Proc. IEEE Int. Conf. Image Processing (IEEE, Piscataway, 1999)

Uncertainties in Bayesian geometric models

Kenneth M. Hanson, Gregory S. Cunningham, and Robert J. McKee
MS P940/DX-3, Los Alamos National Laboratory
Los Alamos, New Mexico 87545 USA
E-mail: {kmh,cunning,mckee}@lanl.gov

Abstract

Deformable geometric models fit very naturally into the context of Bayesian analysis. The prior probability of boundary shapes is taken to proportional to the negative exponential of the deformation energy used to control the boundary. This probabilistic interpretation is demonstrated using a Markov-Chain Monte-Carlo (MCMC) technique, which permits one to generate configurations that populate the prior. One of many uses for deformable models is to solve ill-posed tomographic reconstruction problems, which we demonstrate by reconstructing a two-dimensional object from two orthogonal noisy projections. We show how MCMC samples drawn from the posterior can be used to estimate uncertainties in the location of the edge of the reconstructed object.

1. Introduction

Deformable models fit particularly well into the framework of Bayesian analysis. The smoothness of deformable models is typically controlled by means of a deformation energy function that provides a measure of deviation from smoothness [1]. In the context of Bayesian analysis, the probability of the corresponding geometric configuration is taken to be the negative exponential of this energy function [2]. We will emphasize this interpretation by showing a set of random samples of boundary configurations taken from a prior probability based on a typical form for the deformation energy. These samples are generated using the Markov Chain Monte Carlo (MCMC) technique.

This paper deals with the assessment of uncertainties in estimated models, a capability greatly assisted by the use of Bayesian analysis. The general method we employ here is to generate a sequence of random samples of the posterior probability distribution using MCMC. By fully mimicking the posterior, this sequence of samples can be used to assess the uncertainties in estimated models. As an example of the usefulness of this technique, we consider a problem of reconstructing an object from projections in two directions under the assumption of a known, constant interior density. In the analysis, the boundary of the reconstructed objects are subject to a prior that promotes smoothness. We show how samples from the posterior can be used to estimate uncertainties in the location of the boundary of the reconstructed object.

2. Bayesian Analysis

Bayesian analysis is a model-based approach to analyzing data with a strong emphasis placed on uncertainty assessment. Every aspect of modeling is assigned a probability that indicates our degree of certainty in its value. At the lowest level of analysis, the estimation of the values of parameters for a specified model, a probability density function (PDF) is associated with each continuous parameter. Loosely speaking, the range of a probability distribution indicates the possible range of its associated parameter. The benefit of Bayesian analysis over traditional methods of uncertainty, or error, analysis is that it permits the use of arbitrary probability distributions, not just Gaussian distributions, and of arbitrary measures of uncertainty, not just rms deviation (or variance). Bayesian analysis reveals the use of prior knowledge and assumptions, which other kinds of analysis incorporate, but do not always make explicit. Furthermore, it extends analysis to higher levels of interpretation, e.g., the choice of the strength of the priors used, the rejection of any particular model, and the selection of appropriate models [3–5].

Before conducting an experiment, one starts with some knowledge about the physical object being studied. In addition one often has a model for the object, with associated parameters $\mathbf{x}$, that will be used to analyze the experimental results. In Bayesian analysis, the uncertainties in what is known beforehand are expressed in terms of a PDF on the parameters, $p(\mathbf{x})$, called the prior. This prior knowledge can come from previous measurements, specific information regarding the object itself, or simply general knowledge about the parameters, e.g., that they are nonnegative.

With the experimental data in hand, the prior is modified to yield the posterior, which is the PDF $p(\mathbf{x}|\mathbf{d})$ of the parameters given the observed data $\mathbf{d}$, using Bayes law

$$p(\mathbf{x}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{x})\, p(\mathbf{x}) \ . \tag{1}$$

The probability $p(\mathbf{d}|\mathbf{x})$, called the likelihood, comes from a comparison of the actual data to the data predicted on the basis of the model of the object. The predicted data are generated using a model for how the measurements are related to the object, which we call the measurement model. Under the assumption that the data are degraded by uncorrelated and additive Gaussian noise, it is appropriate to use the exponential of $-\frac{1}{2}\chi^2$ for the likelihood. As usual, χ^2 is the mean squared difference between the actual and the predicted measurements divided by the variance of the measurements. The typical analysis consists of estimating the "best" model to explain the data. Often the model that maximizes the posterior, referred to as the maximum *a posteriori* (MAP) estimate is found, although other estimators can be argued to be more appropriate in some circumstances. While an estimate of the best model is the objective of many analyses, it is only the beginning for the true Bayesian.

The present study is carried out using the Bayes Inference Engine (BIE). We developed the BIE [6] to allow one to easily develop complex models for both the objects under study and the measurement process.

2.1. Geometric Boundary Model

Here we use a deformable model to represent the boundary of the object to be reconstructed. The object's boundary is approximated in discrete terms as a finely-divided polygon. The length of the edges of the polygon can be made short enough to adequately describe a curve at any degree of resolution desired.

A smoothness constraint on the boundary is achieved by placing a prior on the curvature of the boundary. The minus-log-prior is taken to be proportional to $\int \kappa^2(s)\, ds$, where $\kappa(s)$ is the curvature of the curve as a function of the arclength along the curve s. This prior serves to keep the curve smooth because large curvature is penalized. This form for a prior has a physical analog in the formula that describes the potential energy created by bending a stiff rod. To achieve a prior that is related to the shape of the curve, not its size, we multiply the integral by the total arclength of the boundary. From the Bayesian point of view, the prior is interpreted probabilistically. A specific probability is assigned to every closed boundary, which ranks boundary shapes according to their plausibility. This probability expresses the uncertainty about the possible shapes of the object that we have before we take data.

For our discrete polygon model, we replace the integral by a sum of contributions associated with each vertex in combination with half of each neighboring edge of the polygon. We use for the minus-log-prior a discrete approximation to the expression

$$\frac{\alpha}{(2\pi)^2} \int ds \int \kappa^2(s)\, ds \ . \tag{2}$$

The factor of $(2\pi)^{-2}$ is included to normalize the result to unity for a circle when $\alpha = 1$. The parameter α determines the strength of the prior relative to the likelihood. The form of Eq. (2) is such that for equal-sided polygons it is independent of the length of the sides as the length goes to zero. In addition to the smoothness constraint, we find it useful to control the lengths of the sides of the polygon. This control is accomplished in this study by adding to the above minus-log-prior an expression that is quadratic in the deviation of each side of the polygon relative to the average side length. Further details can be found in Ref. [7].

3. Markov Chain Monte Carlo

In Bayesian analysis there is often the need to integrate over the posterior. One way to do that is to use a Monte Carlo technique, i.e. draw random samples from the posterior, The Markov Chain Monte Carlo (MCMC) technique provides a means to sample an arbitrary probability density function (PDF).

A Markov chain is a sequence of states in which the probability of each state depends only on the previous state. In MCMC the objective is to generate a sequence of parameter sets that mimic a specified PDF, let's call it $q(\mathbf{x})$, where $\mathbf{x}$ is a vector of parameters in the relevant parameter space. More precisely, it is desired that the MCMC sequence be in statistical equilibrium with the target PDF $q(\mathbf{x})$, which is achieved when the MCMC sequence is marked by the condition of detailed balance:

$$q(\mathbf{x})\, T(\mathbf{x} \to \mathbf{x}') = q(\mathbf{x}')\, T(\mathbf{x}' \to \mathbf{x}) \ , \tag{3}$$

where $T(\mathbf{x} \to \mathbf{x}')$ is the transition probability for stepping from $\mathbf{x}$ to $\mathbf{x}'$. This equation essentially requires that in a very long sequence the number of steps from $\mathbf{x}$ to $\mathbf{x}'$ is identical to the number from $\mathbf{x}'$ to $\mathbf{x}$. For more information about MCMC, the reader is referred to the excellent book edited by Gilks et al. [8], which represents the best available compendium on MCMC.

A desirable attribute of MCMC is that there are generally no restrictions on the types of PDFs that

Figure 1: The original 2D object used in the example.

can be sampled; no functional form for the PDF is required. In its basic form, MCMC only requires that one be able to calculate $\varphi = -\log(\text{posterior})$, although sometimes the gradient of φ is used.

3.1. Metropolis Algorithm

One of the simplest algorithms used in MCMC calculations is due to Metropolis et al. [9] This algorithm ensures detailed balance (3) for each step in the sequence. One starts at an arbitrary point in the vector space to be sampled, $\mathbf{x}_0$. The general recursion at any point in the sequence $\mathbf{x}_k$ is as follows:

(1) Pick a new trial sample $\mathbf{x}^* = \mathbf{x}_k + \Delta\mathbf{x}$,
where $\Delta\mathbf{x}$ is randomly chosen from a symmetric probability distribution

(2) Calculate the ratio $r = q(\mathbf{x}^*)/q(\mathbf{x}_k)$

(3) Accept the trial sample, that is, set $\mathbf{x}_{k+1} = \mathbf{x}^*$,
if $r \geq 1$,
or with probability r, if $r < 1$,
otherwise don't accept step; set $\mathbf{x}_{k+1} = \mathbf{x}_k$

Repeat steps (1-3) until sequence is long enough.

4. Example: Uncertainty in a Tomographic Reconstruction

We demonstrate the usefulness of deformable models and the versatility of the MCMC technique with an example of tomographic reconstruction from just two views. This problem is an extraordinarily difficult inverse problem. Its solution is made feasible by employing the prior information that the object being reconstructed has constant density and consists of a fairly simple shape with smooth boundaries. Figure 1 displays the object devised for this example. It is fashioned to be representative of a lumen, the cross section of an artery, possessing a sizable occlusion. The calculations here are based on images with a full width of 128×128 pixels, arbitrarily set to a size of 4 mm × 4 mm. For better visualization, all the images shown here display just the central 2.5 mm × 2.5 mm region. To give the scale of the images, the width and height of the object are roughly 64 pixels, or about 2 mm.

Two orthogonal views of the object shown in Fig. 1 are generated, one consisting of the vertical projection and the other of the horizontal. Each projection consists of 128 samples over a distance of 4 mm. Gaussian noise is added to these projections with an rms deviation of 5% of the peak projection amplitude. For this simulation, we ignore blur in the measurement system.

4.1. MAP Reconstruction

For reconstruction the object is modeled in terms of a finely-divided polygon filled with a constant density, which we assume is known beforehand. The polygon has 50 sides to approximate a continuous curve. The parameters in the model consist of the x and y values of the 50 vertices. The –log(posterior) for this problem is the sum of –log(likelihood) = $\frac{1}{2}\chi^2$ and the –log(prior) contributions given in Eq. (2) and the prior of side length. However, the strength of the prior, i.e. α in Eq. (2), must be specified. Ideally, the hyperparameter α that is consistent with the data would be calculated utilizing the next higher level of Bayesian inference [3]. As we are not yet equipped to do that in the BIE, we tried several values for α and selected what seemed to be an appropriate value, $\alpha = 5.0$.

The MAP reconstruction is obtained by using the BIE to find the minimum in $\varphi = -\log$ (posterior) with respect to the 100 variables specifying the polygon model. The BIE accomplishes this in an efficient manner through the use of the Adjoint Differentiation In Code Technique (ADICT) [10] to calculate the gradient of φ with respect to the variables. The initial object is taken to be a circle of diameter 1.6 mm = 51 pixels, for which $\frac{1}{2}\chi^2 = 396.15$ and –log(prior) = 7.88. At the minimum in φ, these values are $\frac{1}{2}\chi^2$ = 119.16, and –log(prior) = 18.24. The resulting reconstructed object compares reasonably well with the original, as shown in Fig. 2. The maximum discrepancy in the position of the two boundaries is about 3.3 pixels, which occurs in the lower lobe. Over the vast majority of the perimeter, the reconstructed boundary lies at most one pixel away from the original.

The reconstruction shown in Fig. 2 is vastly su-

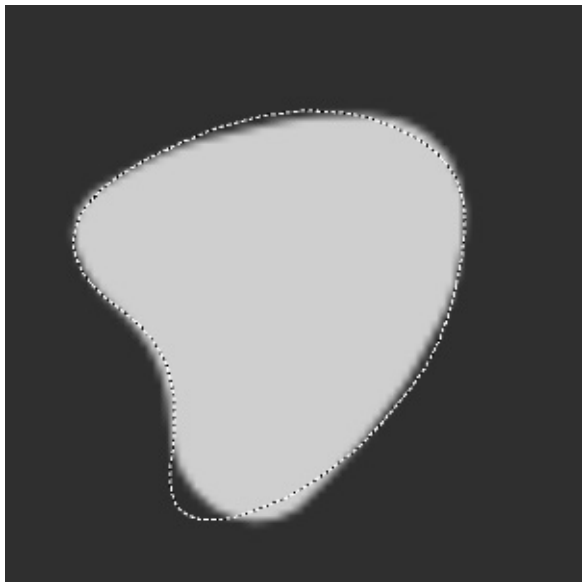

Figure 2: The MAP reconstruction from two orthogonal noisy projections, shown as a grayscale image with the boundary of the original object superimposed.

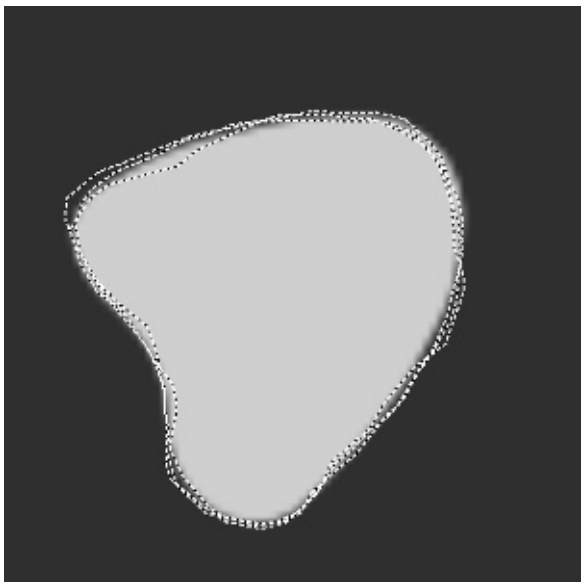

Figure 3: Three representative samples from the posterior shown as curves on top of the grayscale images of the MAP reconstruction.

perior to one that would be obtained using conventional reconstruction algorithms. For example, see Ref. [11] for a reconstruction of a similar object from two views obtained using the multipicative algebraic reconstruction technique (ART), which yields an image that maximizes entropy and incorporates a nonnegativity constraint. We will now prove the reliablity of this reconstruction by using MCMC to assess its uncertainty.

4.2. MCMC Results

The MCMC algorithm described above was used to generate samples from the posterior of this reconstruction problem. In our present example, for each MCMC trial step, the increments in the x and y position of each of the vertices are independently chosen from a Gaussian distribution with an rms step size of 0.06 pixels. Since we are drawing vertex step samples from the space of x and y and the priors outlined in Sect. 2.1 are stated in the θ, L space, we need to transform from $p(\theta, L)$ to $p(x, y)$ using the Jacobian of the transform to properly evaluate the priors for the MCMC algorithm. This Jacobian alters the approximately quadratic form of Eqs. (2) by adding terms that are approximately constant for small θs. We believe these terms amount to relatively minor adjustments in the values of the priors and so ignore them.

In all, 150,000 trial steps were calculated for a total computation time of about 16 hours on a DEC-station 250 with a DEC Alpha processor running at 266 MHz; 42049 steps were accepted, yielding an acceptance rate of about 28%. Three widely-separated samples from the full MCMC sequence are shown in Fig. 3. While it is not possible to get a quantitative estimate of the uncertainty from these three samples, they provide some indication of the amount of variation in the shapes that occupy the posterior. The amount of waviness observed in the boundary is moderate, as can be observed in Fig. 3. The superfluous waviness compared to the original object is evidence that $\alpha = 5$ is a safe choice, i.e. is weak enough that it does not exert an undue influence on the shape of the MAP solution. Comparison to the configurations from the prior shown in Fig. 1 confirms qualitatively that the posterior is much narrower that the prior.

Visual observation of the replayed MCMC sequence indicates that it takes several hundred steps in the sequence for the boundary to move from plus to minus one standard deviation about the mean position, a distance of a few pixels. Roughly speaking, one might expect that it would take on the order of $[2/(0.06\sqrt{2})]^2 \approx 500$ random steps of rms radial distance $0.06\sqrt{2}$ pixels to move a total distance of two pixels.

A quantitative estimate of the characteristics of the posterior is obtained by averaging over the MCMC sequence. Such an average of the grayscale image of the object is shown in Fig. 4, which is calculated as an

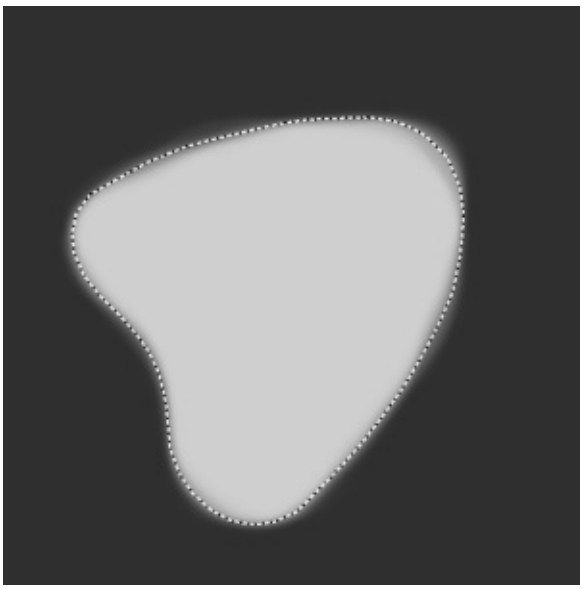

Figure 4: The average of the grayscale images for full MCMC sequence of samples from the posterior with the contour for the MAP reconstruction shown as a dotted line.

overlap of the boundary interior with the pixels of a 512×512 image for increased resolution. Of course, it does not make sense to average the positions of the vertices, because there is nothing to keep the polygon from slipping around the boundary of the object, which has no bearing on the actual object shape. The average MCMC image in Fig. 4 represents the posterior mean image, which may be interpreted as a probability image; the value of each pixel is the posterior probability that the pixel lies inside the boundary of the object. The amount of blur in the edge of the object indicates the variability in the position of that edge allowed by the posterior, i.e. the uncertainty in edge location.

Another way to summarize the uncertainty in boundary position is to display those pixels in the MCMC average image whose value lies between 0.025 and 0.975. When we do this for the present example, we find that indeed 92% of the original boundary lies inside the 95% credible interval [7].

Figure 4 also shows that the MAP reconstruction (the model that maximizes the posterior) appears to be consistent with the contour at half the amplitude of the posterior mean image. This result suggests that the posterior probability distribution is symmetric about its maximum. From the shape of the edge profile of the posterior average we tentatively conclude that the posterior may approximately be a multivariate Gaussian distribution, despite the nonlinear relation between the vertex parameters and the measurements.

An important feature of the MCMC technique is that any feature that one wishes to characterize, e.g., the average edge position and its uncertainty in the above example, is not conditional on the other parameters in the model. In the terms of probability theory, MCMC provides marginalized results, which means that the dependence on the uncertainties in "nuisance variables" is integrated out. In the context of the above example, the uncertainty in the edge position deduced for any particular location of the boundary is independent of the uncertainties in edge positions for the rest of the boundary.

5. Discussion

We have demonstrated their usefulness for solving a limit-angle tomographic problem of reconstructing a simple object from two views and have used MCMC to assess the uncertainties associated with the reconstruction.

We have not addressed in much detail the subject of selection of the hyperparameter α. We have chosen α on the basis of the degree of variation observed in the MCMC samples of the posterior, which indicates that the prior is not restricting the smoothness of the boundary very much. Since the value of α affects the width of the prior probability distribution, it influences the posterior uncertainty. Therefore, it is imperative to place more emphasis on understanding how to choose hyperparameters [3,12]. We note that this problem is not confined to Bayesian analysis.

Another limitation of the present study is that α is fixed. A more appropriate approach would be to consider α as a parameter that should be determined from the data, which goes by the name of empirical Bayesian analysis. By including α in the list of variables to be sampled in the MCMC process, the uncertainty in α would be taken into account in the overall uncertainty analysis.

In regard to our use of a polygon to represent a smooth curve, one would ideally like to reduce the length of the polygon sides to a size where this discretization does not affect the results. However, in the present formalism, shorter sides can lead to an increasing frustration of the MCMC algorithm, for example, because of the prior on side length. The result is that one is forced to use increasingly smaller MCMC steps, which leads to a reduced efficiency of the algorithm. One approach to overcoming this problem is to limit the direction of trial steps of each vertex to be along the bisector of angle between neighboring

sides, somewhat like the suggestion of Lobregt and Viergever[13]. Another approach avoids the use of the prior on polygon side length by using adaptable discretization, either on an edge-by-edge basis[13] or by routinely resampling the full polygon boundary to create equal-length sides. Another means that warrants investigation is to use a multiresolution representation of the boundary. The advantage would be that one could adjust the size of the MCMC steps at each resolution to improve the overall efficiency of the posterior sampling algorithm.

Acknowledgments

We have received valuable help and insight from many people, including John Skilling, Julian Besag, James Gubernatis, and Richard Silver. This work supported by the United States Department of Energy under contract W-7405-ENG-36.

References

[1] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *Inter. J. Comp. Vision*, 1:321–331, 1988.

[2] R. Szeliski and D. Terzopoulos. Physically-based and probabilistic models for computer vision. *Proc. SPIE*, 1570:140–152, 1991.

[3] D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4:415–447, 1992.

[4] D. S. Sivia. *Data Analysis: A Bayesian Tutorial*. Clarendon, Oxford Univ. Press, New York, 1996.

[5] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, London, 1995.

[6] K. M. Hanson and G. S. Cunningham. The Bayes inference engine. In K. M. Hanson and R. N. Silver, editors, *Maximum Entropy and Bayesian Methods*, pages 125–134. Kluwer Academic, Dordrecht, 1996.

[7] K. M. Hanson, G. S. Cunningham, and R. J. McKee. Uncertainty assessment for reconstructions based on deformable models. *Int. J. Imaging Systems and Technology*, 8, 1997.

[8] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 1996.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machine. *J. Chem. Phys.*, 21:1087–1091, 1953.

[10] K. M. Hanson, G. S. Cunningham, and S. S. Saquib. Inversion based on computational simulations. In G. Erickson, J. T. Rychert, and C. R. Smith, editors, *Maximum Entropy and Bayesian Methods*, pages 121–135. Kluwer Academic, Dordrecht, 1998.

[11] K. M. Hanson, G. S. Cunningham, G. R. Jennings, Jr., and D. R. Wolf. Tomographic reconstruction based on flexible geometric models. In *Proc. IEEE Int. Conf. Image Processing, vol. II*, pages 145–147. IEEE, 1994.

[12] G. S. Cunningham, A. Lehovich, and K. M. Hanson. Bayesian estimation of regularization parameters for deformable models. in *Medical Imaging: Image Processing*, K. M. Hanson, ed., *Proc. SPIE*, 3661:562–573, 1999.

[13] S. Lobregt and M. A. Viergever. A discrete dynamic contour model. *IEEE Trans. Med. Imaging*, MI-14:12–24, 1995.