

Bayesian Image Analysis

K. V. MARDIA

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK

Bayes' theorem is a vehicle for incorporating prior knowledge in updating the degree of belief in light of data. For example, the state of tomorrow's weather can be predicted using belief or likelihood of tomorrow's weather given today's weather data. We give a brief review of the recent advances in the area with emphasis on high-level Bayesian image analysis. It has been gradually recognised that knowledge-based algorithms based on Bayesian analysis are more widely applicable and reliable than ad hoc algorithms. Advantages include the use of explicit and realistic statistic models making it easier to understand the working behind such algorithms and allowing confidence statements to be made about conclusions. These systems are not necessarily as time consuming as might be expected. However, more care is required in using the knowledge effectively for a given specific problem; this is very much an art rather than a science.

Keywords: Deformable templates, Markov chain Monte Carlo method, Geometrical models, object recognition, data fusion

HIGH-LEVEL BAYESIAN IMAGE ANALYSIS

In the last fifteen years, statistical approaches to image analysis using the Bayesian paradigm have proved to be very successful. Initially, the methodology was primarily developed for low-level image analysis but is increasingly used for high-level tasks.

Using the Bayesian paradigm one requires a prior model which represents our initial knowledge about the objects in a particular scene and a likelihood model which is the joint probability distribution of the image, dependent on the objects in the scene. By using Bayes' theorem, we obtain the posterior

distribution of the objects in the scene, which can be used for inference, e.g. segmentation and object recognition.

In low-level image analysis, the prior could be say an Ising model, specifying that nearby pixels will tend to have similar grey levels, i.e. the scene is composed mainly of large homogeneous objects. See, for example, Besag (1986), Geman and Geman (1984) and Grenander (1993) for details of the methodology. Indeed, this approach is often referred to as a 'context'-based approach in the remote-sensing literature where the use of 'context' means that neighbouring pixel information has been

The research reported in this paper was first presented at the year long workshop on Mathematics in Medicine held at the International Centre for Mathematical Sciences, Edinburgh during 1994-1995.

used (e.g. Gong and Howarth, 1992; Dattatreya, 1991; Masson and Pieczynski, 1993). Here we shall use ‘context’ to mean neighbouring object information, an approach concentrating on high-level image analysis tasks such as object recognition. However, neighbouring pixel information will still be incorporated in the image model.

An appropriate method for high-level Bayesian image analysis is the use of deformable templates pioneered by Grenander and his colleagues, and our description follows the common theme of Mardia *et al.* (1995). We assume that we are dealing with problems where we have prior knowledge on the composition of the scene to be able to formulate parsimonious geometric descriptions for shapes in the images. For example, in medical imaging, we can expect to know what the image contains, e.g. heart, brain scans, etc. Consider our prior knowledge about the objects under study to be represented by a parameterised ideal prototype or template S_0 . Note that S_0 could be a template of a single object or of many objects in a scene. A probability distribution is assigned to the parameters with density (or probability function) $\pi(S)$, which models the allowed variations S of S_0 . Hence, S is a random vector representing all possible templates with associated density $\pi(S)$. It is through the prior model that we can express the contextual knowledge — for example, in face recognition, the location of the mouth must be approximately half-way between the eyes but not on the same level. Here S is a function of finite number of parameters, say $\theta_1, \dots, \theta_n$. We will denote S by $S(\theta_1, \dots, \theta_n)$.

In addition to the prior model, the image model or full description of image is required. Let the observed image F be the matrix of grey levels x_i where $i = (i_1, i_2) \in \{1, \dots, N\}^2$ are the $N \times N$ pixel locations. This image model or likelihood is the joint probability density function of the grey levels given the parameterised objects S , written as $L(F|S)$. It expresses the dependence of the observed image on the deformed template. It is often convenient to generate an intermediate synthetic image $G = \{g_i, i = 1, \dots, N^2\}$ which specifies how the parameterisation of S determines an image. The

specification could be the mean grey value in a region or texturing within regions, and may need to take into account aspect and occlusion in projections of three-dimensional objects. The intermediate image may be generated deterministically written as $G(S)$ or probabilistically according to the density $\pi(G|S)$. The observed image F differs from G due to, for example, noise or blurring and so does not depend directly on S , except through G , so that $L(F|G) = L(F|G, S)$. In many applications G is the reconstructed ‘true’ image and S is the interpretation. Texture is usually modelled adequately by a MRF (Markov random field) with a suitable choice of the MRF parameters for each object. Hence, in general it is possible to summarise each image feature by a unique set of texture and shape parameter values.

By Bayes’ theorem, the posterior density $\pi(S, G|F)$ of the deformed template S and generated image G , given the observed image F is proportional to

$$L(F|G)\pi(G|S)\pi(S) \quad (1)$$

if the intermediate image is generated stochastically, and

$$L(F|G(S))\pi(S) \quad (2)$$

if deterministic. Note that sometimes the construction of an intermediate image is not necessary and we have

$$\pi(S|F) \propto L(F|S)\pi(S). \quad (3)$$

In all these cases the solution to maximising the expression with respect to S and G is the maximum *a posteriori* (MAP) estimate of the true scene. The MAP is found either by a global search (which is often impractical due to the large number of parameters) or by techniques such as simulated annealing (Geman and Geman, 1984) or iterative conditional modes (ICM) (Besag, 1986). Alternatively, Markov chain Monte Carlo (MCMC) algorithms provide efficient techniques for simulating from any arbitrary posterior density.

First, details on the different aspects of prior information are given followed by discussion of the intermediate image G . Then we describe an

MCMC procedure with illustration of a simple example based on a circle which underlies practical examples, e.g. the iris of the human eye, mushroom, pellet, etc. This is followed by recent methods on multiple objects with occlusions and then a discussion of data fusion. We end with a general discussion.

PRIOR MODELS FOR OBJECTS

The key to successful inclusion of context in Bayesian image analysis is through specification of the prior distribution. Many approaches have been proposed, including methods based on outlines, landmarks, geometric parameters and Gibbs' distributions. The prior can be specified either through a model with known parameters or with parameters estimated from training data.

Grenander and co-workers have constructed a general statistical framework for image understanding using deformable templates. Most frequently, they specify a series of points around the outline of the object and these are connected by straight-line segments. Variability in the template is introduced by premultiplying the line segments by random scale-rotation matrices.

A few of the many applications considered by Grenander's group include identification of leaves in noisy images with landmarks at points of high curvature by Knoerr (1988) and models for chairs and human stomachs (Grenander and Keenan, 1989). Mardia *et al.* (1991) review certain aspects of this work and give the corresponding distribution of the conditional autoregressive model on the landmarks. Kent *et al.* (1996) give details of the underlying conditional cyclic Markov random field.

Sometimes it may be convenient to take a number of equally spaced points around the outline of the object with no identifiable features. Grenander and Miller (1994) use this method for locating mitochondria on micrographs. In this case they assume a block circulant Toeplitz covariance matrix for $(u_1, v_1, \dots, u_k, v_k)^T$ to construct a Gaussian model. Here (u_j, v_j) represents the j th edge of the outline the

object in \mathfrak{R}^2 . A feature of this application was that the number of mitochondria in the image was unknown and so a random number of objects was included in the parameterisation. Green (1996), Baddeley and van Lieshout (1994) and Mardia *et al.* (1997) also consider Bayesian models with unknown numbers of objects (see below).

Another approach is of Cootes and his colleagues where principal components are used to construct a prior model when training data is available. We formulate their principal component model for a configuration $X(2k \times 1)$ of k landmarks in \mathfrak{R}^2 as

$$X = \mu + \sum_{i=1}^p y_i \gamma_i + \epsilon \quad (4)$$

where $y_i \sim N(0, \lambda_i)$, $\epsilon \sim N_{2k}(0, \sigma^2 I)$, independently and the vectors γ_i satisfy

$$\mu^T \gamma_i = 0, \gamma_i^T \gamma_i = 1, \gamma_i^T \gamma_j = 0 (i \neq j),$$

and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. In addition, for invariance under rotation by 90° and for translation, the vectors γ_i satisfy respectively

$$\gamma_i^T v = 0 \text{ and } \gamma_i^T (1, 0, \dots, 1, 0) = 0,$$

$$\gamma_i^T (0, 1, \dots, 0, 1) = 0$$

where $v = (-\beta_1, \alpha_1, \dots, -\beta_k, \alpha_k)^T$ with $\mu = (\alpha_1, \beta_1, \dots, \alpha_k, \beta_k)^T$. Here $p \leq k$ and p is preferably taken to be quite small. Note that this approach allows us to give a model for flexible varying shapes, with often interpretable principal components. The interpretation of each component can be visualized by varying y_i in equation (4) while fixing the other $y_j = 0$, $j \neq i$. In practice, the population parameters must be estimated from a random sample.

GEOMETRIC PARAMETER APPROACH

An alternative approach is to provide a geometric template for S consisting of parametric components, e.g. line segments, circles, ellipses, arcs, etc. For example, Ripley and Sutherland (1990) use a circle

of random radius for the central disc of galaxies. Also, Phillips and Smith (1993) use simple geometric shapes for facial features, following Yuille (1991) and Yuille *et al.* (1992). Baddeley and van Lieshout (1994) use circular templates to locate pellets in an image, where the number of pellets is unknown (see also below).

In these models, distributions are specified for the geometrical parameters, and the hierarchical approach of Phillips and Smith (1993) seems particularly appropriate for context-based vision. Often templates are defined by both global and local parameters. The global parameters are on a more coarse scale and the local parameters give a more detailed description. The idea of a hierarchical model for templates is to specify the marginal distribution of the global parameters and a conditional distribution for the local parameters given the global values. This hierarchical division of parameters can be extended to give a complete description of the local dependence structure between variables. Hence, conditionally, each parameter depends only on variables in the levels immediately above and below it.

In general, we assume that templates can be summarised by a small number of parameters $\theta = (\theta_1, \dots, \theta_n)$ say, where variation in θ will produce a wide range of deformations of the template. By explicitly assigning a prior to θ , we can quantify the relative probability of different deformations. The prior can be based on training data which may not be large. By simulation, we can check the possible shapes that could arise.

For example, consider the mouth template of Phillips and Smith (1993) after Yuille *et al.* (1992). They use marginal normal distributions for (x, y) (location), θ (rotation), b (half the width) and conditional normal distributions for $a|b$ (height given half-width), $c|a$ (depth given height) and $d|a$ (curvature of parabola given height). Here x, y, θ and b are global, a is intermediate, and c and d are local.

In more complicated image scenes where several templates are required, the organisation of the templates can be considered at a higher level of hierarchy. For example, there may be nesting relationships

between the templates, which are subject to constraints. For example, with human face templates there are global constraints such as the requirement that the eyes, mouth and nose must be strictly contained within the head boundary, but this is deterministic not stochastic. We now discuss a specific example relating to a mushroom but it could be the iris of an eye in medical context.

MUSHROOM TEMPLATE MODEL AND ITS PRIOR DENSITIES

For simplicity, we regard a mushroom as a circle (Mardia 1996). A simple two-dimensional template for a circle requires centre (θ_1, θ_2) and log radius θ_3 . This is also a small component of the eye template (see Yuille 1991; Phillips and Smith, 1993). Here we have three parameters.

$$\theta = (\theta_1, \theta_2, \theta_3).$$

Next we discuss the prior distribution for θ . For an image F of size $N \times N$, say, it is simplest to take (θ_1, θ_2) to be uniformly distributed over the square $0 < \theta_1 < N, 0 < \theta_2 < N$ so that the density of (θ_1, θ_2) is simply $1/N^2$. Suppose the radius r has prior mean μ with variance σ^2 . Since $r > 0$, it is preferable to model $\theta_3 = \log r$ by a normal distribution $N(\log \mu, \sigma^2/\mu^2)$ since approximately

$$\text{var}(\theta_3) = (d \log r / d r)_{r=\mu}^2 \text{var}(r) = \sigma^2/\mu^2.$$

That is, θ_3 has a lognormal distribution. Then the joint probability density function of θ is

$$\begin{aligned} \pi^*(\theta) &= C \exp\left\{-\frac{\mu^2}{2\sigma^2}(\theta_3 - \log \mu)^2\right\}, \\ 0 < \theta_1, \theta_2 < N, \theta_3 > 0. \end{aligned} \quad (5)$$

Note that the model can be viewed as hierarchical in the sense that we can write it as

$$P(\theta_1, \theta_2)P(\theta_3|\theta_1, \theta_2)$$

so that the 'global' parameters (the location θ_1, θ_2) are followed by the 'local' parameter (the log radius θ_3). This hierarchy is not very relevant here since

θ_3 is independent of θ_1 and θ_2 , but in general it is helpful to describe location, scale and orientation as global parameters and other parameters as local deformations.

THE INTERMEDIATE IMAGE G

When objects are characterised by a grey-level pattern as well as a shape, this information may be incorporated by generating an intermediate image G which adds grey levels to the shape. One way of doing this is to use a typical grey-level image of the object, called a grey-level template G_0 , which mediates between S and G . For example, if S specifies a deformation of the image plane, then G could be the corresponding deformed version of G_0 . Another way of generating G is by filling regions determined by S with characteristic grey-level patterns.

There are at least three distinct approaches in the area of grey-level templates, eigenfaces, interpolation with control points (e.g. pair of thin-plate splines and finite elements) and warping without control points (Fourier deformation; see Amit *et al.*, 1991). An entirely different method of generating G is by parametric grey-level surfaces. We will describe some of these approaches. For details of these other approaches, see Mardia *et al.* (1995).

Deformed Grey-level Template Image

Mardia and Hainsworth (1993) work with a template image, G_0 , which is a typical (or average) grey-level image of the object and contains a set of labelled landmarks. The coordinates of the observed image are mapped to the coordinates of the template image to make the corresponding landmarks coincide. A pair of thin-plate splines (Bookstein, 1989) provides the deformation Φ and we have the generated image $g(t) = g_0(\Phi(t))$, where g and g_0 are grey levels of the image G and G_0 , respectively.

Here $S(\theta)$ contains the parameters θ as the landmarks plus a deformation parameter controlling the level of deformation. Changes in S_0 are due to landmarks which enclose a region. Once we go from S_0 to S , a deformation of G_0 to G is induced.

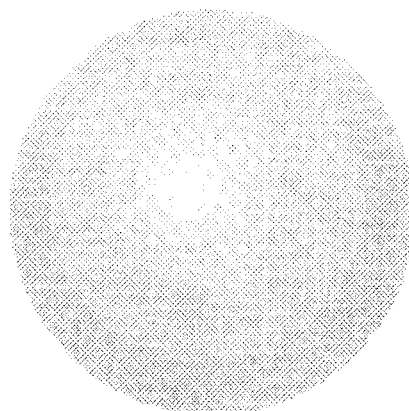


FIGURE 1 Textured image of a mushroom.

Parametric Grey-level Image

In G , bivariate surfaces of grey level positions may be used to assign grey levels to pixels. Ripley and Sutherland (1990) use a function which looks like the bivariate normal distribution for pixels inside the central disc of a galaxy. This model represents the decay of starlight away from the centre of the disc.

We explain a texture model with a simplified mushroom as an example. Recall that a mushroom is a circle, so that the circle S_0 contains a circle of radius 1 centred at the origin. Here S is the shifted scaled circle. In other examples, proper deformation will be allowed. We now allow 'texture' for the mushroom in Figure 1. The texture can be modelled in S by the surface

$$f(i, \tau) = \tau_1 + \tau_2(i_1 - \theta_1) + \tau_3(i_2 - \theta_2) + \tau_4(i_1 - \theta_1)^2 + \tau_5(i_1 - \theta_1)(i_2 - \theta_2) + \tau_6(i_2 - \theta_2)^2 \quad (6)$$

where the site $i = (i_1, i_2)$, (θ_1, θ_2) denotes the centre and τ_1, \dots, τ_6 are the regression parameters. These parameters in practice are fitted, e.g. by the least squares method. We will discuss this example further below.

INFERENCE

Inference about the scene is made through the posterior distribution of S obtained from equations (1)–(3). The full range of Bayesian statistical inference tools can be used and, as stated above,

the maximum of the posterior density (the MAP estimate) is most frequently used. Depending on the number of parameters, maximisation techniques such as a steepest ascent algorithm or simulated annealing could be straightforwardly used. Also, MCMC methods may be useful (e.g. Smith and Roberts, 1993; Green, 1996; Besag *et al.*, 1996) or approximations to the maxima such as ICM (Besag, 1986) could be used. There may be alternative occasions when particular template parameters are of great interest, in which case one would simulate the appropriate marginal posterior densities through MCMC (Phillips and Smith, 1993).

We illustrate these ideas through the mushroom example discussed before. Let us denote the circle template by $S(\theta)$ which segments the image into two regions: inside the circle S and outside S . Suppose the image is subject to observational noise. The simplest possible model is

$$x_i = v_1 + \epsilon_i \text{ if } i \in S, x_i = v_2 + \epsilon_i \text{ if } i \notin S,$$

where $x_i \in \mathcal{R}^2$ denotes the ‘grey level’ at the i th pixel in an $N \times N$ grid. In the simplest version of the model, we suppose that the ϵ are independent

$$N(0, \tau_1^2) \text{ if } i \in S \text{ and } N(0, \tau_2^2) \text{ if } i \notin S.$$

The parameters (v_1, τ_1^2) and (v_2, τ_2^2) summarise the textural difference between the object S and its background. Hence

$$P(x|\theta) = \exp \left[-\frac{1}{2\tau_1^2} \sum_{i \in S} (x_i - v_1)^2 - \frac{1}{2\tau_2^2} \sum_{i \notin S} (x_i - v_2)^2 \right]. \quad (7)$$

More realistic models might include autocorrelation between the errors or an allowance for blurring or both.

Posterior Density

By Bayes’ theorem, the posterior density of S given the data x is

$$\pi(\theta|x) \propto P(x|\theta)\pi^*(\theta).$$

Hence we get

$$\begin{aligned} \pi(\theta|x) \propto & \exp \left\{ -\frac{1}{2\tau_1^2} \sum_{i \in S} (x_i - v_1)^2 - \frac{1}{2\tau_2^2} \sum_{i \notin S} (x_i - v_2)^2 \right\} \\ & \exp \left\{ -\frac{\mu^2}{2\sigma^2} (\theta_3 - \log \mu)^2 \right\}, \end{aligned}$$

with support $0 < \theta_1, \theta_2 < N, \theta_3 > 0$. One possible estimate of θ is given by the mean of θ when the posterior distributions of $\theta|x$. One way to calculate this mean is by a simulation method which does not depend on the complicated normalising constant in $\pi(\theta|x)$. (For discrete values of θ_1, θ_2 and θ_3 , a grid search is an alternative approach.) We now describe an iterative procedure using the Hastings–Metropolis algorithm. This procedure generates a Markov chain whose equilibrium distribution is the posterior distribution of $\theta|x$.

For simplicity write $\pi(\theta|x)$ as $\pi(\theta)$ for this discussion, and choose an arbitrary initial estimate of θ . Then at each iteration, generate θ_{new} , a new set of values from

$$N(\theta_{\text{old}}, \Sigma), \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2), \text{ say,}$$

with density

$$g(\theta_{\text{new}}|\theta_{\text{old}}) \propto \exp \left\{ -\frac{1}{2} \sum_{j=1}^3 (\theta_{\text{new},j} - \theta_{\text{old},j})^2 / \sigma_j^2 \right\}, \quad (8)$$

where θ_{old} denotes the value of θ at the previous iteration. This distribution is called the ‘proposal’ distribution and its parameters $\sigma_1^2, \sigma_2^2, \sigma_3^2$ should be chosen to approximately match the variance of the posterior distribution. Calculate the ‘Hastings ratio’

$$p = \pi(\theta_{\text{new}})g(\theta_{\text{old}}|\theta_{\text{new}}) / \{\pi(\theta_{\text{old}})g(\theta_{\text{new}}|\theta_{\text{old}})\}, \quad (9)$$

where now

$$\begin{aligned} \pi(\theta_{\text{new}}) \propto \exp & \left[-\frac{1}{2\tau_1^2} \sum_{i \in S(\theta_{\text{new}})} (x_i - v_1)^2 \right. \\ & \left. - \frac{1}{2\tau_2^2} \sum_{i \notin S(\theta_{\text{new}})} (x_i - v_2)^2 \right] \\ & \times \exp\{-\mu^2(\theta_{\text{new},3} - \log \mu)^2 / 2\sigma^2\} \quad (10) \end{aligned}$$

for $0 < \theta_{\text{new},1}, \theta_{\text{new},2} < N, \theta_{\text{new},3} > 0$. Note that here the proposal density is symmetric, $g(\theta_{\text{old}}|\theta_{\text{new}}) = g(\theta_{\text{new}}|\theta_{\text{old}})$, so that $p = \pi(\theta_{\text{new}})/\pi(\theta_{\text{old}})$.

We accept $\theta = \theta_{\text{new}}$ with probability $\min(1, p)$ otherwise keep $\theta = \theta_{\text{old}}$, i.e. if $p > 1$, we take $\theta = \theta_{\text{new}}$ whereas if $p < 1$, we perform a further randomization by drawing a random sample from uniform $(0, 1)$, and accept $\theta = \theta_{\text{new}}$ with probability p . Typically, a burn-in period is allowed in initial simulations and the average is taken of the remaining simulations.

For this MCMC algorithm to work in practice, we need suitable choices for σ 's so that the proposal density will roughly approximate the posterior distribution. Also we need to judge when convergence has taken place. (See Besag *et al.*, 1996; Green, 1995.)

Above, we have updated all components of θ at once. Alternatively, it is possible to update the components of θ one at a time using individual proposal densities $g_j(\theta_{\text{new},j}|\theta_{\text{old},j})$, i.e. $\theta_{\text{new},j} \sim N(\theta_{\text{old},j}, \sigma_j^2)$. Hence we change only one component of θ at a time, i.e. in turn to complete a sweep. For this example,

$$p_j = \frac{\pi(\theta_{\text{new},j}|A_j)g(\theta_{\text{old},j}|A_j)}{\pi(\theta_{\text{old},j}|A_j)g(\theta_{\text{new},j}|A_j)},$$

where $A_j = \{\theta_{\text{new},k}, k < j, \theta_{\text{old},k}, k > j\}$, $j = 1, 2, 3$. Then we select $\theta_j = \theta_{\text{new},j}$ with probability $\min(1, p_j)$.

For a 32×32 mushroom image with 31% classification error ($v_1 = 100$, $v_2 = 120$, $\tau^2 = 400$), $\theta_1 = 11.27$, $\theta_2 = 12.67$, $r = 8.39$ and 100 000 iterations with 100 burn-in iterations, the estimated posterior means of θ_1 , θ_2 and r were found to be $\hat{\theta}_1 = 11.21$, $\hat{\theta}_2 = 12.83$, $\hat{r} = 8.47$ respectively with their respective standard errors 0.13, 0.15 and 0.10. The standard errors are the standard deviation of the sampling values of θ_1 , θ_2 , r in the MCMC and these represent an upper bound. This example highlights the strength of the MCMC method in providing information on the whole posterior density.

MULTIPLE OBJECTS AND OCCLUSION

Most of this work has been reviewed in terms of one template. However, a scene can be composed of a

known number of templates or one multiple-object template with straightforward adaptation.

In this section, S_0 will denote a collection of different type of objects with S as their deformed observed version of S . Similarly, G_0 will denote textured templates from S_0 with their observed textured objects G . Again, we will use an illustration of the recognition of image of mushrooms (Figure 2) arising in robotics in a harvesting situation. Here the parameter vector θ will be denoted by different notation depending on its context.

More difficult is the situation where an unknown number of objects are in the scene. The parameters space is then a mixture of discrete and continuous components and suitable techniques based on the Bayesian paradigm have been proposed by Grenander and Miller (1994) (using jump diffusions), Baddeley and van Lieshout (1994) (using spatial birth-death processes) and Green (1995) (using reversible jump MCMC methodology). The computational issues are somewhat complicated but nevertheless can be dealt with in reasonable computational time.

By suitably specifying the prior model and including penalty terms in the likelihood, issues such as overlap of objects or non-allowable neighbouring objects could be built into the procedure. Hence, the specification of high-level contextual information should be reasonably straightforward, although it will be very much application dependent. For example, Grenander and Miller (1994) have built in penalties to prevent overlapping of mitochondria on the micrograph images. Mardia *et al.* (1997) have provided an 'integrated' approach for occluded multiple objects of different types which we now describe. It builds on the work of Baddeley and van Lieshout (1994).

Suppose in the image F with grey levels $x = \{x_i, i \in F\}$, there are m objects c_1, \dots, c_m (m is unknown) which are any combination of q specific types (o_1, \dots, o_q) , e.g. $q = 3$ for circle, ellipse, triangle. For simplicity, let us assume initially that the objects undergo only similarity transformations so that if c_0 is a template then we observe $s(c_0)$ with some error where s denotes a similarity

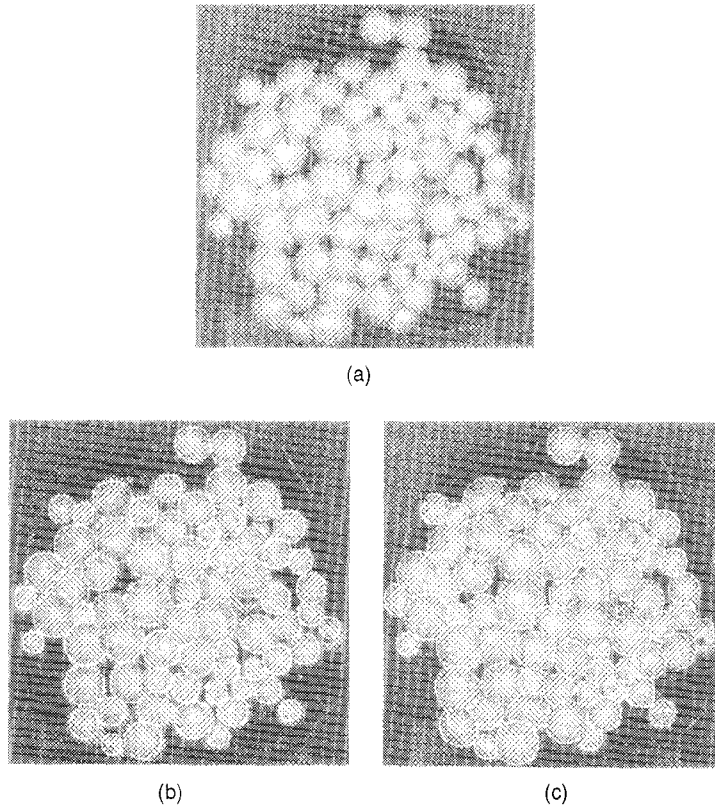


FIGURE 2 (a) Mushroom image. (b) Rigid recognition. (c) Deformed recognition.

transformation. For the m objects, c_k is a deformed version of $s_k(c_k^0)$, $k = 1, \dots, m$. Suppose each object c_k (like the mushroom before) has a textural information with regression parameters τ_i . To allow for occlusion, let there be an object configuration having order $\phi(c)$ in the scene representing for each objective c_i what is visible. Then it is possible to show that the posterior density will be given as (Mardia *et al.*, 1997).

$$\begin{aligned}
 P(c, \phi(c), \tau, \sigma^2 | x) &\propto [\prod_{k=1}^m P(s_k | c_k^0)] P((s_k | c_k^0)) \\
 &\times k = 1, \dots, m | \phi(c) [\prod_{k=1}^m P(c_k | s_k(c_k^0))] \\
 &\times P(\phi(c) | m) P(\tau | c, \phi(c)) P(\sigma^2 | c, \phi(c)) \\
 &\times l(x | c, \phi(c), \tau, \sigma^2), \quad (11)
 \end{aligned}$$

where σ^2 is the variance of the Gaussian noise present in the image F . Each term in this density is the prior information on specific aspects of the image which we now describe.

The first term of equation (11) specifies the prior probability density for the pose of an object. For simplicity, a single object is considered. Let an object be represented by a set of vertices $\{g^0(1), \dots, g^0(n)\} = c_0$. Its template S_0 can represent a number of different types of q objects. Then any affine shape s with shift μ , scale ρ and rotation θ , given by $\{sg^0(1), \dots, sg^0(n)\}$ can be described in the polar system by

$$\begin{aligned}
 sg^0(p) &= \mu + \mu_0 + \rho R^0(p)(\cos\{\theta(p) + \theta\}, \\
 &\sin\{\theta(p) + \theta\})^T, \quad p = 1, \dots, n, \quad (12)
 \end{aligned}$$

where $\mu_0 + R^0(p)(\cos \theta(p), \sin \theta(p))^T$ is the expression for $g^0(p)$. Therefore the first term in equation (11) can be written as

$$P(s | c^0) = P(\mu, \rho, \theta | \mu_0, \mathbf{R}^0) \quad (13)$$

where $\mathbf{R}^0 = \{R^0(1), \dots, R^0(n)\}^T$. This function can be described as the prior density for the pose of an

object. Note that here the parameters θ consist of μ , ρ and θ .

Suppose if we want to incorporate deformation, we can use the following prior which we describe for a single object. Consider $R(p)$, the random realization of $\rho R^0(p)$, $p = 1, \dots, n$, i.e.

$$g(p) = \mu + \mu_0 + R(p)(\cos\{\theta(p) + \theta\}, \\ \sin\{\theta(p) + \theta\})^T.$$

Then we can model $sg(p)$ as $\mathbf{R} = (R(1), \dots, R(n))^T$ with density (Mardia *et al.*, 1997)

$$P(\mathbf{R}|\rho, \mathbf{R}^0, \alpha_1, \alpha_2) \propto \\ \exp \left\{ -\frac{1}{2}\alpha_1 \sum_{p=1}^n \lambda_{1p}[R(p) - \rho R^0(p)]^2 \right. \\ \left. -\frac{1}{2}\alpha_2 \sum_{p=1}^n \lambda_{2p}[R(p) - R(p-1)]^2 \right\}, \quad (14)$$

where $\mathbf{R} = (R^0(1), \dots, R^0(n))^T$, α_1 and α_2 are two deformation parameters, λ_{1p} , λ_{2p} are fixed numbers (both usually taking values one) and reflect the fact that the different parts of the object boundary may have different degrees of susceptibility to deformation. Thus $\{R(p)\}$ is a one-dimensional cyclic Markov random field. Therefore the deformed object c can be expressed as $c = s(c^0) + \epsilon$ where ϵ is noise, dependent on α_1 and α_2 . This constitutes the third term in equation (10).

The second prior of the posterior density deals with modelling the configuration of all observable objects in a scene, where the objects may overlap. Recall that there are m objects, c_k with centers μ_k , $k = 1, \dots, m$ in the scene. It is possible to express the degree of overlap by the following prior,

$$P(c_1, c_2, \dots, c_m) = Z^{-1} \exp\{-\gamma_1 m - \gamma_2 \sum_{c_k \sim c_l} \\ d(c_k, c_l)\}, \quad (15)$$

where Z is the partition function, γ_1 and γ_2 are two parameters where γ_1 describes the potential for the presence of each object and $\gamma_2 d(c_k, c_l)$, the interaction potential between neighbouring pairs c_k and c_l . The sum is over all pairs of neighbouring objects $c_k \sim c_l$, with $k < l$, i.e. for fixed k , only overlapping

objects to c_k (neighbours) contribute to the sum. It should be noted that as γ_1 increases, the number of objects decreases whereas when γ_2 increases, overlap decreases, i.e. $d(\cdot)$ will increase when the area of overlap between two objects increases, which means that higher probabilities are allowed for less overlapping pairs. For more details see Baddeley and van Leishout (1994) and Mardia *et al.* (1997).

The next component of the posterior density describes the prior for the object order. Since objects can overlap each other in a scene, then in an object configuration, some objects may not be fully observable. Therefore it is natural to represent overlapping objects analytically. Given an image, with m objects, it is possible to label the image with an order $\phi(c) = (c_1 < c_2 < \dots < c_m)$. For example, if $c_k < c_l$, $k < l$, then it indicates that some part of the region observable by c_k is unobservable because it is below the region described by c_l . Thus to obtain a labelled image $y = y(c, \phi(c))$, using this ordering the region determined by c_k is set to state k and so on, i.e. pixels in c_k are set to k .

The next part of the posterior density deals with the texture and response functions of objects in an image. The labelled image $y = y(c, \phi(c))$ has $m+1$ states, with $y_i = 0$ representing background pixel and $y_i = k$ representing k th object in $\phi(c)$. Therefore there will be $(m+1)$ response functions $f_k(i, \tau_k)$, $k = 0, 1, \dots, m$ with parameter vector τ_k .

The final part of the posterior density describes the likelihood of an observed image $x = \{x_i, i \in F\}$ which has an underlying configuration $c = (c_1, c_2, \dots, c_m)$ with order $\phi(c)$. Then the conditional density of y given c with order $\phi(c)$ is expressed as $p(x|c, \phi(c), \tau) = \prod_{i \in F} f(x_i|c, \phi(c), \tau)$, where τ is the parameter vector of the model. This model can be a very general 'blur-free' noise model. For simplicity, consider the case of additive Gaussian noise $x_i = f(i, c, \phi(c), \tau) + \epsilon_i$, where the ϵ_i are independent Gaussian variables with zero mean and variance σ^2 and $f(i, c, \phi(c), \tau)$, the noise-free response function at pixel i . Then the labelled image $y = y(c, \phi(c))$ with $(m+1)$ response functions can be written for $y_i = k$ as $x_i = f_k(i, \tau_k) + \epsilon_i$, where τ_k , $k = 0, 1, \dots, m$, is the parameter vector for the

noise-free response function of the k th object in the order. The likelihood of x given $(c, \phi(c))$ is

$$l(x|c, \phi(c), \tau, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i \in F} [x_i - f_{x_i}(i, \tau_{x_i})]^2 \right\}. \quad (16)$$

We now discuss some other priors. For pose, we use the prior for a single mushroom (above) and we use a discrete uniform distribution for the radii. Regarding the prior for the ordering of objects, we will regard all orderings to be equally likely. For texture the prior can be taken as uniform on a restricted range and the prior for the noise variance can be a non-informative prior.

Recognition Procedure

This section describes the general strategy for object recognition with respect to the above model. The recognition process is split into two steps; the first will be called rigid or initial recognition and the second the deformation of the rigid recognition.

Rigid recognition: Assume that there is no deformation, then the reconstruction stage will involve estimating m , c_i^0 , s_i and the order $\phi(s_1(c_1^0), \dots, s_m(c_m^0))$. For this process an iterative algorithm of Baddeley and van Lieshout (1994) is used and briefly described below. The modified posterior density is then proportional to

$$[\prod_{k=1}^m P(s_k|c_k^0)] P(s_k(c_k^0), k = 1, \dots, m) P(x|s_k(c_k^0), \phi(s_k(c_k^0), k = 1, \dots, m)). \quad (17)$$

Consider a new object (s, c^0) with $(s, c^0) \neq (s_k, c_k^0)$, $k = 1, \dots, m$. The decision to accept or delete the new object thus modifying or keeping the object ordering depends if the likelihood increases. (This algorithm can be described as iterative conditional ascent (ICA) algorithm.) That is, accept/insert a new object c^0 if

$$\log[P((s_k, c_k^0), k = 1, \dots, m, c^0, \phi_{(\text{new})}|x)/P((s_k, c_k^0), k = 1, \dots, m, \phi_{(\text{cur})}|x)] > t_0$$

reject/delete the old object c_i^0 if

$$\log[P((s_k, c_k), k = 1, \dots, i-1, i+1, \dots, m, \phi_i|x)/P((s_k, c_k), i = 1, \dots, m, \phi_{(\text{cur})}|x)] > t_0,$$

where t_0 is a threshold which is selected empirically by trial and error, $\phi_{(\text{cur})}$ is the current value of $\phi(c)$, and ϕ_i is ϕ after deleting object i . Thus a new order $\phi(c)$ with its corresponding configuration is obtained. More details can be found in Mardia *et al.* (1997). The steps can be implemented using either the coordinate-wise optimisation or the steepest ascent method.

Deformation of the rigid recognition: Using the information obtained from the initial recognition process, the deformed recognition is carried out by deforming the objects using the prior given by equation (4). In the above procedure, the parameters $\alpha_1, \alpha_2, \gamma_1, \gamma_2, \tau_i$ have been assumed to be known. To estimate these parameters, the general iterative procedure of Baddeley and van Lieshout (1994) can be employed at appropriate stages leading to either a pseudo-likelihood estimator or a maximum likelihood estimator. Full details are given in Qian and Mardia (1995).

Mushroom recognition

Consider an image of mushrooms in a growing bed (Figure 2), captured by a video camera. The resolution is of 512×512 and each pixel is represented in one of the 256 grey levels. The following assumptions are made: (a) the mushroom surfaces can be modelled by response functions given before by equation (6); (b) the mushroom boundaries are deformed versions of the circle so there is only one pure object; (c) all pixels are possible locations, with seven possible radii, ranging from 11 pixels to 17 pixels; and (d) the interaction function in the object process is given by

$$d(\mu_i, \mu_j) = \begin{cases} \infty & \|\mu_i - \mu_j\| < 7 \\ 1 - \frac{\|\mu_i - \mu_j\|}{\rho_i + \rho_j} & 7 \leq \|\mu_i - \mu_j\| < \rho_i + \rho_j \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

As seen in Figure 2(a), mushroom surfaces occupy a large part of the image frame. The initial recognition process was carried out from

the configuration of an empty set, which gave the estimate of the background grey level to be very large. The iterative process was carried out as follows: For each pixel, the average of observations over its 24 neighbours in a 5×5 window was calculated and if this average was > 100 , a mushroom of size 14 at this pixel was assigned. If the distance between two such mushrooms was < 7 , one of them was deleted at random. Thus an arbitrary order for these mushrooms was obtained and a corresponding labelled image was created. Using the method of least squares, the regression coefficients were estimated. The two parameters γ_1 and γ_2 in the Markov object process (equation (15)) were both set to 100, since there were many overlapping objects. The value of $t_0 = 10$ was selected after trial and error.

Figure 2(b) shows the result of the above procedure after ten iterations. In fact, the restored configuration does not change after six iterations. From Figure 2(b) it was observed that almost all the mushrooms were picked out at approximately the right locations. Therefore their deformations in the recognition procedure were considered, but without adding/deleting any object. A search around each current mushroom to find the best location, the best size and the best order was done. The initial order of the configuration is that shown in Figure 2(b). The two deformation parameters α_1 and α_2 are both set to be 1. The result after three iterations is shown in Figure 2(c). Note that satisfactory results of locations and sizes have been obtained, as well as the boundaries of those mushrooms. In particular, from these reconstructions we can now obtain the size distribution of the mushrooms. Further work involves making the algorithm more efficient as well as use of an MCMC method in estimating the size distributions of mushrooms, etc. from the posterior density. Again, note that the same technique applies to multiple objects as in cell recognition.

DATA FUSION

Special problems arise in fusing different modalities. Hurn *et al.* (1996) and Mardia *et al.* (1996a) have

given a review as well as proposing a method of fusing images assuming that they are already registered. The algorithm is developed within a hierarchical Bayesian framework and modelled using a Markov random field (MRF). There are three stages to the hierarchy. At the highest level, it is assumed that there exists a super-population image, Z , which is a fused classification image, and can be described as the 'truth underlying the data'. The prior knowledge that the ground truth is a classified image can be modelled by the Ising model, with a smoothing parameter, β say. The second level of the hierarchy contains ideal images, X_i , which are essentially the super-population image observed under M different modalities. Finally, the lowest level represents the data images, Y_i , which are ideal images that have been degraded in some way due to the process by which the data are recorded. A Gaussian form is used to model the relationship between each level of the hierarchy. The full posterior density is,

$$P(Z, \{X_i\}|\{Y_i\}) \propto \exp \left\{ - \sum_{i=1}^M \frac{1}{2\sigma_i^2} |Y_i - H_i \circ X_i|^2 - \sum_{i=1}^M \frac{1}{2\gamma_i^2} |X_i - K_i(\mu_i(Z))|^2 - \beta \sum_{\langle i,j \rangle} I_{[Z_i \neq Z_j]} \right\},$$

where $i = 1, \dots, M$, $\langle i, j \rangle$ sums over the eight neighbours of pixel i and σ^2 and γ^2 represent the signal noise variance in the data and error variance of the super-population image respectively. H is a simple blurring kernel applied to pixels in the ideal images and K is a mapping operator which enables the comparison of pixels in the ideal images with groups on pixels in the fused image. The solution can be found through iterative conditional modes (ICM).

The method is illustrated using a pair of medical images — Two magnetic resonance images (MRI) obtained from different acquisition techniques. Figures 3(a) and (b) show two MRI data-scan images of size 256×256 obtained using different T_1 - and T_2 -weighted spin echo acquisition, respectively. It can be seen that the contrast between the tissue types, white and grey matter, is different for each

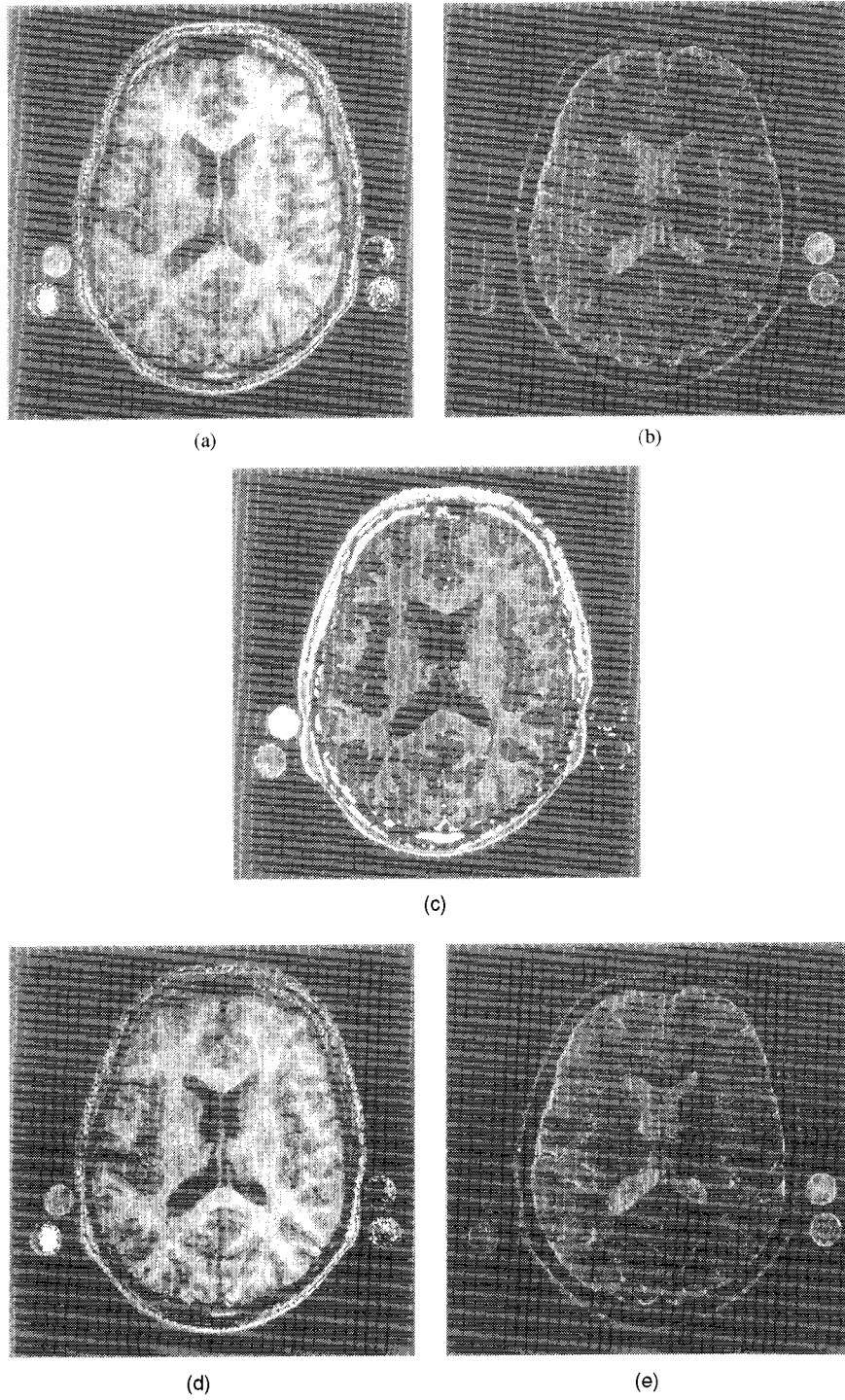


FIGURE 3 Observed Magnetic resonance images: (a) observed 256×256 MRI; (b) observed 256×256 MRI; (c) final classification; (d) reconstruction for first modality; (e) reconstruction for second modality.

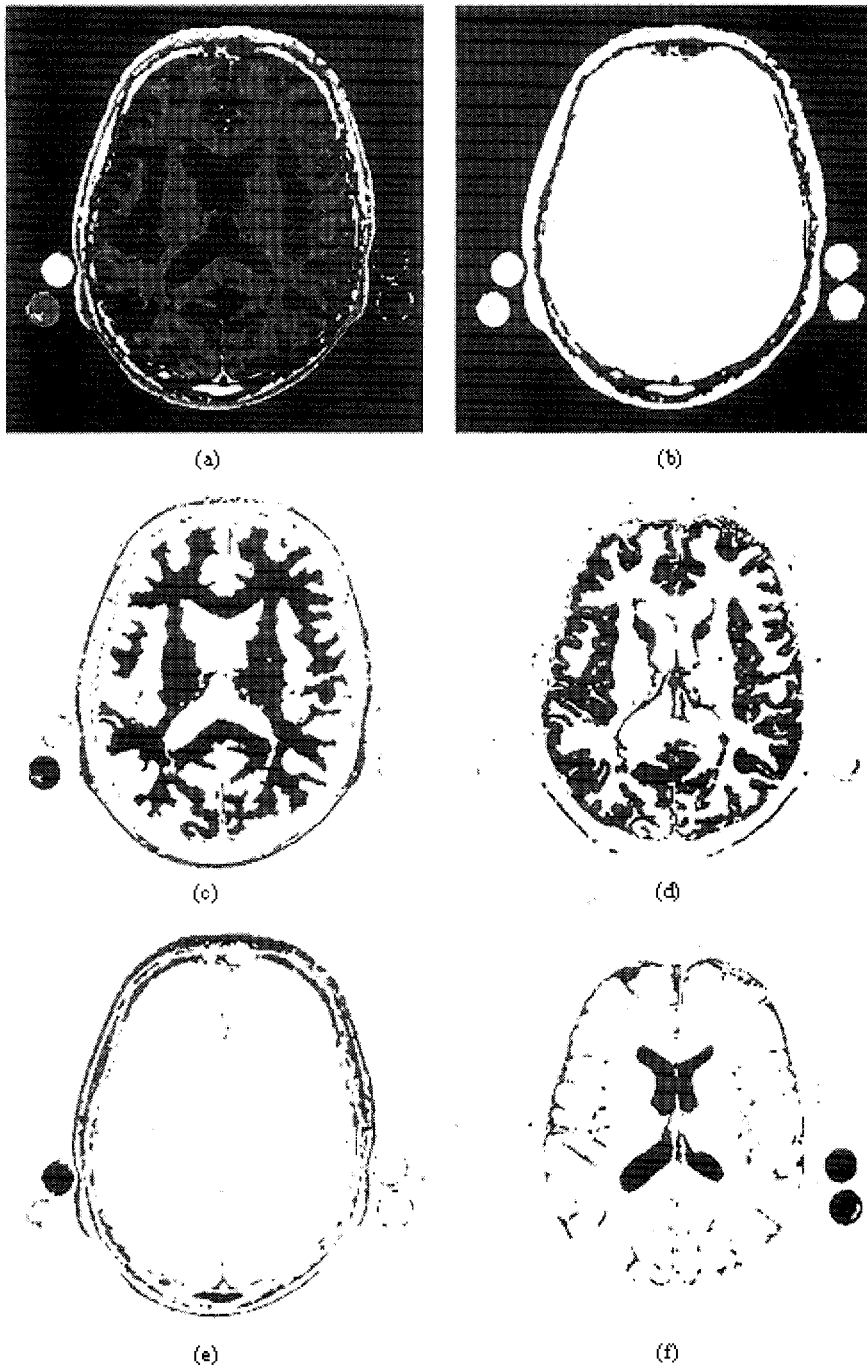


FIGURE 4 Classification showing breakdown of individual classes: (a) classes Z (512×512 pixels); (b) 'air'; (c) 'skull'; (d) 'CSF'; (e) 'grey matter'; (f) 'white matter'.

image and that the cerebrospinal fluid (CSF) (dark grey regions in Figure 3(a)) seems more defined in Figure 3(a) than in (b), due to the contrast between the CSF and the surrounding tissue. Figures 3(d) and (e) show the intermediate ideal images and the final fused classification image is given in Figure 3(c). The individual classes of the fused image Figure 3(c) have been given in Figures 4(b)–(f) and these represent the classes air, skull, CSF, grey matter and white matter, respectively.

DISCUSSION

We have concentrated on some aspects of high-level imaging with parametric deformable templates. There are various other developments such as for image sequences (see, for example, Mardia *et al.*, 1992 and Phillips and Smith, 1993). Special problems arise such as in functional imaging. Performance of methodology and robust methods for imaging is another area (see Haralick and Meer, 1994). Various consortiums are developing a human atlas (see Grenander and Miller, 1994). There is considerable statistical advances in tomography reconstruction problems (see Green, 1994). For other advancements see, for example, Mardia and Kanji (1993), Mardia (1994), Mardia and Gill (1995) and Mardia *et al.* (1996).

For recent work on non-parametric deformable templates see Jain *et al.* (1996) and for graphical templates for model registration see Amit and Kong (1996). For vehicle segmentation and classification using parametric deformable templates see Mardia *et al.* (1992) and Dubuisson Jolly *et al.* (1996). The discussion paper by Besag *et al.* (1996) on MCMC methods is highly recommended. A new innovation in the MCMC strategy is exact sampling with coupled Markov chains of Propp and Wilson (1996) which describes on its own when to stop and that outputs sample in exact accordance with the desired distribution.

Acknowledgements

I am grateful to John Kent, Druti Shah, Ian Dryden, Kevin De Souza and Jayne Kirkbride for their help.

The work is partly supported by research grants from the BBSRC and the EPSRC.

References

- Amit, Y., Grenander, U. and Piccioni, M. (1991) "Structural image restoration through deformable templates," *J. Amer. Statist. Assoc.*, **86**, 376–387.
- Amit, Y. and Kong, A. (1996) "Graphical templates for model registration," *IEEE Trans. Pattern Anal. Machine Intell.*, **18**, 225–236.
- Baddeley, A. J. and van Lieshout, M. N. M. (1994) "Stochastic geometry models in high-level vision." In K. V. Mardia, ed., *Statistics and Images: Vol. 2*. Carfax, Oxford.
- Besag, J. E. (1986) "On the statistical analysis of dirty pictures (with discussion)," *J. Roy. Statist. Soc. B*, **48**, 259–302.
- Besag, J. E., Green, P., Higdon, D. and Mengersen, K. (1996) "Bayesian computation and stochastic systems (with discussion)," *Statistical Science*, **10**, 3–66.
- Bookstein, F. L. (1989) "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Machine Intell.*, **11**, 567–585.
- Dattatreya, G. R. (1991) "Unsupervised context estimation in a mesh of pattern classes for image recognition," *Pattern Recognition*, **24**, 685–694.
- Dubuisson Jolly, M. P., Lakshmanan, S. and Jain, A. K. (1996) "Vehicle segmentation and classification using deformable templates," *IEEE Trans. Pattern Anal. Machine Intell.*, **18**, 293–308.
- Geman, S. and Geman, D. (1984) "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, **6**, 721–741.
- Gong, P. and Howarth, P. J. (1992) "Frequency based contextual classification and gray-level vector reduction for land-use identification," *Photogrammetric Engineering and Remote Sensing*, **58**, 423–437.
- Green, P. J. (1994) "Statistical aspects of medical imaging." Special issue of *Statistical Methods in Medical Research*, **3**, 1–101.
- Green, P. J. (1995) "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, **82**, 711–732.
- Green, P. J. (1996) Markov chain Monte Carlo in image analysis. In W. W. Gilks, S. Richardson and D. Spiegelhalter, eds, *Practical Markov Chain Monte Carlo*. Chapman and Hall, London.
- Grenander, U. (1993) *General Pattern Theory*. Clarendon Press, Oxford.
- Grenander, U. and Keenan, D. M. (1989) "Towards automated image understanding," *J. Appl. Statist.*, **16**, 207–221.
- Grenander, U. and Miller, M. I. (1994) "Representations of knowledge in complex systems (with discussion)," *J. Roy. Statist. Soc. B*, **56**, 285–299.
- Haralick, M. R. and Meer, P. (1994) *Performance versus Methodology, Proceeding NSF/ARPA Workshop '94, Seattle, Washington*.
- Hurn, M. A., Mardia, K. V., Hainsworth, T. J., Kirkbride, J. and Berry, E. (1996) "Bayesian fused classification of medical images," *IEEE Trans. Med. Imaging*. In preparation.
- Jain, A. K., Zhong, Y. and Lakshmanan, S. (1996) "Object matching using deformable templates," *IEEE Trans. Pattern Anal. Machine Intell.*, **18**, 267–278.
- Kent, J. T., Mardia, K. V. and Walder, A. N. (1996) "Conditional cyclic Markov random fields," *J. Appl. Prob.*, **28**, 1–12.

- Knoerr, A. P. (1988) "Global models of natural boundaries: theory and applications," *Reports in Pattern Analysis 148*. Division of Applied Mathematics, Brown University, Providence, RI.
- Mardia, K. V. (ed.) (1994) *Statistics and Images*, Vol II. Carfax, Abingdon.
- Mardia, K. V. (1996). The art and science of Bayesian object recognition. In *Image Fusion and Shape Variability Techniques*, K. V. Mardia, C. A. Gill, and L. L. Dryden, eds, pp. 21-35. Leeds University Press, Leeds.
- Mardia, K. V. and Gill, C. A. (eds) (1995) *Current Issues in Statistical Shape Analysis*. Leeds University Press, Leeds.
- Mardia, K. V. and Hainsworth, T. J. (1993) "Image warping and Bayesian reconstruction with grey-level templates." In K. V. Mardia and G. K. Kanji, eds, *Statistics and Images: Vol. 1*, pp.257-280. Carfax, Oxford.
- Mardia, K. V. and Kanji, G. K. (eds) (1993) *Statistics and Images*, Vol. I, 336 pp. Carfax, Abingdon.
- Mardia, K. V., Kent, J. T. and Walder, A. N. (1991) "Statistical shape models in image analysis." In E. M. Keramidas, ed., *Computer Science and Statistics: Proc. 23rd Symp. Interface*, pp. 550-557. Interface Foundation, Fairfax Station.
- Mardia, K. V., Hainsworth, T. J. and Haddon, J. F. (1992) "Deformable templates in image sequences," *Proc. Int. Conf. Pattern Recognition, Vol. 2, The Hague*, IEEE Computer Society Press, Dos Alamitos, pp. 132-135.
- Mardia, K. V., Rabe, S. and Kent, J. T. (1995) "Statistics, shape and images." In D. M. Titterton, ed., *Complex Stochastic Systems and Engineering*, pp. 85-103. Clarendon Press, Oxford.
- Mardia, K. V., Qian, W., Shah, D. and de Souza, K. (1997) "Deformable template recognition of multiple occluded objects," *IEEE Trans. Pattern Anal. Machine Intell.* In preparation.
- Mardia, K. V., Hainsworth, T. J., Kirkbride, J., Hurn, M. A. and Berry, C. (1996a). Hierarchical Bayesian classification of multimodal medical images. In *Proc. Mathematical Methods in Biomedical Image Analysis*. IEEE Computer Society Press, pp. 53-63.
- Mardia, K. V., Gill, C. A. and Dryden, I. L. (1996b) *Image Fusion and Shape Variability Techniques*. Leeds University Press, Leeds.
- Masson, P. and Pieczynski, W. (1993) "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. on Geoscience and Remote Sensing*, **31**, 618-633.
- Propp, J. G. and Wilson, B. D. (1996) "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Symposium on Discrete Algorithms*. In preparation.
- Phillips, D. B. and Smith, A. F. M. (1993) "Dynamic image analysis using Bayesian shape and texture models." In K. V. Mardia and G. K. Gopal, eds, *Statistics and Images*, Vol. I, pp. 299-322. Carfax, Oxford.
- Qian, W. and Mardia, K. V. (1995) *Recognition of multiple objects with occlusions*. Technical Report 95/01. Department of Statistics, University of Leeds, Leeds.
- Ripley, B. D. and Sutherland, A. I. (1990). "Finding spiral structures in galaxies," *Phil. Trans. Roy. Soc. London A*, **332**, 477-485.
- Smith, A. F. M. and Roberts, G. O. (1993) "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion)," *J. Roy. Statist. Soc. B*, **55**, 3-24.
- Yuille, A. L. (1991) "Deformable templates for face recognition," *J. Cognitive Neuroscience*, **3**, 59-70.
- Yuille, A. L., Hallinan, P. and Cohen, D. (1992) "Feature extraction from faces using deformable templates" *Int. J. Comp. Vision*, **8**, 99-111.