UNSUPERVISED IMAGE SEGMENTATION

Simon A. Barker and Peter J.W. Rayner

Cambridge University Engineering Department,
Trumpington Street, Cambridge CB2 1PZ, U.K.
e-mail sab@eng.cam.ac.uk

ABSTRACT

We present an unsupervised segmentation algorithm comprising an annealing process to select the maximum *a posteriori* (MAP) realization of a Hierarchical Markov Random Field (MRF) Model. The algorithm consists of a sampling framework which unifies the processes of model selection, parameter estimation and image segmentation, in a single Markov Chain. To achieve this, Reversible Jumps are incorporated into the Markov Chain to allow movement between model spaces. By using partial decoupling to segment the MRF it is possible to generate jump proposals efficiently while providing a mechanism for the use of deterministic methods, such as Gabor filtering, to speed up convergence.

1. INTRODUCTION

The classification of noisy or textured images into a number of different regions is a difficult problem. This is compounded when the number of regions into which the image is to be classified is also unknown. If each region is modelled by an individual likelihood function, then unsupervised segmentation may be treated as a model selection problem over a combined model space, where each model comprises a unique number of these functions.

To date, most unsupervised segmentation algorithms either exhaustively search the combined model space, or use approximate techniques consisting of two step processes: the first of these being a coarse image segmentation to find the number of states and estimate the associated model parameters; the second, a high resolution supervised segmentation algorithm, using the model and parameters from the first step. Such procedures make several fundamental assumptions which affect the final MAP segmentation. These include, specifying the minimum image area required to identify an individual state, defining an ad hoc distance measure between states to achieve fuzzy-clustering, or forming region homogeneity tests requiring a further arbitrary parameter.

To overcome such problems, a Markov Chain technique was developed in [1] using reversible jumps [2] to move between model spaces. This allowed direct sampling of the posterior distribution defined over the combined model space, thus reducing the optimisation process to a single annealing run. The reversible jump algorithm consists of a Metropolis-Hastings sampler with a dimension balancing element. This facilitates sampling from different model spaces, by incorporating proposals that increase or decrease model order. A major drawback to this algorithm is its slow speed of convergence which is attributable to the random nature in which new model parameters are proposed. The problem becomes intractable as the use of increasingly more complex models (e.g. when expanding the support of a Gaussian MRF) results in progressively lower rates of acceptance.

The approach adopted in this paper circumvents this problem by adopting a new method of generating proposals which represents a departure from existing algorithms. The basis of our approach is to estimate new model parameters from existing sub-partitions of the observed data when proposing to increase the model order. To achieve this, the image must be partitioned on a finer scale than that of the current segmentation and from a single sub-partition, parameter estimation must be possible. Finally, the algorithm must maintain detailed balance, to ensure samples continue to be drawn from the target distribution.

2. IMAGE MODELS

Algorithms are developed for both Isotropic and Gaussian MRF (GMRF) models. If the observed image, $\mathbf{y}$, is defined on lattice, Ω, then if $\mathbf{x}$ denotes the underlying MRF, its posterior distribution may be written;

$$\pi_k(\mathbf{x} \mid \mathbf{\Psi}_k, \beta, \mathbf{y}) = \frac{1}{Z_k(\mathbf{\Psi}, \beta)} \exp\left\{ U(\mathbf{y} \mid \mathbf{\Psi}_k, \mathbf{x}) + \beta \sum_{i \in \Omega} \sum_{j \in \eta_i} V(x_i, x_j) \right\} \quad (1)$$

where, $Z_k(\cdot)$ is its Gibbs partition function, dependent on both the k'th order likelihood function parameter vector, $\mathbf{\Psi}_k$ and the MRF hyper-parameter, β. $U(\cdot)$ is the free energy associated with the model's likelihood function, η_i is the nearest neighborhood at site i, and $V(x_i, x_j)$ is the potential function for the nearest neighbour clique found at sites i and j, which takes its value from $\{0, 1\}$. For model selection purposes, the partition function is approximated by its pseudo-likelihood equivalent, where the pseudo-likelihood function is defined to be the product of full conditional distributions for each x_i given its neighborhood state vector.

3. PARTIAL DECOUPLING

The sampling process follows a predetermined sequential scan, updating the pixel sites, the model parameters and the model order, respectively. Partial decoupling [3] is used to sample the underlying MRF states, while Metropolis - Hastings sub-chains are incorporated to sample model parameters and the model order.

Partial decoupling is a derivative of the conventional Swendson-Wang algorithm[4] which uses auxiliary variables to improve mixing when sampling from a dynamic system near its critical temperature. The algorithm gives improved mixing over the Swendson-Wang when the system is in the presence of an external field and interactions are strong (as is typically found when sampling the posterior distribution of an MRF). The auxiliary variables introduced by these algorithms may be considered bond variables which determine linkage between neighbouring pixel sites. The status of these variables allows the formation of clusters, consisting of groups of interconnected sites. Then, when sampling from the MRF's posterior distribution using the Swendson-Wang algorithm, the state of each cluster may be updated independently. However, when using partial decoupling, clusters are updated from a distribution conditioned on the surrounding sites. The extent of this conditioning is dependent on a matrix of bond variable hyper-parameters.

Partial decoupling allows the inclusion of prior local knowledge into the conditional distribution for each individual bond variable. Hence, a homogeneity measure for each pair of neighbouring pixels may be incorporated as a prior on the likelihood of bond formation, thus preventing the formation of large clusters in detailed areas of the image.

Using the models defined in the previous section, while denoting the auxiliary bond variables, $\mathbf{u}$, and their associated hyper-parameters, $\boldsymbol{\delta}$, the relevant conditional distributions for the partial decoupling sampler may be written:

$$p(u_{ij} \mid \mathbf{x}, \boldsymbol{\delta}, \beta) \propto \exp\{-\delta_{ij}\beta V(x_i, x_j)\} \times \mathrm{I}\left[0 \leq u_{ij} \leq \exp\{\delta_{ij}\beta V(x_i, x_j)\}\right] \quad (2)$$

$$p(x_i = k, i \in C \mid \mathbf{u}, \boldsymbol{\delta}, \beta, \boldsymbol{\Psi}_k) \propto \exp\left\{\sum_{i \in C}\left[U(y_i \mid \boldsymbol{\Psi}_k) + \sum_{j \in \partial C}(1 - \delta_{ij})\beta V(x_i, x_j)\right]\right\} \quad (3)$$

where I[·] denotes the indicator function taking values from $\{0, 1\}$, C is a set of pixels forming a single cluster, interlinked by bonds and ∂C is the ring of pixels surrounding C, which are also neighbours of the set of sites contained in C. To quantify the status of an auxiliary bond variable, it is clear from equation 2 that if $u_{ij} \geq \exp\{-\delta_{ij}\beta\}$ then the two sites, i an j are said to be bonded. The probability of this occuring, i.e. $1 - 2\exp\{-\delta_{ij}\beta\}$ can be seen to be affected by the hyper-parameter, δ_{ij}, so that greater the value, the greater the probability of bond formation.

The criterion for the generation of $\boldsymbol{\delta}$ is arbitrary. By using deterministic processes on the observed data, e.g. Gabor filtering, to generate $\boldsymbol{\delta}$, a mechanism is introduced that allows the incorporation of more efficient deterministic methods to improve convergence of the MCMC sampling process. The use of data to calculate hyper-parameters would appear to break all Bayesian rules but because $\boldsymbol{\delta}$ is an auxiliary variable matrix, the process chosen affects only convergence and not the target distribution.

For the Isotropic model, we choose the Kolomogorov-Smirnov (KS) distance as a measure of difference in localised grey scale distribution between overlapping regions surrounding neighbouring pixels. When considering the GMRF, an additional component is added indicative of spatial - frequency content. This is achieved by convolving the image with a bank of directionally dependent Gabor filters and calculating the Euclidean distance between the output vectors at each pair of neighbouring sites. To maintain a consistency in bond proliferation when applying the algorithm to different images, the distribution of hyper-parameters is normalised for mean and standard deviation.

Model parameters are sampled from their posterior distributions using the Metropolis-Hastings algorithm. Proposals are drawn from Normal distributions and the posterior distribution is obtained by forming the product of equation 1 with non-informative priors. The robustness of such a methodology would typically be low, since the model's characteristics will change throughout the annealing process, thus leading to the incorrect merging of states early in the optimisation process. However, this is overcome by the use of reversible jumps allowing the birth of fresh states.

4. REVERSIBLE JUMPS

To sample model order the reversible jump alogrithm is used to allow moves within the Markov Chain which increase or decrease model order. The reversible jump sampler uses a Metropolis-Hastings acceptance formula to allow transitions comprising two types of move: the splitting of one state into two and the merging of two states into one. A traditional Metropolis-Hastings sampler preserves detailed balance, so ensuring the ergodicity of the Markov Chain because the old and new model spaces are identical. However, when moving between different models, the associated parameter spaces may be of differing dimension. Hence to preserve detailed balance, the parameter vectors' dimensions are balanced by padding with random variables. If moving between models $m1$ and $m2$, whose parameter vectors are $\boldsymbol{\Psi}_{m1}$ and $\boldsymbol{\Psi}_{m2}$, then by extending these vectors by appending two random vectors, $\mathbf{e}_{m1}$ and $\mathbf{e}_{m2}$, their dimensions become matched.

To propose the new set of parameters, a continuous and invertible mapping between the two extended vectors is defined so that $[\boldsymbol{\Psi}_{m2}, \mathbf{e}_{m2}] = f([\boldsymbol{\Psi}_{m1}, \mathbf{e}_{m1}])$; thus, new parameters are generated by first drawing a random vector, then

applying such a mapping. The mapping must be continuous to allow the calculation of its Jacobian determinent, required in the Metropolis-Hastings acceptance ratio. When considering the splitting of state c into $c1$ and $c2$, this mapping will in general reduce to, $[\boldsymbol{\psi}_{c1}, \boldsymbol{\psi}_{c2}, \mathbf{e}_{m2}] = f([\boldsymbol{\psi}_c, \mathbf{e}_{m1}])$, where $\boldsymbol{\psi}_c$ indicates the parameter vector associated with the likelihood distribution for state c.

The mechanism for generating proposals adopted in this paper is somewhat different to those used in [1] but still satisfies the mapping requirement. Here, the region assigned to the state being split comprises pixels which have previously been grouped (by the partial decoupling algorithm) into smaller clusters. Two of the largest N_c of these clusters are randomly selected to generate two sets of maximum likelihood estimates of model parameters. These parameters are then used to propose a new segmentation of all pixels that are currently allocated to the existing, single state, into separate regions forming the two new states. To meet the requirement for a mapping between extended parameter vectors, thus satisfying the condition for detailed balance, a small random perturbation is added to the estimate; thus, if $\hat{\boldsymbol{\psi}}_{c1}$ and $\hat{\boldsymbol{\psi}}_{c2}$ are the maximum likelihood (ML) parameter estimates generated from the first and second randomly selected clusters and the random vector, $\mathbf{e}_{m1}$ can be expressed as two subcomponents, $[\mathbf{e}_{c1}, \mathbf{e}_{c2}]$, then the new model parameters are given by the mappings, $\boldsymbol{\psi}_{c1} = \hat{\boldsymbol{\psi}}_{c1} + \mathbf{e}_{c1}$ and $\boldsymbol{\psi}_{c2} = \hat{\boldsymbol{\psi}}_{c2} + \mathbf{e}_{c2}$.

For the Isotropic case the generation of the ML parameter estimates is straightforward. However, for the GMRF calculating the ML parameter estimates based on data contained within a single cluster is not so trivial, hence a simpler approach is adopted: to maximize the pseudo-likelihood. To achieve this, define the vector of statistics, $\mathbf{c}$, whose elements are given by,

$$c_\tau = \sum_{i \in C} (y_i - \mu)[(y_{i+\tau} - \mu) + (y_{i-\tau} - \mu)]$$

where τ is the index indicating the translation corresponding to the GMRF correlation parameter, θ_τ. Then define the matrix, $\mathbf{D}$, whose elements are indexed by translations τ_1 and τ_2 and are given by,

$$d_{\tau_1,\tau_2} = \sum_{i=1}^{n} [(y_{i+\tau_1} - \mu) + (y_{i-\tau_1} - \mu)] \times [(y_{i+\tau_2} - \mu) + (y_{i-\tau_2} - \mu)]$$

hence if $\hat{\mu}$ denotes the ML mean estimate for cluster C and n is the cluster size,

$$\hat{\boldsymbol{\theta}} = \mathbf{D}^{-1}\mathbf{c} \text{ and } \hat{\sigma}^2 = \frac{1}{n}\Big[\langle y_i^2 \rangle - \hat{\mu}^2 - 2\hat{\boldsymbol{\theta}}^T \mathbf{c} + \hat{\boldsymbol{\theta}}^T \mathbf{D}\hat{\boldsymbol{\theta}}\Big].$$

The two sets of model parameter proposals may be used as the basis for the re-segmentation of all pixels previously allocated to state c into multiple regions, assigned to both $c1$ and $c2$. The methodology used is simple and somewhat intuitive: clusters are re-allocated using the partial decoupling algorithm's cluster coloring equation 3. However, a state of ignorance exists when considering the conditioning of this equation on each cluster's neighborhood configuration. To overcome this, all pixels of the split state are first assumed to belong to another, different state. Next, clusters are allocated to one of the new pair of states in order of size, the largest first, by Gibbs sampling using their conditional probabilities. The algorithm therefore colours the largest clusters relatively independently but as the size of the clusters decreases and the gaps are filled between the larger, already allocated clusters, the clusters' coloring distributions become conditioned to an ever greater extent on their local neighbourhoods. The probability of proposing a re-segmentation of the label field $\mathbf{x}$ to $\mathbf{x}^+$ is given by,

$$q(\mathbf{x}, \mathbf{x}^+) = \prod_{C \in \boldsymbol{\mathcal{C}}_c} \frac{p(x_i, i \in C \mid \mathbf{u}, \boldsymbol{\delta}, \beta, \boldsymbol{\psi}_{x_i})}{\sum_{c \in \{c1,c2\}} p(x_i = c, i \in C \mid \mathbf{u}, \boldsymbol{\delta}, \beta, \boldsymbol{\psi}_c)}$$

where $\boldsymbol{\mathcal{C}_c}$ is the set of all clusters allocated to class c, before the proposed split, and the distributions, $p(x_i = k, i \in C \mid \mathbf{u}, \boldsymbol{\delta}, \beta, \boldsymbol{\psi}_k)$ are the the cluster coloring partial decoupling conditional distributions, as given by equation 3.

When considering the opposite move, the combining of two states into one, the reverse proposal probability (identical to that if proposing to split the clusters comprising the new merged state into the two original states) must be calculable. Because the order in which the clusters are re-allocated is deterministic, i.e. based on cluster size only, the back calculation of the reverse allocation probability is possible using the above equation.

The resulting acceptance ratio for the splitting the region labelled by class c into regions labelled by $c1$ and $c2$ is given by min[1, R], where,

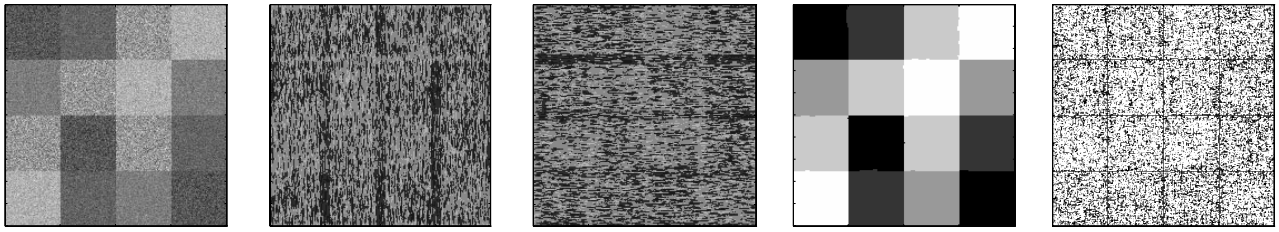
$$\mathrm{R} = \frac{\pi(\mathbf{x}^+, \boldsymbol{\psi}^+ \mid \mathbf{y})}{\pi(\mathbf{x}, \boldsymbol{\psi} \mid \mathbf{y})} \frac{q(c1, c2)}{q(c)} \frac{q(\boldsymbol{\psi}_c \mid \boldsymbol{\psi}_{c1}, \boldsymbol{\psi}_{c2}, \mathbf{x})}{q(\boldsymbol{\psi}_{c1}, \boldsymbol{\psi}_{c2} \mid \mathbf{u}, \mathbf{y})} \frac{1}{q(\mathbf{x}, \mathbf{x}^+)}$$

The first term comprises the ratio of the posterior probabilities after and before the proposed split and the remaining terms are the probabilities of generating the proposals, defined: $q(c)$ and $q(c1, c2)$ are simply the probabilities of selecting the state or states to be split or combined, $q(\boldsymbol{\psi}_c \mid \boldsymbol{\psi}_{c1}, \boldsymbol{\psi}_{c2}, \mathbf{x})$ is the probability of generating new model parameters when combining a state, $q(\boldsymbol{\psi}_{c1}, \boldsymbol{\psi}_{c2} \mid \mathbf{u}, \mathbf{y})$ is that when generating two sets of new parameters when splitting one state into two and $q(\mathbf{x}, \mathbf{x}^+)$ is that of re-segmentation.

The Jacobian determinant corresponding to the transformation between extended parameter vectors in this case is unity and is thus omitted from the expression. The acceptance ratio for the combine move is simply the inverse of the above.

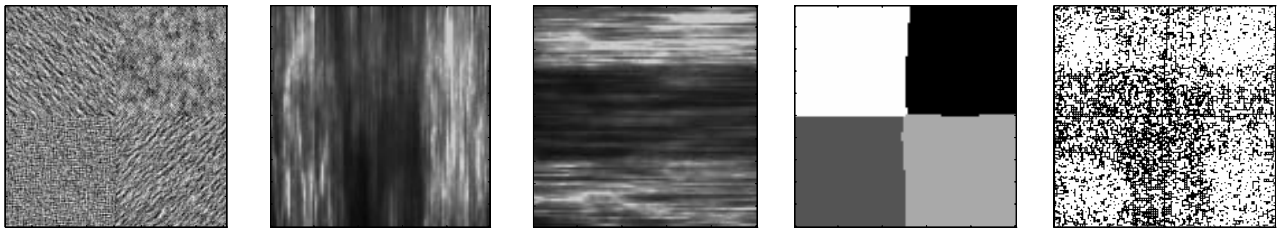
5. RESULTS

Several sets of experimental results are shown, demonstrating the application of the unsupervised algorithm to both isotropic and textured or GMRF models. The annealing schedules used for all experiments are linear. The MRF hyperparameter, β is set *a priori* to values between 1 and 1.8. The



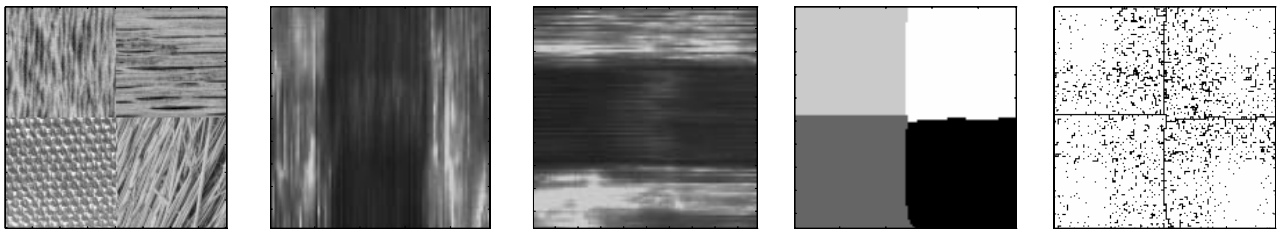
Original Image Horizontal Bond Prior Vertical Bond Prior Final Segmentation Final Bondmap

Figure 1: 300 iteration segmentation experiment using an Isotropic MRF and a synthesized grayscale mosaic.



Original Image Horizontal Bond Prior Vertical Bond Prior Final Segmentation Final Bondmap

Figure 2: 200 iteration segmentation experiment using a 4 parameter GMRF and a synthesized texture mosaic.



Original Image Horizontal Bond Prior Vertical Bond Prior Final Segmentation Final Bondmap

Figure 3: 1000 iteration segmentation experiment using a 22 parameter GMRF and a Brodatz texture mosaic.

effect of altering its value between these margins proved insignificant when considering the mosaic examples given here but when considering more subtle images, a more noticeable effect has been observed.

The original image together with horizontal and vertical priors for the auxiliary bond variable hyper-parameters are shown beside the final segmentation and its accompanying bondmap. When observing the priors, the light coloured regions correspond to areas of homogeneity between pixels, thus giving high bond prior probabilities. The final bondmap effectively comprises a negative of an image containing the borders between regions, with most internal pixels being bonded. The improvement in convergence over [1] is apparent when observing the results for the Isotropic model, shown in Figure 1: the number of iterations required to achieve a similar quality of segmentation has fallen from 500 to 300.

6. CONCLUSION

In this paper we have presented unsupervised segmentation algorithms for both Isotropic and GMRF's. The optimisation process, a single annealing process, is evaluated over the combined model space by incorporating reversible jumps into the Markov Chain to allow movement between models of different order. The MRF sites are updated using partial-decoupling to give improved mixing as the annealing process passes through the critical temperature of the system. The algorithm also facilitates the generation of efficient reversible jump proposals and provides a mechanism for incorporating determinstic methods to improve convergence.

[1] S.A. Barker and P.J.W. Rayner. Unsupervised Image Segmentation using Markov Random Field Models. In *Lecture Notes in Computer Science*, volume 1223. Springer-Verlag, 1997.

[2] P.J. Green. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1996.

[3] D. Higdon. Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications. *unpublished preprint - Duke University, available at http://www.stats.bris.ac.uk/MCMC/*, Jul 1996.

[4] R.H. Swendson and J. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, Jan 1987.