Chapter Twelve

An application of marked point processes to the extraction of linear networks from images

R.S. Stoica[1], M.N.M. van Lieshout[1], X. Descombes[2] and J. Zerubia[2]

[1]*CWI, PO Box 94079, 1090 GB Amsterdam, The Netherlands.*
[2]*Ariana, CNRS/UNSA/INRIA joint research group. PO Box 93, 06902 Sophia Antipolis Cedex, France*

1 Introduction

In this chapter, we show how marked point processes can be used to tackle some high-level image analysis problems. Hereafter, an image refers to a digital image. It consists of a matrix of numbers; the elements are called pixels (for picture element), their values represent a grey level (typically an integer between 0 and 255). Therefore, an image is an element of Λ^S, where Λ is the grey level set and S is a finite subset of $\mathbb{Z}$ (typically a square) called the lattice. The purpose of image analysis is to extract some specific items from the image. The items can have a cartographic interpretation (roads, buildings, ...), a medical relevance (tumours, vessels, ...) or may refer to a lower level of interpretation (edges, homogeneous areas, ...).

One main approach in image analysis is to model the features of interest, and to fit this model to the data. The discrete nature of digital images suggests discrete models, where the variables are defined by the pixels. If one chooses to consider a probabilistic model, it is natural to adopt a Bayesian framework that takes into account both the data and some prior model on the features to be extracted. Typically, the prior is based on homogeneity constraints (the connected components in the resulting image are 'big enough' and have a 'regular' shape), the data term models the grey level distribution within each object. The resulting model, defined on Λ^S, is a random field. Focusing on the correlation of pixel values

leads to so-called Markov random fields (MRF), also known as Gibbs fields in the statistical physics literature, which have played a leading role in Bayesian image analysis [7, 8, 9, 11, 22, 21, 30, 37, 42, 77]. The Markov property states that local interactions suffice to describe the random field. This property is well adapted to images, as neighbouring pixels tend to belong to the same object, hence exhibit strong correlation. The main advantages of this approach are its robustness with respect to noise, due to the prior, and its local interaction structure that is well suited to iterative feature extraction, and parameter estimation algorithms (cf. e.g. [13, 17] as well as the previously mentioned seminal books and papers).

The random field models just described have been applied to a range of image interpretation tasks, especially in the context of remote sensing (by satellites for example) at low or middle range resolutions. Here, a pixel represents a square of 10×10 metres to a square kilometre on the ground. The new generation of sensors provides high resolution images (one metre or less). At such a level of resolution, the object geometry becomes important, and should be used to increase the accuracy of feature detection. MRFs and more generally pixel based models are not well-adapted to high resolution data. Indeed, the notion of object is difficult to model within this framework, and geometric constraints are hard to impose by means of local interactions only. Therefore, the last decade has witnessed increasing interest in other types of models. Still on the pixel-level, Markov connected component fields [49] aim to control the area and perimeter of the image components. Alternatively, the size and shape of regions may be held in check by means of morphological operators [10, 69, 68]. Multi-body and higher order interactions allow control of the length of edges and lines, or some isotropy of objects [16, 14, 66, 74]. MRFs on graphs have also been proposed [2, 75]. The main drawback is the graph definition. If objects are represented by the graph nodes, all of them should have already been detected during the first step.

Polygonal field models can be seen as a first step away from pixel-level modelling, in that they provide a concise description of the image in terms of a polygonal segmentation [1, 48, 51].

In this chapter, following [3, 4, 38, 39, 46, 59, 61] we propose to take the high-level approach, and no longer regard an image as a pixel matrix but rather as a collection of objects defined on a continuous space. The data are considered as a projection of the object silhouette onto the discrete lattice S. The objects may be simple geometrical shapes such as segments [71] or rectangles, but could also be modelled as a random closed set [45] or a deformable template [32, 33, 34, 53, 63, 64, 65]. In all cases, the object locations constitute a point process. The density is controlled by the reference Poisson process whereas interaction between neighbouring objects controls their relative positions. The object parameters (length, width, ...) are random variables representing the marks of the point process. Occlu-

sion may be formalised by ordering scenes according to which objects lie on top of others [43]. The main advantage of such a model is that the geometry of the objects is constrained by the marks. Moreover, the data are taken into account at the object level, which is even more robust with respect to noise than pixel based models.

In section 2, we give an overview of marked point process theory with a view towards image analysis. In section 3, we discuss the computational tools needed later on, including simulation by Markov chain Monte Carlo and parameter estimation. Section 4 is devoted to the problem of extracting linear structures from images. We review the Candy model [71], and discuss how it may be applied [70, 72] to detect road and river networks in satellite images. The paper is concluded with a critical evaluation and prospects for future work.

2 Marked point processes

As mentioned in section 1, in higher level vision, an image is interpreted as a collection of objects. A convenient mathematical formalisation of such a viewpoint is to model the image as an object process [3, 4, 38]. More specifically, we will describe an image by the set

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

of objects it contains. Here $n = 0, 1, 2, \ldots$ denotes the number $n(\mathbf{x})$ of objects in $\mathbf{x}$, and each x_i lies in some space $\mathcal{X}$ determined by the application. If all objects are identical, $\mathcal{X} = K \subset \mathbb{R}^2$ is some compact subset of the plane, and a point $x \in \mathcal{X}$ describes the location of the object in the image; more often, a mark describing the shape, the texture or colour, and in cases of several kinds of objects, its type is required as well. In that case, $\mathcal{X} = K \times M$ where M contains the geometrical and textural marks. Throughout this chapter, we assume that M is a locally compact, second countable Hausdorff space equipped with its Borel σ-algebra $\mathcal{M}$. Clearly, K has the same topological properties. The link to the data is made by assigning to each object $x \in \mathcal{X}$ its projection $R(x) \subseteq S$ in the image; similarly, the projection of an object configuration $\mathbf{x}$ is denoted by $R(\mathbf{x})$. It will prove convenient to separate location information from the marks, so henceforth we shall write $x = (k, m)$, where $k \in K$ and $m \in M$.

In order to define a probabilistic model for random object configurations, we need to provide distributions for the number of objects, as well as for their positions and marks. Thus, let

$$q_n = P(N = n), \qquad n \in \mathbb{N}_0,$$

denote the mass function for the number, and, given that n objects are present, write

$$p_n(x_1, \ldots, x_n)$$

for the joint density of the objects. Since object configurations are unordered sets, $p_n(\cdot, \ldots, \cdot)$ must be a symmetric, measurable function. As an example, suppose $\mu_K(\cdot)$ is a finite, diffuse Borel measure on K, and $\mu_M(\cdot)$ a probability measure on $(M, \mathcal{M})$ (often, $\mu_K(\cdot)$ is the Lebesgue measure and $\mu_M(\cdot)$ the uniform distribution on M, but spatial or geometric non-homogeneity may be taken into account, as in the case study later). Then

$$q_n = e^{-\mu_K(K)} \mu_K(K)^n / n!$$

$$p_n(x_1, \ldots, x_n) = \mu_K(K)^{-n}$$

defines a Poisson process on K with intensity measure $\mu_K(\cdot)$ marked independently according to $\mu_M(\cdot)$. The dominating measure for $p_n(\cdot, \ldots, \cdot)$ is the n-fold product measure $(\mu_K \times \mu_M)^n(\cdot)$ on $\mathcal{X}^n$.

The Poisson process distribution π is a convenient benchmark because of its lack of interactions. More complicated models may be defined by a probability density, say $p(\cdot)$, with respect to π, in which case

$$q_n = \frac{e^{-\mu_K(K)}}{n!} \int_{\mathcal{X}^n} p(\{x_1, \ldots, x_n\}) d\mu_K \times \mu_M(x_1) \cdots d\mu_K \times \mu_M(x_n)$$

and, given $N = n$, the n random objects have joint probability density

$$p_n(x_1, \ldots, x_n) = e^{-\mu_K(K)} p(\{x_1, \ldots, x_n\}) / (n! q_n)$$

with respect to the n-fold product measure of $\mu_K \times \mu_M$. Clearly, since $p(\cdot)$ is permutation invariant, so is $p_n(\cdot, \ldots, \cdot)$. For more details on measurability and the topological structure of object configuration spaces, see [12, 57].

The intuitive concept of interaction can be formalised by means of a symmetric relation $\sim$, cf. [5, 60]. Thus, two objects x and y interact if and only if $x \sim y$, in which case they are called neighbours. The neighbourhood of an object $x \in \mathcal{X}$ is denoted by $\partial(\{x\}) = \{y \in \mathcal{X} : x \sim y\}$. Note that the interaction relation $\sim$ may depend on the marks, but need not. For instance, under the fixed range relation

$$(k, m) \sim (l, n) \Leftrightarrow ||k - l|| \leq r$$

two objects are neighbours whenever their locations are not further than r apart. Especially in the case of objects that may vary in shape, the overlapping object

relation [3]

$$x \sim y \Leftrightarrow R(x) \cap R(y) \neq \emptyset$$

seems more natural. However, specific tasks may call for more tailor-made interaction relations, as for example those that define the Candy line segment model [71] used in the case study in section 4.

Hard object processes are obtained by simply forbidding objects to overlap. Conditioning the Poisson model to this event yields the density

$$p(\{x_1, \ldots, x_n\}) = \begin{cases} 0 & \text{if } R(x_i) \cap R(x_j) \neq \emptyset \quad \text{for some } i \neq j \\ \alpha & \text{otherwise} \end{cases} \tag{106}$$

The normalising constant $\alpha > 0$ ensures the density integrates to unity. Note that the hard object density is hereditary in the sense that if $p(\mathbf{x}) > 0$ for some configuration $\mathbf{x}$ of objects, then also $p(\mathbf{y}) > 0$ for any $\mathbf{y} \subseteq \mathbf{x}$ as whenever no pair of objects in $\mathbf{x}$ overlap each other, the same is true for those objects belonging to $\mathbf{y}$. Moreover, provided $p(\mathbf{x})$ is non-zero, the ratio

$$\frac{p(\mathbf{x} \cup \{\eta\})}{p(\mathbf{x})} = 1\{R(\eta) \cap R(x_i) = \emptyset \text{ for all } x_i \in \mathbf{x}\} \tag{107}$$

depends only on the neighbours of η in $\mathbf{x}$. This observation, simple though it is, has important practical implications, since it means that for any iterative procedure based on small changes to a current configuration, such as adding, deleting or changing an object, only the local neighbourhood has to be taken into account.

Any object process exhibiting the two properties mentioned above (hereditary density, and a likelihood ratio (107) depending only on the neighbours of the added object) is called a Markov object process [3, 4, 38, 39], in accordance with the definition of Markov point processes [5, 60]. By the Hammersley–Clifford theorem, such processes can be characterised by the fact that their density factorises as a product over subsets of interacting objects of functions that describe the attraction or repulsion between these objects. See [60], [5], or [40] for more details. It should be noted that in the physics literature, the term 'Gibbs' [67] is used instead of 'Markov', and an exponential notation of the form

$$p(\mathbf{x}) = p(\emptyset) \exp[-U(\mathbf{x})] \tag{108}$$

is preferred, at least when $p(\cdot)$ is strictly positive.

The Hammersley-Clifford factorisation amounts to expressing the energy function $U(\mathbf{x})$ as a sum of potentials over those subsets of $\mathbf{x}$ that consist of pairwise interacting objects.

3 Tools for manipulating the point processes

The density $p(\cdot)$ of object processes typically contains a normalising constant, for instance α in the hard object model (106), that cannot be calculated explicitly. Moreover, the number of objects is allowed to vary, so that direct sampling from such a model is usually out of the question. On the other hand, the local characteristics (107) are, at least for Markov object process, both low-dimensional and easy to compute. These observations suggest that Markov object processes may be well suited to simulation by Markov chain Monte Carlo. The idea is to run a Markov chain with the density we wish to sample from an equilibrium distribution for a sufficiently long time to be justified in considering the final state of the chain as an approximate realisation of the desired object process. Of course, there are many such Markov chains, the art is to select one that converges fast and is easy to implement. Below, we will describe several well-known types of chains that can be applied quite generally. Nevertheless, huge improvements are possible by tailoring the chain to the model one wants to sample from.

3.1 Spatial birth-and-death processes

The simple form of likelihood ratios such as (107) suggests we build a Markov process whose transitions are births and deaths. If we adopt a continuous time formulation, let $b(\mathbf{x}, u)$ denote the birth rate for a transition from $\mathbf{x}$ to $\mathbf{x} \cup \{u\}$, and $d(\mathbf{x} \setminus \{x_i\}, x_i)$ the death rate for deleting x_i from $\mathbf{x}$. Intuitively speaking, the probability of a birth at $d\eta$ during the time interval $(t, t + dt)$ is $b(x, \eta)dtd\eta$; similarly the probability of deleting $x_i \in \mathbf{x}$ during the same time interval is $d(\mathbf{x} \setminus \{x_i\}, x_i)dt$ for $dt \to 0$, and the probability of multiple transitions is ignored. More formally, regarding the birth kernel, for each $\mathbf{x}$, $b(\mathbf{x}, \cdot)$ is a density of a finite Borel measure on $\mathcal{X}$, and for each Borel set F, $B(\mathbf{x}, F) = \int_F b(\mathbf{x}, (k, m))d\mu_K(k)d\mu_M(m)$ is a measurable function of $\mathbf{x}$; $d(\cdot, \cdot)$ is a jointly measurable function.

Conditions must be imposed to rule out explosion, that is an infinite number of transitions in finite time [55, Prop. 5.1, Thm. 7.1], [5, 47]. Write $B(\mathbf{x})$ for the total birth rate, and $D(\mathbf{x})$ for the total death rate from state $\mathbf{x}$, and 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 $n \geq 1$. Then, if there are relatively few births compared to deaths, in the sense that either $\kappa_n = 0$ for all sufficiently large $n \geq 0$, or that $\kappa_n > 0$ for all $n \geq 1$ but both

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

hold, a unique spatial birth-and-death process with the given rates exists; this process has a unique equilibrium distribution to which it converges in distribution

from any initial state. For a discussion of the rate of convergence see [47].

In our context of Markov object processes, if we require that under the equilibrium density $p(\cdot)$ births are matched exactly by deaths, i.e.

$$\frac{b(\mathbf{x},\eta)}{d(\mathbf{x},\eta)} = \frac{p(\mathbf{x}\cup\{\eta\})}{p(\mathbf{x})} \tag{109}$$

whenever $p(\mathbf{x}\cup\{\eta\}) > 0$, then the birth-and-death process is indecomposable and time reversible, with unique invariant density $p(\cdot)$ (cf. [58]). Thus, it suffices to specify the death rate. A common choice is $d(\mathbf{x},\eta) \equiv 1$, which implies the birth rate $b(\mathbf{x},\eta)$ must be equal to the likelihood ratio for adding η to the object configuration $\mathbf{x}$. If this likelihood ratio is uniformly bounded in both arguments, explosion is ruled out [40].

3.2 Metropolis–Hastings and reversible jump processes

An alternative paradigm for simulating point processes is that of Metropolis and Hastings [26]. Again, the transitions are births and deaths, but, in contrast to the spatial birth-and-death processes discussed in section 3.1, this chain proceeds in discrete time. At each step, a proposal is made to either add or delete an object; subsequently, the proposal is accepted or rejected with probabilities designed in such a way as to ensure convergence to the desired equilibrium distribution.

Thus, let p_b be the probability of choosing to propose a birth, and $p_d = 1 - p_b$ the complementary death proposal probability. For each object configuration $\mathbf{x}$, denote by $b(\mathbf{x},\cdot)$ the proposal density (with respect to $\mu_K \times \mu_M$) for adding a new object to the current configuration $\mathbf{x}$, and let $d(\mathbf{x}\setminus\{x_i\}, x_i)$ be the probability for suggesting we delete x_i from $\mathbf{x}$. By the detailed balance equations (109), the acceptance probability for the transition from $\mathbf{x}$ to $\mathbf{x}\cup\{\eta\}$ is

$$\alpha(\mathbf{x}, \mathbf{x}\cup\{\eta\}) = \min\left\{1, \frac{p_d\, d(\mathbf{x},\eta)\, p(\mathbf{x}\cup\{\eta\})}{p_b\, b(\mathbf{x},\eta)\, p(\mathbf{x})}\right\}; \tag{110}$$

a similar expression holds for death transitions. By the results in [26, section 4], the algorithm converges in total variation to the target density $p(\cdot)$ for p-almost all initial configurations provided $p_b \in (0,1)$ and the proposal densities are strictly positive, but in order to obtain the stronger property of geometric ergodicity, we must assume additionally that the likelihood ratio is uniformly bounded, and that relatively few births are accepted in the sense that [41]

$$\kappa_n = \sup_{\eta\in\mathcal{X}, n(\mathbf{x})=n} \frac{d(\mathbf{x},\eta)}{b(\mathbf{x},\eta)} \to 0; \qquad \delta_n = \inf_{\eta\in\mathcal{X}, n(\mathbf{x})=n} \frac{b(\mathbf{x},\eta)}{d(\mathbf{x},\eta)} \to \infty$$

as $n \to \infty$. The reader is urged to compare this assumption with the non-explosion condition in section 3.1. Note that for the generic choice of uniform proposal distributions, geometric ergodicity holds [25, Prop. 3.3].

The mixing properties of a Metropolis–Hastings chain may be significantly increased by including proposals other than births or deaths. For example, change moves may be used which propose to modify object characteristics [26, 52]. For objects whose shapes are modelled by landmarks scattered around the boundary, the mark space dimension depends on the number of such landmarks, and the Metropolis–Hastings algorithm as described above is no longer applicable. Fortunately, a generalisation exists, known as the reversible jump algorithm [31]. The general idea is the same as that underlying the Metropolis–Hastings sampler, but the acceptance probabilities are modified to compensate for changes in dimension. Typical transition types include births and deaths, but also the splitting of an object n two, object mergers, or changes in the boundary shape [34, 53, 63, 64, 65].

3.3 Coupling from the past

Despite the greater flexibility of the Metropolis–Hastings framework compared to spatial birth-and-death processes, the latter are enjoying a renaissance because of their greater suitability to coupling. Indeed, the proofs for existence and convergence in [55] rely on a coupling to simple birth-and-death processes on $\mathbb{N}_0$.

As mentioned at the beginning of this section, the idea behind Markov chain Monte Carlo methods is to run a Markov process into equilibrium. However, since equilibrium is never actually reached, the main problem the user of such a technique has to face is that of deciding when to stop. No stopping criterion is necessary for coupling from the past algorithms [56]. In order to describe the method, for the moment suppose one wishes to simulate a given probability distribution on some finite state space, and imagine running a series of coupled Markov chains (naturally with the target distribution as their equilibrium) from time $-\infty$, one for each possible initial state. If after some time all chains have reached the same state, clearly the initial state influence has worn off, and they proceed as one. The state reached at time 0 would then be an unbiased sample from the equilibrium distribution. Of course it is not possible to run the chain for an infinite amount of time, but we may start the chain further and further into the past, until coalescence occurs at time 0. The art is to find paths that couple quickly and an efficient way of checking for coalescence. This can be done if a partial order on the state space exists, and the transition kernel respects this order, since then only two chains have to be considered, one associated with the maximal, the other with the minimal state. For object processes, the state space is not finite. Nevertheless, spatial birth-and-death processes may be coupled to a randomly evolving maximum in such a way that exact

samples can be obtained, see [35, 36]. For the Metropolis–Hastings dynamics, the construction of a coupling from the past sampler seems more cumbersome [36].

Finally, it should be mentioned that the simulation algorithms discussed so far, although they seem to be the most widely used ones, are by no means the only options. Alternatives include the jump-diffusion processes in [32], simulated tempering [28, 44], block updating or auxiliary variables schemes. The interested reader is referred to [29] for more information.

3.4 Inference

Next, turn to statistical inference in high-level vision. Given a data image $\vec{\mathbf{y}} = (y_s)_{s \in S}$, with each grey level y_s at pixel s taking a value in the set Λ, the goal is to interpret $\vec{\mathbf{y}}$ in terms of the set of objects $\mathbf{x}$ present in it. As in section 2, we assume that the true set of objects is distributed as a Markov object process. Furthermore, we shall suppose that its density $p(\cdot)$ is strictly positive, hence it can be written in the exponential form (108) with energy $U(\cdot)$. The projection of the configuration $\mathbf{x}$ to the pixel lattice we take to be governed by a strictly positive distribution

$$p_S(\vec{\mathbf{y}}; \mathbf{x}) = \alpha \exp\left[-\tilde{U}(\mathbf{x}, \vec{\mathbf{y}})\right].$$

The reference distribution depends on the value space Λ. For the usual grey levels between 0 and 255, it is the counting measure, hence $p_S(\cdot; \mathbf{x})$ is a mass function. In contrast to $p(\cdot)$, $p_S(\vec{\mathbf{y}}; \mathbf{x})$ is usually of a simple form. For instance, for white noise, $p_S(\vec{\mathbf{y}})$ factorises in independent contributions from individual pixels. Usually, both the Markov object process and the projection density will contain parameters, that is $U(\mathbf{x}) = U(\mathbf{x}; \omega)$ and $\tilde{U}(\mathbf{x}, \vec{\mathbf{y}}) = \tilde{U}(\mathbf{x}, \vec{\mathbf{y}}; \tilde{\omega})$ respectively. Combining all ingredients, we shall perform statistical inference based on the Gibbs process described by its density

$$p(\mathbf{x}; \vec{\mathbf{y}}, \omega, \tilde{\omega}) = \frac{\exp\left[-\left(U(\mathbf{x}; \omega) + \tilde{U}(\mathbf{x}, \vec{\mathbf{y}}; \tilde{\omega})\right)\right]}{Z(\omega, \tilde{\omega})} \tag{111}$$

with respect to π, the law of a Poisson process on K with intensity measure $\mu_K(\cdot)$, marked independently according to $\mu_M(\cdot)$, cf. section 2. The energy function is the sum of two terms: $\tilde{U}(\mathbf{x}, \mathbf{y}; \tilde{\omega})$ describes the goodness-of-fit between the object configuration $\mathbf{x}$ and the data $\vec{\mathbf{y}}$, while $U(\mathbf{x})$ is a regularisation energy constraining the geometry of, and the interactions between, the objects.

Three types of 'parameters' occur in (111), first of all, the objects to be extracted, but also the regularisation parameters ω, and the parameters $\tilde{\omega}$ of the projection model. Clearly, each type of parameter plays a different role. The configuration $\mathbf{x}$ is

meant to provide a concise description of the image. The parameters $\tilde{\omega}$ are crucial in linking the description to the observed image, while the vector ω regulates the strength and type of geometric constraints, or 'smoothness' of the description.

Focus first on the feature extraction problem, and suppose both ω and $\tilde{\omega}$ are fixed. Hence, the penalised maximum likelihood estimator $\hat{\mathbf{x}}$ is obtained by minimising the total energy of (111), that is

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left[U(\mathbf{x};\omega) + \tilde{U}(\mathbf{x},\vec{\mathbf{y}};\tilde{\omega}) \right] \tag{112}$$

where the minimum is over all object configurations $\mathbf{x}$. Naturally, for each application, existence and uniqueness of $\hat{\mathbf{x}}$ have to be established. On rare occasions, the optimisation can be done explicitly. However, the simulation methods discussed in the previous sections stand us in good stead here. Indeed, simulated annealing is based on sampling from the family of distributions defined by the rescaled energies $(U(\cdot;\omega) + \tilde{U}(\cdot,\vec{\mathbf{y}};\tilde{\omega}))/T$. For large positive values of T, the state space is explored; subsequently lowering T gradually to 0 forces the sampler towards the states with minimal energy [38, 77].

Next, suppose $\mathbf{x}$ and ω are fixed. Consider estimation of $\tilde{\omega}$ given the image $\vec{\mathbf{y}}$ and the pattern $\mathbf{x}$. Thus, we must maximise the function $\exp[-\tilde{U}(\mathbf{x},\vec{\mathbf{y}};\tilde{\omega})]/Z(\tilde{\omega})$ with respect to $\tilde{\omega}$. As mentioned above, usually this function is relatively simple. Hence the integration needed to compute $Z(\tilde{\omega})$ can be carried out explicitly, and optimisation is straightforward. Furthermore, if in a practical application the parameters represent known characteristics of the imaging equipment, or may be estimated from training data with the benefit of ground truth knowledge, assigning a value to $\tilde{\omega}$ is no real problem.

Thirdly, let us turn to the regularisation parameters. Since their aim is to penalise undesired characteristics, some authors prefer not to estimate these parameters at all, but to fix them according to the amount of regularisation deemed advisable. If an estimation is carried out, one has to maximise $p(\mathbf{x};\omega)=\exp[-U(\mathbf{x};\omega)]/Z(\omega)$ over ω. Unlike $Z(\tilde{\omega})$, the normalisation constant $Z(\omega)$ with respect to ω is not available in closed form for the vast majority of models. No unique optimum is guaranteed, and even if it does, explicit evaluation of its value is seldom possible. Hence a range of approximate solutions have been proposed, from analytic approximations of the expectation in its right hand side, through various iterative schemes, to a Monte Carlo approach [27, 24]. The latter is based on the observation that the log likelihood ratio with respect to some given value ω_0 can be expressed as

$$\begin{aligned} \log \frac{p(\mathbf{x};\omega)}{p(\mathbf{x};\omega_0)} &= U(\mathbf{x};\omega_0) - U(\mathbf{x};\omega) \\ &\quad - \log E_{\omega_0}\left[\exp[U(\mathbf{x};\omega_0) - U(\mathbf{x};\omega)]\right] \end{aligned}$$

so the expectation in the right hand side may be replaced by its Monte Carlo counterpart in a sample under the parameter value ω_0, and the result optimised with respect to ω. For an overview and critical comparison see [19, 25].

Finally, the inference steps discussed above have to be integrated. Perhaps the first idea that comes to mind is that of iteratively updating each type of parameter in turn. The disadvantage is that such an approach does not necessarily converge, as any asymptotics for each step depend on the output of the previous one. Alternatively, a fully Bayesian approach may be taken, which imposes hyper priors on each of the parameters and samples from the resulting posterior distribution. The obvious problem here is the specification of the hyper priors. No consensus seems to have been reached. We will describe our choice for the case study below.

4 Case study: linear network extraction

The patterns formed by roads, blood vessels, rivers and other waterways, or fissures in materials are known in the image processing community under the common denominator of 'linear networks'. The automatic extraction of such networks would be of use in fields such as cartography or computer-assisted medical diagnosis.

Many current algorithms to extract linear networks are two-step procedures [62, Ch. 10]. The first step consists of using a line detection operator such as the Hough transform, a directional or morphological filter [54, 76]. The result is subjected to a linking procedure performed in the second step to improve its connectivity, which may be based on dynamic programming [6, 20, 50], Markov random fields [76] or information theory [23].

Linear networks have the following characteristic properties:

- a locally homogeneous radiometry;
- contrast between a line segment and its surroundings;
- a slowly varying width;
- piecewise linearity.

While a network extraction algorithm may exploit the above properties, it may also encounter difficulties posed by, for instance, the presence of noise in the image, occlusion by other objects, potential false alarms due to objects with a radiometry that is similar to that of the features to be detected, and dependence of the network topology on its location (for instance, roads tend to be more straight in the United States than in Europe).

Except for the method proposed in [6], the previously mentioned techniques are supervised methods in the sense that the user has to provide a region of interest or points belonging to the network in order to make the detection feasible. These approaches are also pixel oriented, a trait that makes them sensitive to noise and

local optima. The intermediate-level approach to road detection by means of a Markov random field on a graph whose nodes represent road segments is found in [75]. It can be seen as a first attempt in the direction of object based modelling, and has proved its efficiency on radar images. The drawback of the method is the fact that segments that are not detected in the first step cannot be linked in the second one, cf. section 1.

4.1 Candy model

In the object process approach to high-level vision [4], a linear network is regarded as a pattern of connected line segments. Hence, instead of searching the image for pixels belonging to the linear network, one tries to locate directly the different components of the network. To do so, a model is needed for such collections of segments. Here we use the Candy model introduced in [70, 71].

A segment $\eta = (k, (l, \theta))$ is parametrised by its centre $k \in K \subset \mathbb{R}^2$, its length $l \in [l_{\min}, l_{\max}]$, and its orientation $\theta \in [0, \pi)$. Thus, a line segment process can be seen as a marked point process with marks for specifying the lengths and orientations. For linear networks approximated by segments, longer segments ought to be preferred to shorter ones, but there is no preference in orientation.

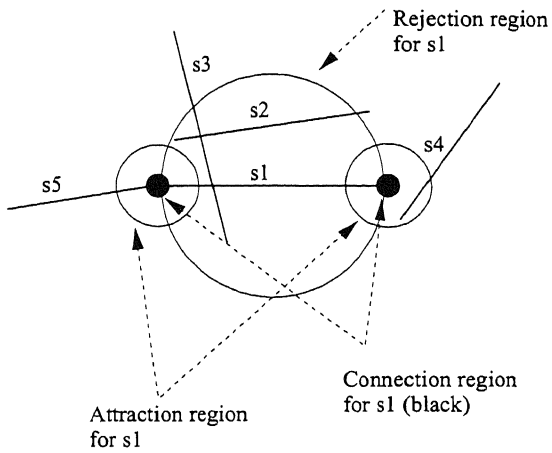


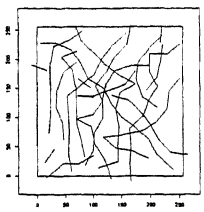
Figure 1: Different types of segment interaction for the Candy model.

Two segments η and ζ are said to be connected if at least one extremity of each segment is closer than some $r_c > 0$ to an extremity of the other, see Figure 1. A

:h no extremities connected is called a free segment. If only one end
ıected, the segment concerned is singly connected, and a segment with
ıints connected is called doubly connected. In a connected network,
gly connected segments typically occur less frequently than doubly
ɜgments, hence the Candy density penalises the former events.
/ery segment $\eta = (k_\eta, (l_\eta, \theta_\eta))$, a rejection region is defined by a circle
$_\eta$ with radius $r_r = l_\eta/2$. If the centre k_η is situated in the rejection
ɜegment ζ and if $||\theta_\eta - \theta_\zeta| - \pi/2| > \delta_{\min}$ for some threshold value
gments η and ζ are said to reject each other, and the Candy model
configurations containing both η and ζ.
.t η has an attraction region which is the union of the two circles cen-
tremities with radius $r_o = l_\eta/4$. If one extremity of a segment η is in
ı region of another segment ζ, if $\min\{|\theta_\eta - \theta_\zeta|, \pi - |\theta_\eta - \theta_\zeta|\} > \tau_{\max}$
$_{ax} > 0$, and if the segments η and ζ are not in the rejection described
are said to reject each other orientationwise. The energy term of the
:l penalises such pairs of segments that are ill-aligned.
illustrates the interaction structure. The black discs around the end
ıre its connection regions, so s_5 is connected to s_1. All other segments
ɜ larger discs indicate the regions of attraction and rejection. Note
ɜ$_2$ reject each other, but that the crossing segments s_1 and s_3 do not.
. and s_4 reject each other orientationwise, but s_1 and s_5 do not as they
orientations.
ry, with parameters $\omega_t \in \mathbb{R}$, and ω_f, ω_s, ω_r, $\omega_o < 0$, the negative
s

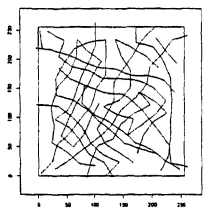
$$-U(\mathbf{x}) = \sum_{(k,(l,\theta))\in\mathbf{x}} \left[\frac{l - l_{\max}}{l_{\max}} + \omega_t\right] + n_f(\mathbf{x})\omega_f + n_s(\mathbf{x})\omega_s + n_r(\mathbf{x})\omega_r + n_o(\mathbf{x})\omega_o \qquad (113)$$

and $n_s(\mathbf{x})$ denote the number of free and singly connected segments
vely, $n_r(\mathbf{x})$ the number of segment pairs in $\mathbf{x}$ that reject each other,
e number of pairs of orientationwise rejecting segments. The domi-
process is a unit rate Poisson process on K marked independently
y. For a full definition, the reader is referred to [70, 71]; regarding the
:ture, see [41].
ical realisations are plotted in figure 2, obtained by the Metropolis–
amics discussed in section 3.2 with as transitions [41] births – both
d prolonging the current network – deaths, and segment changes,
ts components simultaneously or in orientation only. It might seem
ɔropose instead the birth of, say, a singly connected segment, but



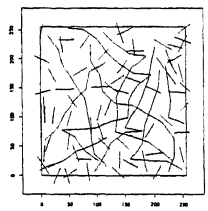
Model parameters	Sufficient statistics
$\omega_t = 2.5$	$n_t = 114$
$\omega_f = -11.0$	$n_f = 4$
$\omega_s = -5.5$	$n_s = 30$
$\omega_r = -2.5$	$n_r = 14$
$\omega_o = -2.5$	$n_o = 13$

a)



Model parameters	Sufficient statistics
$\omega_t = 4.0$	$n_t = 154$
$\omega_f = -12.5$	$n_f = 1$
$\omega_s = -7.0$	$n_s = 26$
$\omega_r = -2.5$	$n_o = 17$
$\omega_o = -2.5$	$n_o = 19$

b)



Model parameters	Sufficient statistics
$\omega_t = 2.5$	$n_t = 145$
$\omega_f = -7.5$	$n_f = 68$
$\omega_s = -5.5$	$n_s = 20$
$\omega_r = -2.5$	$n_r = 13$
$\omega_o = -2.5$	$n_o = 10$

c)

Figure 2: Realisations (left plot) of the Candy model with $l_{\min} = 30$, $l_{\max} = 40$, $\delta_{\min} = 0.05\pi$, and $\tau_{\max} = 0.2\pi$ for the parameters values given in the middle column. The observed values of the sufficient statistics are tabulated at the right.

the complicated nature of the set of such segments forces one to use approximations [70, 71].

Depending on the model parameters, different topologies are obtained. For example, a network with long rectilinear curves is plotted in figure 2a; if we increase the intensity parameter the denser pattern of figure 2b is obtained, while if free segments are penalised less stringently, the network tends to have fewer connected segments as in Figure 2c.

4.2 Data model

In order to extract a linear network from a given image, the data energy term has to check whether a segment belongs to the network or not using the characteristics listed at the beginning of this section. Several situations must be analysed.

First, we verify if the segment is in the middle of a homogeneous region. To do so, we build a one-block mask $[Dh]$ around the segment's main symmetry axis. For a segment $x = (k, m)$, the mask $[Dh]$ is centred at k with a shape depending on the mark m, cf. Figure 3. We write $\vec{y}_{H1}(x)$ for the grey levels within this mask. Using a Gaussian distribution with parameters $\tilde{\omega}_{H1}(x)$, we obtain the log likelihood function, which is denoted by $L_{H1}(x, \vec{y}_{H1}(x); \tilde{\omega}_{H1}(x))$.

Next, we check whether the segment is situated at the frontier of two homogeneous regions, i.e. on an edge. A two-block mask $[Del] - [Der]$ is built around the main symmetry axis of the segment. As in the first case the position of the mask depends on the centre of the segment we analyse, whereas the dimension of the mask depends on the mark of the segment. Again, $\vec{y}_{H2}(x)$ represents the grey levels within the mask $[Del] - [Der]$. The log likelihood is denoted by $L_{H2}(x, \vec{y}_{H2}(x); \tilde{\omega}_{H2}(x))$ and derived as an equal weights mixture of two Gaussian distributions. The parameter vector is denoted by $\tilde{\omega}_{H2}(x)$.

Finally, as in the previous two cases, depending on the centre and the mark of the segment x, we build a three-block mask $[Dsl] - [Ds] - [Dsr]$ around the main symmetry axis of the segment, in order to test whether the segment reflects an actual linear feature. Once again, write $\vec{y}_{H3}(x)$ for the grey levels under the mask, and take an equal weights mixture of, in this case, three Gaussian distributions to arrive at the log likelihood $L_{H3}(x, \vec{y}_{H3}(x); \tilde{\omega}_{H3}(x))$. The parameter vector is $\tilde{\omega}_{H3}(x)$. For an illustration, please consult Figure 3.

monotone with respect to the inclusion relation, implying the optimisation steps in the exact simulation algorithm of [36] would be difficult to implement. Therefore, we opt for a Metropolis–Hastings algorithm, cf. section 3.2.

Suppose we propose to add a segment η to the configuration $\mathbf{x}$. Then, the detailed balance equations (109) hold if

$$\alpha(\mathbf{x}, \mathbf{x} \cup \{\eta\}) = \min \left\{ 1, \frac{p_d \, d(\mathbf{x}, \eta) \exp\left[-(U(\mathbf{x} \cup \{\eta\}) + \tilde{U}(\mathbf{x} \cup \{\eta\}, \vec{\mathbf{y}}))\right]}{p_b \, b(\mathbf{x}, \eta) \exp\left[-(U(\mathbf{x}) + \tilde{U}(\mathbf{x}, \vec{\mathbf{y}}))\right]} \right\}$$

$$= \min \left\{ 1, \frac{p_d \, d(\mathbf{x}, \eta)}{p_b \, b(\mathbf{x}, \eta)} \times \exp[-(\triangle U(\eta) + \triangle \tilde{U}(\eta))] \right\}. \tag{116}$$

Here $\triangle U(\eta) = U(\mathbf{x} \cup \{\eta\}) - U(\mathbf{x})$ represents the difference in regularisation energy caused by the addition of η, and $\triangle \tilde{U}(\eta) = \tilde{U}(\mathbf{x} \cup \{\eta\}, \vec{\mathbf{y}}) - \tilde{U}(\mathbf{x}, \vec{\mathbf{y}})$ that to the data energy. We suppress the dependence on ω and $\tilde{\omega}$ in the notation. Thus, a new segment is accepted depending on its contribution to both $\triangle \tilde{U}(\eta)$ and $\triangle U(\eta)$, that is both the goodness-of-fit to the data image and the geometrical constraints as represented by the regularisation model are taken into account. Moreover, the computation of (116) is 'local', which is one of the main advantages of using the Metropolis–Hastings dynamics. As in section 4.1, we used the proposal densities developed in [41], but with acceptance probabilities modified as in (116) to incorporate the data energy.

4.4 Real data results

We applied the model (111) with (113) and (115) to publicly released SAR images obtained from the NASA/JPL website

`http://southport.jpl.nasa.gov`.

To sample from (111), we used the Metropolis–Hastings algorithm of section 4.3, incorporated into a simulated annealing framework (cf. section 3.4) in order to find the optimal linear network. The cooling schedule was

$$T_k = \frac{T_0}{\log(1+k)} \tag{117}$$

for $T_0 > 0$ and $k = 1, 2, \ldots, n$. At each temperature, 10^3 Metropolis–Hastings iterations were carried out, after which the temperature was decreased. The whole procedure was repeated 3×10^3 times. The initial temperature was $T_0 = 35$. The parameters of the Candy model were, after some trial and error to balance network completeness against false alarms, fixed at $\omega_t = 7.5$, $\omega_f = -92.5$, $\omega_s = -27.5$, $\omega_o = -25.0$ and $\omega_r = -35.0$. The minimum length of a segment

$l_{\min} = 21$, whereas $l_{\max} = 51$. The range of the single segment data potential was $[25.0, 50.0]$. In contrast to the Candy model parameters, which remain unchanged throughout the annealing procedure, the Gaussian mixture parameters $\tilde{\omega} = \tilde{\omega}(\mathbf{x})$ must be re-estimated at each update of the Metropolis–Hastings sampler by their current empirical counterparts.

Figure 4a shows a countryside region close to the city of Muar in Malaysia. The extracted roads are shown in Figure 4b. The main structure of the network is detected. There are some false alarms, which may be caused by the presence of thin, long, crop fields, but most such cases are avoided.

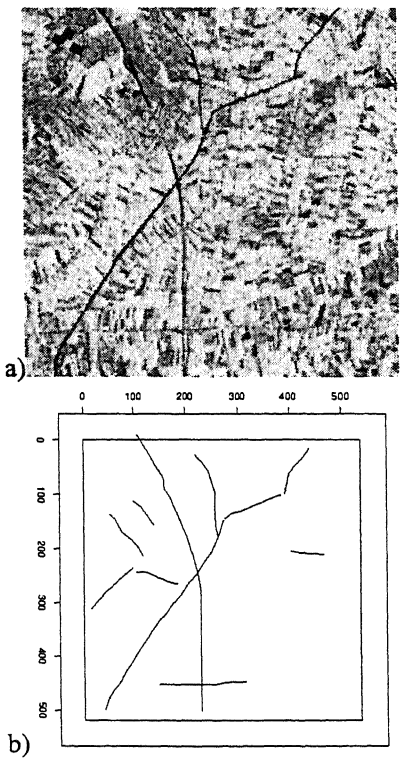


Figure 4: Countryside region in Malaysia: a) original image; b) result of extraction.

From Figure 5, we extract the tracks in between crop fields in an agricultural region of the Ukraine. The result seems quite complete, although some paths are not fully detected, and there are a few spurious instances.

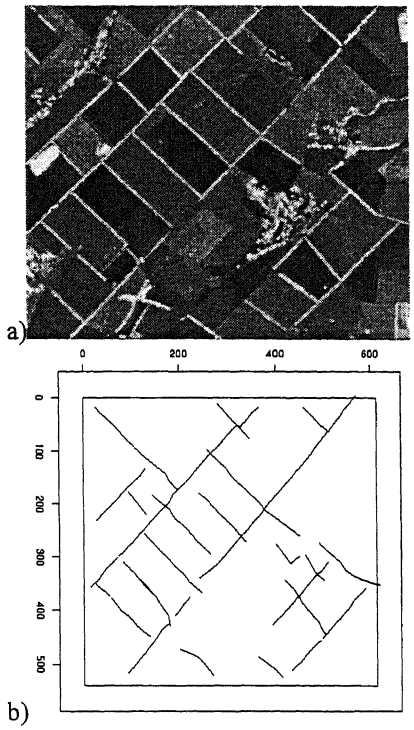


Figure 5: Agricultural region in the Ukraine: a) original image; b) result of extraction.

The third image, shown in Figure 6, depicts a fragment of the Nile delta in Egypt. The lower curve represents one arm of the Nile, the two straighter ones are irrigation canals. The gap in the network can be explained by the low quality of the data in this region over a longer stretch than the typical segment length.

Without access to ground truth data, no truly objective claim can be made about the quality of our results. Nevertheless, most of the more pronounced parts of the linear structure were detected in all three of the images, without introducing an unduly large number of false situations. We experimented with a range of values for the regularisation parameters, and found the results to be quite robust. Indeed, the same values were used for all three images, even though they are quite different.

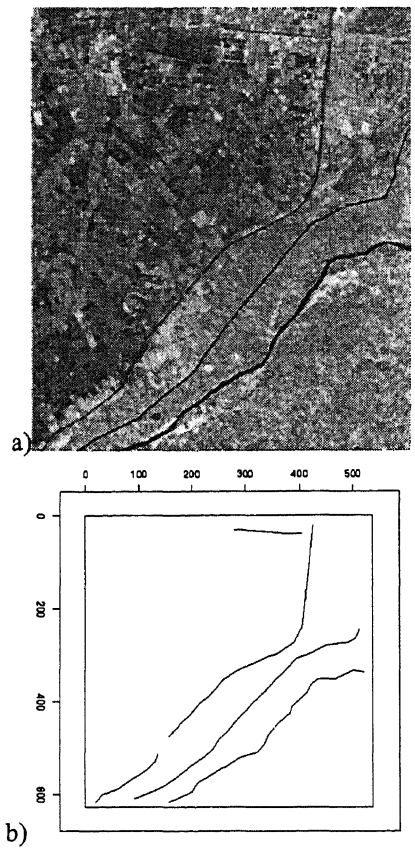


Figure 6: Nile river delta region in Egypt: a) original image; b) result of extraction.

5 Conclusion and future work

In this chapter we presented an application of point processes to image analysis. We suggested a line segment process [15, 41, 70, 71, 72] as a regularisation term in the context of linear network extraction, and presented new results for SAR images.

Several perspectives may be outlined. One possibility would be to investigate exact simulation algorithms for line segment processes; another is to incorporate spatial inhomogeneity [73] into the Candy model. For example, one may detect urban areas and texture parameters linked with the building density. This parameter

could drive the road network density as there are more roads in dense urban areas than in the countryside.

Current work is devoted to the generalisation of line segment processes to 3D. Such a generalisation would be particularly relevant in medical MRA imaging, where one is interested in dectecting blood vessels, and a more complex model based on broken lines instead of segments is under development.

Last but not least, we would like to stress one of the most exacting, therefore one of the most expensive perspectives: scene modelling. As a first step in this direction, in [18], a point process model is applied to building extraction. Indeed, object processes are quite flexible and lend themselves to the high-level modelling of images, hence could be well suited to dealing simultaneously with linear network extraction, texture segmentation and building extraction as well as to similar image interpretation tasks.

Acknowledgements

This research was supported by NWO grant 'Inference for random sets' (613-03-045) and the French-Dutch collaboration programme of NWO and INRIA via ERCIM, grant 'Stochastic geometry for image processing'.

References

[1] T. Arak, P. Clifford and D. Surgailis, Point based polygonal models for random graphs, *Advances in Applied Probability*, **25**, pp. 348-372, 1993.

[2] R. Azencott and C. Graffigne, Non supervised segmentation using multilevel Markov random fields, In *International Conference Pattern Recognition*, pp. 201-204, The Hague, 1992.

[3] A.J. Baddeley and M.N.M. van Lieshout, ICM for object recognition, In Y. Dodge and J. Whittaker, editors, *Computational Statistics*, volume 2, pp. 271-286, Heidelberg-New York, Physica/Springer, 1992.

[4] A.J. Baddeley and M.N.M. van Lieshout, Stochastic geometry models in high-level vision. In K. V. Mardia and G. K. Kanji, editors, *Statistics and Images, Volume 1, Advances in Applied Statistics*, a supplement to *Journal of Applied Statistics*, **20**, pp. 231-256, Abingdon, Carfax, 1993.

[5] A.J. Baddeley and J. Møller, Nearest-neighbour Markov point processes and random sets, *International Statistical Review*, **57**, pp. 89-121, 1989.

[6] M. Barzohar and D.B. Cooper, Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, pp. 707-721, 1996.

[7] J. Besag, Spatial interaction and the statistical analysis of lattice systems,

Journal of Royal Statistical Society, Series B, **36**, pp. 192-326, 1974.

[8] J. Besag, On the statistical analysis of dirty pictures (with discussion), *Journal of the Royal Statistical Society, Series B*, **48**, pp. 259-302, 1986.

[9] R. Chellapa and A. Jain, *Markov Random Fields: Theory and Application*, Boston/San Diego, Academic Press, 1993.

[10] F. Chen and P.A. Kelly, Algorithms for generating and segmenting morphologically smooth binary images, In *Proceedings of the 26th Conference on Information Sciences*, Princeton, 1992.

[11] G.R. Cross and A.K. Jain, Markov random field texture models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5**, pp. 25-39, 1983.

[12] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, New York, Springer-Verlag, 1988.

[13] H. Derin and H. Elliott, Modeling and segmentation of noisy and textured images using random fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, pp. 39-55, 1987.

[14] X. Descombes and E. Pechersky, Isotropic properties of some multi-body interaction models: two quality criteria for Markov priors in image processing, *Research Report No. 3752*, INRIA Sophia Antipolis, 1999.

[15] X. Descombes, M.N.M van Lieshout, R. Stoica and J. Zerubia, Parameter estimation by a Markov chain Monte Carlo technique for the Candy model, In *IEEE Workshop Statistical Signal Processing*, Singapore, 2001.

[16] X. Descombes, J.F. Mangin, E. Pechersky and M. Sigelle, Fine structure preserving Markov model for image processing. In *9th Scandinavian Conference on Image Analysis SCIA'95*, pp. 349-356, Uppsala, 1995.

[17] X. Descombes, R.D. Morris, J. Zerubia and M. Berthod, Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood, *IEEE Transactions on Image Processing*, **8**, pp. 954-963, 1999.

[18] X. Descombes, R. Stoica, L. Garcin and J. Zerubia, A RJMCMC algorithm for object processes in image processing, *Monte Carlo Methods and Applications*, **7**, pp. 149-156, 2001.

[19] P.J. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan and M. Tanemura, On parameter estimation for pairwise interaction processes, *International Statistical Review*, **62**, pp. 99-117, 1994.

[20] M.A. Fischler, J.M. Tenenbaum and H.C. Wolf, Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique, *Computer Graphics and Image Processing*, **15**, pp. 201-223, 1981.

[21] D. Geman, *Ecole d'ete de probabilites de Saint-Flour XVIII - 1988*, volume 1427 of *Lecture Notes in Mathematics*, chapter Random fields and inverse

problems in imaging, Berlin, Springer-Verlag, 1990.

[22] S. Geman and D. Geman, Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, pp. 721-741, 1984.

[23] D. Geman and B. Jedynak, An active testing model for tracking roads in satellite images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, pp. 1-14, 1996.

[24] C.J. Geyer, On the convergence of Monte Carlo maximum likelihood calculations, *Journal of the Royal Statistical Society, Series B*, **56**, pp. 261-274, 1994.

[25] C.J. Geyer, Likelihood inference for spatial point processes, In O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry, Likelihood, and Computation*, Boca Raton, CRC Press/Chapman and Hall, 1999.

[26] C.J. Geyer and J. Møller, Simulation procedures and likelihood inference for spatial point processes, *Scandinavian Journal of Statistics*, **21**, pp. 359-373, 1994.

[27] C.J. Geyer and E.A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B*, **54**, pp. 657-699, 1992.

[28] C.J. Geyer and E.A. Thompson, Annealing Markov chain Monte Carlo with applications to ancestral inference, *Journal of the American Statistical Association*, **90**, pp. 909-920, 1995.

[29] W.R. Gilks, S. Richardson and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, London, Chapman and Hall, 1996.

[30] G.L. Gimel'farb, *Image Textures and Gibbs Random Fields*, Dordrecht, Kluwer, 1999.

[31] P.J. Green, Reversible jump MCMC computation and Bayesian model determination, *Biometrika*, **82**, pp. 711-732, 1995.

[32] U. Grenander and M.I. Miller, Representations of knowledge in complex systems, *Journal of the Royal Statistical Society, Series B*, **56**, pp. 549-603, 1994.

[33] M.B. Hansen, J. Møller and F.Aa. Tøgersen, Bayesian contour detection in a time series of ultrasound images through dynamic deformable template models, Research Report R-00-2007, Department of Mathematical Sciences, Aalborg University, 2000.

[34] M.A. Hurn, Confocal fluorescence microscopy of leaf cells: an application of Bayesian image analysis, *Journal of the Royal Statistical Society, Series C*, **47**, pp. 361–377, 1998.

[35] W.S. Kendall, Perfect simulation for the area-interaction point process, In

L. Accardi and C. Heyde, editors, *Proceedings of the Symposium on Probability Towards the Year 2000*, Berlin, Springer-Verlag, 1998.

[36] W.S. Kendall and J. Møller, Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Advances in Applied Probability (SGSA)*, **32**, pp. 844-865, 2000.

[37] R. Kindermann and J.L. Snell, *Markov Random Fields and Their Applications*, Providence RI, American Mathematical Society, 1980.

[38] M.N.M. van Lieshout, Stochastic annealing for nearest-neighbour point process with application to object recognition, *Advances in Applied Probability*, **26**, pp. 281-300, 1994.

[39] M.N.M. van Lieshout, *Stochastic Geometry Models in Image Analysis and Spatial Statistics*, CWI tract 108. Amsterdam, CWI, 1995.

[40] M. N. M. van Lieshout, *Markov Point Processes and Their Applications*, London/Singapore, Imperial College Press/World Scientific Publishing, 2000.

[41] M.N.M. van Lieshout and R.S. Stoica, *The Candy Model Revisited: Markov Properties and Inference*, CWI Research Report, 2001.

[42] K. Mardia and G.K. Kanji, editors, *Statistics and Images, Advances in Applied Statistics*, a supplement to *Journal of Applied Statistics*, **20**, 1993.

[43] K.V. Mardia, W. Qian, D. Shah and K. de Souza, Deformable template recognition of multiple occluded objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, pp. 1036-1042, 1997.

[44] E. Marinari and G. Parisi, Simulated tempering: a new Monte Carlo scheme, *Europhysics Letters*, **19**, pp. 451-458, 1992.

[45] G. Matheron, *Random Sets and Integral Geometry*, New York, John Wiley and Sons, 1975.

[46] R. Molina and B.D. Ripley, Using spatial models as priors in astronomical image analysis, *Journal of Applied Statistics*, **16**, pp. 193–206, 1989.

[47] J. Møller, On the rate of convergence of spatial birth-and-date process, *Annals of Institute of Statistical Mathematics*, **41**, pp. 565-581, 1989.

[48] J. Møller and Ø. Skare, Bayesian image analysis with coloured Voronoi tesselations and a view to applications in reservoir modelling, Technical Report, Aalborg University, 2000.

[49] J. Møller and R.P. Waagepetersen, Markov connected component fields, *Advances in Applied Probability (SGSA)*, **30**, pp. 1-35, 1998.

[50] N. Merlet and J. Zerubia, New prospects in line detection by dynamic programming, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, pp. 426-431, 1996.

[51] G. Nicholls, Bayesian image analysis with Markov chain Monte Carlo and coloured continuum triangulation models, *Journal of the Royal Statistical*

Society, Series B, **60**, pp. 643-659, 1998.

[52] Y. Ogata and M. Tanemura, Estimation for interaction potentials of spatial point patterns through the maximum likelihood procedure, *Annals of the Institute of Statistical Mathematics*, **33**, pp. 315–338, 1981.

[53] A. Pievatolo and P.J. Green, Boundary detection through dynamic polygons, *Journal of the Royal Statistical Society, Series B*, **60**, pp. 609-626, 1998.

[54] R. Poli and G. Valli, An algorithm for real-time vessel enhancement and detection, *Computer Methods and Programs in Biomedicine*, **52**, pp. 1-22, 1997.

[55] C.J. Preston, Spatial birth-and-death processes, *Bulletin of the International Statistical Institute*, **46**, pp. 371-391, 1977.

[56] J.G. Propp and D.B. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms*, **9**, pp. 223-252, 1996.

[57] R.-D. Reiss, *A Course on Point Processes*, New York, Springer, 1993.

[58] B.D. Ripley, Tests of 'randomness' for spatial point patterns, *Journal of the Royal Statistical Society, Series B*, **41**, pp. 368-374, 1979.

[59] B.D. Ripley, The use of spatial models as image priors, In A. Possolo, editor, *Spatial Statistics and Imaging*, number 20 in Lecture notes – monographs, pp. 309–340, Hayward, California, Institute of Mathematical Statistics, 1991.

[60] B.D. Ripley and F.P. Kelly, Markov point processes, *Journal of the London Mathematical Society*, **15**, pp. 188-192, 1977.

[61] B.D. Ripley and A.I. Sutherland, Finding spiral structures in images of galaxies, *Philosophical Transactions of the Royal Society of London, Series A*, **332**, pp. 477-485, 1990.

[62] A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, volume II. New York, Academic Press, second edition, 1982.

[63] H. Rue and M.A. Hurn, Bayesian object identification, *Biometrika*, **3**, pp. 649-660, 1999.

[64] H. Rue and O.K. Husby, Identification of partly destroyed objects using dynamic polygons, *Statistics and Computing*, **8**, pp. 221-228, 1998.

[65] H. Rue and A.R. Syverseen, Bayesian object recognition with Baddeley's Delta loss, *Advances in Applied Probability (SGSA)*, **30**, pp. 64-84, 1998.

[66] H. Rue and H. Tjelmeland, Fitting Gaussian Markov random fields to Gaussian fields, *Scandinavian Journal of Statistics*, **29**, pp. 31-49, 2002.

[67] D. Ruelle, *Statistical Mechanics*, New York, Wiley, 1969.

[68] K. Sivakumar and J. Goutsias, Morphologically constrained discrete random sets, In D. Jeulin, editor, *Advances in Theory and Applications of Random Sets*, pp. 49-66, Singapore, World Scientific Publishing, 1997.

[69] J. Serra, *Image Analysis and Mathematical Morphology*, London, Academic

Press, 1982.

[70] R. Stoica, *Processus ponctuels pour l'extraction des réseaux linéiques dans les images satellitaires et aériennes*, PhD Thesis (in French), University of Nice–Sophia Antipolis, 2001.

[71] R. Stoica, X. Descombes and J. Zerubia, A Gibbs point process for road extraction in remotely sensed images, *Research Report 3923*, INRIA Sophia Antipolis, 2000.

[72] R. Stoica, X. Descombes and J. Zerubia, Road extraction in remote sensed images using a stochastic geometry framework, In *Proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods*, Gif-sur-Yvette, 2000.

[73] D. Stoyan and H. Stoyan, Non-homogeneous Gibbs process models for forestry - a case study, *Biometrical Journal*, **40**, pp. 521-531, 1998.

[74] H. Tjelmeland and J. Besag, Markov random fields with higher order interactions, *Scandinavian Journal of Statistics*, **25**, pp. 415-433, 1998.

[75] F. Tupin, H. Maître, J.-F. Mangin, J.-M. Nicolas and E. Pechersky, Detection of linear features in SAR images: application to road network extraction, *IEEE Transactions Geoscience and Remote Sensing*, **36**, 1998.

[76] S. Urago, J. Zerubia and M. Berthod, A Markovian model for contour grouping, *Pattern Recognition*, **28**, pp. 683-693, 1995.

[77] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Berlin/Heidelberg, Springer-Verlag, 1995.