Adv. Appl. Prob. **26**, 281–300 (1994)
Printed in N. Ireland

STOCHASTIC ANNEALING FOR NEAREST-NEIGHBOUR POINT PROCESSES WITH APPLICATION TO OBJECT RECOGNITION

M. N. M. VAN LIESHOUT, * *University of Warwick*

Abstract

We study convergence in total variation of non-stationary Markov chains in continuous time and apply the results to the image analysis problem of object recognition. The input is a grey-scale or binary image and the desired output is a graphical pattern in continuous space, such as a list of geometric objects or a line drawing. The natural prior models are Markov point processes found in stochastic geometry. We construct well-defined spatial birth-and-death processes that converge weakly to the posterior distribution. A simulated annealing algorithm involving a sequence of spatial birth-and-death processes is developed and shown to converge in total variation to a uniform distribution on the set of posterior mode solutions. The method is demonstrated on a tame example.

CONVERGENCE IN TOTAL VARIATION; HOUGH TRANSFORM; LIKELIHOOD RATIO; MAXIMUM A POSTERIORI ESTIMATION; NEAREST-NEIGHBOUR MARKOV PROCESSES; SPATIAL BIRTH-AND-DEATH PROCESSES; STOCHASTIC ANNEALING

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60D05
SECONDARY 60J27, 62F15, 68T10

Introduction

Baddeley and Van Lieshout [3], [4], [5] developed a statistical approach to the problem of object recognition in image analysis. A scene composed of possibly overlapping objects is observed subject to blur and noise, and the task is to determine the number of objects and to locate them. Applications include document reading and robot vision. Their approach is formally similar to the well-known Bayesian formulation of low-level image segmentation and classification problems due to Besag [7] and Geman and Geman [9], but differs in that pixel-based Markov random fields are replaced by continuous-space Markov spatial processes borrowed from stochastic geometry [28], [1]. The latter models combine spatial information globally, and are better suited to high-level image interpretation.

The deterministic algorithms presented in [4] are strongly analogous to Besag's ICM method [7]. An obvious question is whether there is also an analogue of stochastic annealing [9], [11] in this context. The present paper explores the

Received 24 May 1993; revision received 29 November 1993.
* Present address: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.
Research carried out at the Free University, Amsterdam, and CWI, Amsterdam.

existence and convergence of certain spatial birth-and-death processes which are the counterparts of stochastic annealing algorithms (see also [3]). Similar questions in autoradiography are apparently studied by Miller et al. [23].

The plan of the paper is as follows. The first section gives detailed background and notation. Convergence in total variation of inhomogeneous Markov processes is the subject of Section 2, while Section 3 studies existence and convergence of spatial birth-and-death processes. The results are used in Section 4 to develop analogues of simulated annealing in object recognition. Finally, Section 5 gives a simple concrete example.

1. Preliminaries

This section gives an overview of notation and the main concepts to be used throughout the paper. For a more detailed description and examples we refer to [3], [4].

1.1. *Notation.* The experimental data consist of an image $\boldsymbol{y} = (y_t; t \in T)$ where the *image space* T is an arbitrary finite set. Apart from the usual two-dimensional rectangular grids, T could be a pair of grids (carrying left and right stereo images), a temporal sequence, etc. The observed value y_t at pixel $t \in T$ ranges over a non-empty set V, typically $\{0, 1\}$ or $\{0, 1, \cdots, 255\}$.

The class U of possible objects is an arbitrary set (*object space*), ranging from simple geometrical figures (lines, discs) specified by a few parameters to completely general closed sets. U is treated as a space in its own right, so that objects are regarded as points u in U, each determining a subset $R(u) \subseteq T$ of image space 'occupied' by the object.

A configuration is simply a finite set of distinct objects $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ where $x_i \in U$, $i = 1, \cdots, n$, $n \geqq 0$. The objects may be in any spatial relation to each other; the number of objects, $n(\boldsymbol{x})$, is variable and may be zero. A configuration $\boldsymbol{x}$ is often associated with its 'silhouette' scene $S(\boldsymbol{x}) = \bigcup_i R(x_i)$ in image space.

The goal is to extract the unobserved underlying pattern $\boldsymbol{x}$ from a given data image $\boldsymbol{y}$.

1.2. *Independent noise models.* The 'true' configuration $\boldsymbol{x}$ gives rise to the observed image $\boldsymbol{y}$ through a known probability distribution with density $f(\boldsymbol{y} \mid \boldsymbol{x})$. This density incorporates both the deterministic influence of $\boldsymbol{x}$ and the stochastic noise inherent in observing $\boldsymbol{y}$. We assume that the values y_t are conditionally independent given $\boldsymbol{x}$ and (without loss of generality) that these conditional distributions belong to a family $\{g(\cdot \mid \theta) : \theta \in \Theta\}$ of probability densities indexed by a parameter space Θ. Hence

$$f(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{t \in T} g(y_t \mid \theta^{(\boldsymbol{x})}(t)).$$

The parameter values $\theta^{(x)}(t)$ form a Θ-valued image called the *signal*, which depends deterministically on $\boldsymbol{x}$. If the signal $\theta^{(x)}(t)$ at site t depends only on whether t belongs to the silhouette $S(\boldsymbol{x})$ or not, the model is called *blur-free*.

1.3. *The Bayesian approach.* A strong motivation for adopting a Bayesian approach in this context is that the MLE tends to contain clusters of almost identical objects, i.e. there is 'multiple response' to each true object. Specifically, in the blur-free case, configurations with equal silhouette cannot be distinguished in likelihood, causing problems when objects occlude each other. Clearly, multiple response is undesirable if it is important to correctly determine the number of objects, or if it is believed that objects do not lie extremely close to one another. For instance in document reading it is known in advance that characters cannot overlap.

In the Bayesian approach, the problem can be solved using a prior model which assigns low probability to configurations in which objects are close to one another. Denoting the prior density by p, the posterior probability for $\boldsymbol{x}$ after observation of data image $\boldsymbol{y}$ is $p(\boldsymbol{x} \mid \boldsymbol{y}) \propto f(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x})$. A *maximum a posteriori* (MAP) estimator of $\boldsymbol{x}$ solves

(1) $$\tilde{\boldsymbol{x}} = \operatorname{argmax}_{\boldsymbol{x}} f(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x}).$$

The prior $p(\boldsymbol{x})$ can be viewed as a penalty assigned to the optimization; because of this interpretation MAP estimation is also known as *penalized maximum likelihood estimation*.

1.4. *Nearest-neighbour Markov object processes.* The basic reference model will be a Poisson process on U with a finite, non-atomic intensity measure μ, $0 < \mu(U) < \infty$. We define a prior model on the set of configurations Ω by its density p with respect to the Poisson process. Note that Ω is hereditary in the sense that $\boldsymbol{x} \in \Omega$ implies $\boldsymbol{y} \in \Omega$ for all subsets $\boldsymbol{y} \subseteq \boldsymbol{x}$ and has probability 1 under the Poisson model.

Natural priors belong to the class of *nearest-neighbour Markov point processes* introduced by Baddeley and Møller [1]. Their essential property is that objects in a given context interact only with their nearest neighbours. Formally, assume that for each $\boldsymbol{x} \in \Omega$ there is a symmetric reflexive relation $\underset{\boldsymbol{x}}{\sim}$ defined on $\boldsymbol{x}$. Related objects are called *$\boldsymbol{x}$-neighbours*; a subset $\boldsymbol{y} \subseteq \boldsymbol{x}$ is called a *clique* in $\boldsymbol{x}$ iff every pair of objects in $\boldsymbol{y}$ are $\boldsymbol{x}$-neighbours.

Definition 1. A nearest-neighbour Markov object process with respect to $\underset{\boldsymbol{x}}{\sim}$ is a stochastic point process whose probability density $p(\cdot)$ on Ω satisfies
(M1) $p(\boldsymbol{x}) > 0$ implies $p(\boldsymbol{y}) > 0$ for all $\boldsymbol{y} \subseteq \boldsymbol{x}$;
(M2) for each $\boldsymbol{x} \in \Omega$ with $p(\boldsymbol{x}) > 0$ and each $u \in U \setminus \boldsymbol{x}$,

$$\frac{p(\boldsymbol{x} \cup \{u\})}{p(\boldsymbol{x})}$$

depends only on u, $N(u \mid \boldsymbol{x} \cup \{u\}) = \left\{x_i \in \boldsymbol{x} : x_i \underset{\boldsymbol{x}\cup\{u\}}{\sim} u\right\}$ and the restrictions of $\underset{\boldsymbol{x}}{\sim}$, $\underset{\boldsymbol{x}\cup\{u\}}{\sim}$ to $N(u \mid \boldsymbol{x} \cup \{u\})$.

Whenever $\underset{\boldsymbol{x}}{\sim}$ does not depend on $\boldsymbol{x}$ the definition is equivalent to that of a Ripley–Kelly process [28]. A useful example of this type is the class of *Markov overlapping object processes* [3], [6]. These are Markov object processes with respect to the relation defined by

$$u \sim u' \Leftrightarrow R(u) \cap R(u') \neq \emptyset.$$

An example where $\underset{\boldsymbol{x}}{\sim}$ does depend on the context is for instance the *Dirichlet object process* for translation models where

$$u \underset{\boldsymbol{x}}{\sim} u' \Leftrightarrow C(u \mid \boldsymbol{x}) \text{ and } C(u' \mid \boldsymbol{x}) \text{ share a common edge.}$$

Here $C(u \mid \boldsymbol{x})$ denotes the Voronoi cell of u in configuration $\boldsymbol{x}$.

An equivalent definition in terms of interactions between objects can be obtained from an analogue of the Hammersley–Clifford theorem ([1], Theorem 4.13).

1.5. *Connection with Hough transform.* As (1.1) is an optimization problem over lists of variable length and additionally the prior is only known up to a normalizing constant, it is generally impossible to compute the MAP estimator analytically. Both the iterative methods in [4] and the alternatives based on spatial birth-and-death processes add and delete objects using log likelihood ratios that can be interpreted as the differences in 'goodness-of-fit' attained by altering the list.

Computing these log likelihood ratios usually involves only pixels 'local' to the altered object. More precisely, for blur-free independent noise models with $g(\cdot \mid \cdot) > 0$

$$L(\boldsymbol{x} \cup \{u\}; \boldsymbol{y}) - L(\boldsymbol{x}; \boldsymbol{y}) = \sum_{t \in R(u) \setminus S(\boldsymbol{x})} h(y_t, \theta_0, \theta_1) \tag{2}$$

where $L(\boldsymbol{x} : \boldsymbol{y}) = \log f(\boldsymbol{y} \mid \boldsymbol{x})$ and

$$h(y_t, \theta_0, \theta_1) = \log \frac{g(y_t \mid \theta_1)}{g(y_t \mid \theta_0)}.$$

The right-hand side of (2) is a generalization of the *Hough transform* [19], [20] used in image processing for detecting simple objects ([14], §4.3).

2. Convergence of inhomogeneous Markov chains

2.1. *Definitions.* Let μ and ν be probability measures on a common measurable space $(\mathcal{S}, \mathcal{A})$. Their *total variation distance* is defined as the maximal difference in mass on measurable subsets $A \in \mathcal{A}$

$$\|\mu - \nu\| = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)|.$$

If $|\mathcal{S}| < \infty$

$$\|\mu - \nu\| = \tfrac{1}{2} \sum_{i \in \mathcal{S}} |\mu(i) - \nu(i)|.$$

Similarly in the continuous case, if both μ and ν are absolutely continuous with respect to some measure m with Radon–Nikodym derivatives f_μ and f_ν,

$$\|\mu - \nu\| = \frac{1}{2} \int_{\mathcal{S}} |f_\mu(s) - f_\nu(s)|\, m(ds).$$

Definition 2. For a transition probability function (stochastic matrix) $P(\cdot, \cdot)$ on $(\mathcal{S}, \mathcal{A})$, *Dobrushin's contraction coefficient* $c(P)$ is defined by

$$c(P) = \sup_{x, y \in \mathcal{S}} \|P(x, \cdot) - P(y, \cdot)\|.$$

We list some properties that are used subsequently (see Dobrushin [8], Section 3).

Lemma 1. *Let* Λ *be the set of all probability measures on* $(\mathcal{S}, \mathcal{A})$. *Then for all transition probability functions* P and Q *and for all* $\mu, \nu \in \Lambda$ *the following hold*:

(i) $c(P) \leqq 1$;

(ii) $\|\mu P - \nu P\| \leqq c(P)\, \|\mu - \nu\|$;

(iii) $c(PQ) \leqq c(P) c(Q)$.

2.2. *Limit theorems.* The main theorem of this section states sufficient conditions under which a sequence of Markov processes converges in total variation to a well-defined limit. The discrete-time case has been studied in [18], [30].

Recall that the transition semi-group $(Q_t)_{t \geqq 0}$ of a Markov process $(Y_t)_{t \geqq 0}$ in continuous time is the semi-group of probability kernels representing its conditional distributions,

$$Q_t(y, F) = \mathbb{P}(Y_t \in F \mid Y_0 = y).$$

Theorem 2. *Let* $(X_t)_{t \geqq 0}$ *be a non-stationary Markov process on a measurable space* $(\mathcal{S}, \mathcal{A})$, *defined by a sequence of transition semi-groups* $(Q_n)_{n \in \mathbb{N}}$. *The process follows the transition rules* Q_n *in the time period* $[t_n, t_{n+1})$, *that is*

$$\mathbb{P}(X_s \in F \mid X_r = y) = (Q_n)_{s-r}(y, F)$$

for $t_n \leqq r < s < t_{n+1}$. *Here* $t_n \nearrow \infty$ *as* $n \to \infty$. *Assume that for each* $n \in \mathbb{N}$, Q_n *has an invariant measure* μ_n, *i.e.*

$$\int_{\mathcal{S}} (Q_n)_t(x, F)\mu_n(dx) = \mu_n(F)$$

for all $F \in \mathcal{A}$ *and* $t \geqq 0$. *Assume moreover that the following hold*:

$$\text{(C)} \qquad \sum_{n=1}^{\infty} \|\mu_n - \mu_{n+1}\| < \infty$$

$$\text{(D)} \qquad c(P_{tt'}) \to 0 \quad \text{as } t' \to \infty \text{ for all } t \geqq 0$$

where $P_{tt'}(x, F) = \mathbb{P}(X_{t'} \in F \mid X_t = x)$. *Then* $\mu_\infty = \lim \mu_n$ *exists and* $\nu P_{0t} \to \mu_\infty$ *in total variation as* $t \to \infty$, *uniformly in the initial distribution* ν.

Proof. Condition (C) implies that (μ_n) is a Cauchy sequence in $\|\cdot\|$ and hence converges in total variation to μ_∞, say.

Define $n(t) = \sup\{n : t_n \leqq t\}$ and choose $0 \leqq t < t_{n(t)+1} < t' < \infty$. Then

$$\begin{aligned} \mu_\infty P_{tt'} - \mu_\infty &= (\mu_\infty - \mu_{n(t)})P_{tt'} + \mu_{n(t)}P_{tt'} - \mu_\infty \\ &= (\mu_\infty - \mu_{n(t)})P_{tt'} + \mu_{n(t)}P_{tt_{n(t)+1}}P_{t_{n(t)+1}t'} - \mu_\infty. \end{aligned}$$

Since $\mu_{n(t)}$ is an invariant measure one sees that

$$\begin{aligned} \mu_\infty P_{tt'} - \mu_\infty &= (\mu_\infty - \mu_{n(t)})P_{tt'} + \mu_{n(t)}P_{t_{n(t)+1}t'} - \mu_\infty \\ &= (\mu_\infty - \mu_{n(t)})P_{tt'} + \sum_{k=n(t)}^{n(t')-1} (\mu_k - \mu_{k+1})P_{t_{k+1}t'} + \mu_{n(t')} - \mu_\infty. \end{aligned}$$

Hence

$$\begin{aligned} \|\mu_\infty P_{tt'} - \mu_\infty\| &\leqq \|\mu_\infty - \mu_{n(t)}\| \, c(P_{tt'}) \\ &\quad + \sum_{k=n(t)}^{n(t')-1} \|\mu_k - \mu_{k+1}\| \, c(P_{t_{k+1}t'}) + \|\mu_{n(t')} - \mu_\infty\| \\ &\leqq 2 \sup_{k \geqq n(t)} \|\mu_k - \mu_\infty\| + \sum_{k=n(t)}^{\infty} \|\mu_k - \mu_{k+1}\| \\ &\to 0 (t \to \infty). \end{aligned}$$

Let $\epsilon > 0$. Choose t such that $\|\mu_\infty P_{tt'} - \mu_\infty\| < \epsilon/2$ for all $t' > t_{n(t)+1}$. Next observe that

$$\begin{aligned} \|\nu P_{0t'} - \mu_\infty\| &= \|(\nu P_{0t} - \mu_\infty)P_{tt'} + \mu_\infty P_{tt'} - \mu_\infty\| \\ &\leqq \|\nu P_{0t} - \mu_\infty\| \, c(P_{tt'}) + \|\mu_\infty P_{tt'} - \mu_\infty\| \\ &\leqq c(P_{tt'}) + \|\mu_\infty P_{tt'} - \mu_\infty\|. \end{aligned}$$

Use condition (D) to choose t' such that $c(P_{tt'}) < \epsilon/4$. Summarizing, we obtain

$$\|\nu P_{0t} - \mu_\infty\| \to 0 \text{ uniformly in } \nu (t \to \infty).$$

A sufficient condition for (D) is given by the next result. It is easier to work with, since only stationary Markov chains have to be considered.

Lemma 3. *Use the same notation as in the previous theorem. If* $c(P_{t_nt_{n+1}}) \leqq 1 - 1/n$ *for all* $n \geqq 2$, *the Dobrushin condition* (*D*) *holds.*

Proof. Write $P_n = P_{t_nt_{n+1}}$. Then

$$-\log c(P_n) \geqq 1 - c(P_n) \geqq \frac{1}{n}.$$

Thus

$$-\sum_{n=2}^{\infty} \log c(P_n) \geqq \sum_{n=2}^{\infty} \frac{1}{n} = \infty$$

or equivalently

$$\prod_{n=2}^{\infty} c(P_n) \to 0 \qquad (N \to \infty).$$

Fix t. Then for $t' > t_{n(t)+2}$

$$\begin{aligned} c(P_{tt'}) &= c(P_{tt_{n(t)+1}} P_{t_{n(t)+1}t_{n(t)+2}} \cdots P_{t_{n(t')}t'}) \\ &\leqq c(P_{tt_{n(t)+1}}) \left[\prod_{i=n(t)+1}^{n(t')-1} c(P_i) \right] c(P_{t_{n(t')}t'}) \\ &\leqq \prod_{i=n(t)+1}^{n(t')-1} c(P_i) \to 0 (t' \to \infty). \end{aligned}$$

3. Fixed temperature sampling

3.1. *Spatial birth-and-death process.* In the context described in Section 1, the natural analogue of the Gibbs sampler [9] is a *spatial birth-and-death process* [26], [24]. This is a continuous-time, pure jump Markov process, whose states are configurations $\boldsymbol{x} \in \Omega$, and for which the only transitions are the birth of a new object (instantaneous transition from $\boldsymbol{x}$ to $\boldsymbol{x} \cup \{u\}$) or the death of an existing one (transition from $\boldsymbol{x}$ to $\boldsymbol{x} \setminus \{x_i\}$). Formally, write $\mathscr{B}$ for the (Borel) σ-algebra on U and let $D(\cdot, \cdot): \Omega \times U \to [0, \infty)$ be a measurable function and $B(\cdot, \cdot): \Omega \times \mathscr{B} \to [0, \infty)$ a finite kernel, i.e. $B(\boldsymbol{x}, \cdot)$ is a finite measure on $(U, \mathscr{B})$ and $B(\cdot, F)$ is a measurable function on Ω. These are called the *death rate* and *birth rate.* The reason is clear from the following. Given the state $\boldsymbol{x}$ at time t,

- the probability of a death $\boldsymbol{x} \to \boldsymbol{x} \setminus \{x_i\}$ during a time interval $(t, t+h)$, $h \to 0$, is $D(\boldsymbol{x} \setminus \{x_i\}, x_i)h + o(h)$;
- the probability of a birth $\boldsymbol{x} \to \boldsymbol{x} \cup \{u\}$ during time $(t, t+h)$, where u lies in a given measurable subset $F \subseteq U$, is $B(\boldsymbol{x}, F)h + o(h)$;
- the probability of more than one transition during $(t, t+h)$ is $o(h)$.

We will assume that $B(\boldsymbol{x}, \cdot)$ has a density $b(\boldsymbol{x}, \cdot)$ with respect to μ on U, so that intuitively $b(\boldsymbol{x}, u)$ is the transition rate for a birth $\boldsymbol{x} \to \boldsymbol{x} \cup \{u\}$. Write

$$B(\boldsymbol{x}) = \int_U b(\boldsymbol{x}, u)\, d\mu(u)$$

for the total birth rate, and similarly define

$$D(\boldsymbol{x}) = \sum_{x_i \in \boldsymbol{x}} D(\boldsymbol{x}\backslash\{x_i\}, x_i).$$

To avoid explosion, an infinite number of transitions occurring in finite time, the rates have to satisfy certain regularity assumptions. Preston ([26], Proposition 5.1, Theorem 7.1) gave sufficient conditions under which there exists a unique spatial birth-and-death process with given rates solving Kolmogorov's backward equations

$$\frac{d}{dt}\mathbb{P}(X_t \in F \mid X_0 = x) = -[B(x) + D(x)]\mathbb{P}(X_t \in F \mid X_0 = \boldsymbol{x}) + \int \mathbb{P}(X_t \in F \mid X_0 = z) R(\boldsymbol{x}, dz)$$

with

$$R(\boldsymbol{x}, F) = B(\boldsymbol{x}, \{u \in U : \boldsymbol{x} \cup \{u\} \in F\}) + \sum_{x_i \in \boldsymbol{x}} 1\{\boldsymbol{x}\backslash\{x_i\} \in F\} D(\boldsymbol{x}\backslash\{x_i\}, x_i)$$

the total rate from pattern $\boldsymbol{x}$ into F. For a given process (X_t) he also found conditions for the existence of a unique invariant probability measure and convergence in distribution (i.e. convergence of $\mathbb{P}(X_t \in F \mid X_0 = \boldsymbol{x})$).

Theorem 4. *For each* $n = 0, 1, \cdots$ *define* $\kappa_n = \sup_{n(\boldsymbol{x})=n} B(\boldsymbol{x})$ *and* $\delta_n = \inf_{n(\boldsymbol{x})=n} D(\boldsymbol{x})$. *Assume* $\delta_n > 0$ *for all* $n \geqq 1$. *If either* (*a*) $\kappa_n = 0$ *for all sufficiently large* $n \geqq 0$, *or* (*b*) $\kappa_n > 0$ *for all* $n \geqq 0$ *and both the following hold*:

$$\sum_{n=1}^{\infty} \frac{\kappa_0 \cdots \kappa_{n-1}}{\delta_1 \cdots \delta_n} < \infty$$

$$\sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_n}{\kappa_1 \cdots \kappa_n} = \infty$$

then there exists a unique spatial birth-and-death process for which $B(\cdot)$ *and* $D(\cdot)$ *are the transition rates*; *this process has a unique equilibrium distribution to which it converges in distribution from any initial state.*

A slightly stronger result given by Møller [24] includes the case $\kappa_0 = 0$, $\kappa_n > 0$ for all $n \geqq 1$ and both

$$\sum_{n=2}^{\infty} \frac{\kappa_1 \cdots \kappa_{n-1}}{\delta_1 \cdots \delta_n} < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_n}{\kappa_1 \cdots \kappa_n} = \infty,$$

still assuming all δ_n positive for $n \geqq 1$.

3.2. *Construction.* Suppose we want to sample from the *temperature modified posterior distribution*

$$p_H(\boldsymbol{x} \mid \boldsymbol{y}) \propto \{f(\boldsymbol{y} \mid \boldsymbol{x}) p(\boldsymbol{x})\}^{1/H}. \tag{3}$$

The purpose of introducing a temperature parameter is to sharpen peaks in posterior probability. For small $H > 0$, configurations with large posterior density are favoured, while others are suppressed. Indeed, if object space U is discretized, $p_H(\cdot \mid y)$ converges pointwise to a uniform distribution on the set of MAP solutions as H tends to zero.

Consider any blur-free independent noise model (Section 1.3) with $g(\cdot \mid \cdot) > 0$ and a nearest-neighbour Markov object prior. The former assumption is needed so that the class $K = \{\boldsymbol{x} : f(y \mid \boldsymbol{x})p(\boldsymbol{x}) > 0\}$ is hereditary. For some fixed $k \in [0, 1]$ set

$$(4) \qquad b_H(\boldsymbol{x}, u) = \begin{cases} \left(\dfrac{f(y \mid \boldsymbol{x} \cup \{u\})p(\boldsymbol{x} \cup \{u\})}{f(y \mid \boldsymbol{x})p(\boldsymbol{x})}\right)^{k/H} & \text{if } f(y \mid \boldsymbol{x})p(\boldsymbol{x}) > 0 \\ 0 & \text{if } f(y \mid \boldsymbol{x})p(\boldsymbol{x}) = 0 \end{cases}$$

for $u \notin \boldsymbol{x}$ and death rate

$$(5) \quad D_H(\boldsymbol{x}\backslash\{u\}, u) = \begin{cases} \left(\dfrac{f(y \mid \boldsymbol{x})p(\boldsymbol{x})}{f(y \mid \boldsymbol{x}\backslash\{u\})p(\boldsymbol{x}\backslash\{u\})}\right)^{(k-1)/H} & \text{if } f(y \mid \boldsymbol{x})p(\boldsymbol{x}) > 0 \\ \delta'_n/n & \text{if } f(y \mid \boldsymbol{x})p(\boldsymbol{x}) = 0, \quad n(\boldsymbol{x}) = n. \end{cases}$$

Here $\delta'_n = \inf\{\sum_{x_i \in \boldsymbol{x}} D_H(\boldsymbol{x}\backslash\{x_i\}, x_i) \mid f(y \mid \boldsymbol{x})p(\boldsymbol{x}) > 0,\ n(\boldsymbol{x}) = n\}$. By convention, the infimum of the empty set equals ∞. Note that by this definition $\delta'_n = \delta_n$, where δ_n is defined as in Theorem 4. The boundary cases $k = 0$ ('constant birth rate') and $k = 1$ ('constant death rate') are well known in spatial statistics, to obtain realizations of a point process. It is widely argued (e.g. [27]) that the constant death rate procedure should be preferred, as under the constant birth rate process there is a high probability that a newly added object will have a large death rate and thus be rapidly deleted again.

For a nearest-neighbour Markov prior the above expressions are typically easy to evaluate, since the normalizing constant is eliminated. Moreover the 'detailed balance' equations

$$(6) \qquad b_H(\boldsymbol{x}, u)p_H(\boldsymbol{x} \mid y) = D_H(\boldsymbol{x}, u)p_H(\boldsymbol{x} \cup \{u\} \mid y)$$

are satisfied whenever $f(y \mid \boldsymbol{x} \cup \{u\})p(\boldsymbol{x} \cup \{u\}) > 0$. Given a spatial birth-and-death process with rates satisfying (6), Ripley [27] remarked that the process is necessarily time reversible and $p_H(\cdot \mid y)$ is the density of its unique invariant probability measure. For each application, however, one should verify that the process just described is well-defined. For instance, the following corollary of Theorem 4 holds.

Corollary 5. *Let* y *and* $H > 0$ *be fixed. For any blur-free independent noise model with* $g(\cdot \mid \cdot) > 0$, *and any nearest-neighbour Markov object process* $p(\cdot)$ *with uniformly bounded likelihood ratios*

$$\frac{p(\boldsymbol{x} \cup \{u\})}{p(\boldsymbol{x})} \leqq \beta < \infty$$

there exists a unique spatial birth-and-death process for which (4) *and* (5) *are the transition rates. The process has unique equilibrium distribution* $p_H(\cdot \mid \boldsymbol{y})$ *and it converges in distribution to* $p_H(\cdot \mid \boldsymbol{y})$ *from any initial state.*

Proof. We will prove the following properties:

1. $\delta_n > 0$, for $n \geqq 1$;
2. If $\kappa_{n_0} = 0$ for some $n_0 \geqq 1$, then $\kappa_n = 0 \; \forall n \geqq n_0$;
3. if $\kappa_n > 0$ for all n, then condition (b) of Theorem 4 holds.

Property 1. Use the representation of the log likelihood ratio as a generalized Hough transform (1.2). Since T is finite, we have upper and lower bounds on the goodness of fit, say $|h(y_t, \theta_0, \theta_1)| \leqq a$ for all t. Hence

$$|L(\boldsymbol{x} \cup \{u\}; y) - L(\boldsymbol{x}; y)| \leqq an(R(u)) \leqq an(T)$$

where n denotes the number of pixels. For $p(\cdot)$ we have by assumption $(p(\boldsymbol{x} \cup \{u\}))/p(\boldsymbol{x}) \leqq \beta$. If $p(\boldsymbol{x}) > 0$ this implies that

$$\begin{aligned} D_H(\boldsymbol{x} \backslash \{u\}, u) &= \left(\frac{f(y \mid \boldsymbol{x})}{f(y \mid \boldsymbol{x} \backslash \{u\})} \frac{p(\boldsymbol{x})}{p(\boldsymbol{x} \backslash \{u\})} \right)^{(k-1)/H} \\ &\geqq \exp\left[\frac{k-1}{H} (|L(\boldsymbol{x} \backslash \{u\}; y) - L(\boldsymbol{x}; y)| + \log \beta) \right] \\ &\geqq \exp\left[\frac{k-1}{H} (an(T) + \log \beta) \right] \\ &=: \delta > 0. \end{aligned}$$

Suppose $p(\boldsymbol{x}) = 0$. If $p(z) = 0$ for all z with $n(z) = n(\boldsymbol{x})$, then $D_H(\boldsymbol{x} \backslash \{u\}, u) = \infty \geqq \delta$. Otherwise $n(\boldsymbol{x}) D_H(\boldsymbol{x} \backslash \{u\}, u) = \inf \{D_H(z) \mid n(z) = n(\boldsymbol{x}),\ p(z) > 0\}$. By the above argument $D_H(z \backslash \{z_i\}, z_i) \geqq \delta$ for all such z and $z_i \in z$. Hence $D_H(z) \geqq \delta n(\boldsymbol{x})$ and $D_H(\boldsymbol{x} \backslash \{u\}, u) \geqq \delta$. Therefore $D_H(\boldsymbol{x}) \geqq \delta n(\boldsymbol{x})$ for all patterns $\boldsymbol{x}$, and hence $\delta_n \geqq \delta n > 0$ for $n \geqq 1$.

Property 2. Use the fact that $K = \{\boldsymbol{x} : f(y \mid \boldsymbol{x}) p(\boldsymbol{x}) > 0\}$ is hereditary.

Property 3. The birth rates are also bounded. For $p(\boldsymbol{x}) > 0$

$$\begin{aligned} b_H(\boldsymbol{x}, u) &= \left(\frac{f(y \mid \boldsymbol{x} \cup \{u\})}{f(y \mid \boldsymbol{x})} \frac{p(\boldsymbol{x} \cup \{u\})}{p(\boldsymbol{x})} \right)^{k/H} \\ &\leqq \exp\left[\frac{k}{H} (an(T) + \log \beta) \right] \\ &=: \kappa > 0. \end{aligned}$$

For $p(\boldsymbol{x}) = 0$, again $b_H(\boldsymbol{x}, u) = 0 \leqq \kappa$. Hence $B_H(\boldsymbol{x}) \leqq \kappa \mu(U)$ and so $\kappa_n \leqq \kappa \mu(U)$.

Using these bounds, one obtains

$$\frac{\kappa_0 \cdots \kappa_{n-1}}{\delta_1 \cdots \delta_n} \leqq \frac{\kappa^n \mu(U)^n}{n!\, \delta^n}.$$

Since $\mu(U)$ is finite by assumption, the first assertion follows. Similarly

$$\frac{\delta_1 \cdots \delta_n}{\kappa_1 \cdots \kappa_n} \geqq \frac{n!\, \delta^n}{\kappa^n \mu(U)^n}$$

which does not converge to zero as $n \to \infty$. The corollary is proved if we combine these properties with Theorem 4.

If the state space is digitized, the situation is easier. Recall that any Markov chain on a finite state space is uniquely defined by its transition rates and if for an irreducible Markov chain with rates $R(i, j)$, $i \neq j$, π is a probability measure satisfying the detailed balance equations

$$R(i, j)\pi(i) = R(j, i)\pi(j) \tag{7}$$

then the chain is time reversible and has unique limit distribution π (see for instance [25], pp. 277–278). Hence, if $f(\boldsymbol{y} \mid \boldsymbol{x}) > 0$ for all $\boldsymbol{x}$ the class K is hereditary due to the Markov property (M1) of $p(\cdot)$. It follows that the birth-and-death process defined above restricted to K is irreducible.

3.3. *Simulation and practical considerations.* To simulate the spatial birth-and-death process we generate the successive states $X^{(k)}$ and the sojourn times $T^{(k)}$ as follows. Given $X^{(k)} = \boldsymbol{x}^{(k)}$, $T^{(k)}$ is exponentially distributed with mean $1/(D_H(\boldsymbol{x}^{(k)}) + B_H(\boldsymbol{x}^{(k)}))$, independent of other sojourn times and of past states. The next state transition is a death with probability $D_H(\boldsymbol{x}^{(k)})/(D_H(\boldsymbol{x}^{(k)}) + B_H(\boldsymbol{x}^{(k)}))$, obtained by deleting one of the existing points x_i with probability

$$\frac{D_H(\boldsymbol{x}^{(k)} \setminus \{x_i\}, x_i)}{D_H(\boldsymbol{x}^{(k)})}.$$

Otherwise the transition is a birth generated by choosing one of the points $u \notin \boldsymbol{x}^{(k)}$ with probability density

$$\frac{b_H(\boldsymbol{x}^{(k)}, u)}{B_H(\boldsymbol{x}^{(k)})}$$

with respect to μ and adding u to the state. The process described above is run for a 'large' time period C to obtain a realization $\boldsymbol{x}^{(L)}$ where

$$L = \min\left\{k = 0, 1, 2, \cdots \;\middle|\; \sum_{i=0}^{k} t^{(i)} > C\right\}.$$

The computational effort per transition is mainly in sampling from $b_H(\boldsymbol{x}, \cdot)/B(\boldsymbol{x})$. Since the birth rate $b_H(\boldsymbol{x}, u)$ is an exponential function of the Hough transform

(2), it tends to have sharp peaks as a function of u when H is small or when $\boldsymbol{x}$ is far from an MAP solution. There is then a high probability that the next transition will add a new object u at one of the locations. This suggests incorporating a search operation. When the dimension of U is large, multiresolution strategies can be used. For more details, see [6].

The main advantage of sampling techniques compared to deterministic methods [4] is the ability to estimate any functional of the (modified) posterior distribution by taking a sufficient number of independent realizations. Examples of useful functionals are: the distribution (mean, variance) of the number of objects; the probability that there is no object in a given subregion of the image and the first-order intensity [29]. In the discrete case the first-order intensity at u is simply the (posterior) probability that u belongs to $\boldsymbol{x}$; it can be regarded as an alternative to the Hough transform.

4. Stochastic annealing for point processes

Instead of making probability statements (probabilities, expectations, functionals) one is sometimes interested in the modes of the posterior distribution. Here we present a technique for solving the MAP equations (1), using the results of Section 2. Assuming the conditions of Lemma 5, for each fixed H we can construct a spatial birth-and-death process with equilibrium distribution $p_H(\cdot \mid \boldsymbol{y})$. Our proposal therefore is to use a stochastic annealing algorithm that simulates these processes consecutively with H gradually dropping to zero.

In the superficially similar context of image segmentation, a simulated annealing algorithm was developed by Geman and Geman [9]. However, the Markov processes involved are rather different. Since in segmentation problems both object and image space are finite pixel grids, a discrete-time Markov chain changing each pixel label in turn suffices.

4.1. *The summability condition.* First we consider the summability condition (C).

Lemma 6. *Let $H_n \searrow 0$ $(n \to \infty)$ and consider the sequence of H_n-modified posterior distributions with densities*

$$p_{H_n}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \{f(\boldsymbol{y} \mid \boldsymbol{x}) p(\boldsymbol{x})\}^{1/H_n}$$

with respect to the reference measure m on Ω (counting measure or law of Poisson process). Assume that $m(\mathcal{M}) > 0$, where $\mathcal{M}$ denotes the set of solutions to the MAP equations (1.1). *Then the sequence p_{H_n} converges in total variation to a uniform distribution on $\mathcal{M}$. Moreover the sequence satisfies condition (C).*

Proof. Since $m(\mathcal{M}) > 0$, $\exists \boldsymbol{x}^{\#}$ attaining the maximum. Denote

$$l_n(\boldsymbol{x}) = \left(\frac{f(\boldsymbol{y} \mid \boldsymbol{x}) p(\boldsymbol{x})}{f(\boldsymbol{y} \mid \boldsymbol{x}^{\#}) p(\boldsymbol{x}^{\#})}\right)^{1/H_n}, \qquad Z_n = \int_\Omega l_n(\boldsymbol{x}) dm(\boldsymbol{x}).$$

It is easily seen that $l_n(x) \to 1\{x \in \mathcal{M}\}$, $(n \to \infty)$, and

$$\lim_{n\to\infty} Z_n = \int_\Omega \lim_{n\to\infty} l_n(x)dm(x) = \int_\Omega 1\{x \in \mathcal{M}\}dm(x) = m(\mathcal{M})$$

by the dominated convergence theorem. Moreover, $Z_n \downarrow m(\mathcal{M})$, $\sup Z_n = Z_1 \leqq m(\Omega) < \infty$.

To prove condition (C), suppress the dependence on x and consider

$$\begin{aligned}\left|\frac{l_n}{Z_n} - \frac{l_{n+1}}{Z_{n+1}}\right| &= (Z_n Z_{n+1})^{-1} |l_n Z_{n+1} - l_{n+1} Z_n| \\ &\leqq (Z_n Z_{n+1})^{-1}\{l_{n+1} \mid Z_{n+1} - Z_n| + Z_{n+1} \mid l_{n+1} - l_n|\} \\ &\leqq \frac{1}{m(\mathcal{M})^2}\{Z_{n+1}(l_n - l_{n+1}) + l_{n+1}(Z_n - Z_{n+1})\} \\ &\leqq \frac{m(\Omega)}{m(\mathcal{M})^2}\{l_n - l_{n+1} + Z_n - Z_{n+1}\}.\end{aligned}$$

Therefore

$$\begin{aligned}&\sum_{n=1}^{N-1} \int_\Omega \left|\frac{l_n(x)}{Z_n} - \frac{l_{n+1}(x)}{Z_{n+1}}\right| dm(x) \\ &\leqq \sum_{n=1}^{N-1} \frac{m(\Omega)}{m(\mathcal{M})^2} \int_\Omega \{l_n(x) - l_{n+1}(x) + Z_n - Z_{n+1}\}\, dm(x) \\ &= \frac{m(\Omega)}{m(\mathcal{M})^2} \int_\Omega \sum_{n=1}^{N-1} \{l_n(x) - l_{n+1}(x) + Z_n - Z_{n+1}\}\, dm(x) \\ &= \frac{m(\Omega)}{m(\mathcal{M})^2} \int_\Omega \{l_1(x) - l_N(x) + Z_1 - Z_N\}\, dm(x) \\ &= \frac{m(\Omega)}{m(\mathcal{M})^2} (1 + m(\Omega))(Z_1 - Z_N).\end{aligned}$$

Letting $N \to \infty$ the right-hand side converges to

$$\frac{m(\Omega)}{m(\mathcal{M})^2} (1 + m(\Omega))(Z_1 - m(\mathcal{M})) < \infty.$$

A more restricted version, $\pi_n \propto \exp[-f/H_n]$ for a bounded, measurable function f was proved in Theorem 3.3a of [18]. The assumption $m(\mathcal{M}) > 0$ is needed; if $m(\mathcal{M}) = 0$ the sequence of modified posterior distributions will not converge in total variation (cf. [18]).

4.2. *The Dobrushin condition.* From now on let f be a blur-free independent noise model with $g(\cdot \mid \cdot) > 0$. Again, let $H_n \searrow 0$ and consider the family $(X^{(n)})_{n\in\mathbb{N}}$ of spatial birth-and-death processes on $K = \{x : f(y \mid x)p(x) > 0\}$ defined by (4) and (5). As K is closed and irreducible, the processes are well-defined and converge to the unique limit $p_H(\cdot \mid y)$ (see Section 3).

Recall the following result by Møller [24], a generalization of earlier work by Lotwick and Silverman [22].

Theorem 7. Let X_t be a spatial birth-and-death process and define κ_n, δ_n as in Theorem 4. Assume moreover that $\delta_n > 0$ for all $n \geqq 1$ and $\kappa_n = 0$ for all $n > n_0$ (condition (a)). Then for all fixed $t_0 > 0$

$$\sup_{x,y} \|P_t(x, \cdot) - P_t(y, \cdot)\| \leqq 2(1 - K(t_0))^{(t/t_0)-1}$$

for all $t > t_0$. The supremum is taken over all initial states x, y containing at most n_0 objects.

Here

$$\tilde{\delta}_m = \min_{i,j:i+j=m} \delta_i + \delta_j; \qquad \tilde{\kappa}_m = \max_{i,j:i+j=m} \kappa_i + \kappa_j; \qquad \tilde{\alpha}_m = \tilde{\delta}_m + \tilde{\kappa}_m;$$

$$K_1(j, t) = [1 - \exp(-\tilde{\alpha}_j t)] \frac{\tilde{\delta}_j}{\tilde{\alpha}_j}; \qquad K_0(n, t_0) = \prod_{j=1}^{n} K_1\left(j, \frac{t_0}{n}\right); \tag{8}$$

$$K(t_0) = \min_{n \leqq n_0} K_0(n, t_0).$$

Therefore, by Lemma 3 we can construct an annealing schedule satisfying condition (D) by requiring

$$2(1 - K(t_0))^{(t/t_0)-1} \leqq 1 - \frac{1}{n}$$

or equivalently

$$t \geqq t_0\left(1 + \frac{\log\left(\frac{1}{2}(1 - (1/n))\right)}{\log(1 - K(t_0))}\right). \tag{9}$$

Under the assumption in Lemma 6, condition (C) also holds and by Theorem 2 the sequence of birth-and-death processes constructed this way converges in total variation to a uniform distribution on the set of global maxima of the posterior distribution, regardless of the initial state.

4.3. *Extensions.* Generalizations to diffusing objects are possible. In the finite cae $|U| < \infty$, write

$$M(x, x_i, u)$$

for the configuration obtained from x by replacing x_i by u. The set of $u \in U$ for which this operation is allowed is denoted by $Q(x, x_i)$. Typically it consists of unoccupied objects close but not identical to x_i. Suppose the diffusion rates are also powers of the log likelihood ratios

$$c_H(x, x_i, u) = \left\{\frac{f(y \mid M(x, x_i, u))p(M(x, x_i, u))}{f(y \mid x)p(x)}\right\}^{k/H}. \tag{10}$$

If detailed balance is required to hold as well, necessarily $k = \frac{1}{2}$.

The Markov chain on K with rates given by (4), (5) and (10) is clearly irreducible, time reversible and has unique equilibrium distribution $p_H(\cdot \mid \boldsymbol{y})$. Writing $C_H(\boldsymbol{x})$ for the total diffusion rate from state $\boldsymbol{x}$;

$$\gamma_i = \max_{n(\boldsymbol{x})=i} C_H(\boldsymbol{x}); \qquad \tilde{\gamma}_m = \max_{i,j:i+j=m} \gamma_i + \gamma_j; \qquad \tilde{\alpha}_m = \tilde{\delta}_m + \tilde{\kappa}_m + \tilde{\gamma}_m$$

and K_1, K_0, K as in (8), for all fixed $t_0 > 0$ and $t > t_0$

$$\max_{\boldsymbol{x},\boldsymbol{y}} \|P_t(\boldsymbol{x}, \cdot) - P_t(\boldsymbol{y}, \cdot)\| \leqq 2(1 - K(t_0))^{(t/t_0)-1}.$$

This can be proved by coupling arguments as in [22], pp. 157–158, using that

$$\mathbb{P}(\text{next event in } Z \text{ occurs before time } t \text{ and is a death} \mid Z_0) \geqq K_1(m+n, t).$$

In the continuous case, n objects can perform a diffusion on U^n

$$d\boldsymbol{x}_t = \nabla \log p_H(\boldsymbol{x} \mid \boldsymbol{y})\, dt + \sqrt{2H}\, dB_t$$

at least if $p_H(\cdot \mid \boldsymbol{y})$ is strictly positive and infinitely differentiable ([27], p. 178). See also [10], [23].

5. Example

Baddeley and Van Lieshout [4] studied a simple example where a scene composed of discs with fixed radius was observed after addition of white Gaussian noise (Figure 1). In the Bayesian approach a Strauss process served as prior distribution on disc configurations. To enable comparison, the same parameter values are used below.

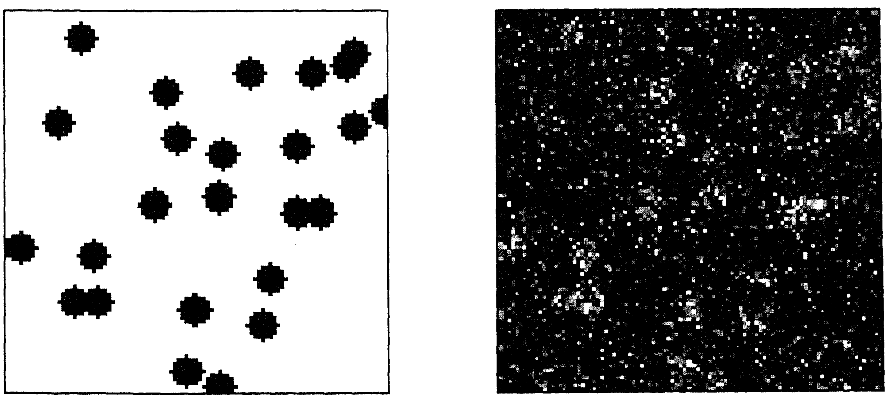

Figure 1. Binary silhouette scene of discs with radius 4 and realization from Gaussian model with $\sigma = 50$, $\theta_1 = 150$ and $\theta_0 = 100$ digitized on a 98×98 square grid

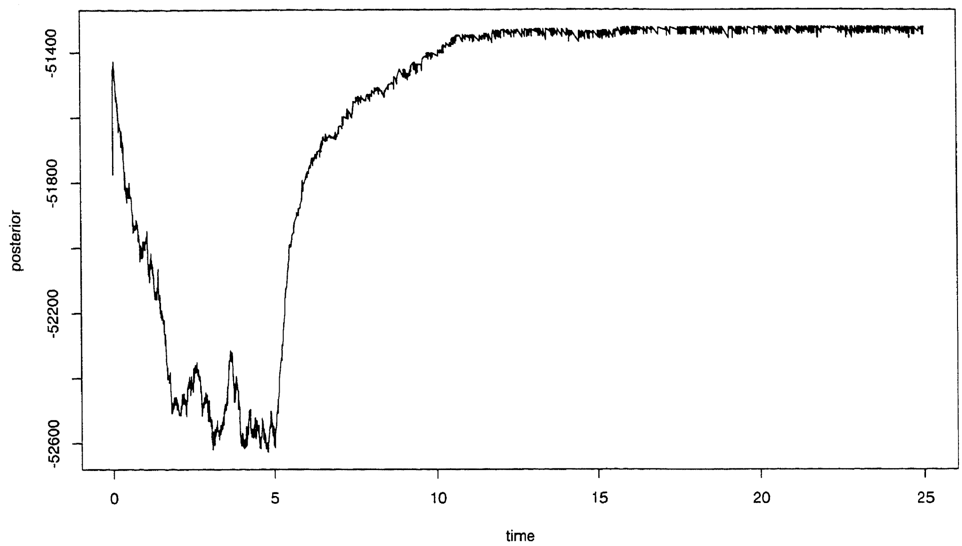

Figure 2. Log posterior likelihood against time for a geometric, cooling schedule starting at $H = 4{\cdot}0$ of rate 0·5 and a Strauss prior with $\beta = 0{\cdot}0025$ and $\gamma = 0{\cdot}25$.

In practice, the theoretical temperature schedule (9) is too slow and one resorts to 'feasible' schemes. Here we chose a geometric cooling of rate 1/2 and initial temperature $H = 4{\cdot}0$. The log posterior likelihood as a function of time is given in Figure 2; Figure 3 graphs the number of objects against time. Finally a typical

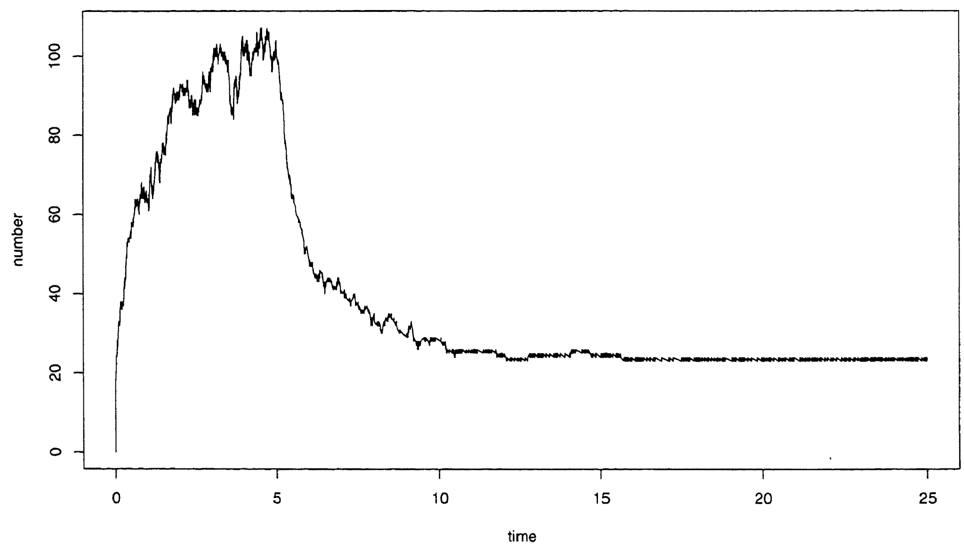

Figure 3. Number of objects against time for a geometric cooling schedule starting at $H = 4{\cdot}0$ of rate 0·5 and a Strauss prior with $\beta = 0{\cdot}0025$ and $\gamma = 0{\cdot}25$

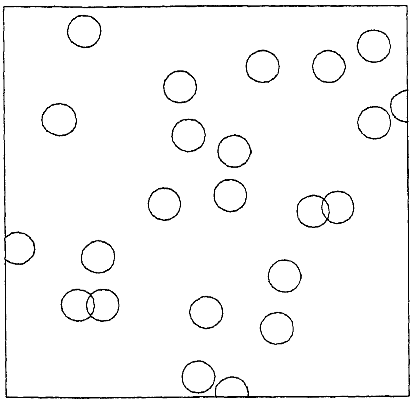

Figure 4. Sample taken at time 25 for a geometric cooling schedule starting at $H = 4{\cdot}0$ of rate $0{\cdot}5$ and a Strauss prior with $\beta = 0{\cdot}0025$ and $\gamma = 0{\cdot}25$

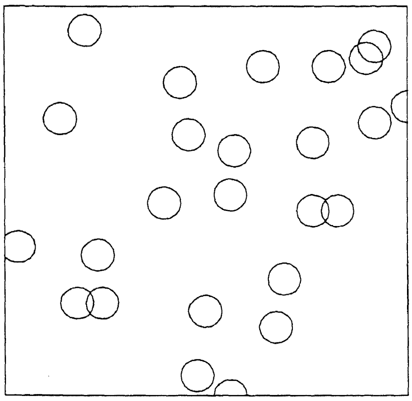

Figure 5. Sample from the posterior distribution using a Strauss prior with $\beta = 0{\cdot}0025$ and $\gamma = 0{\cdot}25$

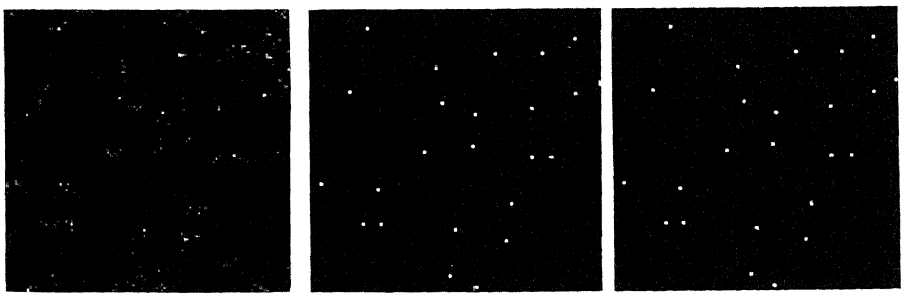

Figure 6. Posterior intensity surfaces at temperature 4·0, 1·0 and 0·25 (from left to right).

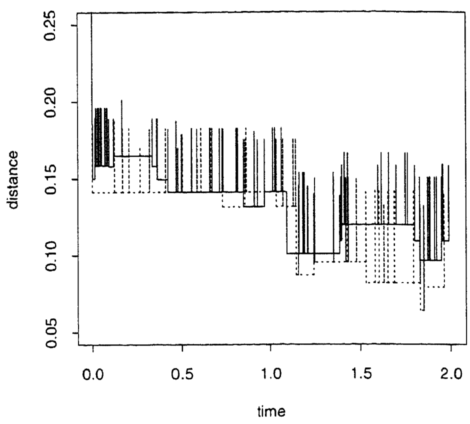

Figure 7. Δ_2 distance between reconstructed and true pattern as a function of time. The cutoff value is 4

reconstruction sampled at time unit 25 and $H = 0{\cdot}25$ is given in Figure 4. Throughout, the constant death rate method was used.

In contrast to ICM, stochastic annealing results in a global maximum, regardless of the initial state. Experiments with several initial states are in accordance with the theory in that similar reconstructions were obtained. One has to be careful though, since too fast a cooling schedule was used. For a discussion on the implications of such *ad hoc* choices, see [13].

A good compromise between ICM and stochastic annealing is to sample at a fixed 'low' temperature (see also [12]). A reconstruction obtained by running the constant death rate procedure at $H = 1$ is given in Figure 5. Using constant birth rate instead was found to behave worse. The latter method tends to add an unlikely object and immediately delete it again. This confirms experience reported in the literature [24], [27].

Estimates of the posterior intensity surface (Figure 6) suggest that the posterior distribution is rather peaked and can be used as an approximation to MAP estimation, apart from being interesting in its own right.

Typical runs of the constant death rate method are illustrated in Figure 7, where the Δ_2 distance [2] to the 'true' pattern is graphed against time. Starting from an empty scene, objects are immediately added to form a plausible reconstruction followed by deletion and immediate reading of one of the objects. Note that the results obtained this way are comparable to steepest ascent reconstructions [4].

Acknowledgements

I am very grateful to my research supervisor Professor A. J. Baddeley for his support and suggestions, to A. G. Steenbeek for technical assistance and to Professor R. D. Gill and Dr J. Møller for helpful feedback.

References

[1] BADDELEY, A. AND MØLLER, J. (1989) Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.* **57**, 89–121.

[2] BADDELEY, A. J. (1992) An error metric for binary images. In *Robust Computer Vision*, ed. W. Forstner and S. Ruwiedel pp. 59–78. Wichmann, Karlsruhe.

[3] BADDELEY, A. J. AND VAN LIESHOUT, M. N. M. (1991) Recognition of overlapping objects using Markov spatial models. CWI Report BS-R9109, March 1991.

[4] BADDELEY, A. J. AND VAN LIESHOUT, M. N. M. (1992) ICM for object recognition. In *Computational Statistics*, ed. Y. Dodge and J. Whittaker, Vol. 2, pp. 271–286, Physica/Springer, Heidelberg.

[5] BADDELEY, A. J. AND VAN LIESHOUT, M. N. M. (1992) Object recognition using Markov spatial processes. In *Proc. 11th IAPR Internat. Conf. on Pattern Recognition*, Volume B, pp. 136–139, IEEE Computer Society Press, Los Alamitos.

[6] BADDELEY, A. J. AND VAN LIESHOUT, M. N. M. (1993) Stochastic geometry models in high-level vision. In: *Statistics and Images*, Vol. 1, ed. K. V. Mardia and G. K. Kanji. Special issue *J. Appl. Statist.* **20,** 233–258.

[7] *BESAG, J.* (1986) *On the statistical analysis of dirty pictures (with discussion). J. R. Statist. Soc. Ser. B* **48,** 259–302.

[8] DOBRUSHIN, R. L. (1956) Central limit theorem for non-stationary Markov chains I, II. *Theory Prob. Appl.* **1,** 65–80, 329–383.

[9] GEMAN, S. AND GEMAN, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6,** 721–741.

[10] GEMAN, S. AND HWANG, C. R. (1986) Diffusions for global optimization. *SIAM. J. Control and Optimization* **24,** 1031–1043.

[11] GEMAN, D. (1990) Random fields and inverse problems in imaging. In *Ecole d'été de Probabilités de Saint-Flour XVIII*—1988. Lecture Notes in Mathematics **1427**. Springer-Verlag, Berlin.

[12] GEMAN, D., GEMAN, S., GRAFFIGNE, C. AND DONG, P. (1990) Boundary detection by constrained optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence* **12,** 609–628.

[13] GREIG, D. M., PORTEOUS, B. T. AND SEHEULT, A. H. (1989) Exact maximum a posteriori estimation for binary images. *J. R. Statist. Soc.* B **51,** 271–279.

[14] GRENANDER, U. (1976) *Lectures on Pattern Theory.* Vol. 1; *Pattern Synthesis.* Applied Mathematical Sciences Vol. 18, Springer-Verlag, Berlin.

[15] GRENANDER, U. (1978) *Lectures on Pattern Theory.* Vol. 2: *Pattern Analysis* Applied Mathematical Sciences Vol. 24, Springer-Verlag, Berlin.

[16] GRENANDER, U. (1981) *Lectures on Pattern Theory.* Vol. 3; *Regular Structures.* Applied Mathematical Sciences, Vol. 33. Springer-Verlag, Berlin.

[17] GRENANDER, U. AND KEENAN, D. M. (1989) A computer experiment in pattern theory. *Commun. Statist.-Stoch. Models* **5,** 531–533.

[18] HAARIO, H. AND SAKSMAN, E. (1991) Simulated annealing process in general state space. *Adv. Appl. Prob.* **23,** 866–893.

[19] HOUGH, P. V. C. (1962) Method and means for recognizing complex patterns. US Patent 3,069,654.

[20] ILLINGWORTH, J. AND KITTLER, J. (1988) A survey of the Hough transform. *Computer Vision, Graphics and Image Processing* **44,** 87–116.

[21] TEN KATE, T. K., VAN BALEN, R., SMEULDERS, A. W. M., GROEN, F. C. A. AND DEN BOER, G. A. (1990) SCILAIM: a multi-level interactive image processing environment. *Pattern Recognition Letters* **11,** 429–441.

[22] LOTWICK, H. W. AND SILVERMAN, B. W. (1981) Convergence of spatial birth-and-death processes. *Math. Proc. Camb. Phl. Soc.* **90,** 155–165.

[23] MILLER, M. et al. (1991) Automated segmentation of biological shapes in electron microscopic autoradiography. In *25th Annual Conf. Inf. Sci. Systems*, ed. F. Davidson and J. Goutsias.

[24] MØLLER, J. (1989) On the rate of convergence of spatial birth-and-death processes. *Ann. Inst. Statist. Math.* **41,** 565–581.

[25] PARZEN, E. (1962) *Stochastic Processes.* Holden-Day, San Francisco.

[26] Preston, C. J. (1977) Spatial birth-and-death processes. *Bull. Internat. Statist. Inst.* **46,** 371–391.

[27] Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B **39,** 172–212.

[28] Ripley, B. D. and Kelly, F. P. (1977) Markov point processes. *J. London Math. Soc.* **15,** 188–192.

[29] Stoyan, D., Kendall, W. S. and Mecke, J. (1987) *Stochastic Geometry and its Applications.* Wiley, Chichester.

[30] Winkler, G. (1990) An ergodic L^2 theorem for simulated annealing in Bayesian image reconstruction. *J. Appl. Prob.* **28,** 779–791.