Stereology and Image Analysis. Ecs10 - Proceedings of the 10th European Congress of ISS, (V.Capasso et al. Eds.), The MIRIAM Project Series, ESCULAPIO Pub. Co., Bologna, Italy, 2009

SEQUENTIAL SPATIAL PROCESSES FOR IMAGE ANALYSIS

MARIE-COLETTE VAN LIESHOUT

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.
Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

ABSTRACT

We give a brief introduction to sequential spatial processes. We discuss their definition, formulate a Markov property, and indicate why such processes are natural tools in tackling high level vision problems. We focus on the problem of tracking a variable number of moving objects through a video stream, and discuss the relationship with the popular Hough transform. A list of pointers to the literature concludes the paper.

Keywords: image analysis, sequential spatial process.

INTRODUCTION

Since the early 1990s, stochastic geometric models, in particular Markov marked point processes (Daley and Vere–Jones, 2003; Lieshout, 2000) have been successfully applied to image interpretation problems. The idea is to parametrise the objects that are present in an image or sequence of images by their location, shape, and colour parameters. Clearly, the parametrisation depends on the context, ranging from simple geometric shapes (Baddeley and Lieshout, 1992; Lieshout, 1994; 1995) through deformable template models (Amit *et al.*, 1991; Hansen *et al.*, 2002; Hurn, 1998; Mardia *et al.*, 1997; Pievatolo and Green, 1998; Rue and Hurn, 1999; Rue and Husby, 1998) to the complex ensembles of simple shapes studied by Lacoste *et al.* (2005), Ortner *et al.* (2007), and Stoica *et al.* (2002; 2004; 2007).

Such methods are especially useful for images of a scene in which the objects do not overlap or have a similar appearance. The purpose of this work is to describe a class of models that can overcome such limitations, namely sequential spatial processes (Lieshout, 2006a;b), and discuss how such models can be applied to the problem of tracking a variable number of moving objects in video data (Lieshout, 2008).

DEFINITIONS

In this section we recall the definition of sequential spatial processes. See (Lieshout, 2006a;b) for further details.

Configurations We assume that the objects that constitute the images we consider are completely described by a reference point that specifies its location, and some parametric features. Under this assumption, the object space is a Cartesian product of the form $D \times M$, where D is a compact set in the plane with non-empty interior for the object location, and M is some Polish space for the object features.

As each image may contain multiple objects, we need to consider ensembles. Generally, one does not know how many objects are contained in each frame of some video stream. Moreover, objects may overlap, and even occlude each other. Thus, we are led to define an object configuration as a vector

$$\vec{\mathbf{x}} = (x_1, \ldots, x_n) = ((d_1, m_1), \ldots, (d_n, m_n)),$$

with $x_i = (d_i, m_i) \in D \times M$. Here n, the number of components, is allowed to range through $\mathbb{N}_0$, with $n = 0$ corresponding to an empty scene without any objects whatsoever. The objects are ordered in terms of proximity to the camera.

Dominating measure The basic reference model for an object configuration is the Poisson object process. Under this model, the sequence length is Poisson distributed, and objects are independent and identically distributed. More formally, if μ is Lebesgue measure on $(D, \mathscr{B}_D)$, with $\mu(D) > 0$, and μ_M a Borel probability measure on the space of object features $(M, \mathscr{B}_M)$, write

$$\nu(F) = \sum_{n=0}^{\infty} \frac{e^{-\mu(D)}}{n!} \int \cdots \int_{(D\times M)^n} \mathbf{1}\{(x_1, \ldots, x_n) \in F\}$$
$$d(\mu \times \mu_M)(x_1) \cdots d(\mu \times \mu_M)(x_n)$$

for F in the σ-algebra on the configuration space generated by the product Borel σ-fields on $(D \times M)^n$. The term for $n = 0$ should be read as

$$\exp[-\mu(D)]\,\mathbf{1}\{\emptyset \in F\}.$$

Sequential spatial processes are defined by giving their Radon–Nikodym derivative (density) with respect to ν. In this form, they first appeared in the context of local scaling (Hahn *et al.*, 2003). In the next section, we present two further examples.

EXAMPLES

Random sequential adsorption model In many physical and biological processes, objects (e.g. monomers, animals, or proteins) arrive in some region and select a position for themselves according to some fixed distribution. If there would be no overlap with an established object, the object is adsorbed at the selected location, otherwise it leaves the system. The process could be run until there is no room left for any object as in the original formulation by Rényi (1958). Alternatively, the number of adsorbed particles could be stochastic, and follow some probability mass function. Such models were dubbed simple sequential inhibition processes in the spatial statistics literature (Diggle *et al.*, 1976) and are known as random sequential adsorption amongst physicists (Evans, 1993), who use them to describe the deposition of colloidal particles onto a substrate.

More specifically, consider the pure jump process on D with empty initial state, time horizon $\theta \geq 0$, and birth rate

$$b(u,\vec{\mathbf{x}}) = \pi(u)\,\mathbf{1}\{d(u,\vec{\mathbf{x}}) > r\}$$

where $\pi(\cdot)$ is the position selection probability distribution and d a metric on D. Note that in this example, the separation distance is fixed. In other words, no marks are attached to locations.

In Figure 1, a realisation in the unit square $D = [0,1]^2$ with $\pi(u) = \pi(z_1,z_2)$ given by

$$\frac{\lambda^2}{4}\exp\left[-\lambda\left(\left|z_1-\frac{1}{2}\right|+\left|z_2-\frac{1}{2}\right|\right)\right]$$

for $\lambda = 25$, $\theta = 1{,}000$, and $r = 0.03$ is presented.

The colour map is such that low indices are represented by a dark colour. Note that the squares cannot overlap. Dark squares dominate the centre of the picture whereas light ones are relatively often found on the outskirts of the point cloud.

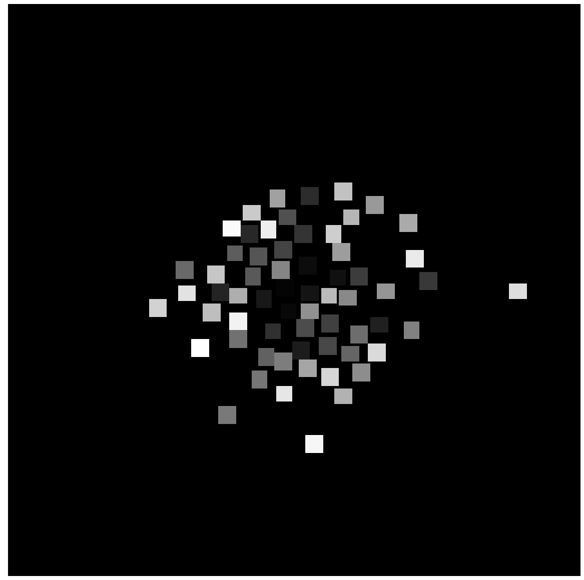

Fig. 1. *Sample from a sequential adsorption process in the unit square with Laplacian location selection with dispersion parameter* $\lambda = 25.0$, *time horizon* $1{,}000.0$, *and hard core distance* $r = 0.030$ *in both coordinates.*

It should be noted that the total birth rate

$$\int_D b(u,\vec{\mathbf{x}})\,du$$

depends on the geometry of $\vec{\mathbf{x}}$, which implies that a meaningful spatial Markov property cannot be expected to hold.

Sequential soft core model An important family of (classic) marked point processes is formed by the pairwise interaction processes, see e.g. (Lieshout, 2000), whose density factorises as a product of terms associated with pairs of neighbouring objects.

An example of a sequential analogue is the soft core model on $D \times \mathbb{R}^+$ defined by $f(\vec{\mathbf{x}})$ proportional to

$$\exp\left[\sum_i\left(\log(\beta)+\log(\gamma)\sum_{j<i}\mathbf{1}\{||d_i-d_j|| \leq m_j\}\right)\right]$$

with respect to the distribution of a sequence of Poisson length with independent components of which the position is uniformly distributed and the mark exponentially. Here $\beta > 0$ is an intensity parameter, and $0 < \gamma < 1$ reflects the strength of interaction, the smaller γ, the stronger the inhibition.

A realisation with $\beta = 100$, $\gamma = 0.611$, and the intensity parameter of μ_M equal to 0.05 is presented in Figure 2. Again a small index is represented by a dark colour. The radii are equal to the mark. Note that light balls tend to avoid being centred in darker ones, but such overlap is not prohibited altogether.

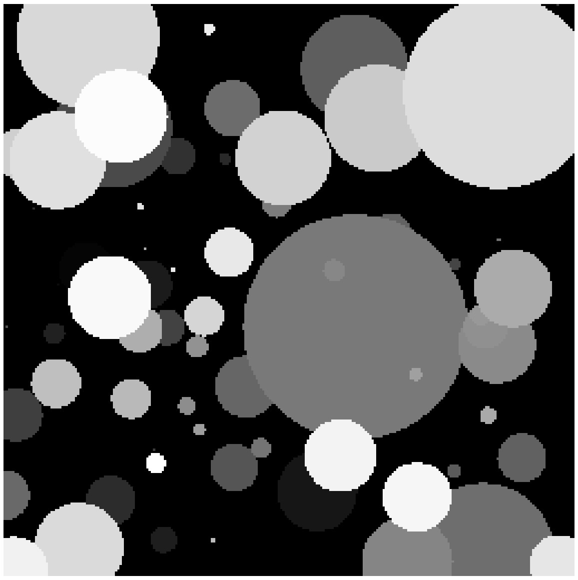

Fig. 2. *Sample from a sequential soft core model with* $\beta = 100$, $\gamma = 0.611$, *and mean radius* 0.050.

Note that the ratio

$$\frac{f(\vec{\mathbf{x}}, (d,m))}{f(\vec{\mathbf{x}})} = \beta \exp\left[\log(\gamma) \sum_j \mathbf{1}\left\{||d - d_j|| \leq m_j\right\}\right]$$

depends only on those $x_j = (d_j, m_j)$ for which $||d - d_j|| \leq m_j$. We say that $f(\cdot)$ is Markov with respect to the relation

$$(d_1, m_1) \sim (d_2, m_2) \Leftrightarrow ||d_1 - d_2|| \leq m_2.$$

By adding objects one at a time to $\emptyset$, it follows that $\beta^{-n} f$ is proportional to the product of

$$\varphi((d_i, m_i), (d_j, m_j)) = \gamma^{\mathbf{1}\{||d_i - d_j|| \leq m_j\}}$$

over pairs of neighbours $(d_i, m_i) \sim (d_j, m_j)$ for object configurations of length n.

The sequential soft core model forms a two-parameter exponential family with sufficient statistics $n(\vec{\mathbf{x}})$, the length of the sequence, and

$$\sum \sum_{j<i} \mathbf{1}\{||d_i - d_j|| \leq m_j\}.$$

For such models, Monte Carlo maximum likelihood estimation methods developed for classic spatial point processes, as reviewed for example by Geyer (1999) or Møller and Waagepetersen (2004), carry over immediately.

Remark Clearly, any sequential spatial process immediately defines a classic object process by ignoring the permutation (Daley and Vere–Jones, 2003; Hahn *et al.*, 2003). The interesting dual property that any finite sequential spatial process can be derived as the time-ordered vector of points in a classic spatio-temporal marked point process can be shown to hold as well. For further details, see (Lieshout, 2006b).

MOTION ANALYSIS

Motion is a prime source of semantic information. Indeed, when objects pass each other, their image projections overlap and their relative distance to the camera can be determined and propagated over frames.

The classical approach to motion tracking is to break the problem up into easier to handle sub-problems (Goodman *et al.*, 1997; Stone *et al.*, 1999; Vihola, 2004). One decides on the number of objects to be tracked, either ad hoc or by some tailor made expert system, and estimates the geokinematic coordinates, that is, position and velocity, by a Kalman filter (Eubank, 2006; Kalman, 1960) and/or Hough transform approach (Hough, 1962; Illingworth and Kittler, 1988). Although the Kalman filter is optimal for the prediction of the unobserved state of a linear system under Gaussian noise, it may not be so for the features extracted from video data. The Hough transform is robust against noise and occlusion but its implementation may require a lot of memory space. More recently, particle filters (Gordon *et al.*, 1993) were proposed. This approach, however, suffers from initialisation problems when the number of objects to be tracked does not remain constant over time (Hue *et al.*, 2002; Vihola, 2004), and does not seem capable of capturing interactions between the objects (Khan *et al.*, 2005).

Below, we shall apply sequential spatial models to the problem of tracking a variable number of interacting objects over video frames, which enables us to implement the sub-tasks outlined above simultaneously, and take into account varying object shapes and sizes, spatio-temporal relationships, and occlusion.

METHODOLOGY

A stochastic model for tracking consists of several ingredients.

Data We model a video sequence as a vector of images

$$\mathbf{y} = (\mathbf{y}^i; i = 1, \ldots, I),$$

$I \in \mathbb{N}$. In turn, each image $\mathbf{y}^i$ is determined by the values it takes on a set of pixels T. In other words,

$$\mathbf{y}^i = (y^i_t \,;\, t \in T).$$

The set T is usually a finite rectangular grid. The observed values y^i_t range over $V = \{0, 1, \ldots, 255\}^d$ with $d = 1$ for grey level and $d = 3$ for colour images.

Objects The object model strongly depends on the application at hand. Here, as before we shall use the generic notation x for a single object, and assume that each object leaves a footprint $R(x) \subseteq T$ in image space, which we call its template. The pixel values in the set $R(x)$ are denoted by $\theta_t(x), t \in R(x)$.

Signal image The signal image is defined as the footprint left by an object configuration rather than a single object in image space T. It is a function of the object templates that takes into account occlusion. More specifically, let

$$\theta_t(\vec{\mathbf{x}}) = \begin{cases} \theta_t(x_j) & \text{if } t \in R(x_j) \setminus \cup_{k<j} R(x_k) \\ \theta_0 & \text{if } t \in T \setminus \cup R(x_j) \end{cases}$$

Here θ_0 is the background value. Thus, among the objects whose template occupies a given pixel, the one with the smallest index, that is, the object closest to the camera, is the one that determines the signal (the lightest in terms of Figure 2). One may think of the signal as an idealised image, the one seen when there were no blur or noise.

Thus, the model explicitly and elegantly accounts for occlusion, in contrast to unordered object processes (Baddeley and Lieshout, 1993; Khan *et al.*, 2005) and in a simpler way than in (Mardia *et al.*, 1997).

Inference Motion analysis aims at inferring a sequence of object configurations $\mathbf{x} = (\vec{\mathbf{x}}^1, \ldots, \vec{\mathbf{x}}^I)$ from a given video stream. We shall take a statistical approach, and treat $\mathbf{x}$ as a parameter to be estimated. In order to do so, we need a probability model for $\mathbf{x}$. An advantage of such an approach over deterministic methods is that the degree of uncertainty about an obtained solution can be quantified.

Hamiltonian Under the Poisson model ν, conditionally given a scene contains n objects, they are independently and uniformly located in D with independently attributed features distributed according to μ_M. Interaction can be introduced by means of a Hamiltonian or energy function U. More precisely, we consider random sequences whose density (Radon–Nikodym derivative) at $\mathbf{x} = (\vec{\mathbf{x}}^1, \ldots, \vec{\mathbf{x}}^I)$ is of the form

$$f(\mathbf{x}) \propto \exp\left[-U(\mathbf{x})\right] \tag{1}$$

with respect to the I-fold product measure ν^I.

REGRESSION MODEL

In the least absolute deviation regression model,

$$U(\mathbf{x}) = \sum_{i=1}^{I} \sum_{t \in T} |y_t^i - \theta_t(\vec{\mathbf{x}}^i)|. \tag{2}$$

The objective is to seek, for data footage $\mathbf{y}$, to maximise (1) or, equivalently, to minimise (2). In general, there is no unique solution, as the addition of extra objects 'behind' the signal of those closer to the camera does not affect U and hence f.

Markov property The potential energy required for adding some object u to, say, configuration $\vec{\mathbf{x}}^i$, giving $u \notin \vec{\mathbf{x}}^i$ the highest index, is given by

$$\sum_{t \in R(u) \setminus \cup_k R(x_k^i)} \left[|y_t^i - \theta_t(u)| - |y_t^i - \theta_0| \right]$$

and depends only on the pixel values in the template $R(u)$ of u and those in $R(x_k^i)$ that overlap $R(u)$. Hence, the single frame energy function defines a Markov sequential object process with respect to the overlapping objects relation

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

Hough interpretation Given a grey scale image $\mathbf{y}$, the (generalised) Hough transform is an integer-valued function on the object space

$$H_{\mathbf{y}}(u) = \sum_{t \in R(u)} y_t$$

that assigns to each object u the total intensity in its template. The Hough transform may be interpreted as follows: each pixel t votes with vigour y_t for all the objects that contain that pixel in their template. Good matches could then be located by finding local maxima of the Hough transform (Illingworth and Kittler, 1988). The intensity y_t may be replaced by $\psi(y_t)$ for some appropriate function $\psi : V \to \mathbb{R}$.

Consider the track of a newly arrived object against an empty background. Write b for its birth frame, d for its death frame, u_i for the object in frame $i = b, \ldots, d$, and $(v_i)_{i=b}^{d-1}$ for the translation vectors between frames. The track thus parametrised will be denoted by $\tilde{u}$. Furthermore, suppose that the template and signal are translation invariant, so that $R(u + \Delta) = R(u) + \Delta$ and $\theta_t(u + \Delta) = \theta_{t-\Delta}(u)$ for all pixels t, Δ with $t - \Delta \in T$. Then, the difference in energy $U(\emptyset) - U(\tilde{u})$ is given by

$$\sum_{t \in R(u_b)} \left[|y_t^b - \theta_0| - |y_t^b - \theta_t(u_b)| \right] + \sum_{i=b}^{d-1} \sum_{t \in R(u_i)} \left[|y_{t+v_i}^{i+1} - \theta_0| - |y_{t+v_i}^{i+1} - \theta_t(u_i)| \right].$$

The first term corresponds to a Hough transform for detecting the initial presence of an object by letting each pixel vote for the objects that contain it with strength $|y_t^b - \theta_0| - |y_t^b - \theta_t(u_b)|$; the second term is a recursive Hough transform voting for the movement from u_i by v_i with strength

$$|y^{i+1}_{t+v_i} - \theta_0| - |y^{i+1}_{t+v_i} - \theta_t(u_i)|$$

for each pair of pixels $(t, t+v_i)$ with $t \in R(u_i)$.

DISCUSSION

In a recent study the ideas described above were applied to sports sequences in which the objects of interest can be described mathematically by geometric objects such as ellipses. In order to avoid over fitting, a regularisation energy was included in the model. It was designed to prevent too much overlap between objects in a single image, to encourage cohesion between objects in adjacent frames, and to include object identifiers in order to keep track of an object's identity as it moves across the frames.

More formally, we introduced a regularisation term that is the sum of two energy functions: the first a purely inhibitive pairwise interaction Markov model with respect to the overlapping objects relation (Baddeley and Lieshout, 1992), the second a Markov chain in 'frame time' inspired by (Lund *et al.*, 1999). The latter's effect is three fold. It penalises objects that are unmatched in the sense of not being identified with objects in adjacent frames, it forces matched objects to have a similar template, and it propagates information on relative proximity to the camera gathered when objects overlap on to adjacent image frames.

Optimisation was carried out by simulated annealing within the Metropolis–Hastings framework (Møller and Waagepetersen, 2004). The method was implemented in the `C++` library `SEQ-MPPLIB` at CWI by Steenbeek and Van Lieshout.

The approach proved very successful and was able to capture simultaneously a variable number of objects, occlusion, depth, and spatial and temporal coherence. For further details, the reader is referred to (Lieshout, 2008).

ACKNOWLEDGEMENTS

This research was supported by the Technology Foundation STW, the applied science division of NWO, and the technology program of the Ministry of Economic Affairs (project CWI.6156 'Markov sequential point processes for image analysis and statistical physics').

The author would like to thank Dr. C. Varekamp (Philips Research) for data and interesting discussions, and A.G. Steenbeek for expert programming assistance.

REFERENCES

Amit Y, Grenander U, Piccioni M (1991). Structural image restoration through deformable templates. J. Amer. Statist. Assoc. 86:376–87.

Baddeley AJ, Lieshout MNM van (1992). Object recognition using Markov spatial processes. In: Proceedings 11th IAPR International Conference on Pattern Recognition B:136–9. Los Alamitos: IEEE Computer Society Press.

Baddeley AJ, Lieshout MNM van (1993). Stochastic geometry models in high-level vision. In: Mardia KV, Kanji GK, eds. Statistics and Images, Volume 1. Advances in Applied Statistics, a supplement to Journal of Applied Statistics 20:231–56. Abingdon: Carfax.

Daley DJ, Vere–Jones D (2003). Introduction to the Theory of Point Processes. Volume I. Elementary Theory and Methods (2nd edn). New York: Springer-Verlag.

Diggle PJ, Besag J, Gleaves JT (1976). Statistical analysis of spatial point patterns by means of distance methods. Biometrics 32:659–67.

Eubank RL (2006). A Kalman Filter Primer. Boca Raton: Chapman & Hall/CRC.

Evans JW (1993). Random and cooperative sequential adsorption. Reviews of Modern Physics 65:1281–329.

Geyer CJ (1999). Likelihood inference for spatial point processes. In: Barndorff-Nielsen O, Kendall WS, Lieshout MNM van, eds. Stochastic Geometry, Likelihood, and Computation. Boca Raton: CRC Press/Chapman and Hall, 79–140.

Goodman IR, Mahler RPS, Nguyen HT (1997). Mathematics of Data Fusion. Volume 39 of Series B: Mathematical and Statistical Methods. Dordrecht: Kluwer.

Gordon N, Salmond D, Smith, A (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings 140:107–13.

Hahn U, Jensen EBV, Lieshout MNM van, Nielsen LS (2003). Inhomogeneous spatial point processes by location dependent scaling. Adv. in Appl. Probab. 35:319–36.

Hansen MB, Møller J, Tøgersen FA (2002). Bayesian contour detection in a time series of ultrasound images through dynamic deformable template models. Biostatistics 3:213–28.

Hough PVC (1962). Method and means for recognizing complex patterns. US Patent 3069654.

Hue C, Le Cadre J-P, Pérez P (2002). Sequential Monte Carlo methods for multiple target tracking and data fusion. IEEE Transactions on Signal Processing 50:309–25.

Hurn MA (1998). Confocal fluorescence microscopy of leaf cells: an application of Bayesian image analysis. J. Roy. Statist. Soc. Ser. C 47:361–77.

Illingworth J, Kittler J (1988). A survey of the Hough transform. Computer Vision, Graphics and Image Processing 44:87–116.

Kalman R (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82:35–45.

Khan Z, Balch T, Dellaert F (2005). MCMC-based particle filtering for tracking a variable number of interacting targets. IEEE Trans. Pattern Anal. Mach. Intell. 27:1805–19.

Lacoste C, Descombes X, Zerubia J (2005). Point processes for unsupervised line network extraction in remote sensing. IEEE Trans. Pattern Anal. Mach. Intell. 27:1568–79.

Lieshout MNM van (1994). Stochastic annealing for nearest-neighbour point processes with application to object recognition. Adv. in Appl. Probab. 26:281–300.

Lieshout MNM van (1995). Stochastic Geometry Models in Image Analysis and Spatial Statistics, CWI Tract 108. Amsterdam: CWI.

Lieshout MNM van (2000). Markov Point Processes and their Applications. London/Singapore: Imperial College Press/World Scientific Publishing.

Lieshout MNM van (2006a). Markovianity in space and time. In: Denteneer D, Hollander F den, Verbitskiy E, eds. Dynamics and Stochastics: Festschrift in Honour of Michael Keane. Lecture Notes Monograph Series 48. Beachwood: Institute for Mathematical Statistics, 154–67.

Lieshout MNM van (2006b). Campbell and moment measures for finite sequential spatial processes. In: Hušková M, Janžura M, eds. Proceedings Prague Stochastics 2006. Prague: Matfyzpress, 215–24.

Lieshout MNM van (2008). Depth map calculation for a variable number of moving objects using Markov sequential object processes. IEEE Trans. Pattern Anal. Mach. Intell. 30:1308–12.

Lund J, Penttinen A, Rudemo M (1999). Bayesian analysis of spatial point patterns from noisy observations. Research Report, Department of Mathematics and Physics. Copenhagen: The Royal Veterinary and Agricultural University.

Mardia KV, Qian W, Shah D, De Souza KMA (1997). Deformable template recognition of multiple occluded objects. IEEE Trans. Pattern Anal. Mach. Intell. 19:1036–42.

Møller J, Waagepetersen RP (2004). Statistical Inference and Simulation for Spatial Point Processes. Boca Raton: Chapman & Hall/CRC Press.

Ortner M, Descombes X, Zerubia J (2007). Building outline extraction from digital elevation models using marked point processes. International Journal of Computer Vision, 72:107–32.

Pievatolo A, Green PJ (1998). Boundary detection through dynamic polygons. J. R. Stat. Soc. Ser. B Stat. Methodol. 60:609–26.

Rényi A (1958). On a one-dimensional problem concerned with random space-filling (in Hungarian). Mag. Tud. Akad. Kut. Mat. Intézet Kozlemenyei 1958:109–27. Reprinted in English in: Turán, ed. Selected Papers of Alfréd Rényi Volume 2. Budapest: Akadémiai Kiadó, 1976.

Rue H, Hurn MA (1999). Bayesian object identification. Biometrika 86:649–60.

Rue H, Husby OK (1998). Identification of partly destroyed objects using deformable templates. Statistics and Computing 8:221–8.

Stoica R, Descombes X, Zerubia J (2004). A Gibbs point process for road extraction in remotely sensed images. International Journal of Computer Vision, 57:121–36.

Stoica RS, Descombes X, Lieshout MNM van, Zerubia J (2002). An application of marked point processes to the extraction of linear networks from images. In: Mateu J, Montes F, eds. Spatial Statistics: Case Studies. Southampton: WIT Press, 287–312.

Stoica RS, Martinez V, Saar E (2007). A three dimensional object point process for detection of cosmic filaments. J. Roy. Statist. Soc. Ser. C 56:459–77.

Stone LD, Barlow CA, Corwin TL (1999). Bayesian Multiple Target Tracking. Norwood: Artech House.

Vihola M (2004). Random Sets for Multitarget Tracking and Data Fusion. Licentiate Thesis, Tampere University of Technology.