CHAPTER 11

Stochastic geometry models in high-level vision

A. J. BADDELEY[1] **& M. N. M. VAN LIESHOUT**[2], [1]*Centre for Mathematics and Computer Science, Amsterdam, and Department of Mathematics and Computer Science, University of Leiden &* [2]*Department of Mathematics and Computer Science, Free University of Amsterdam, and Centre for Mathematics and Computer Science, Amsterdam*

SUMMARY *We survey the use of Markov models from stochastic geometry as priors in 'high-level' computer vision, in direct analogy with the use of discrete Markov random fields in 'low-level' vision. There are analogues of the Gibbs sampler, ICM and simulated annealing, and connections with existing methods in computer vision.*

1 Introduction

Object recognition is the task of interpreting a noisy image to identify certain geometrical features. An object recognition algorithm must decide whether there are any objects of a specified kind in the scene, and if so, determine the number of objects and their locations, shapes, sizes and spatial relationship. The image data are noisy and sometimes blurred. There is a wide field of applications, including industrial robot vision, document reading, interpretation of medical scans (Blokland, 1987), automated cytology (Miller *et al.*, 1991), classification of astronomical features (Molina & Ripley, Chapter 13; Ripley & Sutherland, 1990) and identification of grain structures in materials science.

It is increasingly acknowledged that discrete Markov random fields (MRFs) are not the appropriate prior models to use in object recognition. This is partly because geometrical shapes with smooth boundaries are unlikely to arise as realizations of a discrete MRF; but more importantly because the procedures that result from applying a discrete MRF model do not combine information 'globally' to identify geometrical shape.

This issue is familiar from the computer vision literature as the distinction between 'low-level' and 'high-level' vision. Low-level tasks such as segmentation, classification and tomographic reconstruction call for local (pixel neighbourhood) operations, converting the input image into another raster image. In high-level tasks such as object recognition and scene analysis, we have to interpret the image globally, reducing it to a compact description (e.g. a vector graphics representation) of the scene.

Alternative approaches have been described in recent studies (Dengler & Guckes, 1992; Miller *et al.*, 1991; Molina & Ripley, Chapter 13; Ripley & Sutherland, 1990) on recognizing the shape of an interesting object (hand, galaxy, mitochondrion). The shape is described by a flexible template, typically a polygon, with edge lengths and angles governed by a joint prior distribution, typically a Markov chain.

Our own recent work (van Lieshout, 1991, 1993; Baddeley & van Lieshout, 1991, 1992a, b) has studied the problem of detecting an unknown number of objects of (usually simple) shape in an unknown spatial arrangement, possibly overlapping each other. This requires a prior stochastic model for the spatial arrangement of the objects. We used the Markov object processes studied in stochastic geometry and spatial statistics (Baddeley & Møller, 1989; Ogata & Tanemura, 1984; Ripley, 1988, 1989; Ripley & Kelly, 1977) which have a simple mathematical form, and for which there is a natural analogue of the Gibbs sampler (a spatial birth-and-death process (Baddeley & Møller, 1989; Møller, 1989; Preston, 1977; Ripley, 1979)). This allowed us to develop analogues of the ICM and simulated annealing algorithms. Also, some existing techniques in computer vision turned out to be equivalent to maximum likelihood methods. The use of Markov point process models was also proposed by Molina and Ripley (Chapter 13), Ripley (1991) and Ripley and Sutherland (1990).

This paper describes the (possible) uses of Markov random processes of geometrical objects (Ripley & Kelly, 1977; Baddeley & Møller, 1989) as priors in high-level vision. We claim that the resulting techniques can be applied to various tasks such as object recognition, motion detection and large-scale edge detection and to the identification of spatial clusters in a point pattern.

2 Maximum likelihood object recognition

Object recognition can be formulated as a parameter-estimation problem by direct analogy with the formulation of segmentation and classification (Besag, Chapter 5; Geman, 1990; Geman & Geman, Chapter 4) and in keeping with the general set-up of Grenander (1976, 1978, 1981). In this section, we define the object recognition problem, develop a simple maximum likelihood treatment, and show that this is very similar to some existing techniques in computer vision.

2.1 Set-up

The experimental data consist of an image $\mathbf{y} = (y_t : t \in T)$ where 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 set V that is arbitrary, typically $\{0, 1\}$ or $\{0, 1, \ldots, 255\}$.

The class U of possible objects is an arbitrary set (object space). Typical examples would be the class of all polygons in $\mathbb{R}^2$ or all convex polyhedra in $\mathbb{R}^3$. However, U need not be a class of subsets of $\mathbb{R}^d$ since the specification of an

object may also include properties like colour or surface texture. Section 2.1 discusses this further. Here we assume only that each object $u \in U$ determines a subset $R(u) \subseteq T$ of image space 'occupied' by the object.

An object configuration is simply a finite set of objects

$$\mathbf{x} = \{x_1, \ldots, x_n\}$$

where $x_i \in U, i = 1, \ldots, n, n \geq 0$. The objects may be in any spatial relation to each other; the number of objects is variable and may be zero.

2.2 Noise models

Assume that the observed image $\mathbf{y}$ depends on the true object configuration $\mathbf{x}$ through a known conditional probability density $f(\mathbf{y} \mid \mathbf{x})$. This density describes the 'forward problem' of image formation and includes both the deterministic influence of $\mathbf{x}$ and the stochastic noise inherent in observing $\mathbf{y}$.

Following custom, we assume that the data pixel values y_t are conditionally independent given $\mathbf{x}$. This embraces additive and multiplicative random noise as well as Poisson distributed counts and more general exponential family models. Without loss of generality the conditional distributions of individual pixel values belong to a family of distributions with densities $\{g(\cdot \mid \theta): \theta \in \Theta\}$ indexed by a parameter space Θ. It follows that

$$f(\mathbf{y} \mid \mathbf{x}) = \prod_{t \in T} g(y_t \mid \theta^{(\mathbf{x})}(t)) \tag{1}$$

where $\theta^{(\mathbf{x})}(t) \in \Theta$ is the parameter value of the conditional distribution of y_t given $\mathbf{x}$. Then $\theta^{(\mathbf{x})}(\cdot)$ is a Θ-valued image, deterministically derived from $\mathbf{x}$, which we call the signal.

'Additive noise' is the case where $g(\cdot \mid \theta)$ is a location family of distributions. For example, for additive Gaussian noise $g(\cdot \mid \theta)$ is a Gaussian distribution with mean θ and fixed variance σ^2.

A simple example is the signal

$$\theta^{(\mathbf{x})}(t) = \begin{cases} \theta_1, & \text{if } t \in S(\mathbf{x}) \\ \theta_0, & \text{otherwise} \end{cases}$$

where θ_1, θ_0 are known parameters ('foreground' and 'background' signal) and $S(\mathbf{x})$ is the 'silhouette'

$$S(\mathbf{x}) = \bigcup_{i=1}^{n} R(x_i)$$

formed by taking the union of all objects in the configuration. In other words, under this simple model, each of the objects in the configuration $\mathbf{x}$ is 'painted' on to the scene, and independent random pixel noise is superimposed on the result.

We call this a 'blur-free' independent noise model. It may seem oversimplified, yet we shall show that several standard techniques in computer vision are equivalent to assuming this model. Our practical examples will be 'blurred' models.

2.3 Maximum likelihood estimation

Given observation of $\mathbf{y}$, the likelihood of a configuration $\mathbf{x}$ is defined to be $\ell(\mathbf{x}; \mathbf{y}) = f(\mathbf{y} \mid \mathbf{x})$, and we seek 'the' maximum likelihood estimate of $\mathbf{x}$

$$\hat{\mathbf{x}} = \operatorname{argmax}_{\mathbf{x}} f(\mathbf{y} \mid \mathbf{x}) \tag{2}$$

which may be non-existent or non-unique. Specifically, note that for any blur-free model the likelihood depends on $\mathbf{x}$ only through its silhouette $S(\mathbf{x})$, so configurations with the same silhouette cannot be distinguished in likelihood.

Since the log-likelihood is a sum of individual pixel error terms

$$L(\mathbf{x}; \mathbf{y}) = \log f(\mathbf{y} \mid \mathbf{x}) = \sum_{t \in T} \log g(y_t \mid \theta^{(x)}(t))$$

maximum likelihood estimation is equivalent to regression of $\mathbf{y}$ on the class of signals $\theta^{(\mathbf{x})}(\cdot)$ for all possible $\mathbf{x}$, with pixelwise loss function $-\log g(y_t \mid \cdot)$. For example, least squares regression is the case of Gaussian additive noise; L^1 (least mean absolute deviation) regression is the case of additive two-sided exponential noise.

Squared error and L^1 error have been proposed as optimality criteria for object recognition in their own right (Lin *et al.*, 1990; Maragos, 1988) but we see here that they are special cases of the maximum likelihood approach.

It is also interesting to note that popular 'pre-processing' techniques, such as thresholding or gamma correction, amount to simply modifying the noise model. If the pixel values y_t are subjected to a transformation $y'_t = \phi(y_t)$, then the model (1) is transformed into another model of the same type with g replaced by another density g'. This is true not only for differentiable maps ϕ but also for transformations such as 'clipping' pixel values to an interval $[a, b]$:

$$y'_t = \begin{cases} a, & \text{if } y_t < a \\ y_t, & \text{if } a \le y_t \le b \\ b, & \text{if } y_t > b \end{cases}$$

Treating clipped pixel data as if they arise from additive Gaussian noise is equivalent to assuming additive two-sided exponential noise with signal (=mean) values $\theta_0 = a$, $\theta_1 = b$. These remarks do not hold for more complex pre-processing operations such as neighbourhood filtering, which interfere with the dependence structure of (1).

2.4 Connection with mathematical morphology

Consider a binary image ($y_t = 0, 1$) and a noise model under which pixels in the background are randomly corrupted with probability q (independently of each other) while foreground pixels are not corrupted. Thus $g(\cdot \mid \theta)$ is a Bernoulli(θ) distribution and $\theta_0 = q$, $\theta_1 = 1$.

A maximum likelihood estimator for this model is clearly

$$\begin{aligned} \tilde{\mathbf{x}} &= \{u : y_t = 1 \text{ for all } t \in R(u)\} \\ &= \{u : R(u) \subseteq Y\} \end{aligned} \tag{3}$$

where $Y = \{t : y_t = 1\}$ is the set of foreground pixels in the data image. This $\tilde{\mathbf{x}}$ is the largest solution to the ML estimating equations (2); all solutions are characterized by $\mathbf{x} \subseteq \tilde{\mathbf{x}}$ and $S(\mathbf{x}) = S(\tilde{\mathbf{x}})$.

The operation (3) can be recognized as the erosion operator of mathematical morphology (Serra, 1982) in its general form (Serra, 1988). The 'classical' erosion operator $\ominus$ is the special case where $U = T \subset \mathbb{R}^2$ and $R(u) = u + R$, where R is a fixed subset of T. Then

$$\tilde{\mathbf{x}} = \{u : (u + R) \subseteq Y\} = Y \ominus \check{R}$$

Thus the erosion operator is the MLE for a simple noise model. A similar statement about the dilation operator $\oplus$ is obtained by exchanging foreground and background.

2.5 Iterative methods for MLE

It is usually impossible to solve the ML estimating equations (2) directly, and one has to resort to iterative approximation methods.

The simplest form of iterative adjustment is to add or delete objects one at a time. We start by choosing an initial configuration, such as the empty list $\emptyset$. If the current configuration is $\mathbf{x}$, then we consider adding a new object $u \notin \mathbf{x}$, yielding configuration $\mathbf{x} \cup \{u\}$, if the log likelihood ratio

$$L(\mathbf{x} \cup \{u\}; \mathbf{y}) - L(\mathbf{x}; \mathbf{y}) = \log \frac{f(\mathbf{y} \mid \mathbf{x} \cup \{u\})}{f(\mathbf{y} \mid \mathbf{x})} \tag{4}$$

is sufficiently large; delete an existing object $x_i \in \mathbf{x}$ yielding $\mathbf{x} \backslash \{x_i\}$ if

$$L(\mathbf{x} \backslash \{x_i\}; \mathbf{y}) - L(\mathbf{x}; \mathbf{y}) = \log \frac{f(\mathbf{y} \mid \mathbf{x} \backslash \{u\})}{f(\mathbf{y} \mid \mathbf{x})} \tag{5}$$

is sufficiently large. Two variations of this scheme are to visit the possible objects u sequentially (assuming U is discretized) applying the above rules at each step, or to scan the whole of U to find the object u whose addition or deletion would most increase the likelihood.

Algorithm 1 (Coordinatewise optimization). Initialize $\mathbf{x}^{(0)} = \emptyset$ or some other chosen initial state. When the current reconstruction is $\mathbf{x}^{(m-1)}$, visit every $u \in U$ sequentially in a predetermined order. If $u \notin \mathbf{x}^{(m-1)}$ and $L(\mathbf{x}^{(m-1)} \cup \{u\}; \mathbf{y}) - L(\mathbf{x}^{(m-1)}; \mathbf{y}) \geq w$, where $w \geq 0$ is a fixed threshold, then add u to the configuration, yielding $\mathbf{x}^{(m)} = \mathbf{x}^{(m-1)} \cup \{u\}$. If $u = x_i \in \mathbf{x}$ and $L(\mathbf{x}^{(m-1)} \backslash \{x_i\}; \mathbf{y}) - L(\mathbf{x}^{(m-1)}; \mathbf{y}) \geq w$, then delete x_i yielding $\mathbf{x}^{(m)} = \mathbf{x}^{(m-1)} \backslash \{x_i\}$. Update recursively until one complete scan of the image yields no changes.

Algorithm 2 (Steepest ascent). Initialize $\mathbf{x}^{(0)} = \emptyset$ or some other chosen initial state. Given $\mathbf{x}^{(k-1)}$, determine

$$a = \max_{x_i \in \mathbf{x}^{(k-1)}} \{L(\mathbf{x}^{(k-1)} \backslash \{x_i\}; \mathbf{y}) - L(\mathbf{x}^{(k-1)}; \mathbf{y})\}$$

and

$$b = \sup_{u \in U} \{L(\mathbf{x}^{(k-1)} \cup \{u\}; \mathbf{y}) - L(\mathbf{x}^{(k-1)}; \mathbf{y})\}$$

If $\max\{a, b\} < w$, then stop. Otherwise, if $b \geq a$, add the corresponding object, while if $a > b$, delete the corresponding object.

Both algorithms increase the likelihood at each step. Assuming U is finite, they are guaranteed to converge in finite time to a local maximum (or if $w = 0$ possibly a cycle of states with equal likelihood). The final result depends on the initial state, and for Algorithm 1 on the scanning order as well.

These algorithms bear a very strong resemblance to Besag's ICM algorithm (Chapter 5) except for the lack of a prior distribution. The analogy will be explored in Section 4.

Generalizations of these algorithms are discussed in Section 6.

2.6 Relation to Hough transform

A popular tool for detecting simple objects is the Hough transform (Davies, 1990; Hough, 1962; Illingworth & Kittler, 1988). Given a binary or grey-scale image, the (generalized) Hough transform is an integer-valued function on the object space

$$H_{\mathbf{y}}(u) = \sum_{t \in R(u)} y_t, \qquad u \in U \tag{6}$$

This is often interpreted as a 'vote-counting' operation: each pixel t votes with strength y_t for all the objects that contain that pixel. The optimal match is located typically by finding local maxima of the Hough transform.

The Hough transform is very similar to the log-likelihood ratio (4) for a blur-free model:

$$L(\mathbf{x} \cup \{u\}; \mathbf{y}) - L(\mathbf{x}; \mathbf{y}) = \sum_{R(u) \setminus S(\mathbf{x})} z_t \tag{7}$$

where $z_t = \log g(y_t \mid \theta_1) - \log g(y_t \mid \theta_0)$ is the difference in goodness-of-fit at pixel t. In fact, this is identical to the Hough transform of image $\mathbf{z}$ when the new object u does not overlap any existing object $x_i \in \mathbf{x}$. For example, the Hough transform is the log-likelihood ratio for comparing $\{u\}$, the scene consisting of a single object, against the empty scene $\emptyset$. When objects do overlap, the likelihood ratio (7) is a generalization of the Hough transform, with domain of summation 'masked' by the silhouette of the current configuration. Equivalently, (7) is the Hough transform of the masked image $w_t^{(\mathbf{x})} = z_t 1\{t \notin S(\mathbf{x})\}$.

The similarity between (6) and (7) is even stronger since z_t is linear in y_t for many exponential (and some non-exponential) noise models. For additive Gaussian noise

$$z_t = \frac{\theta_1 - \theta_0}{\sigma^2}\left(y_t - \frac{\theta_0 + \theta_1}{2}\right)$$

and for Poisson noise

$$x_t = y_t \log \frac{\theta_1}{\theta_0} - (\theta_1 - \theta_0)$$

The likelihood ratio is then of the form

$$L(\mathbf{x} \cup \{u\}; \mathbf{y}) - L(\mathbf{x}; \mathbf{y}) = a \sum_{R(u) \setminus S(\mathbf{x})} y_t - b\left|R(u) \setminus S(\mathbf{x})\right|$$

in particular when u does not overlap $\mathbf{x}$ this is a linear adjustment of the Hough transform of $\mathbf{y}$.

Apart from illuminating the meaning of the Hough transform, these results show how to correctly interpret values of the Hough transform when the objects do not have equal area, e.g. in the presence of edge effects. For Gaussian or Poisson additive noise (say), the likelihood ratio is positive when the average value of y_t over $R(u) \setminus S(\mathbf{x})$ exceeds a critical value. The latter is simply the Neyman–Pearson critical value for classifying a single observation y_t as foreground or background ($\theta \in \{\theta_0, \theta_1\}$).

Regarding the practical implementation of Algorithms 1–2, note that each successive addition or deletion of an object u alters (7) only on

$$V(u) = \{v \in U: R(v) \cap R(u) \neq \emptyset\}$$

so that updating is required only for $v \in V(u)$. For example, in a translation model with $U = T = \mathbb{R}^d$ and $R(u) = R_0 + u$ this is the central symmetrization $V(u) = R_0 \oplus \check{R}_0 + u$.

For the general 'blurred' model (1): the likelihood ratio is again similar to the Hough transform

$$L(\mathbf{x} \cup \{u\}; \mathbf{y}) - L(\mathbf{x}; \mathbf{y}) = \sum_{t \in Z(\mathbf{x},u)} h(y_t, \theta^{(\mathbf{x})}(t), \theta^{(\mathbf{x} \cup \{u\})}(t)) \tag{8}$$

where

$$h(y_t, \theta, \theta') = \log \frac{g(y_t \mid \theta')}{g(y_t \mid \theta)}$$

is again the difference in goodness-of-fit at pixel t and

$$Z(\mathbf{x}, u) = \{t: \theta^{(\mathbf{x})}(t) \neq \theta^{(\mathbf{x} \cup \{u\})}(t)\}$$

is the set of pixels where the signal is affected by the addition of object u.

2.7 Example

Figure 1 shows a scanned image ('pellets') taken from the Brodatz texture album (Brodatz, 1966). This is a relatively easy data set for object recognition but helps to illustrate the theory.

We treat the pellets as discs of fixed radius 4 pixels but with blurred boundaries. The grey-level histogram has two distinct peaks at value 8 and 172, suggesting that we can regard the background and foreground signal as roughly constant at these values. Assuming additive Gaussian noise, the noise variance was estimated by thresholding the image and taking the sample variance, giving an estimate of 83.1. Blurring was modelled by assuming that the original blur-free signal was subjected to a 3×3 averaging (linear) filter with relative weights 4 for the central pixel, 2 for horizontal and vertical neighbours, and 1 for diagonal neighbours.

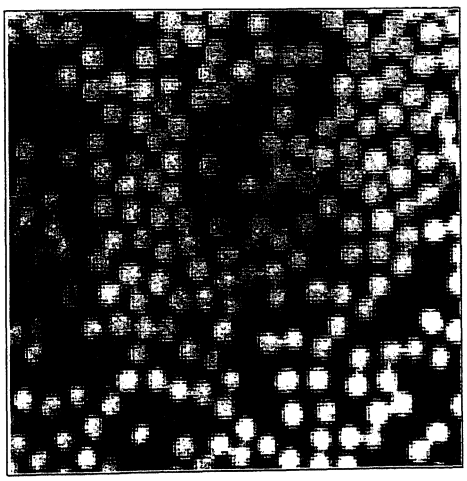

FIG. 1. Pellets image.

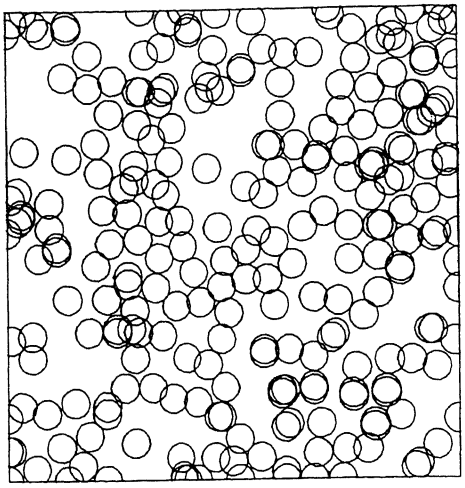

FIG. 2. Maximum likelihood reconstruction by steepest ascent.

Figure 2 shows an approximate MLE computed by steepest ascent (Algorithm 2) from an empty initial configuration. Pellets are correctly identified but there is 'multiple response', i.e. the MLE sometimes contains clusters of objects around the position of each 'true' object.

3 Markov processes of geometrical objects

This section is an overview (with some adaptations) of the theory of random processes of geometrical objects (Stoyan *et al.*, 1987) with emphasis on Markov object processes (Ripley & Kelly, 1977; Baddeley & Møller, 1989).

3.1 Objects

The 'objects' featuring in stochastic geometry range from simple geometrical figures (points, lines, discs) through plane polygons and convex compact sets to completely general closed sets. A given class of objects U is treated as a space in its own right, so that objects are regarded as points in U.

At one extreme, simple geometrical figures can be specified by the values of a few parameters (giving location, orientation, etc.) so that U is isomorphic to a subset of $\mathbb{R}^k$. For example, a disc in $\mathbb{R}^2$ can be specified by its centre (x, y) and radius r so that $U = \mathbb{R}^2 \times \mathbb{R}_+$. At the other extreme, the space $\mathcal{F}$ of all closed subsets of $\mathbb{R}^d$ can be made into a locally compact, second countable Hausdorff space (lcs space) so that a random closed set can be defined as a random element of $\mathcal{F}$ (Matheron, 1975).

It is often useful to represent an object as a 'marked point', i.e. a pair (s, m) consisting of a point $x \in \mathbb{R}^d$ and a 'mark' $m \in \mathcal{M}$, where $\mathcal{M}$ is an arbitrary lcs space. The point s fixes the location of the object and the mark m contains all other information such as size and shape. A disc in $\mathbb{R}^2$ can be regarded as a point (x, y) marked by a radius r. Objects with additional properties such as colour and surface texture can be represented as marked points by choosing an appropriate mark space $\mathcal{M}$. For example, a grey-scale surface texture can be formalized as

an upper semi-continuous function $\mathbb{R}^d \to \mathbb{R}_+$, and the space of all such functions is lcs.

3.2 Poisson processes of objects

For applications, we need only consider spatial patterns in a bounded region $T \subset \mathbb{R}^d$. Let U be the class of compact subsets of T (with the relative myopic topology (Matheron, 1975)). As before, a configuration is a finite set $\mathbf{x} = \{x_1, \ldots, x_n\}$ of objects $x_i \in U$. Write Ω for the set of all configurations. A random process of objects is a random element of Ω or, equivalently, a point process on U consisting of a finite number of 'points' with probability 1.

The basic reference model is the Poisson object process in U with intensity μ, where μ is a finite measure on U. Under this model the total number of objects has a Poisson distribution with mean $\mu(U)$; given that exactly n objects are present, they are independent and identically distributed in U with probability distribution proporitional to μ, i.e. $\mathbb{P}\{x_i \in B\} = Q(B) = \mu(B) \backslash \mu(U)$ for $B \subseteq U$.

Further details can be consulted in Stoyan *et al.* (1987).

3.3 Markov processes of objects

Our interest is in constructing non-Poisson spatial processes exhibiting dependence between neighbouring objects. To do this, we shall specify the probability density of the new process with respect to the Poisson process (thereby restricting attention to processes that are absolutely continuous with respect to the Poisson). The density is an integrable function $p: \Omega \to [0, \infty)$. For the new process, the distribution of the total number of objects is

$$\mathbb{P}\{N = n\} = q_n = \frac{e^{-\mu(U)}}{n!} \int_U \cdots \int_U p(\{x_1, \ldots, x_n\})\, d\mu(x_1) \ldots d\mu(x_n)$$

and given $N = n$, the n random objects have joint probability density $p_n(x_1, \ldots, x_n) = e^{-\mu(U)} \mu(U)^n p(\{x_1, \ldots, x_n\})/(n! q_n)$ with respect to the distribution of n iid objects in U with distribution Q.

Provisionally define two objects u, v to be 'neighbours' if they overlap:

$$u \sim v \Leftrightarrow R(u) \cap R(v) \neq \emptyset \tag{9}$$

This can be replaced by any symmetric ($u \sim v$ iff $v \sim u$) relation between elements of U.

The simplest kind of spatial interaction is that which forbids objects to overlap. Consider a Poisson process of objects in T conditioned on the event that no pair of objects is overlapping. Its density with respect to the original Poisson process is simply

$$p(\mathbf{x}) = \begin{cases} 0, & \text{if } x_i \sim x \text{ for some } i \neq j \\ \alpha, & \text{otherwise} \end{cases} \tag{10}$$

where $\alpha > 0$ is the normalizing constant (=reciprocal of the probability that Poisson process has no overlapping objects). Call this a hard object process by analogy with the better-known hard core point process.

Next consider a pairwise interaction

$$p(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \prod_{x_i \sim x_j} g(x_i, x_j) \tag{11}$$

where $\alpha, \beta > 0$ are constants, n is the number of points in $\mathbf{x}$, and $g: U \times U \to [0, \infty)$. The product is over all pairs of neighbouring objects $x_i \sim x_j$ with $i < j$.

If $g \equiv 1$ then (11) is simply a Poisson process with intensity $\beta\mu$; if $g \equiv 0$ it is the hard object process (10). When $g \equiv \gamma$ for a constant $0 < \gamma < 1$ the model is called a Strauss object process and the density can be written

$$p(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})} \tag{12}$$

where

$$s(\mathbf{x}) = \sum_{i<j} 1\{x_i \sim x_j\}$$

is the number of pairs of neighbouring objects (e.g. number of overlaps) in the configuration. This process exhibits 'repulsion' or 'inhibition' between objects, since $s(\mathbf{x})$ tends to be smaller than under the Poisson model. The density (12) is typically not integrable for $\gamma > 1$.

Just as discrete Markov random fields are closely connected with statistical physics, pairwise interaction processes (11) also occur as models of interacting particle systems. The log probability density of a particular configuration $\mathbf{x}$ can be interpreted as its physical 'energy': it is the sum of a ground potential $\log \alpha$, a potential $\log \beta$ for the presence of each object x_i, and an interaction potential $\log g(u, v)$ between each pair of neighbouring objects u, v.

The density (11) bears a close resemblance to the distribution of a discrete Markov random field with pairwise interaction. However, the number of terms appearing in the product in (11) depends on the realization $\mathbf{x}$. Some configurations have more interaction than others.

Note that if $u \in U$, $u \notin \mathbf{x}$ with $p(\mathbf{x}) > 0$, the ratio

$$\frac{p(\mathbf{x} \cup \{u\})}{p(\mathbf{x})} = \beta \prod_{x_i \sim u} g(u, x_i) \tag{13}$$

depends only on u and on the neighbours of u in $\mathbf{x}$. This important property signifies that all interaction is 'local'. In the statistical physics interpretation, $\log p(\mathbf{x} \cup \{u\}) - \log p(\mathbf{x})$ is the energy required to add a new point u to an existing configuration $\mathbf{x}$. In probabilistic terms, $p(\mathbf{x} \cup \{u\})/p(\mathbf{x})$ is the Papangelou conditional intensity at u given the rest of the pattern $\mathbf{x}$ on $U \setminus \{u\}$, see Daley and Vere-Jones (1988).

Following are definitions and results of Ripley and Kelly (1977) trivially generalized to random object processes (Baddeley & Møller, 1989, Section 3). Let $\sim$ be an arbitrary symmetric relation on U. A random object process with density p is called a Markov object process with respect to $\sim$ if, for all $\mathbf{x} \in \Omega$,

(a) $p(\mathbf{x}) > 0$ implies $p(\mathbf{y}) >$ for all $\mathbf{y} \subset \mathbf{x}$;
(b) if $p(\mathbf{x}) > 0$, then

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

depends only on u and nbd $(u \mid \mathbf{x}) = \{x_i : u \sim x_i\}$.

The name 'Markov' is justified by the fact that a spatial version of the Markov property holds for such processes (Ripley & Kelly, 1977; Baddeley & Møller, 1989; Kendall, 1990).

Define a configuration $\mathbf{x} \in \Omega$ to be a clique if all members of $\mathbf{x}$ are neighbours ($x_i \sim x_j$ for all $i \neq j$). Configurations of 0 or 1 objects are cliques. Then the Ripley–Kelly analogue of the Hammersley–Clifford theorem (Ripley & Kelly, 1977) states that a process with density $p: \Omega \to [0, \infty)$ is Markov iff

$$p(\mathbf{x}) = \prod_{\text{cliques } \mathbf{y} \subseteq \mathbf{x}} q(\mathbf{y}) \tag{14}$$

for all $\mathbf{x} \in \Omega$, where the product is restricted to cliques $\mathbf{y} \subseteq \mathbf{x}$, and $q: \Omega \to [0, \infty)$ is an (arbitrary) function.

An example of an overlapping object model with higher-order interaction terms is the area interaction process defined by

$$p(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \delta^{|S(\mathbf{x})|} \tag{15}$$

with parameters $\beta > 0$, $\delta \geq 1$ and normalizing constant $\alpha > 0$. As usual, $n(\mathbf{x})$ is the number of objects in list $\mathbf{x}$ and $|S(\mathbf{x})|$ is the area (or pixel count) of the silhouette. This model is clearly Markov; it again favours configurations with relatively few overlapping objects.

3.4 Nearest-neighbour Markov object processes

A further extension (Baddeley & Møller, 1989) is to allow interaction behaviour to depend on the realization of the process. For example, in a one-dimensional renewal process, each point can be said to interact with its nearest neighbours to the left and right, regardless of how far distant these neighbours may be. In two dimensions, we would like to construct point processes exhibiting interaction between those pairs of points that are neighbours with respect to the Voronoi (Dirichlet) tessellation generated by the point pattern.

Assume that for each configuration $\mathbf{x}$ we have a symmetric reflexive relation $\underset{\mathbf{x}}{\sim}$ defined on $\mathbf{x}$. We might prefer to think of this as a finite graph whose vertices are the objects $x_i \in \mathbf{x}$. Two consistency conditions are required on the relations $\underset{\mathbf{x}}{\sim}$ for different $\mathbf{x}$ (Baddeley & Møller, 1989, Definition 4.7). Examples of relations satisfying these conditions are

- $x_i \underset{\mathbf{x}}{\sim} x_j$ iff $x_i \sim x_j$ where $\sim$ is any symmetric relation on U (i.e. not depending on the configuration $\mathbf{x}$);
- for points or marked points in $\mathbb{R}^2$, $x_i \underset{\mathbf{x}}{\sim} x_j$ iff x_i, x_j are joined by an edge of the Delaunay triangulation generated by $\mathbf{x}$;
- for compact sets in $\mathbb{R}^d$, $x_i \underset{\mathbf{x}}{\sim} x_j$ iff x_i and x_j belong to the same connected component of the union of the objects.

A random object process with density p shall be called a nearest-neighbour Markov object process (nnMp) with respect to $\{\underset{\mathbf{x}}{\sim}: \mathbf{x} \in \Omega\}$ if, for all $\mathbf{x}$ with $p(\mathbf{x}) > 0$,

- $p(\mathbf{y}) >$ for all $\mathbf{y} \subset \mathbf{x}$;
- the ratio $p(\mathbf{x} \cup \{u\})/p(\mathbf{x})$ depends only on u, on $\text{nbd}(u \mid \mathbf{x} \cup \{u\}) = \{x_i: x_i \underset{\mathbf{x} \cup \{u\}}{\sim} u\}$ and on the relations $\underset{\mathbf{x}}{\sim}$, $\underset{\mathbf{x} \cup \{u\}}{\sim}$ restricted to $\text{nbd}(u \mid \mathbf{x} \cup \{u\})$.

A subset $\mathbf{y} \subseteq \mathbf{x}$ is a clique in $\mathbf{x}$ if all members of $\mathbf{y}$ are $\mathbf{x}$-neighbours of one another ($u \underset{\mathbf{x}}{\sim} v$ for all $u, v \in \mathbf{y}$).

A generalized Hammersley–Clifford theorem holds (Baddeley & Møller, 1989): a process with density p is nnMp iff

$$p(\mathbf{x}) = \begin{cases} \prod_{\text{cliques } \mathbf{y} \subseteq \mathbf{x}} q(\mathbf{y}), & \text{if } q(\mathbf{y}) > 0 \text{ for } \textit{all } \mathbf{y} \subseteq \mathbf{x} \\ 0, & \text{otherwise} \end{cases}$$

where $q: \Omega \to \mathbb{R}_+$ satisfies certain regularity conditions ((I1)–(I2) of Baddeley & Møller, 1989).

Kendall (1990) proved a spatial Markov property for nnMps.

3.5 Spatial birth-and-death processes

The natural analogue of the Gibbs sampler in this context is a spatial birth-and-death process (Møller, 1989). This is a continuous-time Markov process, whose states are configurations $\mathbf{x} \in \Omega$, and for which the only transitions are the birth of a new object (instantaneous transition from $\mathbf{x}$ to $\mathbf{x} \cup \{u\}$) or the death of an existing one (transition from $\mathbf{x}$ to $\mathbf{x} \backslash \{x_i\}$). Given the state $\mathbf{x}$ at time t,

- the probability of a death $\mathbf{x} \to \mathbf{x} \backslash \{x_i\}$ during a time interval $(t, t+h)$, $h \to 0$, is $D(\mathbf{x} \backslash \{x_i\}, x_i)h + o(h)$, where $D(\cdot, \cdot): \Omega \times U \to [0, \infty)$ is a measurable function;
- the probability of a birth $\mathbf{x} \to \mathbf{x} \cup \{u\}$ during time $(t, t+h)$, where u lies in a given set $F \subseteq U$, is $B(\mathbf{x}, F)h + o(h)$ where $B(\mathbf{x}, \cdot)$ is a finite measure on U.
- the probability of more than one transition during $(t, t+h)$ is $o(h)$.

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

Conditions for the existence and convergence of a spatial birth–death process with given transition rates were given by Preston (1977, Property 5.1, Theorem 7.1) and slightly extended by Baddeley and Møller (1989) and Møller (1989) as follows.

Theorem 1. For each $n = 0, 1, \ldots$ define

$$\kappa_n = \sup_{n(\mathbf{x}) = n} B(\mathbf{x})$$

and

$$\delta_n = \inf_{n(\mathbf{x}) = n} D(\mathbf{x})$$

Assume $\delta_n > 0$ for all $n \geq 1$. Assume either (a) that $\kappa_n = 0$ for all sufficiently large $n \geq 0$, or (b) that $\kappa_n > 0$ for all $n \geq 1$ and both the following hold:

$$\sum_{n=2}^{\infty} \frac{\kappa_1 \cdots \kappa_{n-1}}{\delta_1 \cdots \delta_n} < \infty$$

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

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.

Given an object process with density p, if there exists a spatial birth-and-death process with rates satisfying

$$\frac{b(\mathbf{x}, u)}{D(\mathbf{x}, u)} = \frac{p(\mathbf{x} \cup \{u\})}{p(\mathbf{x})}$$

whenever $p(\mathbf{x} \cup \{u\}) > 0$, then the birth-and-death process is indecomposable and time reversible, and its unique invariant measure is the process with density p (cf. Ripley, 1979).

For a Markov (nnMp) object process the above expression is typically easy to evaluate since the normalizing constant α is eliminated. Thus, a simulated realization of a Markov object process can be obtained by running a corresponding spatial birth-and-death process for long enough.

To simulate the birth-death process, we generate the successive states $X^{(k)}$ and the sojourn times $T^{(k)}$ as follows. Given $X^{(k)} = \mathbf{x}^{(k)}$, $T^{(k)}$ is exponentially distributed with mean $1/(n(\mathbf{x}^{(k)}) + B(\mathbf{x}^{(k)}))$, independent of other sojourn times and of past states. The next state transition is a death with probability $n(\mathbf{x}^{(k)})/(n(\mathbf{x}^{(k)}) + B(\mathbf{x}^{(k)}))$, obtained by deleting one of the existing points with equal probability; otherwise the transition is a birth generated by choosing one of the points $u_j \notin \mathbf{x}^{(k)}$ with probability density

$$\frac{b_H(\mathbf{x}^{(k)}, u_j)}{B(\mathbf{x}^{(k)})}$$

and adding u_j to the state. Running this for a 'large' time period C we take $\mathbf{x}^{(K)}$, where

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

For a discussion of the rate of convergence see Møller (1989).

4 Bayesian object recognition

The Bayesian approach to object recognition is formally identical to the case of image segmentation. We first assign a prior probability distribution to object configurations $\mathbf{x}$. The prior shall be one of the nearest-neighbour Markov object processes described in the previous section, with density $p(\mathbf{x})$. Given observation of image $\mathbf{y}$, the posterior probability density for scene $\mathbf{x}$ is then

$$p(\mathbf{x} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) \tag{16}$$

which is also a nearest-neighbour Markov object process. The MAP estimator of the true configuration is

$$\hat{x} = \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x} \mid \mathbf{y}) = \operatorname{argmax}_{\mathbf{x}} f(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) \tag{17}$$

The only unusual feature of this formulation is that (17) is now an optimization over the space Ω of all object configurations.

One argument for adopting a Bayesian approach to object recognition is that the MLE tends to contain clusters of almost identical objects, i.e. there is 'multiple response' to each true object, as noted in Section 2.7. This is related to the fact that the Hough transform has rather flat peaks around the correct object positions. Multiple response is undesirable if it is important to determine the number of objects correctly, and if it is believed that objects do not lie extremely close to one another, say, if it is known in advance that objects cannot overlap. A standard approach in computer vision is to select one object per peak of the Hough transform; but this is very similar to a Bayesian approach using a prior model which assigns low probability to configurations in which objects are close to one another.

Assuming $p(\cdot) > 0$ and rewriting (17) as a penalized version of maximum likelihood:

$$\tilde{\mathbf{x}} = \text{argmax}_{\mathbf{x}} [\log f(\mathbf{y} \mid \mathbf{x}) + \log p(\mathbf{x})] \tag{18}$$

we interpret $-\log f(\mathbf{y} \mid \mathbf{x})$ as a measure of goodness-of-fit to the data, and $-\log p(\mathbf{x})$ as a penalty for the complexity of the configuration $\mathbf{x}$.

A simple prior is the Strauss process (12) which results in a penalty of $-\log \beta$ for the presence of each object $x_i \in \mathbf{x}$ and a penalty of $-\log \gamma$ for each pair of neighbouring objects (e.g. overlapping objects). Modifications which might be useful in this application are

$$p(\mathbf{x}) = \alpha \prod_{i=1}^{n} \beta^{|R(x_i)|} \prod_{i<j} \gamma^{|R(x_i) \cap R(x_j)|} \tag{19}$$

and, for marked objects, to allow the interaction terms to depend on the marks.

Regarding the choice of parameter values, note that if the raster is made finer (say, quadrupling the number of pixels), then the log likelihood typically increases by the same factor. This suggests that to maintain the balance between f and p in (17)–(18) the parameters $\log \beta$ and $\log \gamma$ of a Strauss model should also be multiplied by this factor. Models such as (19) and (15), with interactions expressed in terms of pixel counts, do not require such adjustment.

4.1 ICM for object recognition

Iterative methods are needed in order to find the MAP estimator (17). As in Section 2.5, we consider algorithms which add or delete objects one at a time. The log-likelihood ratio criterion is now replaced by a log posterior probability ratio, so that we add object u to configuration $\mathbf{x}$ if

$$\log \frac{f(\mathbf{y} \mid \mathbf{x} \cup \{u\}) p(\mathbf{x} \cup \{u\})}{f(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})} > 0 \tag{20}$$

For example, taking the Strauss prior (12) and a blur-free signal,

$$\log \frac{f(\mathbf{y} \mid \mathbf{x} \cup \{u\}) p(\mathbf{x} \cup \{u\})}{f(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x})} = \log \beta + r(u; \mathbf{x}) \log \gamma + \sum_{R(u) \setminus S(\mathbf{x})} z_t$$

where $r(u; \mathbf{u}) = s(\mathbf{x} \cup \{u\}) - s(\mathbf{x})$ is the number of neighbours of u in $\mathbf{x}$. The new term in γ is a penalty against adding an object in the vicinity of existing objects. For the area-interaction prior (15) the term involving γ is replaced by $|R(u) \setminus S(\mathbf{x})| \log \delta$, so that we again obtain something very similar to the Hough transform.

Algorithm 1 is now a close analogue of Besag's ICM algorithm (Chapter 5). We scan the (discretized) object space U and add an object u whenever the log posterior probability ratio (20) is positive. Equivalently, we add an object u whenever the posterior conditional probability that $u \in \mathbf{x}$, given all other information on $U \setminus \{u\}$, is greater than $\frac{1}{2}$.

Algorithm 2 is a simple variant of ICM, at least in the discrete case. This algorithm is also defined when U is 'continuous' (any lsc space) but the interpretation is more complex. We add a new object u at the position where the Papangelou conditional intensity of the posterior distribution, given the current configuration $\mathbf{x}$ on $U \setminus \{u\}$, is maximal, provided this is greater than e^w (relative to the reference measure μ).

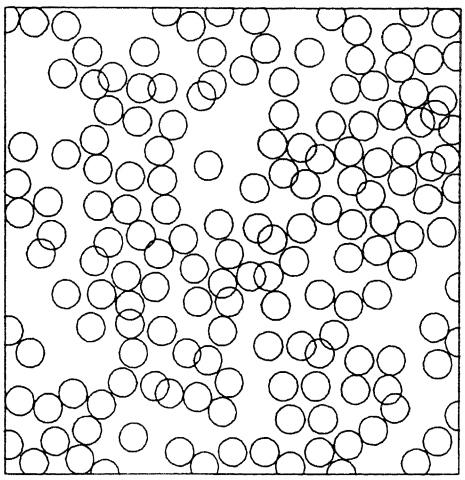

FIG. 3. Approximate MAP reconstruction (steepest ascent) of Brodatz pellets texture.

Figure 3 shows the result of steepest ascent ICM for the pellet texture of Section 2.7. The prior was a Strauss process of overlapping discs (12) with $\log \beta = \log \gamma = -1000$.

4.2 Sampling from the posterior

It is possible to draw samples from the posterior distribution (16) since this is a Markov object process.

Consider any blur-free independent noise model with $g(\cdot \mid \cdot) > 0$ and a nearest-neighbour Markov object prior. Positivity of g is needed so that the class $K = \{\mathbf{x}: f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) > 0\}$ is hereditary; that is, if a configuration $\mathbf{x}$ belongs to K the same is true for all its subsets $\mathbf{y} \subseteq \mathbf{x}$. For some fixed $k \in [0, 1]$ set

$$b(\mathbf{x}, u) = \begin{cases} \left(\dfrac{f(\mathbf{y} \mid \mathbf{x} \cup \{u\})p(\mathbf{x} \cup \{u\})}{f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}\right)^k, & \text{if } f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) > 0 \\ 0, & \text{if } f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) = 0 \end{cases} \tag{21}$$

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

$$D(\mathbf{x}\backslash\{u\}, u) = \begin{cases} \left(\dfrac{f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{f(\mathbf{y} \mid \mathbf{x}\backslash\{u\})p(\mathbf{x}\backslash\{u\})}\right)^{k-1}, & \text{if } f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) > 0 \\ \delta_n'/n, & \text{if } f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) = 0, n(\mathbf{x}) = n \end{cases} \tag{22}$$

Here $\delta_n' = \inf\{\Sigma_{x_i \in \mathbf{x}} D(\mathbf{x}\backslash\{x_i\}, x_i) \mid f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) > 0, n(\mathbf{x}) = n\}$. 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 1. The boundary cases $k = 0$ ('constant birth rate') and $k = 1$ ('constant death rate') are well known in spatial statistics. It is widely argued (e.g. Ripley, 1977) 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 each application, one should verify that the process just described is well defined. For instance, the following corollary of Theorem 1 holds (van Lieshout, 1993).

Lemma 1. Let $\mathbf{y}$ 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

$$\frac{p(\mathbf{x} \cup \{u\})}{p(\mathbf{x})} \leq \beta < \infty,$$

there exists a unique spatial birth-and-death process for which (21) and (22) are the transition rates. The process has unique equilibrium distribution p and converges in distribution to p from any initial state.

Figure 4 shows a sample from the posterior distribution for the pellets texture of Section 2.7 using the same Strauss prior as for Fig. 3.

The main advantage of sampling from the posterior distribution is the ability to estimate any functional of the posterior 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; the distribution of the distance from a given reference point to the nearest object and the first-order intensity (Stoyan *et al.*, 1987). In the discrete case the first-order intensity at u is simply the (posterior) probability that u belongs to $\mathbf{x}$. It can be regarded as an alternative to the Hough transform.

4.3 Stochastic annealing

A MAP solution can also be found by simulated annealing. Assume the conditions of Lemma 1. For $H > 0$ define

$$p_H(\mathbf{x} \mid \mathbf{y}) \propto \{f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})\}^{1/H}$$

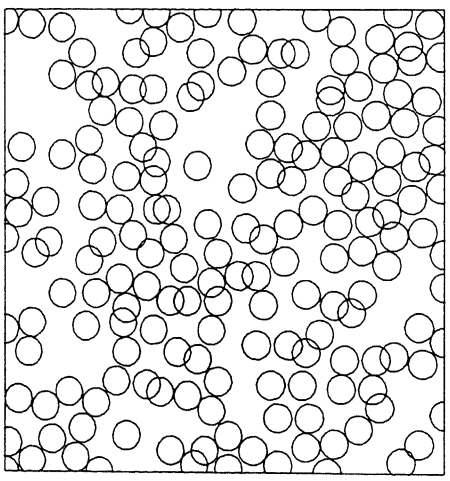

FIG. 4. Sample from posterior distribution for Brodatz pellets texture.

This is the density of a nearest-neighbour Markov object process, and the associated spatial birth-and-death process with rates (21)–(22) exists and converges in distribution to p_H.

As for discrete Markov random fields, H has the interpretation of 'temperature'. If U is discrete, then $p_H(\cdot \mid \mathbf{y})$ converges pointwise as $H \to 0$ to a uniform distribution on the set of MAP solutions.

Take a sequence $H_n \searrow 0$ and consider the corresponding family $(X^{(n)})_{n \in \mathbb{N}}$ of spatial birth–death processes on $K = \{\mathbf{x}: f(\mathbf{y} \mid \mathbf{x})p(\mathbf{x}) > 0\}$. Let $t_n, n \in \mathbb{N}$ be a sequence satisfying

$$t_n \geq t_0\left(1 + \frac{\log\left(\frac{1}{2}\left(1 - \frac{1}{n}\right)\right)}{\log\left(1 - K_n(t_0)\right)}\right)$$

where $K_n(t_0)$ is a certain constant determining the rate of convergence of the nth birth–death process (van Lieshout, 1993). Construct a time-inhomogeneous Markov process X_t, $t > 0$ whose transition rates are those of $X^{(n)}$ during time interval $[s_n, s_{n+1})$ where $s_n = t_0 + \cdots + t_{n-1}$. It is shown in van Lieshout (1993) that if the set of MAP solutions has positive reference measure, and condition (a) in Theorem 1 is satisfied, the sequence of birth–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. For details, see van Lieshout (1993).

For the Brodatz pellet texture of Fig. 1, the results of simulated annealing are very similar to those of posterior sampling (Fig. 4).

The simulation algorithm described in Section 3.5 encounters some numerical difficulties. Since the birth rate is an exponential function of the Hough transform, it may have sharp peaks when H is small or $\mathbf{x}$ is suboptimal. Rejection sampling then becomes impractical because the acceptance probabilities are small.

In the extreme case where H is very close to zero, the required birth–death process behaves like the deterministic Algorithm 2. This suggests using an algorithm which incorporates a search operation over U.

One simple algorithm of this kind is to find the global maximum of birth rate $b^\star = \max_u b_H(\mathbf{x}, u)$, then locate all objects u satisfying $b_H(\mathbf{x}, u) \geq ab^\star$, where $a < 1$. Making a list of all such candidates, we proceed as if these are the only objects in U, computing the total birth rate of the candidates and performing rejection sampling. After each transition the list of candidates has to be recomputed. This algorithm is an approximation to the desired birth–death process: larger values of a increase the speed but decrease the accuracy of the approximation.

It is clear that, whatever strategy is adopted, there will be problems with the 'curse of dimension', i.e. as the dimension of the object space U increases, the cost of searching U increases exponentially. This problem is well known in the context of Hough transforms; it is often named as the major limitation on their performance, and multi-resolution strategies are usually recommended.

We have used a multi-resolution algorithm for simulated annealing. For each 'resolution level' $m = 0, 1, \ldots, M$ conceptually divide object space U into a partition $\mathcal{U}_m = \{U_{1,m}, \ldots, U_{k_m,m}\}$ such that each partition is a refinement of the previous one: $U_{i,m} = \bigcup_j U_{\ell_j, m+1}$ for all i, m. 'Conceptually' means that the subdivision of a block U_{im} into smaller blocks is carried out only when needed. The standard example is the quad tree in which the unit square is divided into $2^m \times 2^m$ smaller squares at stage m.

Interpret each partition $\mathcal{U}_m$ as a class of 'large objects' in T by defining, for any $V \subseteq U$,

$$R(V) = \bigcup_{u \in V} R(u)$$

Define a Hough transform on $\mathcal{U}_m$

$$H_{\mathbf{w}}^{(m)}(V) = \sum_{t \in R(V)} w_t \qquad V \in \mathcal{U}_m$$

where $\mathbf{w}$ is any image. This provides an upper bound for the Hough transform on U, provided $w_t \geq 0$:

$$H_{\mathbf{w}}^{(m)}(V) \geq \max_{u \in V} H_{\mathbf{w}}(u)$$

Furthermore, the Hough transform at any level m is an upper bound for the Hough transform at level $m+1$:

$$H_{\mathbf{w}}^{(m)}(U_{im}) \geq \max_{U_{j,m+1} \subset U_{im}} H_{\mathbf{w}}^{(m+1)}(U_{j,m+1}) \tag{23}$$

Suppose we can find an image $\mathbf{w}$, possibly depending on $\mathbf{x}$, such that

$$w_t \geq \max_{u \in U} h(y_t, \theta^{(\mathbf{x})}(t), \theta^{(\mathbf{x} \cup \{u\})}(t))$$

where h is as in (8). For example, in the blur-free case, we can take

$$w_t = 1\{t \notin S(\mathbf{x})\}(h(y_t, \theta_0, \theta_1))_+$$

where $(a)_+ = \max\{a, 0\}$. Assuming the conditions of Lemma 1 hold, we obtain a decreasing sequence of upper bounds on the birth rates of the stochastic annealing procedure.

$$b(\mathbf{x}, u) \leq b^{(M)}(\mathbf{x}, U_{i_M,M}) \leq \cdots \leq b^{(0)}(\mathbf{x}, U_{i_0,0}) \tag{24}$$

where $u \in U_{i_M,M} \subset \cdots \subset U_{i_0,0}$ and

$$b^{(m)}(\mathbf{x}, V) = [\beta \exp\{H_{\mathbf{w}}^{(m)}(V)\}]^{1/H}$$

where β is as in Lemma 1. The maximum birth rate, and the set of locations where the birth rate is near to maximum, can then be determined by searching the multi-resolution space dynamically in the usual way. This method bears many resemblances to the adaptive Hough transform (Illingworth & Kittler, 1987).

An application of this technique is shown in Fig. 5. This is a synthetic example modelled on the problem of identifying linear features such as long crystals and fission tracks in micrographs of minerals. The objects are line segments of variable length, so that the object space U is four-dimensional, making simple search methods computationally expensive.

The signal is assumed to be a function of the distance $d(t, \mathbf{x})$ from t to $S(\mathbf{x})$, in this case linear decay

$$\theta^{(\mathbf{x})}(t) = \begin{cases} \theta_1 + \dfrac{d(t, \mathbf{x})}{c}(\theta_0 - \theta_1), & \text{for } d(t, \mathbf{x}) \leq c; \\ \theta_0, & \text{otherwise} \end{cases}$$

Figure 5 is a simulation of this model with (arbitrary) line segments of length ranging between 60 and 70, foreground brightness 100, background brightness 254, decay radius $c = 4$ pixel units, and additive Gaussian noise with variance 9.0.

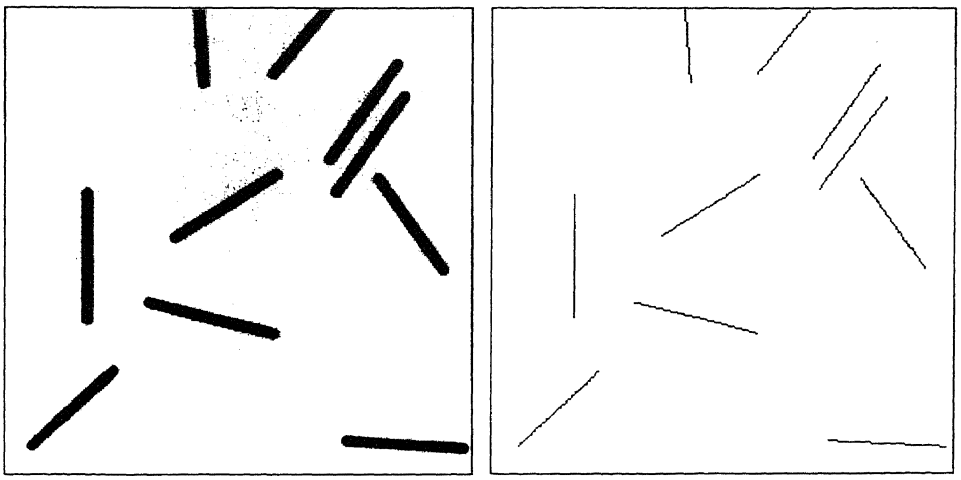

FIG. 5. (a) Synthetic image of blurred line segments, (b) reconstruction by steepest ascent.

The choice of parameterization is important for the computational cost of the multi-resolution algorithm. We choose to parameterize segments by their length ℓ, orientation ρ, and Cartesian coordinates (d, t) of the midpoint of the segment after rotation by $-\phi$, so that d is the distance from the (arbitrary) origin O to the infinite line L containing the segment, and t is the distance from the midpoint of the segment to the foot of the perpendicular OL. Our multi-resolution algorithm splits U at level 0 into blocks of roughly constant (ϕ, d) with a tolerance of 1 degree in ρ and 1 pixel unit in d. That is, we group together all those segments which lie (approximately) along a given infinite line of orientation ρ and distance d from the origin. The objects $R(V)$ at level 0 are thickened lines, whose Hough transform is relatively easy to compute.

We introduce a neighbourhood structure by defining two line segments to be neighbours $u \sim v$ iff their dilations by ball of decay radius have a non-empty intersection. The prior model is a Strauss process with $\beta = 0.002$ and $\gamma = 0.25$. We applied the multi-resolution algorithm with upper bounds (24) computed from $w_t = 1/2\sigma^2\ (y_t - \theta^{(*)}(t))^2$. Note that the choice of parameterization also means that the bounds (24) at level 0 are rather tight. The result of steepest ascent is given in Fig. 5.

5 Spatial clustering and edge detection

There are many similarities between object recognition and the problem of identifying clusters in spatial patterns. We will briefly consider three examples: spatial clustering of point patterns; fitting lines to point patterns; and high-level edge detection.

5.1 Clustering of point patterns

In the simplest case, the data consist of a set of points

$$\mathbf{y} = \{y_1, \ldots, y_m\} \subseteq T$$

where $T \subseteq \mathbb{R}^2$ (say) is the window of observation; and it is required to determine

the locations of an unspecified number of cluster centres $x_1, \ldots, x_n \in T$. Typical applications are the analysis of spatial pattern in the occurrence of rare diseases and the estimation of the positions of ancestors of the current generation of trees in a wood.

The data $\mathbf{y}$ are now a point pattern, but otherwise the structure of the problem is very similar to that in Section 2. Assume $\mathbf{y}$ depends on the unknown pattern $\mathbf{x}$ of cluster centres through a probability density $f(\mathbf{y} \mid \mathbf{x})$. This is the density of a point process with respect to the Poisson point process on T. The analogue of conditional independence is the assumption that, conditional on $\mathbf{x}$, the observed point process is a superposition of independent point processes $N^{(x_i)}$ of 'daughter points' associated with each 'parent point' $x_i \in \mathbf{x}$. Then

$$f(\mathbf{y} \mid \mathbf{x}) = \mathrm{e}^{\mu(T)} \sum_{I_1, \ldots, I_n} \prod_{i=1}^{n} [g(y_{I_i} \mid x_i)\mathrm{e}^{-\mu(T)}] \tag{25}$$

where the sum is over all ordered partitions $I_1, \ldots, I_n$ of $\{1, \ldots, m\}$ into n disjoint sets (allowing $I_i = \emptyset$). Here $y_I = \{y_i : i \in I\}$. The function $g(\cdot \mid x_i)$ is the density of the daughter process $N^{(x_i)}$. Models of this type are frequently considered in spatial statistics (Diggle, 1983; Ripley, 1981; Stoyan *et al.*, 1987) but the present problem is unusual in that we are trying to estimate the parent (cluster centre) process.

It is common in spatial statistics to assume that the daughter processes $N^{(x_i)}$ are identically distributed up to translation:

$$N^{(x_i)} = N^{(0)} + x_i$$
$$g(\mathbf{z} \mid x_i) = g(\mathbf{z} - x_i \mid 0)$$

and that $N^{(0)}$ consists of a random number of points (distribution F) which are iid with density h.

We consider the special case where the number of daughters per parent has a Poisson distribution, so that the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ is an inhomogeneous Poisson process

$$f(\mathbf{y} \mid \mathbf{x}) = \mathrm{e}^{\int_T (1 - \lambda(u))\, \mathrm{d}u} \prod_{j=1}^{m} \lambda(y_j \mid \mathbf{x})$$

with intensity

$$\lambda(u \mid \mathbf{x}) = \sum_{i=1}^{n} h(u - x_i) \tag{26}$$

Then, if the parents also follow a Poisson process, the observed pattern $\mathbf{y}$ is a Neyman–Scott process and is a special case of the well-known Cox process. Examples are the Matérn cluster process in which daughter points are uniform in a disc of radius r around the cluster centre,

$$h(u - x_i) = \frac{\mu}{\pi r^2} 1\{\|u - x_i\| \le r\}$$

and the (modified) Thomas process in which the positions follow a circular Gaussian distribution

$$h(u - x_i) = \mu \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\|u - x_i\|^2/2\sigma^2}$$

(ignoring edge effects in both cases).

A novelty is the possibility of zero-likelihood configurations. For example, in the Matérn process there must be a cluster centre within a distance r of each observed point. A maximum likelihood estimator of $\mathbf{x}$ must be a subset of the morphological closing $\mathbf{y}$ with respect to a disc of radius r (i.e. the r-erosion of the r-dilation (Serra, 1982, page 50)).

The likelihood ratio is

$$\frac{f(\mathbf{y} \mid \mathbf{x} \cup \{u\})}{f(\mathbf{y} \mid \mathbf{x})} = \prod_{j=1}^{m} \left[1 + \frac{h(y_j - u)}{\sum_{i=1}^{n} h(y_j - x_i)} \right] \exp\left\{ - \int_U h(v-u)\, dv \right\}$$

which can be compared to the Hough transform of image data, in the sense that each data point y_j votes with variable strength for a cluster centre at point u.

Generally, the maximum likelihood estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ may run into difficulties similar to those encountered in the context of object recognition. If h is smooth and almost flat near its maximum, and the data pattern is 'dense', the maximum likelihood estimate may tend to contain multiple responses to each true cluster point. In the clustering context, this is precisely what we wish to avoid.

Now introduce a prior distribution for $\mathbf{x}$ in the form of a nearest-neighbour Markov point process (Section 2). The posterior probability ratio

$$\frac{p(\mathbf{x} \cup \{u\} \mid \mathbf{y})}{p(\mathbf{x} \mid \mathbf{y})} = \frac{p(\mathbf{x} \cup \{u\})}{p(\mathbf{x})} \prod_{j=1}^{m} \left[1 + \frac{h(y_j - u)}{\sum_{i=1}^{n} h(y_j - x_i)} \right] \exp\left\{ - \int_U h(v-u)\, dv \right\}$$

is easily computable, so that we can use spatial birth-and-death processes to simulate from the posterior. If the function h is continuous and bounded away from zero, and likelihood ratios of p are uniformly bounded, then the above expression is bounded, so the spatial birth–death processes with constant death rate are well defined and converge weakly to the posterior distribution.

It is then possible to estimate functionals of the posterior, such as the distribution of the number of cluster centres, the probability that there is no cluster in a particular region, and the first-order intensity of cluster positions. One can also perform simulated annealing to find the MAP estimator of $\mathbf{x}$. The convergence results of van Lieshout (1993) still apply.

Lawson (1993) has discussed spatial clustering from a very similar viewpoint but suggested using a Poisson process as the prior distribution. In this case, multiple responses will be suppressed if the intensity β of the prior is small. Lawson proposes a different class of Gibbs sampler techniques for simulating from the posterior. The parameters would be the number of cluster centres, their spatial coordinates, the prior intensity and some radial parameter in the Gaussian forward model. The idea is then to sample in turn from the conditional distribution of one parameter, given the others and the data. Approximations are necessary to make the procedure computationally tractable. For a review of the Gibbs sampler, see Besag and Green (1993) and Smith and Roberts (1993).

5.2 Fitting curves to point patterns

In our second example, the data again consist of a point pattern $\mathbf{y} = \{y_1, \ldots, y_m\} \subseteq T$ with $T \subseteq \mathbb{R}^2$ bounded, but the points are believed to lie

close to a curve or curves, possible not contiguous, and the objective is to estimate the curves. An example is the image analysis task of joining a dot pattern into a curvilinear boundary. An application to spatial statistics is the identification of ancient roads or trade routes given information about the location of archaeological finds (pottery, coins) (Stoyan *et al.*, 1987, p. 139).

Let the curves be parameterized by a small number of real parameters and let U be the corresponding parameter space. The true curve pattern is then a configuration $\mathbf{x} = \{x_1, \ldots, x_n\} \subseteq U$. Again, we can assume an independent cluster model in which each curve x_i gives rise to a point pattern $N^{(x_i)}$ and these daughter patterns are conditionally independent given $\mathbf{x}$. Then (25) holds.

It is no longer appropriate to assume that the daughter patterns $N^{(x_i)}$ are equivalent up to translation, since, for example, the expected number of points may well depend on the length of the curve. However, we can assume $N^{(x_i)}$ is Poisson with intensity $h(\cdot \mid x_i)$, so that the observed point pattern is again a Cox process (Stoyan *et al.*, 1987, p. 138). Particular cases of interest would be the analogues of the Matérn and Thomas models in which distance to the cluster centre is replaced by distance to the nearest point on the cluster curve.

The treatment of this problem is formally equivalent to that in the previous section, the only difference being that the cluster parents now belong to a general family of objects U instead of T. The general methods of Section 4 are required to obtain MAP solutions.

5.3 High-level edge detection

Our final example is the high-level vision problem of identifying large-scale edges in a scene using the output of a low-level edge detector. The 'data' $\mathbf{y}$ consist of a pattern of line segments and the objective is to cluster them around a small number of larger line segments.

Let W denote the set of possible outputs of the low-level edge detector. For example, these may be line segments restricted to have unit length ($=1$ pixel width) and orientation which is a multiple of 45°. As usual, U denotes the space of objects we are looking for, which in this case are also line segments, but have unrestricted length and orientation.

Model $\mathbf{y}$ as a superposition of conditionally independent line segment processes $N^{(x_i)}$ associated with each true line segment x_i. Again, it is not reasonable to suppose that all clusters are identically distributed up to translation (25), but we may assume that they are all Poisson so that $\mathbf{y}$ is a Cox line segment process. Typically, the expected number of line segments in $N^{(x_i)}$ depends on the length of x_i.

The MLE and MAP estimators of $\mathbf{x}$ can then be determined using the techniques we have described above. Maximum likelihood is equivalent to maximizing

$$\sum_{j=1}^{m} \log\left(\sum_{i=1}^{n} h(y_j \mid x_i)\right) - \sum_{i=1}^{n} \int_U h(v \mid x_i)\, dv$$

where $h(\cdot \mid x_i)$ is the intensity of $N^{(x_i)}$.

The possible benefits of a prior model for $\mathbf{x}$ include the ability to encourage long lines and continuity between lines, and to penalize lines that cross one another.

6 Discussion

6.1 Evaluation of the approach

The ultimate objective of this study is (a) to understand the mathematical basis of object recognition, (b) to develop object recognition techniques which outperform existing ones in familiar situations and (c) to develop techniques for problems where there are no satisfactory solutions. Regarding (a), we have noted that some popular *ad hoc* methods for object recognition (Hough transform, erosion) are maximum likelihood techniques for simple statistical models. For (b), we have shown that Bayesian methods perform creditably in simple template matching problems. Since classical methods are typically associated with maximum likelihood, it seems plausible that Bayesian methods are generally better, but no detailed comparison with classical methods has yet been carried out. As regards (c), the general formulation is flexible enough to be immediately adaptable to problems such as subpixel resolution of object shapes, stereo vision, pose estimation and motion detection. However, the algorithms put forward are severely limited by the dimensionality of the object space.

6.2 Why Markov?

An unanswered question is why the prior 'should' be a spatial Markov process. One reason is simplification: the Markov property implies that $p(\mathcal{J}\mathbf{x})/p(\mathbf{x})$ is simple to evaluate, where $\mathcal{J}$ is a prescribed operation that changes the state $\mathbf{x}$, such as flipping the value of a single pixel (for discrete MRFs), adding or deleting an object (for Markov object processes) or introducing a new mosaic fragment (for Arak–Surgailis processes (Arak & Surgailis, 1989)). The associated optimization algorithms and Gibbs sampler consist of successive applications of the operations $\mathcal{J}$ chosen deterministically or stochastically, hence they too are easy to implement.

We may also interpret MAP estimation as penalized or regularized maximum likelihood and ask what kinds of penalty terms are desirable. It is interesting to note that the discrete MRF priors used in image segmentation (Chapters 4 and 5) specify positive association between neighbouring pixel values, while the Markov object processes used in this article exhibit negative association or inhibition between neighbouring objects. In the case of spatial clustering, Lawson (1993) has used Poisson priors, in which the points have zero association, and the effect of the prior is simply to penalize configurations with large numbers of points.

Furthermore, it is unnecessary to 'believe' a prior Markov model, in the sense that good solutions may well have low probability under the prior. In the context of discrete MRFs, Besag (Chapter 5) and others have argued that ICM is preferable to stochastic annealing because we do not believe the 'global' predictions of the prior, and are only using 'local' properties in the optimization. Similar remarks hold in the context of object recognition.

6.3 Alternative algorithms

The algorithms presented here are simple choices and do not do full justice to the theory. There are many alternatives.

Firstly, one can expand the range of transitions $\mathcal{J}$ that can be considered at each stage of ICM or Gibbs sampling. Besides the addition and deletion of objects, one may also consider modifications of an existing object, e.g. translations

and rotations by small amounts. These are equivalent to (small) jumps in the object space U. The deterministic optimization algorithms necessarily converge to 'better' solutions (in the sense of the objective function). The associated stochastic algorithms simulate an interacting particle system in which the existing objects execute (dependent) random walks in the object space as well as the usual births and deaths. In stochastic annealing one expects that, near convergence and at low temperatures the birth and death rates are very low compared to the diffusion rate, so that the number of objects eventually becomes fixed and the existing objects follow random walks converging in distribution to their optimal positions. Miller *et al.* (1991) have described a related class of jump diffusions in which the existing objects execute diffusions in the object space.

Secondly, where simulation is concerned, there are alternatives to rejection sampling of a birth and death process. Geyer and Møller have described a version of the Metropolis–Hastings algorithm suited to this context. 'Auxiliary variable methods' such as the Swendsen–Wang algorithm (Besag & Green, 1993) could also be used to accelerate convergence and inspect the posterior surface more efficiently. Low-temperature sampling is a good compromise between stochastic annealing and ICM.

6.4 Stochastic geometry and image analysis

Finally, it is interesting to note that the standard parameterization of lines in $\mathbb{R}^2$ in stochastic geometry is identical to that used in the Hough transform for straight lines (Hough, 1962; Illingworth & Kittler, 1988). In stochastic geometry, this choice is dictated by properties of the 'invariant' measure for random lines. One may speculate that stochastic geometry can help to suggest the right parameterization for the Hough transform for other classes of geometrical objects, and, in general, suggest the right way to formulate many problems in image analysis.

Acknowledgement

We are grateful to Dr A Lawson for discussions and preprints of his work.

Correspondence: M. N. M. van Lieshout, Centre for Mathematics and Computer Science (CWI), PO Box 94079, NL-1090 GB Amsterdam, The Netherlands.

REFERENCES

ARAK, R. J. & SURGAILIS, D. (1989) Markov random fields with polygonal realizations, *Probability Theory and Related Fields*, 80, pp. 543–579.

BADDELEY, A. J. & MØLLER, J. (1989) Nearest-neighbour Markov point processes and random sets, *International Statistical Review*, 57, pp. 89–121.

BADDELEY, A. J. & VAN LIESHOUT, M. N. M. (1991) Recognition of overlapping objects using Markov spatial processes. Research Report BS-R9109, Centrum voor Wiskunde en Informatica, Amsterdam, March.

BADDELEY, A. J. & VAN LIESHOUT, M. N. M. (1992a) ICM for object recognition. *In*: Y. DODGE & J. WHITTAKER (Eds), *Computational Statistics*, Vol. 2, pp. 271–286 (Heidelberg and New York, Physica/Springer).

BADDELEY, A. J. & VAN LIESHOUT, M. N. M. (1992b) Object recognition using Markov spatial processes. *In*: *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, pp. B136–139 (Los Alamitos, California, IEEE Computer Society Press).

BESAG, J. & GREEN, P. J. (1993) Spatial statistics and Bayesian computation, *Journal of the Royal Statistical Society, Series B*, 55, pp. 25–37.

BLOKLAND, J. A. K. (1987) Detection of elliptical contours: an application in the field of quantitative nuclear cardiology. PhD thesis, University of Leiden, The Netherlands.

BRODATZ, P. (1966) *Texture: a Photographic Album for Artists and Designers* (New York, Dover).

DALEY, D. J. & VERE-JONES, D. (1988) *An Introduction to the Theory of Point Processes* (New York, Springer-Verlag).

DAVIES, E. R. (1990) *Machine Vision: Theory Algorithms, Practicalities* (London, Academic Press).

DENGLER, J. & GUCKES, M. (1992) Estimating a global shape model for objects with badly-defined boundaries. *In*: W. FÖRSTNER & S. RUWIEDEL (Eds), *Robust Computer Vision*, pp. 125–136 (Karlsruhe, Wichmann).

DIGGLE, P. J. (1983) *Statistical Analysis of Spatial Point Patterns* (London, Academic Press).

GEMAN, D. (1990) Random fields and inverse problems in imaging. *In*: *Ecole d'été de probabilités de Saint-Fleur XVIII, 1988*, Vol. 1427 of Lecture Notes in Mathematics (Berlin, Springer-Verlag).

GEYER, C. J. & MØLLER, J. (1993) Simulation procedures and likelihood inference for spatial point processes. Research Report 260, Mathematical Institute, University of Aarhus, January.

GRENANDER, U. (1976) *Lectures on Pattern Theory*, Vol. 1: *Pattern Synthesis*, Applied Mathematical Sciences, Vol. 18 (New York–Berlin, Springer-Verlag).

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

GRENANDER, U. (1981) *Lectures on Pattern Theory*, Vol. 3: *Regular Structures*, Applied Mathematical Sciences, Vol. 33 (New York–Berlin, Springer-Verlag).

HOUGH, P. V. C. (1962) Method and means for recognizing complex patterns. US Patent 3069654.

ILLINGWORTH, J. & KITTLER, J. (1987) The adaptive Hough transform, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9, pp. 690–698.

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

KENDALL, W. S. (1990) A spatial Markov property for nearest-neighbour Markov point processes, *Journal of Applied Probability*, 28, pp. 767–778.

LAWSON, A. (1993) Discussion contribution, *Journal of the Royal Statistical Society, Series B*, 55, pp. 61–62.

VAN LIESHOUT, M. N. M. (1991) A Bayesian approach to object recognition. *In*: U. ECKHARDT, A. HÜBLER, W. NAGEL & G. WERNER (Eds), *Geometrical Problems of Image Processing*, Vol. 4, *Research in Informatics*, pp. 185–190 (Berlin, Akademie Verlag).

VAN LIESHOUT, M. N. M. (1993) Stochastic annealing for nearest-neighbour point processes with application to object recognition. Technical Report BS-9306, Centrum voor Wiskunde en Informatica, Amsterdam, March.

LIN, J.-H., SELKE, T. M. & COYLE, E. J. (1990) Adaptive stack filtering under the mean absolute error criterion, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38, pp. 938–954.

MARAGOS, P. (1988) Optimal morphological approaches to image matching and object detection. *In*: *Proceedings of the IEEE International Conference on Computer Vision 1988*, Tampa, Florida, pp. 659–699.

MATHERON, G. (1975) *Random Sets and Integral Geometry* (New York, Wiley).

MILLER, M. *et al.* (1991) Automated segmentation of biological shapes in electron microscopic autoradiography. *In*: F. DAVIDSON & J. GOUTSIAS (Eds), *Proceedings of the 25th Annual Conference on Information Sciences and Systems*, pp. 637–642.

MØLLER, J. (1989) On the rate of convergence of spatial birth-and-death processes, *Annals of the Institute of Statistical Mathematics*, 41, pp. 565–581.

OGATA, Y. & TANEMURA, M. (1984) Likelihood analysis of spatial point patterns, *Journal of the Royal Statistical Society, Series B*, 46, pp. 496–518.

PRESTON, C. J. (1977) Spatial birth-and-death processes, *Bulletin of the International Statistical Institute*, 46, pp. 371–391.

RIPLEY, B. D. (1977) Modelling spatial patterns (with discussion), *Journal of the Royal Statistical Society, Series B*, 39, pp. 172–212.

RIPLEY, B. D. (1979) On tests of randomness for spatial point patterns, *Journal of the Royal Statistical Society, Series B*, 41, pp. 368–374.

RIPLEY, B. D. (1981) *Spatial Statistics* (Wiley, New York).

RIPLEY, B. D. (1988) *Statistical Inference for Spatial Processes* (Cambridge, Cambridge University Press).

RIPLEY, B. D. (1989) Gibbsian interaction models. *In*: D. A. GRIFFITHS (Eds), *Spatial Statistics: Past, Present and Future*, pp. 1–19 (New York, Image).

RIPLEY, B. D. (1991) The use of spatial models as image priors. *In*: A. POSSOLO (Eds), *Spatial Statistics and Imaging*, Number 20 in Lecture notes—monographs, pp. 309–340. Institute of Mathematical Statistics, Hayward, California

RIPLEY, B. D. & KELLY, F. P. (1977) Markov point processes, *Journal of the London Mathematical Society*, 15, pp. 188–192.

RIPLEY, B. D. & SUTHERLAND, A. I. (1990) Finding spiral structures in images of galaxies, *Philosophical Transactions of the Royal Society of London, Series A*, 332, pp. 477–485.

SERRA, J. (1982) *Image Analysis and Mathematical Morphology* (London, Academic Press).

SERRA, J. (ed.) (1988) *Image Analysis and Mathematical Morphology*, Vol. 2: *Theoretical Advances* (London, Academic Press).

SMITH, A. F. M. & ROBERTS, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society, Series B*, 55, pp. 3–23.

STOYAN, D., KENDALL, W. S. & MECKE, J. (1987) *Stochastic Geometry and its Applications* (Chichester, Wiley).