A Framework for Constructing Probability Distributions on the Space of Image Segmentations

Steven M. LaValle lavalle@cs.uiuc.edu
Seth A. Hutchinson seth@cs.uiuc.edu

The Beckman Institute

and

Dept. of Electrical and Computer Engineering

University of Illinois

Urbana, IL 61801

Abstract

The goal of traditional probabilistic approaches to image segmentation has been to derive a single, optimal segmentation, given statistical models for the image formation process. In this paper, we describe a new probabilistic approach to segmentation, in which the goal is to derive a set of plausible segmentation hypotheses and their corresponding probabilities. Because the space of possible image segmentations is too large to represent explicitly, we present a representation scheme that allows the implicit representation of large sets of segmentation hypotheses that have low probability. We then derive a probabilistic mechanism for applying Bayesian, model-based evidence to guide the construction of this representation. One key to our approach is a general Bayesian method 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 present experimental results for both real and synthetic range data, obtained from objects composed of piecewise planar and implicit quadric patches.

1 Introduction

Image segmentation, the low-level vision task of extracting a set of homogeneous regions from an image, has been a topic of active research since the earliest days of computer vision research. Although considerable effort has yielded many approaches, segmentation is still widely considered to be an unsolved problem. The difficulty of the segmentation problem is due, at least in part, to its underconstrained nature. For example, Horn asserts that one of the primary difficulties in evaluating a segmentation method is the lack of a clear definition of the "correct" segmentation [32]; and Szeliski argues that low-level image models often underconstrain the solution, and advocates the use of uncertainty estimation [63]. Jain and Binford assert that a key problem in vision research is that the segmentation problem is often ignored or is assumed to have been solved [38].

In spite of these criticisms, most, if not all, previous approaches to segmentation have been aimed at deriving a single, optimal segmentation result for a given scene. This segmentation is often used by some higher-level system (e.g., for object recognition). In this paper, we present an alternative approach. For a given input scene, rather than attempting to derive a single, optimal segmentation, we derive a set of plausible segmentation hypotheses and their corresponding probabilities. This is achieved by using statistical image models to determine a Bayesian posterior distribution over a set of alternative image segmentations. In terms of image understanding, this approach can be considered as a step toward breaking down the segregation that usually exists between segmentation and a higher-level system, by allowing a higher-level system to have access to more information at the segmentation level and possibly to influence segmentation-level Bayesian computations.

Our approach provides the following three key advantages, which will be discussed below.

- Multiple alternative segments or segmentations, and their corresponding probabilities, can be utilized by a higher-level system.
- The framework readily supports extensions to incorporate higher-level or other additional models.
- Our system is capable of estimating the amount of information present in the image

under a particular statistical image model.

Consider the first point. One straightforward use of multiple segmentations (or segments) would be to provide ranked alternatives for a system that can repeatedly request a different segmentation (or segment), given that previous solutions led to a failure. 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 a probability of 0.99, then the confidence in the segmentation should be high. If the top ten segmentations have approximately the same probability, some other model may have to be used to further constrain the solution.

Since the space of alternatives is often underconstrained using low-level models [63], a more interesting approach is to introduce additional 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 system, since the single segmentation is formed by making all of the decisions using low-level models, losing all other information. 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 system to constrain the space of segmentations, eventually resulting in a unique solution.

This provides the motivation for our second point. 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. For example, segmentation choices based on surface information could be constrained by introducing a regularity model for segment boundaries. Furthermore, additional evidence can be incorporated directly into posterior distribution on the space of segmentations during the execution of our algorithm. This offers the computational advantage of allowing this ev-

idence to preclude the consideration of numerous alternatives that could later be eliminated. For example, in Section 4.3.2 we provide expressions for combining evidence from multiple independent models when assessing region homogeneity. This additional evidence would also be introduced when computing probabilities for combinations of segments, described in Section 3.3.2. Extensions of this type are currently being investigated in the continuation of this work.

For a typical application, it is useful to know the degree to which a particular image model is providing information regarding the segmentation. When some models are significantly more costly than others, which is typically the case in computer vision, this is particularly true. This is the motivation for our third point. 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 [3]. Several alternative, entropy-based functions can be found in [17],[60].

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 [63]. We presently do not employ entropy measures to guide our algorithms, but we could estimate the amount of information present by applying an entropy measure directly to the probability distributions of segments and segmentations.

In practice, it may not be necessary to construct full segmentations. For instance, for model-based recognition of 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 [23]. Each three-segment group corresponds to a set of segmentations. One advantage of our work is that it allows the consideration of such partial segmentations, and their corresponding probabilities.

We now briefly describe the organization of this paper. Since our work has been influenced by previous statistical approaches, Section 2 presents a brief survey of previous statistical or probabilistic approaches to segmentation.

In Section 3 we will define a probability space on the set of all possible image partitions, which we refer to as the *Segmentation Sample Space*. It is not feasible, from a computational standpoint, to explicitly represent every segmentation in the Segmentation Sample Space. Therefore, in Section 3, we introduce an efficient, approximate representation scheme that allows *explicit* representation of those segmentations with high probability, while *implicitly* representing subsets of segmentations with low probability. The representation is defined in terms of *Segment Sample Spaces*. Each Segment Sample Space represents a set of region groupings in the image. For each such grouping, a probability is computed that reflects the homogeneity of the region grouping. For example, when dealing with range data from objects that are composed of piecewise quadric surfaces, a region grouping will have a high degree of homogeneity if it can be closely approximated by a quadric surface.

To effectively utilize our approximate representations of the Segment and Segmentation Sample Spaces, we have developed probabilistically sound and computationally feasible methods. Section 3.2 describes concepts that relate to the Segment Sample Space. Section 3.3 describes concepts that relate to the Segmentation Sample Space. In Section 4 we present the probability relationships and expressions that are used during the incremental construction of a Segment or Segmentation Sample Space representation, which incorporate evidence from a statistical image model.

In Sections 5 and 6 we describe some algorithm details, and present experimental results. The experimental results are for range data obtained from objects composed of piecewise planar and quadric surfaces. We present probability distributions over alternative segments and segmentations, for both synthetic data and real range data. Conclusions are presented in Section 7. Proofs of the propositions are presented in Appendix A. In this paper we have used the implicit polynomial surface model for defining region-grouping homogeneity, and these details are presented in Appendix B. We have shown in related work that the probabilities that reflect region-grouping homogeneity can also be computed for parametric polynomials for intensity images, and a Markov Random Field model for texture analysis [46].

2 Statistical Segmentation Approaches

Statistical approaches to segmentation can be loosely divided into three categories: statistical clustering, MRF energy optimization, and probabilistic relaxation. In this section, we briefly review work in these areas.

Clustering has been applied to a variety of image types and models. Silverman and Cooper [58] segment intensity images into regions that can be approximated by planar or quadric surfaces. Bell combines clustering with a Monte-Carlo approach to segment radiograph images, to determine manufacturing defects [4]. Four basic components involved in most clustering algorithms are (e.g., [20],[36]): (1) Define a feature metric space. (2) Determine feature values corresponding 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 feature space could, for example, correspond directly to pixel intensities, or could represent a space of polynomial surfaces, as in [58]. The decisions involved in the third step depend on the particular clustering algorithm chosen, such as agglomerative clustering [20],[58] and K-means clustering [24],[67]. Most clustering algorithms require specification of the number of classes, and 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 [40],[67].

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. The general model that is most often used in computer vision was introduced in a seminal paper by Geman and Geman [27], in the context of image restoration. The 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 [25]. The approach has been applied to modeling noise processes and texture [14],[16],[48], color-constancy [15], blurring [27],[39], boundary modeling [28],[39], and locally-dependent nonlinear image transformations [27]. This model has also been considered for range image

segmentation by modeling edges with MRF line processes [18],[37],[49].

In recent years there 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 [27]. 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 [7] 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 [18], 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 [16]. 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 [14].

An alternative approach to the minimization is taken by Chou and Brown [13]. 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 [16]. Cohen and Cooper discuss the problem of adaptively estimating parameters during segmentation in the context of texture models [14]. 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

[62]. An alternative adaptive estimation/optimization scheme has been proposed by Lakshmanan and Derin [44]. Manjunath and Chellappa [48] 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 [44].

The probabilistic relaxation (or relaxation labeling) approach has many similarities to the MRF approach and has been less popular in recent times. One of the early appearances of relaxation labeling is in work by Rosenfeld, Hummel, and Zucker [53]. A *compatibility* measure is used to model the interaction between pairs of pixels, when determining probability assignments. One begins with a prior distribution of labelings, and through an assignment rule based on compatibility, iteratively improves estimates of the "true" probabilities. Peleg provides a 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 determine these probabilities, as in [54]. A more recent discussion of relaxation labeling, applied to the problem of supervised and unsupervised texture classification can be found in [33],[34].

3 Representing Segment and Segmentation Probability Spaces

In this section we describe an approximate representation scheme that allows for explicit representation of segmentation hypotheses with high probability, while implicitly specifying large sets of segmentation hypotheses with low probability. We begin in Section 3.1, by defining several terms that will be used throughout the paper. For convenience, notation that will be used in this paper is summarized in Table 1. In Section 3.2, we describe the mechanism for building approximate representations for a single Segment Sample Space, which is used to maintain a probability distribution over a space of single segments in an image. In Section 3.3, we describe how, given a set of approximate Segment Sample Space representations, we can construct an approximate representation of the Segmentation Sample Space, which describes the probability distribution over all possible segmentations. Finally,

Symbol	Definition	Section
R	A connected subset of the image	3.1
$\mathcal{R}$	The set of all regions	3.1
T	A segment (connected set of regions) from the image	3.1
S	A segmentation (or partition of $\mathcal{R}$)	3.1
Π	The set of all segmentations that can be generated from $\mathcal{R}$	3.1
Θ_i	The set of all segments that include R_i	3.2
$\mathcal{T}_i$	A Segment Sample Space $(\Theta_i, \mathcal{B}_i, P)$	3.2
$\mathcal{B}_i$	The set of all subsets of Θ_i (i.e., the events on a $\mathcal{T}$)	3.2
I	An inclusion set, containing regions	3.2
E	An exclusion set, containing regions	3.2
$\tau(I,E)$	$\{T \in \Theta_i : I \subseteq T, E \cap T = \emptyset\}$	3.2
C	A $\mathcal{T}$-cover (a partition of Θ_i)	3.2
B_ρ	An event chosen for $\mathcal{T}$-refinement	3.2
R_ρ	A region chosen for $\mathcal{T}$-refinement	3.2
$\rho(C, B_\rho, R_\rho)$	The $\mathcal{T}$-refinement mapping	3.2
$\mathcal{S}$	The Segmentation Sample Space $(\Pi, \mathcal{A}, P)$	3.3
$\mathcal{A}$	The set of all subsets of Π (i.e., the events on the $\mathcal{S}$)	3.3
f_i	The segment-to-segmentation mapping	3.3.1
$\sigma(F,I,E)$	$\{S \in \Pi : F \subseteq S\} \bigcap f_i(\tau(I,E))$	3.3.2
$\mathbf{C}$	An $\mathcal{S}$-cover (a partition of Π)	3.3.2
A_ρ	An event chosen for $\mathcal{S}$-refinement	3.3.2
P_I	The probability of including R_ρ	4.1
$\tau_{i\rho}$	$\tau(\{R_i, R_\rho\}, \emptyset)$	4.1
$\mathbf{U}_k$	The parameter space for R_k	4.2
$\mathbf{Y}_k$	The observation space for R_k	4.2
λ_0	A ratio based on prior membership probability	4.2
λ_1	A ratio based on models and observations	4.2

Table 1: Notation used in this paper, with elements sorted by the order of their introduction.

in Section 3.4 we address issues related to defining prior probability distributions for Segment Sample Space and Segmentation Sample Space representations.

3.1 Regions, Segments and Segmentations

The input to our segmentation algorithm is an array of image elements. Associated with each element is its representation. This might may be an intensity value, a set of coordinates in $\Re^3$, other image information, or a combination of these. In Appendix B (in which implicit polynomial models are briefly presented), this represents a point in $\Re^3$, with coordinates $\mathbf{x} = [x_1 \; x_2 \; x_3]$. Since the elements of the image are arranged in a matrix, adjacencies and connectivity can be considered in the usual way.

A *region*, R, is some connected subset of the image. In practice, most region-based segmentation algorithms begin by partitioning the image into an initial set of regions, $\mathcal{R}$

(e.g.,[31],[50],[55],[58]). This provides a computational advantage (since there are not as many potential groupings of data points to consider), and also allows statistical models to be effectively exploited [58]. The only concern in the construction of $\mathcal{R}$ is that each $R \in \mathcal{R}$ should be homogeneous. In practice, however, we do tolerate the existence of several regions that are not homogeneous. In our experiments, we have constructed $\mathcal{R}$ by recursively splitting regions that (a) cannot be closely approximated by a single plane, or (b) contain an edge.

For a given segmentation problem, we work with a pairwise-disjoint set of regions, $\mathcal{R}$, in which every element of the image is contained in exactly one region $R \in \mathcal{R}$. A *segment*, T, is a connected set of regions (e.g., $T = \{R_1, R_2, R_3\}$ is a segment consisting of three regions). A set of regions is *connected* if their union is connected. A *segmentation*, S, denotes a set of segments that forms a partition of $\mathcal{R}$. Note that a segmentation in turn defines a partition of the image.

Given a segmentation, S, and two adjacent segments $T_1, T_2 \in S$, 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. This corresponds to region merging, expressed in our set-theoretic terms. We note, however, that in our formalism regions are not actually *merged* in the traditional sense. Rather, we form groups of regions, without ever explicitly constructing the union of these regions.

Let Π denote the set of all segmentations that can be constructed given $\mathcal{R}$ (i.e., Π is the set of all partitions of $\mathcal{R}$). At one extreme, Π includes the partition induced by the original regions in $\mathcal{R}$. At the other extreme, Π contains the partition that corresponds to the combination of all regions into one segment. The implication of starting with $\mathcal{R}$ is that there are many *image* partitions that are not considered. In the limiting case, the elements of $\mathcal{R}$ are singleton subsets of the image. Hence, using $\mathcal{R}$ as a basis does not impose any inherent limitations on the space of segmentations that can be generated.

3.2 A Segment Sample Space ($\mathcal{T}_i$)

In this section, we define a Segment Sample Space, and describe how approximate Segment Sample Space representations are constructed. A detailed example is described in

Section 3.2.1, which illustrates the definitions and concepts. In Section 3.3, we show how approximate Segmentation Sample Space representations can be constructed, using these approximate Segment Sample Space representations as a starting point.

For some region $R_i \in \mathcal{R}$, let Θ_i be the set of all possible segments that contain R_i. Specifically,

$$\Theta_i = \{T \subseteq \mathcal{R} \mid T \textit{ is connected},\ R_i \in T\}. \tag{1}$$

Note that Θ_i always contains at least two elements: the singleton $\{R_i\}$ and the entire set $\mathcal{R}$ (provided the image is connected). For any such Θ_i, there is a corresponding Segment Sample Space that describes both Θ_i and the probabilities associated with each subset of Θ_i. Specifically, a Segment Sample Space is defined as

$$\mathcal{T}_i = (\Theta_i, \mathcal{B}_i, P_i) \tag{2}$$

in which Θ_i is defined in (1), $\mathcal{B}_i$ is the set of all subsets of Θ_i, and P is a probability mapping on $\mathcal{B}_i$. (Throughout the remainder of the paper, to simplify notation, we omit the subscript on the probability mapping P_i). Since the singleton events are mutually exclusive, the probability for an arbitrary event $B \in \mathcal{B}_i$ can be obtained by summing the probabilities $P(\{T\})$ for each $T \in B_i$.

For real image applications, the set of segments, Θ_i, will be extremely large; the set $\mathcal{B}_i$ is exponentially larger. Therefore, it is infeasible to explicitly enumerate the elements in either Θ_i or $\mathcal{B}_i$. To deal with these combinatoric issues, we now introduce an implicit representation for elements of $\mathcal{B}_i$, a representation for approximations to $\mathcal{T}_i$, and a mechanism by which any given approximation of $\mathcal{T}_i$ can be refined to yield a more accurate approximation.

Each element $B \in \mathcal{B}_i$ corresponds to a set of segments. We can identify this set by specifying (a) the set of all regions common to every segment in B, and (b) a set of regions not included in any segment in B. Specifically, the *inclusion set*, I, is the set of regions common to every segment in B (note that I always includes R_i). The *exclusion set*, E, is a set of regions that are not included in any segment in B. To eliminate redundant representations, we require each element of E to be adjacent to some region in I. Note,

$I \cap E = \emptyset$. Using this notation, we define $\tau(I, E)$, which maps to some $B \in \mathcal{B}_i$, as

$$\tau(I, E) = \{T \in \Theta_i : I \subseteq T,\ E \cap T = \emptyset\}. \tag{3}$$

Thus, $\tau(I, E)$ specifies the set of all segments that include all regions in I, and exclude all regions in E. This representation is efficient when compared to the enumeration of segments in B. For instance, $\tau(\{R_1, R_2\}, \emptyset)$ could represent the set of all segments in an image that contain regions R_1 and R_2, through the simple specification of $I = \{R_1, R_2\}$ and $E = \emptyset$.

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

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

The proof of this and all subsequent propositions are provided in Appendix A.

Given this representation for subsets of Θ_i, we now turn to the construction of approximations of $\mathcal{T}_i$. Ultimately, we would like to construct approximations of $\mathcal{T}_i$ that explicitly represent those segments that have high probability values, while only implicitly specifying large subsets of segments that have low probability values. However, for any approximation to $\mathcal{T}_i$, every segment should be represented, either explicitly or implicitly. To this end, we define a *$\mathcal{T}$-cover*, C, of $\mathcal{T}_i$ to be a set of pairwise-disjoint elements in $\mathcal{B}_i$ that form a partition of Θ_i. If the probabilities for the elements in C are known, we can consider C to represent an approximation of $\mathcal{T}_i$. It is approximate because probabilities are not associated with the singletons in $\mathcal{B}_i$, but only with those elements that are explicitly contained in C. Since the elements of C form a partition of Θ_i, every element of Θ_i is represented in C, either explicitly (in the case of singleton subsets of $\mathcal{B}_i$ in C) or implicitly (in the case of non-singleton subsets of $\mathcal{B}_i$ in C).

The notion of goodness of approximations can be formalized by imposing a partial ordering on $\mathcal{T}$-covers. Given two $\mathcal{T}$-covers, C_1 and C_2, we say that C_1 is better than C_2 if and only if for all $B_1 \in C_1$ there exists some $B_2 \in C_2$ such that $B_1 \subset B_2$. In other words, C_1 is better than C_2 if C_1 can be obtained by partitioning some of the elements of C_2. We denote by C_i^∞ the set of all singletons in $\mathcal{B}_i$. Thus, C_i^∞ is an exact representation of $\mathcal{T}_i$; all of the elements of Θ_i are explicitly represented, and the probability for each is given. Hence,

in this case the entire probability map is fully determined (since the probability for any $B \in \mathcal{B}_i$ can be obtained by summing the probabilities $P(\{T\})$ for each $T \in B_i$). Thus, C_i^∞ is better than C for all other $\mathcal{T}$-covers C. The poorest approximation of $\mathcal{T}_i$ is $C_i^0 = \{\Theta_i\}$. We know that $P(\Theta_i) = 1$; however, the probabilities of the other events in $\mathcal{B}_i$ cannot be directly determined. Thus, C is better than C_i^0 for all other $\mathcal{T}$-covers, C.

Procedurally, in order to construct approximations to $\mathcal{T}_i$, we begin with C_i^0, and derive a sequence of $\mathcal{T}$-covers, C_i^k, such that C_i^{k+1} is better than C_i^{k+1}. Each step in this sequence corresponds to a single *$\mathcal{T}$-refinement* operation. Specifically, given a $\mathcal{T}$-cover C_i^k, an event, $B_\rho = \tau(I_\rho, E_\rho) \in C_i^k$, and a region $R_\rho \notin I_\rho \cup E_\rho$, we define a new $\mathcal{T}$-cover, $C_i^{k+1} = \rho(C, B_\rho, R_\rho)$, by

$$\rho(C_i^k, B_\rho, R_\rho) = (C_i^k \ - \ B_\rho) \cup \{\tau(I_\rho \cup \{R_\rho\}, E_\rho), \tau(I_\rho, E_\rho \cup \{R_\rho\})\}. \tag{4}$$

The region R_ρ is termed the *$\mathcal{T}$-refinement region*. In order to ensure that only connected sets of regions are represented in the new $\mathcal{T}$-cover, we require the $\mathcal{T}$-refinement region, R_ρ to be adjacent to some region in I_ρ. The $\mathcal{T}$-cover, C_i^{k+1}, is termed the *refined $\mathcal{T}$-cover* with respect to C_i^k. The only difference between C_i^k and C_i^{k+1} is the replacement of B_ρ by $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and $\tau(I_\rho, E_\rho \cup \{R_\rho\})$. Thus, the $\mathcal{T}$-refinement operation has the effect of partitioning the event B_ρ into two new subsets of B_ρ: the segments in B_ρ that include R_ρ are in $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and the remaining elements of B_ρ (all those that exclude R_ρ) are in $\tau(I_\rho, E_\rho \cup \{R_\rho\})$. Each singleton in $\mathcal{B}_i$ represents a single segment. We will refer to these events as *ground segment events*, since such events can not be refined.

There is a correspondence between generating a sequence of $\mathcal{T}$-covers, and generating classification and regression trees [10],[12],[26]. Classification and regression trees are used to represent a sample space efficiently, often for the purpose of pattern recognition. In terms of classification and regression trees, $\mathcal{T}$-refinement corresponds to the notion of impurity reduction through partitioning [12]. 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 segment events that have

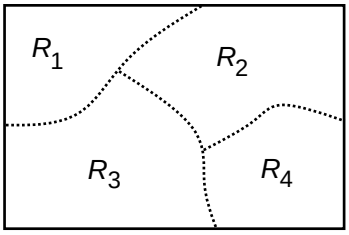


Figure 1: A simple image composed of only four regions is provided as an example.

highest probability.

3.2.1 An Illustrative Example

To help clarify these definitions, we now present an example. Figure 1 shows a hypothetical image consisting of four regions: R_1, R_2, R_3 and R_4. Consider Θ_1, which is the set of all segments that include R_1:

$$\begin{aligned}\Theta_1 \;=\; & \{\{R_1\},\{R_1,R_2\},\{R_1,R_3\},\{R_1,R_2,R_3\},\\ & \{R_1,R_2,R_4\},\{R_1,R_3,R_4\},\{R_1,R_2,R_3,R_4\}\}. \end{aligned} \tag{5}$$

Note that $\{R_1, R_4\} \notin \Theta_1$ since R_1 is not adjacent to R_4. For larger images, the number of segments in Θ_i 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.

In Table 2 we show the representation for a few events in $\mathcal{B}_1$ in terms of $\tau(I, E)$. In rows 4 and 8 to 13 of the table, $\tau(I, E)$ corresponds to a ground segment event. We note that $\tau(I, E)$ provides a compact representation for potentially large subsets of Θ_i. For example, in row 3 of the table, four elements of Θ_1 are implicitly represented by $I = \{R_1, R_2\}$ and $E = \emptyset$. The savings in representation increases as the number of regions in $\mathcal{R}$ increases.

Table 4 depicts one possible sequence of $\mathcal{T}$-refinement operations that could be derived. For convenience, Table 3 names all of the segments of Θ_1. We begin with C_1^0, which corresponds to the set Θ_1. $\mathcal{T}$-cover C_1^1 is constructed by using the mapping $C_1^1 = \rho(C_1^0, \Theta_1, R_2)$. The $\mathcal{T}$-cover C_1^2 is constructed using R_3 and event $\{T_1, T_3, T_6\}$. The iterations are continued

	I	E	$\tau(I,E)$
1	$\{R_1\}$	$\{\}$	Θ_1
2	$\{R_1\}$	$\{R_2\}$	$\{\{R_1\},\{R_1,R_3\},\{R_1,R_3,R_4\}\}$
3	$\{R_1,R_2\}$	$\{\}$	$\{\{R_1,R_2\},\{R_1,R_2,R_4\},\{R_1,R_2,R_3\},\{R_1,R_2,R_3,R_4\}\}$
4	$\{R_1\}$	$\{R_2,R_3\}$	$\{\{R_1\}\}$
5	$\{R_1,R_3\}$	$\{R_2\}$	$\{\{R_1,R_3\},\{R_1,R_3,R_4\}\}$
6	$\{R_1,R_2\}$	$\{R_3\}$	$\{\{R_1,R_2\},\{R_1,R_2,R_4\}\}$
7	$\{R_1,R_2,R_3\}$	$\{\}$	$\{\{R_1,R_2,R_3\},\{R_1,R_2,R_3,R_4\}\}$
8	$\{R_1,R_3\}$	$\{R_2,R_4\}$	$\{\{R_1,R_3\}\}$
9	$\{R_1,R_3,R_4\}$	$\{R_2\}$	$\{\{R_1,R_3,R_4\}\}$
10	$\{R_1,R_2\}$	$\{R_3,R_4\}$	$\{\{R_1,R_2\}\}$
11	$\{R_1,R_2,R_4\}$	$\{R_3\}$	$\{\{R_1,R_2,R_4\}\}$
12	$\{R_1,R_2,R_3\}$	$\{R_4\}$	$\{\{R_1,R_2,R_3\}\}$
13	$\{R_1,R_2,R_3,R_4\}$	$\{\}$	$\{\{R_1,R_2,R_3,R_4\}\}$

Table 2: 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 $\mathcal{B}_1$.

until $\mathcal{T}$-cover C_1^6. No further $\mathcal{T}$-refinements can be performed after C_1^6 since every element in C_1^6 is a ground segment event.

3.3 The Segmentation Sample Space ($\mathcal{S}$)

The Segmentation Sample Space is represented by the probability triple

$$\mathcal{S} = (\Pi, \mathcal{A}, P), \tag{6}$$

in which Π is the set of all segmentations that can be formed using $\mathcal{R}$, $\mathcal{A}$ represents the set

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: Enumeration of segments for the example image

$\mathcal{T}$-cover	R_ρ	B_ρ	Partition of Θ_1
C_1^0	-	-	$\{\{T_1, T_2, T_3, T_4, T_5, T_6, T_7\}\}$
C_1^1	R_2	Θ_1	$\{\{T_2, T_4, T_5, T_7\}, \{T_1, T_3, T_6\}\}$
C_1^2	R_3	$\{T_1, T_3, T_6\}$	$\{\{T_2, T_4, T_5, T_7\}, \{T_1\}, \{T_3, T_6\}\}$
C_1^3	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_1^4	R_4	$\{T_3, T_6\}$	$\{\{T_2, T_5\}, \{T_4, T_7\}, \{T_1\}, \{T_3\}, \{T_6\}\}$
C_1^5	R_4	$\{T_2, T_5\}$	$\{\{T_2\}, \{T_5\}, \{T_4, T_7\}, \{T_1\}, \{T_3\}, \{T_6\}\}$
C_1^6	R_4	$\{T_4, T_7\}$	$\{\{T_2\}, \{T_5\}, \{T_4\}, \{T_7\}, \{T_1\}, \{T_3\}, \{T_6\}\}$

Table 4: A sequence of $\mathcal{T}$-refinements that generates a $\mathcal{T}$-cover, C_1^6, which explicitly represents all segments.

of all subsets of Π (i.e., the power set of Π), and P denotes a probability mapping defined on $\mathcal{A}$. Since the size of Π grows at least exponentially with the size of $\mathcal{R}$, we will only construct approximations to $\mathcal{S}$, in a way similar to that used to construct approximate representations of $\mathcal{T}_i$. Furthermore, we will use approximations of the various $\mathcal{T}_i$ as the basic building blocks in the construction of these approximations to $\mathcal{S}$.

In the remainder of this section, we will develop the relationship of the probability distribution on a particular $\mathcal{T}_i$ to the probability distribution on $\mathcal{S}$, show how approximations to $\mathcal{T}_i$ can be used to construct representations of $\mathcal{S}$, and finally, introduce a method for approximating $\mathcal{S}$.

3.3.1 The segment-to-segmentation mapping

The relationship between a particular $\mathcal{T}_i$ and $\mathcal{S}$ is specified by the function $f_i : \mathcal{B}_i \rightarrow \mathcal{A}$. For a ground segment event, denoted by $\{T\}$, [1] we define f_i by

$$f_i(\{T\}) = \{S \in \Pi : T \in S\}. \tag{7}$$

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

$$f_i(\{T_1\}) \cap f_i(\{T_2\}) = \emptyset \quad \forall T_1, T_2 \in \Theta_i, \quad T_1 \neq T_2. \tag{8}$$

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

In other words, no single segmentation can contain two distinct segments that belong to the same Θ_i, since by the definition of Θ_i, such segments would overlap. Using (7) and (8), we define the mapping for a general event, $B \in \mathcal{B}_i$ as

$$f_i(B) = \bigcup_{T \in B} f(\{T\}) = \bigcup_{T \in B} \{S \in \Pi : T \in S\}, \tag{9}$$

in which both unions occur over disjoint sets. By applying f_i to each ground segment event of $\mathcal{T}_i$ we obtain a set of events that form a partition of Π, with each set in the partition corresponding to some segment from $\mathcal{T}_i$.

The relationship between Θ_i and Π resembles the *refinement* relationship defined by Shafer [57]. In fact, the mapping f_i is very similar to what Shafer terms a *refining mapping*; however, this should not be confused with our use of *refinement* in the context of creating new approximations for $\mathcal{T}_i$.

The relationship between the probability map on $\mathcal{T}_i$ and the probability map on $\mathcal{S}$ is such that the two probability maps coincide on events that are equivalent through f_i. To avoid confusion, in this section and in Section 3.3.2, we will use P_Π to denote the probability map on $\mathcal{S}$, and P_Θ to denote the probability map on $\mathcal{T}_i$. Explicitly, the probability assigned to a ground segment event in $\mathcal{T}_i$ is assigned directly to the corresponding event on $\mathcal{S}$

$$P_\Pi(f_i(\{T\})) = P_\Theta(\{T\}). \tag{10}$$

For an arbitrary event $B \in \mathcal{B}_i$ we use (9) to obtain

$$P_\Pi(f_i(B)) = \sum_{T \in B} P_\Theta(\{T\}) = P_\Theta(B) \tag{11}$$

which holds because ground segment events are disjoint.

Given a probability map defined on $\mathcal{T}_i$, we have constraints only for the corresponding map on $\mathcal{A}$, i.e., we can only determine probabilities for those elements of $\mathcal{A}$ that are in the image of $\mathcal{B}_i$ under f_i. This implies that $\mathcal{T}_i$ can be considered as an approximate representation of $\mathcal{S}$.

3.3.2 Building approximate representations of $\mathcal{S}$

In this section, we show how to construct approximate representations for $\mathcal{S}$ using approximate representations of various $\mathcal{T}_i$'s as building blocks. The idea is to piece together ground

segment events from a number of $\mathcal{T}_i$'s, until the entire image is covered. The probability mapping from ground segment events in $\mathcal{T}_i$ to events in $\mathcal{S}$, given by (11), is used to determine the corresponding probabilities for the events on $\mathcal{S}$. Construction of an approximation to $\mathcal{S}$ can proceed as follows:

1. Let $j = 1$, and let $\mathcal{R}_j = \mathcal{R}$.
2. Select some region, $R_{i_j} \in \mathcal{R}_j$, and construct an approximate representation for $\mathcal{T}_{i_j}$.
3. Refine the representation of $\mathcal{T}_{i_j}$, until a $\mathcal{T}$-cover, C, is obtained such that it includes at least some set, $X \subseteq C$, of ground segment events.
4. Select some segment $T \in X$ and construct $\mathcal{R}_{j+1} = \mathcal{R}_j - T$.
5. Repeat 2-4 until until $\mathcal{R}_j = \emptyset$, i.e., until the entire image has been covered by segments.

Although the algorithm above seems to imply that the $\mathcal{S}$ representation is built from explicit $\mathcal{T}_i$ representations, we show in Sections 3.3.3 and 5.2 that the $\mathcal{S}$ representation can be constructed by a sequence of $\mathcal{S}$ refinements, much in the same manner as for a $\mathcal{T}$ representation.

Given an approximate representation of $\mathcal{S}$, it is possible to compute the probabilities, P_Π, for events on $\mathcal{S}$ that are represented. Consider, for example, the case of two iterations of the above construction process. Let T_1 be a ground segment event from Θ_{i_1}, T_2 be a ground segment event from Θ_{i_2} and let $\mathcal{R}_3 = \mathcal{R} - T_1 - T_2$. We can compute the probability assigned to the event corresponding to all segmentations that contain both T_1 and T_2 as

$$P_\Pi(f_1(\{T_1\}) \cap f_2(\{T_2\})) = P_\Pi(f_2(\{T_2\})|f_1(\{T_1\}))P_\Pi(f_1(\{T_1\})). \tag{12}$$

Assuming that the probability maps on different $\mathcal{T}_i$'s are independent (i.e., there is no statistical dependency between different segments in the same segmentation), we have

$$P_\Pi(f_1(\{T_1\}) \cap f_2(\{T_2\})) = P_\Pi(f_2(\{T_2\}))P_\Pi(f_1(\{T_1\})). \tag{13}$$

Using one segment from each of n $\mathcal{T}_i$'s, and assuming that the probability maps on different $\mathcal{T}_i$'s are independent, we obtain

$$P_\Pi(f_1(\{T_1\}) \cap f_2(\{T_2\}) \cap \ldots \cap f_n(\{T_n\})) = P_\Pi(f_n(\{T_n\})) \cdots P_\Pi(f_2(\{T_2\}))P_\Pi(f_1(\{T_1\})). \tag{14}$$

Segmentation
$\{\{R_1, R_3, R_4\}, \{R_2\}\}$
$\{\{R_1, R_2, R_4\}, \{R_3\}\}$
$\{\{R_1, R_2, R_3\}, \{R_4\}\}$
$\{\{R_1, R_2, R_3, R_4\}\}$
$\{\{R_1, R_3\}, \{R_2, R_4\}\}$
$\{\{R_1, R_3\}, \{R_2\}, \{R_4\}\}$
$\{\{R_1, R_2\}, \{R_3\}, \{R_4\}\}$
$\{\{R_1, R_2\}, \{R_3, R_4\}\}$
$\{\{R_1\}, \{R_2, R_4\}, \{R_3\}\}$
$\{\{R_1\}, \{R_2, R_3\}, \{R_4\}\}$
$\{\{R_1\}, \{R_2, R_3, R_4\}\}$
$\{\{R_1\}, \{R_2\}, \{R_3\}, \{R_4\}\}$
$\{\{R_1\}, \{R_2\}, \{R_3, R_4\}\}$

Table 5: The possible segmentations for four-region example

Cases in which the probability maps on individual $\mathcal{T}_i$'s are not independent are discussed in [45]. This corresponds to a situtation in which additional model-based evidence could be used, causing a statistical dependency *between* segments (as opposed to strictly *within* segments).

For the example presented in Section 3.2.1, Table 5 indicates the set of segmentations that can be derived using this process. For this example, 5 $\mathcal{T}_i$'s were constructed, and all of the 13 possible segmentations are represented in the table.

3.3.3 Compact Representation of Events on $\mathcal{S}$, $\mathcal{S}$-Refinements, and $\mathcal{S}$-covers

In this section, we introduce a representation for events on $\mathcal{S}$ that is analogous to the τ representation introduced in Section 3.2 for events on $\mathcal{T}_i$. Following this, we define an $\mathcal{S}$-cover (which is analogous to a $\mathcal{T}$-cover), and the $\mathcal{S}$-refinement operation (which is analogous to $\mathcal{T}$-refinement).

Any event on $\mathcal{S}$ constructed in the manner presented in Section 3.3.2 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 in the sequence of $\mathcal{T}_i$ constructions. The sets I and E are

the include and exclude sets of $\tau(I, E)$, an event in the current $\mathcal{T}_i$ construction. We will use $(\Theta_i, \mathcal{B}_i, P)$ to denote the current $\mathcal{T}_i$. Formally, we represent an event on $\mathcal{S}$ by a function σ, in which

$$\sigma(F, I, E) = \{S \in \Pi : F \subseteq S\} \bigcap f_i(\tau(I, E)). \tag{15}$$

As defined in Section 3.3.2, f_i is the function that maps events in $\mathcal{B}_i$ to their corresponding events on $\mathcal{S}$. In this context, the corresponding event on $\mathcal{S}$ is

$$f_i(\tau(I, E)) = \bigcup_{T \in \tau(I,E)} \{S \in \Pi : T \in S\}. \tag{16}$$

Thus, (15) represents the set of segmentations that include every segment in F and exactly one segment from $\tau(I, E)$.

An *$\mathcal{S}$-cover*, $\mathbf{C}$, is a set of pairwise-disjoint events in $\mathcal{A}$ that form a partition of Π. As with $\mathcal{T}$-covers, there is a partial ordering on $\mathcal{S}$-covers; and, as with $\mathcal{T}$-covers, it is possible to construct a finer $\mathcal{S}$-cover from an existing $\mathcal{S}$-cover by performing an *$\mathcal{S}$-refinement operation*.

An $\mathcal{S}$-refinement is performed by partitioning the $\mathcal{S}$-refinement event, $A_\rho = \sigma(F_\rho, I_\rho, E_\rho)$, into two finer events, A_I and A_E. This is achieved by applying the $\mathcal{T}$-refinement operation to $\tau(I_\rho, E_\rho)$, for a $\mathcal{T}$-refinement region R_ρ. We impose the constraint that R_ρ be adjacent to some region in I_ρ, and that it not used in F_ρ or I_ρ. For the case in which $\tau(I_\rho, E_\rho)$ is a nonground segment event, the two $\mathcal{S}$-refined events are

$$A_I = \sigma(F_\rho, I_\rho \cup \{R_\rho\}, E_\rho) \tag{17}$$

and

$$A_E = \sigma(F_\rho, I_\rho, E_\rho \cup \{R_\rho\}). \tag{18}$$

This proposition implies that the replacement of A_ρ by A_I and A_E corresponds to a valid $\mathcal{S}$-refinment:

Proposition 2 *If $A_\rho = \sigma(F_\rho, I_\rho, E_\rho)$, and $\tau(I_\rho, E_\rho)$ represents a nonground segment event on $\mathcal{T}_i$, then A_I and A_E, given above, form a disjoint partition of A_ρ.*

It is possible, however, that $\tau(I_\rho, E_\rho)$ may be a ground segment event. In this case, the construction of a new $\mathcal{T}$ must be initiated. We select some region, R_j, that is not in any

of F_ρ, I_ρ, or E_ρ as the initial region for the new $\mathcal{T}$. It is convenient to use an equivalent representation for $\sigma(F_\rho, I_\rho, E_\rho)$, as given in the following proposition:

Proposition 3 *For an event on $\mathcal{S}$, $\sigma(F_\rho, I_\rho, E_\rho)$, in which $\tau(I_\rho, E_\rho)$ is a ground segment event on $\mathcal{T}_i$, and for some region R_j not in F_ρ or I_ρ,*

$$\sigma(F_\rho, I_\rho, E_\rho) = \sigma(F_\rho \cup \{I_\rho\}, \{R_j\}, \emptyset). \tag{19}$$

Although the proof is in the appendix, this equivalence can be seen by noting that $\tau(I_\rho, E_\rho)$ contains only one segment, $\{I_\rho\}$, and the expressions on both sides describe the set of all segmentations that contain the segment $\{I_\rho\}$ and the segments in F_ρ.

Using Proposition 3, when $\tau(I_\rho, E_\rho)$ is a ground segment event on $\mathcal{T}_i$, the $\mathcal{S}$-refined events are

$$A_I = \sigma(F_\rho \cup \{I_\rho\}, \{R_j, R_\rho\}, \emptyset) \tag{20}$$

and

$$A_E = \sigma(F_\rho \cup \{I_\rho\}, \{R_j\}, \{R_\rho\}). \tag{21}$$

Again, these represent a disjoint partition of A_ρ, and hence can be used in an $\mathcal{S}$-refinement.

3.4 Considering Priors

Bayesian approaches require the specification of prior distributions. Our goal when specifying prior distributions is to reflect uniformity, to avoid introducing prior bias. Alternatively, one might want to introduce a prior bias, for instance toward some number of segments per segmentation. In this section, we describe three possible specifications of prior distributions on $\mathcal{T}$ and $\mathcal{S}$ probability spaces. Each of these specifications corresponds to a particular definition of uniformity over $\mathcal{T}$ or $\mathcal{S}$. In Section 4, we discuss how the image models are applied to yield a posterior probability distribution.

The first kind of prior uniformity, termed *segmentation uniformity*, is the condition that all segments have equal prior probability, i.e.,

$$P(\{S\}) = \frac{1}{|\Pi|} \quad \forall S \in \Pi, \tag{22}$$

in which $|\Pi|$ is the number of possible segmentations. This appears to be the most natural definition of uniformity. The difficulty with segmentation uniformity 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.

The second kind of prior uniformity, which will be called *segment uniformity*, specifies that each segment in $\mathcal{T}$ has equal prior probability. Specifically, for a space Θ_i

$$P(\{T\}) = \frac{1}{|\Theta_i|} \quad \forall T \in \Theta_i. \tag{23}$$

Segment uniformity appears to be a natural choice; however, segment uniformity does not imply segmentation uniformity, except for the special case in which $f(\{T\})$ contains the same number of elements, for all $T \in \Theta_i$. This is implied by the probability constraint (11). Thus, in general, with respect to segmentation uniformity, segment uniformity can be considered as a kind of bias.

The third and final type of uniformity that we consider is *membership uniformity*. Membership uniformity reflects the assumption that for any $\mathcal{T}$-refinement, the prior probabilities associated with the two $\mathcal{T}$-refined events are equal. This corresponds to the assumption that the prior probability of including a region in a segment is equal to the prior probability of excluding that region from the segment. Membership uniformity does not imply segment uniformity, except in the special case when $|I \cup E|$ has the same number of regions for all ground segment events, $\tau(I, E) \in \mathcal{T}$. Note that this is not the case for the example of Section 3.2.1, as can be seen from Table 2.

Our experiments indicate that the bias due to priors is readily overcome when evidence is strong. 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 $\mathcal{T}$'s to be constructed than larger segments. The probabilities on $\mathcal{S}$ are obtained from these individual $\mathcal{T}$'s using (14). As the number of segments grows, the prior probability tends to decrease, compensating for the small-segment bias with respect to segment uniformity.

4 Bayesian Probability Assignments for Refinements

Although constructing $\mathcal{T}$-covers by $\mathcal{T}$-refinement has been specified structurally, no attention has yet been given to determining the probability assignments to the events that are created by $\mathcal{T}$-refinement. One primary issue must be considered: we are not given a complete representation of P on $\mathcal{T}$. This would require one probability assignment for every ground segment event.

Recall that each $\mathcal{T}$-refinement removes one event in a $\mathcal{T}$-cover of Θ_i and replaces it with two disjoint events whose union is the original event. The basic strategy in building a representation for some $\mathcal{T}$ is to determine probability assignments for 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 discussed in Section 3.4, there is some prior distribution on the sample space. Also, after the application of evidence, some posterior distribution is obtained. Model-based evidence will be used, along with the prior distribution, to determine probability assignments at the $\mathcal{T}$-refinement step. These issues will be discussed in the remainder of this section.

4.1 Refined-Event Probability Assignments

Using the $\mathcal{T}$-refinement mapping, ρ, successive partitions are constructed for Θ_i, as prescribed by (4). In order to perform a $\mathcal{T}$-refinement operation, we select an event $B_\rho = \tau(I_\rho, E_\rho)$ and a $\mathcal{T}$-refinement region R_ρ, and then partition $\tau(I_\rho, E_\rho)$ 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

$$P(\tau(I_\rho, E_\rho)) = P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) + P(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{24}$$

Prior to the first $\mathcal{T}$-refinement we have $\tau(I_\rho, E_\rho) = \tau(\{R_i\}, \emptyset) = \Theta_i$. Therefore, it is assumed inductively that $P(\tau(I_\rho, E_\rho))$ is known, and that the two probabilities on the right side of (24) must be determined using priors and model-based evidence.

We assume here that $P(\tau(I_\rho, E_\rho))$ is not altered by the $\mathcal{T}$-refinement operation. If this is not the case, it is possible to develop more general models for which $P(\tau(I_\rho, E_\rho))$ is affected

by considering evidence associated with R_ρ. For example, taxonomic hierarchies, analyzed by Pearl [51], allow this to occur. Pearl gives an efficient method for propagating evidence-based, posterior probabilities throughout a hierarchy of events, but the construction of the hierarchy is not considered. 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 [1],[8]. Sarkar and Boyer have proposed Bayesian networks for a hierarchical organization of perceptual features [56].

Since $P(\tau(I_\rho, E_\rho))$ is known inductively, only one of the terms on the right side of (24) is required, say $P(\tau(I_\rho \cup \{R_\rho\}, E_\rho))$. To assist in the application of model-based evidence, we provide the following decomposition:

Proposition 4 *For some refined event, $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$, its probability can be expressed as*

$$P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) = P(\tau(\{R_i, R_\rho\}, \emptyset) \mid \tau(I_\rho, E_\rho))\ P(\tau(I_\rho, E_\rho)). \tag{25}$$

We let, $P_I = P(\tau(\{R_i, R_\rho\}, \emptyset) \mid \tau(I_\rho, E_\rho))$, and refer to P_I as the *membership probability*. This is essentially the probability that R_ρ is a member of the maximal homogeneous segment that contains R_i, given that regions in I_ρ are members and regions in E_ρ are not. Because we will use the event $\tau(\{R_i, R_\rho\}, \emptyset)$ extensively throughout this section, we simplify its notation by letting $\tau_{i\rho} \equiv \tau(\{R_i, R_\rho\}, \emptyset)$. Note that the left hand side of (25) is expressed in a form that explicitly indicates the importance of adding R_ρ to I_ρ. This is the fundamental distinction between the event $\tau(I_\rho, E_\rho)$ and the $\mathcal{T}$-refined event. It is natural to expect that the probability due to evidence will depend directly on the new region that has been brought into consideration, which has been precisely represented by the right hand side of (25). Our membership probability can alternatively be considered as a definition of fuzzy membership, with respect to the segment that contains R_i [19],[43].

4.1.1 IE-independent and IE-dependent models

The probability $P(\tau(\{R_i, R_\rho\}, \emptyset) \mid \tau(I_\rho, E_\rho))$ in (25) depends on R_ρ, I_ρ and E_ρ. If a model uses information from all of these regions, it is termed *IE-dependent*. 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

$$P_I = P(\tau(\{R_i, R_\rho\}, \emptyset) \,|\, \tau(I_\rho, E_\rho)) = P(\tau(\{R_i, R_\rho\}, \emptyset)). \tag{26}$$

The choice between these models depends on the application. 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 $\mathcal{T}$-refinement region. If I_ρ and E_ρ are also considered, then a membership probability computation must be performed for each $\mathcal{T}$-refinement. It is not always be appropriate to use 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 in R_i. Typically, this will imply that better utilization of model evidence will be possible when R_i contains more data points. We note that in general, $\tau(\{R_i, R_\rho\})$ could be statistically dependent on any other region in $\mathcal{R}$; however, for computational reasons, at most the regions in I and E are considered.

4.2 Posterior Evidence-Based Membership Probability

In this section we present the expressions that determine the posterior membership probability for the IE-independent model. The more general, IE-dependent model derivations and expression can be found in Section 4.3.1 and in [45].

For each $R_k \in \mathcal{R}$ we associate the following: a *parameter space*, an *observation space*, a *degradation model*, and a *prior model* (see Table 6). 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 the observer. The observation space defines statistics that are functions of the image elements, and that contain information about the region's parameter value. We could use the image data directly for the observation, or could choose some function (possibly a sufficient statistic, depending on the application) that increases the efficiency of the Bayesian computations.

Although the parameter values are not known in general, a statistical model is introduced which uses two probability density functions (pdf's), forming 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, and can be considered as a model of image noise. Each of these concepts have been used in similar contexts for image segmentation. In fact we have borrowed the term *degradation model* from Geman [29]. 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* [63].

The transformation of the image elements in R_ρ and R_i yield observations of the random variables $\mathbf{Y}_\rho$ and $\mathbf{Y_i}$. These serve as the evidence used to determine the posterior membership probability, which is represented as $P(\tau(\{R_i, R_\rho\}, \emptyset)|\mathbf{y_i}, \mathbf{y}_\rho)$.

Proposition 5 *Given the observations* $\mathbf{y_i}$ *and* $\mathbf{y}_\rho$*, the posterior IE-independent membership probability is*

$$P(\tau(\{R_i, R_\rho\}, \emptyset)|\mathbf{y}_\rho, \mathbf{y_i}) = \frac{1}{1 + \lambda_0 \, \lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)}, \tag{27}$$

in which

$$\lambda_0 \equiv \frac{1 - P_0}{P_0} \tag{28}$$

and

$$\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) = \frac{\left[\int p(\mathbf{y_i}|\mathbf{u_i}) p(\mathbf{u_i}) \mathbf{du_i}\right] \left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho) p(\mathbf{u}_\rho) \mathbf{du}_\rho\right]}{\int p(\mathbf{y_i}|\mathbf{u_{i\rho}}) p(\mathbf{y}_\rho|\mathbf{u_{i\rho}}) p(\mathbf{u_{i\rho}}) \mathbf{du_{i\rho}}}. \tag{29}$$

Parameter space	A random vector, $\mathbf{U_k}$, which could, for instance, represent a space of polynomial surfaces.
Observation space	A random vector, $\mathbf{Y_k}$, which represents the data or functions of the data $\mathbf{x} \in R_k$.
Degradation model	A conditional density, $p(\mathbf{y_k}|\mathbf{u_k})$, which models noise and uncertainty.
Prior Model	An initial parameter space density, $p(\mathbf{u_k})$.

Table 6: The key components in our general statistical framework.

We denote the prior membership probability as P_0. The λ_0 and $\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)$ ratios represent a decomposition of the factors contributing to the membership probability: a prior factor and a posterior factor. The range of values of these ratios is restricted to $0 \leq \lambda_0 < \infty$ and $0 \leq \lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) < \infty$. When one of these ratios takes on the value of 1, it essentially does not bias the posterior membership probability.

The expression (29) has appeared recently in work from the statistics literature, and is termed a *Bayes factor*. Smith and Speigelhalter used a similar ratio for model selection between nested linear parametric models [59]. Aitken has developed a Bayes factor for model comparison that conditions the prior model on the data [2]. Kass and Vaidyanathan present and discuss some asymptotic approximations and sensitivity to varying priors of the Bayes factor [42]. Petit also discusses priors, but with concern for robustness with respect to outliers [52]. The Bayes factor has also been carefully studied for evidence evaluation in a forensic science context [6],[11],[21],[22]. Other references to Bayes factors include [5],[30],[41].

4.3 Other cases and extensions

In this section we briefly present some extensions to the result presented in the previous section. The expressions in Section 4.3.1 represent the more general IE-dependent model. Section 4.3.2 presents the expression obtained when multiple independent parameter spaces and observation spaces are considered. Finally, Section 4.3.3 indicates how the expressions apply to discrete random variable parameter and/or observation spaces. The resulting expressions from these three extensions are similar in appearance to the expressions in Proposition 5, and consequently for detailed derivations the reader is referred to [45].

4.3.1 IE-Dependent Membership Probability

With the IE-dependent model, the probability of homogeneity is expressed as

$$P_I = P(\tau_{i\rho} | \mathbf{y}_\rho, \mathbf{y_1}, \ldots, \mathbf{y_m}, \tau(I_\rho, E_\rho)). \tag{30}$$

This expression is similar to that for the IE-independent model. 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. 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

$$R_I \equiv \bigcup_{R_k \in I_\rho} R_k. \tag{31}$$

In other words there are $\mathbf{Y_I}$ and $\mathbf{U_I}$ for R_I, with a given conditional pdf $p(\mathbf{y_I}|\mathbf{u_I})$. The membership probability is then computed using (29) by replacing R_i with R_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.

Let $I_\rho = \{R_1, \ldots, R_m\}$ be the set of included regions. The observation variables 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. As shown in [45], the membership probability can be represented as

$$P(\tau_{i\rho}|\mathbf{y_1}, \ldots, \mathbf{y_m}, \mathbf{y}_\rho, \tau(I_\rho, E_\rho)) = \frac{1}{1 + \lambda_0\ \lambda_1(\mathbf{y_1}, \ldots, \mathbf{y_m}, \mathbf{y}_\rho)}. \tag{32}$$

in which

$$\lambda_1(\mathbf{y_1}, \ldots, \mathbf{y_m}, \mathbf{y}_\rho) = \frac{\left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}\right]\left\{\int \left[\prod_{\mathbf{k}=1}^{m} p(\mathbf{y_k}|\mathbf{u})\right] p(\mathbf{u})\mathbf{du}\right\}}{\int \left[\prod_{\mathbf{k}=1}^{m} p(\mathbf{y_k}|\mathbf{u})\right] p(\mathbf{y}_\rho|\mathbf{u})p(\mathbf{u})\mathbf{du}}. \tag{33}$$

This form is intuitively pleasing since it is nearly the same as (29) 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.

These two alternatives (creating one large region or considering individual observation spaces), in general, produce different results. 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.

4.3.2 Multiple Independent Models

Since a Bayesian model is used to determine membership probability, 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 shown only for the IE-independent case. The results carry over to the IE-dependent model in a straightforward manner. Consider some set of models, each specified as in Section 4.2. Take some region R_k. There are m models, each with its own parameter space random variables: $\mathbf{U_k^1}, \ldots, \mathbf{U_k^m}$. The superscripts denote different parameter spaces. Also, consider observation variables $\mathbf{Y_k^1}, \ldots, \mathbf{Y_k^m}$, with each $\mathbf{y_k^j}$ corresponding to relevant observations about the parameter space $\mathbf{U_k^j}$.

For each of the m models we receive an observation, yielding the values $\mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m}$ 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

$$P(\tau_{i\rho} | \mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m}, \mathbf{y_i^1}, \ldots, \mathbf{y_i^m}). \tag{34}$$

By using these definitions, the membership probability expression becomes

$$P(\tau_{i\rho} | \mathbf{y}_\rho, \mathbf{y_\rho^1}, \ldots, \mathbf{y_\rho^m}, \mathbf{y_i^1}, \ldots, \mathbf{y_i^m}) = \frac{1}{1 + \lambda_0 \prod_{l=1}^{m} \lambda_l(\mathbf{y}_\rho^l, \mathbf{y}_i^l)}. \tag{35}$$

For the case in which $m = 1$ this specializes to be (27) for the single IE-independent model.

Again, this framework has provided a decomposition of the evidence. The prior component λ_0 behaves as before. 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.

4.3.3 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. The discrete cases are merely notational variants of the previously derived expressions. For instance, with a single IE-independent model with discrete-valued parameter and observation spaces, we have

$$\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) = \frac{\left[\sum_U P(\mathbf{y_i}|\mathbf{u})P(\mathbf{u})\right]\left[\sum_U P(\mathbf{y}_\rho|\mathbf{u})P(\mathbf{u})\right]}{\sum_U P(\mathbf{y_i}|\mathbf{u})P(\mathbf{y}_\rho|\mathbf{u})P(\mathbf{u})}. \tag{36}$$

5 Algorithms to Construct $\mathcal{T}_i$ and $\mathcal{S}$ Representations

The algorithms used to construct approximate representations for $\mathcal{T}_i$'s, and ultimately for $\mathcal{S}$, are primarily implementations of the concepts presented in the preceding sections. In this section, we describe a number of the implementation details related to these algorithms. More detailed descriptions of the algorithms can be found in [45].

As described in Section 3.3.2, an approximate representation of $\mathcal{S}$ is built from ground segment events of approximate representations of $\mathcal{T}_i$. Therefore, in Section 5.1, we describe an algorithm for generating an approximate representation of $\mathcal{T}_i$ in which the n ground segment events that have the highest probability are explicitly represented. Then, in Section 5.2, we describe two methods for generating approximate representations for $\mathcal{S}$. In the first, beam-search is used to generate segmentations. In the second, an explicit representation of the n segmentations that have the highest probability is generated (the latter algorithm is analogous to the algorithm described in Section 5.1).

5.1 Constructing Representations of $\mathcal{T}_i$

As described in Section 3.2, successive approximations to $\mathcal{T}_i$ are constructed by repeated application of the $\mathcal{T}$-refinement operation. At each iteration, $\mathcal{T}$-refinement requires the selection of a $\mathcal{T}$-refinement region, R_ρ, and an event, B_ρ. For a $\mathcal{T}$-cover C, our algorithm selects for $\mathcal{T}$-refinement the element $B \in C$ that has highest probability, since it is most likely to contain ground segment events with larger probabilities. The $\mathcal{T}$-refinement region, R_ρ, which depends on the choice for B_ρ, must also be selected. In general, we prefer to use regions that contain the most information early in the computation. The corresponding membership probability will be close to either zero or one. A membership probability close to zero will assign low probability to the event $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$. When the probability is low, this new event is unlikely to be considered for subsequent $\mathcal{T}$-refinement. The complementary event, $\tau(I_\rho, E_\rho \cup \{R_\rho\})$, will have high probability, and is likely to be selected for further $\mathcal{T}$-refinement. An analogous situation occurs when the membership probability is close to one. We have observed through experimentation that the choice of smaller, less-informed regions early in the computation quickly leads to numerous alternatives, hence regions are sorted by size as candidates for B_ρ.

Recall that a $\mathcal{T}$-cover, C, represents a set of events on $\mathcal{T}_i$ that partition Θ_i. Typically, a $\mathcal{T}$-cover will contain both ground and nonground segment events. At any stage of the algorithm, the current $\mathcal{T}$-cover, C, can be partitioned as

$$C = C_g \cup C_h \qquad \text{and} \qquad C_g \cap C_h = \emptyset, \tag{37}$$

in which C_g contains the ground segment events of C, and C_h contains the remaining, nonground segment events of C. For any ground segment event $\{T\}$ such that $T \in B$ and $B \in C_h$, the posterior probability associated with $\{T\}$ can be no greater than the posterior probability associated with B. This observation is the basis for the termination criterion for our approximation algorithm. This is expressed as

$$P(nth(C_g)) \quad \geq \quad \max_{B \in C_h} P(B), \tag{38}$$

in which $nth(C_g)$ represents the event in C_g that has n^{th} highest probability. When this

condition is met, the representation explicitly contains the n segments that have highest probability in Θ_i.

5.2 Constructing Approximate Representations of $\mathcal{S}$

We have developed two algorithms for deriving approximate representations for $\mathcal{S}$. The first uses a beam-search algorithm [66], and the second generates the n segmentations with the highest posterior probability.

For both $\mathcal{S}$ algorithms, it is necessary, at various stages, to choose a new initial region, R_i, to begin the construction of a new $\mathcal{T}_i$. In our implementation, both algorithms select the largest available region as the new R_i. The motivation for this is that regions containing more "information" tend to cause more extreme values for $\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)$ (tending toward zero or infinity), yielding more compact representations of $\mathcal{T}_i$. Although region size is not a formal measure of information content, in our experiments (reported in Section 6 and in [45]) we have observed that when $\mathcal{T}_i$'s are constructed with small initial regions, the resulting event probabilities tend to remain close to their prior values, and many competing ground segment events on $\mathcal{T}_i$ are obtained.

The beam-search algorithm for constructing an approximate representation of $\mathcal{S}$ begins by generating a set of approximate representations of $\mathcal{T}_i$, each of which explicitly represents the b ground segment events with the highest probabilities. Thus, b is similar to the beam width in a traditional beam-search algorithm. The algorithm then finds the n best segmentations that can be constructed using only those ground segment events that have been derived. When $b = 1$, this algorithm produces a single segmentation in a "greedy" manner.

The second algorithm generates the n best segmentations by performing successive $\mathcal{S}$-refinement operations. Analogous to the partition of $\mathcal{T}$-covers described in Section 5.1, we can partition an $\mathcal{S}$-cover, $\mathbf{C}$, as

$$\mathbf{C} = \mathbf{C}_g \cup \mathbf{C}_h \quad \text{and} \quad \mathbf{C}_g \cap \mathbf{C}_h = \emptyset. \tag{39}$$

in which $\mathbf{C}_g$ contains events that correspond to individual full segmentations, and $\mathbf{C}_h$ contains the events of $\mathbf{C}$ that correspond to sets of segmentations. For any S such that $S \in A$ and

$A \in \mathbf{C}_h$, the posterior probability associated with S can be no greater than the posterior probability associated with A. This observation is the basis for the termination criterion for our approximate $\mathcal{S}$ algorithm. This is expressed by

$$P(nth(\mathbf{C}_g)) \quad \geq \quad \max_{A \in \mathbf{C}_h} P(A), \tag{40}$$

in which $nth(\mathbf{C}_g)$ represents the event in $\mathbf{C}_g$ that has n^{th} highest probability. When this condition is met, the representation explicitly contains the n segmentations that have highest probability in Π.

Since the second algorithm is able to guarantee that the best n segmentations have been represented, one might question the value of the beam-search algorithm. The beam-search algorithm offers the advantage of significantly improving space and time requirements, at the expense of guaranteeing that the best segmentations have been represented. For example, if $b = 3$, and a segmentation that has high probability contains a segment that is ranked fourth in its $\mathcal{T}_i$, then the beam-search algorithm will fail to return it. In practice we have found that the beam-search algorithm can yield reasonable results, particularly when there is a large number of regions, $\mathcal{R}$, and little information per region. Both algorithms have been useful in our experiments.

6 Experiments

In this section we present experimental results on both real and synthetic range data. Experiments on synthetic data were performed so that the effect of noise variance can be explicitly investigated. In this paper we present a few representative examples of the experiments that were performed; additional experimental results and comparisons can be found in [45]. In Section 6.1 we present results obtained using the approximation algorithm for $\mathcal{T}_i$, and in Section 6.2 we present results obtained using the $\mathcal{S}$ approximation algorithm. The evaluations of (29) that were needed for membership probability computations were obtained using numerical techniques presented in [47].

The algorithms were implemented in Common LISP on a SPARC IPC workstation. The execution times varied dramatically, from a few seconds in some cases, to a few hours in

others. Most of the execution time is devoted to computations of the membership probability. We have determined experimentally that the computational cost (in terms of time and space) increases both as the number of regions increases, and as the amount of information per region decreases. For the synthetic data, we have observed that computation time increases as the noise variance is increased (less information per region), because the number of reasonable alternatives rapidly increases. This corresponds to the intuition that the space of segmentations becomes increasingly underconstrained as the amount of uncertainty is increased.

6.1 $\mathcal{T}$ Representations

In this section we present experimental results that show $\mathcal{T}_i$ representations on some real range images and one synthetic image. Each approximate representation was obtained using the algorithm described in Section 5.1 to derive the 20 best ground segment events. For some results, we show only the first 8 of the 20 best ground segments events. 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 of two regions is homogeneous, there are many quadric surfaces that well-approximate one region and not the other, causing $\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)$ to be low.

The synthetic image consists of 10000 data points (100 x 100). Figure 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 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. Figure 2.(b) shows the range image after applying noise with $\sigma^2 = 0.1$. Figure 2.(c) shows the range image after applying noise with $\sigma^2 = 1.0$. Figure 3 shows the set of regions, $\mathcal{R}$, that was presented to the algorithms.

If we select R_{42} as the initial region and $\sigma^2 = 1.0$, we obtain the top twenty segments shown in Figure 4. Due to the high level of noise, the "correct" segment does not obtain

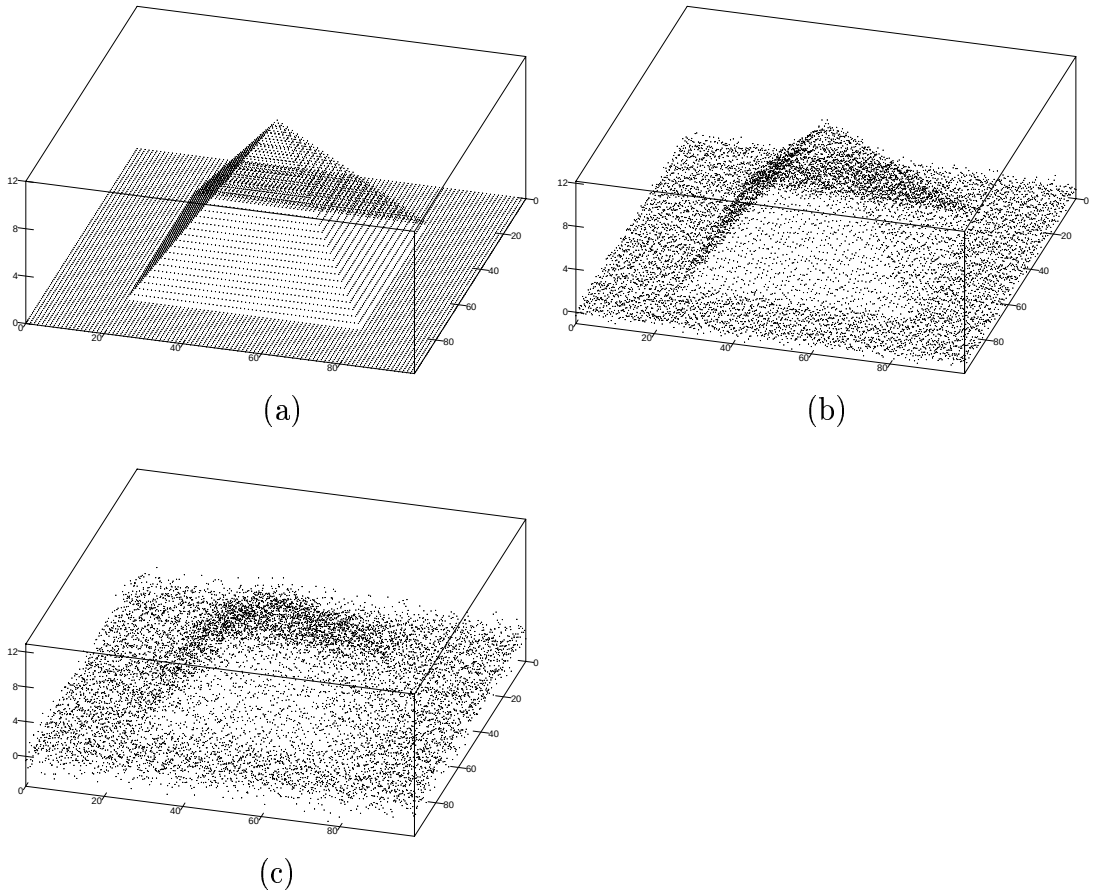


Figure 2: (a) The range data without noise; (b) the data with $\sigma^2 = 0.1$; (c) the data with $\sigma^2 = 1.0$

highest rank; however, due to the representation of alternatives it appears in the 8^{th} position. Figure 5 shows the result when R_{73} is selected as an initial region and $\sigma^2 = 0.1$. For this experiment, the noise level is significantly reduced, and for this problem the "correct" segment appears first and has very high probability. Hence we infer (as one would expect) that the image model is significantly more powerful when the noise level is low.

Figure 6 describes a set of range data of a polyhedral object with two initial regions that were used for experiments. For this range image we present an artificial rendering of the data so the reader can clearly see the object. The regions, $\mathcal{R}$, are obtained by combining the edge maps from edge detection and a recursive splitting algorithm which splits a region

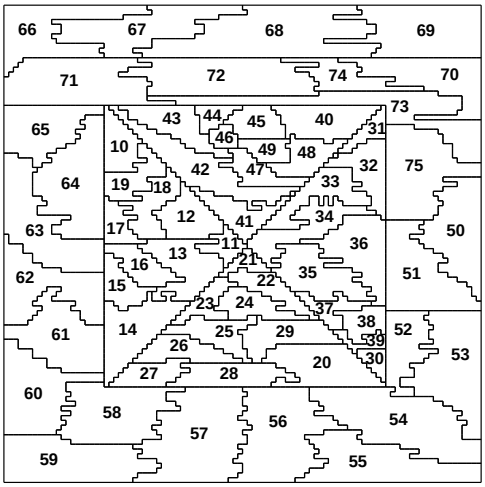


Figure 3: The set of regions, $\mathcal{R}$. This image shows the region boundaries projected into the x_1-x_2 plane, and the regions are labeled with integers for reference.

if the sum-of-squares error is too large using the optimal planar parameter estimate. The data sets are noisy, and the region maps that are presented contain many small regions that correspond to invalid data. This causes many specks to appear inside the segments. Figure 7 shows the resulting distribution using R_{264} as the initial region. Note that regions that are close to the planar boundaries tend to be excluded. This is due to the fact that these regions are not truly homogeneous. Most of the alternative segments are very similar because there are many small competing regions. Figure 8 shows a result in which small regions are excluded from consideration to emphasize the segment differences.

An image of a piecewise quadric object is presented in Figure 9. An intermediate step was performed when generating the initial regions, which merges regions whose union is planar with very high probability. The resulting set of segments using R_{783} as the initial region is shown in Figure 10.

6.2 $\mathcal{S}$ Representations

In this section we present three experimental results which show representations of $\mathcal{S}$, and individual segmentations on images that were used in Section 6.1. The algorithms presented in Section 5.2 are demonstrated here. The first two approximate representations of $\mathcal{S}$ correspond to generating the 20 best segmentations. The final approximate representation

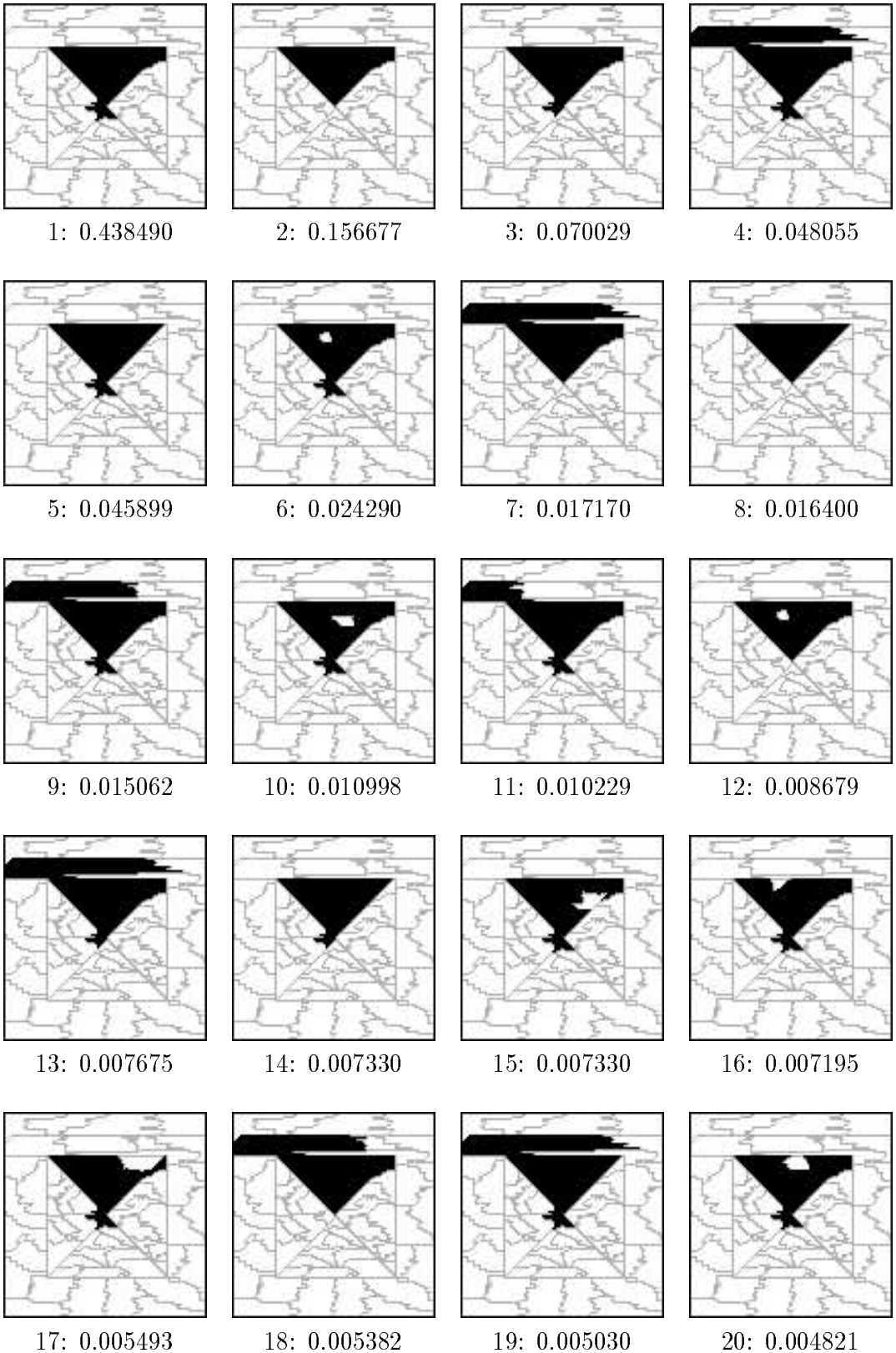


Figure 4: Twenty segments that have highest probability in Θ_{42}. There were 216 events in the final $\mathcal{T}$-cover, with 36 ground segment events. In this experiment, $\sigma^2 = 1.0$, and the IE-independent model is in use.

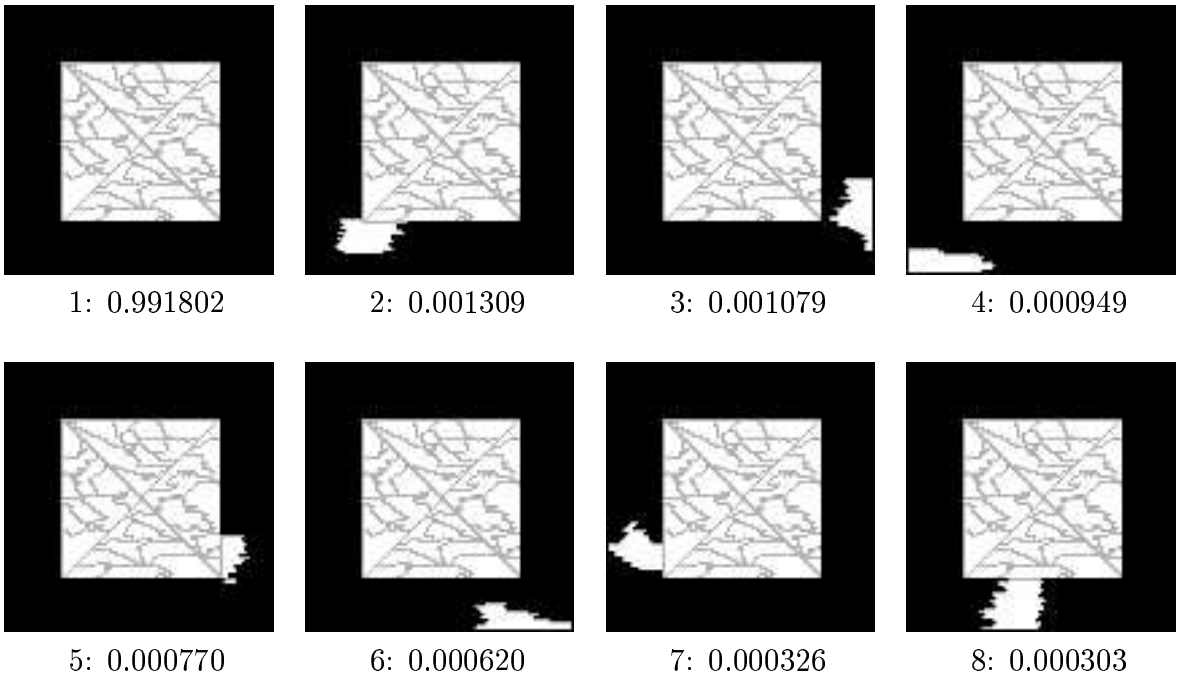


Figure 5: The first eight of the twenty ground segment events that were determined to have highest probability in Θ_{73}. There were 550 events in the final $\mathcal{T}$-cover, with 29 ground segment events. In this experiment, $\sigma^2 = 0.1$, and the IE-independent model is in use.

of $\mathcal{S}$ was obtained using the beam-search algorithm, with $b = 5$. The result in Figure 11 corresponds to little noise, and the "correct" segmentation is at the top of the ranking with many similar alternatives also appearing. The result in Figure 12 does not even represent the "correct" segmentation in the top twenty due to the high level of noise. The alternatives are, however, similar to the "correct" segmentation. Figure 13 shows twenty segmentations obtained on one of the real range images.

Figure 14 shows some compiled results after performing numerous experiments on the synthetic image. Figure 14.a shows how the probability distribution over the top twenty segmentations changes as the noise variance is increased from 0.01 to 1.0. For each value of the variance, ten trials were performed in which the synthetic image was regenerated each time according to the given noise model; the results presented in the figure were obtained by averaging over the trials. When the variance is small, the first segmentation (being the correct one in this instance) receives a probability near one. As the variance increases,

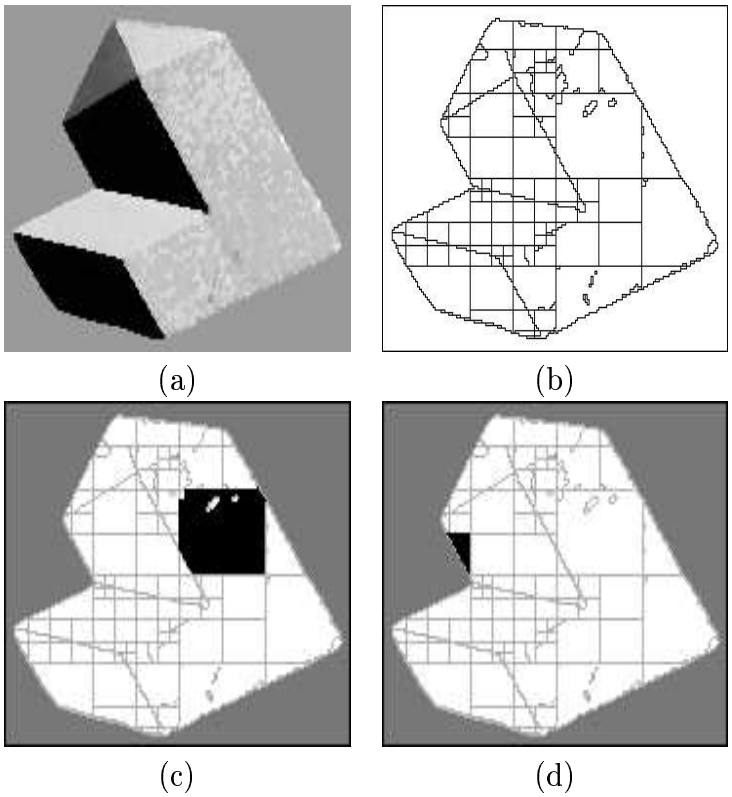


Figure 6: (a) A rendering of the data set, (b) the set of regions, $\mathcal{R}$, (c) initial region R_{264}, (d) initial region R_{224}.

the first segmentation receives much less probability, and the probability distribution is closer to being uniform. We have also observed that the entropy increases as the variance increases, quantifying the increase of uncertainty. In Figure 14.b we observe that the number of refinements performed (and hence computation time and space) increases as the variance increases, due to the consideration of more reasonable alternatives.

7 Conclusion

We have developed a general approach to constructing representations of probability distributions of image segments and segmentations, which are conditioned on statistical image models. From the experiments we conclude that segmentation need not be treated as an isolated process with an optimal solution, but can be considered practically as a set of low-level models that can yield a probability distribution over the space of alternatives. A higher-level

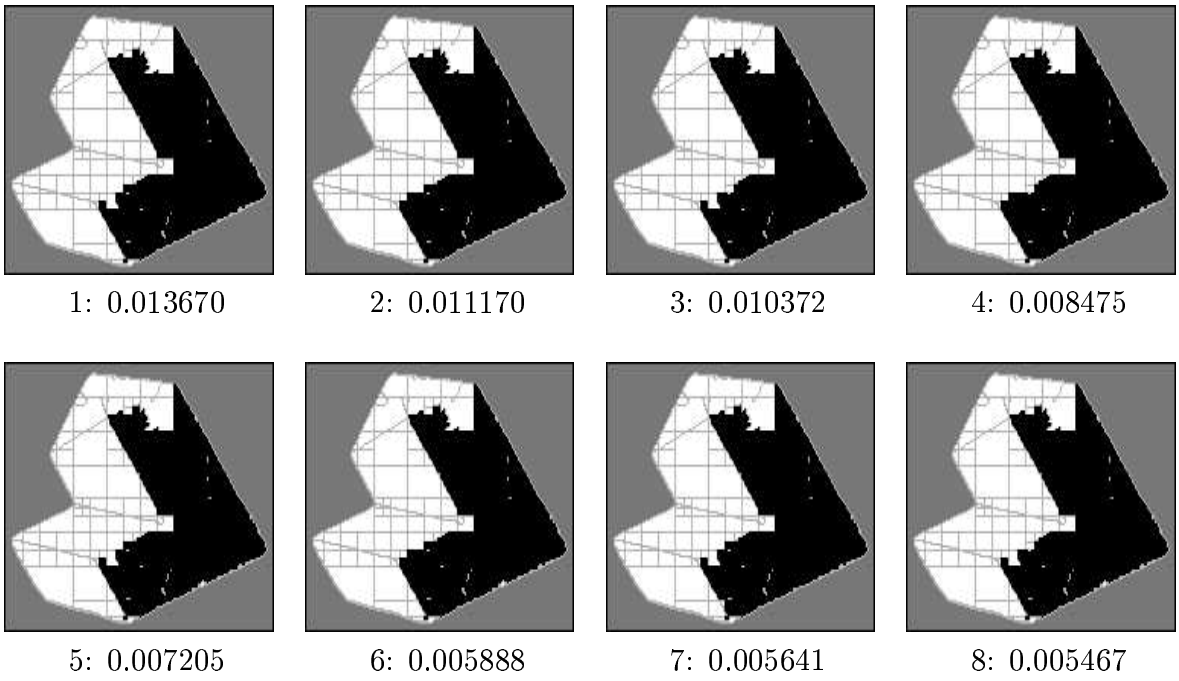


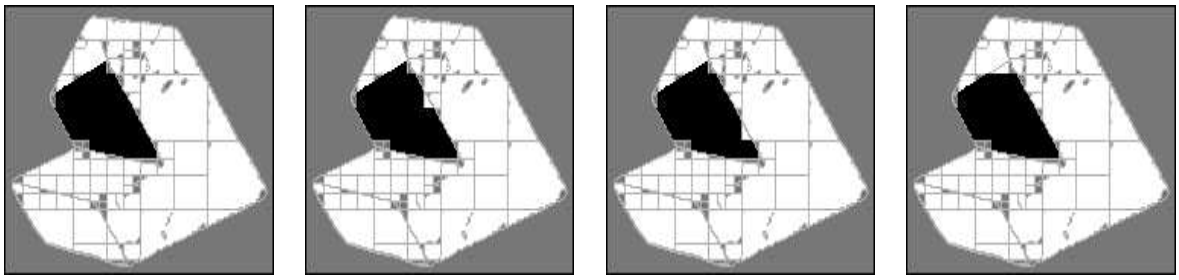
Figure 7: The first eight of the twenty ground segment events that were determined to have highest probability in Θ_{264} There were 2294 events in the final $\mathcal{T}$-cover, with 20 ground segment events. In this experiment, the IE-independent model is in use.

system can request alternative segmentations, or extensions could be made to incorporate higher-level or other additional models. A higher-level system can also measure the amount of information present in the image under the application of a statistical image model.

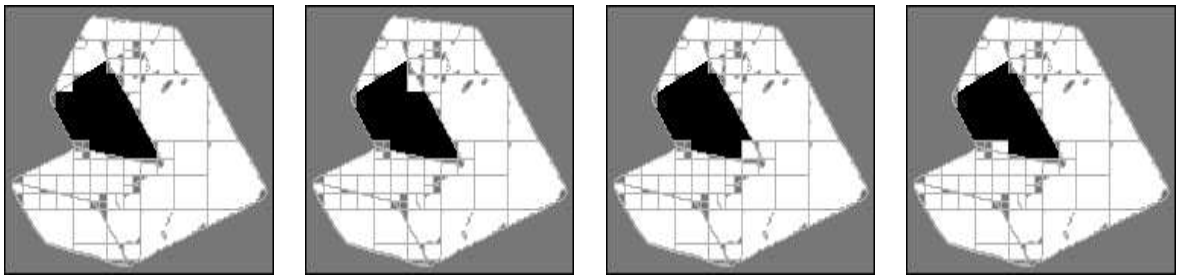
We hope that this contribution will change some of the focus in segmentation research toward the consideration of distributions of segments and segmentations and stronger Bayesian models, and away from the determination of single, optimal (or near-optimal) segmentations from underconstraining models. Progress toward this goal will require the ability to successfully integrate other types of image models into the Bayesian computations, which is presently being investigated.

Acknowledgements

We thank Pat Flynn and the MSU Pattern Recognition and Image Processing Lab for providing us with range images. John Mark Agosta, Becky Castaño, Didier Dubois, Ken



1: 0.785845 2: 0.060649 3: 0.042663 4: 0.034070



5: 0.014327 6: 0.009798 7: 0.009176 8: 0.004976

Figure 8: The first eight of the twenty ground segment events that were determined to have highest probability in Θ_{224} There were 255 events in the final $\mathcal{T}$-cover, with 28 ground segment events. In this experiment, the IE-independent model is in use. Regions, R_k, such that $|R_k| < 24$ have been removed.

Moroney, Kevin Nickels and Ross Shachter have provided helpful comments and suggestions regarding this work. We are also grateful for the useful comments provided by the anonymous reviewers. This work was sponsored by NSF under grant #IRI-9110270.

A Proofs of the Propositions

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). 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

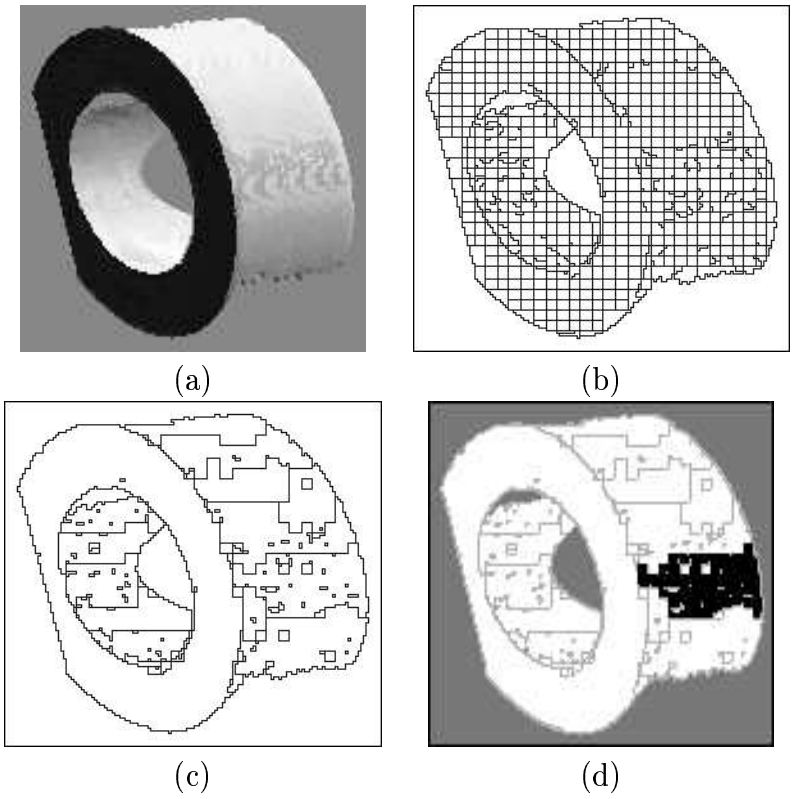


Figure 9: (a) The rendering of the data set; (b) the set of regions used in a preprocessing step; (c) the set of regions, $\mathcal{R}$; (d) The initial region.

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. $\square$

Proposition 2 If $A_\rho = \sigma(F_\rho, I_\rho, E_\rho)$, and $\tau(I_\rho, E_\rho)$ represents a nonground event on $\mathcal{T}_i$, then A_I and A_E, given by (17) and (18), form a disjoint partition of A_ρ.

Proof Using the definition of σ, (15), the intersection of A_I and A_E is

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap f_i(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) \bigcap f_i(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{41}$$

This set is empty since $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$ and $\tau(I_\rho, E_\rho \cup \{R_\rho\})$ represent events for $\mathcal{T}$-refinement (which are known to be disjoint), and the application of f_i to each of them yields disjoint sets of segmentations (recall (9)).

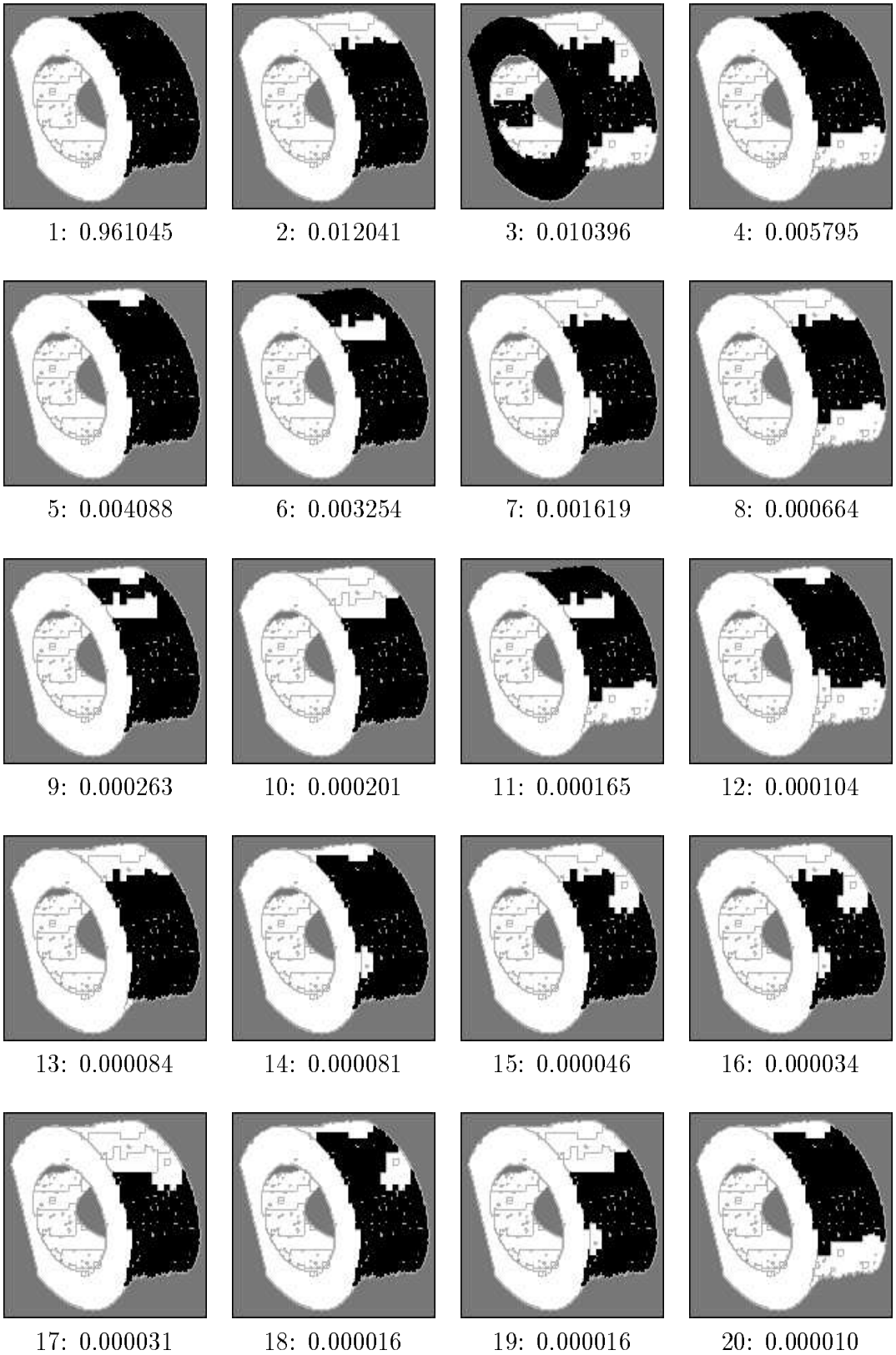


Figure 10: Twenty segments that have highest probability in Θ_{783}. There were 150 events in the final $\mathcal{T}$-cover, with 37 ground segment events. In this experiment, the IE-dependent model is in use.

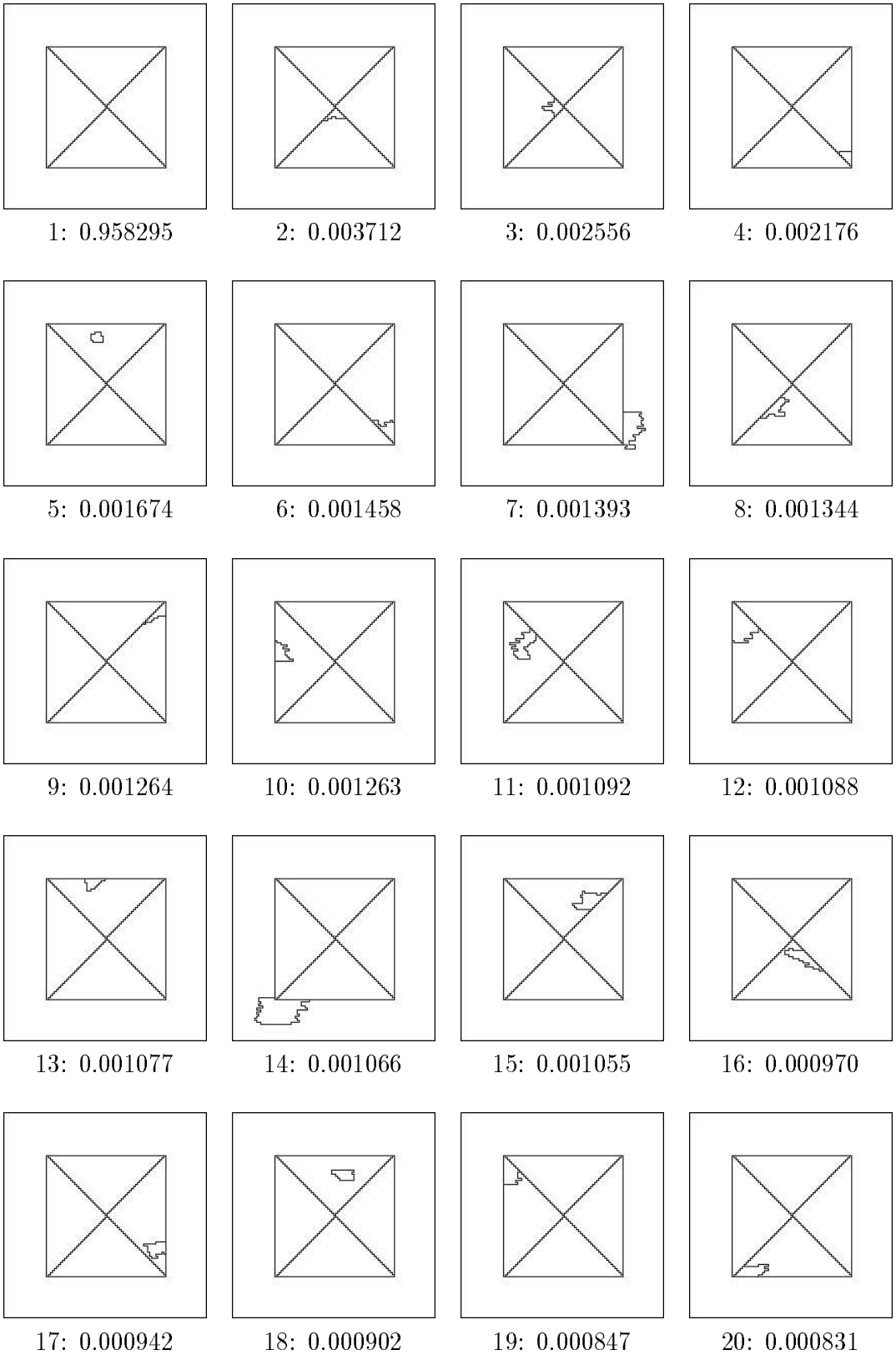


Figure 11: Twenty segmentations that have highest probability in Π. In this experiment, $\sigma^2 = 0.1$, and the IE-independent model is in use.

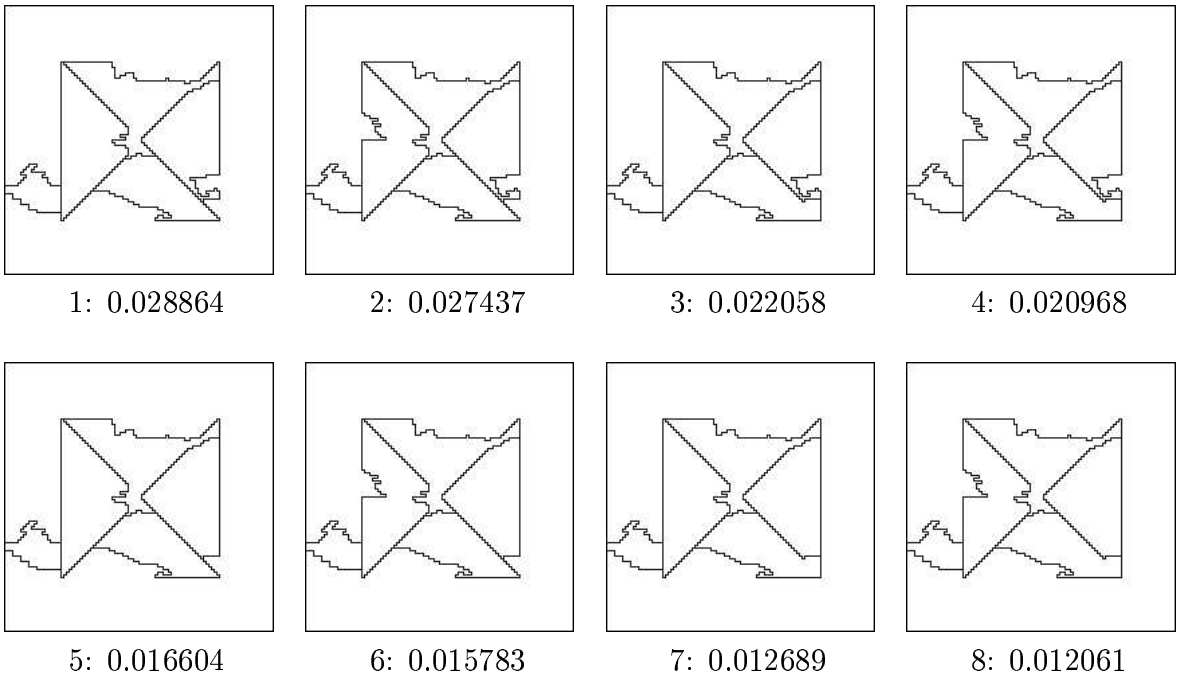


Figure 12: The first eight of the twenty segmentations that were determined to have highest probability in Π. In this experiment, $\sigma^2 = 1.0$, and the IE-independent model is in use.

Taking the union of A_I and A_E we obtain

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap \left[f_i(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) \bigcup f_i(\tau(I_\rho, E_\rho \cup \{R_\rho\})) \right]. \tag{42}$$

This is equivalent to

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap f_i(\tau(I_\rho, E_\rho)) = A_\rho \tag{43}$$

since

$$f_i(\tau(I_\rho, E_\rho)) = f_i(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) \cup f_i(\tau(I_\rho, E_\rho \cup \{R_\rho\})). \tag{44}$$

□

Proposition 3 For an event $\sigma(F_\rho, I_\rho, E_\rho)$, on $\mathcal{S}$, in which $\tau(I_\rho, E_\rho)$ is a ground segment event on $\mathcal{T}_i$, and for some region R_j not in F_ρ or I_ρ,

$$\sigma(F_\rho, I_\rho, E_\rho) = \sigma(F_\rho \cup \{I_\rho\}, \{R_j\}, \emptyset). \tag{45}$$

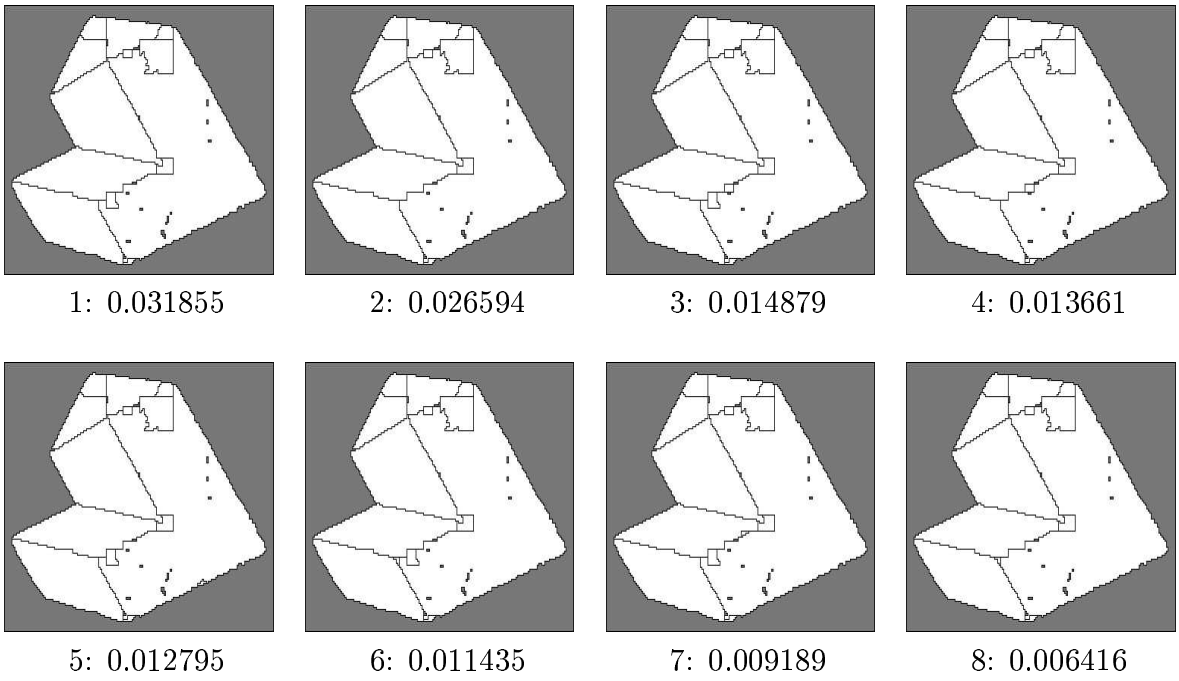


Figure 13: Eight segmentations that were obtained from a beam-search with $b = 5$. In this experiment, the IE-independent model is in use.

Proof On the right side, F_ρ represents a set of segments, and I_ρ and E_ρ represent an additional segment (a ground segment event on $\mathcal{T}_i$). 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 $\mathcal{T}_i$ with the include set, $\{R_j\}$, and exclude set, $\emptyset$.

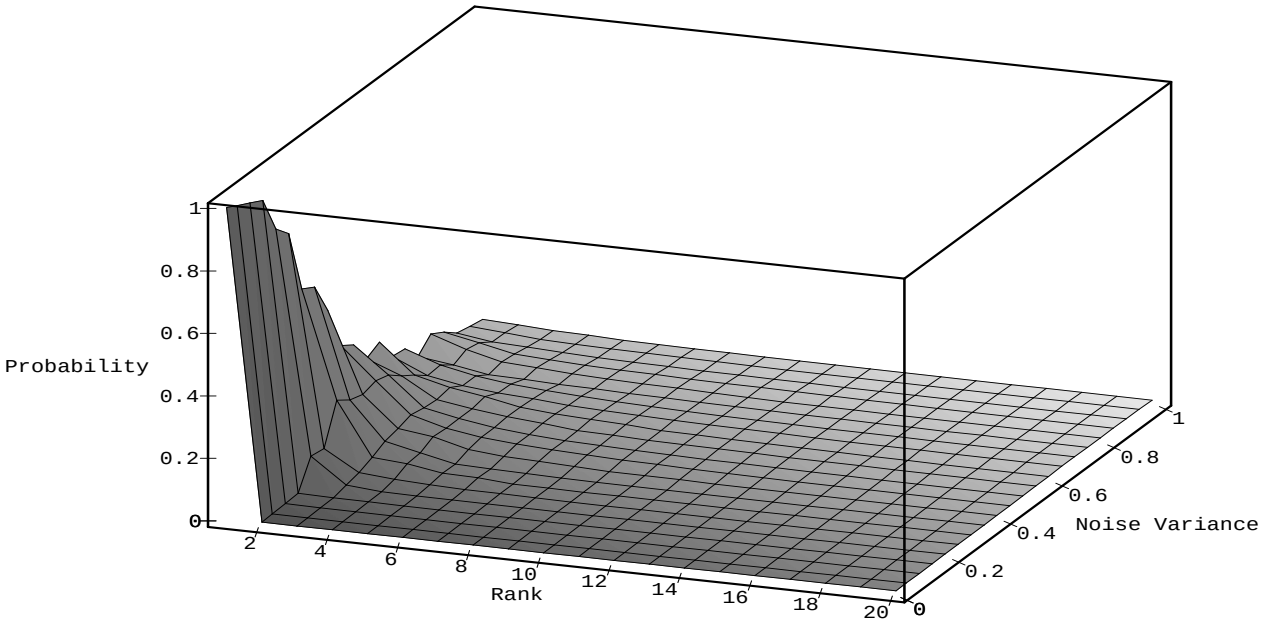
This equivalence becomes clearer when the definition of σ is applied to left and right sides of (45) to obtain

$$\{S \in \Pi : T \in S \; \forall T \in F_\rho\} \bigcap f_i(\tau(I_\rho, E_\rho)) \tag{46}$$

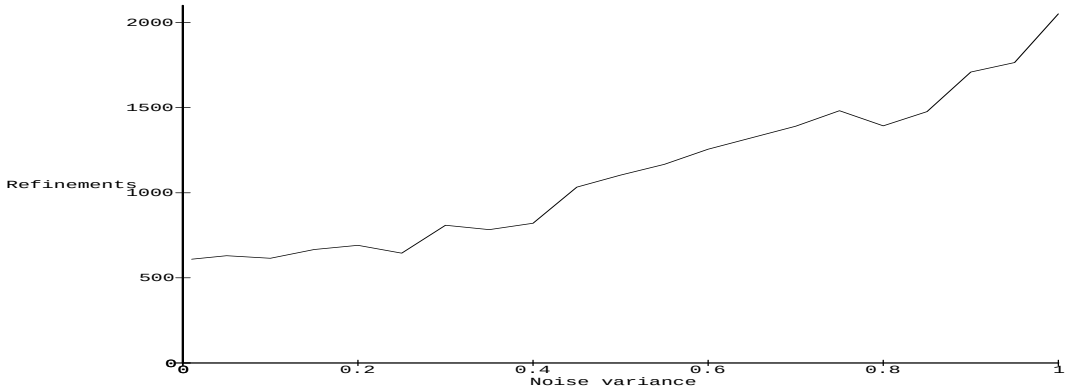
and

$$\{S \in \Pi : T \in S \; \forall T \in (F_\rho \cup \{I_\rho\})\} \bigcap f_i(\tau(\{R_j\}, \emptyset)), \tag{47}$$

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 segment 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



(a)



(b)

Figure 14: (a) The probability distribution over the top twenty segmentations, plotted against variance; (b) The number of refinments performed vs. increasing variance

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 event on $\mathcal{S}$. $\square$

Proposition 4 For some refined event, $\tau(I_\rho \cup \{R_\rho\}, E_\rho)$, its probability can be expressed as

$$P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) = P(\tau(\{R_i, R_\rho\},\emptyset) \,|\, \tau(I_\rho, E_\rho))\; P(\tau(I_\rho, E_\rho)). \tag{48}$$

Proof Since $\tau(I_\rho \cup \{R_\rho\}, E_\rho) \subset \tau(I_\rho, E_\rho)$, we have

$$P(\tau(I_\rho \cup \{R_\rho\}, E_\rho)) = P(\tau(I_\rho \cup \{R_\rho\}, E_\rho) \,|\, \tau(I_\rho, E_\rho))\; P(\tau(I_\rho, E_\rho)). \tag{49}$$

By using the following lemma with $I_1 = I_\rho$, $I_2 = \{R_i, R_\rho\}$, $E_1 = E_\rho$, $E_2 = \emptyset$, we have

$$P(\tau(I_\rho \cup \{R_\rho\}, E_\rho) \,|\, \tau(I_\rho, E_\rho)) = P(\tau(\{R_i, R_\rho\},\emptyset) \,|\, \tau(I_\rho, E_\rho)). \tag{50}$$

Lemma 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 $\tau(I, E) = \tau(I_1, E_1) \cap \tau(I_2, E_2)$. $\square$

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

$$\tau(I_1 \cup I_2, E_1 \cup E_2) = \{T \in \Theta_i : (I_1 \cup I_2) \subseteq T, (E_1 \cup E_2) \cap T = \emptyset\}. \tag{51}$$

This can equivalently be expressed as

$$\{T \in \Theta_i : I_1 \subseteq T, I_2 \subseteq T, E_1 \cap T = \emptyset, E_2 \cap T = \emptyset\}. \tag{52}$$

This is the same as

$$\{T \in \Theta_i : I_1 \subseteq T\} \cap \{T \in \Theta_i : I_2 \subseteq T\} \cap \{T \in \Theta_i : E_1 \cap T = \emptyset\} \cap \{T \in \Theta_i : E_2 \cap T = \emptyset\}. \tag{53}$$

But this is equivalent to

$$\{T \in \Theta_i : I_1 \subseteq T, E_1 \cap T = \emptyset\} \cap \{T \in \Theta_i : I_2 \subseteq T, E_2 \cap T = \emptyset\}, \tag{54}$$

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

Proposition 5 Given the observations $\mathbf{y}_\rho$ and $\mathbf{y_i}$, the posterior membership probability is

$$P(\tau_{i\rho} | \mathbf{y}_\rho, \mathbf{y_i}) = \frac{1}{1 + \lambda_0\; \lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)}, \tag{55}$$

in which

$$\lambda_0 = \frac{1 - P_0}{P_0} \tag{56}$$

and

$$\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) = \frac{\left[\int p(\mathbf{y_i}|\mathbf{u_i})p(\mathbf{u_i})\mathbf{du_i}\right]\left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho\right]}{\int p(\mathbf{y_i}|\mathbf{u_{i\rho}})p(\mathbf{y}_\rho|\mathbf{u_{i\rho}})p(\mathbf{u_{i\rho}})\mathbf{du_{i\rho}}}. \tag{57}$$

Proof In order to determine the probability of merging two regions, it will be necessary to consider a statement of the form $H(R_i \cup R_\rho) = true$, which corresponds to the condition that $R_i \cup R_\rho$ is homogeneous, and $H(R_i \cup R_\rho) = false$, which corresponds to the condition that $R_i \cup R_\rho$ is not homogeneous. We will use $\tau_{i\rho}$ to represent the condition $H(R_i \cup R_\rho) = true$, and $\tau_{i\rho}^C$ to represent $H(R_i \cup R_\rho) = false$. Note that if H is true then R_i and R_ρ share the same parameter value.

Next, we state that if two regions, R_i and R_ρ, share the same parameter value, $\mathbf{u_{i\rho}}$, which is given, then

$$p(\mathbf{y_i}, \mathbf{y}_\rho|\mathbf{u_{i\rho}}) = p(\mathbf{y_i}|\mathbf{u_{i\rho}})p(\mathbf{y}_\rho|\mathbf{u_{i\rho}}). \tag{58}$$

The region's parameter value (and not observations from neighboring regions) is all that is needed to predict the observation. This is equivalent to asserting that nothing is learned about the degradation model when observations are made from other regions having the same given parameter value.

We next make the assumption that the observations $\mathbf{Y_i}$ and $\mathbf{Y}_\rho$ are conditionally independent, given that $R_i \cup R_\rho$ is not homogeneous. Formally, this is stated as

$$p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho}^C) = p(\mathbf{y_i})p(\mathbf{y}_\rho). \tag{59}$$

We note that this assumption is not necessary, which is further discussed in [45].

The observations serve as the evidence used to determine the Bayesian probability of homogeneity, which is represented as $P(\tau_{i\rho}|\mathbf{y_i}, \mathbf{y}_\rho)$. We can apply Bayes' rule to obtain

$$P(\tau_{i\rho}|\mathbf{y_i}, \mathbf{y}_\rho) = \frac{p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho})P(\tau_{i\rho})}{p(\mathbf{y_i}, \mathbf{y}_\rho)} \tag{60}$$

$$= \frac{p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho})P(\tau_{i\rho})}{p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho})P(\tau_{i\rho}) + p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho}^C)P(\tau_{i\rho}^C)}. \tag{61}$$

The denominator of (61) is the standard normalizing factor from Bayes' rule, over the binary sample space, $\{\tau_{i\rho}, \tau_{i\rho}^C\}$. The expression $P(\tau_{i\rho})$ represents the prior probability of homogeneity, i.e., the probability that two adjacent regions should be merged, when $\mathbf{y_i}$ and $\mathbf{y}_\rho$ have not been observed, and in practice we usually take $P(\tau_{i\rho}) = P(\tau_{i\rho}^C) = 1/2$. This represents a uniform distribution over the binary sample space.

We can write (61) as

$$P(\tau_{i\rho}|\mathbf{y_i}, \mathbf{y}_\rho) = \frac{1}{1 + \lambda_0\ \lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)} \tag{62}$$

in which

$$\lambda_0 = \frac{1 - P(H)}{P(H)} \quad \text{and} \quad \lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) = \frac{p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho}^C)}{p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho})}. \tag{63}$$

Substituting (59) into the expression for $\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho)$ in (28), we obtain

$$\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) = \frac{p(\mathbf{y_i})p(\mathbf{y}_\rho)}{p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho})}. \tag{64}$$

The condition $\tau_{i\rho}$ is equivalent to asserting that $\mathbf{u_i} = \mathbf{u}_\rho$. Using a common prior density $p(\mathbf{u_{i\rho}})$, which is equal to both $p(\mathbf{u_i})$ and $p(\mathbf{u}_\rho)$, we can write the denominator of (64) as a marginal with respect to $U_{i\rho}$:

$$p(\mathbf{y_i}, \mathbf{y}_\rho|\tau_{i\rho}) = \int p(\mathbf{y_i}, \mathbf{y}_\rho|\mathbf{u_{i\rho}})p(\mathbf{u_{i\rho}})\mathbf{du_{i\rho}}. \tag{65}$$

Using (58) in the denominator, and using the marginal over $\mathbf{U_k}$ for each term of the numerator, we obtain:

$$\lambda_1(\mathbf{y_i}, \mathbf{y}_\rho) = \frac{\left[\int p(\mathbf{y_i}|\mathbf{u_i})p(\mathbf{u_i})\mathbf{du_i}\right]\ \left[\int p(\mathbf{y}_\rho|\mathbf{u}_\rho)p(\mathbf{u}_\rho)\mathbf{du}_\rho\right]}{\int p(\mathbf{y_i}|\mathbf{u_{i\rho}})p(\mathbf{y}_\rho|\mathbf{u_{i\rho}})p(\mathbf{u_{i\rho}})\mathbf{du_{i\rho}}}. \tag{66}$$

□

B Implicit-Polynomial Surface Model Expressions

In this section we briefly present the polynomial formulation of the membership probability that we used for our experiments. A more complete discussion can be found in [45].

For this model, an image element is represented by a point in $\Re^3$, specified by $\mathbf{x} = [x_1, x_2, x_3]$. We will describe points that belong to some region R_k by $\mathbf{x} \in R_k$. This model

declares that all points in a region R_k came from the same polynomial surface patch, with some noise occurring in the observation process. The segmentation goal in this context is to determine maximal connected sets of regions that belonged to the same polynomial surface (before noise was applied).

B.1 The parameter manifold

We will now introduce a 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 [64]. A polynomial can be considered as a linear combination of monomial basis functions.

An implicit polynomial equation is represented as

$$\phi(\cdot, \mathbf{u}) = \sum_{j=1}^{N} u_j x_1^{a_j} x_2^{b_j} x_3^{c_j} = 0 \tag{67}$$

with

$$a_N = b_N = c_N = 0. \tag{68}$$

The constants a_j, b_j, and c_j are integers, representing the exponents of each variable. The $\cdot$ used here indicates that we have an implicit function with $\mathbf{x}$ as the variables. In later expressions we will refer to $\phi(\mathbf{x}, \mathbf{u})$, which yields a nonzero value 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 present formulation, there are redundant representations of the solution sets (i.e., there are many parameter vectors that describe the same surface in $\Re^3$). It is profitable to choose some restriction of the parameter space that facilitates the integrations in (29), but maintains full expressive power. We use the constraints, $\|\mathbf{u}\| = 1$ and $u_1 > 0$, to constrain the parameter space to a half-hypersphere, Σ^N, termed the *parameter manifold*.

B.2 The observation space

The observation considered here is a function of the signed distances of the points $\mathbf{x} \in R_k$ from the surface determined by $\mathbf{u_k}$, termed as *displacements*. Define $\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))$ to be the

displacement of the point $\mathbf{x}$ to 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.

We consider the following observation space definition, and others are discussed in [45]:[2]

$$y_k(R_k, \mathbf{u_k}) = \sum_{\mathbf{x} \in R_k} [\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))]^2. \tag{69}$$

This function of the displacements is often used for polynomial parameter estimation.

Although we have defined the observation space in terms of the displacements, a closed-form expression for the displacement of a point to a polynomial surface does not exist in general. We use a displacement estimate presented by Taubin and Cooper [65]:

$$\hat{\delta}(\mathbf{x}, \phi(\cdot, \mathbf{u_k})) = \frac{\phi(\mathbf{x}, \mathbf{u_k})}{\|\nabla_x \phi(\cdot, \mathbf{u_k})\|}. \tag{70}$$

B.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. We use a probability model for range-scanning error, used and justified by Bolle and Cooper [9], and also used by Taubin [64]. The model asserts that density, $p(\delta|\mathbf{u})$, of the displacement of an observed point from the surface, $\phi(\mathbf{x}, \mathbf{u})$, is a Gaussian random variable with zero mean and some known variance, σ^2. This degradation model is merely chosen as a representative of possible models that can be used. 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 [35].

Since taking the sum of squares of Gaussian densities yields the chi-square density, the degradation density, for unit variance, using (69) is

$$p(y_k|\mathbf{u_k}) = \chi^2_{m_k}(y_k) = \frac{1}{2^{m_k/2}\Gamma(m_k/2)} y_k^{m_k/2-1} e^{-y_k/2}. \tag{71}$$

Here y_k is the sum-of-squares for a given region, R_k, and parameter value $\mathbf{u_k}$, given by (69). Also, $\Gamma(\cdot)$ is the standard gamma function and $m_k = |R_k|$ (the number of elements in R_k).

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

B.4 The prior model

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 density on the space of implicit surfaces. If other constraints were used on the parameter space, and we assumed a constant-valued pdf, the density 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 just the inverse of the surface area of the half hypersphere that defines the parameter manifold, which is straightforward to compute. The prior model is $p(\mathbf{u_k}) = A_N^{-1}$, in which A_N represents the area of the N parameter manifold. This quantity can be determined through a straightforward integral transformation [61].

References

[1] 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*, pages 16–27, San Diego, CA, July 1991.

[2] M. Aitkin. Posterior Bayes factors. *J. Royal Statistical Society*, B53(1):111–142, 1991.

[3] R. B. Ash. *Information Theory.* Dover Publications, New York, NY, 1990.

[4] Z. W. Bell. A Bayesian/Monte Carlo segmentation method for images dominated by Gaussian noise. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(9):985–989, September 1989.

[5] J. O. Berger and M. Delampady. Testing precise hypotheses (with discussion). *Statist. Sci.*, 2:317–335, 1989.

[6] D. A. Berry, I. W. Evitt, and R. Pinchin. Statistical inference in crime investigations using deoxyribonucleic acid profiling. *J. Royal Statistical Society*, C41:499–531, 1992.

[7] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statistical Society*, B48:259–302, 1986.

[8] 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*, pages 86–92, 1987.

[9] 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.*, 8(5):619–638, September 1986.

[10] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Wadsworth, Belmont, CA, 1984.

[11] K. P.-S. Chan and C. G. G. Aitkin. Estimation of the bayes' factor in a forensic science problem. *J. Statist. Comput. Simul.*, 33:249–264, 1991.

[12] P. A. Chou. Optimal partitioning for classification and regression trees. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(4):340–354, April 1991.

[13] P. B. Chou and C. M. Brown. The theory and practice of Bayesian image labeling. *Int. J. Comput. Vis.*, 4:185–210, 1990.

[14] 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.*, 9(2):195–219, March 1987.

[15] M. Daily. Color image segmentation using Markov Random Fields. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 304–312, June 1989.

[16] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(1):39–55, January 1987.

[17] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach.* Prentice/Hall Publications, Englewood Cliffs, NJ, 1982.

[18] R.C. Dubes, A.K. Jain, S.G. Nadabar, and C.C Chen. MRF model-based algorithms for image segmentation. In *Proc. IEEE 10th Int. Conf. on Pattern Recognition*, pages 808–814, Atlantic City, NJ, June 1990.

[19] D. Dubois and M.-C. Jaulent. A general approach to parameter evaluation in fuzzy digital pictures. *Pattern Recognition Letters*, 6:251–259, September 1987.

[20] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* Wiley, New York, NY, 1973.

[21] I. W. Evitt. A quantative theory for interpreting transfer evidence in criminal cases. *Applied Ststistics*, 33:25–32, 1984.

[22] I. W. Evitt, P. E. Cage, and C. G. G. Aitkin. Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases. *Applied Ststistics*, 36:174–180, 1987.

[23] O. D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *Int. J. of Robot. Res.*, 5(3):27–52, Fall 1986.

[24] K. Fukunaka. *Introduction to Statistical Pattern Recognition.* Academic Press, New York, NY, 1972.

[25] D. Geiger and F. Girosi. Parallel and deterministic algorithms from MRF's: Surface reconstruction. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(5):401–412, May 1991.

[26] 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.*, 13(2):163–174, February 1991.

[27] D. Geman and S. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721–741, November 1984.

[28] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):609–628, July 1990.

[29] S. Geman. Experiments in Bayesian image analysis. *Bayesian Statistics*, 3:159–172, 1988.

[30] I. J. Good. Weight of evidence: A brief survey. *Bayesian Statistics*, 2:249–270, 1985.

[31] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Comp. Vision, Graphics, and Image Process.*, 29:100–132, January 1985.

[32] B. K. P. Horn. *Robot Vision.* MIT Press, Cambridge, MA, 1986.

[33] J. Y. Hsi and A. A. Sawchuk. Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(12):1279–1292, December 1989.

[34] J. Y. Hsi and A. A. Sawchuk. Unsupervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. *Comp. Vision, Graphics, and Image Process.*, 48:1–21, 1989.

[35] K. Ikeuchi and T. Kanade. Modeling sensors: Toward automatic generation of object recognition program. *Comp. Vision, Graphics, and Image Process.*, 48:50–79, 1989.

[36] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data.* Prentice Hall, Inc., Englewood Cliffs, NJ, 1988.

[37] A. K. Jain and S. G. Nadabar. MRF model-based segmentation of range images. In *Proc. Int. Conf. on Computer Vision*, pages 667–671, 1990.

[38] R. C. Jain and T. O. Binford. Ignorance, myopia, and naiveté in computer vision systems. *Comp. Vision, Graphics, and Image Process.*, 53:112–117, 1991.

[39] V.E. Johnson, W.H. 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.*, 13(5):413–425, May 1991.

[40] J. Jolion, P. Meer, and S. Bataouche. Robust clustering with applications in computer vision. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(8):791–801, August 1991.

[41] R. E. Kass, L. Tierney, and J. B. Kadane. Asymptotics in Bayesian computation. *Bayesian Statistics*, 3:263–278, 1988.

[42] R. E. Kass and S. K. Vaidyanathan. Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binonmial proportions. *J. Royal Statistical Society*, B54:129–144, 1992.

[43] J.T. Kent and K.V. Mardia. Spatial classification using fuzzy membership models. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(5):659–671, September 1988.

[44] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(48):799–813, August 1989.

[45] S. M. LaValle. A Bayesian framework for considering probability distributions of image segments and segmentations. Master's thesis, Univ. of Illinois, Urbana/Champaign, December 1992.

[46] S. M. LaValle and S. A. Hutchinson. A Bayesian segmentation methodology for parametric image models. *IEEE Trans. Pattern Anal. Machine Intell.* To appear.

[47] S. M. LaValle and S. A. Hutchinson. Methods for numerical integration of high-dimensional posterior densities with application to statistical image models. In *Proc. of the SPIE Conf. on Stochastic Methods in Signal Processing, Image Processing, and Computer Vision*, pages 292–303, San Diego, CA, July 1993.

[48] B. S. Manjunath and R. Chellappa. Unsupervised texture segmentation using Markov random field models. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(5):478–482, May 1991.

[49] S. G. Nadabar and A. K. Jain. Parameter estimation in MRF line process models. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 528–533, June 1992.

[50] T. Pavlidis. *Structural Pattern Recognition.* Springer-Verlag, New York, NY, 1977.

[51] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., San Mateo, California, 1988.

[52] L. I. Pettit. Bayes factors for outliers models using the device of imaginary observations. *J. American Statistical Association*, 87:541–544, 1992.

[53] A. Rosenfeld, R. Hummel, and S. Zucker. Scene labeling by relaxation algorithms. *IEEE Trans. Syst., Man, Cybern.*, 6:420–433, 1976.

[54] W. S. Rutkowski, S. Peleg, and A. Rosenfeld. Shape segmentation using relaxation. *IEEE Trans. Pattern Anal. Machine Intell.*, 3(4):368–375, July 1981.

[55] B. Sabata, F. Arman, and J. K. Aggarwal. Segmentation of 3D range images using pyramidal data structures. *Comp. Vision, Graphics, and Image Process.*, 57:373–387, May 1993.

[56] S. Sarkar and K. L. Boyer. Integration, inference, and management of spatial information using Bayesian networks: Perceptual organization. *IEEE Trans. Pattern Anal. Machine Intell.*, 15(3):256–274, March 1993.

[57] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ, 1976.

[58] J. F. Silverman and D. B. Cooper. Bayesian clustering for unsupervised estimation of surface and texture models. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(4):482–496, July 1988.

[59] A. F. M. Smith and D. J. Spiegelhalter. Bayes factors and choice criteria for linear models. *J. Royal Statistical Society*, B42:213–220, 1980.

[60] H. E. Stephanou and S. Y. Lu. Measuring consensus effectiveness by a generalized entropy criterion. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(4):544–554, July 1988.

[61] A. H. Stroud. *Approximate Calculation of Multiple Integrals.* Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.

[62] J. Subrahmonia, Y.P. Hung, and D.B. 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*, pages 390–397, Atlantic City, NJ, June 1990.

[63] R. Szeliski. Bayesian modeling of uncertainty in low-level vision. *Int. J. Comput. Vis.*, 5(3):271–301, December 1990.

[64] 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.*, 13(11):1115–1137, November 1991.

[65] G. Taubin and D.B. 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.

[66] P. H. Winston. *Artificial Intelligence.* Addison-Wesley, Reading, MA, 1984.

[67] 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.*, 12(10):1009–1017, October 1990.