

THE USE OF SPATIAL MODELS AS IMAGE PRIORS

B.D. Ripley

Department of Statistics
University of Strathclyde
26 Richmond Street
Glasgow G1 1XH, UK

SUMMARY

Bayesian methods of image analysis depend on representing prior information about an image formally by probability models. Since the most useful prior information concerns the spatial and geometrical aspects of the image, it is appropriate to consider how useful the models of spatial statistics can be in this context. This is investigated in this paper, with particular reference to three examples, the deconvolution of images of galaxies, segmenting binary patterns and in extracting shapes of objects within images.

1. Introduction

The title of this symposium links two areas of statistics which have grown enormously in the last twenty years. Although image processing has a long history, it is only in the 1980's that statisticians have begun to regard imaging as a legitimate part of their subject. It is still reasonable to ask the question "what have statisticians to add to what has already been done by physicists, electronic engineers and computer scientists?" One contribution has been the formal use of Bayesian statistics to produce a unified philosophical approach to a wider spread of problems than had previously been considered. (Ulf Grenander and Donald and Stuart Geman have been particularly influential in developing this approach.) It seems to me that the prior models which have been used up to now have been chosen primarily for computational simplicity, and that it may be possible to produce better analyses by making more use of the more than a decade of experience in modelling spatial phenomena by the methods of spatial statistics.

The previous paragraph gives a very broad programme, so to illustrate it in the later sections of this paper we will concentrate on three rather restricted problems. The first is a commentary on the simple Markov random field models which have proved to be so popular. These are considered for a binary segmentation, in which the image is divided into black and white regions. Even in this simplest of problems the choice of prior model gives considerable flexibility. Figure 1 shows an example, a micrograph of a nematode. The classification of nematodes depends on the shapes of their bodies and organs, and the first step in automatic classification is to extract the nematode from the background and move it to a standard position. In many such problems the images are too noisy and are digitized to too few grey levels for conventional edge-detection methods (Rosenfeld & Kak, 1982) to be useful. It is essential to use prior information on the scale of several pixels.

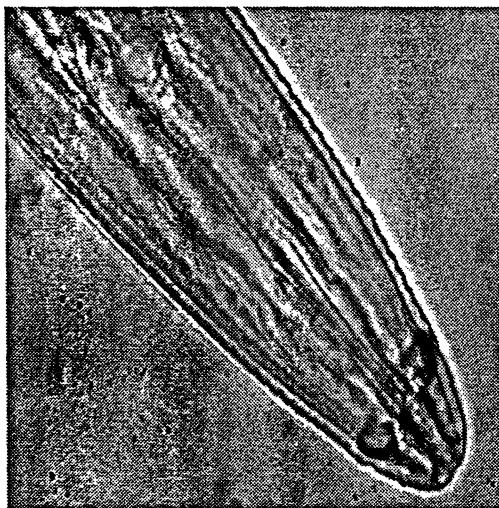


Figure 1. A micrograph of a nematode at 512×512 resolution. The original had 256 greylevels, but only 16 are shown here.

The second example comes from optical astronomy. Modern astronomy no longer uses photographic plates, but CCD (charge-coupled device) detectors. These count indirectly the number of photons of light falling on each cell of a grid (typically about

1024×700). Since astronomers will always be most interested in celestial objects at the limit of their hardware, a number of sources of noise and distortion are important, including the discrete nature of photons, electronic noise in the CCD device, and blurring caused by the motion of the earth's atmosphere during the "exposure" of several hours. The physics behind these distortion sources is quite well understood, and there is also prior information about the sources being imaged. These are made up of point sources (stars) and objects of smoothly varying luminosity (galaxies) against an almost black background. Furthermore, interest is concentrated on features in the galaxies at low luminosities.

Our third example extends the first, involving the classification of several objects in an image. Figure 2 illustrates the problem; we need both to recognize objects meeting a prior specification and 'foreign bodies' of unknown type. Eventually the astronomical problem will also become of this type, as the aim is to automatically classify galaxies according to type (ellipsoidal, spiral...).



Figure 2. A TV scan of a plate of peas with a foreign body present. The true resolution is 256×192 .

By no means all statistical methods make explicit use of prior information. One prominent exception is Shepp & Vardi's (1982) work on emission tomography. There the "maximum likelihood" solution is actually the result of a partially completed iterative procedure whose limit satisfies the likelihood equations. There are also "regularization" methods (e.g. Titterington, 1985) which use an explicit smoothness penalty. The latter can often be interpreted as using an implicit spatial prior, but I find the former unsatisfactory. There is a folklore belief that Bayesian methods are very time consuming. This may have been so for early experiments, but in many cases the algorithms have been improved to the point where their computational cost is no longer a serious consideration.

My recent essay (Ripley, 1988) gives some further examples and background for the ideas discussed here.

2. The Bayesian paradigm

The general Bayesian paradigm applies directly to abstract imaging problems. We suppose that we have a description \mathbf{S} of the image, and an observed image \mathbf{Z} which is a (known) function of \mathbf{S} and a noise process ε . The prior knowledge gives us a probability distribution $P(\mathbf{S})$ over images, and knowledge of the noise and distortion process gives us $P(\mathbf{Z} | \mathbf{S})$. From Bayes' theorem we base our conclusions on

$$P(\mathbf{S} | \mathbf{Z}) \propto P(\mathbf{Z} | \mathbf{S})P(\mathbf{S})$$

If $P(\mathbf{S})$ is expressed as a Gibbs distribution we will have

$$P(\mathbf{S}) \propto \exp -\beta U(\mathbf{S})$$

(which is just a formal change of description to an ‘energy’ U), so

$$L = -\ln P(\mathbf{S} | \mathbf{Z}) = \text{const} - \text{likelihood} + \beta U(\mathbf{S})$$

Further, in many cases the likelihood will have a scale factor ϕ representing the scale of the noise process, so

$$L = \text{const} - \text{likelihood}(\phi = 1)/\phi + \beta U(\mathbf{S})$$

and Bayesian inference is based on

$$L_0 = [-\text{likelihood}(\phi = 1)] + \beta\phi[U(\mathbf{S})] \quad (1)$$

where the two terms in brackets represent the infidelity of \mathbf{Z} to \mathbf{S} , and the “roughness” of \mathbf{S} , the extent to which it deviates from our prior perceptions. A common philosophy is to report the mode of $P(\mathbf{S} | \mathbf{Z})$, the MAP (maximum a posteriori) estimator, which corresponds to minimizing L or L_0 .

These formal manipulations are valuable in that they allow us to interpret other procedures from the Bayesian viewpoint. We can regard $\lambda = \beta\phi$ in the definition of L_0 as a Lagrange multiplier, so that (at least under convexity assumptions) minimizing L_0 is equivalent to

$$\min \quad \text{infidelity} \quad \text{subject to} \quad \text{roughness} \leq \text{constraint}$$

and to

$$\min \quad \text{roughness} \quad \text{subject to} \quad \text{infidelity} \leq \text{constraint}$$

Many methods derived from other principles fit into one of these forms, including many implementations of maximum entropy principles. Viewing $[\exp -\beta \text{roughness}]$ as a prior distribution can be a very illuminating way to consider what assumptions are being made. (This is illustrated in §4.)

This formal paradigm is so appealing that it is easy to skate over a number of crucial points in its implementation, and I believe that these have not yet been given sufficient consideration.

The description of \mathbf{S}

The ‘signal’ \mathbf{Z} will inevitably comprise values on a fixed grid of pixels, and by far the most common approach has been to represent \mathbf{S} on the same grid. Consider again figures 1 and 2. The scale of the pixels in the grey-scale images is based purely on physical constraints of the imaging device. The underlying phenomena occur in a continuum, but a description on a much coarser scale than the grey-scale pixels may be the most useful. Continuum-based models may be impossibly complicated, but it is always possible to represent the image on a finer grid than the observations. This can be particularly helpful in reconstructing boundaries and in classification problems in remote sensing where a single pixel can cover parts of several landuses. We give an example of restoration at a smaller pixel scale in §3.

Once the space for \mathbf{S} is fixed the model may need unobserved hierarchical components. This has been exploited by Grenander (e.g. 1983) in his ‘Pattern Theory’. The idea (as in figure 2) is to model a higher-level description of the image than the value at each pixel.

The loss function

It is far from clear that the MAP estimate is intrinsically desirable, and this is the subject of some considerable controversy (e.g. Marroquin, Mitter & Poggio, 1987). The MAP estimator corresponds to the loss function which penalizes all incorrect choices for \mathbf{S} equally. Another choice which has considerable support is MPM (marginal posterior mode) which chooses the modal value for S_i for each pixel i . This is appropriate for a loss function which counts the number of misclassifications if the image is described by a discrete set of values at each pixel.

One attraction of the MAP estimate is that it depends on $P(\mathbf{S} \mid \mathbf{Z})$ whereas the MPM estimator needs $P(S_i \mid \mathbf{Z})$. Grenander (1983) pointed out that the MPM estimator can often be found by sampling repeatedly from $P(\mathbf{S} \mid \mathbf{Z})$ and taking for each pixel i the most frequently occurring value of S_i . This is often computationally feasible when each pixel takes a small number of values, and is illustrated in §3.

Note that both MAP and MPM can depend crucially on the representation chosen for \mathbf{S} . A rather simple example is the astronomical deconvolution. Here we commonly look at the image on logarithmic scale, and the MAP estimator for $\ln \mathbf{S}$ is *not* the logarithm of the MAP estimator for \mathbf{S} . This problem will arise whenever we have a probability density, and we might hope to avoid it in discrete problems where we can work with actual probabilities. However, we have already commented that a discrete representation is usually an approximation to a underlying continuum image, and the character of the MAP estimate of the approximation can depend crucially on how that approximation was done. Finally, it should be obvious that measuring image characteristics on a MAP estimate does not necessarily give a good estimator of those characteristics.

Quite often MAP and MPM estimates are very similar. Where they are not there seems to be no general rule as to which to prefer. Wherever possible a more realistic loss function should be used.

Parameters in the prior

Spatial models used as image priors almost invariably contain parameters θ . In a fully Bayesian approach these should be given a prior distribution (sometimes known as

a ‘hyperprior’), and $P(\mathbf{S} | \mathbf{Z})$ is then evaluated by averaging over $P(\mathbf{S} | \mathbf{Z}, \theta)$ over θ . This has to my knowledge never been done, probably because its computational complexity can be prohibitive. One approximation is to take a fixed value of θ , that is a hyperprior without uncertainty. In practice this can work well when we have to analyze a series of similar images and have found an acceptable value of θ by experimenting on early images in the series. One way to elicit a value of θ is to experiment with simulations of $P(\mathbf{S})$. This can be misleading, as features which dominate the appearance of unconditional simulations can be swamped in the conditional distribution $P(\mathbf{S} | \mathbf{Z}, \theta)$. (A particular example is the Ising model considered in §3.)

The alternative to full Bayesian estimation is to estimate the parameter θ from the data \mathbf{Z} , an *empirical Bayes* procedure. This can also be applied to parameters (such as ϕ) in the likelihood. As (1) shows, the posterior distribution depends only on the Lagrange multiplier λ for the case of a Gibbs prior with a completely specified energy function U . In that particular case a large number of more or less *ad hoc* procedures have been proposed to estimate λ (Titterington, 1989). In the examples below we prefer to estimate β and ϕ separately as parameters.

One way to estimate β is to include the human expert in the ‘feedback loop’, and to show reconstructions with different β until the expert finds one which reflects his preconceptions. This is possibly the most effective parameter estimation method at present; in principle it could be refined to define a loss function on images \mathbf{S} and so choose β automatically.

One problem with using Markov random field priors is that their parameters are not immediately interpretable as characteristics of images on the desired scale. Theory relating medium-scale features to local parameters would be very helpful in this context (and also in others such as conditional simulation in petroleum reservoir characterization).

It is also important to remember that spatial models can be used as image priors in less formal ways than the Bayesian paradigm. In particular they can be used to select a series of image-processing operations for production analyses, by checking the effect of this filtering on simulated images to verify the designer’s understanding.

3. Simple image segmentation

We return to our first example. An image is presented made up of k classes, and we want a classified final image. Statisticians have considered this problem within the context of satellite images (e.g. Switzer, 1980; Hjort & Mohn, 1984), but the problem occurs in many other contexts, such as figure 1. Figure 3 shows an artificial example. An map of the British Isles was digitized from the output of a TV camera at 256×256 resolution to 16 greylevels. The fine detail is unreliable due to jitter in the camera and digitizer. Figure 3a shows a thresholded binary image, and figure 3b is a ‘maximum likelihood’ reconstruction of the data from aggregated data on a 64×64 grid, which we will attempt to restore at a sub-pixel scale.

The description of our image is a label S_i for each pixel on the sub-pixel size restoration grid. The observed data Z_I are continuously distributed on the large pixel grid. We assume that each large pixel is made up of r equally sized small pixels. Then

$$Z_I \sim N(\mu_I, \kappa) \quad (2)$$

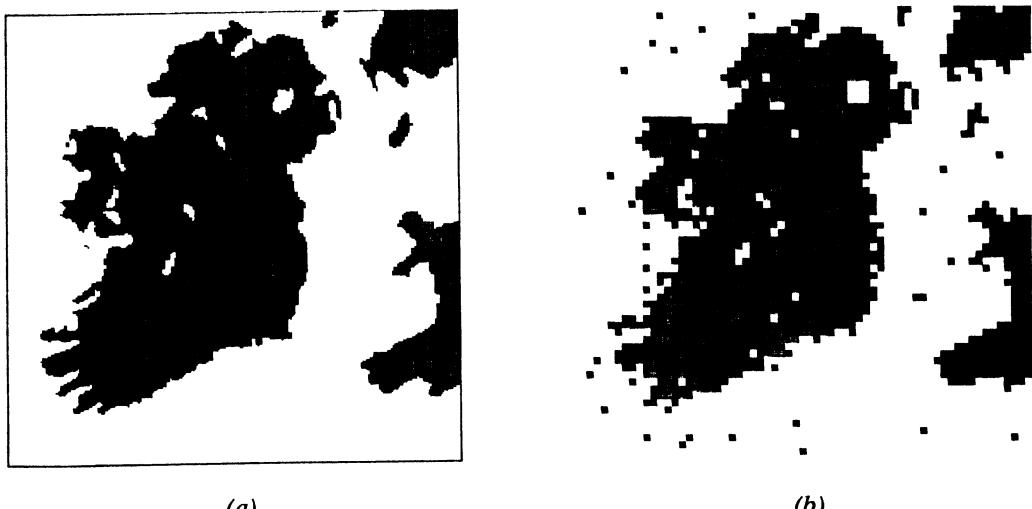


Figure 3. (a) test image on original 256×256 scale. (b) 64×64 non-spatial reconstruction.

independently for each large pixel, given \mathbf{S} . Further,

$$\mu_I = \frac{1}{r} \sum_{j \in i} \nu_S, \quad (3)$$

for *known* means ν_c for each class.

The simplest Markov random field that has been proposed as an image prior (Besag, 1983, 1986; Geman & Geman, 1984) is

$$P(S) \propto \exp \beta \# \text{ pairs of neighbors of like class}$$

In the statistical literature this model was proposed by Strauss (1977) (and also by Besag in a 1976 Princeton technical report); in statistical physics it is named after Potts (1952). The definition of ‘neighbor’ is very wide; all that is needed is a graph whose edges give exactly those pairs of pixels which are to be neighbors. In the case of four neighbors (horizontal and vertical) this process reduces to the celebrated Ising model. Figure 4 shows some simulations for this prior with four and eight neighbors (in the latter case including diagonal neighbors with equal weight). Onsager showed that for the Ising model there was a critical point β_c such that for $\beta > \beta_c$ the process possesses long-range behavior whereas its behavior for $\beta < \beta_c$ is strictly local. The exact statements are for the process defined on an infinite lattice as a limit of finite-lattice processes. Then for super-critical β realizations almost surely contain infinite patches of contiguous black pixels (and of white pixels) and the correlation between S_i and S_j does *not* tend to zero as j is moved arbitrarily far away from i . Pickard (1987) gave a lucid account of the properties of this process; some further details are given, rather cryptically, by Bartlett (1975). The value of β_c is known to be $\sinh^{-1}(1) \approx 0.88$. For values of β much larger than those shown in figure 4 the realizations are usually all black or all white. (For $\beta = 1.0$ in the 8-neighbor case we expect $64^2/(1 + e^8) \approx 1.37$ isolated pixels of the other color.)

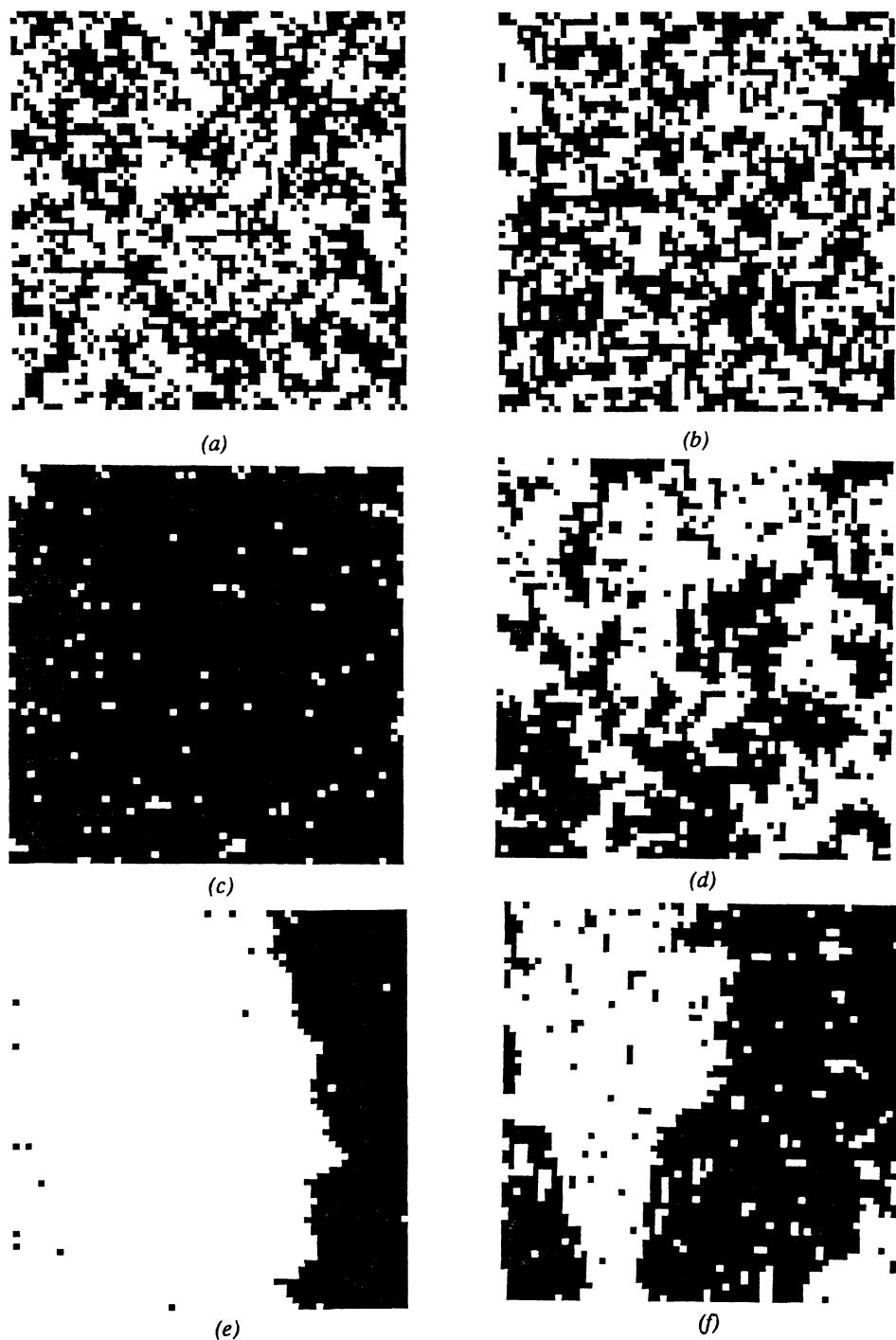


Figure 4. Simulations of the Strauss model at 64×64 pixels. (a) 8 neighbours, $\beta = 0.25$. (b) 4 neighbours, $\beta = 0.5$. (c) 8 neighbours, $\beta = 0.5$. (d) 4 neighbours, $\beta = 0.8$. (e) 8 neighbours, $\beta = 0.75$. (f) 8 neighbours, $\beta = 1.0$.

It has been suggested that these long-range correlations make the super-critical Ising model unsuitable for use as a prior in image analysis. Although formal proofs are unknown (to me), it is believed that the Potts model with both four and eight neighbors also has a critical point of the same type, so the same objection would also apply to the Strauss/Potts prior. I believe that this criticism has very limited validity. The “phase transition” applies only to the model with no “external field” in the terminology of statistical physics. An external field corresponds to altering the energy function to

$$-\left[\beta \# (\text{like neighbor pairs}) + \sum_i \alpha_i\right]$$

where $\alpha_i \neq 0$ biases the conditional and marginal distributions of pixel i towards black or white. When the pixel grids for S and Z coincide the posterior distribution $P(S | Z)$ is of this form. Further, even a very small degree of conditioning can radically alter the realizations of a Strauss prior. Thus it is not usually the long-range behavior of the prior which influences the MAP and MMP estimators. The situation is analogous to the usually benign effect of improper priors in univariate Bayesian statistics.

Simulation

The simulation of these processes is rather critical, and is part of the algorithms normally used to produce both MAP and MPM estimates. The Gibbs sampler (Geman & Geman, 1984) visits each pixel in turn and replaces S_i by a sample from $P(S_I | Z, \text{other } S_j)$. This can converge in distribution *extremely* slowly when β is super-critical and the information in Z is weak. Figure 4 is the result of around 10^{2-4} sweeps over the image. One trick which can speed up the computation enormously for large β is the *clock* method (Ripley, 1988). It is instructive to watch the Gibbs sampler in action. Almost invariably only a small proportion of the pixels change class on each sweep. Unless its neighbors have changed class, the probability that S_i changes class is constant from sweep to sweep, and the time to the next change is geometrically distributed. The clock method exploits this by setting a geometrically distributed clock for each pixel, and not considering that pixel again until its clock has timed out *or* has been reset because a neighbor has changed class.

In this form the clock method is directly applicable to simulation from $P(S)$ (to visualize the prior) and $P(S | Z)$ (to find the MPM estimator). Simulated annealing simulates from the distribution

$$P_T(S | Z) \propto P(S | Z)^{1/T}$$

with $T \searrow 0$ as the sweep number increases. Then the probability that S_i changes does vary from sweep to sweep. However, provided S_I is in its most probable class, the probability of change will decrease and so the next change will occur after the next change at constant T (if the same pseudo-random number stream is used in both cases). In my experiments I tend to reduce T in discrete steps every 10-100 sweeps. This allows the clock method to be used, just by resetting the clock on the (few) pixels not in their most probable class whenever T is reduced, and noting that when a clock does time out, the change will have a probability less than one if the clock was started with a higher T .

Restoration at the original pixel scale

The details of using a Strauss prior at the original pixel scale are well documented (Geman & Geman, 1984; Ripley, 1986, 1988). Figure 5 illustrates the results for our map on a 64×64 grid. (This is not strictly restoration, since the map was on a much finer scale originally.) With the relatively low value of β used the few isolated errors shown in figure 3b are removed, and the MAP and MPM estimates are identical with four neighbors and nearly so with eight. (The algorithms used give approximations to both MAP and MPM, so exact equivalence is a matter of chance.) The values of β used is already super-critical, but the simulations of $P(\mathbf{S} | \mathbf{Z})$ show no long-range behavior, and the iterative process merely flips single pixels on the black/white boundary without ever deviating far from the maps shown.

The number of neighbors does have an effect on the shape of the class boundaries in realizations of the prior (see figure 4) and this is reflected in the differences between figures 5a,b and 5e,f. In order to compare the two priors we matched

$$P(S_i = \text{black} \mid \text{all neighbors are white})$$

which gives double the values of β with four neighbors as with eight neighbors. Much more extensive experiments than those presented here have shown this to be a good basis for comparison. The value of β was chosen from past experience and from the geometrical arguments given in Ripley (1986). In applications such as figure 1 we want to lose the fine detail of the boundaries and so would deliberately choose a larger value of β such as is shown in figure 5c,d,g,h. Here we find much greater differences between MAP and MPM estimators, and between four and eight neighbors. The signal \mathbf{Z} no longer dominates the prior, and the simulation-based algorithms are beginning to have difficulty. Different MPM runs on 1000 sweeps can give quite different estimates, and the MAP estimate shown is the best of several runs (in the sense of giving the highest posterior probability).

K-class patterns with a Strauss prior and additive white noise are probably the most-explored example of the use of spatial priors in imaging, yet even here much more needs to be understood. My belief is that many of the reported problems (e.g. Greig, Porteous & Seheult, 1986, 1989) only occur when unrealistically noisy observations are used, allowing the prior to dominate. In those situations we need to consider alternatives to the Gibbs sampler because of its slow convergence. The clock method alleviates the problem by allowing many more sweeps to be done. Some initial experiments with multi-grid simulation methods are promising, and there are also other simulation algorithms for the prior (e.g. Swendsen & Wang, 1987)* which might be able to contribute.

Sub-pixel restoration

We now turn to our original aim, to refine the detail in figure 3b. From (2) and (3) we can deduce that

$$\ln P(S_i \mid \mathbf{Z}, \text{other } S_j) = \text{const} + \beta \#(\text{nhbrs of class i}) - \frac{1}{2\kappa} (Z_I - \frac{1}{r} \sum_{j \in I} \nu_{S_j})$$

* I am grateful to Stuart Geman for this reference.



Figure 5. Restorations at 64×64 pixels. Left column MAP, right column MMP. Rows are 8 neighbours, $\beta = 0.7$, 8 neighbours, $\beta = 2.0$, 4 neighbours, $\beta = 1.4$, and 4 neighbours, $\beta = 4.0$.



Figure 5 Continued.

where I is the large pixel containing i . In the binary case we have the log odds on black,

$$\begin{aligned} \ln P(S_i = \text{black} \mid \mathbf{Z}, \text{other } S_j) - \ln P(S_i = \text{white} \mid \mathbf{Z}, \text{other } S_j) \\ = \beta \left[\#\text{black nhbrs} - \#\text{white nhbrs} \right] \\ + \frac{1}{r^2 \kappa} \left[r Z_I - \sum_{j \neq i \in I} \nu_S, -1/2 \right] \end{aligned} \quad (4)$$

This is only marginally more complex than the case $r = 1$ where the pixel grids coincide. Equation (4) is sufficient to simulate from $P_T(\mathbf{S} \mid \mathbf{Z})$ and hence to find MAP and MPM estimators.

Figure 6 shows the reconstruction on a 128×128 grid with the same low values of β used in the 64×64 case. Once again the number of neighbors shows through in the shapes of the boundaries. A considerable amount of useful data has been recovered, although for several features one of the restorations is more appropriate than the other. This is an inevitable problem with using a prior on an image which was constructed by non-stationary mechanisms.

Discussion

These examples illustrate that the Strauss model can be very useful even though its unconditional realizations seem to bear no overall resemblance to the prior knowledge being modelled. It is the shapes of the (rare) boundaries between classes in super-critical Strauss processes which we use to fill in uncertainties in the data.

Note that the choice between four and eight neighbors and the level of boundary smoothing used (governed by $\beta\kappa$) really is a matter of prior assumption, and that in many examples (including ours) there is very little information in \mathbf{Z} about β . Further (a point we shall echo in §4), what information there is in \mathbf{Z} about β may reflect features of the prior other than those we use in the boundary smoothing.



Figure 6. Restoration at 128×128 . (a) Eight neighbours. (b) Four neighbours.

Despite all these *caveats*, the Strauss model has proved to be a very useful prior for k-class images. I suspect that if priors could be found which more accurately match our preconceptions, the computational problems with multiple maxima which bedevil MAP and MPM estimation would be greatly reduced.

The computation of restorations with simple MRF priors has been greatly improved over the years. This is not an appropriate place to go into the details of the algorithms used in our examples. It is worth noting, however, that when using simulated annealing I used an adaptive schedule based on the rate of pixel change, with a tendency towards a geometrical decrease in temperature rather than the logarithmic needed for eventual convergence. Often, too, I run several annealings with different pseudo-random number seeds and compare the posterior probabilities achieved. One will of course report the image with highest posterior probability.

Other models

The full Bayesian approach with a Markov random field prior is not the only one adopted by statisticians for the k-class classification problem. An earlier idea to introduce context into pixel classification was to choose S_i from $\{Z_j\}$ in a neighborhood of i . Thus we base inference on

$$P(S_i = c \mid Z_j \text{ for } j \text{ in a neighborhood of } i)$$

which needs a local dependence model for S . Several such models have been proposed (surveyed by Ripley, 1988, §5.5). Some researchers have used rather specialized Markov random fields such as the unilateral Markov mesh processes (Derin *et al* 1984; Devijver, 1985). Owen (1984) and Hjort & Mohn (1984) both use a simple model of a straight boundary passing through the neighborhood. Owen used Switzer's (1965) Poisson mosaic model as a prior within the current five-pixel neighborhood.

4. Deconvolution in Optical Astronomy*

We have already sketched the problem in §1. The main concern is to remove the effects of electronic and photon noise in the detector and to remove the blurring caused by atmospheric motion so as to reveal details of galactic structure. (Both distortions can be seen in figure 7.) The blurring is a linear process and so can be described by a point-spread function h which will be radially symmetric. Previous studies (Moffat, 1969, Bounanno *et al.*, 1983) have suggested the form

$$h(r) = (\beta/\pi R^2)[1 + (r/R)^2]^{-\beta} \quad (5)$$

for β about 3. Stars can usually be easily identified in displays of the digital image, so they can be extracted and the parameters in (5) fitted by (weighted) least squares. Our studies found $R \approx 3.5$, so a point source is effectively spread over a few hundred pixels. The fit of (5) was remarkably consistent from star to star and image to image.

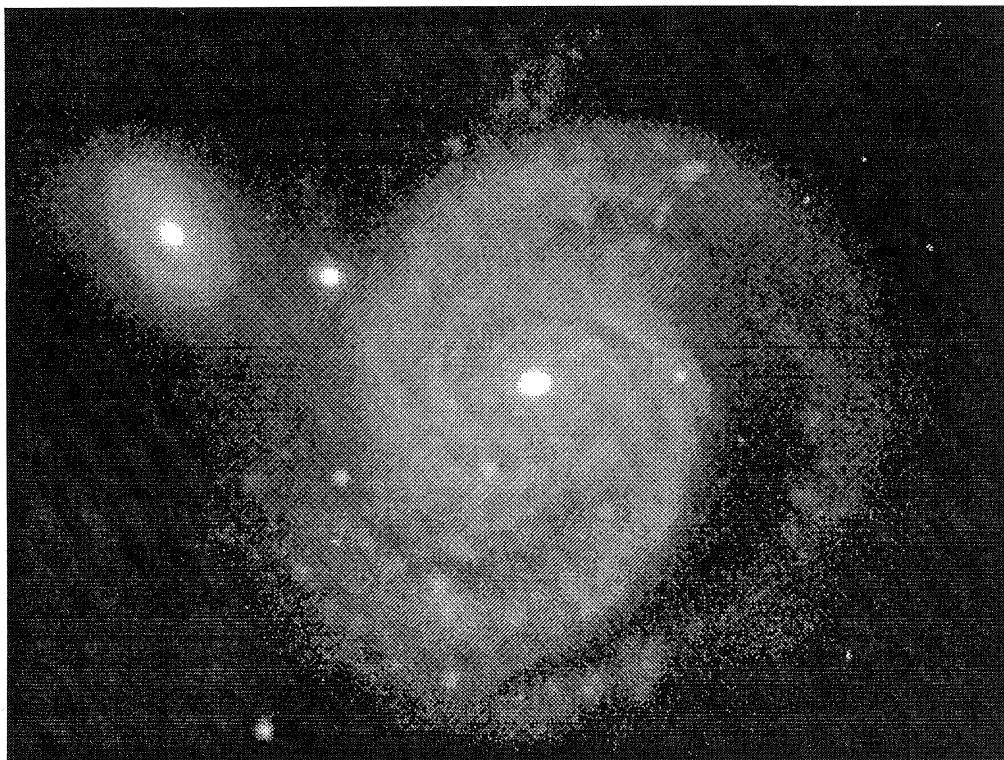


Figure 7a. A 329×256 section of a digital astronomical image. Greylevels. Counts above 2000 are shown as white.

* This section reports joint work with Rafael Molina. Later work is now published in

Molina & Ripley (1989).

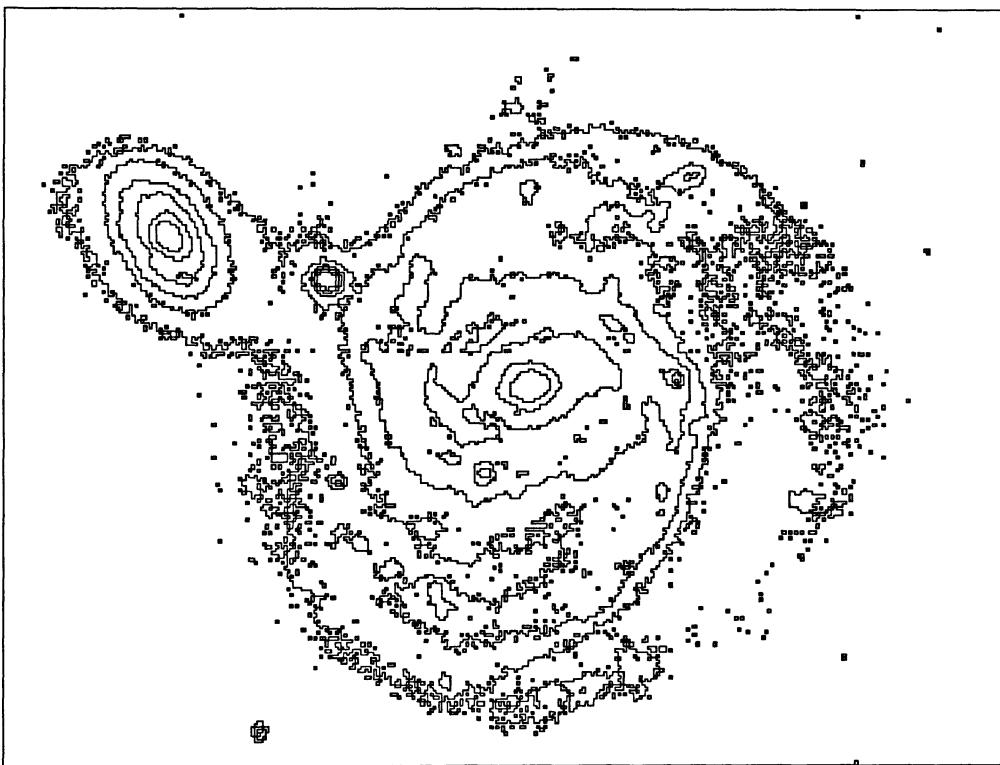


Figure 7b. Contour plot, with contours at 100, 200, 400, 800, 1600 and 3200.

The deconvolution problem is to reconstruct the image \mathbf{S} which would have been observed by a perfect detector under ideal viewing conditions. It is well-known that this is an ill-posed problem and that some smoothness constraint on \mathbf{S} is needed to achieve a satisfactory reconstruction. Common choices are a “roughness” constraint (in the sense of §2) based either on entropy or on the size of derivatives of \mathbf{S} . We chose rather to introduce an explicit spatial model for \mathbf{S} , thereby re-deriving some classical procedures and producing some new ideas. A fuller account of the procedures and detailed analysis of actual images will appear elsewhere. Here we sketch the process and concentrate on the effect of the spatial prior. The background theory is expanded in (Ripley, 1988, §5.3), and further details of the astronomical application will appear in Molina & Ripley (1989).

The noise process

The physics of the detector suggests that the main sources of noise will be Poisson noise from the discrete nature of photons, and noise from the process whereby each photon displaces electrons in the CCD cell. Thus we expect

$$\text{variance}(Z_i) = a + b \times \text{mean}(Z_i) \quad (6)$$

The true image \mathbf{S} will be non-negative, but the charge counts from the CCD detector are recorded with a DC offset and so can be negative (and often are).

Viewing the raw data revealed a number of other problems. There are lines and parts of lines with charge counts of zero or 2^{16} and isolated pixels with counts far removed from those of their neighbors. (Lines can be seen in figure 7a.) The presence of a psf ensures that these features cannot arise from the signal \mathbf{S} and so must be regarded as part of the noise process. Let $H\mathbf{S}$ denote the result of convolution of \mathbf{S} by the psf. We expect this to be quite constant at pixel scale, so

$$\mathbf{Z}^N = \mathbf{Z} - \frac{1}{r}N\mathbf{Z}$$

will have mean close to zero. (Here N denotes the neighbor incidence matrix, each pixel having r neighbors.) Then Z_i^N has mean zero and variance $(1+1/r)\sigma^2(H\mathbf{S}_i)$, but with dependent errors from pixel to pixel. We used a systematic sample of pixels and a robust measure of variance to estimate the constants a and b in (6). Difficulties arose at large mean values, where the curvature of $H\mathbf{S}$ is appreciable, but for moderate mean counts the relationship between variance and mean was acceptably linear.

Spatially smooth priors

The simplest priors for a spatially smooth digital surface are spatial autoregressions (Bartlett, 1975; Ripley, 1981). The conditional autoregression (CAR) is defined by a symmetric matrix C such that $I - C$ is positive definite, and has

$$\mathbf{S} \sim N(\mu, \kappa(I - C)^{-1})$$

A simultaneous autoregression (SAR) is defined by a matrix S such that $(I - S)$ is non-singular, and

$$(I - S)(\mathbf{S} - \mu) \sim N(0, \kappa)$$

As an SAR model (on a finite grid of pixels) can always be written as a CAR model we will only consider the CAR form in the formulae.

A simple choice for C or S which has proved to be useful in agricultural field trials is $C = \phi N$ where ϕ is less than $1/r$. For astronomical images we need ϕ very close to $1/r$ to obtain an apparently smooth realization, so C is very close to a finite-difference matrix. The corresponding SAR has

$$C = [2\phi N - \phi^2(N^2 - rI)]/(1 + \phi^2 r)$$

and corresponds to second differencing.

Realizations of these autoregressions differ from our preconceptions of the true image in at least two ways. They do not respect the constraint $\mathbf{S} \geq 0$ and demand that peaks in \mathbf{S} are as smooth as fluctuations at low levels of luminance. These deficiencies can be alleviated by using an autoregressive model for a transform of \mathbf{S} (such as $\ln \mathbf{S}$).

Algorithms

One case is particularly simple. Suppose we have additive Gaussian white noise of constant variance κ_n and a CAR prior with variance determined by κ_s . Then

$$2L_0 = \|\mathbf{Z} - H\mathbf{S}\|^2 + \frac{\kappa_n}{\kappa_s} \mathbf{S}^T(I - C)\mathbf{S}$$

which is a quadratic function of \mathbf{S} . Let $\lambda = \kappa_n/\kappa_s$, and $\alpha = \kappa_s/(\kappa_n + \kappa_s)$. Then the constrained minimum of L_0 with $\mathbf{S} \geq 0$ satisfies

$$(H^T H + \lambda I - \lambda C)\mathbf{S} \geq H^T \mathbf{Z} \text{ and } \mathbf{S} \geq 0 \quad (7)$$

with at least one equality for each pixel. The unconstrained solution is equality in the first inequality of (7), and corresponds to the Wiener filter solution in signal processing. We can transform (7) to

$$\mathbf{S} = \max(0, (1 - \alpha)C\mathbf{S} + \alpha[H^T(\mathbf{Z} - H\mathbf{S}) + \mathbf{S}]) \quad (8)$$

If the constraint is omitted equation (7) can be solved rapidly by Fourier methods, and so is taken as the starting point of our iterative schemes. Note that equation (8) is a classical constrained regularized solution, derived from a Bayesian point of view. We can solve it by first finding the unconstrained solution and then iterating (8), possibly with relaxation. (Conditions under which this scheme is stable are derived in Molina & Ripley, 1989; usually a few iterations suffice.)

The section on the noise process suggested that the noise variance κ_n depended on the mean level $H\mathbf{S}_i$. Let W denote a diagonal matrix of these variances (which thus depends, weakly, on \mathbf{S}). We find, approximately,

$$\mathbf{S} = \max(0, C\mathbf{S} + \kappa_s H^T W^{-1}(\mathbf{Z} - H\mathbf{S})) \quad (9)$$

Equation (9) omits the derivative term expressing the dependence of W^{-1} on \mathbf{S} , which is negligible. (It is exact for Poisson noise with $\kappa_n(\mu) = \mu$.) Once again we can use an iterative scheme to solve (9).

The range of observed counts is large (–200 to 20,000, say) and we noticed that astronomers tend to look at the raw data on logarithmic scale, by choosing contour levels in a geometric progression. Further, the features which interest them are often around a few hundred counts in magnitude and so are dominated by the peaks in linear plots. This suggested that we should use a CAR or SAR prior for $\mathbf{Y} = \ln(\mathbf{S} + p)$, the constant p being used to reflect the fact that very low counts are examined on linear scale. Then

$$2L_0 = \text{const} + \frac{\kappa_n}{\kappa_s} \mathbf{Y}^T (I - C) \mathbf{Y} + \|\mathbf{Z} - H\mathbf{S}\|^2$$

which leads to

$$\mathbf{Y} \geq C\mathbf{Y} + \lambda[H^T(\mathbf{Z} - H\mathbf{S})]e^{\mathbf{Y}} \quad (10)$$

where once again we have either equality or $S_i = 0$. We can combine the noise process of (9) with the prior on transformed scale to yield

$$\mathbf{Y} \geq C\mathbf{Y} + \kappa_s [H^T W^{-1}(\mathbf{Z} - H\mathbf{S})]e^{\mathbf{Y}} \quad (11)$$

All of (9) to (11) can be solved by suitable iterative schemes. Note that in (11) the effect of the factors W^{-1} and $e^{\mathbf{Y}}$ approximately cancel out, since under our noise model these are for large S_i , $1/(bS_i)$ and $S_i + p$ respectively. Relying on the local nature of H , we start the iteration for (11) from

$$\mathbf{Y} = C\mathbf{Y} + (\kappa_s/b)[H^T(\mathbf{Z} - H\mathbf{S})] \quad (11\text{bis})$$

which once again is linear and can be solved by Fourier methods.

Note that we have found the MAP estimator of \mathbf{Y} , and this does not transform to the MAP estimator of \mathbf{S} . The latter can be found from a equation similar to (11). However, since astronomers view images on logarithmic scale, we believe that the MAP estimator on that scale is most appropriate.

We are still in the early stages of experimenting with iterative schemes for the solution of these equations, and it may prove better to use direct optimization techniques. Stability does not hold for the non-linear schemes, but *with a judiciously chosen starting point* we have to date encountered few problems and only needed a few tens of iterations at most.

Other generalizations are possible within this general scheme. We mentioned that the detector faults form part of the noise process. Perhaps the simplest way to safeguard against them is to replace the least-squares “infidelity” arising from the Gaussian noise assumption by a loss function ρ from robust statistics (Huber, 1981; Hampel *et al.*, 1986). This yields versions of the same equations with $(\mathbf{Z} - \mathbf{HS})_i$ replaced by $\sigma_i \psi((\mathbf{Z} - \mathbf{HS})_i / \sigma_i)$ where σ_i is the standard deviation appropriate to that pixel, and ψ is the derivative of ρ . Again, this fits easily into the iterative form of solution.

The other component of our prior knowledge, stars, has been ignored up to now. In practice these are easy to spot and remove, so they cause no difficulty. We believe it will be possible to model our prior knowledge about stars by a Gibbsian marked point process (e.g. Ripley, 1988) and add this as a further component to the formal prior, but the full details remain to be considered.

Even in the simplest case of constant noise variance and CAR prior for \mathbf{S} the MPM estimator is not easily computed, since the distribution of S_i conditional on \mathbf{Z} is not of a simple form.

One-dimensional examples

It is much easier to see how well a reconstruction algorithm is performing in one dimension, so we start with some artificial examples. Figure 8 shows our test image, together with the effects of convolution and added noise. The spike at pixel 115 would normally be recognized as a star and removed, but is used here to see what its effects might be.

Figure 9 shows some reconstructions under the simplest model (8) with parameter κ_n calculated from the data but κ_s chosen to reflect the range of possible smoothing obtained. Figure 10 is similar, but with a prior on $\mathbf{Y} = \ln(\mathbf{S} + 100)$. (For comparison, the results are plotted on linear scale.) This is very successful, and the estimate of \mathbf{S} is quite insensitive to the value used for κ_s (here 0.05^2).

Parameter estimation

Our priors have an unknown parameter κ_s . At the beginning of this work we believed it would be possible to estimate this from the observations \mathbf{Z} . However, the experiments in (Ripley, 1988, §5.3) showed grave difficulty in estimating κ_s from \mathbf{S} ! The problem is that we do not expect the prior to represent the behavior of the true image on all scales. The parameter κ_s is the conditional variance

$$\kappa_s = \text{variance}(S_i | S_j, \forall \text{ neighbors of } j)$$

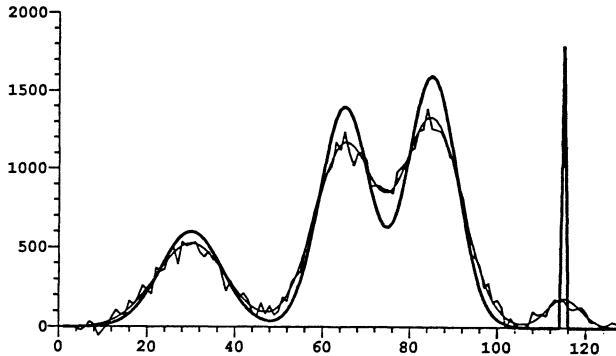


Figure 8. A test image, with scale and noise levels similar to figure 7. The thick line is the original, the thin smooth and rough lines the results of convolution and added noise respectively.

and so measures the *local* variability of \mathbf{S} . Both maximum likelihood and pseudo-likelihood estimators of κ_s are based on local characteristics, and when applied to images such as figure 8 produce rather small estimates of κ_s , which are far too low to explain the large-scale variability observed.

Ironically, this problem is eased when estimating from \mathbf{Z} . The noise process swamps the high-frequency behavior of the prior, which is also filtered severely by the convolution, so we must estimate κ_s from the low frequency behavior of \mathbf{Z} . For simplicity, consider only the simplest situation, that leading to (8). Suppose that $C = \phi N$ and that the CAR process is defined on a circle or torus and so is stationary. Spectral methods are then applicable. We can write down the log-likelihood of κ_n and κ_s as

$$-2L(\kappa_n, \kappa_s) = \sum_s \ln f_Z(2\pi s/n) + 2\pi \sum_s |\tilde{Z}_s|^2 / f_Z(2\pi s/n)$$

where n is the number of pixels, \tilde{Z} the discrete Fourier transform of \mathbf{Z} and

$$f_Z(\omega) = \frac{\kappa_n}{2\pi} + |h(\omega)|^2 f_S(\omega) \quad (12)$$

is the spectral density of \mathbf{Z} . Minimizing (12) gives estimators of κ_n and κ_s ; clearly that for κ_s is based on the lower frequencies of \mathbf{Z} .

An alternative approach can be based on moments. The maximum likelihood estimator of κ_s based on \mathbf{S} is $\hat{\kappa}_s = \mathbf{S}^T(I - C)\mathbf{S}$. Now

$$\begin{aligned} E\mathbf{Z}^T(I - C)\mathbf{Z} &= E[\text{tr}(I - C)\mathbf{Z}\mathbf{Z}^T] \\ &= \text{tr}[(I - C)(\kappa_s H(I - C)H^T + \kappa_n)] \\ &= n\kappa_n + \kappa_s \text{tr}(H^T H) \end{aligned}$$

This can be expressed more concisely in the Fourier domain. Let $c(\omega)$ denote the discrete Fourier transform of c , and $\omega_s = 2\pi s/n$. Then

$$\mathbf{Z}^T(I - C)\mathbf{Z} = \sum_s (1 - c(\omega_s)) |\tilde{Z}(\omega_s)|^2 \quad (13)$$

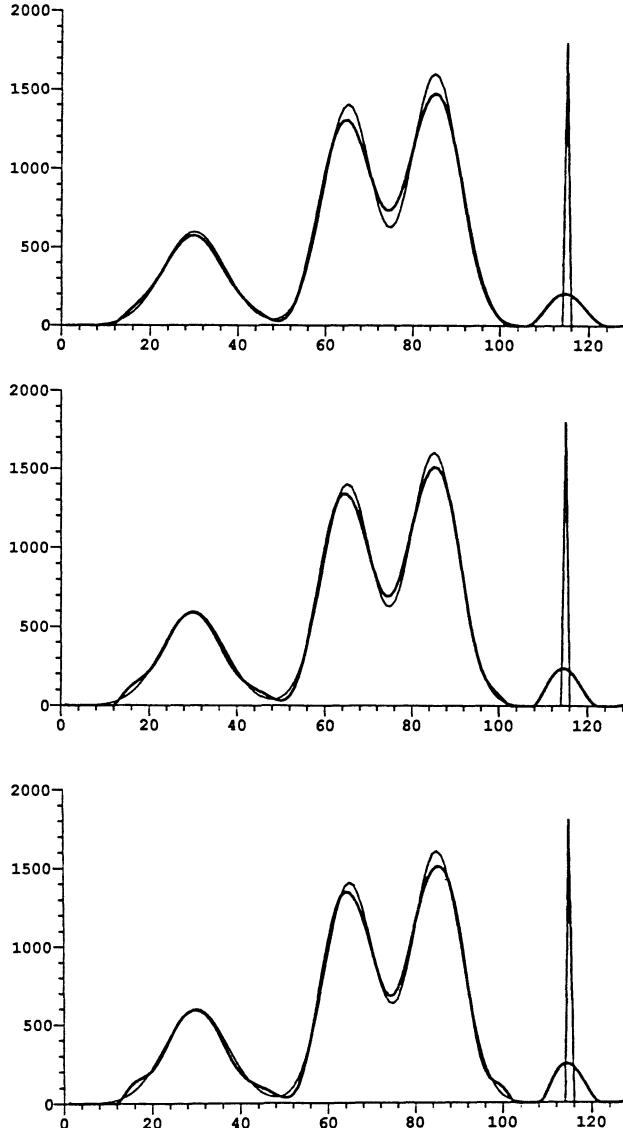


Figure 9 (a-c). Reconstruction of figure 8 by the simplest model with a range of values of $\kappa_s = 1000, 4000, 9000$. The heavy line is the reconstruction, the thin line the true image.

which has mean

$$n\kappa_n + \kappa_s \operatorname{tr}(H^T H) = \sum_s (h(\omega_i)^2 \kappa_s + \kappa_n)$$

and each term in the sum is the mean of the corresponding term in (13). This suggests an estimator of the form

$$\tilde{\kappa}_s = \sum_{0 < \omega_i < \omega_c} \left[(1 - c(\omega_i)) (\|\tilde{Z}(\omega_i)\|^2 - \kappa_n) \right] / \sum_{0 < \omega_i < \omega_c} h(\omega_i)^2$$

Both these suggestions appear to work well in cases where the constant variance and

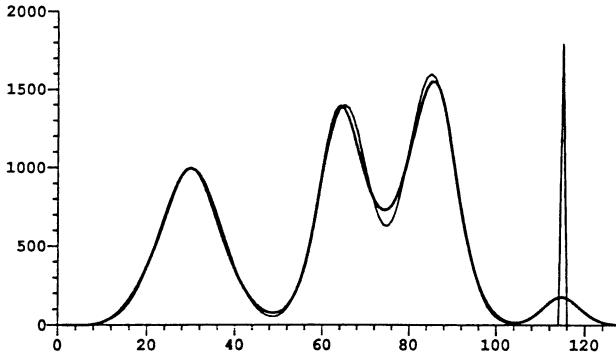


Figure 10. Reconstruction of figure 8 with a CAR prior for $\mathbf{Y} = \ln(\mathbf{S} + 100)$.

CAR prior for \mathbf{S} are appropriate. Neither generalize to the extensions needed for the astronomical problem.

Overall it seems that the best we are currently able to do with parameter estimation for κ_s is to produce a “ball park” figure, which may already be available from past experience. For many aspects of the image we are unable to chose κ_s at all accurately by trial-and-error viewing of the reconstruction, so it seems unreasonable to expect to be able to do so automatically.

Why not maximum entropy?

Various maximum entropy principles have been quite widely proposed as *the* solution to regularization problems in astronomy. The entropy functional (Gull & Skilling, 1985) is $SE = -\sum p_i \ln p_i$, where $p_i = \alpha S_i$, scaled so that $\sum p_i = 1$ and α is thus the total flux. The proposed methods either maximize entropy subject to a fidelity constraint $C \leq C_{aim}$ or minimize a Lagrangian form $-S + \lambda C$. As we have seen in §2, this can be interpreted as taking $K e^{SE}$ as a prior probability distribution on images \mathbf{S} . This prior depends only on the marginal distribution of greylevels $\{S_i\}$ and *not* on their spatial locations.

We maintain that this prior does not conform to our preconceptions about the image \mathbf{S} in that it lacks spatial smoothness. This follows from the axiomatic derivation of SE . One of the axioms states that if we partition the image, our prior for the two parts will be independent; it is this axiom which we reject. Although it is appealing when applied to well-separated parts of the image, it must also be applied to contiguous pixels to derive SE . Given that SE does not impose spatial smoothness, it may be surprising that almost all published examples of maximum entropy solutions appear to be spatially smooth. This is related to blurring, as without a blurring matrix H , maximum entropy solutions are not spatially smooth (figure 11).

Figure 12 shows maximum entropy fits for a range of Lagrange multipliers λ for our test example. Notice that the effect of the constraint $\mathbf{S} > 0$ is to inflate the solution in areas where the true value is zero, and that for a particular degree of fit maximum entropy gives a rougher solution than our spatial priors. These conclusions appear to be qualitatively correct, but not too much faith should be placed in the optimizations shown in figures 11 and 12. (They used a modified form of the ‘integral equation’ method of Gull & Daniel, starting from a smooth solution. Convergence was checked by using

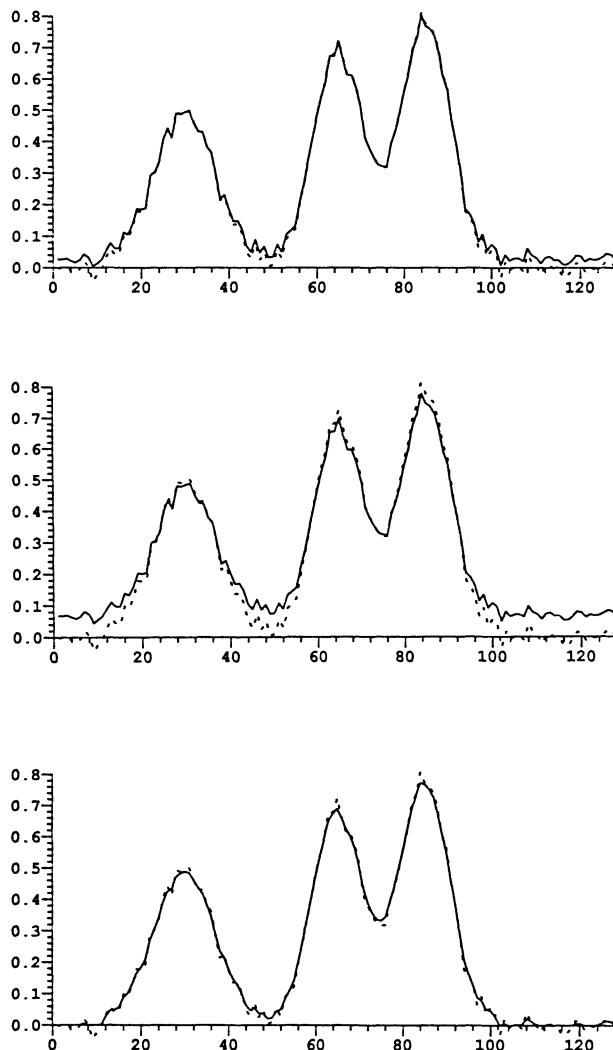


Figure 11. Reconstructions of a noisy image (shown by the dashed line) without convolution. (a–b) Maximum entropy with two values of λ . (c) Via spatial smoothing.

a number of starting points. Skilling & Gull, 1985, discuss optimization algorithms for this problem.)

Discussion

The battery of methods discussed in the sub-section on algorithms provide a number of useful ways to remove noise and blurring from an astronomical image. Our present implementations take from a few minutes to a few hours on a Sun 4 for a 512×512 image. Figure 13 shows a reconstruction of figure 7. This removes the noise very effectively and somewhat sharpens the features, except the object in the upper left surrounded by

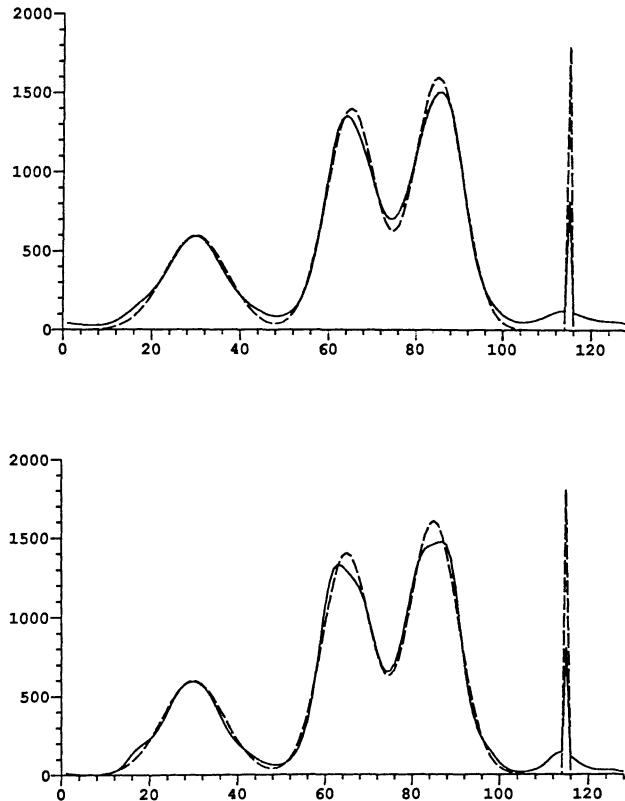


Figure 12 (a-b). Reconstructions of figure 8 (shown dashed) by maximum entropy for two values of λ .

a black ring. This may be a star or a very compact galaxy, so is too close to a point source for our methods. We believe these methods are now close to being a useful tool for our astronomer clients.

The role of the spatial prior is small in the final result, but was very important in its derivation. We could have proposed all our methods as regularization principles, but it is unlikely that we would have done so. Viewing the “roughness” penalty as a prior distribution over images has proved to be very illuminating at times during this work.

5. Object recognition

Figure 2 showed one recognition problem in computer vision, and figures 14 and 15 show some others from the same area. In each case the aim is a higher-level description of the image than the classification of each pixel. Not only do we want to say that pixel (27,63) is ‘carrot’, we want also to describe the position, size and so forth for each carrot. Note that this is not necessary for the factory inspection problem, where all that is necessary is to identify the pixels which are definitely *not* ‘carrot’ and on the

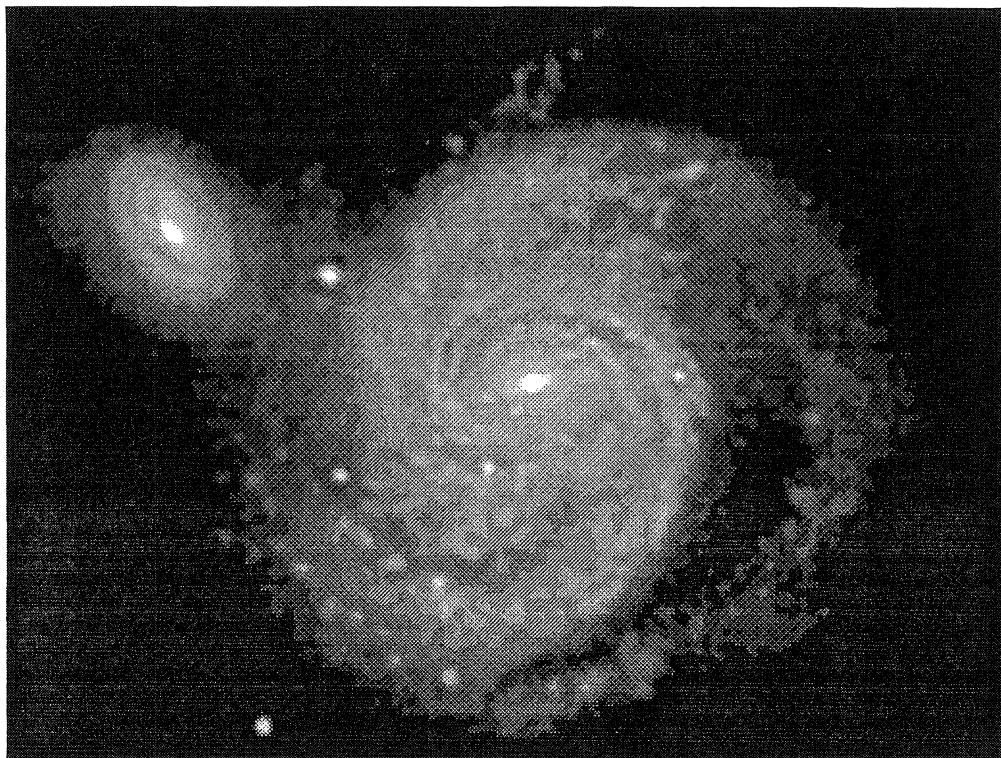


Figure 13a. A spatially smooth reconstruction of figure 7. Details as for figure 7.

basis of these make an accept/reject decision on the whole image.

Pixel-based models do not seem promising for object recognition. The only successful attempt to use them that I know is Grenander's (1983) model for horizontal, vertical L, T and + structures. Even these must be aligned with the pixel grid and have 'arms' of width exactly three pixels. The models of stochastic geometry (e.g. Stoyan, Kendall & Mecke, 1987) are a much more natural domain from which to choose our priors. In particular, our images of foods could be modeled by 'germ-grain' processes, in which 'centers' of each object are distributed by a specified point process, and the object descriptions are drawn independently. For example, the carrot slices in figure 14b could be modeled by a point process of non-overlapping discs of random radii, and the texture of each carrot by a sine-wave pattern of uniformly distributed orientation. (These models differ from the 'Boolean schemes' of Matheron and Serra (Serra, 1982) in two ways. The important one is that point processes can be much more general than a Poisson process; also we retain the knowledge of which 'grain' is uppermost when they overlap.)

Perhaps the most useful class of point processes in this context are Gibbs point processes, specified by their probability density (Radon-Nikodym derivative) with respect to a suitable Poisson process. In the spirit of stochastic geometry we think of a point process of *objects* and 'merely' have to specify an energy function U on each configuration of objects. There are iterative simulation schemes which move or replace objects at each step and have the required Gibbs point process as their equilibrium distribution

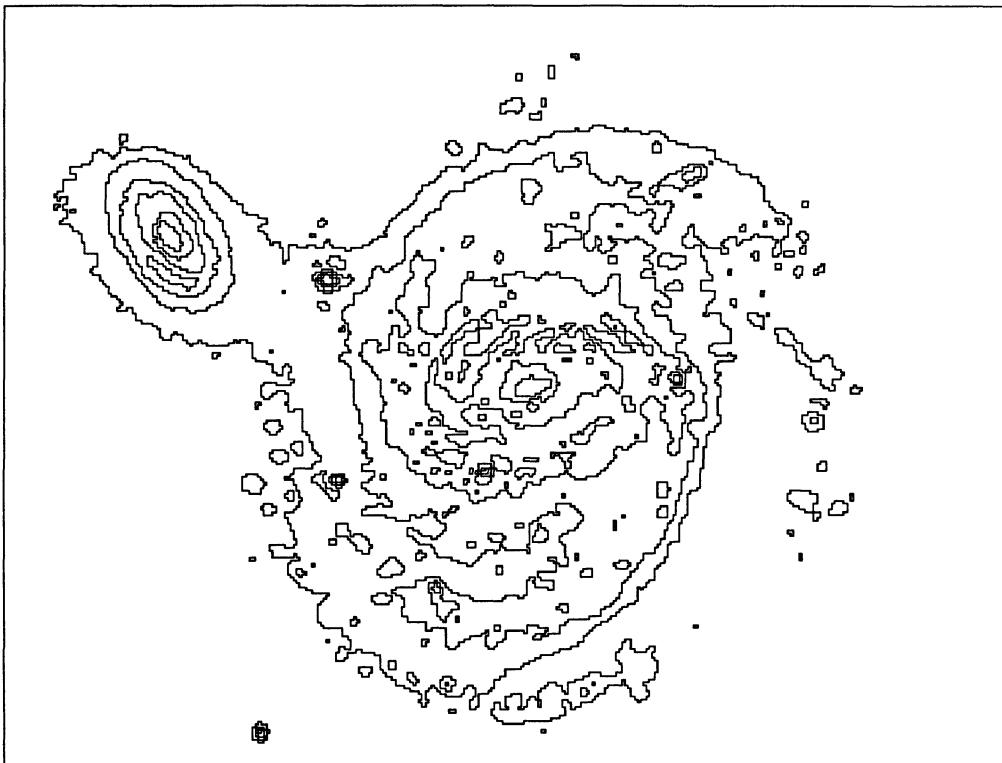


Figure 13b. A spatially smooth reconstruction of figure 7. Details as for figure 7.

(Ripley, 1977). Unlike the examples of §3 and §4, here it is the observation process $P(\mathbf{Z} | \mathbf{S})$ which it is hard to specify. Camera noise is easy to model, but an integration is needed in going from the continuum-based \mathbf{S} to the pixel-based \mathbf{Z} (although this could be ignored). The real problem is to specify the errors which can occur, since these are foreign bodies of completely unspecified form. The strategy adopted here is to leave the foreign bodies unspecified, and to examine images manually where the best fit of \mathbf{S} to \mathbf{Z} is poor. It would also be possible to include in the prior model the common types of unwanted objects (such as unsliced carrots in figure 15a); this is the approach taken in Grenander's study.

Once again it is important to choose a good starting point for the optimization of the MAP estimate. (Since the point process description of the image will involve a few tens of continuous parameters, we can use conventional optimization techniques.) To do so we return to the other use of prior models mentioned at the end of §2, the design of filters. One use of filtering would be to enhance the textures seen in figures 2 and 14. For example, in figure 15b both the peas and olives have round shapes, but the peas are shiny and so have characteristic highlights. Our experiments concentrated on the *shapes* of the objects, and used the filters of mathematical morphology (Serra, 1982; Ripley, 1988). The basic building blocks are the *erosion*

$$E_T(A) = \{\mathbf{x} \mid \mathbf{x} + T \subset A\}$$

and ‘*dilatation*’

$$D_T(A) = \{\mathbf{x} \mid \mathbf{x} + T \text{ intersects } A\}$$

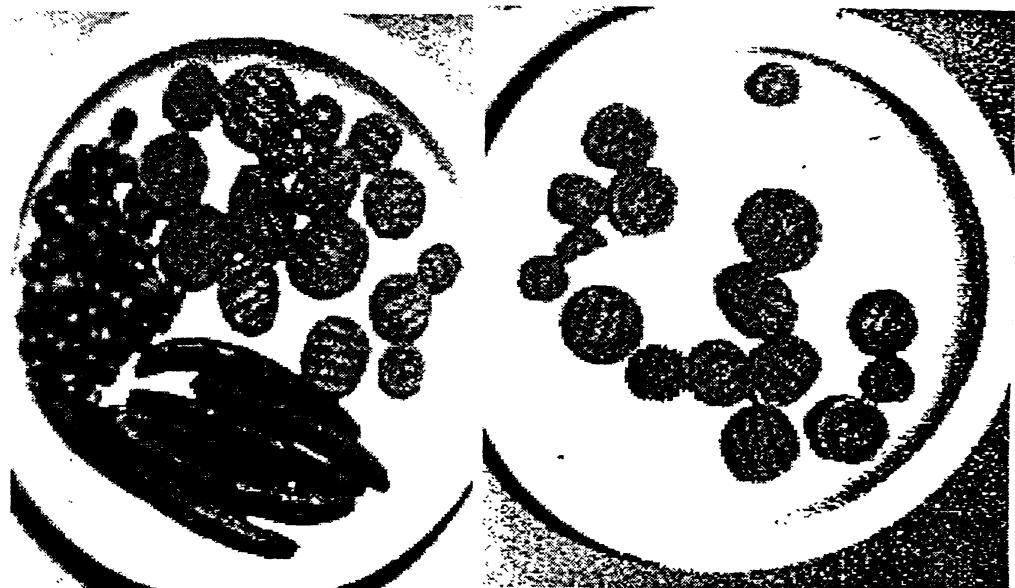


Figure 14. TV scans of foodstuffs. (a) Carrots, peas and green beans. (b) Sliced carrots (one broken).

which operate on a binary set A by a ‘test set’ T . From these we can construct the *opening*

$$O_T(A) = D_{-T}(E_T(A))$$

and *closing*

$$C_T(A) = E_{-T}(D_T(A))$$

In particular, $O_T(A)$ is contained in A and represents the ‘T-like’ part of A . There are extensions of these concepts to greylevel images by Sternberg; see Sternberg (1986) and Haralick, Sternberg & Zhuang (1987).

To illustrate these filters consider figure 14b. Figure 16a shows a binary image obtained by thresholding the greylevel image. Our prior knowledge is that cut carrots have almost perfectly circular shapes within a range of sizes. Figure 16b shows the opening by a disc of radius 8 pixels, the minimum acceptable size of a carrot slice. It is *very* easy to fit a germ-grain model to this filtered image. Further, we can mask out all pixels fitting the model (figure 16c) to identify the area(s) of poor fit. The MAP fit of a full point process model is virtually identical.

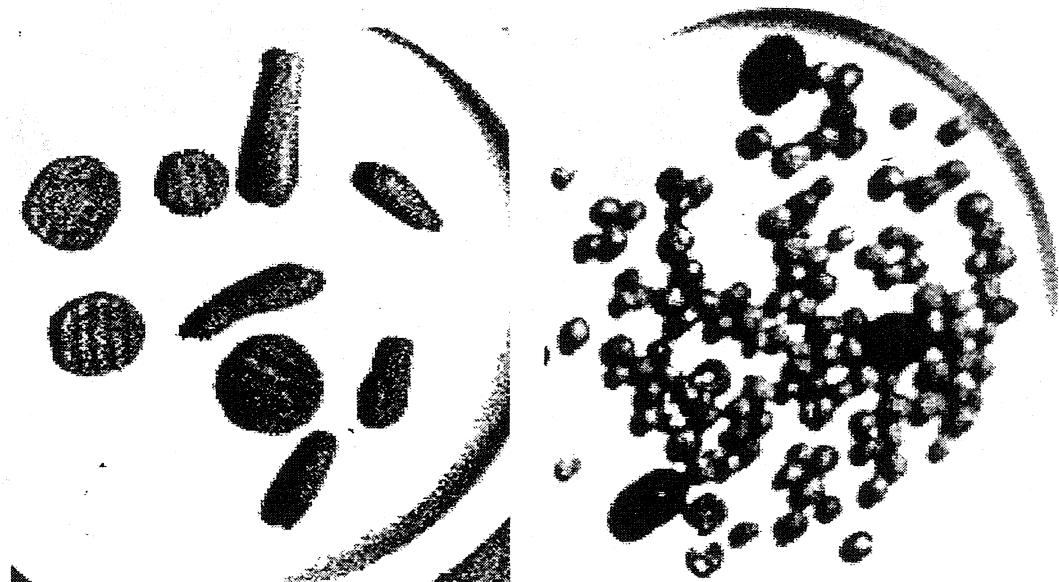


Figure 15. (a) Sliced and unsliced carrots. (b) Peas and olives.

Another example is shown in figure 17. In this case some texture (highlights on the uncut carrots) was still visible after thresholding, so a closing with radius 2 pixels was used (figure 16a). Once again an opening, this time with radius 12 to reflect a change in magnification, filters out the sliced carrots.

One advantage of the approach developed above is that it does lead, at least in theory, to a probability distribution over descriptions S of the scene. It should be possible to assign probabilities to the several most plausible descriptions and so to use a Bayes' rule with a realistic loss function to act on the description.

Continuum-based models are in their very early stages of development, but I believe will become an important part of the use of spatial models as image priors.

6. Conclusions

The use of spatial statistical models to express prior knowledge about images is in its infancy, so reaching 'conclusions' may be premature. The remarks in this section may be controversial, but they should be seen as a basis for thought and discussion.

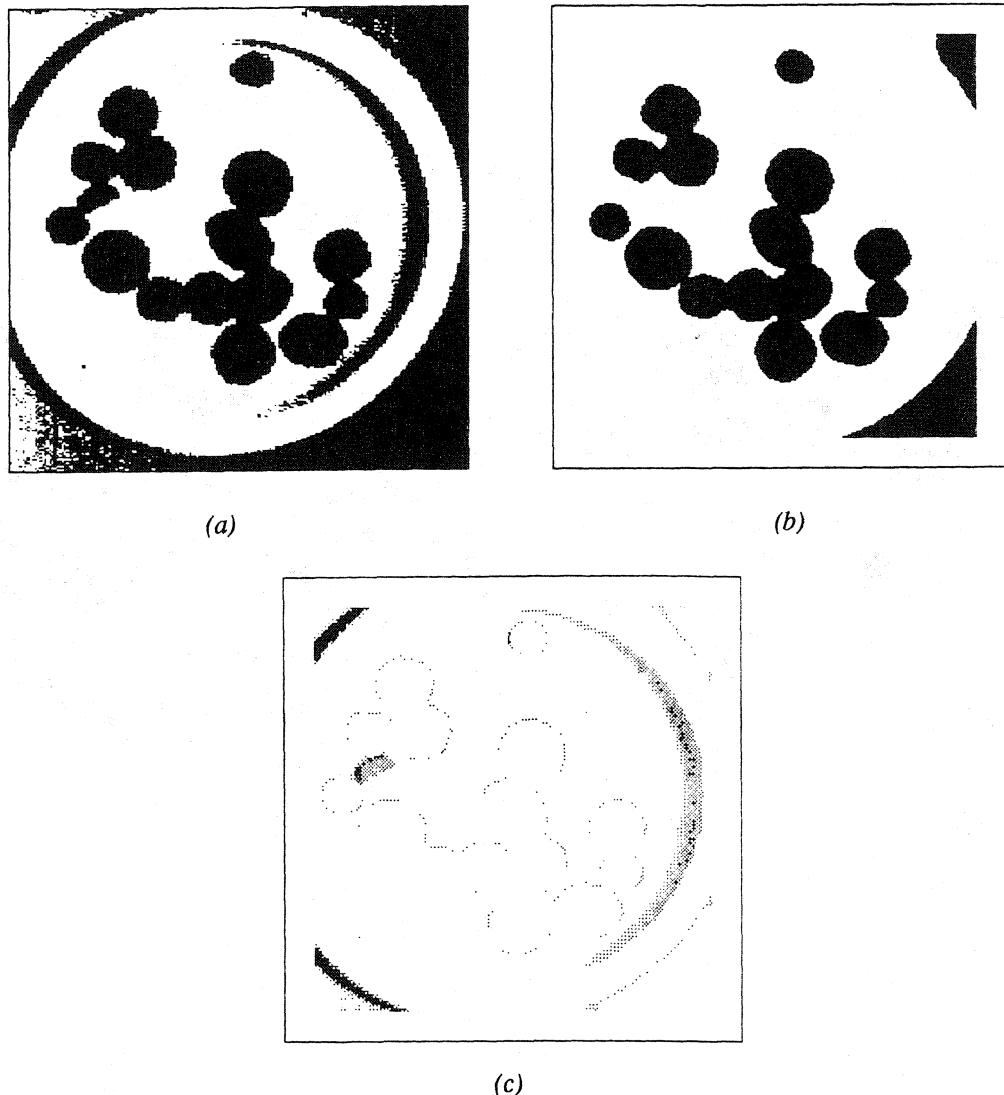


Figure 16. Filters on figure 14b. (a) Binary image obtained from thresholding. (b) Opening by a disc. (c) Masked by the opening.

- *The prior used will affect the estimate obtained.* Perhaps obvious, but not clear from the literature which pays little attention to the choice of prior.
- *The important aspects of the prior are those which cannot be estimated from the observed image.* Thus it is essential to use ‘world knowledge’ in choosing the prior and its critical parameters.
- *Not very realistic priors can be useful.* This has also been the experience with the use of spatial lattice models in field trials. Perhaps when a model is only a means to another end, its inadequacies are only of second-order importance.
- *Unrealistic models can cause severe computational difficulties in finding MAP estimates.*

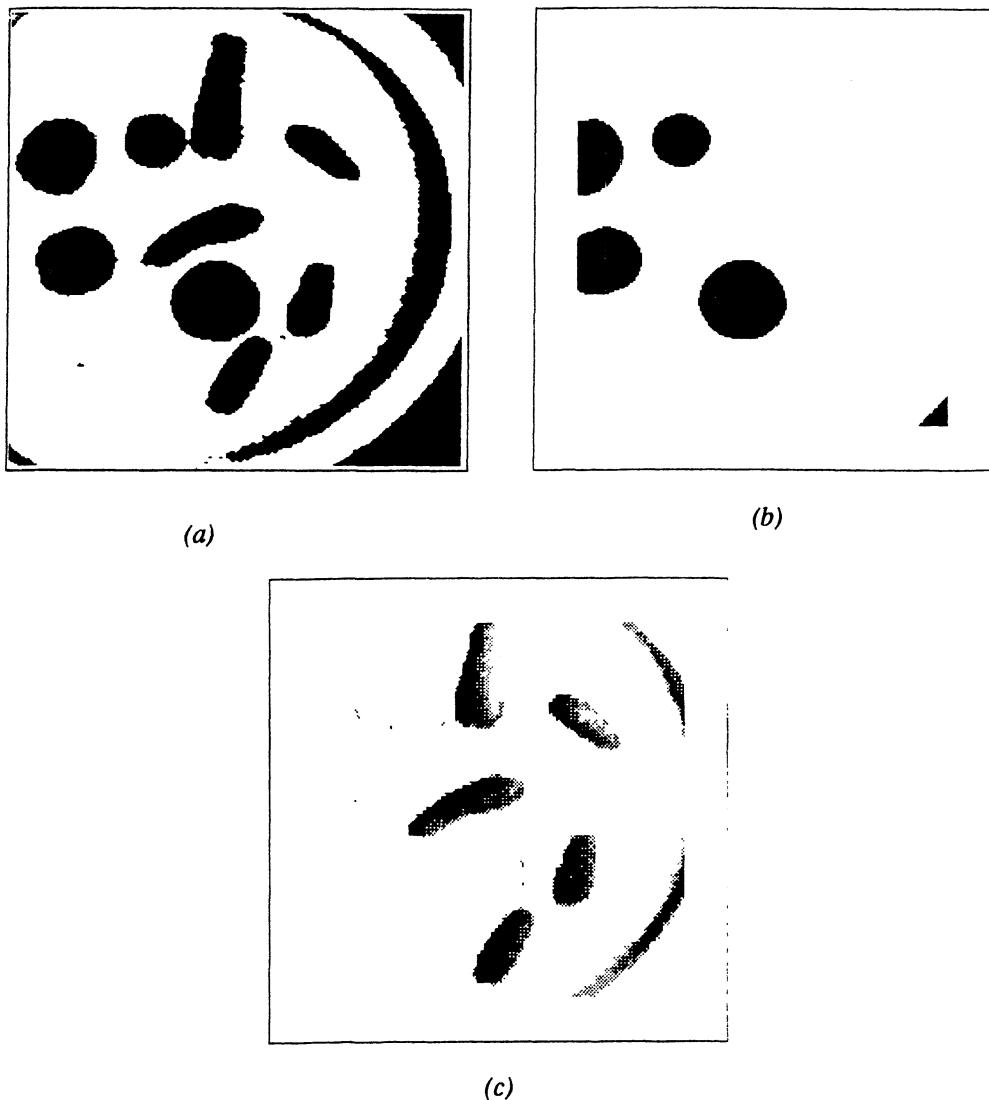


Figure 17. Filters on figure 15a. (a) Binary image obtained from thresholding and closing. (b) Opening by a disc. (c) Masked by the opening.

- *It may be better to solve the real problem than to do image reconstruction.* For example, in tomography I believe it will be easier to segment the image than to reconstruct its greylevels.
- *Better understanding of the models can lead to better algorithms.* Parallel computation is frequently mentioned as the answer to the computational complexity of image processing. However, in most current algorithms very little useful work is being done at most pixels most of the time, so naive implementations on SIMD architectures can be very wasteful. (For example, the clock method of §3 does not fit at all well with SIMD.) Load balancing should be possible on flexible MIMD machines (such as some of those based on Inmos transputers).
- *Hierarchical models will become increasingly important.*

I believe that it has already been demonstrated that the models of spatial statistics do have uses in imaging, and that they will be used more extensively in the future. Further, their uses in imaging are posing new questions about the models, such as medium-scale behavior of conditioned Markov random fields, so the interaction between the two fields will be fruitful for both.

Acknowledgements

The work described here is part of a project supported by the Science and Engineering Research Council under its ‘Complex Stochastic Systems’ Initiative. The section on astronomical deconvolution reports joint work with Rafael Molina of the University of Granada and the Instituto de Astrofísica de Andalucía, whose colleagues provided the astronomical data and the data for figure 1.

References

- Bartlett, M. S. (1975). *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London.
- Besag, J. (1983). Discussion of invited papers. *Bulletin International Statistical Institute* **50** (3), 302-309.
- Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society B* **48**, 259-302.
- Buonanno, R., Buscema, G., Corsi, C. E., Ferraro, I., & Iannicola, G. (1983). Automated photographic photometry of stars in globular clusters. *Astronomy & Astrophysics* **126**, 278-282.
- Derin, H., Elliott, H., Cristi, R., & Geman, D. (1984). Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*-**6**, 707-720.
- Devijver, P. A. (1985). *Probabilistic labeling in a hidden second order Markov mesh*. Phillips Research report.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*-**6**, 721-741.
- Greig, D. M., Porteous, B. T., & Seheult, A. H. (1986). Contribution to the discussion of Besag (1986).
- Greig, D. M., Porteous, B. T., & Seheult, A. H. (1989). Exact M.A.P. estimation for binary images. *Journal of the Royal Statistical Society B* **51**, 271-279.
- Grenander, U. (1983). *Tutorial in Pattern Theory*. Division of Applied Mathematics, Brown University.
- Gull, S. F., & Skilling, J. (1985). The entropy of an image. Smith, C. R., & Gandy, W. T. Jr. (1985, eds.), *Maximum-Entropy and Bayesian Methods in Inverse Problems*, Reidel, Dordrecht, 287-301.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel W. A. (1986). *Robust Statistics: The Approach based on Influence Functions*. Wiley, New York.

- Haralick, R. M., Sternberg, S. R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**, 532-550.
- Hjort, N. L., & Mohn, E. (1984). A comparison of some contextual methods in remote sensing classification. *Proc. 18th Int. Symp. Remote Sens. Env.* CNES, Paris.
- Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.
- Marroquin, J., Mitter, S., & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of American Statistical Association* **82**, 76-89.
- Moffat, A. F. J. (1969). A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry. *Astronomy & Astrophysics* **3**, 455-461.
- Molina, R., & Ripley, B. D. (1989). Using spatial models as priors in astronomical image analysis. *Journal of Applied Statistics* **16**, 193-206.
- Owen, A. (1984). A neighborhood-based classifier for LANDSAT data. *Canadian Journal of Statistics* **12**, 191-200.
- Pickard, D. K. (1987). Inference for discrete Markov random fields: the simplest nontrivial case. *Journal of American Statistical Association* **82**, 90-96.
- Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society* **48**, 106-109.
- Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society B* **39**, 172-212.
- Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.
- Ripley, B. D. (1986). Statistics, images and pattern recognition (with discussion). *Canadian Journal of Statistics* **14**, 83-111.
- Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.
- Rosenfeld, A., & Kak, A. C. (1982). *Digital Picture Processing*. Academic Press, Orlando.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.
- Shepp, L. A., & Vardi, Y. (1982). Maximum likelihood reconstruction in positron emission tomography. *IEEE Transactions on Medical Imaging* **1**, 113-122.
- Skilling, J., & Gull, S. F. (1985). Algorithms and applications. Smith, C. R., & Gandy, W. T. Jr. (1985, eds.), *Maximum-Entropy and Bayesian Methods in Inverse Problems*, Reidel, Dordrecht, 83-132.
- Sternberg, S. R. (1986). Grayscale morphology. *Computer Vision Graphics Image Processing* **35**, 333-355.
- Stoyan, D., Kendall, W. S., & Mecke, J. (1987). *Stochastic Geometry*. Akademie-Verlag-Wiley, Berlin-Chichester.

- Strauss, D. J. (1977). Clustering on colored lattices. *Journal of Applied Probability* **14**, 135-143.
- Swendsen, R. H., & Wang, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physics Review Letters* **58**, 86-88.
- Switzer, P. (1965). A random set process in the plane with a Markov property. *Annals Mathematical Statistics* **36**, 1859-1863.
- Switzer, P. (1980). Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Mathematical Geology* **12**, 367-376.
- Titterington, D. M. (1985). Common structure of smoothing techniques in statistics. *International Statistical Review* **53**, 141-170.
- Titterington, D. M. (1989). Choice of regularization parameter in image restoration. *This volume?*