PRIOR DISTRIBUTIONS FOR BAYESIAN IMAGE ANALYSIS

M N M van Lieshout

Department of Statistics
University of Warwick
Coventry CV4 7AL, United Kingdom

In a wide range of disciplines including robot vision, microscopy and material science, it is required to extract objects of interest such as industrial parts, cell structures and minerals from a noisy and blurred image. We will formulate this task as a Bayesian estimation problem, paying particular attention to the choice of the prior distribution, and discuss how Markov chain Monte Carlo techniques can be used for statistical inference.

1 Definitions and notation

Following Baddeley and Van Lieshout [3, 4], we assume that the objects to be recognised are represented by a finite number of real parameters that determine size, shape and location. Writing U for the space of possible parameter vector values, a point $u \in U$ represents an object $R(u)$ in the digitised image T. To allow for images with multiple objects, define an object configuration to be a finite unordered set $\mathbf{x} = \{x_1, \cdots, x_n\}, n \geq 0$, of object parameters $x_i \in U$. In mathematical terms, $\mathbf{x}$ is a realisation of a *finite point process* on U.

In practice, the digitised objects $R(x_i)$, $x_i \in \mathbf{x}$, are corrupted during the imaging process. We model the deterministic blurring component by mapping the object configuration $\mathbf{x}$ to an image $\theta_t^{(\mathbf{x})}$, $t \in T$, called the signal; the remaining random noise component is modelled by a probability density $f(\mathbf{y} \mid \theta^{(\mathbf{x})})$. It is convenient to assume that the pixel values $\mathbf{y} = (y_t)_{t \in T}$ are conditionally independent given the signal but texture models such as Markov random fields could be used as well at a slight increase in computational cost.

As an illustrative example, consider a blur-free signal $\theta_t^{(\mathbf{x})} = \mathbf{1}\{t \in S(\mathbf{x}) = \bigcup_{i=1}^n R(x_i)\}$, $t \in T$, and a binary noise component that randomly swaps background pixels from value 0 to 1 with some fixed probability $p \in (0, 1)$ and leaves pixels in the foreground $S(\mathbf{x})$ unchanged. To extract the objects from a data image $\mathbf{y}$, regard $\mathbf{x}$ as a parameter and estimate it by maximising the likelihood $f(\mathbf{y} \mid \theta^{(\mathbf{x})})$ over $\mathbf{x}$. Letting Y be the set of data pixels with value 1, this likelihood is nonzero only when $S(\mathbf{x}) \subseteq Y$ and the log likelihood is a linear function of $|S(\mathbf{x})|$, the total area occupied by objects. Hence, one solution of the maximum likelihood equations is $\hat{\mathbf{x}}_{\max} = \{u \in U : R(u) \subseteq Y\}$, the (generalised) erosion operator of mathematical morphology (cf.[36]). The other solutions are the subsets $\mathbf{x} \subseteq \hat{\mathbf{x}}_{\max}$ with $S(\mathbf{x}) = S(\hat{\mathbf{x}}_{\max})$.

From the example above it is clear that maximum likelihood techniques suffer from multiple response due to occlusion. This could well be undesirable and suggests a Bayesian approach with a prior distribution penalising scenes with many overlapping objects.

2 Markov prior models

As advocated in [3, 4, 23, 33], a class of models for object configurations suitable for penalising overlap and encouraging 'smooth' objects is that of *Markov overlapping object processes* [32, 6, 3, 4]. These processes are defined by their density $p(\cdot)$ with respect to a Poisson process on U with some non-atomic

finite intensity measure $\mu(\cdot)$ – for instance if U is a bounded subset of $\mathbb{R}^d$, the usual choice for $\mu(\cdot)$ is d–dimensional Lebesgue measure – and satisfy the Ripley and Kelly [32] Markov property that, for all configurations $\mathbf{x}$ with $p(\mathbf{x}) > 0$, (a) $p(\mathbf{y}) > 0$ for all $\mathbf{y} \subseteq \mathbf{x}$; and (b) the conditional intensity $\frac{p(\mathbf{x} \cup \{u\})}{p(\mathbf{x})}$ depends only on u and those $x_i \in \mathbf{x}$ with $R(x_i) \cap R(u) \neq \emptyset$.

An advantage of Markov overlapping object processes is that they are easy to interpret in terms of interactions between intersecting objects: a density $p(\cdot)$ is Markov iff $p(\mathbf{x})$ can be factorised as

$$p(\mathbf{x}) = \prod_{\text{cliques } \mathbf{y} \subseteq \mathbf{x}} \phi(\mathbf{y}) \tag{1}$$

where the product is restricted to cliques $\mathbf{y} \subseteq \mathbf{x}$ of mutually overlapping objects and $\phi(\cdot)$ are non-negative functions. See [32] for details.

By the factorisation (1), we can define a model by specifying the interaction functions $\phi(\cdot)$. However, some care must be taken to ensure that the result is well-defined and can be normalised into a probability density. A sufficient condition is that all $\phi(\cdot)$ are bounded above by 1; in that case the model is said to be purely inhibitory.

2.1 Within object interaction

As we mentioned in Section 1, any $u \in U$ describes an object in terms of a few parameters. In the simplest case where all objects are identical up to translation, two location parameters suffice. Thus the prior term $\phi(\{u\})$ in (1) may favour some locations over others, or simply be constant $\phi(\{u\}) \equiv \beta$. Since each object contributes a term β to the prior, values of $\beta > 1$ favour scenes with many objects while $\beta < 1$ will penalise too many objects.

Often, there will be more than just location parameters. For instance if the image contains both squares and discs, a type indicator is needed and the prior may give different weights to the two types. Size may be taken into account by a factor $\beta^{|R(u)|}$ [4].

More complicated templates have been considered as well. For example Yuille [39] and Philips and Smith [27, 28] model the human mouth by global parameters describing location, orientation and size, while local parameters specify the depths of the upper and lower lip. The prior may include a Gaussian term for each of the global parameters and conditional Gaussians for the local parameters (see also Aykroyd and Green [1]).

Another class of parametrisations models object boundaries by polygons with a fixed or variable number of vertices [13]. The prior may enforce smoothness or penalise too many sides [26, 30, 34] for instance by a cyclic Markov random field.

$$\phi(u) = \exp[-\alpha_R \sum (R(i) - R(i-1))^2 - \alpha_\theta \sum (\theta(i) - \theta(i-1) - \pi)^2]$$

where $\alpha_R, \alpha_\theta > 0$ are smoothness parameters, and u has vertices with (centred) polar coordinates $(R(i), \theta(i))$.

2.2 Between object interaction

Next we turn attention to the interaction functions for cliques of mutually overlapping objects. If no objects are allowed to overlap at all, a *hard object process* [4] with $\phi(x_i, x_j) = \mathbf{1}\{R(x_i) \cap R(x_j) = \emptyset\}$ would be appropriate. In the absence of other interaction terms, this is just a Poisson process conditioned on having no intersecting objects. Applications can be found in [4, 16, 34].

The hard object process is an example of a *pairwise interaction model*

$$p(\mathbf{x}) = \alpha \prod_i \phi(x_i) \prod \phi(x_i, x_j)$$

where α is the normalising constant, and the second product is over all pairs of overlapping objects x_i, x_j. Other special cases include the *Strauss object process* $\phi(x_i, x_j) \equiv \gamma$, or $\phi(x_i, x_j)$ a function of the distance between the object centers as in Qian and Mardia [30]. In both cases, $\phi(\cdot, \cdot)$ must be bounded by 1 to ensure that the model is well-defined.

The Qian and Mardia model takes into account the amount of overlap between the objects. Similar ideas lie behind the *area-interaction model* [40, 5]

$$p(\mathbf{x}) = \alpha\gamma^{-|S(\mathbf{x})|}\prod_i \phi(x_i) \tag{2}$$

where $S(\mathbf{x})$ again denotes the union $\cup_i R(x_i)$. Contrary to the pairwise interaction models discussed above, $p(\cdot)$ is well-defined for all values of $\gamma > 0$; realisations tend to have few intersections for $\gamma < 1$ and clusters of overlapping objects for $\gamma > 1$. Note that (2) exhibits interactions between arbitrarily many objects.

Various generalisations of (2) have been proposed recently. Baddeley et al. [2] study quermass interaction processes which replace area by perimeter and other fundamental geometric measures. Van Lieshout and Molchanov [22] consider models based on the coverage function $c_{\mathbf{x}}(a) = \sum_{i=1}^{n} 1\{a \in R(x_i)\}$, counting the number of objects covering pixel a. For instance,

$$p(\mathbf{x}) = \alpha\gamma^{-|\{t:c_{\mathbf{x}}(t)=1\}|}\prod_i \phi(x_i)$$

with $\gamma < 1$ penalises realisations containing many overlapping objects. As for the area-interaction process, the model is well-defined for all $\gamma > 0$.

Finally, note that all models discussed above treat objects equally. However, as suggested in Qian and Mardia [30], occlusion can be taken into account by imposing an ordering on object configurations to describe which objects lie on top.

3 Random set based and morphological priors

In some applications it is not necessary to identify individual objects, only to separate foreground from background. For segmentation problems like this, pixel-based Markov random fields have been proposed as prior distributions to encourage neighbouring pixels to belong to the same segment. See [7] and the references therein for details. Recently, more 'region-based' priors have been proposed. Since the natural phenomena underlying many images are continuous in nature, in this Section we will consider their random set based analogues.

We will restrict attention to random sets X defined by their density with respect to a Boolean model on a compact window W with typical grain distribution $\nu(\cdot)$ and intensity $\lambda > 0$ [37]. Note that the density must depend on the union of the grains only. The role of within object interactions is taken over by $\nu(\cdot)$.

Møller and Waagepetersen [25] introduced *Markov connected component fields*, where the density factorises into terms associated with the connected components in an image rather than with cliques of neighbouring pixels (cf. [6]). They prove that if the model is also a Markov random field with respect to horizontal and vertical neighbours, then the density must be defined in terms of the area and perimeter of the components. If diagonal neighbours are included, Euler characteristics as well as the numbers of corners and discontinuities must be included. Hence, the random set analogues are the quermass random sets of [2].

Recently, Chen and Kelly [8] proposed

$$p(X) = \alpha\gamma^{|X\setminus X\circ B|} \tag{3}$$

to favour images that are morphological smooth, i.e. do not have narrow isthmuses, small islands or sharp capes. Here X is the set of 1−pixels and $X \circ B$ the opening of X by a finite structuring element B. By duality, $|X \setminus X \circ B|$ may be replaced by $|X \bullet B \setminus X|$, where the closing $X \bullet B$ penalises small holes [36]. Being defined in terms of area, generalisation to continuous random sets is straightforward.

More generally, size constraints can be imposed, as studied by Sivakumar and Goutsias [35]. In stochastic geometry, the size of voids left open by a stationary random set may be measured by the *empty space function* $F_B(r) = \mathbb{P}(0 \in X \oplus rB)$, $r \geq 0$ or the *contact distribution function*

$$G_B(r) = \mathbb{P}(0 \in X \oplus rB \mid 0 \notin X) = \frac{F_B(r) - F_B(0)}{1 - F_B(0)}, \qquad r \geq 0.$$

Here, the structuring element B is a convex, compact set containing a neighbourhood of the origin, usually the unit ball, and $X \oplus rB$ [36] is the dilation of X with structuring element rB. F_B and G_B are easy to interpret and have analogues in the theory of point processes, however, from a morphological point of view, dilation is not an anti-granulometry since it fails to satisfy the 'sieving condition' $(X \oplus rB) \oplus sB = (X \oplus sB) \oplus rB = X \oplus (\max(r,s))B$, $r, s \geq 0$. Replacing dilation by closing, a proper size distribution is obtained [36].

Thus, given a compact, convex set B, define a size distribution of voids by

$$G_\bullet(r) = \mathbb{P}(0 \in X \bullet rB \mid 0 \notin X) = 1 - \frac{\mathbb{P}(0 \notin X \bullet rB)}{\mathbb{P}(0 \notin X)}, r \geq 0$$

where $X \bullet rB$ denotes the morphological closing of X with structuring element rB. Here we assume that the coverage fraction $p_X = \mathbb{P}(0 \in X)$ lies in the open interval $(0,1)$. Similarly, reversing the roles of foreground and background, we obtain the size distribution of X itself as

$$G_\circ(r) = 1 - \mathbb{P}(0 \in X \circ rB \mid 0 \in X) = 1 - \frac{\mathbb{P}(0 \in X \circ rB)}{\mathbb{P}(0 \in X)}, r \geq 0$$

where $X \circ rB$ denotes the morphological opening of X by rB (again provided the coverage fraction $p_X \in (0,1)$). Note that $G_\circ(r)$ can be interpreted as the probability of elimination by rB. Ripley [31] suggested to plot the estimated coverage fractions of $X \oplus rB, X \ominus rB, X \circ rB$ and $X \bullet rB$ for a range of r-values as an exploratory data analysis tool. For details on mathematical morphology, see for instance Serra [36].

If both the size distribution of the foreground and the background are of interest, we can consider the distribution function

$$G(r) = \begin{cases} 1 - \mathbb{P}(0 \in X \circ rB) & r \geq 0 \\ 1 - \mathbb{P}(0 \in X \bullet -rB) & r < 0 \end{cases}$$

defined on $\mathbb{R}$.

Sivakumar and Goutsias [35] defined a discrete morphologically constrained random field by

$$p(X) = \alpha \exp[-\sum_{i=0}^{I} \beta_i |X \circ iB \setminus X \circ (i+1)B| - \sum_{j=1}^{J} \gamma_j |X \bullet jB \setminus X \bullet (j-1)B|]. \tag{4}$$

Noting that (4) is based on the naive empirical distribution

$$\hat{G}_X(r) = \begin{cases} 1 - \frac{|W \cap (X \circ rB)|}{|W|} & r \geq 0 \\ 1 - \frac{|W \cap (X \bullet -rB)|}{|W|} & r < 0 \end{cases}$$

we can rewrite $p(X) = \alpha \exp[-\int_{-J}^{I} f(s) d\hat{G}_X(s)]$ where $f(\cdot)$ is a step function taking values β_i, γ_j. Generalisations may be obtained by letting $f(\cdot)$ be any bounded function. The interaction range clearly depends on the range of $f(\cdot)$ and on I, J. Note that the function $f(\cdot)$ may favour some sizes and penalise others. For instance, if $f(\cdot)$ is the indicator function of $[-I, I]$, it will encourage the fore- and background to exceed size I.

Note that the naive estimator $\hat{G}_X(\cdot)$ does no take into account edge effects. For improved estimators, see Hansen et al. [10], Chiu and Stoyan [9] and the references therein.

4 Statistical inference

Statistical inference for the 'true' scene $\mathbf{x}$ can be based on the posterior distribution $p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{x}) f(\mathbf{y} \mid \mathbf{x})$. Since the normalising constant is usually not available in analytical form, Markov chain Monte Carlo techniques are needed (see for instance [12], [11, 24] or the 1993 special issue of the Journal of the Royal Statistical Society for more details).

For a Markov object process, ratios of the form $p(\mathbf{x}')/p(\mathbf{x})$ are easy to compute for $\mathbf{x}'$ obtained from $\mathbf{x}$ by adding, deleting or modifying an object. For such changes, the likelihood ratio term is also

straightforward. For instance, consider an independent noise model and a blur-free signal with values θ_1 for the foreground and θ_0 for the background. Then adding object $u \in U$ to configuration $\mathbf{x}$ yields a log likelihood ratio $\sum_{i \in R(u)} \log \frac{g(y_i|\theta_1)}{g(y_i|\theta_0)}$ that is a sum over pixels in $R(u)$ only, hence related to the *Hough transform* [15] in computer vision (see Illingworth and Kittler [17] for an overview).

Monte Carlo samplers can then be build by repeatedly performing these operations. Inference can be based on maximising the posterior distribution [19] or some other optimality criterion [34]. Multiresolution techniques are useful if the dimension of U becomes large. Also, simple iterative techniques similar to Besag's ICM algorithm can be constructed [7, 3, 30].

Recently, Propp and Wilson [29] realised that under certain monotonicity assumptions, it is possible to detect whether a Markov chain has reached equilibrium. The main idea is to run two Markov chains backwards in time until coalescence. The common state then yields an exact sample from the equilibrium distribution. In a point process context, exact samplers have been studied for the area-interaction model (2) by Kendall [18] and Häggström et al. [14].

Finally, the framework described in this paper is quite general and can be adapted easily to a wide range of problems involving e.g. subpixel resolution of objects, stereo pairs, motion tracking [28], edge detection and clustering of image features [20, 21] or signal analysis [38].

Acknowledgement

I am grateful to Professor Mardia for his encouragement.

References

[1] R.G. Aykroyd and P.J. Green. Global and local priors, and the location of lesions using gamma-camera imagery. *Philosophical Transactions of the Royal Society of London, Series A*, 337:323–342, 1991.

[2] A.J. Baddeley, W.S. Kendall and M.N.M. van Lieshout. Quermass-interaction processes. Research Report 293, University of Warwick, 1996.

[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, pages 271–286, Physica/Springer, Heidelberg-New York, 1992.

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

[5] A.J. Baddeley and M.N.M. van Lieshout. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, 47:601–619, 1995.

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

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

[8] 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.

[9] S.N. Chiu and D. Stoyan. Estimators of distance distributions for spatial patterns. To appear in Advances in Applied Probability, 1996.

[10] M.B. Hansen, R.D. Gill and A.J. Baddeley. Kaplan-Meier type estimators for linear contact distributions. *Scandinavian Journal of Statistics*, 23:129–155, 1996.

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

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

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

[14] O. Häggström, M.N.M. van Lieshout and J Møller. Characterisation and simulation results for two Markovian point process models. Technical Report R96-2040, University of Aalborg, 1996.

[15] P.V.C. Hough. Method and means for recognizing complex patterns. US Patent 3069654, 1962.

[16] M. A. Hurn. Bayesian image analysis in confocal microscopy. Technical Report 9601, University of Bath, 1996.

[17] J. Illingworth and J. Kittler. A survey of the Hough transform. *Computer Vision, Graphics and Image Processing*, 44:87–116, 1988.

[18] W.S. Kendall. Perfect simulation for the area-interaction point process. To appear in *L. Accardi and C. Heyde, editors, Proceedings of the Symposium on Probability towards the year 2000*. Springer-Verlag, 1997.

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

[20] M.N.M. van Lieshout. *Stochastic geometry models in image analysis and spatial statistics*. CWI tract 108, Amsterdam, 1995.

[21] M.N.M. van Lieshout and A.J. Baddeley. Markov chain Monte Carlo methods for clustering of image features. In *Proceedings 5th IEE International Conferenc on Image Processing and its Applications. IEE Conference Publication*, 410, pp. 241–245. London: IEE Press, 1995.

[22] M.N.M. van Lieshout and I.S. Molchanov. Shot-noise-weighted processes: a new family of spatial point processes. CWI Report BS-R9527, 1995.

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

[24] J. Møller. Markov chain Monte Carlo and spatial point processes. In: *Proceedings Seminaire Européen de Statistique, "Stochastic geometry, likelihood, and computation"*. O. Barndorff–Nielsen, W.S. Kendall and M.N.M. van Lieshout, editors, Chapman and Hall, 1997. To appear.

[25] J. Møller and R. Waagepetersen. Markov connected component fields. Research Report 96-2009, Department of Mathematics and Computer Science, Aalborg University, 1996.

[26] A. Pievatolo and P.J. Green. Object restoration through dynamic polygons. Technical Report S-95-12, University of Bristol, 1995.

[27] D.B. Philips and A.F.M. Smith. Dynamic image analysis using Bayesian shape and texture models. In K. Mardia and G.K. Kanji, editors, *Statistics and images, Advances in Applied Statistics, a supplement to Journal of Applied Statistics*, 20:299–322, 1993.

[28] D.B. Philips and A.F.M. Smith. Object tracking using stochastic deformable templates. Technical report TR-94-11, Imperial College, 1994.

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

[30] W. Qian and K.V. Mardia. Recognition of multiple objects with occlusions. Technical report STAT 95/01, University of Leeds, 1995.

[31] B.D. Ripley. *Statistical inference for spatial processes*. Cambridge University Press, 1988.

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

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

[34] H. Rue and A.R. Syversveen. Bayesian object recognition with Baddeley's Delta loss. Technical Report Statistics 8, University of Trondheim, 1995.

[35] K. Sivakumar and J Goutsias. Morphologically constrained discrete random sets. In: *D. Jeulin, editor, Advances in Theory and Applications of Random Sets*. Fontainebleau, 1996.

[36] J. Serra. *Image analysis and mathematical morphology*. Academic Press, London, 1982.

[37] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. Akademie-Verlag, Berlin, 1987. Second edition, 1995.

[38] A. Trubuil, H. Stryhn and M. Hoebeke. Segmentation d'images: Approche par simulation d'un processus de diffusion et de saut. Proceedings INRA, 1995.

[39] A.L. Yuille. Deformable templates for face recognition. *Journal of Cognative Neuroscience*, 3:59–70, 1991.

[40] B. Widom and J.S. Rowlinson. New model for the study of liquid-vapor phase transitions. *The Journal of Chemical Physics*, 52:1670–1684, 1970.