A Bayesian Segmentation Methodology for Parametric Image Models

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

The Beckman Institute
and
Dept. of Electrical and Computer Engineering
University of Illinois
Urbana, IL 61801

Abstract

Region-based image segmentation methods require some criterion for determining when to merge regions. This paper presents a novel approach by introducing a Bayesian probability of homogeneity in a general statistical context. Our approach does not require parameter estimation, and is therefore particularly beneficial for cases in which estimation-based methods are most prone to error: when little information is contained in some of the regions and, therefore, parameter estimates are unreliable. We apply this formulation to three distinct parametric model families that have been used in past segmentation schemes: implicit polynomial surfaces, parametric polynomial surfaces, and Gaussian Markov random fields. We present results on a variety of real range and intensity images.

1 Introduction

The problem of image segmentation, partitioning an image into a set of homogeneous regions, is a fundamental problem in computer vision. Approaches to the segmentation problem can be grouped into region-based methods, in which image subsets are grouped together when they share some property (e.g., [26]); edge-based methods, in which dissimilarity between regions is used to partition the image (e.g., [9]); and combined region- and edge-based methods (e.g., [22]). In this paper, we present a new, Bayesian region-based approach to segmentation.

A standard approach to region-based segmentation is to characterize region homogeneity using parameterized models. With this approach, two regions are considered to be homogeneous if they can be explained by a single instance of the model, i.e., if they have a common parameter value. For example, in range image applications, object surfaces are often modeled as being piecewise algebraic (e.g., [30]). The parameters of such a surface are the coefficients of the corresponding polynomial. Two regions are homogeneous, and thus should be merged, if they belong to a single polynomial surface (i.e., if the coefficients for their corresponding polynomials are the same).

In practice, a region's parameters cannot be observed directly, but can only be inferred from the observed data and knowledge of the imaging process. In statistical approaches, this inference is made using Bayes' rule and the conditional density, $p(\mathbf{y_k}|\mathbf{u_k})$, which expresses the probability that certain data (or statistics derived from the data), $\mathbf{y_k}$, will be observed, given that region k has the parameter value $\mathbf{u_k}$. In typical statistical region merging algorithms (e.g., [27]) point estimates in the parameter space are obtained for different regions, and merging decisions are based on the similarity of these estimates. Often the maximum a posteriori (MAP) estimate is used, which is obtained by maximizing $p(\mathbf{y_k}|\mathbf{u_k})$.

An inherent limitation of nearly all estimation-based segmentation methods reported to date is that they do not explicitly represent the uncertainty in the estimated parameter values, and therefore, are prone to error when parameter estimates are poor (one notable exception to this is the work of Szeliski [29], in which both optimal estimates and the variance in the estimates are computed). To overcome this problem, we present a Bayesian *probability of homogeneity* that directly exploits all of the information contained in the statistical image models, as opposed to computing point parameter estimates.

The probability of homogeneity is based on the ability to formulate a prior probability density on the parameter space, and assess homogeneity by taking the expectation of the data likelihood

over a posterior parameter space. This type of expectation was also used by Cohen and Fan to formulate a data likelihood for segmentation, applied to the Gaussian Markov random field model [5]. In their work, segmentations are defined by a space of pixel labelings, and through window-based iterative optimization, a segmentation is determined that maximizes the data likelihood. By considering the region-based probability of homogeneity, we introduce a different decomposition and prior on the space of segmentations.

Our probability of homogeneity can also be considered as a function of the *Bayes factor* from recent statistical literature [1, 15, 23, 28], which has been developed for statistical decision making, such as model selection. A detailed description of our model and the derivation of the Bayesian probability of homogeneity are given in Section 2.

In addition to providing an explicit accounting of the uncertainty associated with a segmentation (which could feasibly be used in higher level vision processes, such as recognition), our method extends in a straightforward way to allow application of multiple, independent image models. Furthermore, our framework does not require the specification of arbitrary parameters (e.g., threshold values), since context dependent quantities can be statistically estimated.

We have applied our Bayesian probability of homogeneity to segmentation problems using three popular model families: implicit polynomial surfaces, in Section 3; parametric (explicit) polynomial surfaces, in Section 4; and Gaussian Markov random fields for texture segmentation, in Section 5. In Section 7 we present experimental results from each of the model families. These results were obtained using the algorithm described in Section 6.

Further, we have developed special numerical computation methods for directly computing the probability of homogeneity using the parametric models presented in this paper [17], without using large data set, asymptotic assumptions. For this reason, we were able to consider small region sizes for the implicit polynomial results presented in Section 7. Previous techniques that obtain expectations over the parameter space have used some form of this assumption [3, 5, 27].

In principle, our Bayesian probability of homogeneity could be applied to most region-based segmentation algorithms. In related work we have used the probability of homogeneity as a key component for generating probability distributions of alternative segments and segmentations [18].

2 The General Probability of Homogeneity

This section provides the formulation and derivation of the general probability of homogeneity. The version presented here determines the probability that the union of *two* regions is homoge-

neous; probabilistic treatment of more general region sets appears in [16]. Section 2.1 defines the random variables and densities used in our general statistical context. In Section 2.2 we derive expressions for the probability of homogeneity.

2.1 General Model Definitions

The elements of an image, D, are arranged in a 2D array. A given point $D[i,j]$ will have a set of *neighbors*. Using standard four-neighbors, this set is: $D[i-1,j]$, $D[i+1,j]$, $D[i,j-1]$, $D[i,j+1]$. A *region*, R_k, is some connected subset of D. Two regions, R_1 and R_2, will be called *adjacent* if there exists some $D[i_1,j_1] \in R_1$ and $D[i_2,j_2] \in R_2$ that are neighbors.

It is often profitable to begin with some initial partition of the image into small regions, and to construct new segmentations by combining these regions. This is a standard approach taken in the region merging paradigm. For instance, Sabata et al. initially generate an image of *sels*, which corresponds to regions that have near-constant differential properties [26], and Silverman and Cooper begin with an initial grid of small regions [27]. We denote the initial set of regions as $\mathcal{R}$, which represents a partition of D.

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 1). 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), yielding the prior model and the degradation model. The prior model is represented by a density on the parameter space (usually uniform), before any

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 1. The key components in our general statistical framework.

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. These components have been used in similar contexts for image segmentation [8, 29].

In order to determine the probability of homogeneity, it will be necessary to consider a statement of the form $H(R_1 \cup R_2) = true$, which corresponds to the condition that $R_1 \cup R_2$ is homogeneous, and $H(R_1 \cup R_2) = false$, which corresponds to the condition that $R_1 \cup R_2$ is not homogeneous. We will use H to represent the condition $H(R_1 \cup R_2) = true$, and $\neg H$ to represent $H(R_1 \cup R_2) = false$. Note that if H is true then R_1 and R_2 share the same parameter value.

2.2 Probability of Homogeneity Derivation

In this section we derive an expression for the Bayesian probability of homogeneity, given observations from R_1 and R_2. The result is an expression requiring three integrations on the parameter space, given by (2) and (5). The vectors $\mathbf{Y_1}$ and $\mathbf{Y_2}$ represent the observation spaces of R_1 and R_2 respectively. In other words, the random vector $\mathbf{Y_1}$ corresponds to applying functions to the data variables, $D[i, j]$, which belong to R_1. Similarly, $\mathbf{Y_2}$ is obtained from R_2. The observations serve as the evidence used to determine the Bayesian probability of homogeneity, which is represented as $P(H|\mathbf{y_1}, \mathbf{y_2})$. We can apply Bayes' rule to obtain

$$P(H|\mathbf{y_1}, \mathbf{y_2}) = \frac{p(\mathbf{y_1}, \mathbf{y_2}|H)P(H)}{p(\mathbf{y_1}, \mathbf{y_2})} = \frac{p(\mathbf{y_1}, \mathbf{y_2}|H)P(H)}{p(\mathbf{y_1}, \mathbf{y_2}|H)P(H) + p(\mathbf{y_1}, \mathbf{y_2}|\neg H)P(\neg H)}. \tag{1}$$

The denominator of (1) is the standard normalizing factor from Bayes' rule, over the binary sample space, $\{H, \neg H\}$. The expression $P(H)$ represents the prior probability of homogeneity, i.e., the probability that two adjacent regions should be merged, when $\mathbf{y_1}$ and $\mathbf{y_2}$ have not been observed, and in practice we usually take $P(H) = P(\neg H) = 1/2$. This represents a uniform distribution over the binary sample space. The implications of this and other prior distributions is discussed in [18].

We can write (1) as

$$P(H|\mathbf{y_1}, \mathbf{y_2}) = \frac{1}{1 + \lambda_0 \, \lambda_1(\mathbf{y_1}, \mathbf{y_2})} \tag{2}$$

in which

$$\lambda_0 = \frac{1 - P(H)}{P(H)} \quad \text{and} \quad \lambda_1(\mathbf{y_1}, \mathbf{y_2}) = \frac{p(\mathbf{y_1}, \mathbf{y_2}|\neg H)}{p(\mathbf{y_1}, \mathbf{y_2}|H)} = \frac{p(\mathbf{y_1})p(\mathbf{y_2})}{p(\mathbf{y_1}, \mathbf{y_2}|H)}. \tag{3}$$

This utilizes the reasonable assumption that $p(\mathbf{y_1}, \mathbf{y_2}|\neg H) = p(\mathbf{y_1})p(\mathbf{y_2})$, which is further discussed in [16]. The λ_0 and $\lambda_1(\mathbf{y_1}, \mathbf{y_2})$ ratios represent a decomposition of the factors contributing to the

posterior probability of homogeneity. When either of these ratios takes on the value of 1, it essentially does not bias the posterior probability of homogeneity.

Using a common prior density $p(\mathbf{u_{12}})$, and an assumption that the observations y_1 and y_2 are independent when given the common parameter value, u_{12}, we can write the denominator of $\lambda_1(\mathbf{y_1}, \mathbf{y_2})$ as a marginal with respect to U_{12}:

$$p(\mathbf{y_1}, \mathbf{y_2}|H) = \int p(\mathbf{y_1}, \mathbf{y_2}|\mathbf{u_{12}}) p(\mathbf{u_{12}}) \mathbf{du_{12}} = \int p(\mathbf{y_1}|\mathbf{u_{12}}) p(\mathbf{y_2}|\mathbf{u_{12}}) p(\mathbf{u_{12}}) \mathbf{du_{12}}. \tag{4}$$

Using (4), and the marginal over $\mathbf{U_k}$ for each term of the numerator, we obtain:

$$\lambda_1(\mathbf{y_1}, \mathbf{y_2}) = \frac{\left[\int p(\mathbf{y_1}|\mathbf{u_1}) p(\mathbf{u_1}) \mathbf{du_1}\right] \left[\int p(\mathbf{y_2}|\mathbf{u_2}) p(\mathbf{u_2}) \mathbf{du_2}\right]}{\int p(\mathbf{y_1}|\mathbf{u_{12}}) p(\mathbf{y_2}|\mathbf{u_{12}}) p(\mathbf{u_{12}}) \mathbf{du_{12}}}. \tag{5}$$

The ratio above (and similar forms) 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 [28]. Aitkin has developed a Bayes factor for model comparison that conditions the prior model on the data [1]. Kass and Vaidyanathan present and discuss some asymptotic approximations and sensitivity to varying priors of the Bayes factor [15]. Pettit also discusses priors, but with concern for robustness with respect to outliers [23].

Our approach extends in a straightforward way to the case in which we have m independent observation spaces and parameter spaces. In this case, the posterior probability of homogeneity can be expressed as [16]:

$$P(H|\mathbf{y}_1^1, \ldots, \mathbf{y}_1^m, \mathbf{y}_2^1, \ldots, \mathbf{y}_2^m) = \frac{1}{1 + \lambda_0 \prod_{l=1}^{m} \lambda_l(\mathbf{y}_1^l, \mathbf{y}_2^l)} \tag{6}$$

in which $\lambda_l(\mathbf{y}_1^l, \mathbf{y}_2^l)$ is similar to (5).

3 Implicit Polynomial Surfaces for Range Data

Surface models that correspond to the solution sets of implicit algebraic equations are treated in this section, and parametric (or explicit) polynomial models are treated in Section 4. Bolle and Cooper have modeled objects appearing in ranges images with patches of planes, spheres, and cylinders for position estimation [3]. Faugeras and Hebert have used implicit quadric and planar models for object modeling, segmentation, and recognition [7]. Taubin and Cooper have developed an efficient estimation procedure for implicit polynomial curves and surfaces of arbitrary order, with application to object recognition [30].

For this model $D[i,j]$ represents a point in $\Re^3$, specified by $[x_1, x_2, x_3]$ coordinates. For simplicity of notation, we will denote some element of the image D by $\mathbf{x}$ instead of $D[i,j]$. Rather than using the i,j indices, we will index points that belong to some region R_k by $\mathbf{x} \in R_k$.

3.1 The parameter manifold

An implicit polynomial equation is represented as

$$\phi(\cdot, \mathbf{u}) = \sum_{j=1}^{N} u_j x_1^{a_j} x_2^{b_j} x_3^{c_j} = 0 \quad \text{with} \quad a_N = b_N = c_N = 0. \tag{7}$$

The constants a_j, b_j, and c_j are positive integers, representing the exponents of each variable. The $\cdot$ used here indicates that we have an implicit function with $\mathbf{x}$ as the variables.

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 (5), 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.*

3.2 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 *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 (others are mentioned in [16]), and distance approximation [3, 30]:

$$y_k(R_k, \mathbf{u_k}) = \sum_{\mathbf{x} \in R_k} [\delta(\mathbf{x}, \phi(\cdot, \mathbf{u_k}))]^2 \approx \sum_{\mathbf{x} \in R_k} \frac{\phi(\mathbf{x}, \mathbf{u_k})^2}{\|\nabla_x \phi(\cdot, \mathbf{u_k})\|^2}. \tag{8}$$

Note that we use y_k instead of $\mathbf{y_k}$ when the observation space is scalar. We chose to use the sum of squares since we obtained improved integration efficiency with similar segmentation results when experimentally compared to using the displacements directly. The distance approximation is good for small displacements, and for our approach good approximations are only required for small displacements. Large displacement errors will not cause difficulty because of the approximately zero tail values of the chi-square pdf, which will be presented in Section 3.3.

3.3 The degradation model

To define the degradation model, we first need to express the density corresponding to the displacement of an observed point from a given surface. We use a probability model for range-scanning error used and justified in [3], and also used in [30]. The model asserts that the 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 known variance, σ^2.

This degradation model is merely chosen as a representative of possible models that can be used. In practice, for different imaging systems, other models may be more appropriate. For instance, Mirza and Boyer use a t-distribution to model the degradation for robustness with respect to outliers [20]. Ikeuchi and Kanade provide a detailed discussion of the modeling of a variety of range-imaging sensors [12].

Since taking the sum of squares of independent standard Gaussian random variables yields a chi-squared random variable, the degradation density using (8) is

$$p(y_k|\mathbf{u_k}) = \chi^2_{m_k}(y_k) = \frac{1}{2^{m_k/2}\Gamma(m_k/2)}(\frac{y_k}{\sigma^2})^{m_k/2-1}e^{-y_k/2\sigma^2}. \tag{9}$$

Here y_k is the sum-of-squares for a given region, R_k, and parameter value $\mathbf{u_k}$, given by (8). Also, $\Gamma(\cdot)$ is the standard gamma function and $m_k = |R_k|$ (the number of elements in R_k). The variance σ^2 is estimated, and considered as part of the specified degradation model.

3.4 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 distribution on the space of surfaces. If other constraints were used on the parameter space, and we assumed a constant-valued pdf, the distribution would be somewhat different from the one we have selected here. Once some information is present, i.e., some observed data points, this distinction becomes less important.

Since the density over the parameter manifold must integrate to one, the uniform density is just the inverse of the surface area of the half hypersphere that defines the parameter manifold. The prior model is $p(\mathbf{u_k}) = A_N^{-1}$, in which A_N represents the area of the N parameter manifold, which can be obtained through straightforward integration techniques [16].

4 Parametric (Explicit) Polynomial Surfaces

Parametric polynomial surfaces have been used in past segmentation work to model surface patches in range imagery and sets of intensities in intensity images. In early work by Haralick and Watson, the *facet model* was introduced in which intensity image subsets were approximated by polynomials, representing an idealized image [10]. Besl and Jain have used polynomials of variable order for segmentation and select the best model by analyzing fitting-error signs and the mean-square error [2]. Leonardis et al. have also used bivariate polynomials of variable order and select an appropriate image description through a cost/benefit objective function to obtain a segmentation [19]. Silverman and Cooper have used explicit quadric and planar equations to model surfaces patches in intensity images for clustering-based segmentation [27]. Sabata et al. have used parametric polynomials to model surfaces in a hierarchical range image segmentation scheme [26].

The general form of the parametric polynomial model is

$$\psi(\mathbf{u}; i, j) = u_1 + \sum_{m=2}^{N} u_m i^{a_m} j^{b_m}, \tag{10}$$

in which a_m and b_m are positive integers.

The observation space, $\mathbf{Y_k}$, represents a vector of point-to-surface displacements of the intensities in R_k, given a parameter value $\mathbf{u_k}$. For degradation, we use an additive Gaussian iid zero-mean noise model, as considered in [27]. Hence, there is a Gaussian pdf associated with each element of the observation space vector. Due to the independence of the noise model, the joint density is obtained by taking the product of the individual displacement densities:

$$p(\mathbf{y_k}|\mathbf{u_k}) = (2\pi\sigma^2)^{-\frac{|R_k|}{2}} exp\left\{\frac{-1}{2\sigma^2}\sum_{x[i,j]\in R_k}[x[i,j] - \psi(\mathbf{u_k}; i, j)]^2\right\}. \tag{11}$$

We define the prior model by assigning a uniform density to a compact portion of the parameter space. The problem of selecting bounds for a uniform prior has been known to lead to difficulty in Bayesian analysis, referred to as Lindley's paradox [28]. As the volume over which the unfiorm density is defined increases, the ratio (5) decreases. We select $P(H)$ in our experiments to appropriately cancel the effects of the volume; however, it must be understood that the choice of prior in this case significantly affects the probability of homogeneity. For our purposes, a problem of this type only changes λ_1 by some scaling factor, leading to the correct ordering of likely merges, but an ambiguous termination criterion.

5 Texture Segmentation using a Gaussian MRF Model

Models of texture have been used extensively for segmentation. In this section, we consider the application of our general probability of homogeneity to a Gaussian Markov random field (GMRF) model for the problem of unsupervised texture segmentation. This problem has been considered in numerous contexts, and an extensive survey that covers fractal models, operator models, structural texture methods, and frequency domain techniques is provided in [25]. For our parameter space, we use a special MRF formulation known as the SAR model, which is described in [14]. This model has been applied to texture segmentation of intensity images in [4, 5, 27], and has recently been extended to texture modeling and segmentation of color images [21]. In particular, Cohen and Fan have considered maximizing likelihoods formulated through the integration on the GMRF parameter space [5], which is similar to the approach taken here; however, we are interested in iteratively merging region pairs that maximize the probability of homogeneity.

An image element, $D[i,j]$ represents a single intensity, $X[i,j]$, treated as a random variable. We have an N-dimensional parameter space, which represents the interaction of a pixel with a local set of neighboring pixels. The *order* of an MRF indicates the size of the local neighborhood that is considered. In a first order MRF, $N = 4$, corresponding to interactions of $X[i,j]$ with $X[i+1,j]$, $X[i-1,j]$, $X[i,j+1]$, $X[i,j-1]$. For any general order of MRF interactions, the image element of the l^{th} parameter interaction is denoted by $T_l(x)$. Hence, in general at some point $X[i,j] = x$, the model is

$$x - \mu_k = \sum_{l=1}^{N} u_l(T_l(x) - \mu_k). \tag{12}$$

We could also consider the intensity mean in R_k, μ_k, as part of the parameter space; instead, we chose to estimate the mean using the region data for our experiments.

The observation space, $\mathbf{Y_k}$, is defined as a vector that corresponds to all of the intensity data, $x[i,j]$, in some region R_k. Hence, the dimension of $\mathbf{Y_k}$ is equal to the number of pixels in R_k.

We assume that the noise process that occurs in the linear prediction (12) is Gaussian. The joint density that we use over the points in R_k is not a proper pdf; however, it has been considered as a reasonable approximation and used in previous segmentation schemes [4, 21, 27]. We obtain the degradation model by taking the product of the density expressions over each of the individual pixels:

$$p(\mathbf{y_k}|\mathbf{u_k}) = (2\pi\sigma_k^2)^{-\frac{|R_k|}{2}} exp\left\{\frac{-1}{2\sigma_k^2}\sum_{x\in R_k}\left[x - \mu_k - \sum_{l=1}^{N} u_l(T_l(x) - \mu_k)\right]^2\right\} \tag{13}$$

in which σ_k^2 represents the variance over R_k. The variance could also be considered as part of the parameter space; however, we estimate the variance for each region.

For the texture model we used that same prior that was used in Section 4.

6 Computation Issues

We provide an outline of the algorithm that was used to generate the experiments presented in Section 7. Our algorithm resembles agglomerative clustering [27]; however, the standard metric-based merging criterion is replaced by our probability of homogeneity:

1. For each pair of adjacent regions R_i, $R_j \in \mathcal{R}$, compute $P(H(R_i \cup R_j)|\mathbf{y_i}, \mathbf{y_j})$, and store the result in a priority queue with elements sorted by probability.
2. Remove the first pair from the queue, R_{m_1},R_{m_2}, and update $\mathcal{R}$ by adding $R_m \equiv R_{m_1} \cup R_{m_2}$ and removing R_{m_1} and R_{m_2}.
3. For each R_i adjacent to R_m, compute $P(H(R_i \cup R_m)|\mathbf{y_i}, \mathbf{y_m})$, and insert the result into the priority queue.
4. If the probability of the first pair in the queue is less than P_c (or alternatively, the number of regions in $\mathcal{R}$ is c) then terminate
5. Go to 2

With regard to Line 4, many clustering algorithms require the specification of the number of final clusters. Some recent work has been done specifically addressing the problem of determining the number of clusters (known as *cluster validation*), in the context of image segmentation applications [13, 32].

The integrals arising from (5) were computed using a specialized Monte Carlo-based technique for the implicit surface model, and an ellipsoidal decomposition technique on the parameter spaces for parametric polynomials and MRFs. These computation methods are discussed in [17].

7 Experiments

This section presents experimental results using the models presented in Sections 3-5. For each of the three models, we have performed segmentation experiments on dozens of real images.

Figure 1 shows eight range image results, using either an implicit planar or quadric model. For each result, we first show the intensity image (or a synthetic rendering) for the range image, and then the segmentation result. Figure 1.a also shows the initial region set, $\mathcal{R}$.

The variance in the degradation model was only estimated once for a given range image set. To yield accurate placement of points that are close to segment boundaries, we performed maximum-likelihood supervised clustering on the segmentation output from our merging algorithm. We first

discard very small final segments, and then for each image point, $\mathbf{x}$, we choose the region label, l, such that

$$p(\mathbf{x}|\hat{\mathbf{u}}_\mathbf{l}) = \max_{R_k \in \mathcal{R}_a} p(\mathbf{x}|\hat{\mathbf{u}}_\mathbf{k}), \tag{14}$$

in which $\mathcal{R}_a$ represents the set of regions containing the points that are adjacent to $\mathbf{x}$, and $\hat{\mathbf{u}}_\mathbf{k}$ is a least-squares parameter estimate in region R_k.

The initial region set, $\mathcal{R}$, was obtained by combining a small grid with the edge map produced by running the Canny edge detector on the corresponding intensity image. When building an initial region set we would like as few nonhomogeneous initial regions as possible, and the application of the edge detector provides slightly improved performance near boundaries. The edge detector was applied to synthetic renderings of the range data (generated by the method discussed by Sabata et al. [26]) for the images in Figures 1a-c, and a true, corresponding intensity image was used for the remaining range images. We performed no parameter tuning with the Canny edge detector; therefore, there are many missed edges and extra edges in the initial region maps.

The first three images presented in Figure 1 belong to the MSU range image set, which is often used for evaluation of segmentation algorithms. Hoffman and Jain performed smoothing on this data, and then iteratively clustered the range points based on position and surface normal estimation [11]. A conservative clustering is obtained, and additional merging occurs after surface type classification and boundary analysis is performed. Sabata et al. also provide some results on this imagery [26]. They also perform smoothing on the data, use a pyramidal clustering algorithm on synthetic renderings of the range data, and finally merge regions using a squared-error criterion.

We obtained the remaining images in Figure 1 in our lab using the K^2T GRF range scanner setup. This range image set is more suitable for demonstrating our framework, since typical segments have approximately 400-500 points, while the typical segment sizes in the MSU images have around 4000-5000 points (the noise levels for the two sets are comparable). This leads to greater uncertainty in the probabilities; however, good segmentations were obtained.

The next two images show the application of the parametric polynomial model to the segmentation of intensity images. We do not necessarily propose parametric polynomials as the most appropriate model for intensity-based segmentation, but instead are demonstrating the success of our methodology for this given model family, which has been considered previously [27]. Figure 2.a is an image of a tape dispenser, and Figure 2.b shows a plastic slinky. Since the model family is not as accurate (i.e., the images are not likely to have an underlying polynomial model with

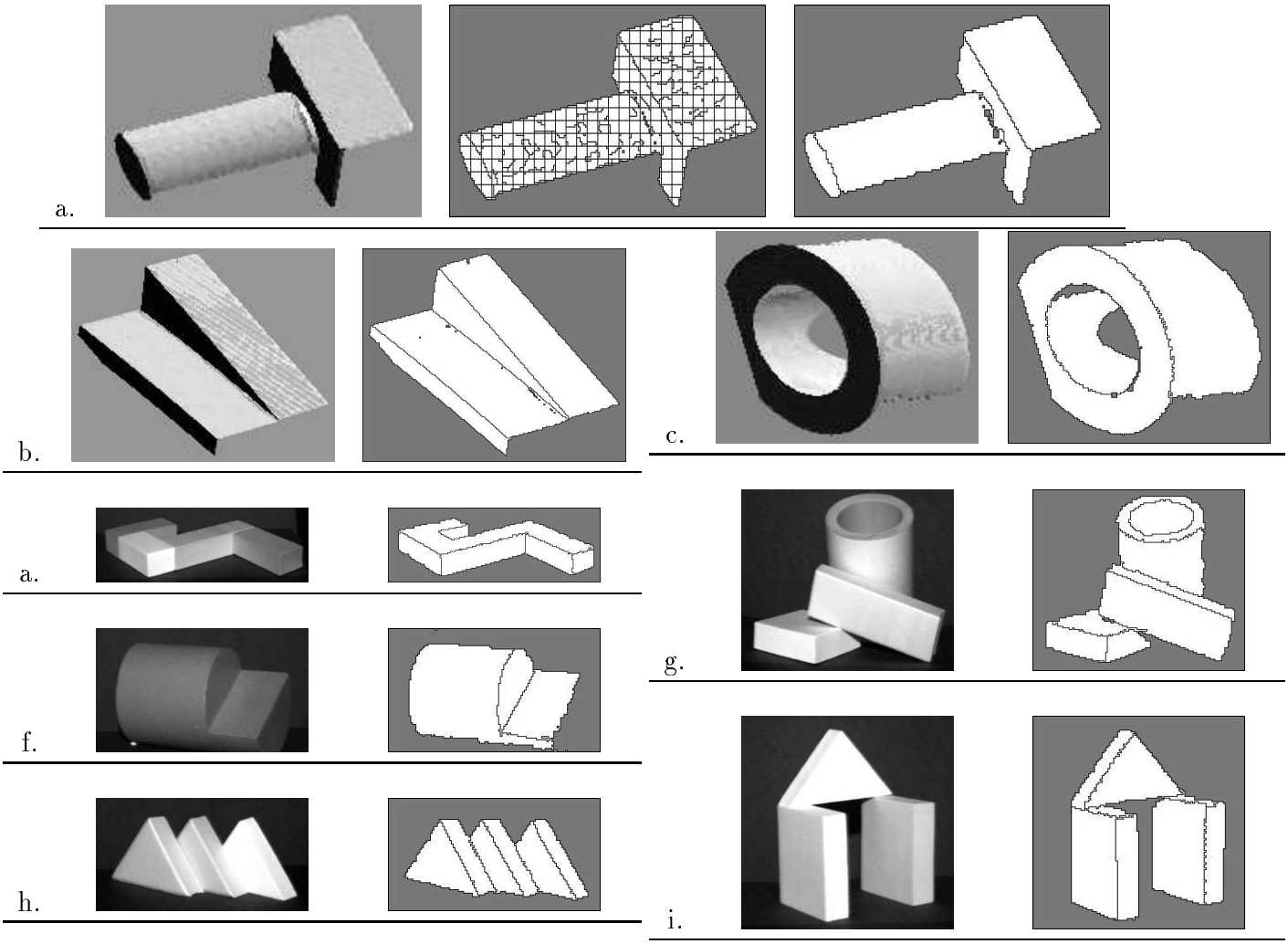

Figure 1. Some of our range image segmentation results.

additive Gaussian noise) the number of classes is more difficult to select; therefore, we show results in this section for user-specified, fixed class numbers.

Figure 2.c-f show four texture results, which were obtained using a third-order MRF. Figure 2.c is composed of Brodatz textures. Figure 2.d is a texture image that was constructed for testing texture segmentation algorithms [6]. Figure 2.e shows a four-class texture result from an image that was obtained by piecing together photographs of different quilts. Figure 2.f is an image of NASA Magellan space probe data of Venusian terrain. Some recent discussion and comparisons of models for texture segmentation can be found in [6, 24, 25], and some texture segmentation experiments on similar imagery appear in [4, 27, 31, 32].

We have only considered coarse segmentations, which are obtained from an initial region set, $\mathcal{R}$, that is formed by partitioning the image into square blocks. We present these coarse segmentations

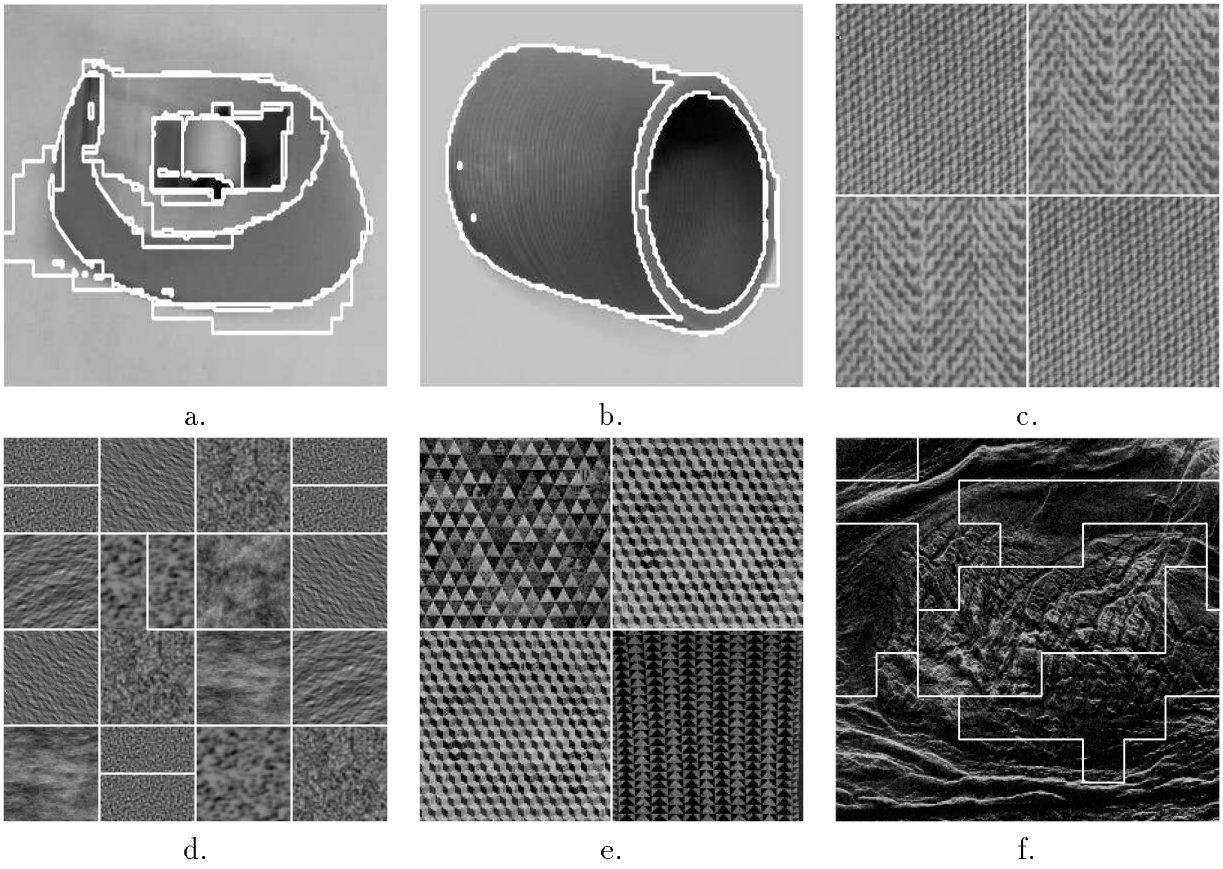

Figure 2. Some of our parametric polynomial and texture segmentations.

since: 1) some minimal region size in needed before the model contains useful information [25]; 2) supervised methods exist for providing good boundary localization, given a coarse segmentation [4, 5, 6].

8 Conclusions

In this paper, we have presented a new approach to region-based segmentation. The key to our approach is a new formulation for the probability that the union of two regions is homogeneous. Our approach does not require parameter estimation, and is therefore particularly beneficial for cases in which estimation-based methods are most prone to error. Our experiments provide strong support for our Bayesian formalism based on the quality of the segmentation results, and the broad class of models considered, which indicates the general applicability of our methods. The segmentations that we obtained with the highest-probability-first algorithm are good, in the context of other recent segmentation results for each of the model types. Further, our Bayesian formalism has been used in algorithms that generate probability distributions of alternative segments and segmentations on real imagery [18].

Acknowledgements

We thank Ken Moroney for his work on some experiments. We also thank Pat Flynn, the MSU Pattern Recognition and Image Processing Lab, Hans du Buf, Becky Castaño, Fang Liu, and Narendra Ahuja, and the anonymous reviewers. This work was sponsored by NSF #IRI-9110270.

References

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

[2] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(2):167–191, March 1988.

[3] 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.

[4] 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.

[5] F. S. Cohen and Z. Fan. Maximum likelihood unsupervised textured image segmentation. *Comp. Vision, Graphics, and Image Process.*, 54(3):239–251, May 1992.

[6] J. M. H. du Buf, M. Kardan, and M. Spann. Texture feature performance for image segmentation. *Pattern Recognition*, 23:291–309, 1990.

[7] 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.

[8] 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.

[9] 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.

[10] R. M. Haralick and L. Watson. A facet model for image data. *Comp. Vision, Graphics, and Image Process.*, 15:113–129, February 1981.

[11] R. Hoffman and A. K. Jain. Segmentation and classification of range images. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(5):608–620, September 1987.

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

[13] 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.

[14] R. L. Kashyap and R. Chellappa. Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Trans. Information Theory*, 29(1):60–72, January 1983.

[15] 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.

[16] 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.

[17] 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*, San Diego, CA, July 1993.

[18] S. M. LaValle and S. A. Hutchinson. On considering uncertainty and alternatives in low-level vision. In *Proc. of the Annual Conference on Uncertainty in Artificial Intelligence*, Washington, D.C., July 1993.

[19] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation as the search for the best description of the image in terms of primitives. In *Proc. Int. Conf. on Computer Vision*, pages 121–125, 1990.

[20] M. J. Mirza and K. L. Boyer. An information theoretic robust sequential procedure for surface model order selection in noisy range data. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 366–371, 1992.

[21] D. K. Panjwani and G. Healey. Unsupervised segmentation of textured color images using markov random fields. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 776–777, New York, June 1993.

[22] T. Pavlidis and Y. Liow. Integrating region growing and edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(3):225–233, March 1990.

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

[24] R. W. Picard, T. Kabir, and F. Liu. Real-time recognition with the entire brodatz texture database. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 638–639, New York, June 1993.

[25] T. R. Reed and J. M. Hans du Buf. A review of recent texture segmentation and feature extraction techniques. *Comp. Vision, Graphics, and Image Process.*, 57:359–372, May 1993.

[26] 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.

[27] 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.

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

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

[30] 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.

[31] M. Tüceryan and A. K. Jain. Texture segmentation using Voronoi polygons. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(2):211–216, February 1990.

[32] 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.