Coloured continuum triangulation models in the Bayesian analysis of two dimensional change point problems.

Geoff Nicholls
Mathematics Department, The University of Auckland, Auckland, New Zealand

May 1997

Abstract

We consider Bayesian image segmentation from a continuum parametric model. This is a class of problems in which a great deal of prior information is generally available. This information is often hierarchical in character, concerning the appearance of a range of different types of composite structures, including edges, vertices, regions, and the topology of the boundary graphs. Bayesian inference proceeds most simply, and flexibly, when the states of the chosen prior model are described in terms of the same elements used to specify prior information. We explore a model in which coloured continuum triangulations are used to represent image structure. We show how an initial model may be modified to take account of image structure. Changes to the density are incorporated in a Metropolis Hasting MCMC algorithm by slight modifications of generation and/or acceptance probabilities, which can require just a few lines of code.

1 History and Motivation

Spatial patterns may include "unexpected" objects. Such objects will not be represented by one of a limited set of "high-level" object templates; we hope, however, that they may be represented in terms of some set of more elementary "intermediate-level" components, such as edges, edge junctions and regions. For this reason, high level inference may prove to be more conveniently conducted from intermediate level elements than directly from the data. This is the familar trade off (Marr, 1982) between between bottom-up and top-down reasoning, which a Bayesian/sampling approach to image analysis can encompass.

Mosaic models of spatial data are "vector based", continuum models of general planar patterns. A state is a coloring $\chi(x)$ of the points in a region $x \in \mathcal{D}$ of the plane. It is specified by giving a graph, and then saying how it is to be imbedded, or drawn, in the plane, and giving the coloring of each face. We use τ to denote the set of points in a particular imbedded graph, so that $\tau \subset \mathcal{D}$. We may model topological properties of the graph, geometric properties of the imbedding, and the details of the face coloring. The considerations in the first paragraph motivate the construction of mosaic models.

The second, and more immediate motivation for constructing mosaic models, is that it may be convenient for a statistician to interpret spatial data at the intermediate level. In this case the transformation from data to continuum mosaic can be thought of as a Bayesian approach to surface regression, in which change curves in the surface are modeled in an explicit way. This will be the useful when spatial datasets are complex, perhaps combining several different types of measurement over a single region. Thus the approach to image analysis described in this paper is distinct from the approach taken in, for example, Grenander and Miller (1994), or Baddeley and van Lieshout (1994). We have no object-process. Instead we aim to build an intermediate level representation of image structure.

Recent developments in Bayesian image analysis have been driven primarily by the three or four hundred fold increase in computing power since the late eighties. At that time "The option of modeling the continuous process, and basing inference on the sampled or aggregated data is usually computationally prohibitive" (Ripley 1988). Much of the sampling and model theory on which recent work depends was available, and simple continuous-time reversible jump Monte carlo was used in papers such as Ripley and Sutherland (1990). The sheer complexity of the sampling problem was another limiting factor. In this respect, a series of generalisations of the Metropolis Hastings scheme for Markov chain Monte Carlo (notably the reversible jump Monte carlo of Geyer and Moller (1994) and Green (1995) and the Hybrid sampler of Tierney (1994) in the form described in Green 1995) have made it possible to simulate complex systems with managable algorithms and relatively simple modular sampling software. The main complication remaining in implementing these MCMC algorithms is the problem of maintaining a representation of the evolving state. Clifford and Nicholls (1994) give a case study.

2 Triangulation Mosaics

We illustrate intermediate level inference from a colored continuum triangulation mosaic model of the kind described in Nicholls (1996). Consider the segmentation depicted in Figure 1A. A number of triangulations of this segmentation are possible, and one is shown in Figure 1C. If points may be added, the number is infinite. A triangulation mosaic measure does not directly assign probability to a particular segmentation, as in Figure 1A, but rather to a colored triangulation. This means that the probability associated with a segmentation is the marginal over all equivalent colored triangulations. Let $\mathcal{D}$ be the set of points in an open subset of the plane defined by a simple polygon with boundary $\partial\mathcal{D}$. A triangulation is formed by adding node points on the boundary and in the interior of $\mathcal{D}$, and joining the nodes with straight edges until no more can be added without intersecting an edge already in place. The corner nodes of the polygon $\mathcal{D}$ are included in this process. An example is given in Figure 1C. Let $\Gamma_{\mathcal{D}}$ be the set of all distinct triangulations which can be constructed in $\mathcal{D}$ by such a process, and let τ be a triangulation in $\Gamma_{\mathcal{D}}$. Let $\mathcal{J}$ be a vector of N color levels, $\mathcal{J} = (c_1, c_2 \ldots c_N)$. A colored triangulation is formed by assigning to each point in each triangular face in τ a color in $\mathcal{J}$. For each triangulation τ, let $C(\tau)$ be the set of all distinct face colorings of τ and let c be a vector $c \in C(\tau)$ with entries

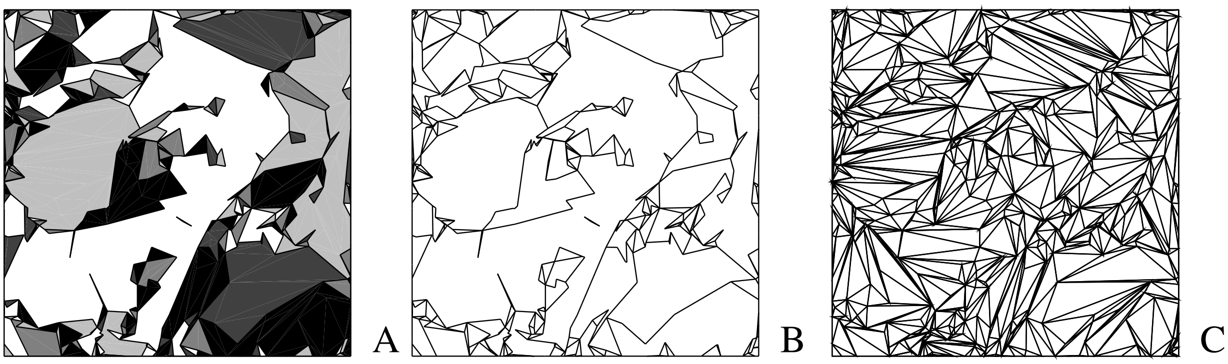


Figure 1: (A) A mosaic state, $\chi_{\tau,c} \in \Omega_{\mathcal{D}}$, with five colors. (B) The discontinuity set $\partial\chi_{\tau,c}$ of the state in (A). (C) The triangulation, τ underlying the state in (A).

recording the color in each face in τ. If τ is composed of F triangles then there will be F^N face color vectors c in $C(\tau)$. The pair (τ, c), with $c \in C(\tau)$, attaches a color $\chi(x) \in \mathcal{J}$ to each point $x \in \mathcal{D}$, and constitutes a triangulation mosaic, as in Figure 1A. Let $\chi_{\tau,c}(x)$ denote the coloring at the point x given by the pair (τ, c). Let $\Omega_{\mathcal{D}}$ be the set of all triangulation mosaics, $\chi_{\tau,c}$, and let χ denote a generic state in $\Omega_{\mathcal{D}}$. Let $\partial\chi$ denote the discontinuity set of the triangulation mosaic χ, as depicted in Figure 1B. $\partial\chi$ is the set of points in edges of τ having adjacent faces of unequal color. $\partial\chi$ is the primary object for modeling since it gives the boundary edges of a segmentation of $\mathcal{D}$ into general polygonal regions of uniform color. Notice that a given segmentation has an infinite number of representations in terms of triangulation mosaics.

Let $d\tau$ be the measure on $\Gamma_{\mathcal{D}}$ defined in Nicholls (1996), so that $d\tau/\mathcal{Z}$ is a probability measure for some finite normalising constant $\mathcal{Z}$. This measure depends on a parameter β giving the intensity of the point process generating node points of the triangulation on $\partial\mathcal{D}$, the region boundary. Increasing β has the effect of shifting the probability mass in $d\tau/\mathcal{Z}$ onto triangulations with a greater number of node points, so that the typical triangulation is more complex. We take the intensity of the point process generating node points in the interior to be β^2. Let $d\chi_{\tau,c} \equiv d\tau/F^N\mathcal{Z}$ be the probability measure derived from $d\tau$ by taking a uniform distribution over the set of colorings of a triangulation. We define a pair of densities $f_L(\chi;\theta_L)$ and $f_E(\chi;\theta_E)$, $\chi \in \Omega_{\mathcal{D}}$. For a mosaic with discontinuity set $\partial\chi$ these densities penalise the overall length $L(\partial\chi)$ of the discontinuity, and the total number of edges $E(\partial\chi)$ in the discontinuity in the coloring. We take

$$\begin{aligned} f_L(\chi;\theta_L) &= \exp(-\theta_L L(\partial\chi)) && (1) \\ f_E(\chi;\theta_E) &= \exp(-\theta_E E(\partial\chi)). && (2) \end{aligned}$$

The second density corresponds to taking an Ising ($N = 2$) or Potts ($N > 2$) model with parameter θ_E on the lattice of faces for a given triangulation. In the first density, the same applies, though the lattice coupling parameter for the color at adjacent faces is now inhomogeneous, and proportional to the length of the shared edge. The probability measures $f_L(\chi;\theta_L)d\chi_{\tau,c}$ and $f_E(\chi;\theta_E)d\chi_{\tau,c}$ each depend on $N+2$ parameters, viz. the lumping parameter θ_L or θ_E, a node intensity β and the N values in the vector $\mathcal{J}$ of color levels. In the following we will treat the entries c_i in $\mathcal{J}$ as iid uniform random variables, $c_i \in [0,1]$, so that $\mathcal{J} \in [0,1]^N$. In the context of a grey level image this means that we assume the image to contain just N distinct grey levels, however the values of these N levels are unknown, and we wish to recover them. In this case we model with just two parameters, a lumping parameter θ_L or θ_E, and a node intensity β.

The processes we have defined are not Markov, even though they are based on *additive* potentials. The potential function $L(\partial\chi)$ is additive since for any open set S in $\mathcal{D}$, with complement $S^c = \mathcal{D} \setminus \mathcal{S}$, the potential for the whole coloring is the sum of the potentials for the subregion colorings, $L(\partial\chi) = L(S \cap \partial\chi) + L(S^c \cap \partial\chi)$. The potential $E(\partial\chi)$ is additive if we count edges by counting, for example, edge midpoints. Additivity is desirable in models which must be analyzed by MCMC sampling, as it ensures local modifications to the state impose a function evaluation which depends only on state properties in the update region.

Useful values for the parameter β will depend on the amount of detail to be distinguished in the data. Useful values for θ_L range from $\theta_L \simeq \beta$ up to a few times β. Simulations from the prior show that for $\theta_L > \beta$, states are dominated by a single color. Indeed there is a sharp transition from disordered to ordered states at around $\theta_L = \beta$. Samples from the prior with density f_L are shown in Figure 2. For θ_E the corresponding useful range is from $\theta_E \simeq 0.5$ to a few times that value, indpendent of β.

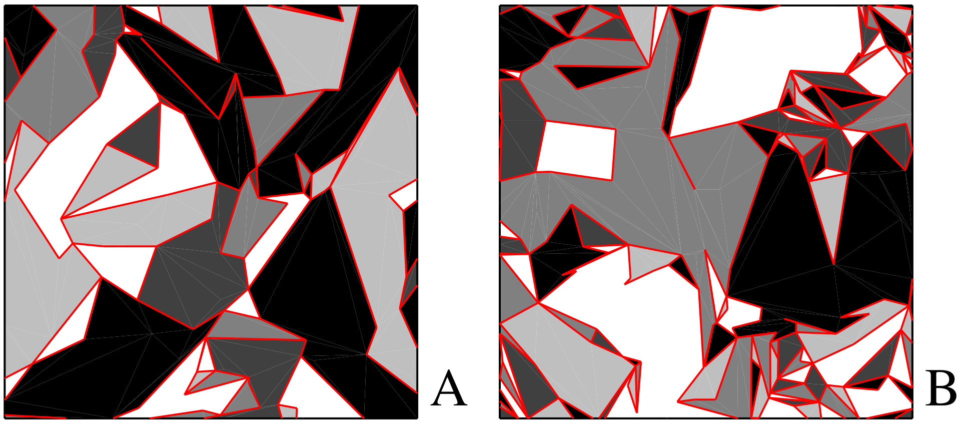


Figure 2: Samples from the prior with density f_L. (A) The distribution conditioned on the restricted vertex set of Figure 5, $\Omega^*_{\mathcal{D}} = \Omega^{*,2}_{\mathcal{D}}$. $\beta = 10$, and $\theta_L = 8$. (B) A sample from the prior using the unrestricted state space, $\Omega^*_{\mathcal{D}} = \Omega_{\mathcal{D}}$. $\beta = 10$, and $\theta_L = 11$. Parameter values have been chosen to generate interesting states.

3 Sampling the Posterior distribution

We illustrate our model by making a study of a Synthetic Aperture Radar (SAR) image. The observation process in SAR is complex. We begin by assuming that the observed grey-level g_i at the pixel i, $i = 1, 2 \ldots N$ corresponding to the point y_i in $\mathcal{D}$, is a Gaussian random variable, with mean equal the unknown true grey-level, $\psi(y_i)$, at y_i, and standard deviation σ. This observation model is not accurate, so we allow for some overdispersion; Let $G_T(\chi, y, g) =$

$\sum_{n=1}^{N} \min(9\sigma^2, (\chi(y_i) - g_i)^2/2\sigma^2)$, with y and g vectors with components respectively the location of, and grey-level at, a pixel. Our likelihood has a density equal $\exp(-G_T(\chi, y, g)$. The posterior measure on colorings is

$$\Pr(d\chi|g) \propto exp(-V(\chi) - G_T(\chi, y, g))$$

where $V(\chi)$ is the prior potential, $L(\partial\chi)$ or $E(\partial\chi)$. We refer to $V(\chi) + G_T(\chi, y, g)$ as the *posterior potential*.

The MCMC sampler we use to sample the state is described in Nicholls (1996). However we will outline the algorithm, and add to the description already given. We wish to define a stochastic process generating a sequence $X_0, X_1 \dots X_t \dots$ of states $X_t \in \Omega_{\mathcal{D}}$ with the property that X_t has a distribution $\Pr(X_t \in d\chi_{\mathcal{T},c}|X_0 = \chi_0)$ which tends in the limit $t \to \infty$ to $\Pr(d\chi|g)$. We use Metropolis Hastings MCMC to define a transition generating X_{t+1} given $X_t = \chi_{\mathcal{T},c}$. There are four steps. Let $\Omega^*_{\mathcal{D}}$ be the set of allowed coloring states. For the present this set is identical to $\Omega_{\mathcal{D}}$.

1. A candidate state $\chi_{\mathcal{T}',c'}$ is generated according to some simple stochastic process. For any pair of states it must be possible to genarate a sequence of states connecting the pair. To achieve this we select operations from a set of *move-types* labeled $m, m = 1, 2 \dots 6$. These are shown in Figure 3. Each move operates in some subregion S of $\mathcal{D}$ formed from the

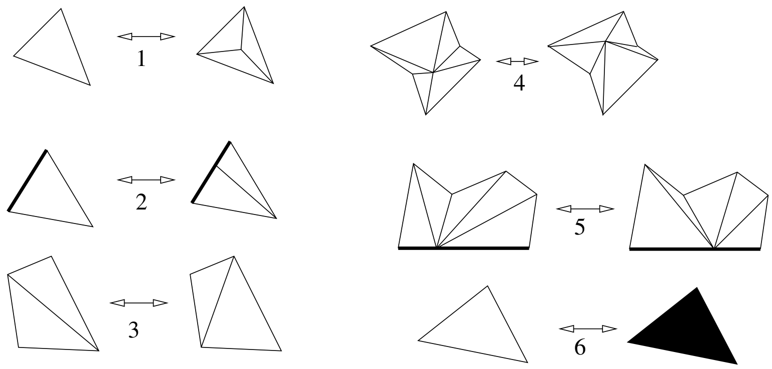


Figure 3: The set of moves used to explore the space of triangulation mosaics, $\Omega_{\mathcal{D}}$. The color at a point is not changed by moves 1-3. In moves 4 and 5 triangles maintain their colour as their vertex is shifted. The density f_L is modified by moves 4-6, while the density f_E is modified by move 6 only.

union of a small number of triangles. We generate $\chi_{\mathcal{T}',c'}$ by applying a randomly selected operation of type m, with probability measure $q_m(d\chi_{\mathcal{T}',c'}|\chi_{\mathcal{T},c})$.

2. We check the generated state is *legal*, *ie* that $\chi_{\mathcal{T}',c'} \in \Omega^*_{\mathcal{D}}$. If the state is not in $\Omega^*_{\mathcal{D}}$ then it is rejected, and we set $X_{t+1} = \chi_{\mathcal{T},c}$.

3. We compute the ratio

$$R(\chi_{\mathcal{T}',c'}|\chi_{\mathcal{T},c}) = \frac{\exp(-V(\chi_{\mathcal{T}',c'}) - G_T(\chi_{\mathcal{T}',c'}, y, g))d\chi_{\mathcal{T}',c'}q_m(d\chi_{\mathcal{T},c}|\chi_{\mathcal{T}',c'})}{\exp(-V(\chi_{\mathcal{T},c}) - G_T(\chi_{\mathcal{T},c}, y, g))d\chi_{\mathcal{T},c}q_m(d\chi_{\mathcal{T}',c'}|\chi_{\mathcal{T},c})}.$$

This requires only local computations, since contributions to the posterior potential from outside S cancel in the above ratio, $V(\chi_{\tau,c}) - V(\chi_{\tau',c'}) = V(S \cap \chi_{\tau,c}) - V(S \cap \chi_{\tau',c'})$. One must check that R is indeed a density, ie that the measures in the numerator and denominator have a common dominating measure (Green 1995).

4. The acceptance probability is $\alpha(\chi_{\tau',c'}|\chi_{\tau,c}) = \min(1, R(\chi_{\tau',c'}|\chi_{\tau,c}))$. With probability α, we accept the move (and set $X_{t+1} = \chi_{\tau',c'}$). Otherwise the move is rejected and we set $X_{t+1} = \chi_{\tau,c}$.

There are two reasons for our choice of a Metropolis Hastings sampler. The first is now widely acknowledged. Continuous time birth and death processes, used in the past to simulate processes in which the state vector has randomly variable dimension, share with the Gibbs sampler the undesirable feature that a normalisation must be maintained over the set of states into which a jump may be made. When this "local" normalisation is available, it should be exploited, as continuous time processes in particular are observed to converge to equilibrium on multi-modal distributions where a comparable Metropolis Hastings process does not converge (Nicholls 1994). However, the local normalisation is not readily calculated, and we avoid the calculation using a Metropolis Hastings scheme. There is a related consideration. When we wish to introduce constraints on the sampled state, modifying its state space by conditioning out certain classes of state, we can often do so in just a few lines of code, by rejecting any illegal states generated. The constraint modifies the set of legal states $\Omega^*_{\mathcal{D}}$. If an illegal state can be identified efficiently, and illegal states are not generated too frequently, so that ergodicity is lost in theory or in practice, and it becomes necessary to modify the generation process, then this kind of conditioning on a state property is straightforward. This is possible in the alternative MCMC processes on the condition that the modification to the state space can be straightforwardly incorporated in the local normalisations.

As a first example, we wish to eliminate the atom of probability that exists in the model for an edge in $\partial\chi$ to be incedent a corner node. Since there is a node point on the boundary of $\mathcal{D}$ at the corners, there is a finite probability for a triangulation to have an edge connected at exactly that point, and consequently a finite probability for the coloring discontinuity of a state $\chi_{\tau,c}$ to attach to the boundary at a corner. To remove this undesirable feature, we let $\Omega^{*;1}_{\mathcal{D}}$ be the set of states in $\Omega_{\mathcal{D}}$ in which discontinuities $\partial\chi_{\tau,c}$ do not attach at corner nodes, and take $\Omega^*_{\mathcal{D}} = \Omega^{*;1}_{\mathcal{D}}$ in the algorithm above. We add, in our implementation of step 2, a test, which is applied to each vertex in the update region S: if the vertex is a corner node, the update is rejected. This is equivalent to sampling the posterior distribution $\Pr(d\chi|g, Z)$, conditioned on the information "Z" that a corner node is not a vertex.

4 Modeling with contextual information

Consider the detail in Figure 4, where data in a subregion ω of a larger region $\mathcal{D}$ is displayed, along with a possible line interpretation. We are not concerned with what the true interpretation of this image is, but with the set of plausible interpretations. Support for the conjectured boundaries is weak, and depends on the extrapolation of lines from surrounding boundaries. Suppose that there is a state $\psi \in \Omega_{\mathcal{D}}$ corresponding to the "true" grey level in this image. Let $\partial\psi$ be

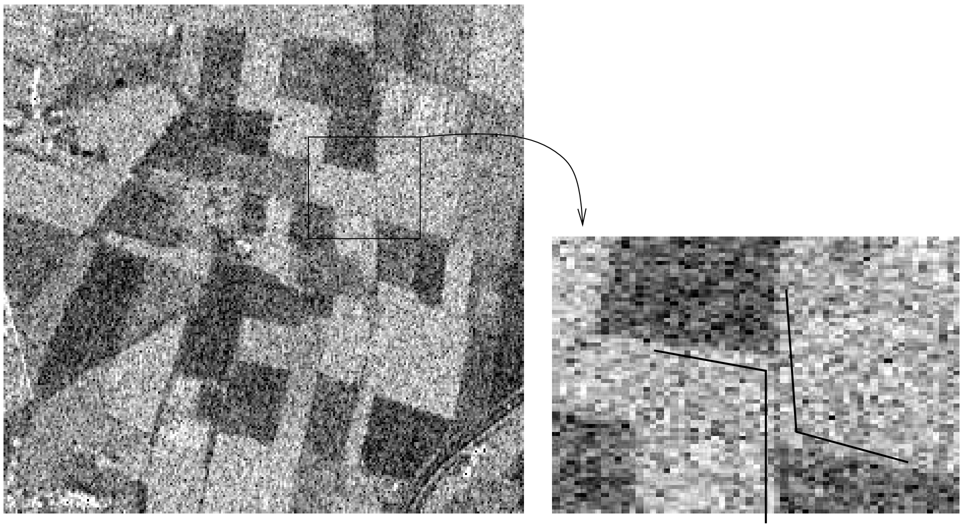


Figure 4: Log-transformed SAR image of Cambridgeshire crops, and detail, obtained via BIOSS, Scotland, from the Joint Reasearch Center, Ispra. The lines indicate a conjectured segmentation in a small patch.

the discontinuity set of this unknown true coloring. We have topological information about the nature of the change curve. In particular we expect vertex events in an image of fields to be one of the set listed in Figure 5, the V, T, Y and X vertices. The I vertex occurs only where an

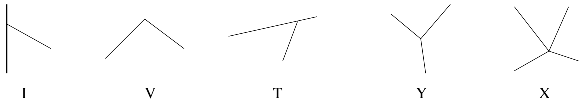
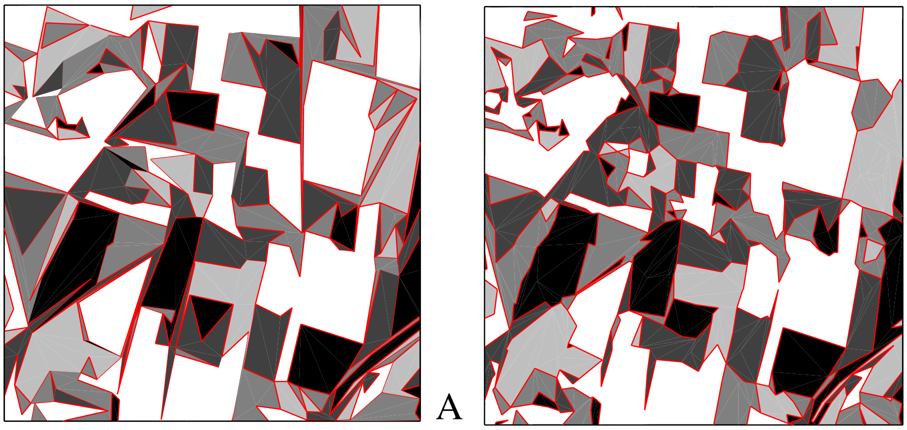


Figure 5: The set of allowed vertices in the model of the SAR data.

edge in $\partial\chi$ meets the boundary. This corresponds to a constraint on the topology of the graph $\partial\chi$. Note that vertices of the triangulation τ which are not on coloring discontinuities may have any degree. We let $\Omega^*_{\mathcal{D}} = \Omega^{*,2}_{\mathcal{D}}$ be those states in $\Omega^{*,1}_{\mathcal{D}}$ with vertices of degree two, three or four, or one if on the boundary. We must check that "irreducibility" is retained (or strictly, ergodicity in the sense of Harris). We check that the empty state is accessable, with finite probability in a finite number of steps, from any state.

A pair of states sampled from the posterior, with prior densities f_L and f_E, are shown in Figure 6. Note that these images were sampled using the full SAR image, a 256×250 image. MCMC convergence was not satisfactory on these images, for runs of the order of one day in length. Convergence on smaller datasets is easilly attained. A prior penalising length tends to give a more rounded shape to segemented regions, whilst the prior penalising edge number is particularly good at extropolating edges through unclear regions. For the SAR image we have more fundamental reasons for favouring f_E as a model. The fundamental event is the edge, corresponding to a field boundary. The physical boundary creation process created entire straight edges. The model should reflect this, and avoid breaking a line into a chain of colinear edges. This is part of common modeling practice: the elements of the model should simulate the physical events generating the data.



A B

Figure 6: (A) A sample from the posterior obtained using the density f_E, with $\beta = 10$ and $\theta_N = 4$. (B) A sample obtained using the density f_L, with $\beta = 20$ and $\theta_L = 100$.

References

Baddeley, A. J. and van Lieshout, M. N. M. (1994). Stochastic geometry models in high-level vision. In Mardia, K. V., editor, *Statistics and Images: Vol. 2*. Carfax, Oxford.

Clifford, P. and Nicholls G. K. (1994). A Metropolis sampler for polygonal image reconstruction. Technical Report, Dept. Statistics, University of Oxford, UK.

Green, P. (1995). Reversible jump MCMC and Bayesian model determination. *Biometrika*, 82:711–732.

Grenander, U. and Miller, M. (1994). Representation of knowledge in complex systems (with discussion). *Journal of the Royal Statistical Society B*, 56:549–603.

Marr, D. (1982). *Vision*. W. H. Freeman and Co, San Francisco.

Nicholls G. K. (1996). Bayesian image analysis with Markov Chain Monte Carlo and colored continuum triangulation mosaics. *Journal of the Royal Statistical Society B*, to appear.

Nicholls G. K. (1994). Comparison of Birth-and-Death and Metropolis-Hastings Markov chain Monte Carlo for the Strauss process. Technical Report, Dept. Mathematics, University of Auckland, New Zealand.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge University Press, Cambridge.

Ripley, B. D. and Sutherland, A. I. (1990) Finding spiral structures in images of Galaxies 332:477–485