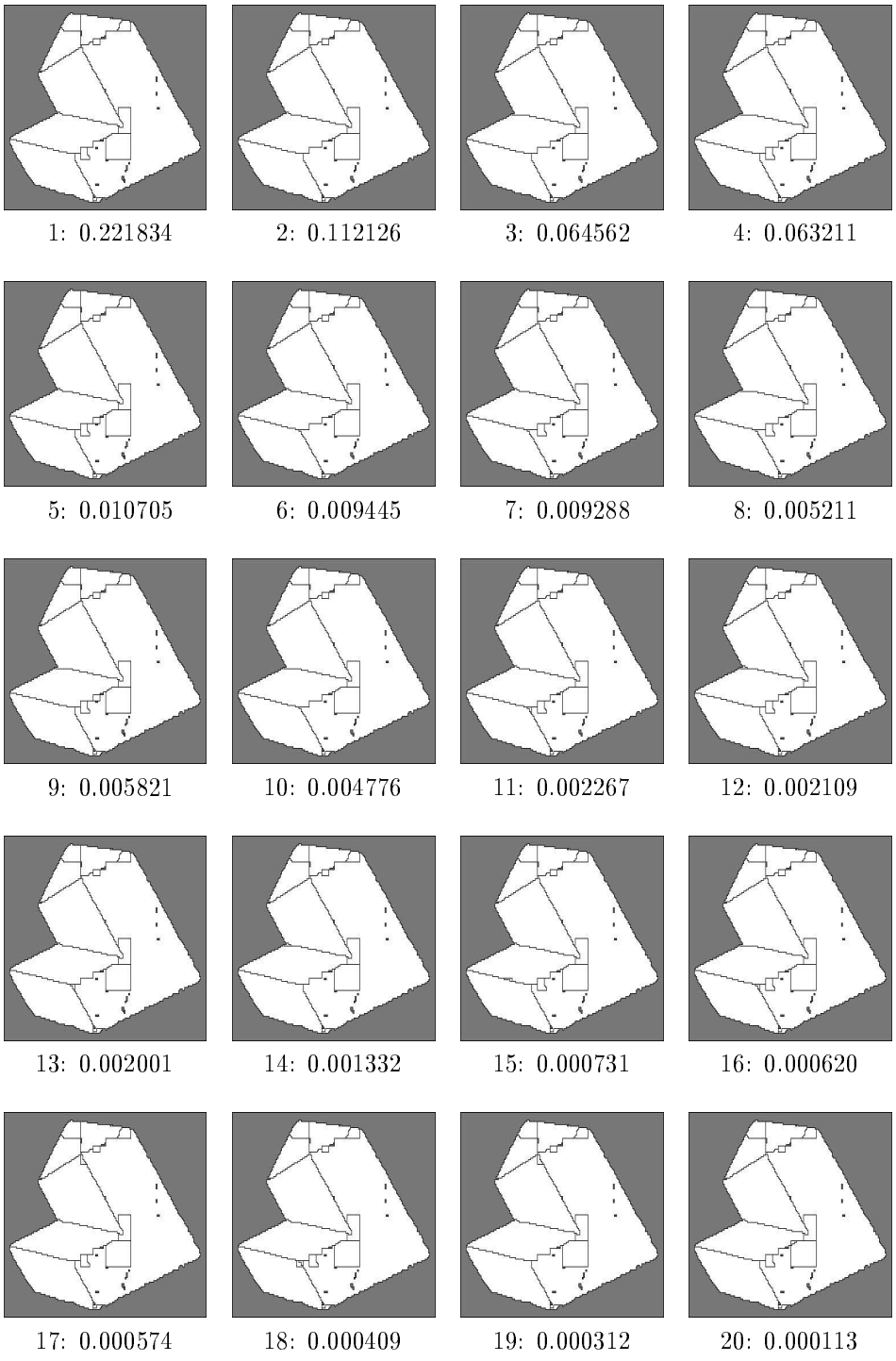


Figure 8.49 Twenty segmentations obtained from a beam-search with $b = 5$. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.

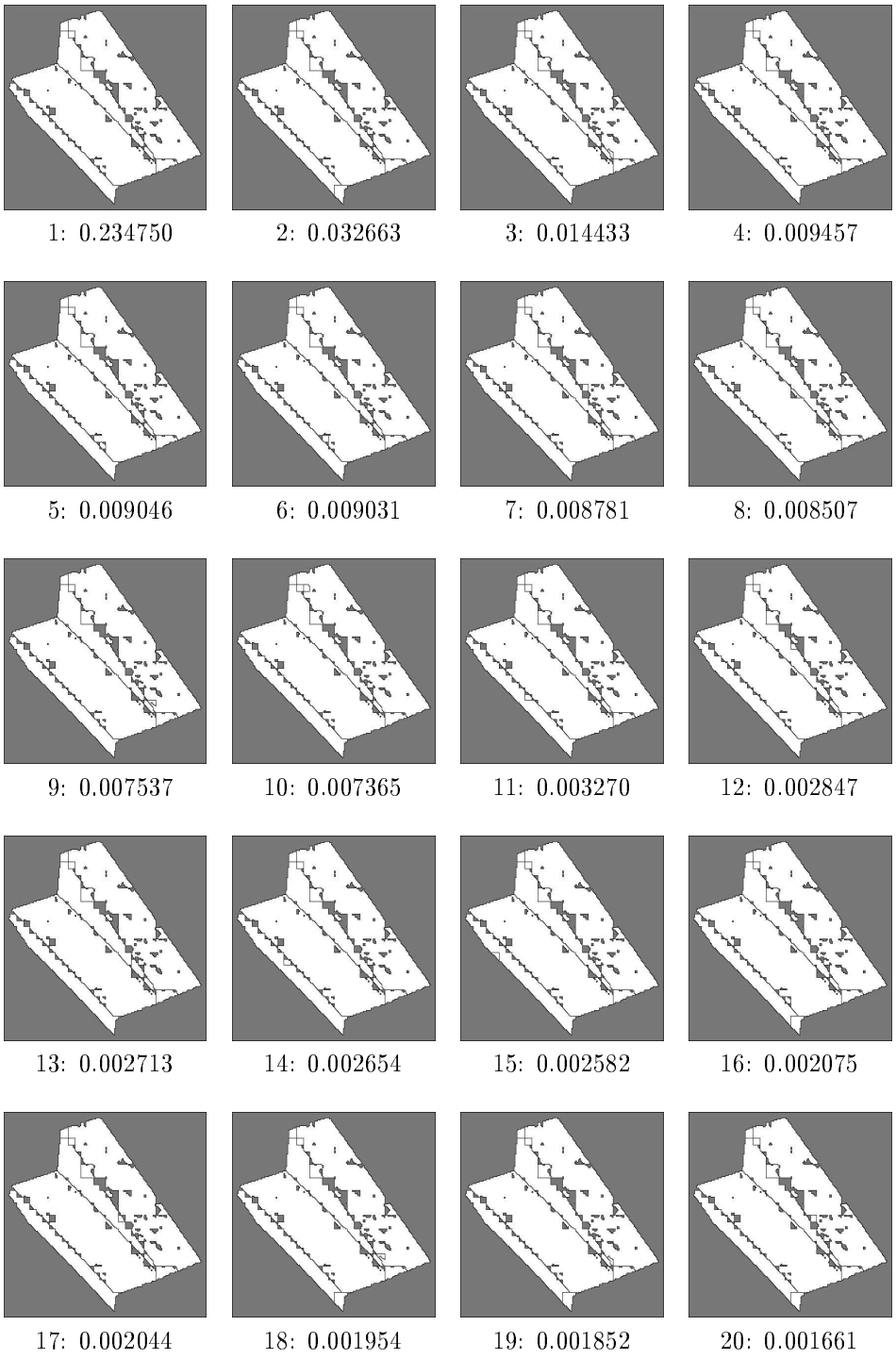


Figure 8.50 Twenty segmentations obtained from a beam-search with $b = 5$. In this experiment, the sum-of-squares observation space and IE-dependent model are in use.

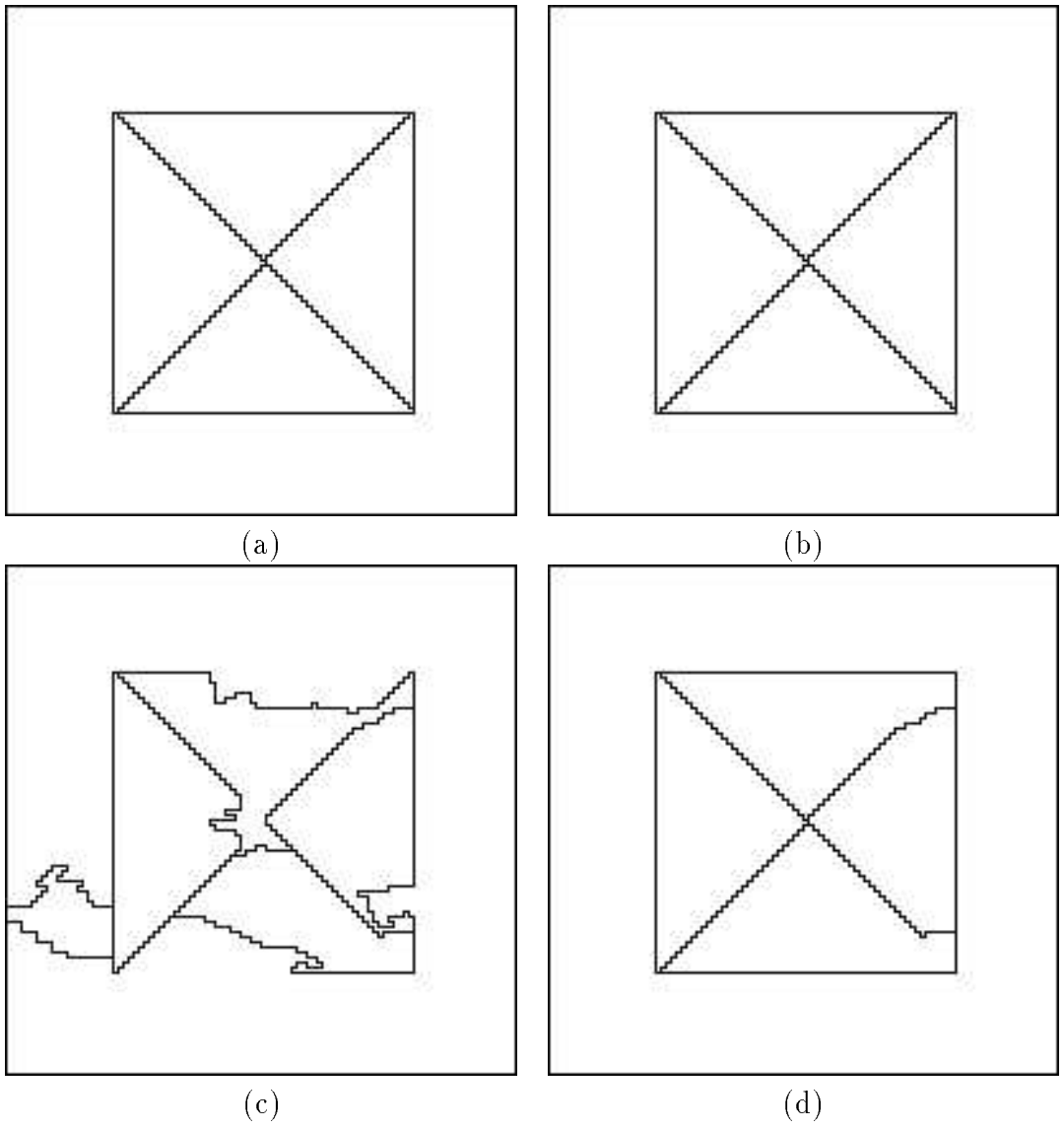


Figure 8.51 Greedy segmentations: (a) with $\sigma^2 = 0.1$ and the IE-independent model; (b) with $\sigma^2 = 0.1$ and the IE-dependent model; (c) with $\sigma^2 = 1.0$ and the IE-independent model; (d) with $\sigma^2 = 1.0$ and the IE-dependent model

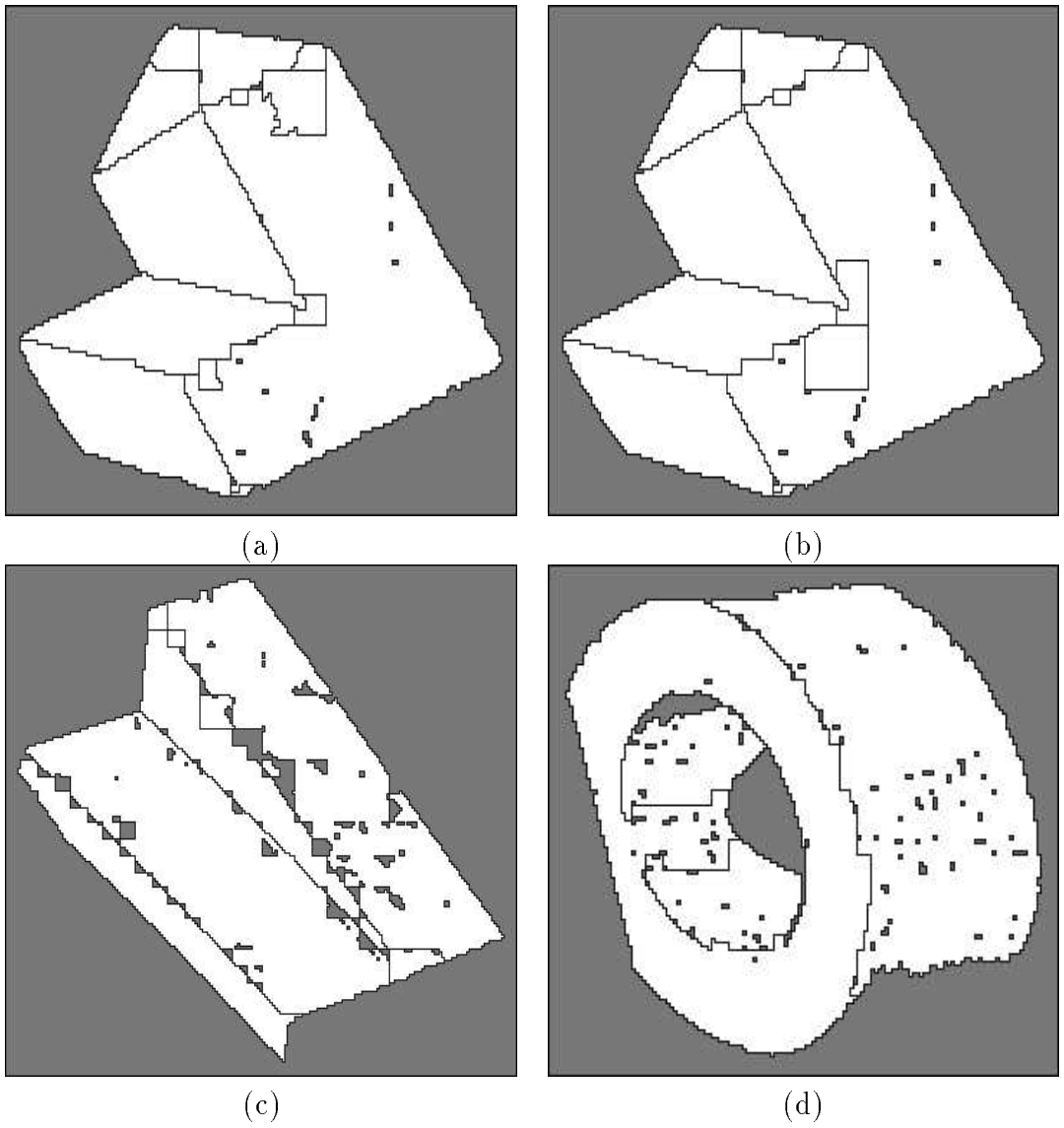


Figure 8.52 Greedy segmentations: (a) using the IE-independent model; (b) using the IE-dependent model; (c) using the IE-dependent model; (d) using the IE-dependent model.

CHAPTER 9

DISCUSSION

This chapter concludes the thesis by reviewing the concepts and objectives presented, and discussing some future research potentials. Section 9.1 provides a brief summary of the thesis and how the originally stated objectives have been satisfied. In Section 9.2, some areas for future research are discussed. Finally, in Section 9.3, a few concluding remarks are made.

9.1 Review

We have presented a general Bayesian framework for representing and considering probability distributions over spaces of segments and segmentations. A concise representation of events in the segment sample space (TSS) was developed through the use of the $\tau(I, E)$ representation. Through the use of the refinement operation, representations of the TSS and SSS for a particular image can be iteratively constructed. To build this representation, we had to determine the probabilities of the refined events, as well as a sequence of refinement operations that efficiently yield a TSS or SSS representation.

Since we have developed a probabilistic approach, the probability assignments made at each refinement are extremely important, causing our focus to shift to the membership probability. We presented a general Bayesian model for making these probability

assignments, expressed in terms of a general parameter space and observation space. This model was applied to implicit polynomial surfaces in which the parameter space represented the space of implicit surfaces of some fixed degree, and the observation space represented functions of point-to-surface displacements. This model required integration over a half-hypersphere parameter space. A Monte Carlo-based scheme was employed to compute the membership probability, since the complexity of the method is independent of dimension.

Since the combinatorics of the TSS and SSS spaces can be unmanageable, several algorithms were developed to select the refinements and obtain representations with moderate computational effort. Algorithms were presented that are guaranteed to produce a list of the best n segments or segmentations and their corresponding probabilities. Experiments were performed on range data using these algorithms, with the Bayesian membership model applied to planar and quadric surfaces.

We now discuss how the objectives stated in Section 1.2 have been achieved by this research.

To develop a system capable of representing any number of alternative segments or segmentations that have corresponding probabilities.

This goal has been the primary focus of our approach and has been demonstrated in theory in Chapters 3-7, and with experiments in Chapter 8. The TSS and SSS representations, discussed in Chapter 3, permit the consideration of multiple segments and segmentations. Also, consistent probabilities are defined for these representations, and in Chapter 4, a general statistical model was introduced which derived these probabilities from image models. With the exception of GREEDY-SEGMENTATION, the algorithms in Chapter 7 allow the specification of any number of segments or segmentations to be represented.

To derive the framework completely from underlying statistical models, which can be experimentally determined.

The Bayesian membership model is expressed in terms of statistical components: the parameter space, observation space, prior model, and the degradation model. The application in Chapter 5 presented a degradation density that could be estimated from experimentation. The parts of our work that may appear arbitrary are the prior model and the prior membership probability. It is important to note, however, that these become dominant only in the presence of little information, and if no information is truly available, then is seems natural to resort to some uniform, noninformative assignments.

To develop a framework capable of handling complex statistical image models (not necessarily those restricted to local dependencies).

This was demonstrated by the application of the Bayesian membership model to implicit polynomial surfaces, in Chapter 5. These implicit surfaces were combined with the degradation model to provide a useful and challenging statistical model. The surface model encodes dependencies that are not necessarily local. The experiments in Chapter 8 demonstrated how this statistical model can be used in practice for image segmentation applications, with reasonable computational performance.

To build a system capable of estimating the amount of information present in the image under a particular statistical image model.

This was a natural consequence of the ability to represent a distribution of segments or segmentations. We can consider an entropy measure over a set of ground segments, ground segmentations, or a binary membership event. If the Bayesian membership model

is applied to other image models, the information content induced by the different models can be compared.

To develop a framework that applies to the most general images and models possible.

We have been as general as possible, but restricted some of the presentation for notational convenience. The concepts formulated in Chapter 3 apply to any segmentation into connected regions, with any adjacency relation between the image elements. The membership probability is expressed in terms of a general piece of evidence, e, until Chapter 4. The algorithms presented in Chapter 7 assume that some method of computing membership probability has been provided, and apply at the same level of generality as Chapter 3.

The Bayesian membership model was presented in a very general form, in terms of image elements, the parameter space, the observation space, the prior density, and the observation density. IE-independent, IE-dependent, and multiple independent models were considered. Variations for discrete-valued variables and the Dirac delta-function approximation were also given.

In Chapters 5 and 6, an application of the Bayesian membership model was presented for implicit polynomial models. This encompasses a large class of surface models, although higher-degree models are limited by computational expense. The dimension of the integration increases (although the number of Monte-Carlo iterations remains fixed), and the size of the matrices in the quadratic form grows quadratically with the number of basis functions.

To develop a framework that readily supports extensions to incorporate higher-level models.

As stated in Chapter 1, the Bayesian framework provides a natural way to combine evidence from multiple models, and this was demonstrated by the membership model in Chapter 4. Also, since an entire distribution of segments and segmentations is maintained, it should be possible to construct other models, which attempt to prune the alternatives. The next section discusses an extension that considers segment-interdependent evidence.

To explain the implications of computationally based simplifications and, when possible, to allow monotonic improvement of accuracy by increasing computation.

This goal was achieved in several instances. The performance of the basic Monte-Carlo procedure degraded linearly with the number of points. The accuracy of the membership probability computation can be arbitrarily improved, by expending extra computational cost. The Dirac delta-function approximation allowed great computational savings by making the assumption that the parameter value for one of the regions can be reliably estimated. This is an efficiency-based assumption, and in general the Monte-Carlo method always be used at the cost of increased computation.

With the exception of GREEDY-SEGMENTATION, the algorithms in Chapter 7 have options that allow monotonic increase in performance with increase of computation. First consider GET-TOP-SEGMENTS (or GET-TOP-SEGMENTS-DEP). If the termination condition, Proposition 3, cannot be affordably reached, there is a computational tradeoff between the quality of the obtained list of segments and the expense of obtaining the list. A list of n segments can be obtained quickly, and as other ground events are determined, this list monotonically improves. When a new ground event is encountered, if its probability is greater than the smallest of the n best segments, it will be inserted in the queue, C_g. The previous n^{th} best segment will become the $(n+1)^{th}$ best segment, and no longer be considered. Hence, only improvements can be made to

the list, and eventually the optimal list is obtained. These same observations hold for BEAM-SEARCH-SEGMENTATIONS and GET-TOP-SEGMENTATIONS. The performance of the beam-search method can also be improved monotonically by increasing the beam width, at the expense of increased computation. Any segmentation considered with a beam size of b will also be considered with a beam size of $b+1$.

The approximations and simplifications made to the distance function and sum-of-squares in Section 5.4.3 could be greatly improved. However, we believe it would be difficult to utilize improved estimates efficiently since this computation is only a single iteration in the integration procedure. Removing the simplifications may not, in our formulation of the problem, be computationally manageable. One could, however, use some iterative method to compute the correct sum-of-squares displacements [111], and the results would improve at the expense of increased computation.

To avoid prior specification or estimation of the number of segments in order to perform segmentation.

The number of segments was not required to construct a TSS or SSS representation. In fact, in a resulting SSS distribution, we have seen that varying numbers of segments can appear in the represented segmentations. Although the number of segments is not specified, some indirect bias can be introduced by choosing different membership prior probabilities. A lower prior membership probability induces smaller segments, and a higher prior membership probability induces larger segments.

To design efficient algorithms that are capable of handling large images with realistic probability models.

This goal was accomplished with the algorithms presented in Chapter 7. Because of our multiple-segment and multiple-segmentation approach, this goal is challenging. The combinatorics of considering all alternatives can be unmanageable, and by focusing on the probabilistically interesting alternatives only, some efficient algorithms were obtained. The experiments in Chapter 8 demonstrated their utility for real and synthetic images.

To design algorithms and representations that allow for straightforward parallelization.

The two major components of computation that were presented are parallelizable. The Monte-Carlo integration method is trivially parallelizable since the samples are independent and the summation can be decomposed. Also, the algorithms presented in Chapter 7 can be significantly parallelized. Each refinement operation can in theory spawn two independent computation processes. The computations resulting from performing refinements on two refined events from the same refinement operation are independent. For efficiency, some concern must, however, be given for attempting to represent the priority queues of nonground events in a TSS or the SSS.

9.2 Prospects

The purpose of this section is to provide some areas for future investigation based on our framework. Section 9.2.1 indicates how the membership probability model can also be used for agglomerative clustering. Section 9.2.2 describes a generalization of the membership model presented in Chapter 4 to incorporate estimation into the general model definitions. Section 9.2.3 mentions some issues that arise when modeling dependencies between segments.

9.2.1 Agglomerative clustering with the Bayesian membership model

In this section we present general expressions that could be used to apply the Bayesian membership model to the agglomerative clustering approach to segmentation. The approach described here is closely related to the work of Silverman and Cooper for explicit polynomial surfaces with intensity images [43], and hence is treated as a continuation of the discussion at the end of Section 2.1.

Recall the required steps in the agglomerative clustering algorithm from Section 2.1. The first step builds the initial clusters and performs parameter estimation. In our approach, the set of regions, $\mathcal{R}$, is used as the initial set of clusters, and estimation is not necessary.

The second step considered all possible pairs of adjacent clusters to merge into a single cluster. With the membership model, this involves computing the membership probability for all possible adjacent pairs of regions.

The third step chose the most likely pair to merge. In our case this is the pair with the highest membership probability. With a given set of regions, $\mathcal{R}$, the best pair to merge, R_{m_1}, R_{m_2}, satisfies

$$P(\tau(\{R_{m_1}, R_{m_2}\}, \emptyset)|\mathbf{y_{m_1}}, \mathbf{y_{m_2}}) \leq P(\tau(\{R_i, R_j\}, \emptyset)|\mathbf{y_i}, \mathbf{y_j}) \quad \forall R_i, R_j \in \mathcal{R} \ \ (R_i \neq R_j). \tag{9.1}$$

The fourth step tested the termination criterion. If another iteration occurs, a new set of regions is defined by

$$\mathcal{R}' = [\mathcal{R} \ \cup \ \{R_{m_1} \cup R_{m_2}\}] \ - \ R_{m_1} \ - \ R_{m_2}. \tag{9.2}$$

The set $\mathcal{R}'$ is used in the first step of the next iteration, and membership probabilities that reference $R_{m_1} \cup R_{m_2}$ are computed.

We terminate when

$$P(\tau(\{R_{m_1}, R_{m_2}\}, \emptyset)|\mathbf{y_{m_1}}, \mathbf{y_{m_2}}) < \frac{1}{2}. \tag{9.3}$$

This occurs when even the best merge possible is not likely to produce a truly homogeneous region. Note that this criterion can be biased by varying the prior membership probability. If a high prior membership probability is assigned, then more merges will occur and there will be fewer segments. If a low probability is assigned, there will be fewer merges, resulting in more segments.

9.2.2 Generalized estimation/integration membership model

As discussed in Chapter 2, nearly all segmentation algorithms rely on some sort of parameter estimation. The model presented in Chapter 4 resulted in integration over the parameter space for the purpose of segmentation. It is possible, however, to unify estimation and integration into a single membership model.

In this thesis, there have been instances in which estimation and integration have been combined. Chapter 5 involved a mixture of estimation and integration, with the nonhomogeneous parameter space. The issue came about by deciding whether to estimate the extra parameter d, or to treat is as part of the parameter space. Further, the delta-function approximation, presented in Section 4.5, estimated all parameters for a region. The purpose of this section is to briefly discuss a generalization of model developed in Chapter 4 that easily incorporates these considerations.

We begin by generalizing the parameter space to include variables that can be either estimated, or included in the integrations. We consider a vector of random variables, $\mathbf{G_k}$, to be a *generalized parameter space* associated with some region R_k. This space has a similar interpretation as the parameter space in Chapter 4. We are also given an observation space, a degradation model, and a prior model.

The distinction between a generalized parameter space, $\mathbf{G_k}$, and a parameter space, $\mathbf{U_k}$ is that *all* the parameters in $\mathbf{U_k}$ are integrated in a membership probability computation; however, *some* of the parameters in $\mathbf{G_k}$ are treated as random variables, and the remaining ones are estimated.

This requires dividing $\mathbf{G_k}$ into two orthogonal subspaces, which we denote by $\mathbf{U_k}$ and $\mathbf{V_k}$. The vector $\mathbf{U_k}$ represents the parameter space, as before, and $\mathbf{V_k}$ represents a vector of parameters that must be estimated.

These parameters may appear anywhere in the functional expressions for the degradation and prior densities. For the nonhomogeneous parameter space, $\mathbf{V_k}$ is represents the scalar, d. The remaining parameters comprise $\mathbf{U_k}$. Under the delta function model all parameters belong to $\mathbf{V_k}$.

A new example can be considered by extending the example presented at the end of Chapter 4. Consider the ratio obtained in (4.111). In this example we can consider both u and σ to represent the generalized parameter space, $\mathbf{G_k}$. In (4.111), u represents the parameter space, $\mathbf{U_k}$, and σ represents $\mathbf{V_k}$. We can alternatively consider both u and σ to comprise $\mathbf{U_k}$. Since σ is treated as a random variable, we represent it with some pdf, $p(\sigma)$. This results in the following alternative expression of the evidence-based ratio:

$$\frac{\left[\int_0^\infty \int_0^1 exp\left[\frac{N_\rho(y_\rho - u)^2}{2\sigma^2}\right] p(\sigma) du d\sigma\right] \left[\int_0^\infty \int_0^1 exp\left[\frac{N_i(y_i - u)^2}{2\sigma^2}\right] p(\sigma) du d\sigma\right]}{\int_0^\infty \int_0^1 exp\left[\frac{N_\rho(y_\rho - u)^2}{2\sigma^2}\right] exp\left[\frac{N_i(y_i - u)^2}{2\sigma^2}\right] p(\sigma) du d\sigma} \tag{9.4}$$

The task that remains is to decide, for each region R_k, which portion of $\mathbf{G_k}$ should comprise $\mathbf{U_k}$ and which should comprise $\mathbf{V_k}$. When a parameter can be reliably estimated, it should belong to $\mathbf{V_k}$; otherwise, it should be treated as a random variable in $\mathbf{U_k}$. Formulating some decision criterion for this task, in a particular application, poses an interesting research problem.

9.2.3 Segment interdependent models

When multiple TSS representations are constructed, we have assumed that segments can be combined to yield independent probabilities. By modeling dependencies between segments, a higher-level class of models can be considered. Given some distribution or database of possible objects appearing in a scene, some combinations of segments may be more favorable than others. In addition, since multiple segments are considered, information from image elements at the boundaries between segments can be considered. Hence, this is a natural place to incorporate edge-based models.

As an indication of how to proceed with this task, we present several Bayesian expressions relating to some general piece of evidence. These expressions are at the same level of abstraction as those of Chapter 3, where the membership probability is first introduced. We leave the investigation of feasibly incorporating statistical models into these expressions, analogous to the work of Chapters 4-6, for future research.

Recall from Section 3.5.2 that the SSS probability is determined by multiplying the probabilities of each of the individual segments. This is equivalent to assuming statistical independence between the segments. Suppose we consider the two probabilities to be dependent. Abstractly, for two segments $T_1 \in \Theta_1$ and $T_2 \in \Theta_2$, this is represented as

$$P(\{T_1\}, \{T_2\}) = P(\{T_1\})P(\{T_2\}|\{T_1\}). \tag{9.5}$$

Consider some general piece of evidence, e_{12}, that depends on both T_1 and T_2, but on neither individually (e.g., this could represent some edge-based model, in which the information is derived from the boundary between T_1 and T_2). Consider the expression above in the presence of this evidence,

$$P(\{T_1\}, \{T_2\}|e_{12}) = P(\{T_1\}|e_{12})P(\{T_2\}|\{T_1\}, e_{12}). \tag{9.6}$$

Since e_{12} depends on neither T_1 nor T_2 individually, we will have $P(\{T_1\}|e_{12}) = P(\{T_1\})$ and $P(\{T_2\}|e_{12}) = P(\{T_2\})$.

The second part of the product above can be expanded with Bayes' rule to obtain

$$P(\{T_2\}|\{T_1\}, e_{12}) = \frac{P(e_{12}|\{T_1\}, \{T_2\})P(\{T_2\}|\{T_1\})}{P(e_{12}|\{T_1\})}. \tag{9.7}$$

By substitution, we obtain

$$P(\{T_1\}, \{T_2\}) = P(\{T_1\})P(\{T_2\}) \left[\frac{P(e_{12}|\{T_1\}, \{T_2\})}{P(e_{12}|\{T_1\})}\right]. \tag{9.8}$$

Both $P(\{T_1\})$ and $P(\{T_2\})$ are given by the probability maps on $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. The final part of the product above can be considered as another evidence-based ratio. The denominator of this ratio is a normalizing factor and can be expanded into

$$P(e_{12}) = \sum_{T_2 \in \Theta_2} P(e_{12}|\{T_1\}, \{T_2\})P(\{T_2\}|\{T_1\}) = \sum_{T_2 \in \Theta_2} P(e_{12}|\{T_1\}, \{T_2\})P(\{T_2\}). \tag{9.9}$$

9.3 General Conclusions

From this general framework we conclude that segmentation does not have to be treated as an isolated process with an optimal solution. Instead it can be treated as a package of low-level models that are used to constrain the space of segments and segmentations, resulting in a probability distribution of alternatives. This work has demonstrated that this view of segmentation can be realized in an efficient and sufficiently general computational structure. We hope that this contribution will change some of the focus in segmentation research toward the consideration of distributions of segments and segmentations and stronger Bayesian models, and away from the determination of single, near-optimal segmentations from underconstraining models.

APPENDIX A

MORE ON PARAMETER SPACE INTEGRATION

A.1 Converting the Hypersphere Surface Integral into a Volume Integral

Since Σ^N is a relatively simple manifold, we begin by generalizing techniques from standard vector calculus. Recall a standard form for the surface integral in $\Re^3$ [103],

$$\int_{surface} (\mathbf{v} \cdot \mathbf{n}) d\sigma = \int \int v_1 du_2 \wedge du_3 + v_2 du_3 \wedge du_1 + v_3 du_1 \wedge du_2. \qquad (A.1)$$

Here $\mathbf{v} \equiv [v_1\ v_2\ v_3]$ represents a vector field over $\Re^3$, and $\mathbf{n}$ represents the unit normal to the surface. The u_1, u_2, and u_3 are the point coordinates in $\Re^3$ (the space in which the surface is defined). The $\wedge$ operator is a standard notation from calculus of differential forms used to distinguish the manifold differentials from ordinary volume differentials. One can imagine them as an ordinary product of differentials, but with the distinction that their ordering forces a certain surface orientation (i.e. $du_1 \wedge du_2 = -du_2 \wedge du_1$). This surface integral will be generalized in dimension and then applied to Σ^N in $\Re^{N+1}$.

The equivalent of (A.1) applying to a manifold of dimension $N-1$ in $\Re^N$ is

$$\int_{manifold} (\mathbf{v} \cdot \mathbf{n}) d\sigma = \int \cdots \int \sum_{i=1}^{N} v_i sgn(i, k_1^i, k_2^i, \ldots, k_{N-1}^i)\, du_{k_1^i} \wedge \cdots \wedge du_{k_{N-1}^i}. \qquad (A.2)$$

The notation used above requires some explanation. The product of differentials represents all possible ways to choose $N-1$ coordinate differentials with N variables

(i.e., there are N of them). The subscripts k_j^i represent possible subscripts which index the coordinate variables of $\mathbf{u}$. They are written in this form so that we are not required to give the exact ordering of the product of differentials (we are required only to choose $N-1$ of them each time). A change in the ordering can change the surface orientation, causing the sign of the integration to change. This will become clearer when we perform transformations (A.9). The *sgn* function compensates for sign changes due to the ordering of the indices. If the sign of the permutation of the indices is even, then *sgn* returns 1, otherwise -1. The value, v_i, represents one component of our vector-valued function $\mathbf{v}$.

For our specific application, consider a vector-valued function of $\mathbf{u}$ in $\Re^N$ defined at every $\mathbf{u}$ by

$$\mathbf{v} = h(\mathbf{u})[u_1\ u_2 \ldots\ u_N]. \tag{A.3}$$

This is a scalar function $h(u_1, \ldots, u_N)$ multiplied across a vector consisting of the $\mathbf{u}$ coordinates. Next, consider the equation of Σ^N, given in implicit form,

$$\alpha(u_1, \ldots, u_N) \equiv \sum_{i=1}^{N} u_i^2\ - 1 = 0. \tag{A.4}$$

Its unit normal is

$$\mathbf{n} = \frac{\nabla\alpha}{\|\nabla\alpha\|} = [u_1\ u_2\ \ldots\ u_N]^T. \tag{A.5}$$

It can be seen that the vector-valued function $\mathbf{v}$ is h in the direction of the normal to the Σ^N when $\|\mathbf{u}\| = 1$. The h is defined as the value of the function that is desired at that point on the manifold.

We now want to express the integration on Σ^N by specializing (A.2). First, the left side can be written as

$$\int_{\Sigma^N} (\mathbf{v}\cdot\mathbf{n}) d\sigma = \int h(u_1, \ldots, u_N)[u_1\ \ldots\ u_N][u_1\ \ldots\ u_N]^T d\sigma = \int_{\Sigma^N} h(\mathbf{u}) d\sigma. \tag{A.6}$$

By replacing $d\sigma$ as in (A.2), the integral becomes

$$\int_{\Sigma^N} h(\mathbf{u}) d\sigma = \int \cdots \int h(\mathbf{u}) \sum_{i=1}^{N} u_i sgn(i, k_1^i, k_2^i, \ldots, k_{N-1}^i)\, du_{k_1^i} \wedge \cdots \wedge du_{k_{N-1}^i}. \tag{A.7}$$

Since this integration takes place over Σ^N, it is desirable to reparameterize this integral from a difficult region of integration, to an $(N-1)$-dimensional rectangle. This will cause all of the bounds of integration to be constants, keeping the integrand from growing more complex with each iterated integration.

The hypersphere, Σ^N in $\Re^N$, can be parameterized as [101]

$$\begin{aligned}
u_1 &\equiv g_1(\mathbf{t}) &&= cos(t_1) \\
u_2 &\equiv g_2(\mathbf{t}) &&= sin(t_1)cos(t_2) \\
u_3 &\equiv g_3(\mathbf{t}) &&= sin(t_1)sin(t_2)cos(t_3) \\
&\vdots \\
u_{N-1} &\equiv g_{N-1}(\mathbf{t}) &&= sin(t_1)sin(t_2)sin(t_3)\dots cos(t_{N-1}) \\
u_N &\equiv g_N(\mathbf{t}) &&= sin(t_1)sin(t_2)sin(t_3)\dots sin(t_{N-1}).
\end{aligned} \tag{A.8}$$

For the case of $N = 2$ or $N = 3$, this parameterization specializes to standard polar and spherical coordinates, respectively. For the full hypersphere we have $0 \leq t_j \leq \pi$ for $0 \leq j \leq N-2$ and $0 \leq t_{N-1} \leq 2\pi$. Since we are interestred in only half of the hypersphere, we take $0 \leq t_{N-1} \leq \pi$. These equations can be substituted into (A.7), and some transformation Jacobians are necessary to relate the new parameters to the previous differentials. Each manifold differential is transformed in the sum of (A.2) by

$$du_{k_1^i} \wedge \cdots \wedge du_{k_{N-1}^i} = \frac{\partial(u_{k_1^i}, \dots, u_{k_{N-1}^i})}{\partial(t_1, \dots, t_{N-1})} dt_1 \dots dt_{N-1}. \tag{A.9}$$

The partial functions above represent the determinant of an $N-1 \times N-1$ Jacobian matrix. It has ∇u_j from the equations in (A.8) as each of its rows. Also, from this expression one can see the importance of the permutation sign since a transposition of differentials in the wedge product corresponds to swapping rows in the Jacobian, changing the sign of the determinant.

When the differential transformation is substituted into (A.7), the manifold integral becomes

$$\int \cdots \int h(g_1(\mathbf{t}) \cdots g_N(\mathbf{t})) \sum_{i=1}^{N} g_i(\mathbf{t}) sgn(i, k_1^i, k_2^i, \dots, k_{N-1}^i) \frac{\partial(u_{k_1^i}, \dots, u_{k_{N-1}^i})}{\partial(t_1, \dots, t_{N-1})} dt_1 \dots dt_{N-1}. \tag{A.10}$$

It can be shown that when the determinants are evaluated, and simplifications are made over the sum through trigonometric identities, the resulting integral is significantly simplified to

$$\int_0^{\pi} \cdots \int_0^{\pi} h(g_1(\mathbf{t}), \ldots, g_N(\mathbf{t})) sin^{N-2}(t_1) sin^{N-3}(t_2) \cdots sin^2(t_{N-3}) sin(t_{N-2}) dt_1 dt_2 \ldots dt_{N-1}. \tag{A.11}$$

A.2 Integration by Polynomial Approximation

In this section we present a deterministic method for performing the integration on the parameter space. This method will work if the integrand is polynomial or can be well-approximated in the least-squares sense by a polynomial. From the degradation models used in our experiments, the Monte Carlo method is superior; however, this method is presented as a possible alternative.

Suppose now that $h(u_1, \ldots, u_N)$ is a multivariate polynomial of the form

$$h(u_1, \ldots, u_N) = \sum_{k=1}^{Q} a_k u_1^{d_{1,k}} u_2^{d_{2,k}} \cdots u_N^{d_{N,k}}. \tag{A.12}$$

The number of terms in the polynomial is Q, and the $d_{i,k}$ are the exponents of the variables; k is the index of the k^{th} term in the sum and i is the i^{th} variable. The a_k are just fixed scalar constants.

We can substitute in the parameterization equations (A.8) for the integration in the next step, to obtain

$$h(u_1, \ldots, u_N) = \sum_{k=1}^{Q} a_k [g_1(\mathbf{t})]^{d_{1,k}} [g_2(\mathbf{t})]^{d_{2,k}} \cdots [g_n(\mathbf{t})]^{d_{N,k}}. \tag{A.13}$$

Substituting (A.12) into (A.11) yields

$$\int_0^{\pi} \cdots \int_0^{\pi} \sum_{k=1}^{Q} a_k [g_1(\mathbf{t})]^{d_{1,k}} [g_2(\mathbf{t})]^{d_{2,k}} \cdots [g_N(\mathbf{t})]^{d_{N,k}} \; sin^{N-2}(t_1) \cdots sin(t_{N-2}) d\mathbf{t}. \tag{A.14}$$

By linearity of integration we convert it to

$$\sum_{k=1}^{Q} a_k \int_0^{\pi} \cdots \int_0^{\pi} [g_1(\mathbf{t})]^{d_{1,k}} [g_2(\mathbf{t})]^{d_{2,k}} \cdots [g_N(\mathbf{t})]^{d_{N,k}} \; sin^{N-2}(t_1) \cdots sin(t_{n-2}) d\mathbf{t}. \tag{A.15}$$

The g_j functions have been used to concisely denote that the variables have been replaced by the transformation equations (A.8). To simplify the discussion, consider only a single term in this summation of integrals. We will give a fast method for evaluating a single term in this summation (i.e., each multiple integral). The single-term computation is then applied iteratively to every term in the sum, yielding the aggregate result. A single term in the sum (A.15) with the index variable k dropped can be represented as

$$a \int_0^{\pi} \cdots \int_0^{\pi} [g_1(\mathbf{t})]^{d_1} [g_2(\mathbf{t})]^{d_2} \cdots [g_N(\mathbf{t})]^{d_N} \; sin^{N-2}(t_1) \cdots sin(t_{N-2}) d\mathbf{t}. \tag{A.16}$$

When the g functions are expanded, and the variables in t_i are grouped, the integral is of the form

$$a \int_0^{\pi} \cdots \int_0^{\pi} \prod_{i=1}^{N} sin^{l_i}(t_i) cos^{m_i}(t_i) d\mathbf{t}. \tag{A.17}$$

By adding up the exponents from products of identical trigonometric functions, each of the l_i amd m_i is given by

$$\begin{aligned} l_1 &= d_2 + d_3 \ldots + d_N + N - 2 \\ l_2 &= d_3 + \ldots + d_N + N - 3 \\ &\vdots \\ l_{N-2} &= d_{N-1} + d_N + 1 \\ l_{N-1} &= d_N \end{aligned} \tag{A.18}$$

and

$$\begin{aligned} m_i &= d_i \text{ for } 1 \leq i \leq N-1 \\ m_N &= 0. \end{aligned} \tag{A.19}$$

Since (A.17) is written as a product of a functions of a single variable, the multiple integral can be separated into

$$a \prod_{i=1}^{n} \int_0^{\pi} sin^{l_i}(t_i) cos^{m_i}(t_i) dt_i. \tag{A.20}$$

From standard tables of definite integration, a single integral evaluates to

$$\int_0^{\pi} sin^l(t_i)cos^m(t_i)dt_i = \begin{cases} 0 & \text{if } m > 1 \text{ is odd} \\ \frac{2(l-1)(l-3)\cdots 4\cdot 2}{(l+m)(l+m-2)\cdots(m+3)(m+1)} & \text{if } l > 1 \text{ is odd} \\ \frac{\pi(l-1)(l-3)\cdots 3\cdot 1\cdot(m-1)(m-3)\cdots 3\cdot 1}{(l+m)(l+m-2)\cdots 4\cdot 2} & \text{if both are even.} \end{cases} \quad \text{(A.21)}$$

There are similar expressions for cases in which l or m is zero or one. The integral problem has now been solved in a closed form. For a given multivariate polynomial h as in (A.12), the required manifold integration reduces to a large sum of products of elements from (A.21).

A further simplification can be made by considering the implications of (A.19) and the fact the (A.21) is zero if m_i is odd. This implies that the entire product of (A.20) will simplify to zero if *any* of the exponents of the variables in the original term of the polynomial are odd. Depending on the distribution of the odd exponents, this result can cause numerous terms to disappear from the final sum, simply by inspection of the exponents.

APPENDIX B

SEGMENTATION DATA

B.1 Data Points for Example Regions, R_1 and R_2

The point coordinates are (x_1 x_2 x_3).

The points in R_1 are:

```
(24.0 10.0 -1.5369262) (24.0  9.0 -0.2597535) (24.0  8.0  1.4579037)
(24.0  7.0  1.3218407) (24.0  6.0 -1.1364636) (23.0 12.0  0.8938043)
(23.0 11.0 -1.1381520) (23.0 10.0  0.2881298) (23.0  9.0 -0.5505101)
(23.0  8.0 -0.3093603) (23.0  7.0  0.0335365) (23.0  6.0  1.2867588)
(22.0 12.0  0.2791380) (22.0 11.0  0.7047143) (22.0 10.0  0.8048603)
(22.0  9.0  1.1842969) (22.0  8.0 -0.6484462) (22.0  7.0  1.0783443)
(22.0  6.0 -0.7410597) (21.0 12.0 -1.0307879) (21.0 11.0 -0.1422201)
(21.0 10.0 -0.3741038) (21.0  9.0  1.8272568) (21.0  8.0 -0.8696967)
(21.0  7.0  0.4275988) (21.0  6.0 -1.554798) .
```

The points in R_2 are:

```
(28.0 29.0 -1.0400078) (27.0 30.0 -1.4599035) (27.0 29.0  0.7520007)
(27.0 28.0  0.0888681) (27.0 27.0 -1.5862793) (26.0 30.0  0.9179294)
(26.0 29.0  1.9066130) (26.0 28.0 -1.3594628) (26.0 27.0 -0.5189850)
```

(25.0 31.0 -0.3602752) (25.0 30.0 -0.8257680) (25.0 29.0 -1.7867581)
(25.0 28.0 -0.6265615) (25.0 27.0 0.7143770) (24.0 31.0 1.5510347)
(24.0 30.0 -0.9501506) (24.0 29.0 -1.0527672) (24.0 28.0 -0.3747505)
(24.0 27.0 -0.2552269) (23.0 31.0 -0.0135852) (23.0 30.0 -1.3428557)
(23.0 29.0 -1.5470942) (23.0 28.0 0.1422386) (23.0 27.0 -0.0165717)
(22.0 31.0 -0.1671206) (22.0 30.0 0.6013317) (22.0 29.0 1.8095361)
(22.0 28.0 1.1711003) (22.0 27.0 -0.5246841) (21.0 31.0 -0.2063034)
(21.0 30.0 0.1388412) (21.0 29.0 0.6763243) (21.0 28.0 -0.4390843)
(21.0 27.0 -1.0159935) .

REFERENCES

[1] B. K. P. Horn, *Robot Vision.* Cambridge, MA: MIT Press, 1986.

[2] R. Szeliski, "Bayesian modeling of uncertainty in low-level vision," *Int. J. Comput. Vis.*, vol. 5, pp. 271–301, Dec. 1990.

[3] J. L. Horowitz and T. Pavlidis, "Representation and segmentation of a cluttered scene using fused edge and surface data," in *Proc. IEEE 2nd Int. Conf. on Pattern Recognition*, pp. 424–433, 1974.

[4] T. Pavlidis, *Structural Pattern Recognition.* New York, NY: Springer-Verlag, 1977.

[5] A. L. Vickers and J. W. Modestino, "A maximum likelihood approach to texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 4, pp. 61–68, Jan. 1982.

[6] U. Grenander and M. I. Miller, "Recognition of global shapees in electron-mircographs," tech. rep., Monograph of the Electronic Signals and Systems Research Laboratory at Washington University, 1991.

[7] S. D. Cochran and G. Medioni, "3-D surface description from binocular stereo," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 981–994, Oct. 1992.

[8] H. Maitre and W. Luo, "Using models to improve sterro reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 269–277, Feb. 1992.

[9] J. Weng, N. Ahuja, and T. S. Huang, "Matching two perspective views," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 802–825, Aug. 1992.

[10] I. S. Kweon and T. Kanade, "High-resolution terrain map from multiple sensor data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 278–292, Feb. 1992.

[11] K. Ikeuchi and T. Kanade, "Modeling sensors: Toward automatic generation of object recognition program," *Comp. Vision, Graphics, and Image Process.*, vol. 48, pp. 50–79, 1989.

[12] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 679–698, 1986.

[13] J. R. Beveridge, J. Griffith, R. R. Kohler, A. R. Hanson, and E. M. Riseman, "Segmenting images using localized histograms and region merging," *Int. J. Comput. Vis.*, vol. 2, pp. 311–347, Jan. 1989.

[14] S. W. Zucker, "Region growing: Childhood and adolescence," *Comp. Graphics and Image Process.*, vol. 5, pp. 382–399, 1976.

[15] C. Chu and J. K. Aggarwal, "The integration of region and edge-based segmentation," in *Proc. Int. Conf. on Computer Vision*, pp. 117–120, 1990.

[16] J. F. Haddon and J. F. Boyce, "Image segmentation by unifying region and boundary information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 929–948, Oct. 1990.

[17] T. Pavlidis and Y. Liow, "Integrating region growing and edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 225–233, Mar. 1990.

[18] N. Yokoya and M. D. Levine, "Range image segmentation based on differential geometry: A hybrid approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 643–649, June 1989.

[19] A. Pérez and R. C. Gonzalez, "An iterative thesholding algorithm for image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 742–751, Nov. 1987.

[20] R. M. Haralick and L. Watson, "A facet model for image data," *Comp. Vision, Graphics, and Image Process.*, vol. 15, pp. 113–129, Feb. 1981.

[21] F. S. Cohen and D. B. Cooper, "Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 195–219, Mar. 1987.

[22] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 39–55, Jan. 1987.

[23] B. S. Manjunath and R. Chellappa, "Unsupervised texture segmentation using Markov random field models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 478–482, May 1991.

[24] J. Y. Hsi and A. A. Sawchuk, "Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 1279–1292, Dec. 1989.

[25] J. Y. Hsi and A. A. Sawchuk, "Unsupervised textured image segmentation using feature smoothing and probabilistic relaxation techniques," *Comp. Vision, Graphics, and Image Process.*, vol. 48, pp. 1–21, 1989.

[26] A. Perry and D. G. Lowe, "Segmentation of textured images," in *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 319–324, June 1989.

[27] T. R. Reed and H. Wechsler, "Segmentation of textured images and Gesalt organization using spatial/spatial frequency representations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1–12, Jan. 1990.

[28] M. Tüceryan and A. K. Jain, "Texture segmentation using Voronoi polygons," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 211–216, Feb. 1990.

[29] A. Khotanzad and J. Chen, "Unsupervised segmentation of textured images by edge detection in multidimensional features," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 414–421, Apr. 1989.

[30] D. Geman and S. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, Nov. 1984.

[31] J. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: A stepwise optimization approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 150–163, Feb. 1989.

[32] M. Bister, J. Cornelis, and A. Rosenfeld, "A critical view of pyramid segmentation algorithms," *Pattern Recognition Lett.*, vol. 11, pp. 605–617, 1990.

[33] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 99–113, Feb. 1991.

[34] W. I. Grosky and R. Jain, "A pyramid-based approach to segmentation applied to region matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 639–650, Sept. 1986.

[35] L. M. Lifshitz and S. M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 529–540, June 1990.

[36] P. B. Chou and C. M. Brown, "The theory and practice of Bayesian image labeling," *Int. J. Comput. Vis.*, vol. 4, pp. 185–210, 1990.

[37] M. I. Sezan and A. M. Tekalp, "Survey of recent developments in digital image restoration," *Opt. Eng.*, vol. 29, no. 5, pp. 393–404, 1990.

[38] C. Brice and C. Fennema, "Scene analysis using regions," *Art. Intell.*, vol. 1, pp. 205–226, 1970.

[39] R. B. Ash, *Information Theory.* New York, NY: Dover Publications, 1990.

[40] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach.* Englewood Cliffs, NJ: Prentice/Hall Publications, 1982.

[41] H. E. Stephanou and S. Y. Lu, "Measuring consensus effectiveness by a generalized entropy criterion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 544–554, July 1988.

[42] D. Geiger and A. Yuille, "A common framework for image segmentation," in *Proc. IEEE 10th Int. Conf. on Pattern Recognition*, (Atlantic City, NJ), pp. 502–507, June 1990.

[43] J. F. Silverman and D. B. Cooper, "Bayesian clustering for unsupervised estimation of surface and texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 482–496, July 1988.

[44] S. Geman, "Experiments in Bayesian image analysis," *Bayesian Statistics*, vol. 3, pp. 159–172, 1988.

[45] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, Aug. 1989.

[46] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York, NY: Wiley, 1973.

[47] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice Hall, Inc., 1988.

[48] J. Jolion, P. Meer, and S. Bataouche, "Robust clustering with applications in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 791–801, Aug. 1991.

[49] M. Celenk, "A color clustering technique for image segmentation," *Comp. Vision, Graphics, and Image Process.*, vol. 52, pp. 145–170, 1990.

[50] K. Fukunaka, *Introduction to Statistical Pattern Recognition.* New York, NY: Academic Press, 1972.

[51] J. Zhang and J. W. Modestino, "A model-fitting approach to cluster validation with applications to stochastic model-based image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1009–1017, Oct. 1990.

[52] Z. W. Bell, "A Bayesian/Monte Carlo segmentation method for images dominated by Gaussian noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 985–989, Sept. 1989.

[53] P. K. Sahoo, S. Soltani, A. K. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Comp. Vision, Graphics, and Image Process.*, vol. 41, pp. 233–260, Feb. 1988.

[54] E. Ising *Zeitschrift Physik*, vol. 31, p. 253, 1925.

[55] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 401–412, May 1991.

[56] D. Geiger and A. Yuille, "A common framework for image segmentation," *Int. J. Comput. Vis.*, vol. 6, no. 3, pp. 227–243, 1991.

[57] M. Daily, "Color image segmentation using Markov Random Fields," in *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 304–312, June 1989.

[58] V. Johnson, W. Wong, X. Hu, and C. Chen, "Image restoration using Gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameter," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 413–425, May 1991.

[59] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 609–628, July 1990.

[60] J. Besag, "Spatial interaction and the statistical analyis of lattice systems," *J. Royal Statistical Society*, vol. B36, pp. 192–236, 1974.

[61] B. D. Ripley, *Statistical Inference for Spatial Processes.* New York, NY: Cambridge Press, 1988.

[62] J. W. Modestino and J. Zhang, "A Markov random field model-based approach to image interpretation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 606–615, June 1992.

[63] J. Goutsias and J. M. Mendel, "Simultaneous optimal segmentation and model estimation of nonstationary noisy images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 990–998, Sept. 1989.

[64] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statistical Society*, vol. B48, pp. 259–302, 1986.

[65] R. Dubes, A. Jain, S. Nadabar, and C. Chen, "MRF model-based algorithms for image segmentation," in *Proc. IEEE 10th Int. Conf. on Pattern Recognition*, (Atlantic City, NJ), pp. 808–814, June 1990.

[66] J. Subrahmonia, Y. Hung, and D. Cooper, "Model-based segmentation and estimation of 3D surfaces from two or more intensity images using Markov random fields," in *Proc. IEEE 10th Int. Conf. on Pattern Recognition*, (Atlantic City, NJ), pp. 390–397, June 1990.

[67] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene labeling by relaxation algorithms," *IEEE Trans. Syst., Man, Cybern.*, vol. 6, pp. 420–433, 1976.

[68] S. Peleg, "A new probabilistic relaxation scheme," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 2, pp. 362–369, 1980.

[69] A. Rosenfeld and A. C. Kak, *Digital Picture Processing.* New York, NY: Academic Press, 1982.

[70] W. S. Rutkowski, S. Peleg, and A. Rosenfeld, "Shape segmentation using relaxation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 3, pp. 368–375, July 1981.

[71] O. D. Faugeras and M. Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 3, pp. 412–424, July 1981.

[72] S. A. Kuschel and C. V. Page, "Augmented relaxation labeling and dynamic relaxation labeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 4, pp. 676–682, Nov. 1982.

[73] H. Don and K. Fu, "A parallel algorithm for stochastic image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 594–603, Sept. 1986.

[74] D. Jeong and P. M. Lapsa, "Unified approach for early-phase image understanding using a general decision criterion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 357–371, Apr. 1989.

[75] P. J. Flynn and A. K. Jain, "CAD-Based computer vision: From CAD models to relational graphs," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 114–132, Feb. 1991.

[76] S. M. LaValle and S. A. Hutchinson, "Representing probability distributions of image segments and segmentations," in *Proc. IEEE Int. Conf. on Syst., Man, Cybern.*, (Chicago), pp. 1552–1557, Oct. 1992.

[77] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comp. Vision, Graphics, and Image Process.*, vol. 29, pp. 100–132, Jan. 1985.

[78] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees.* Belmont, CA: Wadsworth, 1984.

[79] P. A. Chou, "Optimal partitioning for classification and regression trees," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 340–354, Apr. 1991.

[80] S. B. Gelfand, C. S. Ravishankar, and E. J. Delp, "An iterative growing and pruning algorithm for classification tree design," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 163–174, Feb. 1991.

[81] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, California: Morgan Kaufmann Publishers, Inc., 1988.

[82] S. M. LaValle and S. A. Hutchinson, "Considering multiple surface hypotheses in a Bayesian hierarchy," in *Proc. of the SPIE Conf. on Stochastic Methods in Signal Processing, Image Processing, and Computer Vision*, (San Diego, CA), pp. 2–15, July 1991.

[83] J. M. Agosta, "An example of a Bayes network of relations among visual features," in *Proc. of the SPIE Conf. on Stochastic Methods in Signal Processing, Image Processing, and Computer Vision*, (San Diego, CA), pp. 16–27, July 1991.

[84] T. O. Binford, T. S. Levitt, and W. B. Mann, "Bayesian inference in model-based machine vision," in *Proceedings of the AAAI Workshop on Uncertainty in Artificial Intelligence*, pp. 86–92, 1987.

[85] S. Sarkar and K. L. Boyer, "Integration, inference, and management of spatial information using Bayesian networks: Perceptual organization," Tech. Rep. SAMPL-91-08, Ohio State Dept. of Elec. Engin., Nov. 1991.

[86] O. D. Faugeras and M. Hebert, "The representation, recognition, and locating of 3-D objects," *Int. J. of Robot. Res.*, vol. 5, pp. 27–52, Fall 1986.

[87] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.

[88] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*. New York, NY: Macmillan, 1978.

[89] P. Williams, *Probability with Martingales*. New York, NY: Cambridge Press, 1991.

[90] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. Englewood Cliffs, NJ: Prentice Hall, Inc., 1986.

[91] P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 167–191, Mar. 1988.

[92] A. Leonardis, A. Gupta, and R. Bajcsy, "Segmentation as the search for the best description of the image in terms of primitives," in *Proc. Int. Conf. on Computer Vision*, pp. 121–125, 1990.

[93] G. Taubin, "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 1115–1137, Nov. 1991.

[94] R. M. Bolle and D. B. Cooper, "On optimally combining pieces of information, with application to estimating 3-D complex-object position from range data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 619–638, Sept. 1986.

[95] G. Taubin and D. Cooper, "Recognition and positioning of 3D piecewise algebraic objects using Euclidean invariants," in *Proc. Workshop on the Integration of Numerical and Symbolic Computing Methods*, (Saratoga Springs, NY), July 1990.

[96] R. T. Chin and C. R. Dyer, "Model-based recognition in robot vision," *Comput. Surv.*, vol. 18, pp. 67–108, Mar. 1986.

[97] D. J. Kreigman and J. Ponce, "On recognizing and positioning curved 3D objects from image contours," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1127–1137, Dec. 1990.

[98] R. M. Bolle and B. C. Vemuri, "On three-dimensional surface reconstruction methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 1–12, Jan. 1991.

[99] H. W. Guggenheimer, *Differential Geometry.* New York, NY: Dover Publications, 1977.

[100] P. Meer, J. Jolion, and A. Rosenfeld, "A fast parallel algorithm for blind estimation of noise variance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 216–223, Feb. 1990.

[101] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds.* New York, NY: Dover Publications, 1980.

[102] A. H. Stroud, *Approximate Calculation of Multiple Integrals.* Englewood Cliffs, NJ: Prentice Hall, Inc., 1971.

[103] W. Kaplan, *Advanced Calculus.* Menlo Park, CA: Addison Wesley, 1984.

[104] S. Yakowitz, J. E. Krimmel, and F. Szidarovsky, "Weighted Monte Carlo integration," *SIAM J. Numer. Anal.*, vol. 15, pp. 1289–1300, Dec. 1978.

[105] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods.* New York, NY: Wiley, 1986.

[106] E. Masry and S. Cambanis, "Trapezoidal Monte Carlo integration," *SIAM J. Numer. Anal.*, vol. 27, pp. 225–246, Feb. 1990.

[107] J. H. Zar, "Approximations for the percantage points of the chi-squared distribution," *J. Appl. Statist.*, vol. 27, no. 3, pp. 280–290, 1978.

[108] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. of Appl. Math.*, vol. 11, pp. 431–441, 1963.

[109] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *An Introduction to Algorithms.* Cambridge, MA: MIT Press, 1990.

[110] P. H. Winston, *Artificial Intelligence.* Reading, MA: Addison-Wesley, 1984.

[111] S. Sullivan, J. Ponce, and L. Sandford, "Exact geometric distance minimization for surface deformation, and pose estimation from range, CT, and video images," submitted for publication, 1992.