

Short Papers

Approximate Bayes Factors for Image Segmentation: The Pseudolikelihood Information Criterion (PLIC)

Derek C. Stanford, *Member, IEEE*, and
Adrian E. Raftery

Abstract—We propose a method for choosing the number of colors or true gray levels in an image; this allows fully automatic segmentation of images. Our underlying probability model is a hidden Markov random field. Each number of colors considered is viewed as corresponding to a statistical model for the image, and the resulting models are compared via approximate Bayes factors. The Bayes factors are approximated using BIC (Bayesian Information Criterion), where the required maximized likelihood is approximated by the Qian-Titterton pseudolikelihood. We call the resulting criterion PLIC (Pseudolikelihood Information Criterion). We also discuss a simpler approximation, MMIC (Marginal Mixture Information Criterion), which is based only on the marginal distribution of pixel values. This turns out to be useful for initialization and it also has moderately good performance by itself when the amount of spatial dependence in an image is low. We apply PLIC and MMIC to a medical image segmentation problem.

Index Terms—BIC, color image quantization, ICM algorithm, image segmentation, Markov random field, medical image, mixture model, posterior model probability, pseudolikelihood, satellite image.

1 INTRODUCTION

In this paper, we consider the problem of determining the number of colors or gray levels to be used in presenting or interpreting an image. We are motivated primarily by problems in medical image segmentation. There, a segmentation is often desired to delineate organs, tumors or other features in pixelized X-ray, MRI, PET, CAT, or other images. If the image is segmented into too many colors, it may make the border finding problem harder, while too few colors may result in border information being lost. The idea is thus to find the number of colors needed to represent the information in the image (e.g., [1], [2]). The use of no more than the number of gray levels needed facilitates the presentation and analysis of mammograms, for example, [3].

Here, we introduce a new, model-based approach to this problem using approximate Bayes factors. We model the image in terms of a Markov random field and then each number of colors considered corresponds to a different statistical model for the image data. By doing this, we recast the problem of determining the number of colors in the image as a problem of statistical model comparison, and for this we use the standard Bayesian approach of Bayes factors. We propose an approximation to the Bayes factor based on the pseudolikelihood, called PLIC (for PseudoLikelihood Information Criterion). We also discuss a simpler approximate Bayes factor approach based on the marginal distribution of pixel values [4], called MMIC (for Marginal Mixture Information Criterion). This simpler approach is useful for initializing the PLIC method; MMIC takes no account of spatial structure, whereas PLIC takes account of it explicitly.

In Section 2, we review the Bayesian image models on which our work is based. In Section 3, we review the basic ideas behind

- D.C. Stanford is with Insightful Corp., 1700 Westlake Ave. North, Seattle, WA 98109. E-mail: stanford@insightful.com.
- A.E. Raftery is with the Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322. E-mail: raftery@stat.washington.edu.

Manuscript received 8 Mar. 2001; revised 21 Oct. 2001; accepted 19 Dec. 2001. Recommended for acceptance by P. Meer.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113772.

Bayes factors and present the PLIC and MMIC methods. In Section 4, we give the example of a medical image segmentation problem; the PLIC approach yields a good result.

2 BAYESIAN IMAGE MODELING

2.1 Markov Random Fields with Noise

The model that underlies our work is a standard one in Bayesian image analysis: a Markov random field with observation noise. We denote the value observed at pixel i by Y_i which will be a scalar for gray-scale images, and a vector for color or multispectral ones. For each observed Y_i , there is a corresponding discrete-valued unobservable state, X_i which determines the distribution of Y_i ; the set of all the X_i values is called the "true scene." Each possible state for X_i corresponds to a particular distribution of Y_i . The Y_i are assumed to be conditionally independent given the X_i and, so, dependence among the Y_i variables occurs only through dependence among the X_i values.

We impose a dependence structure on the X_i by using a Markov random field to model the true state of each pixel. This is a hidden Markov random field model because it is observable only through the Y_i values. Suppose that there are K possible states, so that X_i is an integer between 1 and K . We define $I(X_i, X_j)$ as an indicator function equal to 1 when $X_i = X_j$ and to zero otherwise. We let $N(X_i)$ be the set of neighbors of X_i . Here, we will take these to be the eight pixels adjacent to pixel X_i . We let $U(N(X_i), k)$ denote the number of points in $N(X_i)$ which have state k , so that $U(N(X_i), X_i)$ is the number of neighbors of pixel i which have the same state as pixel i . For our work here, we use the Potts model, defined as follows:

$$p(X) \propto \exp\left(\phi \sum_{i \sim j} I(X_i, X_j)\right), \quad (1)$$

where the sum is over all neighbor pairs, $i \sim j$. This leads to the following conditional distribution:

$$p(X_i = m | N(X_i), \phi) = \frac{\exp(\phi U(N(X_i), m))}{\sum_k \exp(\phi U(N(X_i), k))}. \quad (2)$$

The parameter ϕ expresses the amount of spatial homogeneity in the model. A positive value of ϕ means that neighboring pixels tend to be similar, while a negative value would mean that neighboring pixels tend to be dissimilar. If $\phi = 0$, then the pixels are independent. Note that pixels on the boundary of an image will not have a full set of observed neighbors. For simplicity and because the boundary is only a small fraction of the data, we exclude boundary pixels from the analysis except in their use as neighbors of interior pixels.

2.2 Parameter Estimation

The Iterated Conditional Modes (ICM) algorithm was introduced by [5]. It can be used as a method of image reconstruction when local characteristics of the true image can be modeled as a Markov random field. In particular, this can be used with the model described in (1). The algorithm begins with an initial estimate of the true scene X and proceeds iteratively to provide an estimate of the parameters of the conditional distribution of Y_i given X_i , as well as ϕ and X . An initial estimate of X , which is required for ICM, can be found through the simple marginal mixture EM segmentation method [4].

3 APPROXIMATE BAYES FACTORS FOR CHOOSING THE NUMBER OF COLORS OR TRUE GRAY LEVELS: PLIC AND MMIC

Our general approach to the problem of choosing the number of colors or true gray levels in an image, K , is to recast it as a statistical model selection problem and then to use the standard

Bayes factor approach to choose the appropriate model. The Markov random field model in Section 2 is viewed as defining not one model, but several, one for each value of K considered. Here, we review the basic ideas of Bayes factors and then introduce our pseudolikelihood-based approximation, PLIC, and the simpler marginal mixture approximation, MMIC.

3.1 Bayes Factors

The Bayesian approach to model comparison and model selection is based on *posterior model probabilities*. Given a set of models considered, $\{M_K : K = 1, \dots, K_{\max}\}$ and data Y , the posterior probability of model M_K is

$$p(M_K|Y) = \frac{p(Y|M_K)p(M_K)}{\sum_{L=1}^{K_{\max}} p(Y|M_L)p(M_L)}, \quad (3)$$

where $p(Y|M_K)$ is the integrated likelihood of model M_K and $p(M_K)$ is the prior probability of model M_K . Here, we will take the models to be equally likely a priori, so that

$$p(M_K) = 1/K_{\max} \quad (K = 1, \dots, K_{\max}).$$

The integrated likelihood, $p(Y|M_K)$, is defined by

$$p(Y|M_K) = \int p(Y|\theta_K, M_K)p(\theta_K)d\theta_K, \quad (4)$$

where θ_K is the parameter (vector) for model M_K , $p(Y|\theta_K, M_K)$ is the (usual) likelihood and $p(\theta_K)$ is the prior distribution. The quantity

$$B_{KL} = \frac{p(Y|M_K)}{p(Y|M_L)} \quad (5)$$

is known as the *Bayes factor* for model M_K against model M_L . See [6] and [7] for reviews of Bayes factors.

Evaluating the integral in (4) is often hard and much of the research in this area has focused on ways of doing it. A simple, but often reasonably good approximation is

$$2 \log p(Y|M_K) \approx BIC = 2 \log p(Y|\hat{\theta}_K, M_K) - d_K \log(N), \quad (6)$$

with $\hat{\theta}_K$ being the maximum-likelihood estimator of θ_K and d_K denoting $\dim(\theta_K)$, the number of parameters in model M_K . Under regularity conditions that are roughly those that guarantee consistency and asymptotic normality of $\hat{\theta}_K$, the error in this approximation is $O(1)$ regardless of the prior $p(\theta_K|M_K)$ ([8], [9]). If, in addition, the prior $p(\theta_K|M_K)$ is a unit information prior (i.e., a multivariate normal prior distribution centered at the MLE with variance matrix equal to the inverse of the Fisher information matrix for one observation), then the error is $O(N^{-1/2})$ [10]. Raftery [11] argues that this prior, while proper, is conservative and, so, may be appropriate as the basis for a baseline reference analysis. Differences of 10 or more in BIC values between competing models are conventionally viewed as constituting strong evidence for the favored model over its competitor [6].

3.2 Penalized Pseudolikelihood Criterion: PLIC

We wish to conduct inference for K , the number of segments in the image, using the Bayesian approach in Section 3.1. Here, M_K refers to the Potts model (1) with K components. Thus, θ_K consists of ϕ and the parameters of the conditional distribution of Y_i given X_i . We propose the use of a criterion derived from the *BIC*. A similar approach was used by [12] for choosing between different Markov random field models in the case where the true scene is directly observed and the number of segments is known in advance; neither of these is the case in the applications we have in mind. When, as here, the true scene is not observed, this would require evaluation of the likelihood of the observed data, $L(Y|K)$, namely,

$$L(Y|K) = \sum_x p(Y|X=x, K)p(X=x|K). \quad (7)$$

The sum in (7) involves all possible configurations of the hidden states. With N pixels and K states, there are K^N possible configurations, which is huge, making this approach intractable. Instead, we approximate the likelihood term by a pseudolikelihood proposed by [13] and [14] which maintains computational feasibility.

The basic idea of this pseudolikelihood is that instead of summing over all possible configurations of X , we consider only configurations that are close to the ICM estimate of X , denoted by \hat{X} . Specifically, we consider each pixel Y_i , in turn, and condition on \hat{X}_{-i} , which is \hat{X} excluding the value at X_i . We then obtain the following conditional likelihood, in which $N(\hat{X}_i)$ denotes the neighbors of \hat{X}_i :

$$L(Y_i|\hat{X}_{-i}, K) = \sum_{j=1}^K p(Y_i|X_i=j)p(X_i=j|N(\hat{X}_i)). \quad (8)$$

The first term in the sum, $p(Y_i|X_i=j)$, simply requires evaluation of the conditional density of Y_i given X_i ; the second term, $p(X_i=j|N(\hat{X}_i))$ is evaluated using (2). The conditional likelihoods from (8) are combined to form the Qian-Titterton pseudolikelihood of the image, as follows:

$$\begin{aligned} L_{\hat{X}}(Y|K) &= \prod_i f(Y_i|\hat{X}_{-i}, \hat{\phi}) \\ &= \prod_i \sum_{j=1}^K f(Y_i|X_i=j)p(X_i=j|N(\hat{X}_i), \hat{\phi}). \end{aligned} \quad (9)$$

Returning to the formula for *BIC*, we replace the intractable $L(Y|K)$ by the easily computable $L_{\hat{X}}(Y|K)$ from (9), to obtain our new approximation

$$PLIC(K) = 2 \log(L_{\hat{X}}(Y|K)) - d_K \log(N). \quad (10)$$

Because we do not expect the model assumptions to hold for values of K very far from the true value, we adopt a sequential approach to choosing K to maximize $PLIC(K)$. We begin by computing $PLIC(K)$ for $K=1$, and then incrementally increase the value of K . At each step, we compare $PLIC(K)$ with $PLIC(K-1)$ and stop the process when the smaller model is preferred. In other words, as we increase K incrementally from $K=1$, we take the first local maximum of $PLIC(K)$ to be our choice for the number of segments K .

3.3 MMIC: A Simpler Bayes Factor Approximation for Initialization and Fast Computation

A faster approximate Bayes factor approach to choosing the number of colors or true gray levels is available by just considering the empirical marginal distribution of pixel values and ignoring their spatial locations. MMIC is much faster computationally than the PLIC method and it can be used as an initialization step for PLIC. If the data are generated by the Markov random field model (1), then the marginal distribution of pixels is a finite mixture of K distributions, each equal to the conditional distribution of Y_i given that X_i takes on one of its K possible values. The basic idea of MMIC is to compute the BIC value for each number of components in this finite mixture distribution.

Suppose we have observations $Y = (Y_1, \dots, Y_N)$ from a mixture model with K components. Let P_j denote the mixture proportion of the j th component. Let Φ be the mixing density (for example, for a Gaussian mixture, Φ is a single Gaussian density) with $\theta = (\theta_1, \dots, \theta_K)$ giving the parameters for the K components. The marginal density for a single observation, Y_i , is

$$f(Y_i|K, \theta) = \sum_{j=1}^K P_j \Phi(Y_i|\theta_j). \quad (11)$$

Then, the loglikelihood for Y , assuming that all of the Y_i are independent, is

$$\log p(Y|K, \theta) = \sum_{i=1}^N \log \left(\sum_{j=1}^K P_j \Phi(Y_i|\theta_j) \right). \quad (12)$$

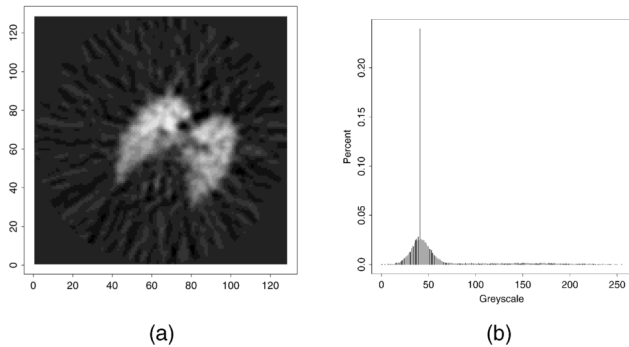


Fig. 1. (a) PET Image of a dog lung. (b) Marginal histogram of the dog lung image.

After choosing θ to maximize the loglikelihood from (12), the maximized loglikelihood can be used in the *BIC* formula (6). We refer to the resulting approximation as the Marginal Mixture Information Criterion (MMIC):

$$MMIC(K) = 2 \log p_{MLE}(Y|K) - d_K \log(N), \quad (13)$$

where d_K is the number of parameters in the mixture model with K components.

We view MMIC as only a heuristic guideline because the likelihood on which it is based is correct only if the Y_i are independent, which is not the case for images, and also because the regularity conditions for BIC to approximate the Bayes factor do not hold for mixture models. However, it is known that choosing a model based on BIC produces consistent density estimates when the data are independent and one-dimensional [15] and a proof of the consistency of BIC as a model selection criterion for mixture models has been produced [16]. In addition, BIC has given good results for choosing the number of components in a wide range of applications of mixture models ([15], [17], [18], [4], [19]).

4 EXAMPLE: MEDICAL IMAGE SEGMENTATION

Fig. 1a shows a PET image of a dog lung. This image was obtained from Dr. H.T. Robertson at the University of Washington Division of Pulmonary and Critical Care. The goal of the analysis here is to delineate the lung in the image automatically.

It is clear from Fig. 1a that the actual image area is circular, with the corners of the image filled in with a constant gray value; this reflects the way the PET image is taken. This sort of artifact can be removed easily with a mixture model; one of the components converges to a spike for that gray level, separating it from the rest of the data. This can be clearly seen in the marginal histogram of the image shown in Fig. 1b.

TABLE 1
PLIC and MMIC Results for the Dog Lung Image

Segments	$\hat{\phi}$	Logpseudolikelihood	PLIC	MMIC
1		-80641.02	-161311.13	-166007.1
2	1.14	-61081.56	-122221.33	-137193.0
3	1.20	-59542.47	-119172.25	-137019.8
4	1.67	-50961.33	-102039.06	-128746.2
5	1.07	-52444.73	-105034.97	-135323.5
6	1.11	-52090.24	-104355.08	-135333.6

The preferred number of segments is shown in boldface for each method.

Table 1 shows that PLIC decisively chooses four segments for this image. In the context of PET imagery, the choice of four segments is quite reasonable for this image. Two segments are needed for the background: one to model the spike (due to the corner artifact) and one for the general background. Since the image is constructed based on radioactive emissions from gas in the lung, it is not surprising to see two segments for the lung itself to account for the high gas density in the interior of the lung and the somewhat lower gas density around the periphery. For this case, MMIC also chooses four segments using only the marginal grayscale values from the image.

The results of the marginal mixture EM segmentation are shown in Fig. 2a. This does a reasonable job of separating the lung from the background, especially given that it makes no use of spatial context. It provides a good initialization for the segmentation methods that do take account of spatial context.

The ICM refinement, shown in Fig. 2b, does a good job of reducing clutter in the image and gives a qualitatively satisfactory answer. We also applied morphological smoothing [20] to Fig. 2b; the final result is shown in Fig. 2c. This smoothes the outer edge of the lung, and it also removes most of the other clutter in the image. The small spot separate from the lung and below could be removed by simply considering the lung to be the largest connected component. The small void in the center of the lung is not artifactual; it is real.

Subjective evaluation of this result by the researchers who provided us with this image found that our final segmentation provided a good delineation of the lung as a region of interest for further analysis; also, the segmentation based on four segments was preferred to those based on three or five segments, thus providing some practical validation of our method in this case.

The final result shows that this segmentation method has promise for the purpose of automatically processing a database of lung images to separate out the region of interest (the lung) from the

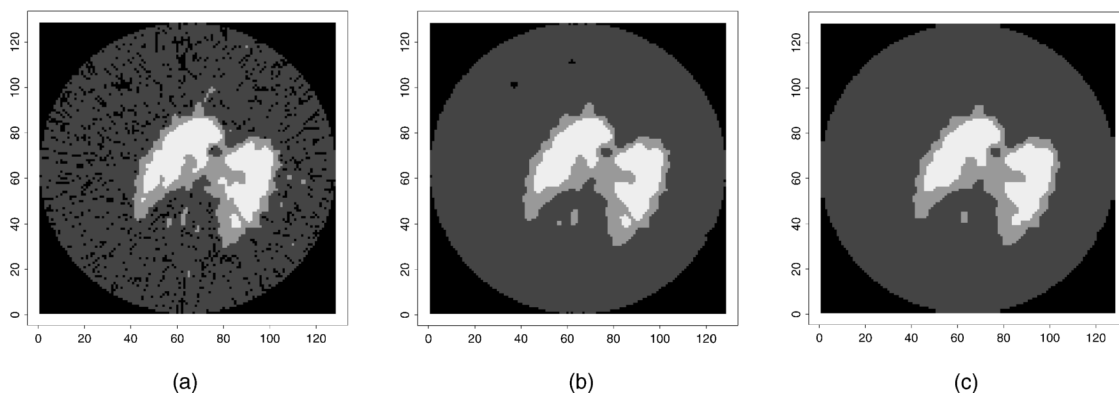


Fig. 2. (a) Marginal mixture EM segmentation of the dog lung image into four segments. (b) Segmentation of the dog lung image into four segments by the ICM algorithm, initialized by the marginal mixture EM segmentation result. (c) Final segmentation of the dog lung image into four segments after morphological smoothing (opening and closing, conditional on the edge pixels).

background. Currently, the most widely used method for this sort of segmentation is for a human expert to manually outline the lung with an interactive computer program, a process which is quite tedious and can take a long time for a large database of images. The automatic segmentation algorithm can obviate the need for the manual process, requiring only human inspection of the results.

5 DISCUSSION

We have proposed an approximate Bayesian method for determining the number of colors or gray levels in a noisy image from the image data themselves. This number has usually been either predetermined or chosen in an ad hoc manner, but the literature shows that there are many instances where a more formal, data-based, method could be useful. Our method works well in our medical image segmentation example and seems to be potentially useful for a variety of other applications, including automatic segmentation of satellite images, color and gray-level image quantization, and the use of cooccurrence matrices. Color image quantization is the process by which an original color image is mapped into an output image with a limited number of colors, while attempting to preserve the image quality; this arises because real-world images typically come in many colors, whereas output devices can often display far fewer. In some cases, the number of colors used in an image can be reduced by a factor of 1,000 or more without much decrease in quality [21]. Cooccurrence matrices, which are used for assessing texture in images ([22], [23]), are defined in terms of groups of pixel values called "bins" or "colors." Choice of the number of bins has an effect on the effectiveness of cooccurrence matrices in characterizing textures [24].

Our method is fully defined for multispectral and color images, but we have shown examples of its use only for gray-scale images. For multispectral images, the noise distribution would often be taken to be multivariate normal. For initialization, the marginal mixture EM segmentation method is still available, for example, using model-based clustering [18]. If the number of pixels is too large for this to be efficient, a subsample of the pixels could be used for the clustering, and then discriminant analysis applied to classify the remaining pixels [25], or an efficient method based on the minimal spanning tree could be used [26].

Our approach has been to compute approximate Bayes factors using the ICM algorithm and the pseudolikelihood of [13]. Another approach would be to carry out a fully Bayesian analysis of the image model using Markov chain Monte Carlo, as suggested by [27], and then to estimate the integrated likelihood using one of the available methods for doing this from MCMC (e.g., [28], [29], [30], [31]). This would conform more fully to the Bayesian paradigm, but could be extremely expensive computationally.

Our method could also be used for choosing between competing higher-order interaction Markov random field models for texture in images, e.g., between special cases of the Chien model of [32] or the models of [33]. In some situations, one might expect there to be a relationship between the number of colors used and the complexity of the texture model needed to describe the resulting pattern. Our approach could be generalized to choosing both of these simultaneously by identifying each combination of texture model and number of colors with the corresponding probability model and then comparing the resulting probability models using approximate Bayes factors, calculated using methods based on those described here.

ACKNOWLEDGMENTS

The authors would like to thank Dr. H.T. Robertson of the University of Washington Division of Pulmonary and Critical Care for helpful discussions and for providing us with the dog lung image. Dr. Stanford was supported by the US Defense Advanced Research Projects Agency under grant DAAH01-99-C-R248. Dr. Raftery was supported by the US Office of Naval Research under grants N00014-96-1-0192 and N00014-96-1-0330.

REFERENCES

- [1] S.E. Umbaugh, R.H. Moss, W.V. Stoecker, and G.A. Hance, "Automatic Color Segmentation Algorithms with Application to Skin Tumor Feature Identification," *IEEE Eng. in Medicine and Biology Magazine*, vol. 12, pp. 75-82, 1993.
- [2] G.A. Hance, S.E. Umbaugh, R.H. Moss, and W.V. Stoecker, "Unsupervised Color Image Segmentation," *IEEE Eng. in Medicine and Biology Magazine*, vol. 15, pp. 104-111, 1996.
- [3] J.W. Byng, J.P. Critten, and M.J. Yaffe, "Thickness-Equalization Processing for Mammographic Images," *Radiology*, vol. 203, pp. 564-568, 1997.
- [4] D.C. Stanford, "Fast Automatic Unsupervised Image Segmentation and Curve Detection in Spatial Point Processes," PhD thesis, Univ. of Washington, 1999.
- [5] J. Besag, "Statistical Analysis of Dirty Pictures," *J. Royal Statistical Soc., Series B*, vol. 48, pp. 259-302, 1986.
- [6] R.E. Kass and A.E. Raftery, "Bayes Factors," *J. Am. Statistical Assoc.*, vol. 90, pp. 773-795, 1995.
- [7] A.E. Raftery, "Bayesian Model Selection in Social Research," *Sociological Methodology*, P.V. Marsden, ed., pp. 111-163, Cambridge: Blackwells, 1995.
- [8] D.R. Cox and D. Hinkley, *Problems and Solutions in Theoretical Statistics*. London: Chapman and Hall, 1978.
- [9] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, pp. 461-464, 1978.
- [10] R.E. Kass and L. Wasserman, "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion," *J. Am. Statistical Assoc.*, vol. 90, pp. 928-934, 1995.
- [11] A.E. Raftery, "Bayes Factors and BIC—Comment on 'A Critique of the Bayesian Information Criterion for Model Selection'," *Sociological Methods and Research*, vol. 27, pp. 411-427, 1999.
- [12] C. Ji and L. Seymour, "A Consistent Model Selection Procedure for Markov Random Fields based on Penalized Pseudolikelihood," *Annals of Applied Probability*, vol. 6, pp. 423-443, 1996.
- [13] W. Qian and D.M. Titterton, "Estimation of Parameters in Hidden Markov Models," *Philosophical Trans. Royal Soc. of London*, vol. 337, pp. 407-428, 1991.
- [14] W. Qian and D.M. Titterton, "Stochastic Relaxations and EM Algorithms for Markov Random Fields," *J. Statistical Computing and Simulations*, vol. 40, pp. 55-69, 1991.
- [15] K. Roeder and L. Wasserman, "Practical Bayesian Density Estimation Using Mixtures of Normals," *J. Am. Statistical Assoc.*, vol. 92, pp. 894-902, 1997.
- [16] C. Keribin, "Consistent Estimate of the Order of Mixture Models," *Comptes Rendus de l'Academie des Sciences*, vol. 326, pp. 243-248, 1998.
- [17] A. Dasgupta and A.E. Raftery, "Detecting Features in Spatial Point Processes with Clutter via Model-Based Clustering," *J. Am. Statistical Assoc.*, vol. 93, pp. 294-302, 1998.
- [18] C. Fraley and A.E. Raftery, "How Many Clusters? Which Clustering Method?—Answers via Model-Based Cluster Analysis," *Computer J.*, vol. 41, pp. 578-588, 1998.
- [19] D.C. Stanford and A.E. Raftery, "Finding Curvilinear Features in Spatial Point Processes: Principal Curve Clustering with Noise," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 601-609, 2000.
- [20] J. Serra, *Image Analysis and Mathematical Morphology*. New York: Academic Press, 1982.
- [21] S. Dixit, "Quantization of Color Images for Display Printing on Limited Color Output Devices," *Computers and Graphics*, vol. 15, pp. 561-567, 1991.
- [22] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 3, pp. 610-621, 1973.
- [23] R.M. Haralick, "Statistical and Structural Approaches to Texture," *Proc. IEEE*, vol. 67, pp. 786-804, 1979.
- [24] F.M.J. Valckx and J.M. Thijssen, "Characterization of Echographic Image Texture by Cooccurrence Matrix Parameters," *Ultrasound in Medicine and Biology*, vol. 23, pp. 559-571, 1997.
- [25] J.D. Banfield and A.E. Raftery, "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, vol. 49, pp. 803-821, 1993.
- [26] C. Posse, "Hierarchical Model-Based Clustering for Large Datasets," *J. Computational and Graphical Statistics*, to appear, 2000.
- [27] J. Besag, J. York, and A. Mollie, "Bayesian Image Restoration, with Two Applications in Spatial Statistics," *Annals of the Inst. Math. Statistics*, vol. 43, pp. 1-59, 1991.
- [28] M.A. Newton and A.E. Raftery, "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap (with discussion)," *J. Royal Statistical Soc., Series B*, vol. 56, pp. 3-48, 1994.
- [29] S. Chib, "Marginal Likelihood from the Gibbs Output," *J. Am. Statistical Assoc.*, pp. 1313-1321, 1995.
- [30] A.E. Raftery, "Hypothesis Testing and Model Selection," *Markov Chain Monte Carlo in Practice*, W.R. Gilks, D.J. Spiegelhalter, and S. Richardson, eds., pp. 163-188, London: Chapman and Hall, 1996.
- [31] T.J. DiCiccio, R.E. Kass, A.E. Raftery, and L. Wasserman, "Computing Bayes Factors by Combining Simulation and Asymptotic Approximations," *J. Am. Statistical Assoc.*, vol. 92, pp. 903-915, 1997.
- [32] X. Descombes, R. Morris, and J. Zerubia, "Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood," Technical Report 3015, INRIA, France, 1996.
- [33] H. Tjelmeland and J. Besag, "Markov Random Fields with Higher Order Interactions," *Scandinavian J. Statistics*, vol. 25, pp. 415-434, 1998.