Methods for Numerical Integration of High-Dimensional Posterior Densities with Application to Statistical Image Models

Steven M. LaValle — lavalle@cs.uiuc.edu
Kenneth J. Moroney — moroney@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

Numerical computation with Bayesian posterior densities has recently received much attention both in the statistics and computer vision communities. This paper explores the computation of marginal distributions for models that have been widely considered in computer vision. These computations can be used to assess homogeneity for segmentation, or can be used for model selection. In particular, we discuss computation methods that apply to a Markov random field formulation, implicit polynomial surface models, and parametric polynomial surface models, and present some demonstrative experiments.

1 Introduction

One difficulty that is often encountered when applying Bayesian analysis is the expensive computation that is incurred. In a standard application of Bayes' rule, an integral (or summation) is required to marginalize one set of the random variables with respect to another. This can be costly when the dimension of the random variables is high, as is often the case with statistical image models. In this paper we present numerical methods for efficiently evaluating the marginalizing integrals for popular statistical image models, discuss applications, and present some experimental results that directly use the methods.

Numerical computation on the posterior space that results from a Bayesian analysis has been a subject of active interest in the statistics community. The typical high dimensionality of posteriors in Bayesian analysis has led to recent computation techniques that have increased its applicability. Gibbs sampling is a Markov chain-based technique that allows sampling from the joint posterior distribution of an exponential family, and has proven successful in image processing applications [11]. A recent discussion and comparison of Markov chain methods that use Monte-Carlo simulation, which includes the Gibbs sampler, appears in [3, 24]. Smith has provided a more general survey of Bayesian computation methods, including analytic approximations to the integrals, parametrizations and quadrature rules, and some adaptive sampling techniques [23].

For our general Bayesian analysis, we consider two vectors of random variables, $\mathbf{U}$ and $\mathbf{Y}$. The vector $\mathbf{U}$ represents a continuous parameter space, and $\mathbf{Y}$ represents the observations. The observations can be the image data or some statistics of the image data. There is a noise model, $p(\mathbf{y}|\mathbf{u})$, which represents the anticipated observation for a given parameter value, and a prior density on the parameter space, $p(\mathbf{u})$.

We will next describe three important Bayesian contexts in which this expression occurs:

$$\int p(\mathbf{y}|\mathbf{u})p(\mathbf{u})\mathbf{d}\mathbf{u}. \tag{1}$$

This expression represents the marginalization of $\mathbf{y}$ with respect to $\mathbf{u}$, and the computaiton of its value for statistical image models is the primary concern of this paper.

Model order selection has been a subject of interest in the computer vision community [4, 19], particularly for application to segmentation. This addresses the problem of deciding which model, $\mathbf{U}'$ or $\mathbf{U}$, is appropriate for a given data set, when the models are nested, $\mathbf{U}' \subseteq \mathbf{U}$. For example, $\mathbf{U}'$ could represent a linear model, and $\mathbf{U}$ a quadratic model. For two nested parameter spaces, the following ratio of marginals has been used extensively for model selection in a Bayesian framework [1, 8, 25]:

$$\frac{\int p(\mathbf{y}|\mathbf{u}')p(\mathbf{u}')d\mathbf{u}'}{\int p(\mathbf{y}|\mathbf{u})p(\mathbf{u})d\mathbf{u}}. \tag{2}$$

As (2) increases, confidence in $\mathbf{U}'$ also increases, favoring the simpler model.

In a Bayesian estimation context, one is interested in selecting the $\mathbf{u}$ that maximizes the likelihood, $p(\mathbf{y}|\mathbf{u})p(\mathbf{u})$. This is done since the application of Bayes' rule,

$$p(\mathbf{u}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{u})p(\mathbf{u})}{\int p(\mathbf{y}|\mathbf{u})p(\mathbf{u})d\mathbf{u}}, \tag{3}$$

implies that $p(\mathbf{u}|\mathbf{y})$ is maximized. By computing the denominator of (3), one can use the equation directly to obtain a normalized pdf value for a parameter, given the observations, $p(\mathbf{u}|\mathbf{y})$. Using this, comparisons can be made to the prior pdf values, $p(\mathbf{u})$.

The third application of this marginal computation is for the assessment of region homogeneity for image segmentation. For two subsets, R_1 and R_2, of an image, it has been shown that the ratio,

$$\frac{\left[\int p(\mathbf{y_1}|\mathbf{u_1})p(\mathbf{u_1})d\mathbf{u_1}\right]\left[\int p(\mathbf{y_2}|\mathbf{u_2})p(\mathbf{u_2})d\mathbf{u_2}\right]}{\int p(\mathbf{y_1}|\mathbf{u_{12}})p(\mathbf{y_2}|\mathbf{u_{12}})p(\mathbf{u_{12}})d\mathbf{u_{12}}}, \tag{4}$$

can be used along with Bayes' rule to obtain the probability that the data in R_1 and R_2 were generated by the same parameter value (for some given parameter space) [15]. The variable $\mathbf{u_{12}}$ refers to a combined parameter space that is associated with both R_1 and R_2.

The two ratios, (2) and (4), (and similar forms) have appeared recently in work from the statistics literature, and are termed *Bayes factors*. Smith and Speigelhalter used this ratio for model selection between nested linear parametric models [25]. Aitken 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 [14]. Petit also discusses priors, but with concern for robustness with respect to outliers [21]. The Bayes factor has also been carefully studied for evidence evaluation in a forensic science context [2, 9].

Section 2 discusses some popular statistical image models to which this form of Bayesian analysis applies. Section 3 describes an integration method that pertains to models in which $p(\mathbf{y}|\mathbf{u})p(\mathbf{u})$ is a function of a quadratic in $\mathbf{u}$, and the region of integration is $\Re^N$. Section 4 alternatively describes a more general, but less efficient, Monte Carlo-based integration method. The method of Section 3 applies to parametric polynomial model (Section 2.1) and a Markov random field model (Section 2.2), and the method of Section 4 applies to the implicit surface model (Section 2.3). In Section 5 we show some segmentation results that were obtained using the models in Section 2. Finally, some conclusions are presented in Section 6.

2 Specific Model Applications

We will present the general methods of numerical integration in Section 3 and 4, and in this section we first describe image models to which these methods apply. For each application, sufficient information is given to form the integrand of (1), $p(\mathbf{y}|\mathbf{u})p(\mathbf{u})$. In each section we refer to a set of image elements as R, which could be a set of intensities or range coordinates, depending on the image type.

2.1 Parametric (Explicit) Polynomial Models

The general form of the parametric polynomial model is

$$\psi(\mathbf{u}; i, j) = u_1 + \sum_{m=2}^{N} u_m i^{a_m} j^{b_m}, \tag{5}$$

in which a_m and b_m are positive integers. The *degree* of the model is the maximum over m of $a_m + b_m$. The parameter space is thus spanned by the coefficients that are usually selected for surface estimation.

The observations, $\mathbf{Y}$, are represented by a vector of point-to-surface displacements of the intensities in R, given a parameter value $\mathbf{u}$. We denote a single displacement as

$$\delta(x[i,j], \mathbf{u}) = x[i,j] - \psi(\mathbf{u}; i, j) \tag{6}$$

in which $x[i,j]$ is the image intensity at the i^{th} row and j^{th} column. The dimension of $\mathbf{Y}$ is equal to the number of pixels in R.

We assume an additive Gaussian iid zero-mean noise model, as considered in [22]. The joint density is obtained by taking the product of the individual displacement densities:

$$p(\mathbf{y}|\mathbf{u}) = (2\pi\sigma^2)^{\frac{|R|}{2}} exp\left\{\frac{-1}{2\sigma^2} \sum_{x[i,j]\in R} [\delta(x[i,j], \mathbf{u})]^2\right\}. \tag{7}$$

We define the prior model by assigning a uniform density to a bounded parameter space. For regions that we have considered, a rectangular portion of the parameter space can always be identified that encloses nearly all of the probability mass that contributes to the integrals in (4), and using the integration method of Section 3, we are actually not required to specify bounds to perform the integration (all of $\Re^N$ is used). The problem of selecting bounds for a uniform prior has been known to lead to difficulty in Bayesian analysis, and is referred to as Lindley's paradox [17]. As the volume over which the uniform density is defined increases, the ratio (4) decreases.

2.2 A Markov Random Field Model

For our parameter space, we use a special MRF formulation, which is introduced in [13]. This model has been applied to texture segmentation of intensity images in [10, 22], and has been recently extended to texture modeling and segmentation of color images [20].

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. Figure 1 shows the neighbor set that is used for the MRF orders considered in our experiments.

We use μ and will use σ^2 to represent the mean and variance in R, respectively. 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

$$x - \mu = \sum_{l=1}^{N} u_l(T_l(x) - \mu). \tag{8}$$

We could also consider μ and σ^2 as part of the parameter space. This would require the selection of appropriate prior densities, $p(\mu)$ and $p(\sigma^2)$, and require them to be integrated in (4).

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

	4	3	4	
4	2	1	2	4
3	1	x	1	3
4	2	1	2	4
	4	3	4	

Figure 1: The MRF pixel neighborhood with $X[i, j]$ located in the center. For an n^{th}-order MRF, the pixels in boxes with numbers less than or equal to n comprise the neighborhood.

We assume that the noise process that occurs in the linear prediction (8) is Gaussian. The joint density that we use over the points in R is not a proper pdf; however, it has been considered as a reasonable approximation and used in previous segmentation schemes [10, 20, 22].

We obtain the complete noise model by taking the product of the density expressions over each of the individual pixels:

$$p(\mathbf{y}|\mathbf{u}) = (2\pi\sigma^2)^{\frac{|R|}{2}} exp \left\{ \frac{-1}{2\sigma^2} \sum_{x \in R} \left[x - \mu - \sum_{l=1}^{N} u_l (T_l(x) - \mu) \right]^2 \right\}. \tag{9}$$

For the texture model we also use a uniform prior density on a bounded parameter space.

2.3 Implicit Polynomial Models

For this model each image element represents a point in $\Re^3$, specified by $[x_1, x_2, x_3]$ coordinates, which we concisely denote by $\mathbf{x}$. An implicit polynomial equation is represented as

$$\phi(\cdot, \mathbf{u}) \equiv \sum_{j=1}^{N} u_j x_1^{a_j} x_2^{b_j} x_3^{c_j} = 0 \tag{10}$$

with

$$a_N = b_N = c_N = 0. \tag{11}$$

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. We will later 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 $a_j + b_j + c_j$.

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 (4), 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, termed the *parameter manifold*.

The observation considered here is a function of the signed distances of the points $\mathbf{x} \in R$ from the surface determined by $\mathbf{u}$, termed *displacements*. Define $\delta(\mathbf{x}, \phi(\cdot, \mathbf{u}))$ to be the displacement of the point $\mathbf{x}$ to the surface described by the zero set $\{\mathbf{x} : \phi(\mathbf{x}, \mathbf{u}) = 0\}$. The function $\delta(\mathbf{x}, \phi(\cdot, \mathbf{u}))$ 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 mentioned in [15]:

$$y(R,\mathbf{u}) = \sum_{\mathbf{x}\in R} [\delta(\mathbf{x}, \phi(\cdot,\mathbf{u}))]^2. \tag{12}$$

Note that we use y instead of $\mathbf{y}$ when the observation is a scalar.

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 in [26]:

$$\hat{\delta}(\mathbf{x}, \phi(\cdot,\mathbf{u})) = \frac{\phi(\mathbf{x},\mathbf{u})}{\|\nabla_x \phi(\cdot,\mathbf{u})\|} \tag{13}$$

To define the noise model, we 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 [6]. 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 some known variance, σ^2.

Since taking the sum of squares of Gaussian densities yields the chi-square density, the density using (12) is

$$p(y|\mathbf{u}) = \chi^2_m(y) = \frac{1}{2^{m/2}\Gamma(m/2)} y^{m/2-1} e^{-y/2}. \tag{14}$$

Here y is the sum-of-squares for a given region, R, and parameter value $\mathbf{u}$, given by (12). Also, $\Gamma(\cdot)$ is the standard gamma function and $m = |R|$ (the number of elements in R).

We assign $p(\mathbf{u})$ to be a uniform prior on the parameter manifold.

3 Integration of an N-variate Function of a Quadratic

In this section, we consider an integral of the form

$$\int f(g(\mathbf{u}))d\mathbf{u}, \tag{15}$$

in which $\mathbf{U} = \Re^N$ and g is a scalar, real-valued quadratic function:

$$w \equiv g(\mathbf{u}) = \mathbf{u}^T M\mathbf{u} + b\mathbf{u} + c, \tag{16}$$

and f is a positive continuous function.

We will transform the N-variate integral above into a single integral by decomposing $\mathbf{U}$ into subsets on which $f(g(\mathbf{u}))$ is approximately constant. This is accomplished by considering fixed values, w, for $g(\mathbf{u})$, and the quadaric surfaces in $\mathbf{U}$ that result from using (16). Hence we consider transforming the domain of integration from $\mathbf{u}$ to w, yielding

$$\int f(w)dg^{-1}(w), \tag{17}$$

thus collapsing the N-dimensional integral into one dimension.

Now we consider the set of all points in the parameter space that map between w_{i-1} and w_i (see Figure 2):

$$A_i = \{\mathbf{u} : w_{i-1} < g(\mathbf{u}) \leq w_i\}. \tag{18}$$

In a summation, the differential $dg^{-1}(w)$ is represented by the Lebesgue measure (or area) of A_i. Hence we can write

$$\int f(w)dg^{-1}(w) = \lim_{[w_i - w_{i-1}]\to 0} \sum_i f(w_i)\mu(A_i), \tag{19}$$

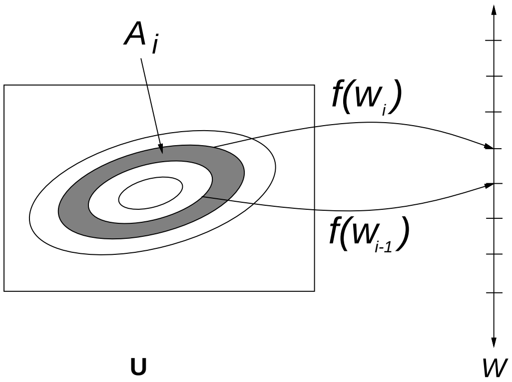

Figure 2: Decomposing the parameter space into concentric ellipsoids.

in which $\mu(A_i)$ represents the measure of A_i.

Since g is quadratic, A_i is a bounded set iff $g(\mathbf{u}) = w_i$ is the equation of an ellipsoid. If A_i is unbounded, then it can be seen in (19) that the integral in (15) is infinite; therefore, we are only concerned with cases in which $g(\mathbf{u}) = w_i$ represents an ellipsoid.

The measure of A_i is found by taking the set difference of two concentric ellipsoids that are rotated and translated away from the origin, as depicted in Figure 2. Recall that the volume of an ellipse is proportional to its axis lengths. To compute $\mu(A_i)$, we center the ellipsoids at the origin with their axes aligned with the coordinate axes.

By using an affine transformation on $g(\mathbf{u}) = w_i$), described in [5], we obtain the quadratic form $\mathbf{u'}^T M' \mathbf{u'} = 1$, in which M' is diagonal. The resulting standardized ellipse equation is

$$\sum_j (\frac{u'_j}{a_j})^2 = 1 \tag{20}$$

in which

$$a_j = \sqrt{\frac{k}{\lambda_j}} \tag{21}$$

and

$$k = \sum_j \frac{(b'_j)^2}{4\lambda_j} - c + w. \tag{22}$$

The vector $\mathbf{b'}$ is computed by the inner product, $R \cdot \mathbf{b}$, in which R is the corresponding matrix of eigenvectors of the matrix M. Also, λ_i represents the i-th eigenvalue of the matrix M. The ellipse volume is

$$C_N \prod_{j=1}^{N} a_j \tag{23}$$

in which

$$C_N = \begin{cases} \dfrac{2\pi^{(N/2)}}{N(N/2-1)!} & \text{if } N \text{ is even} \\ 2^{(N+1)/2} \left[\dfrac{\pi^{(N-1)/2}}{N(N-2)(N-4)\cdots 1} \right] & \text{if } N \text{ is odd .} \end{cases} \tag{24}$$

The coefficient C_N can be determined by performing an elliptical coordinate transformation, which is a generalization of standard spherical coordinates [15].

In practice we compute the integral (15) by considering a finite approximation of the sum in (19):

$$\int f(g(\mathbf{u}))d\mathbf{u} = \int f(w)dg^{-1}(w) \approx \sum_{i=0}^{k} f(w_i)\mu(A_i). \tag{25}$$

In general numerical quadrature formulas can also be applied; however, we have obtained satisfactory performance by directly using the sum.

We select starting and ending points, w_0 and w_k in (25) by making the assumption that

$$\int_{-\infty}^{w_0} f(w)dg^{-1}(w) \approx 0 \tag{26}$$

and

$$\int_{w_k}^{\infty} f(w)dg^{-1}(w) \approx 0. \tag{27}$$

Hence the performance of this method is affected by the rate at which f approaches the origin.

4 Monte Carlo-based integration

In this section we describe a more general method for high dimensional integration, which does not require a particular form for the integrand; however, the integrations are less efficient. We first briefly describe the general Monte Carlo integration method. In Section we discuss a specialized technique that obtains significant improvement in computational performance for the model introduced in Section 2.3.

The Monte-Carlo integration method iteratively approximates a definite integral by uniformly sampling from the domain of integration, and averaging the function values at the samples. The integrand is treated as a random variable. A sampling/averaging scheme yields a parameter estimate of the mean, or expected value of the random variable. We use the most basic Monte Carlo method. More elaborate schemes with faster convergence rates are discussed in [27]; however, improvement in the convergence rate for these methods is possible only for low-dimensional cases ($N \leq 3$). For our purposes, since the dimension is high and the number of needed samples does not depend on dimension, the basic Monte Carlo integration method is appropriate. For a complete introduction to Monte-Carlo integration, see [12].

In the derivation that follows, we treat the region of integration as a vector of random variables, denoted by $\mathbf{T}$, defined on a unit cube. If a different region of integration is needed, an appropriate integral transformation can be applied. Take $h : \mathbf{T} \to \Re$ and $h \in L^2$.[1] The integral (1) is represented as

$$I(h) = \int_T h(\mathbf{t})d\mathbf{t}. \tag{28}$$

Take a set of n independent samples $\mathbf{t_1}, \mathbf{t_2}, \cdots, \mathbf{t_n}$, drawn uniformly from the space $\mathbf{T}$. The n^{th} estimate of $I(h)$ is

$$I_n(h) = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{t_i}). \tag{29}$$

By the strong law of large numbers, $\|I_n(h) - I(h)\| \to 0$ as $n \to \infty$, with probability one. Consider the variance of the estimate, $\sigma_n^2 = E[I_n(h) - I(h)]^2$. From (29), we observe by linearity that $E[I_n(h)] = E[I(h)] = I(h)$. From this observation we obtain an expression for the variance of the estimate [18],

$$\sigma_n^2 = E[I_n^2] - 2E[I_n(h)]E[I(h)] + E[I(h)]^2 = \frac{1}{n}\{I(h^2) - [I(h)]^2\}. \tag{30}$$

[1] By $h \in L^2$, we mean that $\int h^2 < \infty$.

This indicates that the error is reduced at a rate of $1/n$.

4.1 Importance sampling for peaked integrands

The Monte Carlo integration method is known to have difficulty when the integrand is peaked [12]. The reason is that most of the area that contributes to the integral is concentrated in a small region, which is missed by most of the samples. This has led to methods that attempt to concentrate the samples in the peaked area(s), and is termed *importance sampling.*

For the model in Section 2.3, we identify a rectangular region in the domain of integration which contains nearly all of the points that significantly contribute to the integral. The random sampling is only performed inside the rectangular box, and the number of samples required is significantly reduced (by a factor of thousands in many practical cases).

First we consider the case of evaluating the integral (1) for some region R with size n. If we use the expressions (12), (13), and (14), and an approximation from [26], the integrand can be expressed as

$$f(\frac{\mathbf{u}^T M \mathbf{u}}{\mathbf{u}^T Q \mathbf{u}}), \tag{31}$$

in which M and Q are positive definite symmetric matrices, and f represents the chi-square density. The Monte Carlo sampling is applied to a volume integral that is obtained by transforming the integral over the parameter space into a volume integral over the unit hypercube; this transformation is a generalization of the spherical coordinate transformation [15].

We are evaluating the chi-square at various points throughout the integration. We can take some maximum value, $k_1 >> n\sigma^2$, such that sample points that yield a sum-of-squares value greater than k_1 contribute relatively little to the integration, since either density asymptotically approaches zero. We use the Cornish-Fisher approximation [28] to the chi-square cumulative distribution function to obtain a k_1 at the 99.9th percentile for some n. The left side of the equation below represents the set of all parameter values that yield sum-of-squares less than k_1. Note that this is a subset of the right side:

$$\left\{\mathbf{u} : \frac{\mathbf{u}^T M \mathbf{u}}{\mathbf{u}^T Q \mathbf{u}} < k_1\right\} \subseteq \left\{\mathbf{u} : \frac{\mathbf{u}^T M \mathbf{u}}{max(\mathbf{u}^T Q \mathbf{u})} < k_1\right\}. \tag{32}$$

Since $\|\mathbf{u}\| = 1$,

$$max(\mathbf{u}^T Q \mathbf{u}) = maxeig(Q) \equiv k_2, \tag{33}$$

in which $maxeig(Q)$ represents the maximum eigenvalue of Q. Therefore, the right side below describes a solid ellipsoid, centered at $\mathbf{u} = \mathbf{0}$, which encloses all the points in the parameter space that significantly contribute to the integration:

$$\left\{\mathbf{u} : \frac{\mathbf{u}^T M \mathbf{u}}{\mathbf{u}^T Q \mathbf{u}} < k_1\right\} \subseteq \left\{\mathbf{u} : \mathbf{u}^T M \mathbf{u} < k_1 k_2\right\} \tag{34}$$

Let $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ denote the eigenvalues of M, in order of increasing magnitude. Also, let S denote the corresponding eigenvector matrix, which is a rotation matrix that aligns the ellipsoid with the coordinate axes (diagonalizing M). Take $\mathbf{u} = S\mathbf{v}$, and we obtain

$$\mathbf{u}^T M \mathbf{u} = (S\mathbf{v})^T M (S\mathbf{v}) = \mathbf{v}^T S^T M S \mathbf{v} = \mathbf{v}^T \Lambda \mathbf{v} \tag{35}$$

in which $\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$. Using $\mathbf{v}$, the ellipsoid equation becomes

$$\sum_i \lambda_i v_i^2 < k_1 k_2. \tag{36}$$

The half lengths of the principle ellipsoid axes are

$$b_i = \sqrt{\frac{k_1 k_2}{\lambda_i}}. \tag{37}$$

The rectangular subset of $\Re^N$ that has corners located at coordinates $\pm b_i$ encloses the ellipsoid, and the rectangular faces are tangent to the ellipsoid surface. We can apply the inverse of the spherical coordinate transformation to map the corners of the box into T. These form a rectangular subset, T', of T in which the corners have coordinates we denote by $1/2 \pm c_i$.[2]

The portion of parameter space that significantly contributes to the resulting integral lies within T', hence we only need to draw samples uniformly from T'. Using these results, the integral (1) can be computed by

$$\frac{1}{nF} \sum_{\mathbf{v}} f\left(\frac{\sum_i \lambda_i v_i^2}{\mathbf{v}^T S^T Q S \mathbf{v}}\right). \tag{38}$$

Above, F represents the ratio of the area of T' to T, and n is the number of samples, $\mathbf{v}$, that are used. The F also represents the factor by which the number of required samples is reduced.

To compute an integral in the denominator of (4), we use the smallest rectangular region, T', (and corresponding rotation S) of the two regions R_1 and R_2. If that region is R_1 then the integral is computed by

$$\frac{1}{nF} \sum_{\mathbf{v}} f\left(\frac{\sum_i \lambda_i v_i^2}{\mathbf{v}^T S^T Q_1 S \mathbf{v}}\right) f\left(\frac{\mathbf{v}^T S^T M_2 S \mathbf{v}}{\mathbf{v}^T S^T Q_2 S \mathbf{v}}\right) \tag{39}$$

in which the λ_i are the eigenvalues of M_1 and S is its eigenvector matrix.

5 Experimental examples

In this section we briefly present some segmentation results that were obtained with a clustering algorithm that begins with a partition of the image into small regions, and iteratively groups regions that have the highest probability of homogeneity, as prescribed by (4). These segmentations, and other experiments that we have performed, require high-dimensional (from 3 to 20) parameter space integrations using the models presented in Section 2.

The first two rows of Figure 5 show two range image experiments that use the implicit quadric surface model (Section 2.3). Figures 5.a and 5.d show synthetic renderings of the 3D data sets. Figures 5.b and 5.e show an (automated) initial partition of the image, on which the clustering is performed. The final segmentation results, shown in Figures 5.c and 5.f, are obtained after performing the clustering and a simple boundary localization operation [16]. Figures 5.g-i show texture segmentation results on intensity images (using the model in Section 2.2) that were obtained by clustering on images that were initially partitioned with square grid. Figures 5.j-l show some segmentations of intensity images by applying a quadric parametric polynomial model of Section 2.1 directly to the intensities.

6 Conclusions

We have presented integration methods that pertain to a wide class of statistical image models. In particular, these methods have been successfully applied to the implicit polynomial surface model, the parametric (explicit) polynomial surface model, and a Markov random field model. These integration methods are

[2] Some of rectangular faces in the parameter space may lie outside the unit hypersphere. When the first axis is found that is outside, the remaining c_i are set to their maximum value, 1/2.

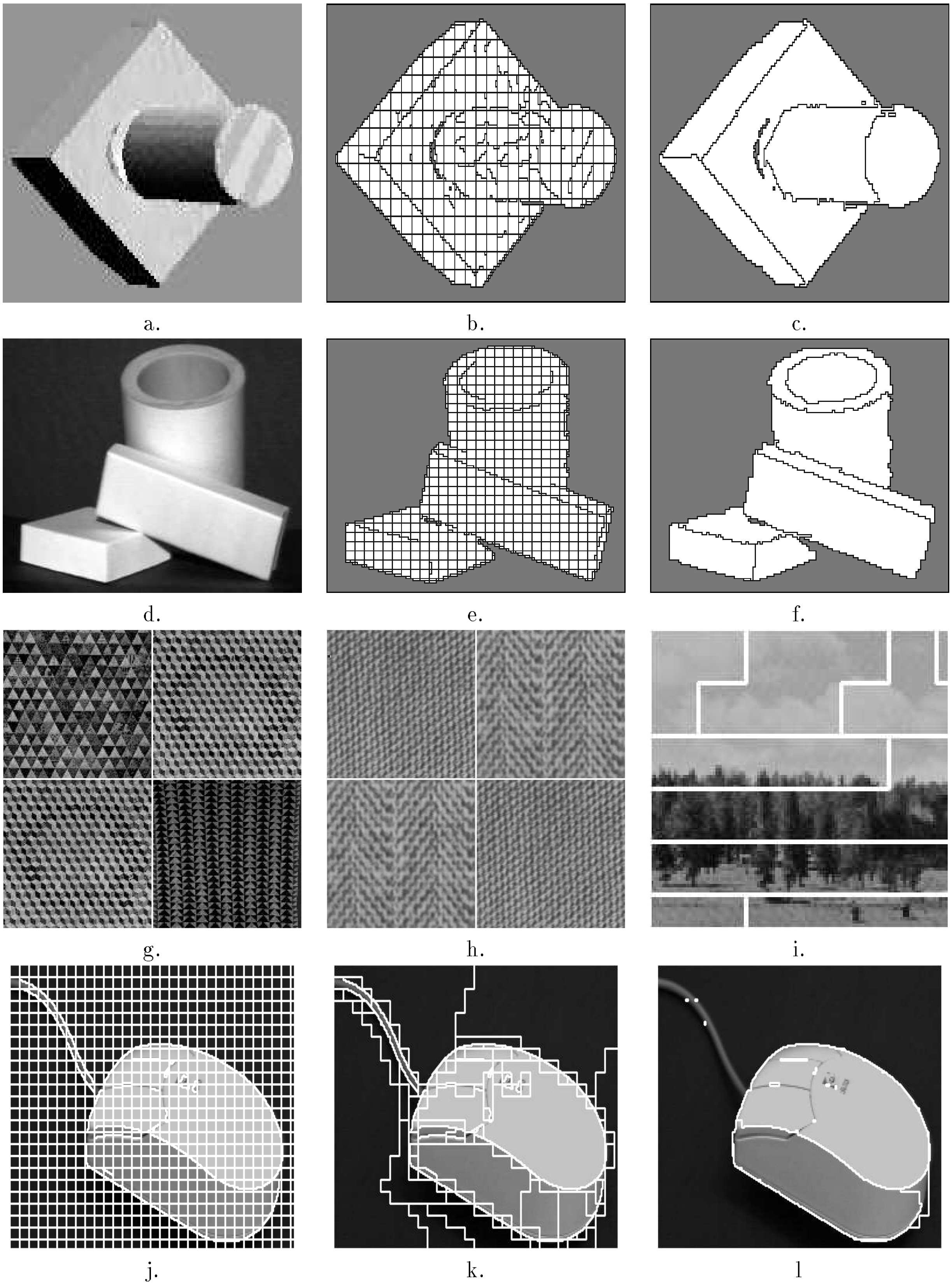

Figure 3: Some image segmentation results.

crucial for the Bayesian computation required in our related segmentation work [15], which use (4). Since the models presented in Section 2 are nested families, it would be interesting in future work to study the application of (2) for model selection.

7 Acknowledgements

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

References

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

[2] 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.

[3] J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *J. Royal Statistical Society*, B55(1):25–37, 1993.

[4] 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.

[5] D. M. Bloom. *Linear Algebra and Geometry.* Cambridge University Press, 1978.

[6] 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.

[7] R. M. Bolle and B. C. Vemuri. On three-dimensional surface reconstruction methods. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(1):1–12, January 1991.

[8] B. P. Carlin, R. E. Kass, F. Javier Lerch, and B. R. Hugeunard. Predicting working memory failure: A subjective Bayesian approach to model selection. *J. American Statistical Association*, 87:319–327, 1989.

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

[10] 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.

[11] 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.

[12] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods.* Wiley, New York, NY, 1986.

[13] 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.

[14] 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.

[15] 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.

[16] S. M. LaValle and S. A. Hutchinson. A bayesian segmentation methodology for parametric image models. *In preparation*, 1993.

[17] D. V. Lindley. A statistical paradox. *Biometrika*, 44:187–192, 1957.

[18] E. Masry and S. Cambanis. Trapezoidal Monte Carlo integration. *SIAM J. Numer. Anal.*, 27(1):225–246, February 1990.

[19] 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.

[20] 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.

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

[22] 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.

[23] A. F. M. Smith. Bayesian computational methods. *Phil. Trans. R. Soc. Lond. A*, 337:369–386, 1991.

[24] A. F. M. Smith and G. O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Royal Statistical Society*, B55(1):3–23, 1991.

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

[26] 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.

[27] S. Yakowitz, J. E. Krimmel, and F. Szidarovsky. Weighted Monte Carlo integration. *SIAM J. Numer. Anal.*, 15(6):1289–1300, December 1978.

[28] J. H. Zar. Approximations for the percentage points of the chi-squared distribution. *J. Appl. Statist.*, 27(3):280–290, 1978.