Determining the Number of Colors or Gray Levels in an Image Using Approximate Bayes Factors: The Pseudolikelihood Information Criterion (PLIC)

Derek C. Stanford and Adrian E. Raftery [1]

Technical Report no. 390
Department of Statistics
University of Washington

February 15, 2001

[1]Derek C. Stanford is a Research Scientist at Mathsoft, Inc., 1700 Westlake Ave N, Seattle, WA 98109. E-mail: stanford@statsci.com. Adrian E. Raftery is Professor of Statistics and Sociology at the Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322, USA. E-mail: raftery@stat.washington.edu. Web: http://www.stat.washington.edu/raftery. 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. Raftery was supported by the Office of Naval Research under grants N00014-96-1-0192 and N00014-96-1-0330.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
15 FEB 2001		**00-02-2001 to 00-02-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Determining the Number of Colors or Gray Levels in an Image Using Approximate Bayes Factors: The Pseudolikelihood Information Criterion (PLIC)	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington,Department of Statistics,Box 354322,Seattle,WA,98195-4322**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**25**	

Abstract

We propose a method for choosing the number of colors, or true gray levels, in an image. This is motivated by medical and satellite image segmentation, and may also be useful for color and gray scale image quantization, the display and storage of computer-generated holograms, and the use of cooccurrence matrices for assessing texture in 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, where the required maximized likelihood is approximated by the Qian-Titterington 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 also to have moderately good, albeit suboptimal, performance in its own right. We apply PLIC to three examples: a simulated two-band image, a medical segmentation problem, and a satellite image, and in each case it gives good results in practice.

Keywords: BIC; Color image quantization; Cooccurrence matrix; Hologram; ICM algorithm; Image segmentation; Markov Random Field; Medical image; Mixture model; Posterior model probability; Pseudolikelihood; Satellite image.

Contents

List of Tables

List of Figures

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. Umbaugh et al. (1993); Hance et al. (1996)). Using no more than the number of gray levels needed facilitates the presentation and analysis of mammograms, for example Byng et al. (1997).

A second area of application of this problem is color image quantization, 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 1000 or more without much decrease in quality Dixit (1991). Cluster analysis has been extensively used to find the best color mapping (Dixit (1991); Xiang and Joy (1994); Xiang (1997); Scheunders (1997b); Scheunders (1997a); Papamarkos (1999)). Clearly, the fewer colors that can be used without degrading the image, the better for many purposes. However, the number of colors used has in general been subjectively defined by the user.

A third area of application is the display and storage of computer-generated holograms. Bokor and Papp (1996) investigated the effect of the number of gray levels used on image fidelity. They found, rather surprisingly, that image fidelity was maximized in their experiments when binary holograms (with 2 gray levels) were used. Burr et al. (1998) showed that 5 gray levels is optimal for hologram storage in the absence of signal-dependent noise, but that because of the presence of signal-dependent noise, 3 gray levels will most likely be the optimal choice for a practical system. These results suggest that the use of small numbers of gray levels is advantageous, and that the appropriate number may depend on the situation, pointing to the need for a data-based way of choosing the number.

A fourth area of application arises in the use of cooccurrence matrices for assessing texture in images (Haralick et al. (1973); Haralick (1979)). The cooccurrence matrix of an image at distance $d = (d_1, d_2)$ is the empirical discrete joint distribution of the pixel value at location (x_1, x_2) and the value at location $(x_1 + d_1, x_2 + d_2)$, after the observed pixel values have

been grouped into G bins or "colors". Valckx and Thijssen (1997) considered the effect of the number of bins on the effectiveness of cooccurrence matrices in characterizing textures in echograms. They concluded that in gray-scale echograms with 256 gray levels, far fewer were needed to differentiate textures; they found 64 bins to be enough.

There thus seems to be a widespread feeling in the literature that image analysis can often gain by using a relatively small number of colors or "true gray levels". Also, it often seems that the underlying subject matter theory or knowledge is not enough to completely determine the number of colors to be used, even though it may often suggest a reasonable range. This points to the need to use the image data themselves to help make this determination. However, there is a striking lack of systematic data-based approaches to deciding how many colors should be used. Umbaugh et al. (1992) used "automatic induction" to determine the number of colors, but they did not describe the method in detail. Godtliebsen and Chu (1995) used a method based on kernel density estimation from the marginal gray-scale histogram.

Here we introduce a new, model-based, approach to this problem that uses 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 Pseudolikehood Information Criterion).

We also discuss a simpler approach based on the marginal distribution of pixel values. We use the fact that this is a finite mixture distribution, and we use approximate Bayes factors to choose the number of colors. This is based on the idea of marginal mixture EM segmentation Stanford (1999), and we call it MMIC (for Marginal Mixture Information Criterion). This simpler approach is useful for initializing our preferred PLIC method, and also works moderately well in its own right. The big difference between the two methods is that 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 Bayes factors and give the marginal mixture EM segmentation method. We then introduce the PLIC approximation. In Section 4, we give three examples. The first is a simulated two-color example, and there PLIC gives a good answer, whereas MMIC is less satisfactory; this illustrates the value of taking account of the spatial structure. We then consider a medical image segmentation problem, and finally a

satellite image. In both of these latter examples, PLIC gives good answers.

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, 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 8 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:

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

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

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

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) algorthm was introduced by Besag (1986). 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 Stanford (1999). This is based on the fact that if, for example, the conditional distribution of Y_i given X_i is Gaussian, then the marginal distribution of the observed pixel values, Y_i, is a finite mixture of Gaussians with K components. The EM algorithm Dempster et al. (1977) can be used to fit a Gaussian mixture model to this distribution, and a maximum likelihood classification of each of the pixels into one of the K components yields an initial estimate of X.

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. In Section 3.1 we review the basic ideas of Bayes factors, in Section 3.2 we introduce our pseudolikelihood-based approximation, PLIC, and in Section 3.3 we discuss 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, $\{M_K : K = 1, \ldots, K_{\max}\}$ and data D, the posterior probability of model M_K is

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

where $p(D|M_K)$ 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 $p(M_K) = 1/K_{\max}$ $(K = 1, \ldots, K_{\max})$.

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

$$p(D|M_K) = \int p(D|\theta_K, M_K)p(\theta_K)d\theta_K, \tag{4}$$

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

$$B_{KL} = \frac{p(D|M_K)}{p(D|M_L)} \tag{5}$$

is known as the *Bayes factor* for model M_K against model M_L. See Kass and Raftery (1995), Raftery (1995), and Hoeting et al. (2000) for reviews of Bayesian model selection and 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

$$2 \log p(D|M_K) \approx BIC \tag{6}$$

where

$$BIC = 2 \log p(D|\hat{\theta}_K, M_K) - N \log(\dim(\theta_K)), \tag{7}$$

with $\hat{\theta}_K$ being the maximum likelihood estimator of θ_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 $p(\theta_K|M_K)$ (Cox and Hinkley (1978); Schwarz (1978)). 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 $O(N^{-1/2})$ Kass and Wasserman (1995). Raftery (1999) 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 consituting strong evidence for the favored model over its competitor Kass and Raftery (1995).

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. An initial thought would be to use BIC, as defined by (7). A proposal along these lines was made by Ji and Seymour (1996) for choosing between different Markov random field models in the case where the true scene is directly observed and the number of colors or true gray levels is known in advance. Neither of these is the case in the applications we have in mind, and Ji and Seymour (1996) did not consider the choice of the number of segments. When, as here, the true scene is not observed, this would require evaluation of the likelihood of the observed data, $L(Y|K)$, namely

$$L(Y|K) = \sum_x p(Y|X = x, K)p(X = x|K). \tag{8}$$

The sum in (8) 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 Qian and Titterington (1991a) and Qian and Titterington (1991b), 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 $\hat{X}_i$:

$$L(Y_i|\hat{X}_{-i}, K) = \sum_{j=1}^{K} p(Y_i|X_i = j)p(X_i = j|N(\hat{X}_i)). \tag{9}$$

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 (9) are combined to form the Qian-Titterington pseudolikelihood of the image, as follows:

$$L_{\hat{X}}(Y|K) = \prod_i f(Y_i|\hat{X}_{-i}, \hat{\phi}) \tag{10}$$

$$= \prod_i \sum_{j=1}^{K} f(Y_i|X_i = j)p(X_i = j|N(\hat{X}_i), \hat{\phi}). \tag{11}$$

We replace the intractable $L(Y|K)$ by the easily computable $L_{\hat{X}}(Y|K)$ from (11), to obtain our new approximation

$$PLIC(K) = 2\log(L_{\hat{X}}(Y|K)) - D_K \log(N). \tag{12}$$

Ideally, one could compute $PLIC(K)$ for a large range of K values and then choose K to maximize $PLIC(K)$. However, we do not expect the model assumptions to hold for values of K very far from the true value. Because of this, we adopt a sequential approach. 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. This is much faster computationally than the PLIC method. It turns out to give results that are often reasonable at least for initializing the more demanding PLIC method, although it has the clear disadvantage relative to PLIC that it does not take account of salient spatial information.

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

$$f(Y_i|K, \theta) = \sum_{j=1}^{K} P_j \Phi(Y_i|\theta_j). \tag{13}$$

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

$$\log p(Y|K, \theta) = \sum_{i=1}^{N} \log(\sum_{j=1}^{K} P_j . \Phi(Y_i|\theta_j)). \tag{14}$$

After choosing θ to maximize the loglikelihood from (14), the maximized loglikelihood can be used in the BIC formula (7). We refer to the resulting approximation as the Marginal

Mixture Information Criterion (MMIC):

$$MMIC(K) = 2 \log p_{MLE}(Y|K) - d_K \log(N), \tag{15}$$

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

We view MMIC as a heuristic guideline rather than a rigorously justified approximation, because the likelihood on which it is based is correct if the Y_i are independent, which is not the case, and also because the usual 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 Roeder and Wasserman (1997), and a proof of the consistency of BIC as a model selection criterion for mixture models has been produced Keribin (1998). In addition, BIC has given good results for choosing the number of components in a wide range of applications of mixture models (Roeder and Wasserman (1997); Dasgupta and Raftery (1998); Fraley and Raftery (1998); Stanford (1999); Stanford and Raftery (2000)).

4 Examples

4.1 Simulated Two-Segment Image

The simulated two-segment image shown in Figure 1 is made up of two solid bands, with mean greyscale values of 120 and 140. Independent Gaussian noise with mean 0 and standard deviation 10 was added to each pixel, and then the values were rounded to integers.

The two bands are close enough together on average that the marginal information alone makes it hard to see that there are actually two segments here. The marginal histogram of the pixels in the simulated image is shown in Figure 2, and this is quite unimodal. Visually it is quite clear from the image, Figure 1, that there are two bands, but this is due to the spatial arrangement of the pixels; this is a case where there should be advantage to taking account of the spatial structure.

Table 1 shows the PLIC and MMIC values for this image. PLIC, which takes account of the spatial information, correctly and quite decisively determines that there are two segments, while MMIC identifies only one segment (although the difference of about 3 between the MMIC values for one and two segments would be viewed as indecisive).

Compare this to the scrambled image in Figure 3, which is a random reordering of the pixels from Figure 1. Since these two images contain exactly the same pixels (just in a different order), their marginal information will be the same. Thus the histogram of the

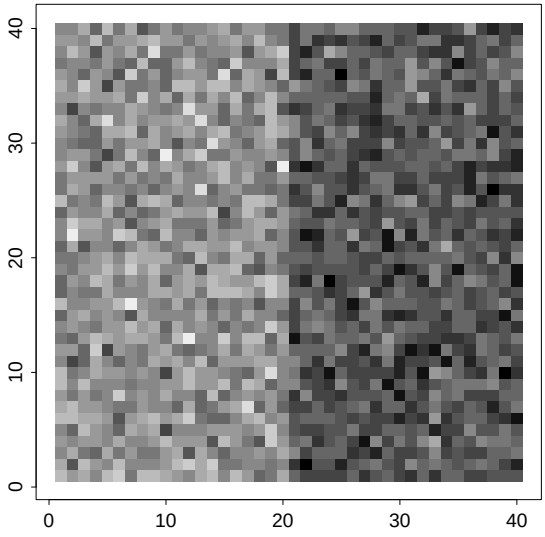

Figure 1: Simulated Two-Segment Image.

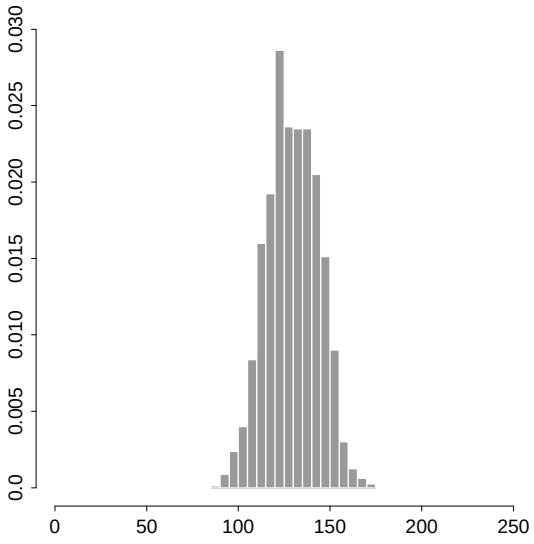

Figure 2: Marginal Histogram of the Simulated Two-Segment Image in Figure 1.

Table 1: PLIC and MMIC Results for the Simulated Two-Segment Image and the Scrambled Image. The number of segments preferred by a method for the given image is shown in boldface.

	Two-Segment Image			Scrambled Image		
Segments	$\hat{\phi}$	PLIC	MMIC	$\hat{\phi}$	PLIC	MMIC
1	–	-11731.28	**−12998.79**	–	**−11756.49**	**−12998.79**
2	2.06	**−10784.69**	-13002.17	0.05	-11833.84	-13002.17
3	0.45	-10948.62	-13021.64	0.19	-11868.36	-13021.64

values in the scrambled image, Figure 3, is the same as for the original image in Figure 1, and so is also shown in Figure 2. Table 1 shows that both PLIC and MMIC choose one segment for the scrambled image – PLIC quite decisively and MMIC just barely. From most points of view, this is the correct answer in this case. Again, PLIC seems to perform better, because it chooses the right answer decisively, whereas MMIC is uncertain between the two answers.

The estimated ϕ values used in computing PLIC are also shown in the table. The large ϕ for the two segment fit of the two-segment image indicates that a large degree of spatial association is found; this results in a much higher value of PLIC than is obtained for either one or three segments.

4.2 Medical Image Segmentation: Dog Lung

Figure 4 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 Figure 4 that the actual image area is circular, with the corners of the image filled in with a constant grey value; this reflects the way the PET image is taken. This sort of artifact can be removed quite easily with a mixture model; one of the components converges to a spike for that grey level, effectively separating it from the rest of the data. This can be clearly seen in the marginal histogram of the image shown in Figure 5.

Table 2 shows that PLIC quite 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

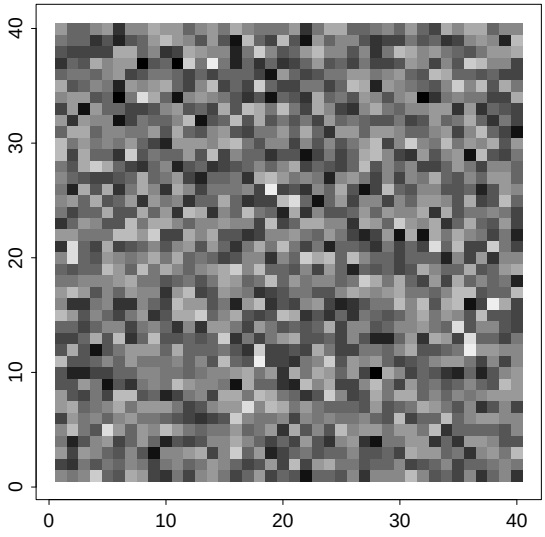

Figure 3: Scrambled Version of Figure 1.

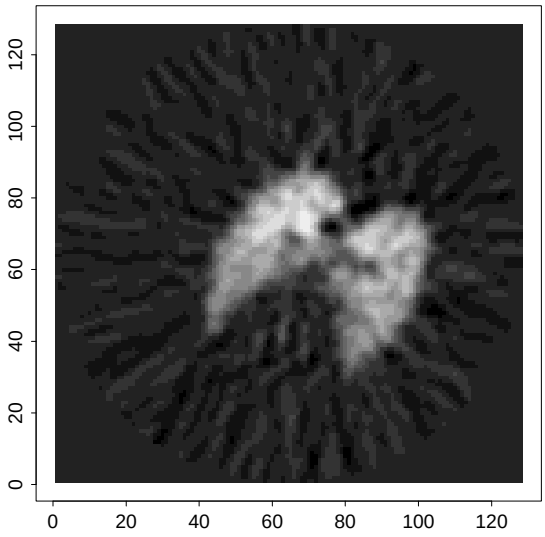

Figure 4: PET Image of a Dog Lung, Before Processing.

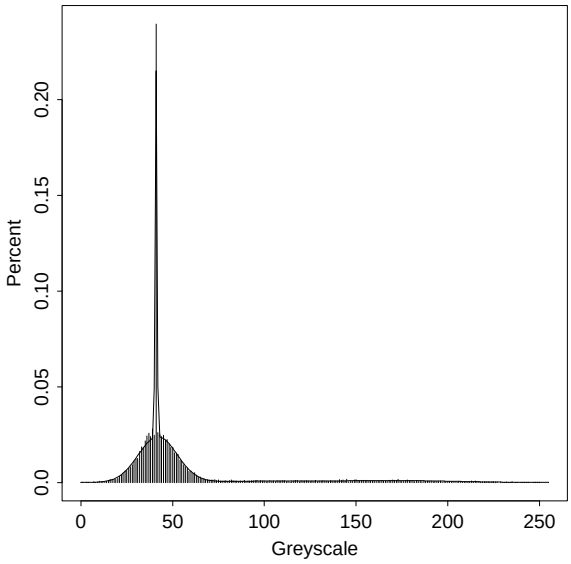

Figure 5: Marginal Histogram of the Dog Lung Image.

emissions from gas in the lung, it is not at all 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 4 segments using only the marginal greyscale values from the image.

The results of the marginal mixture EM segmentation are shown in Figure 6. 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.

Table 2: PLIC and MMIC Results for the Dog Lung Image. The preferred number of segments is shown in boldface for each method.

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

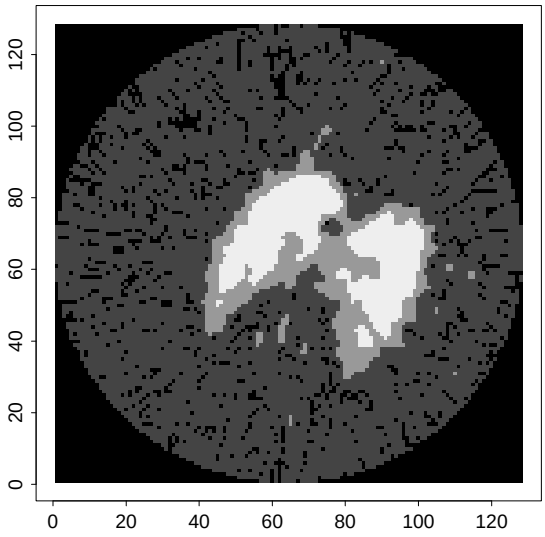

Figure 6: Marginal Mixture EM Segmentation of the Dog Lung Image into Four Segments.

The ICM refinement, shown in Figure 7, does a good job of reducing clutter in the image, and gives a qualitatively satisfactory answer. We also applied morphological smoothing Matheron (1975); Serra (1982) to Figure 7; the final result is shown in Figure 8. This gives additional smoothing to 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 easily 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.

We shared our results with the researchers who provided us with this image and asked for their subjective evaluation of the segmented image. They found that our final segmentation provided a good delineation of the lung as a region of interest for further analysis and felt that the segmentation based on four segments was more useful than those based on three or five segments, thus providing some clinical 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 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,

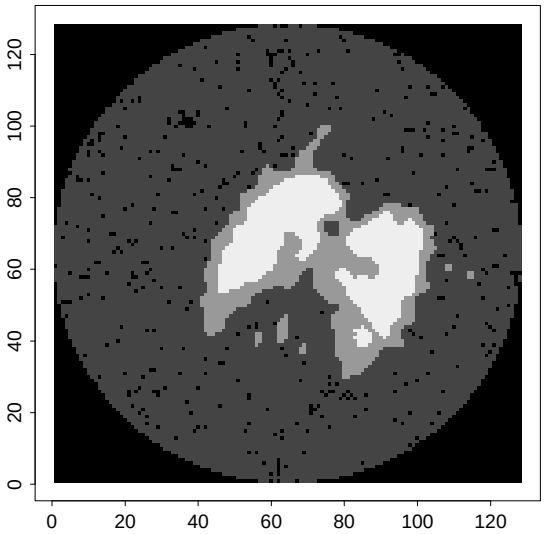

Figure 7: Segmentation of the Dog Lung Image into 4 Segments by the ICM Algorithm. The algorithm was initialized by the marginal mixture EM segmentation result.

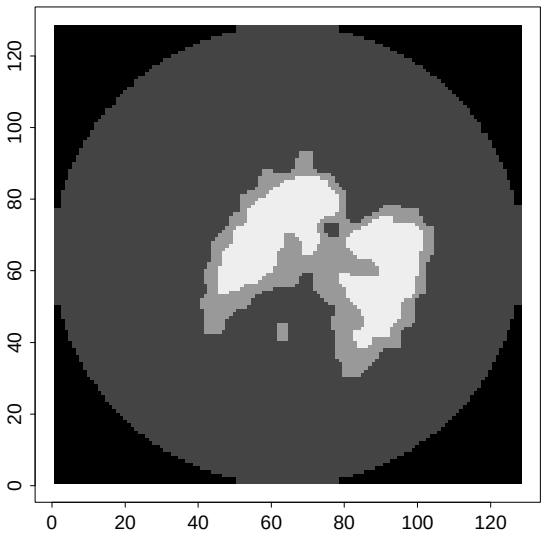

Figure 8: Final Segmentation of the Dog Lung Image into 4 Segments After Morphological Smoothing (opening and closing, conditional on the edge pixels).

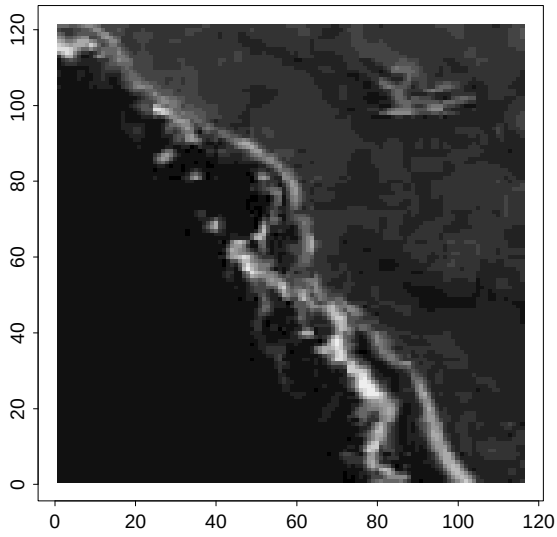

Figure 9: Satellite Image of Washington Coast

requiring only human inspection of the results.

4.3 Satellite Image Segmentation: Washington Coast

Figure 9 is a satellite image of a section of the Pacific coast of Washington state's Olympic peninsula; this image, provided by the US Geological Service as part of the National Aerial Photography Program (NAPP), was obtained from the Terraserver web page, terraserver.microsoft.com. The resolution is 8 meters per pixel, and it is a 256-level greyscale image.

Table 3 shows that the maximum value of PLIC occurs at $K = 6$ segments, so we regard this as the optimal choice of K for this model. For comparison, MMIC values are also shown; the maximum occurs at 4 segments. The marginal histogram of this image is shown in Figure 10; the large spike is due to the small variance in the greyscale value of the water.

For the 7 segment case, the EM algorithm led to a solution with only 6 segments. This causes convergence problems and numerical instability; if one of the mixture proportions approaches zero, the meaning of the corresponding estimates of mean and variance becomes unclear. These problems make the results for the 7 segment case somewhat unreliable; we regard this simply as evidence that the 7-segment model does not fit well.

The 6-segment results of the marginal mixture EM segmentation are shown in Figure 11. The water is quite well classified as a single segment, while the land is mostly contained

Table 3: PLIC and MMIC Results for the Washington Coast Image. The 7 segment case, noted with †, suffered convergence problems because the EM algorithm led to a solution with only 6 segments; this is discussed in the text. The favored result from each method is shown in boldface.

Segments	$\hat{\phi}$	Logpseudolikelihood	PLIC	MMIC
1	–	-64521.78	-129072.20	-133475.0
2	1.23	-56721.70	-113500.70	-125273.0
3	0.99	-45975.06	-92036.07	-117607.8
4	0.99	-45514.45	-91143.49	**−117528.6**
5	0.99	-42993.06	-86129.36	-117564.9
6	1.00	-42763.60	**−85699.10**	-117537.5
7	0.94†	-71772.18†	-143744.89†	-123280.5†

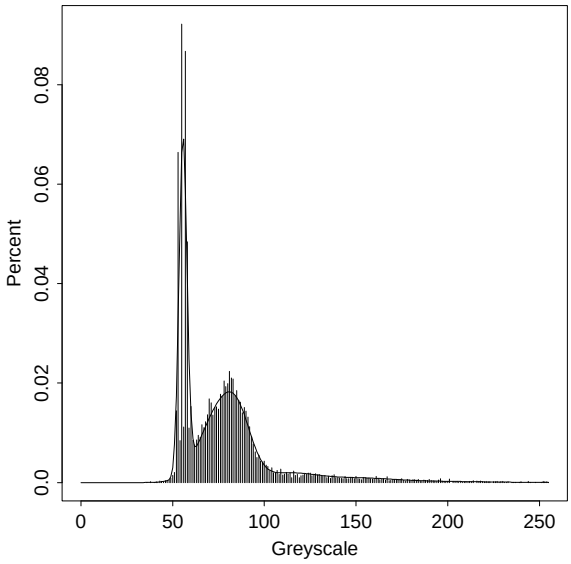

Figure 10: Marginal Histogram of the Washington Coast Image.

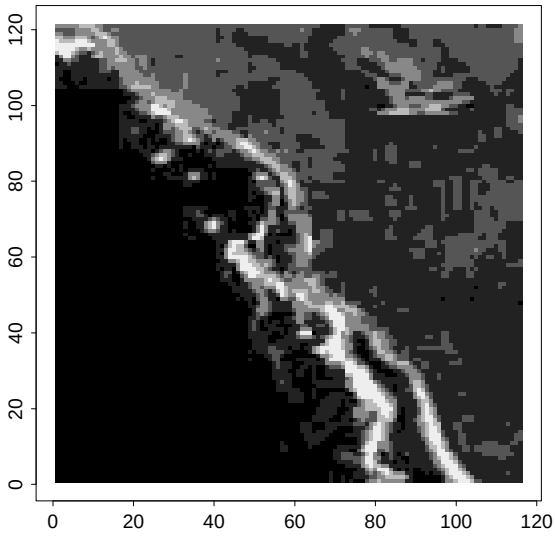

Figure 11: Marginal Mixture EM Segmentation of the Washington Coast Image Into Six Segments.

in two segments. The tideline accounts for most of the variability in the image. Figure 12 gives the ICM refinement of the segmentation; we can see that much of the noise in the land interior has been smoothed out, while the tideline is still quite clear.

In our final ICM segmentation (Figure 12), the water is well-characterized by the darkest segment. The second-darkest segment corresponds to shallow tidewater near the bright beach, as well as combining with the third-darkest segment to characterize most of the dry land. Although the second-darkest segment comprises both dry land and shallow tidewater, these two land types are spatially separated by the beach. The two brightest segments correspond to the beach, which is quite reflective. The third-brightest segment is transitional; it fills in between dry land and water in regions where the beach is not evident, as well as capturing the small hill in the upper right hand corner of the image.

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 by the user 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

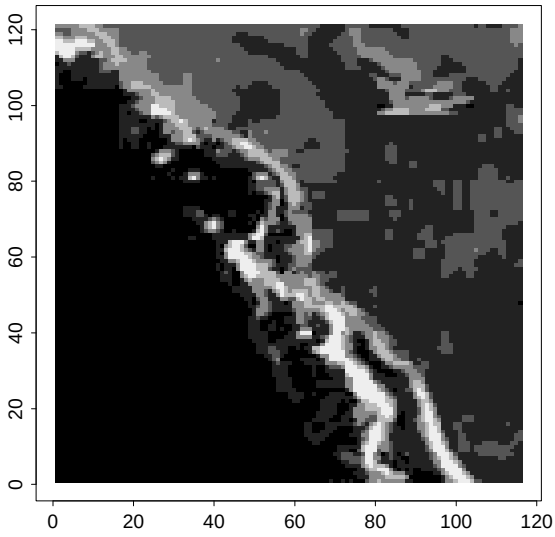

Figure 12: Segmentation of the Washington Coast Image into Six Segments Via ICM.

method works well in several examples, and seems to be potentially useful in the automatic segmentation of medical and satellite images, and perhaps also for color and gray-level image quantization, and the operational use of cooccurrence matrices for characterizing texture. Its use in these latter contexts remains to be explored, however.

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 Fraley and Raftery (1998). 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 Banfield and Raftery (1993), or an efficient method based on the minimal spanning tree could be used Posse (2000).

Our approach has been to compute approximate Bayes factors using the ICM algorithm and the pseudolikelihood of Qian and Titterington (1991a). Another approach would be to carry out a fully Bayesian analysis of the image model using Markov chain Monte Carlo, as suggested by Besag et al. (1991), and then to estimate the integrated likelihood using one of the available methods for doing this from MCMC (e.g. Newton and Raftery (1994); Chib (1995); Raftery (1996); DiCiccio et al. (1997); Oh (1999)). 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 Descombes et al. (1995) and Descombes et al. (1996) or the models of Tjelmeland and Besag (1998). 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 simultaneously. This could be done by identifying each combination of texture model and number of colors with the corresponding probability model, and then comparing all the resulting probability models using approximate Bayes factors, calculated using methods based on those described here.

References

Banfield, J. D. and Raftery, A. E. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, 49, 803–821.

Besag, J. (1986). Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society, Series B*, 48, 259–302.

Besag, J., York, J., and Mollie, A. (1991). Bayesian Image Restoration, with Two Applications in Spatial Statistics. *Annals of the Institute of Mathematical Statistics*, 43, 1–59.

Bokor, N. and Papp, Z. (1996). Computational Method for Testing Computer-Generated Holograms. *Optical Engineering*, 35, 2810–2815.

Burr, G. W., Barking, G., Coufal, H., Hoffnagle, J. A., Jefferson, C. M., and Neifeld, M. A. (1998). Gray-Scale Data Pages for Digital Holographic Data Storage. *Optics Letters*, 23, 1218–1220.

Byng, J. W., Critten, J. P., and Yaffe, M. J. (1997). Thickness-equalization Processing for Mammographic Images. *Radiology*, 203, 564–568.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association*, pages 1313–1321.

Cox, D. R. and Hinkley, D. (1978). *Problems and Solutions in Theoretical Statistics*. Chapman and Hall, London.

Dasgupta, A. and Raftery, A. E. (1998). Detecting Features in Spatial Point Processes with

Clutter via Model-Based Clustering. *Journal of the American Statistical Association*, 93, 294–302.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39, 1–38.

Descombes, X., Mangin, J. F., Pechersky, E., and Sigelle, M. (1995). Fine Structures Preserving Markov Model for Image Processing. *Proceedings of the 9th Scandinavian Conference on Image Analysis*, pages 349–356.

Descombes, X., Morris, R., and Zerubia, J. (1996). Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood. Technical Report 3015, INRIA, France.

DiCiccio, T. J., Kass, R. E., Raftery, A. E., and Wasserman, L. (1997). Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association*, 92, 903–915.

Dixit, S. (1991). Quantization of Color Images for Display Printing on Limited Color Output Devices. *Computers and Graphics*, 15, 561–567.

Fraley, C. and Raftery, A. E. (1998). How many clusters? Which clustering method? - Answers via Model-Based Cluster Analysis. *Computer Journal*, 41, 578–588.

Godtliebsen, F. and Chu, C. K. (1995). Estimation of the Number of True Grey Levels, Their Values, and Relative Frequencies in a Noisy Image. *Journal of the American Statistical Association*, 90, 890–899.

Hance, G. A., Umbaugh, S. E., Moss, R. H., and Stoecker, W. V. (1996). Unsupervised Color Image Segmentation. *IEEE Engineering in Medicine and Biology Magazine*, 15, 104–111.

Haralick, R. M. (1979). Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 67, 786–804.

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3, 610–621.

Hoeting, J. A., Madigan, D., Raftery, A., and Volinsky, C. (2000). Bayesian Model Averaging: A Tutorial (with discussion). *Statistical Science*, 14, 382–401.

Ji, C. and Seymour, L. (1996). A Consistent Model Selection Procedure for Markov Random Fields Based on Penalized Pseudolikelihood. *Annals of Applied Probability*, 6, 423–443.

Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Asso-*

ciation, 90, 773–795.

Kass, R. and Wasserman, L. (1995). A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion. *Journal of the American Statistical Association*, 90, 928–934.

Keribin, C. (1998). Consistent estimate of the order of mixture models. *Comptes Rendus de l'Academie des Sciences*, 326, 243–248.

Matheron, G. (1975). *Random Sets and Integral Geometry*. New York: Wiley.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B*, 56, 3–48.

Oh, M. S. (1999). Estimation of Posterior Density Functions from a Posterior Sample. *Computational Statistics and Data Analysis*, 29, 411–427.

Papamarkos, N. (1999). Color Reduction Using Local Features and a Kohonen Self-Organized Feature Map Neural Network. *International Journal of Imaging Systems and Technology*, 10, 404–409.

Posse, C. (2000). Hierarchical Model-Based Clustering for Large Datasets. *Journal of Computational and Graphical Statistics*, to appear.

Qian, W. and Titterington, D. M. (1991a). Estimation of Parameters in Hidden Markov Models. *Philosophical Transactions of the Royal Society of London*, 337, 407–428.

Qian, W. and Titterington, D. M. (1991b). Stochastic Relaxations and EM Algorithms for Markov Random Fields. *Journal of Statistical Computing and Simulations*, 40, 55–69.

Raftery, A. (1995). Bayesian Model Selection in Social Research. In Marsden, P., editor, *Sociological Methodology 1995*, pages 111–163. Blackwells, Cambridge.

Raftery, A. E. (1996). Hypothesis Testing and Model Selection. In Gilks, W., Spiegelhalter, D., and Richardson, S., editors, *Markov Chain Monte Carlo in Practice*, pages 163–188. Chapman and Hall, London.

Raftery, A. E. (1999). Bayes factors and BIC - Comment on 'A critique of the Bayesian Information Criterion for model selection'. *Sociological Methods and Research*, 27, 411–427.

Roeder, K. and Wasserman, L. (1997). Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association*, 92, 894–902.

Scheunders, P. (1997a). A Comparison of Clustering Algorithms Applied to Color Image

Quantization. *Pattern Recognition Letters*, 18, 1379–1384.

Scheunders, P. (1997b). A Genetic C-Means Clustering Algorithm Applied to Color Image Quantization. *Pattern Recognition*, 30, 859–866.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6, 461–464.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. New York: Academic Press.

Stanford, D. C. (1999). *Fast Automatic Unsupervised Image Segmentation and Curve Detection in Spatial Point Processes*. PhD thesis, University of Washington. Available at http://www.isomorphic.org/papers/.

Stanford, D. C. and Raftery, A. E. (2000). Finding Curvilinear Features in Spatial Point Processes: Principal Curve Clustering with Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 601–609.

Tjelmeland, H. and Besag, J. (1998). Markov Random Fields with Higher Order Interactions. *Scandinavian Journal of Statistics*, 25, 415–434.

Umbaugh, S. E., Moss, R. H., and Stoecker, W. V. (1992). An Automatic Color Segmentation Algorithm with Application to Identification of Skin Tumor Borders. *Computerized Medical Imaging and Graphics*, 16, 227–235.

Umbaugh, S. E., Moss, R. H., Stoecker, W. V., and Hance, G. A. (1993). Automatic Color Segmentation Algorithms with Application to Skin Tumor Feature Identification. *IEEE Engineering in Medicine and Biology Magazine*, 12, 75–82.

Valckx, F. M. J. and Thijssen, J. M. (1997). Characterization of Echographic Image Texture by Cooccurrence Matrix Parameters. *Ultrasound in Medicine and Biology*, 23, 559–571.

Xiang, Z. G. (1997). Color Image Quantization by Minimizing the Maximum Intercluster Distance. *ACM Transactions on Graphics*, 16, 260–276.

Xiang, Z. G. and Joy, G. (1994). Color Image Quantization by Agglomerative Clustering. *IEEE Computer Graphic and Applications*, 14, 44–48.