CHAPTER

6

Segmentation

J. Ashburner and K. Friston

INTRODUCTION

This chapter describes a method of segmenting magnetic resonance (MR) images into different tissue classes, using a modified Gaussian mixture model. By knowing the prior spatial probability of each voxel that is grey matter, white matter or cerebrospinal fluid, it is possible to obtain quite robust classifications. A probabilistic framework is presented that enables image registration, tissue classification, and bias correction to be combined within the same generative model. A derivation of a log-likelihood objective function for this unified model is provided. The model is based on a mixture of Gaussians, and is extended to incorporate a smooth intensity variation and non-linear registration with tissue probability maps. A strategy for optimizing the model parameters is described, along with the requisite partial derivatives of the objective function.

Segmentation of brain images usually takes one of two forms. It can proceed by adopting a tissue classification approach, or by registration with a template. The aim of this chapter is to unify these procedures into a single probabilistic framework.

- The first approach rests on tissue classification, whereby voxels are assigned to a tissue class according to their intensities. In order to make these assignments, the intensity distribution of each tissue class needs to be characterized, often from voxels chosen to represent each class. Automatic selection of representative voxels can be achieved by first registering the brain volume to some standard space, and automatically selecting voxels that have a high probability of belonging to each class. A related approach involves modelling the intensity distributions by a mixture of Gaussians, but using tissue probability maps to weight the classification according to Bayes' rule.
- The other approach involves some kind of registration, where a template brain is warped to match the brain volume to be segmented (Collins *et al*., 1995). This need not involve matching volumes: some methods that are based on matching surfaces (MacDonald *et al*., 2000; Pitiot *et al*., 2004) would also fall into this category. These approaches allow regions that are predefined on the template to be overlaid, allowing different structures to be identified automatically.

A paradigm shift is evident in the field of neuroimaging methodology, away from simple sequential processing, towards a more integrated generative modelling approach. The model described in this chapter (which is implemented in SPM5) is one such example (see also Fischl *et al*., 2004). Both approaches combine tissue classification, bias correction and non-linear warping within the same framework. Although the integrated frameworks have some disadvantages, these should be outweighed by more accurate results. The main disadvantage is that the approaches are more complex and therefore more difficult to implement. In addition, the algorithms are integrated, making it difficult to mix and match different programs within 'pipeline' procedures (Zijdenbos *et al*., 2002; Fissell *et al*., 2003; Rex *et al*., 2003). A perceived disadvantage of these combined models is that execution time is longer than it would be for sequentially applied procedures. For example, optimizing two separate models with 100 parameters is likely to be faster than optimizing a combined single model with 200 parameters. However, the reason a combined model takes longer to run is because it actually completes the optimization. There are usually conditional correlations among parameters of the different models, which sequential processing discounts. The advantage of unified models is that they are more accurate, making better use of the information available in the data. Scanning time is relatively expensive, but computing time is relatively cheap. Complete

Statistical Parametric Mapping, by Karl Friston et al.
ISBN–13: 978-0-12-372560-8 ISBN–10: 0-12-372560-7

models may take longer to run, but they should add value to the raw data.

THE OBJECTIVE FUNCTION

In this section, we describe the segmentation model and how it is used to define an objective function. In the next section, we will show how this function is used to estimate the parameters of interest. The objective function, minimized by the optimum parameters, is derived from a mixture of Gaussians model. We show how this objective function can be extended to model smooth intensity non-uniformity. Tissue probability maps are used to assist the classification, and we describe how the objective function accommodates deformations of these maps, so that they best match the image to segment. The section ends by explaining how the estimated non-uniformity and deformations are constrained to be spatially smooth. The end-point of these model elaborations is a generative model whose inversion segments, spatially normalizes and intensity corrects a given image.

Mixture of Gaussians

A distribution can be modelled by a mixture of K Gaussians. This is a standard technique (see e.g. Bishop, 1995), which is used widely by many tissue classification algorithms. For univariate data, the kth Gaussian is modelled by its mean (μ_k), variance (σ_k^2) and mixing proportion (γ_k, where $\sum_{k=1}^{K}\gamma_k = 1$ and $\gamma_k \geq 0$). Fitting a mixture of Gaussians (MOG) model involves maximizing the probability of observing the I elements of data $\mathbf{y}$, given the parameterization of the Gaussians. In a simple MOG, the probability[1] of obtaining a datum with intensity y_i given that it belongs to the kth Gaussian ($c_i = k$), and that the kth Gaussian is parameterized by μ_k and σ_k^2 is:

$$P(y_i|c_i = k, \mu_k, \sigma_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}}\exp\left(-\frac{(y_i-\mu_k)^2}{2\sigma_k^2}\right) \quad \textbf{6.1}$$

The prior probability of any voxel, irrespective of its intensity, belonging to the kth Gaussian, given the proportion of voxels that belong to that Gaussian is simply:

$$P(c_i = k|\gamma_k) = \gamma_k \quad \textbf{6.2}$$

Using Bayes' rule, the joint probability of cluster k and intensity y_i is:

$$\begin{aligned} &P(y_i, c_i = k|\mu_k, \sigma_k, \gamma_k) \\ &\quad = P(y_i|c_i = k, \mu_k, \sigma_k)P(c_i = k|\gamma_k) \end{aligned} \quad \textbf{6.3}$$

By integrating over all k Gaussians, we obtain the probability of y_i given the parameters:

$$P(y_i|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \sum_{k=1}^{K} P(y_i, c_i = k|\mu_k, \sigma_k, \gamma_k) \quad \textbf{6.4}$$

The probability of the entire dataset $\mathbf{y}$ is derived by assuming that all elements are independent:

$$\begin{aligned} P(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) &= \textstyle\prod_{i=1}^{I} P(y_i|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) \\ &= \textstyle\prod_{i=1}^{I}\left(\sum_{k=1}^{K}\frac{\gamma_k}{(2\pi\sigma_k^2)^{\frac{1}{2}}}\exp\left(-\frac{(y_i-\mu_k)^2}{2\sigma_k^2}\right)\right) \end{aligned} \quad \textbf{6.5}$$

This probability is maximized with respect to the unknown parameters ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$), when the following cost function (ε) is minimized (because the two are monotonically related):

$$\begin{aligned} \varepsilon &= -\log P(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) \\ &= -\textstyle\sum_{i=1}^{I}\log\left(\sum_{k=1}^{K}\frac{\gamma_k}{(2\pi\sigma_k^2)^{\frac{1}{2}}}\exp\left(-\frac{(y_i-\mu_k)^2}{2\sigma_k^2}\right)\right) \end{aligned} \quad \textbf{6.6}$$

The assumption that voxels are independent is clearly implausible, but the priors embody a certain degree of spatial dependency. This means that the conditional probability that a voxel belongs to a tissue class shows spatial dependencies, even though the likelihood in Eqn. **6.5** does not.

Intensity non-uniformity

MR images are usually corrupted by a smooth, spatially varying artefact that modulates the intensity of the image (bias). There are a number of sources of this artefact, which are reviewed by Sled *et al*. (1998). These artefacts, although not usually a problem for visual inspection, can impede automated processing of the images. Early bias correction techniques involved homomorphic filtering, but these have generally been superseded. A review of

[1] Strictly speaking, it is a probability density rather than a probability. The mathematical notation used is $P(\cdot)$ for both probabilities and probability densities. Some authors make a distinction by using $P(\cdot)$ for probabilities and $p(\cdot)$ for probability densities.

bias correction approaches is presented in Belaroussi *et al*. (2006). Most current methods can be broadly classed as those that use parametric representations of image intensity distributions (such as mixtures of Gaussians), and those that use non-parametric representations (such as histograms).

- Non-parametric models usually involve image intensity histograms. Some authors have proposed using a multiplicative model of bias, and optimizing a function that minimizes the entropy of the histogram of the bias corrected intensities. One problem with this is that the entropy is minimized when the bias field is uniformly zero, resulting in a single bin containing all the counts. This was a problem (pointed out by Arnold *et al*. (2001)) for the bias field correction in SPM99 (Ashburner and Friston, 2000), where there was a tendency for the correction to reduce the mean intensity of brain tissue in the corrected image. The constraint that the multiplicative bias should average to unity resulted in a bowl-shaped dip in the estimated bias field.

  To counter this problem, Mangin (2000) minimized the entropy of the histograms, but included an additional term in the cost function to minimize the squared difference between the original and restored image mean. A related solution was devised by Likar *et al*. (2001). In addition to modelling a multiplicative bias field, the latter method also modelled a smooth additive bias. These represent partial solutions to the problem, but are not ideal. When the width of a Gaussian (or any other distribution) is multiplied by a factor of ρ, then the entropy of the distribution is increased by $\log \rho$. Therefore, when scaling data by some value, the log of this factor needs to be considered when developing an entropy-based cost function.

  An alternative solution is to minimize the entropy of the histogram of log-transformed intensities. In addition to being generally better behaved, this also allows the bias fields to be modelled as an additive effect in log-space (Sled *et al*., 1998). In order to work with log-transformed data, low intensity (and negative valued) voxels are excluded so that numerical problems are not introduced. This exclusion motivates a more generic model of all regional effects.
- Parametric bias correction models are often an integral part of tissue classification methods, many of which are based upon modelling the intensities of different tissues as a mixture of Gaussians. Other clustering methods can also be used, such as k-means and fuzzy c-means. Additional information is often encoded, in these approaches, using Markov random field models to embed knowledge that neighbouring voxels are likely to belong to the same tissue class. Most algorithms assume that the bias is multiplicative, but there are three commonly used models of how the bias interacts with noise.

  In the first parametric model, the observed signal (y_i) is assumed to be an uncorrupted signal (μ_i), scaled by some bias (ρ_i) with added Gaussian noise (n_i) that is independent of the bias (Pham and Prince, 1999; Shattuck *et al*., 2001). The noise source is assumed to be from the scanner itself:

$$y_i = \mu_i/\rho_i + n_i \qquad \textbf{6.7}$$

  The second model is similar to the first, except that the noise is added before the signal is scaled. In this case, the noise is assumed to be due to variations in tissue properties. This model is the one used in this chapter:

$$y_i = (\mu_i + n_i)/\rho_i \qquad \textbf{6.8}$$

  A combination of the scanner and tissue noise models has been adopted by Fischl *et al*. (2004). This would probably be a better model, especially for images corrupted by a large amount of bias. The single noise source model was mainly chosen for its simplicity.

  A third approach involves log transforming the data first, allowing a multiplicative bias to be modelled as an additive effect in log-space (Wells *et al*., 1996b; Garza-Jinich *et al*., 1999; Van Leemput *et al*., 1999a; Styner, 2000; Zhang *et al*., 2001). The cost function for these approaches is related to the entropy of the distribution of log-transformed bias corrected data. As with the non-parametric model based on log-transformed data, low intensity voxels have to be excluded to avoid numerical problems. The generative model is of a form similar to:

$$\log y_i = \log \mu_i - \log \rho_i + n_i$$
$$y_i = \mu_i e^{n_i}/\rho_i \qquad \textbf{6.9}$$

  Sometimes these methods do not use a consistent generative model throughout, for example when alternating between the original intensities for the classification steps, and the log-transformed intensities for the bias correction (Wells *et al*., 1996a).

In the model described here, bias correction is included in the MOG by extra parameters that account for smooth intensity variations. The field modelling the variation at element i is denoted by $\rho_i(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of unknown parameters. Intensities from the kth cluster are assumed to be normally distributed with mean $\mu_k/\rho_i(\boldsymbol{\beta})$, and variance $(\sigma_k/\rho_i(\boldsymbol{\beta}))^2$. Therefore, the probability of

obtaining intensity y_i from the kth cluster, given its parameterization is:

$$P(y_i|c_i = k, \mu_k, \sigma_k, \boldsymbol{\beta}) = \rho_i(\boldsymbol{\beta}) \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \exp\left(-\frac{(\rho_i(\boldsymbol{\beta})y_i - \mu_k)^2}{2\sigma_k^2}\right) \quad \textbf{6.10}$$

The tissue classification objective function is now:

$$\varepsilon = -\sum_{i=1}^{I} \log\left(\rho_i(\boldsymbol{\beta}) \sum_{k=1}^{K} \frac{\gamma_k}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \exp\left(-\frac{(\rho_i(\boldsymbol{\beta})y_i - \mu_k)^2}{2\sigma_k^2}\right)\right) \quad \textbf{6.11}$$

The model employed here parameterizes the bias as the exponential of a linear combination of low-frequency basis functions. A small number of basis functions are used, as bias tends to be spatially smooth. Positivity is ensured by the exponential.

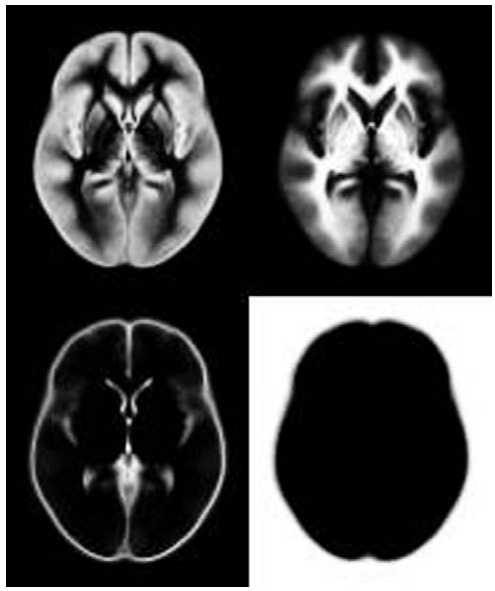

FIGURE 6.1 The tissue probability maps for grey matter, white matter, CSF and 'other'.

Spatial priors

Rather than assuming stationary prior probabilities based upon mixing proportions, additional information is used, derived from tissue probability maps from other subjects' brain images. Priors are usually generated by registering a large number of subjects together, assigning voxels to different tissue types and averaging tissue classes over subjects. The data used by SPM5 are a modified version of the ICBM Tissue Probabilistic Atlas.[2] They consist of tissue probability maps of grey and white matter, and of CSF (Figure 6.1). A fourth class is also used, which is simply one minus the sum of the first three. These maps give the prior probability of any voxel in a registered image being of any of the tissue classes – irrespective of its intensity. The implementation uses tissue probability maps for grey matter, white matter and CSF, although maps for additional tissue types (e.g. blood vessels) could also be included. The simple model of grey matter being all of approximately the same intensity could also be refined by using tissue probability maps for various internal grey matter structures (Fischl *et al.*, 2002).

The model in Eqn. **6.11** is modified to account for these spatial priors. Instead of using stationary mixing proportions ($P(c_i = k|\boldsymbol{\gamma}) = \gamma_k$), the prior probabilities are allowed to vary over voxels, such that the prior probability of voxel i being drawn from the kth Gaussian is:

$$P(c_i = k|\boldsymbol{\gamma}) = \frac{\gamma_k b_{ik}}{\sum_{j=1}^{K} \gamma_j b_{ij}} \quad \textbf{6.12}$$

where b_{ik} is the tissue probability for class k at voxel i. Note that $\boldsymbol{\gamma}$ is no longer a vector of true mixing proportions, but for the sake of simplicity, its elements will be referred to as such.

The number of Gaussians used to represent the intensity distribution for each tissue class can be greater than one. In other words, a tissue probability map may be shared by several Gaussians. The assumption of a single Gaussian distribution for each class does not hold for a number of reasons. In particular, a voxel may not be purely of one tissue type, and instead contain signal from a number of different tissues (partial volume effects). Some partial volume voxels could fall at the interface between different classes, or they may fall in the middle of structures, such as the thalamus, which may be considered as being either grey or white matter. Various image segmentation approaches use additional Gaussians to model such partial volume effects. These generally assume that a pure tissue class has a Gaussian intensity distribution, whereas intensity distributions for partial volume voxels are broader, falling between the intensities of the pure classes. Most of these models assume that a mixing combination of, e.g. 50/50, is just as probable as one of 80/20 (Laidlaw *et al.*, 1998; Shattuck *et al.*, 2001; Tohka *et al.*, 2004), whereas others allow a spatially varying prior probability for the mixing combination, which is dependent upon the contents of neighbouring voxels (Van Leemput *et al.*, 2001). Unlike these partial volume segmentation approaches, the model adopted here simply assumes that the intensity distribution of each class may not be Gaussian, and assigns

[2] Avaliable from http://www.loni.ucla.edu/ICBM/ICBM_Probabilistic.html

belonging probabilities according to these non-Gaussian distributions. Selecting the optimal number of Gaussians per class is a model order selection issue, and will not be addressed here. Typical numbers of Gaussians are three for grey matter, two for white matter, two for CSF, and five for everything else.

Deformable spatial priors

The above formulation (Eqn. **6.12**) is refined further by allowing the tissue probability maps to be deformed according to parameters $\boldsymbol{\alpha}$. This allows registration to a standard space to be included within the generative model.

$$P(c_i = k|\boldsymbol{\gamma}, \boldsymbol{\alpha}) = \frac{\gamma_k b_{ik}(\boldsymbol{\alpha})}{\sum_{j=1}^{K} \gamma_j b_{ij}(\boldsymbol{\alpha})} \quad \textbf{6.13}$$

After including the full priors, the objective function becomes:

$$\mathcal{E} = -\sum_{i=1}^{I} \log \left(\frac{\rho_i(\boldsymbol{\beta})}{\sum_{k=1}^{K} \gamma_k b_{ik}(\boldsymbol{\alpha})} \sum_{k=1}^{K} \gamma_k b_{ik}(\boldsymbol{\alpha}) (2\pi\sigma_k^2)^{-\frac{1}{2}} \times \exp\left(-\frac{(\rho_i(\boldsymbol{\beta}) y_i - \mu_k)^2}{2\sigma_k^2} \right) \right) \quad \textbf{6.14}$$

There are many ways of parameterizing how the tissue probability maps could be deformed. The implementation in SPM5 uses a low-dimensional approach, which parameterizes the deformations by a linear combination of about a thousand cosine transform bases (Ashburner and Friston, 1999). This is not an especially precise way of encoding deformations, but it can model the variability of overall brain shape. Evaluations have shown that this simple model can achieve a registration accuracy comparable to other fully automated methods with many more parameters (Hellier *et al*., 2001, 2002). This deformation means that inversion of the generative model can, implicitly, normalize images. Indeed, this is the prefered method of spatial normalization in SPM5.

Regularization

One important issue relates to the distinction between intensity variations that arise because of bias artefact due to the physics of MR scanning, and those that arise due to different tissue properties. The objective is to model the latter by different tissue classes, while modelling the former with a bias field. We know *a priori* that intensity variations due to MR physics tend to be spatially smooth, whereas those due to different tissue types tend to contain more high frequency information. A more accurate estimate of a bias field can be obtained by including prior knowledge about the distribution of the fields likely to be encountered by the correction algorithm. For example, if it is known that there is little or no intensity non-uniformity, then it would be wise to penalize large values for the intensity non-uniformity parameters. This regularization can be placed within a Bayesian context, whereby the penalty incurred is the negative logarithm of a prior probability for any particular pattern of non-uniformity. Similarly, it is possible for intensity variations to be modelled by incorrect registration. If we had some knowledge about a prior probability distribution for brain shape, then this information could be used to regularize the deformations. It is not possible to determine a complete specification of such a probability distribution empirically. Instead, the SPM5 approach (as with most other non-linear registration procedures) uses an educated guess for the form and amount of variability likely to be encountered. Without such regularization, the pernicious interactions (Evans, 1995) among the parameter estimates could be more of a problem. With the regularization terms included, fitting the model involves maximizing:

$$P(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = P(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma}) P(\boldsymbol{\beta}) P(\boldsymbol{\alpha}) \quad \textbf{6.15}$$

This is equivalent to minimizing:

$$\mathcal{F} = -\log P(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \mathcal{E} - \log P(\boldsymbol{\beta}) - \log P(\boldsymbol{\alpha}) \quad \textbf{6.16}$$

In the SPM5 implementation, the probability densities of the spatial parameters are assumed to be zero-mean multivariate Gaussians ($P(\boldsymbol{\alpha}) = N(\mathbf{0}, \mathbf{C}_{\boldsymbol{\alpha}})$ and $P(\boldsymbol{\beta}) = N(\mathbf{0}, \mathbf{C}_{\boldsymbol{\beta}})$). For the non-linear registration parameters, the covariance matrix is defined such that $\boldsymbol{\alpha}^T \mathbf{C}_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\alpha}$ gives the bending energy of the deformations (see Ashburner and Friston, 1999 for details). The prior covariance matrix for the bias is based on the assumption that a typical bias field could be generated by smoothing zero mean random Gaussian noise by a broad Gaussian smoothing kernel (about 70 mm FWHM, full width at half maximum), and then exponentiating (i.e. $\mathbf{C}_{\boldsymbol{\beta}}$ is a Gaussian Toeplitz matrix).

OPTIMIZATION

This section describes how the objective function from Eqns **6.14** and **6.16** is minimized (i.e. how the model is inverted). There is no closed-form solution for the parameters, and optimal values for different parameters depend upon the values of others. An iterated conditional modes (ICM) approach is used. It begins by assigning

starting estimates for the parameters, and then iterating until a locally optimal solution is found. Each iteration involves alternating between estimating different groups of parameters, while holding the others fixed at their current 'best' solution (i.e. conditional mode). The mixture parameters are updated using expectation maximization (EM), while holding the bias and deformations fixed at their conditional modes. The bias is estimated while holding the mixture parameters and deformation constant. Because intensity non-uniformity is very smooth, it can be described by a small number of parameters, making the Levenberg-Marquardt (LM) scheme ideal for this optimization. The deformations of the tissue probability maps are re-estimated while fixing the mixture parameters and bias field. A low-dimensional parameterization is used for the spatial deformations, so the LM strategy is also applicable here.

The model is only specified for brain, as there are no tissue probability maps for non-brain tissue (scalp etc). Because of this, there is a tendency for the approach to stretch the probability maps so that the background class contains only air, but no scalp. A workaround involves excluding extra-cranial voxels from the fitting procedure. This is done by fitting a mixture of two Gaussians to the image intensity histogram. In most cases, one Gaussian fits air, and the other fits everything else. A suitable threshold is then determined, based on a 50 per cent probability. Fitting only the intra-cranial voxels also saves time.

Mixture parameters ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\gamma}$)

It is sufficient to minimize ε with respect to the mixture parameters because they do not affect the prior or regularization terms in $\mathcal{F}$ (see Eqn. **6.16**). For simplicity, we summarize the parameters of interest by $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}\}$. These are optimized by EM (see e.g. Dempster *et al*., 1977; Bishop, 1995 or Neal and Hinton, 1998), which can be considered as using some distribution, q_{ik}, to minimize the following upper bound on $\mathcal{E}$:

$$\varepsilon \leq \varepsilon_{EM} = -\sum_{i=1}^{I} \log P(y_i|\boldsymbol{\theta}) + \sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik} \log\left(\frac{q_{ik}}{P(c_i = k|y_i, \boldsymbol{\theta})}\right) \quad \textbf{6.17}$$

EM is an iterative approach, and involves alternating between an E-step (which minimizes $\mathcal{E}_{EM}$ with respect to q_{ik}), and an **M**-step (which minimizes $\mathcal{E}_{EM}$ with respect to $\boldsymbol{\theta}$). The second term of Eqn. **6.17** is a Kullback-Leibler distance, which is at a minimum of zero when $q_{ik} = P(c_i = k|y_i, \boldsymbol{\theta})$, and Eqn. **6.17** becomes an equality ($\mathcal{E} = \mathcal{E}_{EM}$). Because q_{ik} does not enter into the first term, the E-step of iteration n consists of setting:

$$q_{ik}^{(n)} = P(c_i = k|y_i, \boldsymbol{\theta}^{(n)}) = \frac{P(y_i, c_i = k|\boldsymbol{\theta}^{(n)})}{P(y_i|\boldsymbol{\theta}^{(n)})} = \frac{p_{ik}}{\sum_{j=1}^{K} p_{ij}} \quad \textbf{6.18}$$

where

$$p_{ik} = \frac{\gamma_k b_{ik}(\boldsymbol{\alpha})}{\sum_{j=1}^{K} \gamma_j b_{ij}(\boldsymbol{\alpha})} (2\pi\sigma_k^2)^{-\frac{1}{2}} \times \exp\left(-\frac{(\rho_i(\boldsymbol{\beta})y_i - \mu_k)^2}{2\sigma_k^2}\right) \quad \textbf{6.19}$$

The M-step uses the recently updated values of $q_{ik}^{(n)}$ in order to minimize ε with respect to $\boldsymbol{\theta}$. Eqn. **6.17** can be reformulated[3] as:

$$\varepsilon = \varepsilon_{EM} = -\sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik} \log P(y_i, c_i = k|\boldsymbol{\theta}) + \sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik} \log q_{ik} \quad \textbf{6.20}$$

Because the second term is independent of $\boldsymbol{\theta}$, the M-step involves assigning new values to the parameters, such that the derivatives of the following are zero:

$$\begin{aligned} -\sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik} \log P(y_i, c_i = k|\boldsymbol{\theta}) = \\ \sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik}\left(\log\left(\sum_{j=1}^{K}\gamma_j b_{ij}(\boldsymbol{\alpha})\right) - \log\gamma_k\right) \\ + \sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik}\left(\frac{1}{2}\log(\sigma_k^2) + \frac{1}{2\sigma^2}(\rho_i(\boldsymbol{\beta})y_i - \mu_k)^2\right) \\ + \sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik}\left(\frac{1}{2}\log(2\pi) - \log(\rho_i(\boldsymbol{\beta})b_{ik}(\boldsymbol{\alpha}))\right) \end{aligned} \quad \textbf{6.21}$$

Differentiating Eqn. **6.21** with respect to μ_k gives:

$$\frac{\partial\mathcal{F}}{\partial\mu_k} = \frac{\partial\varepsilon}{\partial\mu_k} = \sum_{i=1}^{I}\frac{q_{ik}^{(n)}}{\sigma_k^2}(\mu_k - \rho_i(\boldsymbol{\beta})y_i) \quad \textbf{6.22}$$

This gives the update formula for μ_k by solving for $\frac{\partial\mathcal{E}}{\partial\mu_k} = 0$

$$\mu_k^{(n+1)} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)}\rho_i(\boldsymbol{\beta})y_i}{\sum_{i=1}^{I} q_{ik}^{(n)}} \quad \textbf{6.23}$$

[3] Through Bayes' rule, and because $\sum_{k=1}^{K} q_{ik} = 1$, we obtain $\log P(y_i|\boldsymbol{\theta}) = \log\left(\frac{P(y_i, c_i = k|\boldsymbol{\theta})}{P(c_i = k|y_i, \boldsymbol{\theta})}\right) = \sum_{k=1}^{K} q_{ik}\log\left(\frac{P(y_i, c_i = k|\boldsymbol{\theta})}{P(c_i = k|y_i, \boldsymbol{\theta})}\right)$.

Similarly, differentiating Eqn. **6.21** with respect to σ_k^2:

$$\frac{\partial \mathcal{F}}{\partial \sigma_k^2} = \frac{\partial \mathcal{E}}{\partial \sigma_k^2} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)}}{2\sigma_k^2} - \frac{\sum_{i=1}^{I} q_{ik}^{(n)} (\mu_k - \rho_i(\boldsymbol{\beta}) y_i)^2}{2(\sigma_k^2)^2} \quad \textbf{6.24}$$

This gives the update formula for σ_k^2:

$$(\sigma_k^2)^{(n+1)} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)} (\mu_k^{(n+1)} - \rho_i(\boldsymbol{\beta}) y_i)^2}{\sum_{i=1}^{I} q_{ik}^{(n)}} \quad \textbf{6.25}$$

Differentiating Eqn. **6.21** with respect to γ_k:

$$\frac{\partial \mathcal{F}}{\partial \gamma_k} = \frac{\partial \varepsilon}{\partial \gamma_k} = \sum_{i=1}^{I} \frac{b_{ik}(\boldsymbol{\alpha})}{\sum_{j=1}^{K} \gamma_j b_{ij}(\boldsymbol{\alpha})} - \frac{\sum_{i=1}^{I} q_{ik}^{(n)}}{\gamma_k} \quad \textbf{6.26}$$

Deriving an exact update scheme for γ_k is difficult, but the following ensures convergence:[4]

$$\gamma_k^{(n+1)} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)}}{\sum_{i=1}^{I} \frac{b_{ik}(\boldsymbol{\alpha})}{\sum_{j=1}^{K} \gamma_j^{(n)} b_{ij}(\boldsymbol{\alpha})}} \quad \textbf{6.27}$$

Bias ($\boldsymbol{\beta}$)

The next step is to update the estimate of the bias field. This involves holding the other parameters fixed, and improving the estimate of $\boldsymbol{\beta}$ using an LM optimization approach (see Press *et al.*, 1992 for more information). Each iteration requires the first and second derivatives of the objective function, with respect to the parameters. In the following scheme, $\mathbf{I}$ is an identity matrix and λ is a scaling factor. The choice of λ is a trade-off between speed of convergence, and stability. A value of zero for λ gives the Newton-Raphson or Gauss-Newton optimization scheme, which may be unstable. Increasing λ will slow down the convergence, but increase the stability of the algorithm. The value of λ is usually decreased slightly after iterations that decrease (improve) the cost function. If the cost function increases after an iteration, then the previous solution is retained, and λ is increased in order to provide more stability.

$$\boldsymbol{\beta}^{(n+1)} = \boldsymbol{\beta}^{(n)} - \left(\frac{\partial^2 \mathcal{F}}{\partial \boldsymbol{\beta}^2}\Big|_{\beta^{(n)}} + \lambda \mathbf{I} \right)^{-1} \frac{\partial F}{\partial \boldsymbol{\beta}}\Big|_{\beta^{(n)}} \quad \textbf{6.28}$$

The prior probability of the parameters is modelled by a multivariate Gaussian density, with mean $\boldsymbol{\beta}_0$ and covariance $\mathbf{C}_{\boldsymbol{\beta}}$.

$$-\log P(\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \mathbf{C}_{\boldsymbol{\beta}}{}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \text{const} \quad \textbf{6.29}$$

[4] The update scheme was checked empirically, and found to always reduce $\mathcal{E}$. It does not fully minimize it though, which means that this part of the algorithm is really a generalized EM.

The first and second derivatives of $\mathcal{F}$ (see Eqn. **6.16**) with respect to the parameters are therefore:

$$\frac{\partial \mathcal{F}}{\partial \boldsymbol{\beta}} = \frac{\partial \varepsilon}{\partial \boldsymbol{\beta}} + \mathbf{C}_{\boldsymbol{\beta}}{}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \text{ and } \frac{\partial^2 \mathcal{F}}{\partial \boldsymbol{\beta}^2} = \frac{\partial^2 \varepsilon}{\partial \boldsymbol{\beta}^2} + \mathbf{C}_{\boldsymbol{\beta}}{}^{-1} \quad \textbf{6.30}$$

The first and second partial derivatives of ε are:

$$\frac{\partial \varepsilon}{\partial \beta_m} = -\sum_{i=1}^{I} \frac{\partial \rho_i(\boldsymbol{\beta})}{\partial \beta_m} \times \left(\rho_i(\boldsymbol{\beta})^{-1} + y_i \sum_{k=1}^{K} \frac{q_{ik}(\mu_k - \rho_i(\boldsymbol{\beta}) y_i)}{\sigma_k^2} \right) \quad \textbf{6.31}$$

$$\frac{\partial^2 \varepsilon}{\partial \beta_m \beta_n} = \sum_{i=1}^{I} \frac{\partial \rho_i(\boldsymbol{\beta})}{\partial \beta_m} \frac{\partial \rho_i(\boldsymbol{\beta})}{\partial \beta_n} \left(\rho_i(\boldsymbol{\beta})^{-2} + y_i^2 \sum_{k=1}^{K} \frac{q_{ik}}{\sigma_k^2} \right) - \sum_{i=1}^{I} \frac{\partial^2 \rho_i(\boldsymbol{\beta})}{\partial \beta_m \partial \beta_n} \times \left(\rho_i(\boldsymbol{\beta})^{-1} + y_i \sum_{k=1}^{K} \frac{q_{ik}\,(\mu_k - \rho_i(\boldsymbol{\beta}) y_i)}{\sigma_k^2} \right) \quad \textbf{6.32}$$

The bias field is parameterized by the exponential of a linear combination of smooth basis functions:

$$\rho_i(\boldsymbol{\beta}) = \exp\left(\sum_{m=1}^{M} \beta_m \psi_{im} \right), \frac{\partial \rho_i(\boldsymbol{\beta})}{\partial \beta_m} = \psi_{im} \rho_i(\boldsymbol{\beta}), \text{ and } \frac{\partial^2 \rho_i(\boldsymbol{\beta})}{\partial \beta_m \partial \beta_n} = \psi_{im} \psi_{in} \rho_i(\boldsymbol{\beta}) \quad \textbf{6.33}$$

Therefore, the derivatives used by the optimization are:

$$\frac{\partial \varepsilon}{\partial \beta_m} = -\sum_{i=1}^{I} \psi_{im} \left(1 + \rho_i(\boldsymbol{\beta}) y_i \sum_{k=1}^{K} \frac{q_{ik}(\mu_k - \rho_i(\boldsymbol{\beta}) y_i)}{\sigma_k^2} \right)$$

$$\frac{\partial^2 \varepsilon}{\partial \beta_m \beta_n} = \sum_{i=1}^{I} \psi_{im} \psi_{in} \left((\rho_i(\boldsymbol{\beta}) y_i)^2 \sum_{k=1}^{K} \frac{q_{ik}}{\sigma_k^2} - \rho_i(\boldsymbol{\beta}) y_i \sum_{k=1}^{K} \frac{q_{ik}\,(\mu_k - \rho_i(\boldsymbol{\beta}) y_i)}{\sigma_k^2} \right) \quad \textbf{6.34}$$

Deformations ($\boldsymbol{\alpha}$)

The same LM strategy (Eqn. **6.28**) is used as for updating the bias. Schemes such as LM or Gauss-Newton are usually used only for registering images with a mean squared difference objective function, although some rare exceptions exist where LM has been applied to information-theoretic image registration (Thévenaz and Unser, 2000). The strategy requires the first and second derivatives of the cost function, with respect to the parameters that define the deformation. In order to simplify deriving the

derivatives, the likelihood component of the objective function is re-expressed as:

$$\mathcal{E} = -\sum_{i=1}^{I} \log\left(\sum_{k=1}^{K} f_{ik} l_{ik}\right) - \sum_{i=1}^{I} \log \rho_i(\boldsymbol{\beta}) \qquad \textbf{6.35}$$

where

$$f_{ik} = \frac{b_{ik}(\boldsymbol{\alpha})}{\sum_{j=1}^{K} \gamma_j b_{ij}(\boldsymbol{\alpha})} \qquad \textbf{6.36}$$

and

$$l_{ik} = \gamma_k (2\pi\sigma_k^2)^{-\frac{1}{2}} \exp\left(-\frac{(\rho_i(\boldsymbol{\beta}) y_i - \mu_k)^2}{2\sigma_k^2}\right) \qquad \textbf{6.37}$$

The first derivatives of ε with respect to α are:

$$\frac{\partial \mathcal{E}}{\partial \alpha_m} = -\sum_{i=1}^{I} \frac{\sum_{k=1}^{K} \frac{\partial f_{ik}}{\partial \alpha_m} l_{ik}}{\sum_{k=1}^{K} f_{ik} l_{ik}} \qquad \textbf{6.38}$$

The second derivatives are:

$$\frac{\partial^2 \mathcal{E}}{\partial \alpha_m \partial \alpha_n} = \sum_{i=1}^{I} \frac{\left(\sum_{k=1}^{K} \frac{\partial f_{ik}}{\partial \alpha_m} l_{ik}\right)\left(\sum_{k=1}^{K} \frac{\partial f_{ik}}{\partial \alpha_n} l_{ik}\right)}{\left(\sum_{k=1}^{K} f_{ik} l_{ik}\right)^2} - \sum_{i=1}^{I} \frac{\sum_{k=1}^{K} \frac{\partial^2 f_{ik}}{\partial \alpha_m \partial \alpha_n} l_{ik}}{\sum_{k=1}^{K} f_{ik} l_{ik}} \qquad \textbf{6.39}$$

The following is needed in order to compute derivatives of ε with respect to α:

$$\frac{\partial f_{ik}}{\partial \alpha_m} = \frac{\frac{\partial b_{ik}(\boldsymbol{\alpha})}{\partial \alpha_m}}{\sum_{j=1}^{K} \gamma_j b_{ij}(\boldsymbol{\alpha})} - \frac{b_{ik}(\boldsymbol{\alpha}) \sum_{j=1}^{K} \gamma_j \frac{\partial b_{ij}(\boldsymbol{\alpha})}{\partial \alpha_m}}{\left(\sum_{j=1}^{K} \gamma_j b_{ij}(\boldsymbol{\alpha})\right)^2} \qquad \textbf{6.40}$$

The second term in Eqn. **6.39** is ignored in the optimization (Gauss-Newton approach), but it could be used (Newton-Ràphson approach). These gradients and curvatures enter the update scheme as in Eqn. **6.28**.

The chain rule is used to compute derivatives of f_{ik}, based on the rate of change of the deformation fields with respect to changes of the parameters, and the tissue probability map gradients sampled at the appropriate points. Trilinear interpolation could be used as the tissue probability maps contain values between zero and one. Care is needed when attempting to sample the images with higher degree B-spline interpolation (Thévenaz *et al*., 2000), as negative values should not occur. B-spline interpolation (and other generalized interpolation methods) require coefficients to be estimated first. This essentially involves deconvolving the B-spline bases from the image (Unser *et al*., 1993a, b). Sampling an interpolated value in the image is then done by re-convolving the coefficients with the B-spline. Without any non-negativity constraints on the coefficients, there is a possibility of negative values occurring in the interpolated probability map.

One possible solution is to use a maximum-likelihood deconvolution strategy to estimate some suitable coefficients. This is analogous to the iterative method for maximum-likelihood reconstruction of PET (positron emission tomography) images (Shepp and Vardi, 1982), or to the way that mixing proportions are estimated within a mixture of Gaussians model. A second solution is to add a small background value to the probability maps, and take a logarithm. Standard interpolation methods could be applied to the log-transformed data, before exponentiating again. Neither of these approaches is really optimal. In practice, 3rd degree B-spline interpolation is used, but without first deconvolving. This introduces a small, but acceptable, amount of additional smoothness to the tissue probability maps.

Example

The segmentation accuracy is illustrated for data generated by the BrainWeb MR simulator (Kwan *et al*., 1996; Cocosco *et al*., 1997; Collins *et al*., 1998). The simulated images were all of the same subject, had dimensions of $181 \times 217 \times 181$ voxels of $1 \times 1 \times 1$ mm and had 3 per cent noise (relative to the brightest tissue in the images). The contrasts of the images simulated T1-weighted, T2-weighted and proton density (PD). The T1-weighted image was simulated as a spoiled FLASH sequence, with a 30° flip angle, 18 ms repeat time, 10 ms echo time. The T2 and PD images were simulated by a dual echo spin echo, early echo technique, with 90° flip angle, 3300 ms repeat time and echo times of 35 and 120 ms. Three different levels of image non-uniformity were used: 0 per cent RF (which assumes that there is no intensity variation artefact), 40 per cent RF, and 100 per cent RF (Figure 6.2). Three components were considered: grey matter, white matter and whole brain (grey and white matter). Because the causes of the simulated images were available, it was possible to compare the segmented images with images of 'true' grey and white matter using the Dice metric, which is used widely for evaluating segmentation algorithms (e.g. Van Leemput *et al*., 1999b; Shattuck *et al*., 2001). The probabilities were thresholded at 0.5 in order to compute the number of misclassifications. If TP refers to the number of true positives, FP to false positives and FN to false negatives, then the Dice metric is given by:

$$\text{Dice metric} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \qquad \textbf{6.41}$$

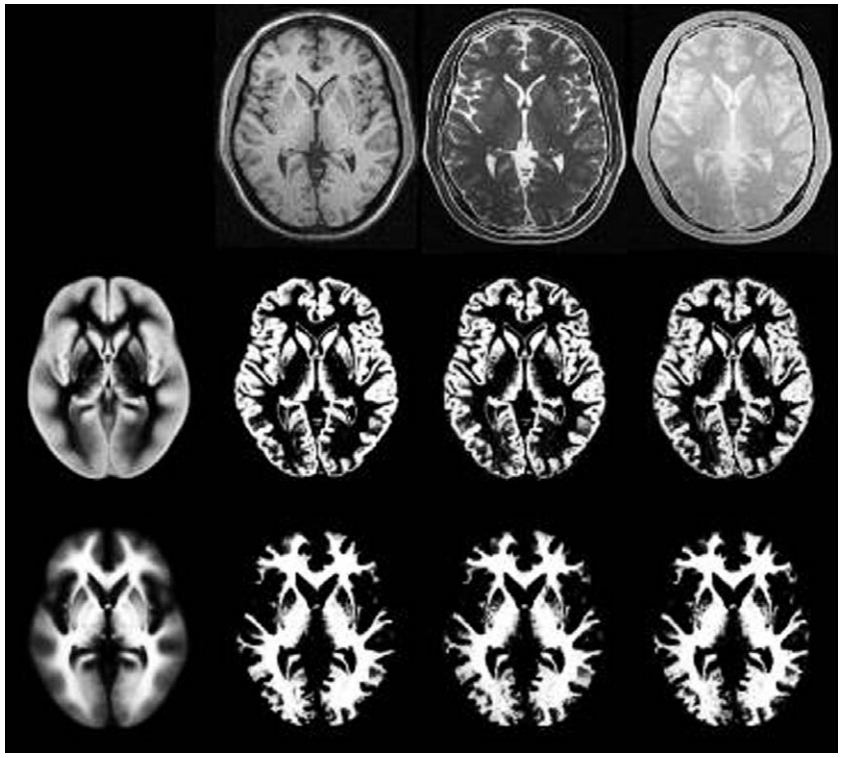

FIGURE 6.2 Results from applying the method to the BrainWeb data. The first column shows the tissue probability maps for grey and white matter. The first row of columns two, three and four show the 100 per cent RF BrainWeb T1, T2 and PD images after they are warped to match the tissue probability maps (by inverting the spatial transform). Below the warped BrainWeb images are the corresponding segmented grey and white matter.

Our results are shown in Table 6-1. Values range from zero to one, where higher values indicate better agreement.

Discussion

This chapter illustrates a framework whereby tissue classification, bias correction and image registration are integrated within the same generative model. The objective was to explain how this can be done, rather than focus on the details of a specific implementation. The same framework could be used for a more sophisticated implementation. When devising a model, it is useful to think about how that model could be used to generate data. The distribution of randomly generated data should match the distribution of any data the model has to explain. There are a number of aspects of our model that could be improved in order to achieve this goal.

The SPM5 implementation assumes that the brain consists of grey and white matter, and is surrounded by a thin layer of CSF. The addition of extra tissue probability maps should improve the model. In particular, grey matter classes for internal structures may allow them to be segmented more accurately.

It is only a single channel implementation, which can segment a single image, but is unable to make optimal use of information from two or more registered images of the same subject. Multispectral data may provide more accurate results by allowing the model to work with joint intensity probability distributions. For two registered images of the same subject, one form of objective function would use axis-aligned multivariate Gaussians (with σ_{k1}^2 and σ_{k2}^2 are diagonal elements of a 2×2 covariance matrix).

TABLE 6-1 Dice metrics computed from segmented BrainWeb images

	T1			T2			PD		
Dice metric	0%	40%	100%	0%	40%	100%	0%	40%	100%
Grey matter	0.932	0.934	0.918	0.883	0.881	0.880	0.872	0.880	0.872
White matter	0.961	0.961	0.939	0.916	0.916	0.915	0.923	0.928	0.923
Whole brain	0.977	0.978	0.978	0.967	0.966	0.965	0.957	0.959	0.955

$$\mathcal{E} = -\sum_{i=1}^{I} \log\left(\frac{\rho_{i1}(\boldsymbol{\beta})\rho_{i2}(\boldsymbol{\beta})}{\sum_{k=1}^{K}\gamma_k b_{ik}(\boldsymbol{\alpha})}\right) - \sum_{i=1}^{I} \log\left(\sum_{k=1}^{K}\gamma_k b_{ik}(\boldsymbol{\alpha}) \times \frac{\exp\left(-\frac{(\rho_{i1}(\boldsymbol{\beta})y_{i1}-\mu_{k1})^2}{2\sigma_{k1}^2}\right)}{(2\pi\sigma_{k1}^2)^{\frac{1}{2}}} \frac{\exp\left(-\frac{(\rho_{i2}(\boldsymbol{\beta})y_{i2}-\mu_{k2})^2}{2\sigma_{k2}^2}\right)}{(2\pi\sigma_{k2}^2)^{\frac{1}{2}}}\right) \quad \textbf{6.42}$$

Multispectral classification usually requires the images to be registered together. Another possible extension of the framework could be to include within-subject registration (Xiaohua *et al*., 2004).

The generative model contains nothing to encode the probability that neighbouring voxels are more likely to belong to the same class. The inclusion of such priors should make the generative model more realistic. One solution could be to include a Markov random field (MRF) (Besag, 1986) in the model. Another strategy for making the model more realistic may be to have crisper tissue probability maps, and more precise warping.

Objective functions, such as the mean squared difference or cross-correlation, can only be used to register MR images generated using the same sequences, field strengths etc. An advantage that they do have over information theoretic measures (such as mutual information), is that they are also appropriate for registering to smooth averaged images. One of the benefits of the approach is that the same averaged tissue probability maps can be used to normalize spatially (and segment) images acquired with a wide range of different contrasts (e.g. T1-weighted, T2-weighted etc). This flexibility could also be considered a weakness. If the method is only to be used with images of a particular contrast, then additional constraints relating to the approximate intensities of the different tissue types could be included (Fischl *et al*., 2002). Alternatively, the MR parameters could be estimated within the model (Fischl *et al*., 2004), and the cluster means constrained to be more realistic. Rather than using fixed intensity distributions for the classes, a better approach would invoke some kind of hierarchical modelling, whereby prior probability distributions for the cluster parameters are used to inform their estimation.

The hierarchical modelling scheme could be extended in order to generate tissue probability maps and other priors using data from many subjects. This would involve a very large model, whereby many images of different subjects are simultaneously processed within the same hierarchical framework. Strategies for creating average (in both shape and intensity) brain atlases are currently being devised (Ashburner *et al*., 2000; Avants and Gee, 2004; Joshi *et al*., 2004). Such approaches could be refined in order to produce average shaped tissue probability maps and other data for use as priors.

REFERENCES

Arnold JB, Liow JS, Schaper KA *et al*. (2001) Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effect. *NeuroImage* **13**: 931–43

Ashburner J, Andersson J, Friston KJ (2000) Image registration using a symmetric prior – in three-dimensions. *Hum Brain Mapp* **9**: 212–25

Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* **7**: 254–66

Ashburner J, Friston KJ (2000) Voxel-based morphometry – the methods. *NeuroImage* **11**: 805–21

Avants B, Gee JC (2004) Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage* **23**: S139–S50

Belaroussi B, Milles J, Carme S *et al*. (2006) Intensity non-uniformity correction in MRI: Existing methods and their validation. *Med Image Anal* **10**: 234–46

Besag J (1986) On the statistical analysis of dirty pictures. *J R Stat Soc Ser B* **48**: 259–302

Bishop CM (1995) *Neural networks for pattern recognition*. Oxford University Press, Oxford

Cocosco CA, Kollokian V, Kwan RK-S *et al*. (1997) Brainweb: online interface to a 3D MRI simulated brain database. *NeuroImage* **5**: S425

Collins DL, Evans AC, Holmes C *et al*. (1995) Automatic 3D segmentation of neuro-anatomical structures from MRI. In *Proc Information Processing in Medical Imaging (IPMI)*, Bizais Y, Barillot C, Di Paola R (eds). Kluwer Academic Publishers, Dordrecht

Collins DL, Zijdenbos AP, Kollokian V *et al*. (1998). Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imag* **17**: 463–68

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B* **39**: 1–38

Evans AC (1995) Commentary. *Hum Brain Mapp* **2**: 165–89

Fischl B, Salat DH, Busa E *et al*. (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**: 341–55

Fischl B, Salat DH, van der Kouwe AJW *et al*. (2004) Sequence-independent segmentation of magnetic resonance images. *NeuroImage* **23**: S69–S84

Fissell K, Tseytlin E, Cunningham D, *et al*. (2003) Fiswidgets: a graphical computing environment for neuroimaging analysis. *Neuroinformatics* **1**: 111–25

Garza-Jinich M, Yanez O, Medina V *et al*. (1999) Automatic correction of bias field in magnetic resonance images. In *Proc International Conference on Image Analysis and Processing* IEEE Computer Society, CA.

Hellier P, Ashburner J, Corouge I *et al*. (2002) Inter subject registration of functional and anatomical data using SPM. In *Proc Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 2489 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin and Heidelberg

Hellier P, Barillot C, Corouge I *et al*. (2001) Retrospective evaluation of inter-subject brain registration. In *Proc Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Niessen WJ, Viergever MA (eds), vol. 2208 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin and Heidelberg, pp 258–65

Joshi S, Davis B, Jomier M *et al*. (2004) Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* **23**: S151–S60

Kwan RK-S, Evans AC, Pike GB (1996) An extensible MRI simulator for post-processing evaluation. In *Proc Visualization in Biomedical Computing*, Springer Verlag

Laidlaw DH, Fleischer KW, Barr AH (1998) Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Trans Med Imag* **17**: 74–86

Likar B, Viergever MA, Pernuš F (2001) Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE Trans Med Imag* **20**: 1398–410

MacDonald D, Kabani N, Avis D *et al*. (2000) Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage* **12**: 340–56

Mangin J-F (2000) Entropy minimization for automatic correction of intensity nonuniformity. In *Proc IEEE Workshop on Mathematical Methods in Biomedical Image Analysis* IEEE Computer Society, CA

Neal RM, Hinton GE (1998) A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, Jordan MI (ed.) Kluwer Academic Publishers, Dordrecht, pp 355–68

Pham DL, Prince JL (1999) Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans Med Imag* **18**: 737–52

Pitiot A, Delingette H, Thompson PM *et al*. (2004) Expert knowledge-guided segmentation system for brain MRI. *NeuroImage* **23**: S85–S96

Press WH, Teukolsky SA, Vetterling WT *et al*. (1992) *Numerical Recipes in C*, 2nd edn. Cambridge University Press, Cambridge

Rex DE, Maa JQ, Toga AW (2003) The LONI pipeline processing environment. *NeuroImage* **19**: 1033–48

Shattuck DW, Sandor-Leahy SR, Schaper KA *et al*. (2001) Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* **13**: 856–76

Shepp LA, Vardi Y (1982) Maximum likelihood reconstruction in positron emission tomography. *IEEE Trans Med Imag* **1**: 113–22

Sled JG, Zijdenbos AP, Evans AC (1998) A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Trans Med Imag* **17**: 87–97

Styner M (2000) Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Trans Med Imag* **19**: 153–65

Thévenaz P, Blu T, Unser M (2000) Interpolation revisited. *IEEE Trans Med Imag* **19**: 739–58

Thévenaz P, Unser M (2000) Optimization of mutual information for multiresolution image registration. *IEEE Trans Image Process* **9**: 2083–99

Tohka J, Zijdenbos A, Evans A (2004) Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage* **23**: 84–97

Unser M, Aldroubi A, Eden M (1993a) B-spline signal processing: Part I – theory. *IEEE Trans Signal Process* **41**: 821–33

Unser M, Aldroubi A, Eden M (1993b) B-spline signal processing: Part II – efficient design and applications. *IEEE Trans Signal Process* **41**: 834–48

Van Leemput K, Maes F, Vandermeulen D *et al*. (1999a) Automated model-based bias field correction of MR images of the brain. *IEEE Trans Med Imag* **18**: 885–96

Van Leemput K, Maes F, Vandermeulen D *et al*. (1999b) Automated model-based tissue classification of MR images of the brain. *IEEE Trans Med Imag* **18**: 897–908

Van Leemput K, Maes F, Vandermeulen D *et al*. (2001) A statistical framework for partial volume segmentation. In *Proc Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Niessen WJ, Viergever MA (eds) vol. 2208 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin and Heidelberg, pp 204–12

Wells WM III, Grimson WEL, Kikinis R *et al*. (1996a) Adaptive segmentation of MRI data. *IEEE Trans Med Imag* **15**: 429–42

Wells WM III, Viola P, Atsumi H *et al*. (1996b) Multi-modal volume registration by maximisation of mutual information. *Med Image Anal* **1**: 35–51

Xiaohua C, Brady M, Rueckert D (2004) Simultaneous segmentation and registration for medical image. In *Proc Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Barillot C, Haynor DR, Hellier P (eds) vol. 3216 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin and Heidelberg, pp 663–70

Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imag* **20**: 45–57

Zijdenbos AP, Forghani R, Evans AC (2002) Automatic 'pipeline' analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans Med Imag* **21**: 1280–91