UNSUPERVISED SEGMENTATION OF TEXTURED SATELLITE AND AERIAL IMAGES WITH BAYESIAN METHODS

Simon P. Wilson
Dept. of Statistics, Trinity College,
Dublin 2, Ireland
Tel: +353 1 6081759; fax: +353 1 6615046
e-mail: simon.wilson@tcd.ie

Josiane Zerubia
INRIA, BP 93, 2004 Route des Lucioles,
06902 Sophia-Antipolis Cedex, France
Tel: +33 4 92 38 78 65; fax: +33 4 92 38 76 43
e-mail: Josiane.Zerubia@sophia.inria.fr

ABSTRACT

We investigate Bayesian solutions to unsupervised image segmentation based on the double Markov random field model. Inference on the number of classes in the image is done with reversible jump Metropolis moves. These moves are implemented by splitting and merging classes. Tests are conducted on satellite and aerial images.

1 INTRODUCTION

A double Markov random field is a hierarchical model for an image composed of several textures (or classes), where both the grey levels of each class and the class labels are described by Markov random fields. It has been extensively used in Bayesian approaches to image segmentation, see [8] for instance. Markov random fields have many applications to segmentation, see [3, 9] for example.

Consider a rectangular lattice of pixel sites $\mathcal{S}$. An image consists of an array of grey values $(x_s)_{s\in\mathcal{S}}$ and labels $(y_s)_{s\in\mathcal{S}}$, identifying the texture type present. Let there be R textures in the image and each texture, defined on all of $\mathcal{S}$, is a Markov random field T^r, parameterised by θ_r, with neighbourhood system having set of cliques $\mathcal{C}_r$. The label process is another Markov random field Y, parameterised by β and with neighbourhood system having set of cliques $\mathcal{C}_Y$. All the fields are independent conditional on model parameters and their distributions have the following Gibbs representation:

$$P(T^r = t \,|\, \theta_r) \;=\; \frac{\exp[-U_r(t;\theta_r)]}{Z_r(\theta_r)}, \qquad (1)$$

$$P(Y = y \,|\, R, \beta) \;=\; \frac{\exp[-U_Y(y;R,\beta)]}{Z_Y(R,\beta)}. \qquad (2)$$

The partition functions Z_r and Z_Y cannot usually be evaluated. In this paper, we will define the $P(T^r \,|\, \theta_r)$ to be Gaussian Markov random fields (GMRFs), so that for a GMRF having K clique types in its neighbourhood system, $\theta_r = (\mu_r, \sigma_r^2, \theta_{r1}, \ldots, \theta_{rK})$, and $P(Y = y \,|\, R, \beta)$ to be a Potts model, so $\beta > 0$.

The *double Markov random field* is a probability distribution over X and Y of the form

$$\begin{aligned} &P(X = x, Y = y \,|\, R, \theta_1, \ldots, \theta_R, \beta) \\ &\quad = \; P(Y = y \,|\, R, \beta) \prod_{r=1}^{R} P(T^r_{S_r} = x_{S_r} \,|\, \theta_r), \end{aligned} \qquad (3)$$

where $S_r = \{s \in \mathcal{S} \,|\, y_s = r\}$ and $T^r_{S_r}$, x_{S_r} denote T^r and x restricted to S_r, and the distributions of T^r and Y are given in Equations 1 and 2.

In what follows we assume that the classes are ordered by their mean value i.e. $\mu_r \le \mu_{r+1}$. The ordering removes the non-identifiability of a segmentation, in the sense that, without the ordering, the method cannot distinguish between segmentations where the class labels are permuted. The non-identifiability would complicate the method that we propose for determining R.

2 USING THE MODEL FOR SEGMENTATION

In a Bayesian approach to unsupervised segmentation, the goal is to calculate the distribution of Y, R and the model parameters if unknown, given X, that is:

$$\begin{aligned} &P(Y, R, \theta_1, \ldots, \theta_R, \beta \,|\, X) \\ &\quad \propto \; P(X, Y \,|\, R, \theta_1, \ldots, \theta_R, \beta)\, P(R, \theta_1, \ldots, \theta_R, \beta), \end{aligned} \qquad (4)$$

where $P(R, \theta_1, \ldots, \theta_R, \beta)$ is a prior distribution on the parameters. The parameters are usually assumed independent *a priori*. In the GMRF case, these are assumed to be: uniform prior distribution for μ_r over the range $[0, 255]$, inverse gamma prior for σ_r^2 with parameters a and b, uniform prior for $(\theta_{r1}, \ldots, \theta_{rK})$ over the allowable range of values (see [7] for this range in the case of a second order neighbourhood system) and geometric prior on R with parameter p. The only available technique to evaluate this posterior distribution is Monte Carlo Markov chain (MCMC) simulation, usually the Gibbs sampler, where one simulates from the full conditional distributions of each Y_s, $s \in \mathcal{S}$, and those of β and θ_r, $r = 1, \ldots, R$. To simulate from R, other methods are needed, as described below.

For the double Markov random field, the full conditional distribution for the pixel labels is not easy to evaluate. We use an approximation to it; in [8] it was shown that the pseudo-likelihood performed well,

$$P(X, Y \mid R, \beta, \theta_1, \ldots, \theta_R) \approx P(Y \mid R, \beta) \prod_{s \in \mathcal{S}} P(X_s \mid \theta_{Y_s}, X_t; t \in \eta_s), \quad (5)$$

where η_s is the neighbourhood of s. For each s, it is assumed that the texture Y_s holds in the entire neighbourhood of s. This approximation is developed in [2]. In the case of the Potts-Gaussian model, the full conditional of Y_s is:

$$P(y_s = r | x, y_j, j \neq s, \theta_r, \beta) \propto \frac{\exp\left[-\frac{\left\{x_s - \mu_r - \sum_{k=1}^{K} \sum_{j|<s,j>_k} \theta_{rk}(x_j - \mu_r)\right\}^2}{2\sigma_r^2}\right]}{\sqrt{2\pi\sigma_r^2}} \times \exp\left[\beta \sum_{j \in \eta_s} I(r = y_j)\right], \quad (6)$$

where $< s, j >_k$ means that pixels s and j form a clique of type k (horizontal, vertical neighbours, etc.), η_s is the neighbourhood of s in the Potts model and $I()$ is the indicator function.

The Gibbs sampler repeatedly samples from each full conditional in turn, forming a Markov chain whose stationary distribution is the posterior distribution. As a solution, we take the MAP or most likely posterior segmentation. This is found by running Gibbs sampling in tandem with simulated annealing to find the maximum.

3 THE REVERSIBLE JUMP APPROACH TO SAMPLING FROM R

Reversible jump has been used for identifying R in image segmentation by Markov random fields [1, 6]. Reversible jump was proposed in [4], which we refer to for a full description of the approach. The idea is to sample R by a Metropolis move. This requires an acceptance probability, and in this application the only moves that are both practical and easy to propose with a non-zero acceptance probability are to increase or decrease R by 1. Increasing R by 1 requires that we generate parameter values for a new class, and a new segmentation that involves the new class. The new parameter values must be generated according to a 1-1 transformation between the new set of values and the old parameters together with random numbers necessary to generate the parameters for the new class. So a move from $\Theta_R = (\theta_1, \ldots, \theta_R)$ to $\Theta^*_{R+1} = (\theta^*_1, \ldots, \theta^*_R, \theta^*_{R+1})$ is achieved with a set random numbers u of the same dimension as θ^*_{R+1} such that there is a 1-1 function $(\Theta_R, u) \leftrightarrow \Theta^*_{R+1}$. The usual conditions of reversibility and irreducibility must be maintained. Decreasing R by 1 requires that we eliminate one of the classes, and propose a segmentation where the deleted class is not present. The reversibility means that this move must be done using the same 1-1 function between Θ^*_{R+1} and (Θ_R, u). Once either move has been done, an acceptance probability can be computed. For the move to increase R to $R+1$, with a change in parameters from Θ_R to Θ^*_{R+1}, and a change in segmentation from Y to Y^*, this acceptance probability is the minimum of 1 and

$$\frac{P(X, Y^* \mid R+1, \Theta^*_{R+1}, \beta)}{P(X, Y \mid R, \Theta_R, \beta)} \frac{P(\Theta^*_{R+1})}{P(\Theta_R)} \frac{P(R+1)}{P(R)} \times \frac{1}{P(u)} \frac{P(\text{decrease } R+1)}{P(\text{increase } R)} \left|\frac{\partial(\Theta_R, u)}{\partial \Theta^*_{R+1}}\right|, \quad (7)$$

where the final partial derivative is the Jacobian of the transformation $(\Theta_R, u) \rightarrow \Theta^*_{R+1}$, and $P(\text{decrease } R+1)$ and $P(\text{increase } R)$ refer to any other probabilities involved in decreasing or increasing the number of classes (for example, randomly choosing which class to delete).

The above is a strategy for sampling for R. Sampling from the other parameter values and Y is from the full conditionals, as described in Section 2. Repeated sampling like this in tandem with simulated annealing produces our solution.

4 SPLIT AND MERGE MOVES

Increasing R by 1 is done by choosing an existing class and splitting it into two. Several split moves for increasing R have been considered. Finally we propose the following, which is a compromise between computational complexity and arriving at a proposed new segmentation that has a reasonable acceptance probability. To increase the number of classes from R to $R+1$, we propose to:

1. Randomly pick a non-empty class c to split.

2. Generate new parameters θ^*_{c1} and θ^*_{c2} from θ_c and random numbers u by a 1-1 transformation such that $\dim(\theta_c, u) = \dim(\theta^*_{c1}, \theta^*_{c2})$. For GMRF texture models, we use the following, designed to be simple perturbations of the current parameters:

$$\begin{aligned} \mu^*_{c1} &= \mu_c - u_1\sigma_c & \mu^*_{c2} &= \mu_c + u_1\sigma_c \\ \sigma^{*2}_{c1} &= \sigma^2_c(1 + u_2) & \sigma^{*2}_{c2} &= \sigma^2_c(1 - u_2) \\ \theta^*_{c1,k} &= \theta_{ck} + u_{2+k} & \theta^*_{c2,k} &= \theta_{ck} - u_{2+k}, \end{aligned} \quad (8)$$

   where the u_1 is uniform(0,1), u_2 is uniform(-1,1), u_{2+k} is uniform(-0.1,0.1) and $k = 1, \ldots, K$. The Jacobian of the transformation in Equation 8 is $4\sigma_c^3 2^K$.

3. Check to see that the new means are still ordered (that is, $\mu_{c-1} < \mu^*_{c1} < \mu^*_{c2} < \mu_{c+1}$) and lie in the prior range $[0, 255]$. If not, reject move immediately. If so, continue.

4. Assign all labels currently assigned to class c to classes $c1$ or $c2$. This is done as follows. Labels in the set $\mathcal{S}_c$ are assigned randomly to either $c1$ or $c2$. Then several Gibbs sampling sweeps are taken over $\mathcal{S}_c$ only, using classes $c1$ and $c2$ and the full conditionals of Equation 6. The Gibbs sampling hopefully produces homogeneous regions in $\mathcal{S}_c$ of each new class.

5. Compute the acceptance probability (see below) and accept or reject the move. If accept, relabel parameters and classes to include $c1$ and $c2$.

The choices for the u at step 2 are a balance between allowing the possibility of reasonably large changes in the properties of old and new classes, and restricting changes so that the chance of accepting the move is not always 0.

Reducing R is done by choosing two existing classes and merging them into one. The merge move is to choose a class $c1$, select the class $c2$ that has closest mean to it (from the ordering on means property, this is either $c1+1$ or $c1-1$) and propose new parameters for the merged class c that are the inverse of the transformation in Equation 8. We use the following transformation of parameters for the reversal of the split transformation:

$$\begin{aligned} \mu_c^* &= \frac{n_1}{n_1+n_2}\mu_{c1} + \frac{n_2}{n_1+n_2}\mu_{c2}; \\ \sigma_c^{*2} &= \frac{\sigma_{c1}^2+\sigma_{c2}^*}{2}; \quad \theta_{ck}^* = \frac{\theta_{c1,k}+\theta_{c2,k}}{2}, \end{aligned} \tag{9}$$

where n_1 and n_2 are the number of pixels assigned to $c1$ and $c2$ respectively.

The accept probability is rather complicated but it can be shown that, using the pseudolikelihood approximation and following Equation 7, in the case of using 2nd order GMRFs for textures (so $K=4$) and 1st order Potts model for Y, it is the minimum of 1 and

$$\begin{aligned} &\left\{ \prod_{s\in\mathcal{S}_c} \frac{\exp\left[-\left((\epsilon^*)^2/2\sigma_{y_s^*}^{*2}\right) + \beta\sum_{j\in\eta_s}\delta(y_s^*-y_j^*)\right]}{\exp\left[-\left(\epsilon^2/2\sigma_c^2\right) + \beta\sum_{j\in\eta_s}\delta(y_s-y_j)\right]} \right. \\ &\quad \left. \times \frac{\sqrt{2\pi\sigma_c^2}\,\sum_{r=1}^{R}\exp\left[\beta\sum_{j\in\eta_s}\delta(r-y_j)\right]}{\sqrt{2\pi\sigma_{y_s^*}^{*2}}\,\sum_{r=1}^{R+1}\exp\left[\beta\sum_{j\in\eta_s}\delta(r-y_j^*)\right]} \right\} \\ &\times \frac{|A_\theta|}{255}\,\frac{a^b}{\Gamma(b)}\left(\frac{\sigma_c^2}{\sigma_{c1}^{*2}\sigma_{c2}^{*2}}\right)^{b+1}\exp\left(a\left(\frac{1}{\sigma_c^2}-\frac{1}{\sigma_{c1}^2}-\frac{1}{\sigma_{c2}^2}\right)\right) \\ &\quad \times (1-p)\times\frac{1}{1\times 0.5\times 5^4} \\ &\times \frac{m_{R+1}\times \text{no. of non-empty classes}}{s_R\times(1+\text{no. of non-empty classes})\,P_{\text{alloc}}}\times 64\sigma_c^3, \end{aligned} \tag{10}$$

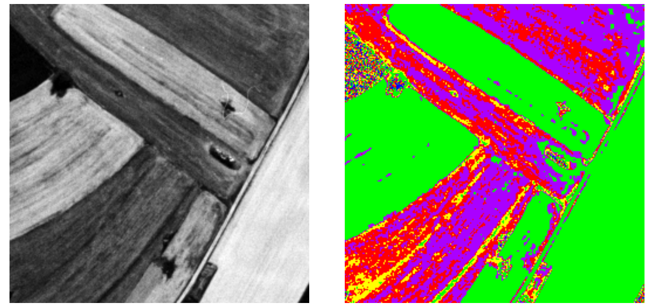

Figure 1: An aerial image on the left, with MAP segmentation, using splitting and merging of non-empty classes.

where $|A_\theta|$ is the volume of the allowable range of clique parameters (see [7]),

$$\begin{aligned} \epsilon^* &= x_s - \mu_{y_s^*}^* - \sum_{k=1}^{4}\sum_{j|<s,j>_k}\theta_{y_s^*,k}^*(x_j-\mu_{y_s^*}^*), \\ \epsilon &= x_s - \mu_c - \sum_{k=1}^{4}\sum_{j|<s,j>_k}\theta_{ck}(x_j-\mu_c), \end{aligned}$$

and P_{alloc} is the probability associated with the Gibbs sampling sweeps on the new textures, and is given by:

$$P_{\text{alloc}} = \prod_{s\in\mathcal{S}_c}\frac{\exp\left[-\left((\epsilon^*)^2/2\sigma_{y_s^*}^{*2}\right)+\beta\sum_{j\in\eta_s}\delta(y_s^*-y_j^*)\right]}{C_s\times\sqrt{2\pi\sigma_{y_s^*}^{*2}}}, \tag{11}$$

for normalising constant summing over the two new classes:

$$C_s = \sum_{r=c1,c2}\frac{\exp\left[-\left((\epsilon^*)^2/2\sigma_r^{*2}\right)+\beta\sum_{j\in\eta_s}\delta(r-y_j^*)\right]}{\sqrt{2\pi\sigma_r^{*2}}}.$$

We refer to [10] for details of the derivation of Equation 10. The accept probability for the merge move is the maximum of 1 and the inverse of Equation 10, considering the merge as a split from c to $c1$ and $c2$.

5 EXPERIMENTS

Figure 1 shows an aerial image, with the MAP solution from the reversible jump algorithm. The algorithm finally converged to 10 classes. This is an oversegmentation, in the sense of discriminating the main classes in the image: different fields, trees and the road. It has also segmented each field into different classes that correspond to different colourations in the image.

Figure 2 is an analysis of a satellite image of an area of Holland. In this case we end up with 24 classes. This represents another oversegmentation; for example, the reversible jump method of [6] finds 10 classes. The image consists of fields, which the algorithm classifies into

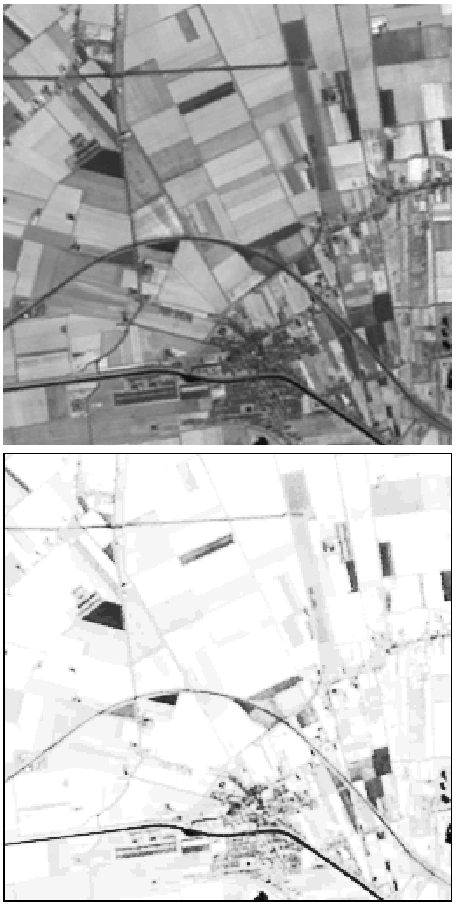

Figure 2: Image of an agricultural area (top), with MAP segmentation (bottom).

different classes according to mean intensity, and urban areas, which the algorithm segments according to mean intensity, ignoring the similar texture over all the urban areas. We also find that 90% of pixels are assigned to only 4 classes. In spite of this, the algorithm did not merge the classes assigned to only a small number of pixels. We refer to [10] for more details.

6 CONCLUSIONS

In this paper we have developed a segmentation algorithm based on Markov random fields and Bayesian inference, implemented with MCMC. The split/merge moves proposed here seem to oversegment the images, with many classes being assigned to small proportions of the image and not merged. Possibilities for reducing the number of classes are: larger neighbourhood orders for the labels, fixing β to be reasonably large, running an MPM (marginal posterior mode) sampler or some form of post-processing. Other MCMC approaches that infer the number of classes may also provide ideas, particularly those using stochastic geometry (see [5] for work in Projet Ariana). These will be topics for future work.

7 ACKNOWLEDGEMENTS

This work was done while the first author was a visiting researcher at Projet Ariana in INRIA Sophia-Antipolis during April – July 2000, supported by the European Science Foundation "Highly Structured Stochastic Systems" Programme, and has formed part of the European Union Research and Training Network "MOUMIR". Project Ariana is a joint project of CNRS, INRIA and Université de Nice-Sophia Antipolis. The authors would like to thank the French Space Agency (CNES) for providing SPOT satellite images and the French Mapping Institute (IGN) for providing the aerial data.

8 REFERENCES

[**1**] S. A. Barker and P. J. W. Rayner. Unsupervised image segmentation using Markov random field models. *Pattern Recognition*, 33:587–602, 200.

[**2**] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. B*, 36:192–236, 1974.

[**3**] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibb's random fields. *IEEE Trans. Patt. Anal. Machine Intell.*, 9:39–55, 1987.

[**4**] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[**5**] M. Imberty and X. Descombes. Simulation de processus objets: etude de faisabilité pour une application à la segmentation d'image. *Research Report* no. 3881, INRIA, Sophia Antipolis, 2000.

[**6**] Z. Kato. Bayesian color image segmentation using reversible jump Markov chain Monte Carlo. Technical Report 01/99-R055, European Research Consortium for Informatics and Mathematics, 1999.

[**7**] S. Lakshmanan and H. Derin. Valid parameter space for 2-d Gaussian Markov random fields. *IEEE Trans. Inform. Theory*, 39:703–709, 1993.

[**8**] D. E. Melas and S. P. Wilson. Double Markov random fields and Bayesian image segmentation. *IEEE Trans. Sig. Proc.*, 50:357–365, 2002.

[**9**] C. S. Regazzoni, A. N. Venetsanopoulos. Group-membership reinforcement for straight edges based on Bayesian networks. *IEEE Trans. Image Proc.*, 7:1–19, 1998.

[**10**] S. P. Wilson and J. Zerubia. Segmentation of textured satellite and aerial images by Bayesian inference and Markov random fields. *Research Report* no. 4336, INRIA, Sophia-Antipolis, 2001.