Fully Bayesian Image Segmentation - an Engineering Perspective

Robin Morris, Xavier Descombes, Josiane Zerubia

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Fully Bayesian Image Segmentation – an Engineering Perspective

Robin Morris, Xavier Descombes and Josiane Zerubia

N˚ 3017

October 1996

THÈME 3

ISSN 0249-6399

Fully Bayesian Image Segmentation – an Engineering Perspective

Robin Morris, Xavier Descombes and Josiane Zerubia

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet PASTIS

Rapport de recherche n ° 3017 — October 1996 — 18 pages

Abstract: Developments in Markov Chain Monte Carlo procedures have made it possible to perform fully Bayesian image segmentation. By this we mean that all the parameters are treated identically, be they the segmentation labels, the class parameters or the Markov Random Field prior parameters. We perform the analysis by sampling from the posterior distribution of all the parameters. Sampling from the MRF parameters has traditionally been considered if not intractable then at least computationally prohibitive. In the statistics literature there are descriptions of experiments showing that the MRF parameters may be sampled by approximating the partition function. These experiments are all, however, on 'toy' problems – for the typical size of image encountered in engineering applications phase transition behaviour of the models becomes a major limiting factor in the estimation of the partition function. Nevertheless, we show that with some care, fully Bayesian segmentation can be performed on realistic sized images. We also compare the fully Bayesian approach with the approximate pseudolikelihood method.

Key-words: Image segmentation, Markov chain Monte Carlo, parameter estimation, Markov Random Field, partition function

(Résumé : tsvp)

* R.D. Morris was supported by a grant from the Commission of the European Communities under the HCM program.

** email: *name*@sophia.inria.fr

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Segmentation d'image Bayesienne considéré sous l'aspect des sciences pour l'ingénieur

Résumé : Grâce au développement récent des méthodes MCMC, il est maintenant possible de réaliser des segmentations d'image qui soient complètement bayesiennes. Ce qui signifie que tous les paramètres sont traités de façon identique, qu'il s'agisse des étiquettes de segmentation, des paramètres de classe ou encore des paramètres relatifs à la probabilité a priori du champ de Markov. L'analyse est effectué par échantillonnage à partir de la distribution a posteriori de tous les paramètres. Echantillonner à partir des paramètres du champ de Markov est traditionnellement considéré comme prohibitif d'un point de vue informatique. En statistique, des papiers décrivent des expériences où les paramètres du champ de Markov sont échantillonnés par approximation de la fonction de partition. Ces expériences sont néanmoins réalisées sur des problèmes très simples. Pour la taille des images utilisées dans des applications réelles, le problème de la transition de phase est un facteur majeur de limitation eu égard à l'estimation de la fonction de partition. Néanmoins, nous montrons qu'une segmentation complètement bayesienne est possible sur des images de tailles assez grandes, pourvu que l'on prenne quelques précautions. Nous comparons aussi l'approche complètement bayesienne avec une approche qui utilise la pseudo-vraisemblance, qui n'est qu'une approximation.

Mots-clé : segmentation d'image, champ de Markov, estimation de paramètre, MCMC, fonction de partition

Contents

1 Introduction

Image segmentation (and here we are primarily concerned with the segmentation of SPOT images of agricultural regions) is a classic case of an ill-posed problem [13] – the data themselves are not sufficient to unambiguously define a solution. Using only the data and no smoothness assumptions leads to a very noisy segmentation, rather than the well defined, homogeneous regions that are sought [11, 12].

That prior information as to the desired solution is available (namely that the segmentation should consist of homogeneous regions) naturally suggests using a Bayesian approach, where this information may be easily and naturally incorporated into the a-priori distribution for the segmentation. The use of a Gibbs distribution (or equivalently, a Markov Random Field (MRF)) [4] for this a-priori distribution is something that has now been common for some time.

The main problem with the use of MRFs as regularisers has been in the estimation of the model's parameters. This is due to the analytic intractability of the partition function, and has lead to a number of approximate methods being proposed [1, 2] for this purpose. Recently methods have been proposed which estimate the partition function with sufficient accuracy to enable likelihood inference for the model's parameters to be performed [6, 7], and also enable the posterior distribution for the model's parameters to be sampled in a Markov chain Monte Carlo scheme, resulting in a 'fully Bayesian' analysis [9].

In this paper we discuss some of these statistical ideas from an engineering standpoint. We show that aspects of the models considered make this fully Bayesian approach computationally expensive for the sizes of images encountered in real-world problems. We compare the results of substituting the pseudolikelihood when sampling the MRF parameters, and conclude with some comments about the models.

2 The Models

We consider the well-known Potts model as the regulariser. The Potts model is defined by

$$\begin{aligned} p(\mathbf{x}|\beta) &= \frac{1}{Z(\beta)} \exp\left(\beta \sum_{i\sim j} \delta(x_i - x_j)\right) \\ &= \frac{1}{Z(\beta)} \exp\left(\beta \times \mathrm{NHC}(\mathbf{x})\right) \end{aligned} \tag{1}$$

where $Z(\beta)$ is the partition function, $\sum_{i\sim j}$ indicates a sum over all neighbouring pairs of pixels and $\delta(\cdot)$ is the delta function. This reduces to the number of homogeneous cliques in the realisation, NHC($\mathbf{x}$).

This model has been extensively studied in the statistical physics literature, and a number of analytical results obtained. Of interest to us here is that (in the infinite size limit) the model exhibits a phase transition at

$$\beta_c = \frac{1}{2}\log(1+\sqrt{n})$$

where n is the number of classes [5]. In the finite size case, the phase transition manifests as an abrupt change of behaviour over a small range of values of β close to β_c. As the size of the lattice and especially the number of classes considered increases, the abruptness of this change of behaviour becomes more pronounced. The Markov chain methods used to simulate realisations of the model also become much more slow to converge close to the phase transition. This is the main cause of the difficulty in applying the fully Bayesian approach to real engineering sized problems – to demonstrate that the method works in principle, an image of size 32×32 pixels with 2 classes is sufficient; in a real application we must deal with images of at least 256×256 (and often much larger) and many more classes (in this paper we have 5 classes).

2.1 Sampling the Potts Model

We use the Swendsen-Wang (S-W) algorithm [15] to simulate realisations of the Potts model. This algorithm mixes (ie explores the state space) much more freely than algorithms which update only single sites at a time (such as the usual implementations of the Gibbs Sampler or the Metropolis-Hastings algorithm). The Swendsen-Wang algorithm is outlined in appendix A.

Figure 1 shows the evolution of the number of homogeneous cliques in the realisations of the S-W algorithm for a Potts model of size 32×32 with 2 classes, after the first 200 iterations were disgarded. Clearly the algorithm converges quickly to the equilibrium distribution, and then explores freely the region of the state space with appreciable probability mass. Contrast figure 2, which shows the evolution for the Potts model of size 256×256 with 5 classes (note the horizontal scale). There is a narrow band of values where the exploration of the state space is sluggish. Clearly this is a potential cause of extremely inaccurate estimates – for values of β close to this phase transition many iterations must be disgarded to ensure that the algorithm has converged to the equilibrium distribution, and then an extremely large number

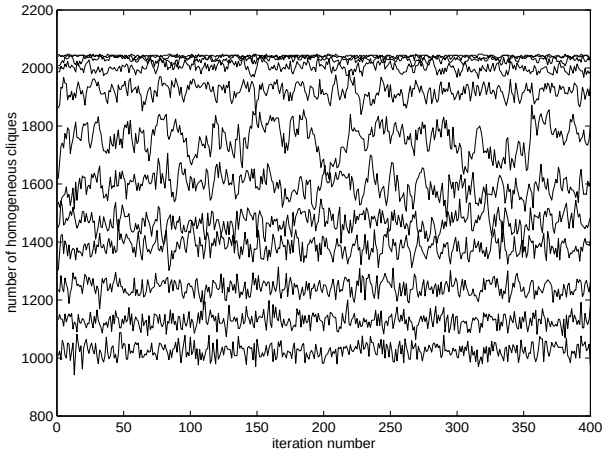

Figure 1: The evolution of the number of homogeneous cliques in the realisations of the Potts model by the Swendsen-Wang algorithm, for an image of size 32×32 with 2 classes, for various values of β

of realisations must be used when forming any Monte Carlo estimates, to ensure that the state space is adequately sampled. Figure 1 indicates that these considerations do not hold with the same force for the size of images used in the 'toy' problems discussed in the literature previously.

Consider the likelihood for the parameter β given a segmentation $\hat{\mathbf{x}}$.

$$p(\beta|\hat{\mathbf{x}}) = \frac{1}{Z(\beta)} \exp(\beta \times \mathrm{NHC}(\hat{\mathbf{x}}))$$

where

$$Z(\beta) = \sum_{\{x\}} \exp(\beta \times \mathrm{NHC}(\mathbf{x}))$$

The maximum likelihood estimate is given by

$$\frac{\partial \log p(\beta|\hat{\mathbf{x}})}{\partial \beta} = \mathrm{NHC}(\hat{\mathbf{x}}) - \frac{Z'(\beta)}{Z(\beta)} = 0$$

and $Z'(\beta)/Z(\beta) =< \mathrm{NHC}(\mathbf{x}) >_\beta$, that is, the mean number of homogeneous cliques under the distribution defined by the value of β. The maximum likelihood estimate is thus the value of β for which

$$\mathrm{NHC}(\hat{\mathbf{x}}) =< \mathrm{NHC}(\mathbf{x}) >_\beta$$

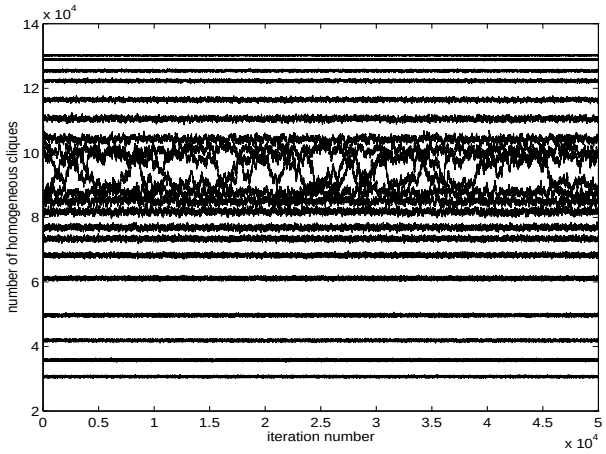

Figure 2: The evolution of the number of homogeneous cliques in the realisations of the Potts model by the Swendsen-Wang algorithm, for an image of size 256×256 with 5 classes, for various values of β. Note the different horizontal scale from figure 1.

The graphs in figures 3 and 4 show $< \mathrm{NHC}(\mathbf{x}) >_\beta$ from Monte Carlo simulations. Clearly for the 'toy' problem the maximum likelihood estimate of β can easily be found, and the variance of the realised number of homogeneous cliques is sufficiently small that the estimates are reliable. For the simulations of an image of size 256×256 with 5 classes, however, it is clear that very many images will fall into the zone near the phase transition (as this includes a very large range of values of $\mathrm{NHC}(\hat{\mathbf{x}})$).

2.2 The Class Conditional Probabilities

These distributions describe what we expect to observe at a pixel, knowing that it is in a particular class. A common assumption in the segmentation of satellite imagery is that the grey scale values associated with a particular class are gaussian distributed, with a mean and variance depending on the class, and that they are conditionally independent given the segmentation. This gives

$$p(y_i|\mathbf{x}, \{\mu, \sigma\}) = \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left(-\frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2}\right) \tag{2}$$

where μ_{x_i} and σ_{x_i} are the mean and variance associated with the class indicated by the label x at location i and $\mathbf{y}$ is the observed image.

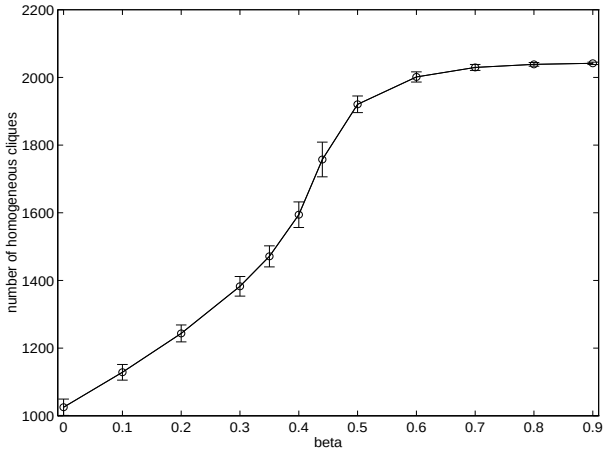

Figure 3: The mean and standard deviation of the number of homogeneous cliques in the realisations of the Potts model by the Swendsen-Wang algorithm, for an image of size 32×32 with 2 classes, as a function of β

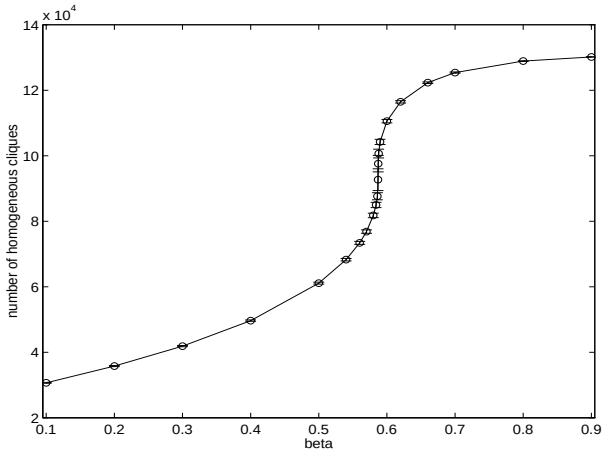

Figure 4: The mean and standard deviation of the number of homogeneous cliques in the realisations of the Potts model by the Swendsen-Wang algorithm, for an image of size 256×256 with 5 classes, as a function of β

2.3 The Posterior Distribution

It is by analysis of the posterior distribution that we will draw inferences about the segmentation, the parameters of the class conditional distributions and the a-priori segmentation model's hyperparameters. Thus we are interested in sampling from $p(\mathbf{x}, \{\mu, \sigma\}, \beta|\mathbf{y})$. Using Bayes Theorem and natural independence assumptions we write this as

$$p(\mathbf{x}, \{\mu, \sigma\}, \beta|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}, \{\mu, \sigma\})p(\mathbf{x}|\beta)p(\{\mu, \sigma\})p(\beta)}{p(\mathbf{y})}$$

The first term is given by the class conditional probabilities (equation 2) and the second is the Potts model prior on the segmentation (equation 1). The third term is the priors on the mean and variance of each class. We use a uniform distribution on the means, and a Jeffrey's prior [14] on the variances. For the prior on β we also use a uniform distribution. It could be argued that we should choose a prior which favours larger β for increased smoothness, but for simplicity this was not done. The term $p(\mathbf{y})$ is constant for a given image.

3 Sampling the Parameters

In section 2 we have defined the posterior distribution for all the parameters of interest, that is, the joint posterior of the segmentation labels, the class means and variances, and the hyperparameters of the MRF models. This is the distribution we wish to sample, and from these samples we will make our inferences about the parameters.

To sample from the posterior we use either the Metropolis-Hastings (M-H) algorithm or the Gibbs sampler, depending on convenience – if simple expressions for the full conditional distributions needed in the Gibbs sampler are available and random variates from these distributions can be easily and efficiently generated, then the Gibbs sampler is used. If not, then the more general Metropolis-Hastings algorithm is the algorithm of choice, requiring only the calculation of the ratio between the probabilities of two different states. The Swendsen-Wang algorithm is also used for sampling from the prior on the segmentation, these samples being used to estimate the acceptance probabilities when sampling the model's hyperparameter.

3.1 Sampling the segmentation labels

The task of sampling the segmentation labels is the standard task in any algorithm using MRFs, and is performed either by using the M-H algorithm, proposing a change

to the label at the current pixel, and then using the M-H acceptance criterion to determine the next state at that pixel, or using the Gibbs sampler, by computing explicitly the conditional distribution over the labels at that pixel, and then sampling from that discrete distribution.

The local conditional distribution involved is

$$p(x_i = s|\ldots) \propto \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(y_i - \mu_s)^2}{2\sigma_s^2} + \beta \sum_{j \in \mathcal{N}_i} \delta(x_j - s) \right) \tag{3}$$

where s is the class label and $\mathcal{N}_i$ indicates the neighbours of i.

3.2 Sampling the class mean and variance parameters

From the posterior distributions the following marginal distributions for the mean and variance parameters for each class are easily derived. For the means we have

$$\begin{aligned} p(\mu_s|\ldots) &\propto \exp\left(-\sum_{i:x_i=s} \frac{(y_i - \mu_s)^2}{2\sigma_s^2} \right) \\ &= \mathrm{N}\left(\frac{1}{N_s} \sum_{i:x_i=s} y_i, \frac{\sigma_s^2}{N_s} \right) \end{aligned}$$

where N_s is the number of pixels in class s and $\mathrm{N}(\cdot,\cdot)$ denotes the Normal distribution. For the variances the conditional distribution is

$$\begin{aligned} p(\sigma_i|\ldots) &\propto \frac{1}{\sigma_s} \prod_{i:x_i=s} \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(y_i - \mu_s)^2}{2\sigma_s^2} \right) \\ &\propto \sigma_s^{-(N_s+1)} \exp\left(-\sum_{i:x_i=s} \frac{(y_i - \mu_s)^2}{2\sigma_s^2} \right) \end{aligned}$$

The means are Gaussian distributed, and hence the new values can be generated easily. The distributions for the variances are inverse-chi distributions, with N_s degrees of freedom, and may be generated by a transformation from variables with a Gamma distribution with $N_s/2$ degrees of freedom [14]. The transformation being the multiplication of the Gamma variates by the factor $\sqrt{(\sum_{i:x_i=s}(y_i - \mu_s)^2)/2}$.

Whilst theoretically the initialisation of the means and variances is not important, practically, for convergence in a reasonable amount of computation (or at all if the probability of some configurations is rounded to zero by the effects of finite precision

computer arithmetic), it is necessary to initalise the means and variances close to their true values. Treating the data just as a Gaussian mixture, and analysing this mixture to determine the parameters of the components is a difficult problem, and the domain where closely spaced components can be differentiated is restricted [16]. In [3] a method was proposed which uses the contextual information available in an image to broaden the domain of applicability, and it is this method which we use to obtain initial values for the means and variances.

3.3 Sampling the model's hyperparameter

Sampling the hyperparameter exactly is intractable. However, it is possible to estimate the partition function to a sufficient degree of accuracy to make it possible to sample values for the hyperparameter. As explained above, this is not trivial for the image sizes we are interested in. Indeed in [8] the authors instead sample from the pseudolikelihood, claiming that they could see no way to make sampling from the posterior for the parameters computationally reasonable. Consider the marginal distributions for the hyperparameters.

$$p(\beta|\ldots) \propto \frac{1}{Z(\beta)} \exp(\beta \times \mathrm{NHC}(\mathbf{x})) \tag{4}$$

Clearly it is the partition function $Z(\beta)$ which causes the computation difficulty. We now show how it is possible to estimate $Z(\beta)$ with sufficient accuracy, and sufficiently efficiently to enable the posterior distribution of the parameters to be sampled from. We have that

$$\begin{aligned} Z(\beta) &= \sum_{\{x\}} \exp(\beta \times \mathrm{NHC}(\mathbf{x})) \\ &= \sum_{\{x\}} \exp((\beta - \phi) \times \mathrm{NHC}(\mathbf{x})) Z(\phi) p(\phi) \\ &= Z(\phi) < \exp((\beta - \phi) \times \mathrm{NHC}(\mathbf{x})) >_{\phi} \\ \frac{Z(\beta)}{Z(\phi)} &\simeq \frac{1}{N} \sum_{j=1\ldots N} \exp((\beta - \phi) \times \mathrm{NHC}(\mathbf{x}^j)) \end{aligned}$$

ie, that it is possible to use samples $\mathbf{x}^j$ drawn from the distribution $p(\mathbf{x}|\phi)$ to estimate the ratio of the partition functions for two different values of the parameter. If we use a Metropolis-Hastings algorithm to sample β, using a proposal distribution that is uniformly distributed in a small range either side of the current value, then the

acceptance probability for the new value, β', is

$$A(\beta', \beta) = \min\left(1, \frac{Z(\beta)}{Z(\beta')} \exp((\beta' - \beta)\mathrm{NHC}(\mathbf{x}))\right)$$

The ratio $Z(\beta')/Z(\beta)$ can be estimated in a similar manner using samples drawn from the distribution defined by the parameter ϕ. What is important is that the distributions overlap sufficiently that the samples from $p(\mathbf{x}|\phi)$ have appreciable probability under the other two distributions. From figures 1 and 2 we can see that if we make the proposed change small enough, that this will always be possible. Indeed, there is a range of values of β where large changes can be made, where the variance of the number of homogeneous cliques is large.

In [9, 10] a slightly different approach was taken. Because of the small sizes of the images considered, it was possible to define a small set of value of β which, when sampled, would adequately cover the entire range of β of interest. These samples were then used to construct a mixture distribution which had appreciable probability everywhere [6]. This distribution was then used when sampling β. From figure 2 it is clear that for an image of appreciable size the number of values of β which would have to be considered becomes prohibitively large. Also, when using the mixture distribution, the vast majority of the samples will have a probability very close to zero under the actual value of β considered. It is for these reasons that for the sizes of images we are concerned with here that we have advocated the simpler scheme outlined above. This scheme can also be implemented adaptively – if we do not have samples from a distribution adequately close to the values of β under consideration, we can generate these samples on-line (and then store them for re-use).

3.4 Sampling the Pseudolikelihood

Because of the difficulties described above when sampling from the distribution for the hyperparameter using Monte Carlo estimates of the partition function, the pseudolikelihood was introduced as a simple approximation [1]. The pseudolikelihood is constructed as the product of the conditional distributions at each pixel, that is

$$p_{pl}(\beta|\ldots) = \prod_i p(x_i|x_{\mathcal{N}_i}, \beta)$$

which is simple to compute.

In the experiments, we used the M-H algorithm, with a proposal uniformly distributed in a small range either side of the current value. This leads to the acceptance

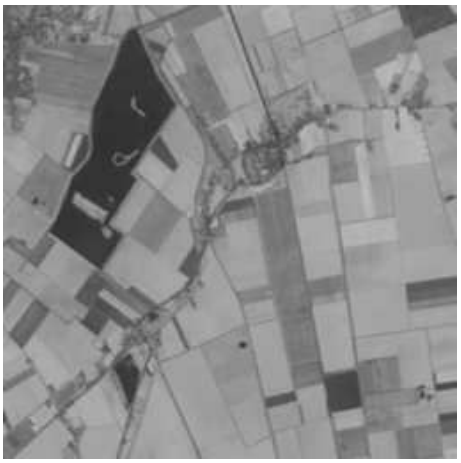

Figure 5: Section from a SPOT image

probability for the new value being simply

$$\min(1, p_{pl}(\beta_{\text{new}})/p_{pl}(\beta_{\text{old}}))$$

4 Results and Conclusions

Figure 5 shows a section from a SPOT image which we wish to segment into five classes. Figure 6 shows the evolution of β when sampled from the pseudolikelihood, and figure 7 shows the evolution of β when sampled using estimates of the ratios of the partition functions. The evolution of the means and variances of the class conditional distributions is not shown. The variation in β is much smaller for the 'fully Bayesian' case, and also the pseudolikelihood significantly over estimates the parameter's value – from figure 4 we can see that a value of $\beta = 1.1$ corresponds to a prior which is almost completely uniform. This does not correspond well with the image. The values of β sampled from the true likelihood are centered around $\beta = 0.603$. This is a little above the critical value, and from figure 4 indicates a more reasonable modellisation, in that the number of homogeneous cliques in the segmentation is very close to the number of homogeneous cliques implied by the prior.

The segmentation results in figures 8 and 9 are very similar, however (these are MPM estimates from 800 iterations after disgarding 250 iterations). This is because of the high quality of the data – the image in figure 5 is relatively noise free and the

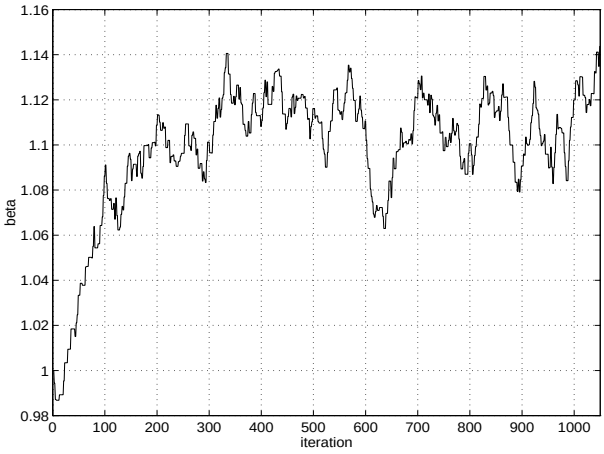

Figure 6: The evolution of β when sampled from the pseudolikelihood.

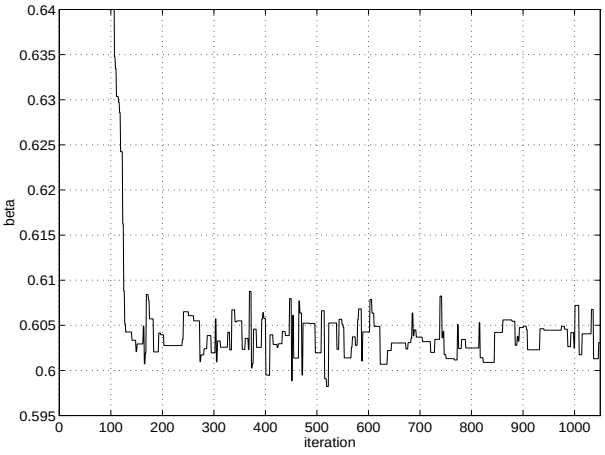

Figure 7: The evolution of β when sampled using estimates of $Z(\beta)$.

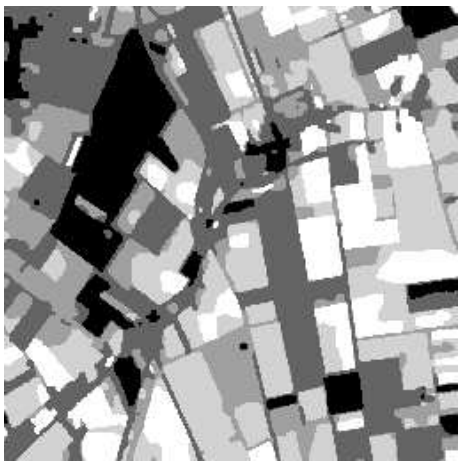

Figure 8: The segmentation result using the pseudolikelihood

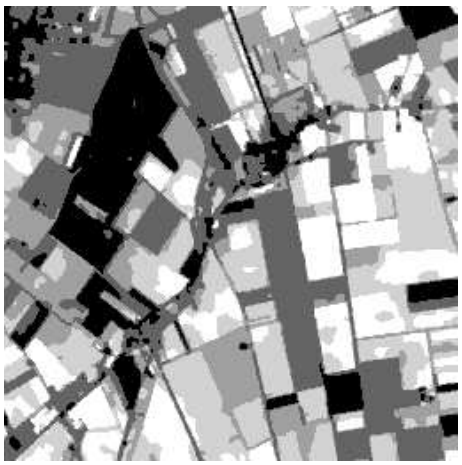

Figure 9: The segmentation result using the 'fully Bayesian' approach

classes are for the most part easily distinguishable. The differences are mainly at the edges of the regions. Because the pseudolikelihood over smoothes the segmentation, it results in homogeneous zones. The 'fully Bayesian' segmentation leaves many isolated pixels at the edges of the zones. This is because, with a value of β around 0.6, the Potts model does not penalise isolated pixels sufficiently. This indicates that the Potts model is not an adequate prior, when our prior information is that we have uniform zones and do not wish to have isolated pixels remaining in the segmentation.

A The Swendsen-Wang Algorithm

For simulating the Potts model the Swendsen-Wang (SW) algorithm was developed, which converges rapidly to the equilibrium distribution and then moves relatively freely within this distribution. It is an example of an auxiliary variable algorithm. The idea behind auxiliary variable methods is the following:

It is desired to simulate a distribution $\pi(\mathbf{x})$. Auxiliary variables $\mathbf{u}$ are introduced, with conditional distribution $\pi(\mathbf{u}|\mathbf{x})$. This gives a joint distribution $\pi(\mathbf{x},\mathbf{u}) = \pi(\mathbf{u}|\mathbf{x})\pi(\mathbf{x})$, with the desired marginal distribution for $\mathbf{x}$ of $\pi(\mathbf{x})$. Simulation of this distribution is generally performed by alternately updating $\mathbf{u}$ and $\mathbf{x}$ – the idea being to define $\pi(\mathbf{u}|\mathbf{x})$ such that the updates cause rapid mixing. The realisations of $\mathbf{x}$ are those desired.

For the SW algorithm the distribution $\pi(\mathbf{u}|\mathbf{x})$ is defined such that the u_{ij} are independent, and is

$$p(u_{ij}|\mathbf{x}) \propto \exp\left(-\beta \mathrm{I}[x_i = x_j]\right) \mathrm{I}[0 \leq u_{ij} \leq \exp\left(\beta \mathrm{I}[x_i = x_j]\right)] \tag{5}$$

where u_{ij} can be considered as a continuous 'bond' variable between the pixels x_i and x_j and $\mathrm{I}[\cdot]$ is the indicator function. This results in

$$\pi(\mathbf{x}|\mathbf{u}) \propto \prod_{i \sim j} \mathrm{I}[0 \leq u_{ij} \leq \exp\left(\beta \mathrm{I}[x_i = x_j]\right)] \tag{6}$$

What does this choice of distribution give us? Considering first $p(u_{ij}|\mathbf{x})$, for $u_{ij} > 1$ we must have $\exp\left(\beta \mathrm{I}[x_i = x_j]\right) > 1$, or equivalently, $x_i = x_j$. Thus $u_{ij} > 1$ constrains x_i and x_j to be in the same state. Conversely, if x_i and x_j are in the same state, what is the probability of $u_{ij} > 1$? From the conditional distribution in equation 5 we have that

$$p(u_{ij} > 1 | x_i = x_j) = 1 - \exp(-\beta) \tag{7}$$

Since it is only important whether u_{ij} is greater or less than one we may think of the u_{ij} as binary bond variables. From equation 7 the bond variable is present

between two pixels in the same state with probability $1-\exp(-\beta)$. To sample $\mathbf{u}$ thus involves placing bonds between neighbouring pixels of the same state with probability $1-\exp(-\beta)$ and omitting bonds between neighbouring pixels of differing states.

Once the bonds are in place, the conditional distribution $\pi(\mathbf{x}|\mathbf{u})$ says that all configurations where bonded pixels are of the same state are equally probable. Thus to update $\mathbf{x}$ we form clusters of connected pixels and assign to all pixels of the cluster the same state, chosen uniformly from the allowed states. This scheme allows potentially large clusters of pixels to change state at each iteration, allowing the Markov chain to explore the distribution freely.

For small β the links will only rarely be placed, resulting in a large number of small clusters. In this case standard single-site updating algorithms are more efficient. For large β (and especially β close to the critical value (see section 2)) the improved mobility of the SW algorithm more than compensates for the extra computation involved.

References

[1] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statistical Soc. B*, 48(3):259–302, 1986.

[2] H. Derin and H. Elliott. Modelling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans PAMI*, 9:39–55, 1987.

[3] X. Descombes. Application of stochastic techniques in image processing for automatic tissue classification in MRI and blood vessel restoration in MRA. Technical report, Technical report, Laboratory for Medical Imaging Research (ESAT-Radiology), 1996.

[4] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans PAMI*, 6(6):721–741, November 1984.

[5] H-O Georgii. *Gibbs Measures and Phase Transitions*. de Gruyter, Berlin, 1988.

[6] C.J. Geyer. Reweighting Monte Carlo mixtures. Technical report, No. 568, School of Statistics, University of Minnesota, 1991.

[7] C.J. Geyer and E.A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *JRSS-B*, 54:657–699, 1992.

[8] J. Heikkinen and H. Högmander. Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, 43:569–582, 1994.

[9] D. Higdon. *Spatial Applications of Markov Chain Monte Carlo for Bayesian Inference*. PhD thesis, University of Washington, 1994.

[10] D.M. Higdon, V.E. Johnson, T.G. Turkington, J.E. Bowsher, D.R. Gilland, and R.J. Jaaszczak. Fully Bayesian estimation of Gibbs hyperparameters for emission coupled tomography data, 1995. Preprint. Institute of Statistics and Decision Sciences, Duke University.

[11] R. Hu and M.M. Fahmy. Texture segmentation based on a hierarchical markov random field model. *Signal Processing*, 26:285–305, 1992.

[12] Z. Kato, J. Zerubia, and M. Berthod. Bayesian image classification using Markov random fields. In G. Demoments, editor, *Maximum entropy and Bayesian Methods*, pages 375–382. Kluwer, 1993.

[13] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82(397):76–89, March 1987.

[14] J.J.K. O Ruanaidh and W.J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag, Berlin, 1996.

[15] R.H. Swendsen and J-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.

[16] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, 1985.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399