Chapter 5

Image Segmentation

Contents

5.1 Introduction

Healthy brain tissue can generally be classified into three broad tissue types on the basis of an MR image. These are grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF). This classification can be performed manually on a good quality T1 image, by simply selecting suitable image intensity ranges which encompass most of the voxel intensities of a particular tissue type. However, this manual selection of thresholds is highly subjective.

Some groups have used clustering algorithms to partition MR images into different tissue types, either using images acquired from a single MR sequence, or by combining information from two or more registered images acquired using different scanning sequences or echo times (eg. proton-density and T2-weighted). The approach adopted here is a modified version of one of these clustering algorithms. The clustering algorithm of choice is the maximum likelihood 'mixture model' algorithm (Hartigan, 1975), which has been extended to include spatial maps of prior belonging probabilities, and also a correction for image intensity non-uniformity that arises for many reasons in MR imaging. Because the tissue classification is based on voxel intensities, partitions derived without the correction can be confounded by these smooth intensity variations.

The model assumes that the MR image (or images) consists of a number of distinct tissue types (clusters) from which every voxel has been drawn. The intensities of voxels belonging to

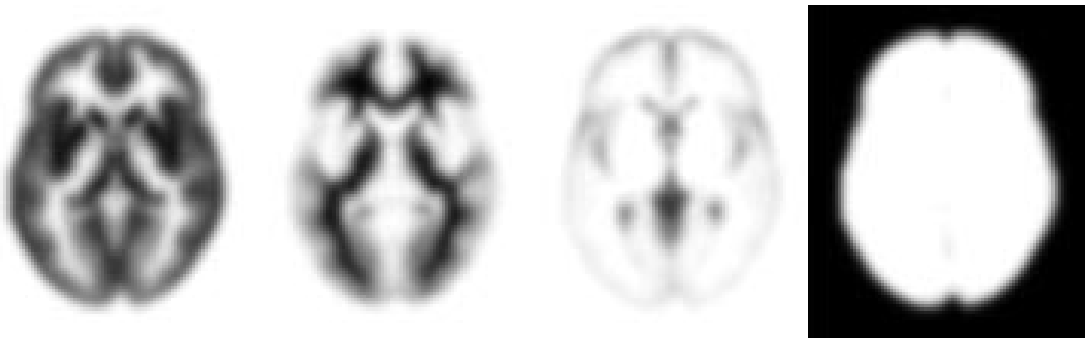

Figure 5.1: The *a priori* probability images of GM, WM, CSF and non-brain tissue. Values range between zero (white) and one (black).

each of these clusters conform to a multivariate normal distribution, which can be described by a mean vector, a covariance matrix and the number of voxels belonging to the distribution. In addition, the model has approximate knowledge of the spatial distributions of these clusters, in the form of prior probability images.

Prior to classifying an image, it is necessary to determine the spatial transformation that maps from each voxel in the image to its equivalent location in the *a priori* probability images, thus allowing simple 'on-the-fly' sampling of the probability images. The mapping is normally achieved by least squares matching with template images in the same stereotactic space as the prior probability images. This can be done using nonlinear warping, but the examples provided in this chapter were done using affine registration.

One of the greatest problems faced by tissue classification techniques is non-uniformity of the images intensity. Many groups have developed methods for correcting intensity non-uniformities, and the scheme developed here shares common features. There are two basic models describing image noise properties: multiplicative noise and additive noise. The multiplicative model describes images that have noise added before being modulated by the non-uniformity field (i.e., the standard deviation of the noise is multiplied by the modulating field), whereas the additive version models noise that is added after the modulation (standard deviation is constant). The current method uses a multiplicative noise model, which assumes that the errors originate from tissue variability rather than additive Gaussian noise from the scanner. Figure 5.2 illustrates the model used by the classification.

Non-uniformity correction methods all involve estimating a smooth function that modulates the image intensities. If the function is is not forced to be smooth, then it will begin to fit the higher frequency intensity variations due to different tissue types, rather than the low frequency intensity non-uniformity artifact. Spline (Yan & Karp, 1995; Sled *et al.*, 1998) and polynomial (Van Leemput *et al.*, 1999a; Van Leemput *et al.*, 1999b) basis functions are widely used for modelling the intensity variation. In these models, the higher frequency intensity variations are restricted by limiting the number of basis functions. In the current method, a Bayesian model is used, where it is assumed that the modulation field ($\mathbf{U}$) has been drawn from a population for which the *a priori* distribution is known, thus allowing high frequency variations of the modulation field to be penalised.

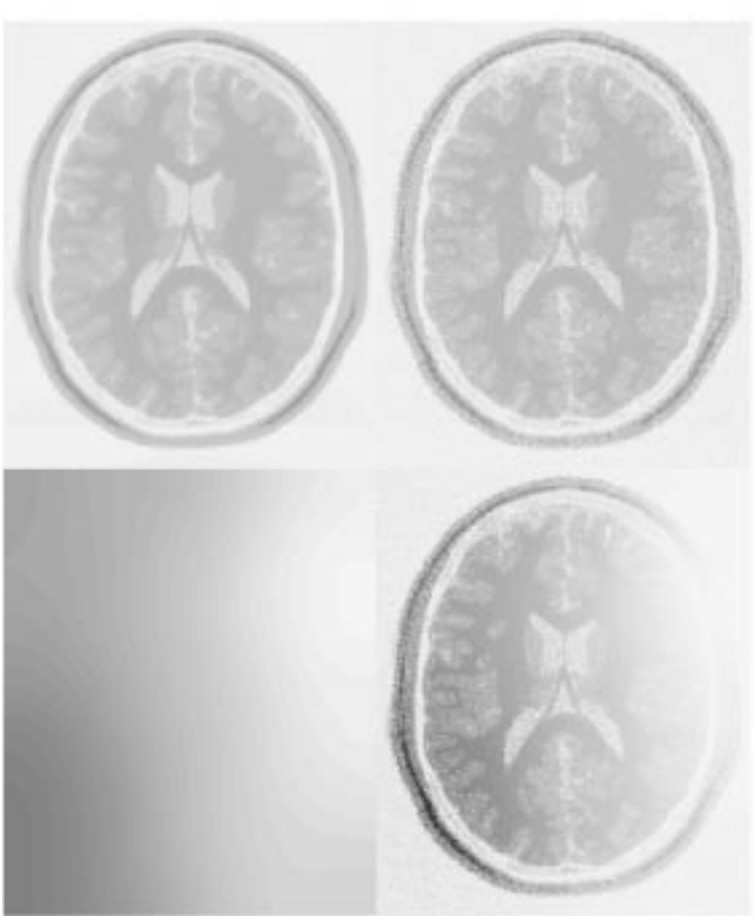

Figure 5.2: The MR images are modelled as a number of distinct clusters (top left), with different levels of Gaussian random noise added to each cluster (top right). The intensity modulation is assumed to be smoothly varying (bottom left), and is applied as a straightforward multiplication of the modulation field with the image (bottom right).

5.2 Methods

Although a three dimensional implementation of the tissue classification method has been developed, that can also be applied to multi-spectral images, the explanation of the algorithm will be simplified by describing its application to a single two dimensional image.

The tissue classification model makes a number of assumptions. The first is that each of the $I \times J$ voxels of the image ($\mathbf{F}$) has been drawn from a known number (K) of distinct tissue classes (clusters). The distribution of the voxel intensities within each class is normal (or multi-normal for multi-spectral images) and initially unknown. The distribution of voxel intensities within cluster k is described by the number of voxels within the cluster (h_k), the mean for that cluster (v_k), and the variance around that mean (c_k).

Because the images are matched to a particular stereotactic space, prior probabilities of the voxels belonging to the grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) classes are known. This information is in the form of probability images – provided by the Montréal Neurological Institute (Evans *et al.*, 1992; Evans *et al.*, 1993; Evans *et al.*, 1994) as part of the ICBM, NIH P-20 project (Principal Investigator John Mazziotta). They were derived from scans of 152 young healthy subjects (66 female and 86 male, 129 right handed, 14 left handed and 9 unknown handedness, aged between 18 and 44, with a mean age of 25 and median age of 24) that were segmented into binary images of GM, WM and CSF, and all normalised into the same space using a 9 parameter (3 translations, 3 rotations and 3 orthogonal zooms) affine transformation. The images were originally classified using a neural network approach, and misclassified non-brain tissue was removed by a masking procedure. The probability images are the means of these binary images, and contain values in the range of zero to one. They represent the *a priori* probability of a voxel being either GM, WM or CSF after an image has been normalised to the same space using a 9 parameter affine transformation (see Figure 5.1). To increase the stability of the classification with respect to small registration errors, the prior probability images have been convolved with an 8mm full width at half maximum Gaussian smoothing kernel. The prior probability of a voxel at co-ordinate i, j belonging to cluster k is denoted by b_{ijk}[1].

The final assumption is that the intensity and noise associated with each voxel in the image has been modulated by multiplication with an unknown smooth scalar field.

There are many unknown parameters to be determined by the classification algorithm, and estimating any of these requires knowledge of the others. Estimating the parameters that describe a cluster (h_k, v_k and c_k) relies on knowing which voxels belong to the cluster, and also the form of the intensity modulating function. Estimating which voxels should be assigned to each cluster requires the cluster parameters to be defined, and also the modulation field. In turn, estimating the modulation field needs the cluster parameters and the belonging probabilities.

The problem requires an iterative algorithm (see Figure 5.3). It begins by assigning starting estimates for the various parameters. The starting estimate for the modulation field is typically uniformly one. Starting estimates for the belonging probabilities of the GM, WM and CSF partitions are based on the prior probability images. Since there are no prior probability maps for background and non-brain tissue clusters, they are estimated by subtracting the prior probabilities

[1]Note that ij subscripts are used for voxels rather than the single subscripts used in the previous chapters. This is to facilitate the explanation of how the modulation field is estimated for 2D images as described in Section 5.2.3.

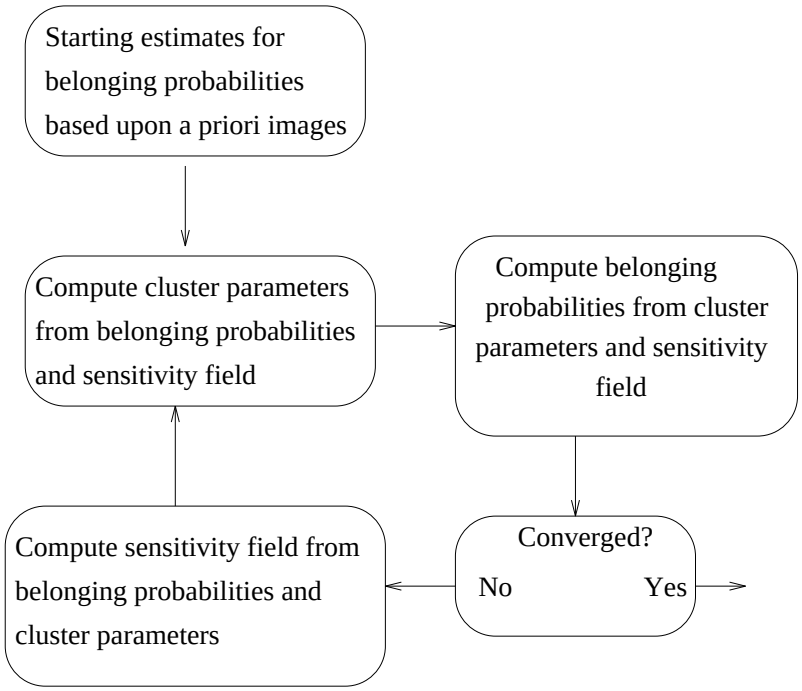

Figure 5.3: A flow diagram for the tissue classification.

for GM, WM and CSF from a map of all ones, and dividing the result equally between the remaining clusters [2].

Each iteration of the algorithm involves estimating the cluster parameters from the non-uniformity corrected image, assigning belonging probabilities based on the cluster parameters, checking for convergence, and re-estimating and applying the modulation function. With each iteration, the parameters describing the distributions move towards a better fit and the belonging probabilities ($\mathbf{P}$) change slightly to reflect the new distributions. This continues until a convergence criterion is satisfied. The parameters describing clusters with corresponding *a priori* probability images tend to converge more rapidly than the others. This may be partly due to the better starting estimates. The final values for the belonging probabilities are in the range of 0 to 1, although most values tend to stabilise very close to one of the two extremes. The algorithm is in fact an *expectation maximisation* (EM) approach, where the *E-step* is the computation of the belonging probabilities, and the *M-step* is the computation of the cluster and non-uniformity correction parameters. The individual steps involved in each iteration are now described in more detail.

5.2.1 Estimating the Cluster Parameters

This stage requires the original image to be intensity corrected according to the most recent estimate of the modulation function. Each voxel of the intensity corrected image is denoted by g_{ij}, and the current estimate of the probability of voxel i, j belonging to class k is denoted by p_{ijk}. The first step is to compute the number of voxels ($\mathbf{h}$) belonging to each of the K clusters

[2] Where identical prior probability maps are used for more than one cluster, the affected cluster parameters need to be modified slightly. This is typically done after the first iteration, by assigning different values for the means uniformly spaced between zero and the intensity of the white matter cluster.

as:

$$h_k = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} \text{ over } k = 1..K. \tag{5.1}$$

Mean voxel intensities for each cluster ($\mathbf{v}$) are computed. This step effectively produces a weighted mean of the image voxels, where the weights are the current belonging probability estimates:

$$v_k = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} g_{ij}}{h_k} \text{ over } k = 1..K. \tag{5.2}$$

Then the variance of each cluster ($\mathbf{c}$) is computed in a similar way to the mean:

$$c_k = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} (g_{ij} - v_k)^2}{h_k} \text{ over } k = 1..K. \tag{5.3}$$

5.2.2 Assigning Belonging Probabilities

The next step is to to re-calculate the belonging probabilities. It uses the cluster parameters computed in the previous step, along with the prior probability images and the intensity modulated input image. Bayes rule is used to assign the probability of each voxel belonging to each cluster:

$$p_{ijk} = \frac{r_{ijk} s_{ijk}}{\sum_{l=1}^{K} r_{ijl} s_{ijl}} \text{ over } i = 1..I,\ j = 1..J \text{ and } k = 1..K. \tag{5.4}$$

where p_{ijk} is the *a posteriori* probability that voxel i, j belongs to cluster k given its intensity of g_{ij}, r_{ijk} is the likelihood of a voxel in cluster k having an intensity of g_{ik}, and s_{ijk} is the *a priori* probability of voxel i, j belonging in cluster k.

The likelihood function is obtained by evaluating the probability density functions for the clusters at each of the voxels:

$$r_{ijk} = (2\pi c_k)^{-1/2} \exp\left(\frac{-(g_{ij} - v_k)^2}{2c_k}\right) \text{ over } i = 1..I,\ j = 1..J \text{ and } k = 1..K. \tag{5.5}$$

The prior (s_{ijk}) is based on two factors: the number of voxels currently belonging to each cluster (h_k), and the prior probability images derived from a number of images (b_{ijk}). With no knowledge of the *a priori* spatial distribution of the clusters or the intensity of a voxel, then the *a priori* probability of any voxel belonging to a particular cluster is proportional to the number of voxels currently included in that cluster. However, with the additional data from the prior probability images, a better estimate for the priors can be obtained:

$$s_{ijk} = \frac{h_k b_{ijk}}{\sum_{l=1}^{I} \sum_{m=1}^{J} b_{lmk}} \text{ over } i = 1..I,\ j = 1..J \text{ and } k = 1..K. \tag{5.6}$$

Convergence is ascertained by following the log-likelihood function:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \log\left(\sum_{k=1}^{K} r_{ijk} s_{ijk}\right) \tag{5.7}$$

The algorithm is terminated when the change in log-likelihood from the previous iteration becomes negligible.

5.2.3 Estimating and Applying the Modulation Function

To reduce the number of parameters describing an intensity modulation field, it is modelled by a linear combination of low frequency discrete cosine transform (DCT) basis functions (see Section 3.2.3), which were chosen because there are no constraints at the boundary. The sensitivity correction field is computed by estimating the coefficients ($\mathbf{q}$) of the DCT basis functions that minimise the weighted sum of squared differences between the data and the model, and also a penalty function based on the smoothness of the modulation field. This can be expressed using matrix terminology as a regularised weighted least squares fitting:

$$\mathbf{q} = \left(\mathbf{A_1}^T\mathbf{A_1} + \mathbf{A_2}^T\mathbf{A_2} \cdots + \mathbf{C_0}^{-1}\right)^{-1} \left(\mathbf{A_1}^T\mathbf{b_1} + \mathbf{A_2}^T\mathbf{b_2} \cdots + \mathbf{C_0}^{-1}\mathbf{q_0}\right) \tag{5.8}$$

where $\mathbf{q_0}$ and $\mathbf{C_0}$ are the means and covariance matrices describing the *a priori* distribution of the coefficients. Matrix $\mathbf{A_k}$ and column vector $\mathbf{b_k}$ are constructed for cluster k from:

$$\mathbf{A_k} = \text{diag}\left(\mathbf{p_k} c_k^{-1/2}\right) \text{diag}\left(\mathbf{f}\right) \mathbf{D} \text{ and } \mathbf{b_k} = \mathbf{p_k} c_k^{-1/2} v_k \tag{5.9}$$

where matrix $\mathbf{D}$ contains the two or three dimensional DCT basis functions, and $\mathbf{p_k}$ refers to the belonging probabilities for the kth cluster considered as a column vector. Image $\mathbf{F}$ is similarly considered as a column vector $\mathbf{f}$. This model aims to find the smooth modulating function (described by its DCT coefficients), that will bring the voxel intensities of each cluster as close as possible (in the least squares sense) to the cluster means, where the vectors $\mathbf{p_k} c_k^{-1/2}$ are voxel by voxel weighting functions.

A two (or three) dimensional discrete cosine transform (DCT) is performed as a series of one dimensional transforms, which are simply multiplications with the DCT matrix. The elements of a matrix ($\mathbf{D}$) for computing the first M coefficients of the one dimensional DCT of a vector of length I is given by:

$$\begin{aligned} d_{i1} &= \frac{1}{\sqrt{I}} \quad i = 1..I \\ d_{im} &= \sqrt{\frac{2}{I}} \cos\left(\frac{\pi(2i-1)(m-1)}{2I}\right) \quad i = 1..I, m = 2..M \end{aligned} \tag{5.10}$$

The matrix notation for computing the first $M \times N$ coefficients of the two dimensional DCT of a modulation field $\mathbf{U}$ is $\mathbf{Q} = \mathbf{D_1}^T\mathbf{U}\mathbf{D_2}$, where the dimensions of the DCT matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ are $I \times M$ and $J \times N$ respectively, and $\mathbf{U}$ is an $I \times J$ matrix. The approximate inverse DCT is computed by $\mathbf{U} \simeq \mathbf{D_1}\mathbf{Q}\mathbf{D_2}^T$. An alternative representation of the two dimensional DCT is obtained by reshaping the $I \times J$ matrix $\mathbf{U}$ so that it is a vector ($\mathbf{u}$). Element $i + (j-1) \times I$ of the vector is then equal to element i, j of the matrix. The two dimensional DCT can then be represented by $\mathbf{q} = \mathbf{D}^T\mathbf{u}$, where $\mathbf{D} = \mathbf{D_2} \otimes \mathbf{D_1}$ (the Kronecker tensor product of $\mathbf{D_2}$ and $\mathbf{D_1}$), and $\mathbf{u} \simeq \mathbf{D}\mathbf{q}$.

Computing $\mathbf{A_k}^T\mathbf{A_k}$ and $\mathbf{A_k}^T\mathbf{b_k}$ could be potentially very time consuming, especially when applied in three dimensions. However, this operation can be greatly speeded up using the properties of Kronecker tensor products (see Chapter 3). Figure 5.4 shows how this can be done in two dimensions using Matlab as a form of pseudo-code.

The Prior Probability Distribution

In Eqn. 5.8, $\mathbf{q_0}$ and $\mathbf{C_0}$ represent a multi-normal *a priori* probability distribution for the basis function coefficients. The mean of the *a priori* distribution is such that it would generate a field

```
alpha_k = zeros(M*N,M*N);
beta_k  = zeros(M*N,1);
weight  = P_k*(c_k^(-0.5));
img1    = weight.*F;
img2    = weight*v_k;
for j = 1:J,
   tmp     = (img1(:,j)*ones(1,M)).*D1;
   alpha_k = alpha_k + kron(D2(j,:)'*D2(j,:), tmp'*tmp);
   beta_k  = beta_k  + kron(D2(j,:)', tmp'*img2(:,j));
end;
```

Figure 5.4: The algorithm for computing $\mathbf{A_k}^T\mathbf{A_k}$ ($alpha\_k$) and $\mathbf{A_k}^T\mathbf{b_k}$ ($beta\_k$) in two dimensions using Matlab as a pseudocode (c.f., Figure 3.5). The symbol "$*$" refers to matrix multiplication, whereas "$.*$" refers to element by element multiplication. "$'$" refers to a matrix transpose and "$\hat{}$" to a power. The jth row of matrix "$D2$" is denoted by "$D2(j,:)$", and the jth column of matrix "$img2$" is denoted by "$img2(:,j)$". The functions "$zeros(a,b)$" and "$ones(a,b)$" would produce matrices of size a$\times$b of either all zero or all one. A Kronecker tensor product of two matrices is represented by the "$kron$" function. Matrix "F" is the $I \times J$ non-uniformity corrected image. Matrix "P_k" is the $I \times J$ current estimate of the probabilities of the voxels belonging to cluster k. Matrices "$D1$" and "$D2$" contain the DCT basis functions, and have dimensions $I \times M$ and $J \times N$. "v_k" and "c_k" are scalers, and refer to the mean and variance of the kth cluster.

that is uniformly one. For this, all the elements of the mean vector are set to zero, apart from the first element that is set to $\sqrt{IJ}$.

The covariance matrix $\mathbf{C_0}$ is such that $(\mathbf{q}-\mathbf{q_0})^T{\mathbf{C_0}}^{-1}(\mathbf{q}-\mathbf{q_0})$ produces an "energy" term that penalises modulation fields that would be unlikely *a priori*. There are many possible forms for this penalty function (see Section 3.2.4). Some widely used simple penalty functions include the "membrane energy" and the "bending energy", which (in three dimensions) have the forms $h = \sum_i \sum_{j=1}^{3} \lambda \left(\frac{\partial u(\mathbf{x}_i)}{\partial x_{ji}}\right)^2$ and $h = \sum_i \sum_{j=1}^{3} \sum_{k=1}^{3} \lambda \left(\frac{\partial^2 u(\mathbf{x}_i)}{\partial x_{ji} \partial x_{ki}}\right)^2$ respectively. In these formulae, $\frac{\partial u(\mathbf{x}_i)}{\partial x_{ji}}$ is the gradient of the modulating function at the ith voxel in the jth orthogonal direction, and λ is a user assigned constant. However, for the purpose of modulating the images, a smoother cost function is used that is based on the squares of the third derivatives (third order regularisation):

$$h = \sum_i \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{l=1}^{3} \lambda \left(\frac{\partial^3 u(\mathbf{x}_i)}{\partial x_{ji} \partial x_{ki} \partial x_{li}}\right)^2 \tag{5.11}$$

This model was chosen because it produces slowly varying modulation fields that can represent the variety of non-uniformity effects that are likely to be encountered in MR images (see Figure 5.5). In two dimensions it can be computed from:

$$\begin{aligned} {\mathbf{C_0}}^{-1} = \quad & \lambda \left(\dddot{\mathbf{D}}_{\mathbf{2}}{}^T \dddot{\mathbf{D}}_{\mathbf{2}}\right) \otimes \left(\mathbf{D_1}^T\mathbf{D_1}\right) + 3\lambda \left(\ddot{\mathbf{D}}_{\mathbf{2}}{}^T \ddot{\mathbf{D}}_{\mathbf{2}}\right) \otimes \left(\dot{\mathbf{D}}_{\mathbf{1}}{}^T \dot{\mathbf{D}}_{\mathbf{1}}\right) + \\ & 3\lambda \left(\dot{\mathbf{D}}_{\mathbf{2}}{}^T \dot{\mathbf{D}}_{\mathbf{2}}\right) \otimes \left(\ddot{\mathbf{D}}_{\mathbf{1}}{}^T \ddot{\mathbf{D}}_{\mathbf{1}}\right) + \lambda \left(\mathbf{D_2}^T\mathbf{D_2}\right) \otimes \left(\dddot{\mathbf{D}}_{\mathbf{1}}{}^T \dddot{\mathbf{D}}_{\mathbf{1}}\right) \end{aligned} \tag{5.12}$$

where the notation $\dot{\mathbf{D}}_{\mathbf{1}}$, $\ddot{\mathbf{D}}_{\mathbf{1}}$ and $\dddot{\mathbf{D}}_{\mathbf{1}}$ refer to the first, second and third derivatives (by differentiating Eqn. 5.10 with respect to i) of $\mathbf{D_1}$, and λ is a user specified hyper-parameter.

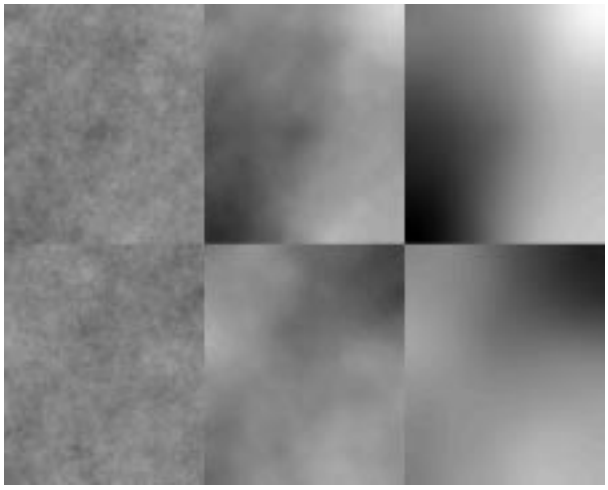

Figure 5.5: Randomly generated modulation fields generated using the membrane energy cost function (left), the bending energy cost function (centre) and the squares of the third derivatives (right). These can be referred to as 1st, 2nd and 3rd order regularisation.

Applying the Correction

Finally, once the coefficients have been estimated, then the modulation field $\mathbf{U}$ can be computed from the estimated coefficients ($\mathbf{Q}$) and the basis functions ($\mathbf{D_1}$ and $\mathbf{D_2}$).

$$u_{ij} = \sum_{n=1}^{N} \sum_{m=1}^{M} d_{2jn} q_{mn} d_{1im} \text{ over } i = 1..I \text{ and } j = 1..J. \tag{5.13}$$

The new estimate for the sensitivity corrected images are then obtained by a simple element by element multiplication with the modulation field.

$$g_{ij} = f_{ij} u_{ij} \text{ over } i = 1..I \text{ and } j = 1..J. \tag{5.14}$$

5.3 Evaluation

In order to provide a qualitative example of the classification, Figure 5.6 shows a single sagittal slice through six randomly chosen T1-weighted images. The initial registration to the prior probability images was via the 12-parameter affine transformation described in Section 3.2.2. The images were automatically classified using the method described here, and contours of extracted grey and white matter are shown superimposed on the images.

Tissue classification was also evaluated using a number of simulated images ($181 \times 217 \times 181$ voxels of $1 \times 1 \times 1$ mm) of the same brain generated by the BrainWeb simulator (Cocosco *et al.*, 1997; Kwan *et al.*, 1996; Collins *et al.*, 1998) with 3% noise (relative to the brightest tissue in the images). The contrasts of the images simulated T1-weighted, T2-weighted and proton density (PD) images (all with 1.5 Tesla field strength), and they were classified individually and in a multi-spectral manner [3]. The T1-weighted image was simulated as a spoiled FLASH sequence,

[3] Note that different modulation fields that account for non-uniformity were assumed for each image of the multi-spectral data-sets.

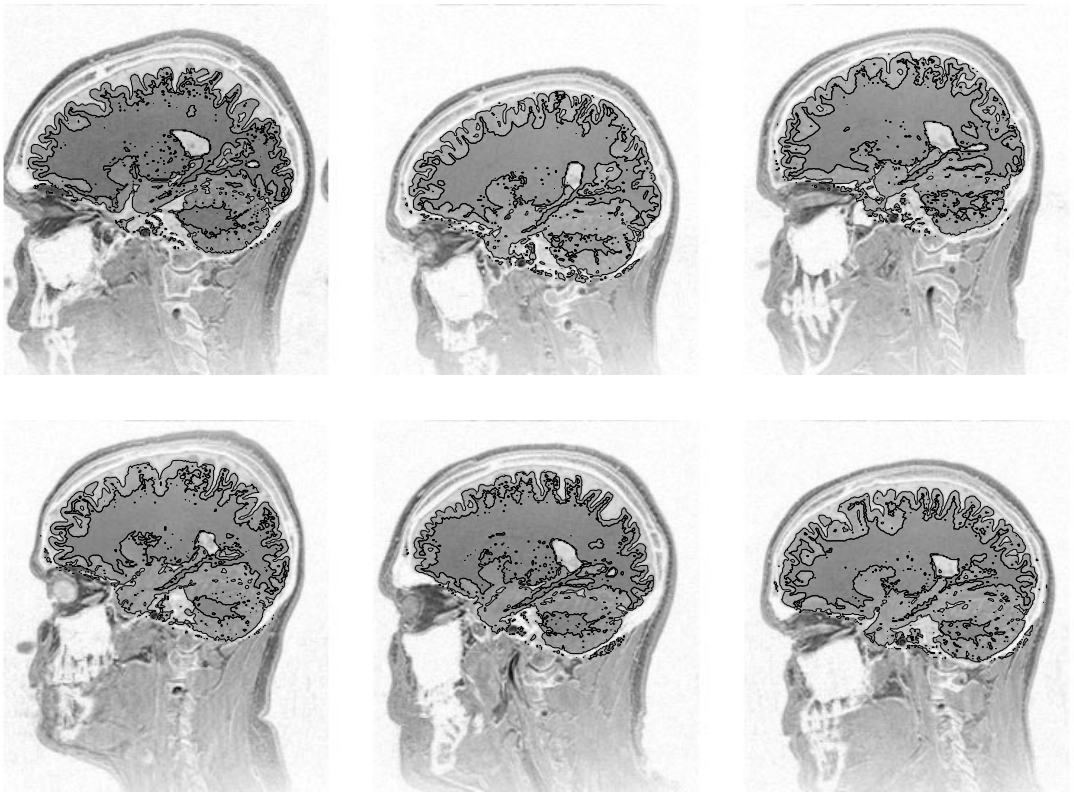

Figure 5.6: A single sagittal slice through six T1-weighted images (2 Tesla scanner, with an MPRAGE sequence, 12^o tip angle, 9.7ms repeat time, 4ms echo time and 0.6ms inversion time). Contours of extracted grey and white matter are shown superimposed on the images.

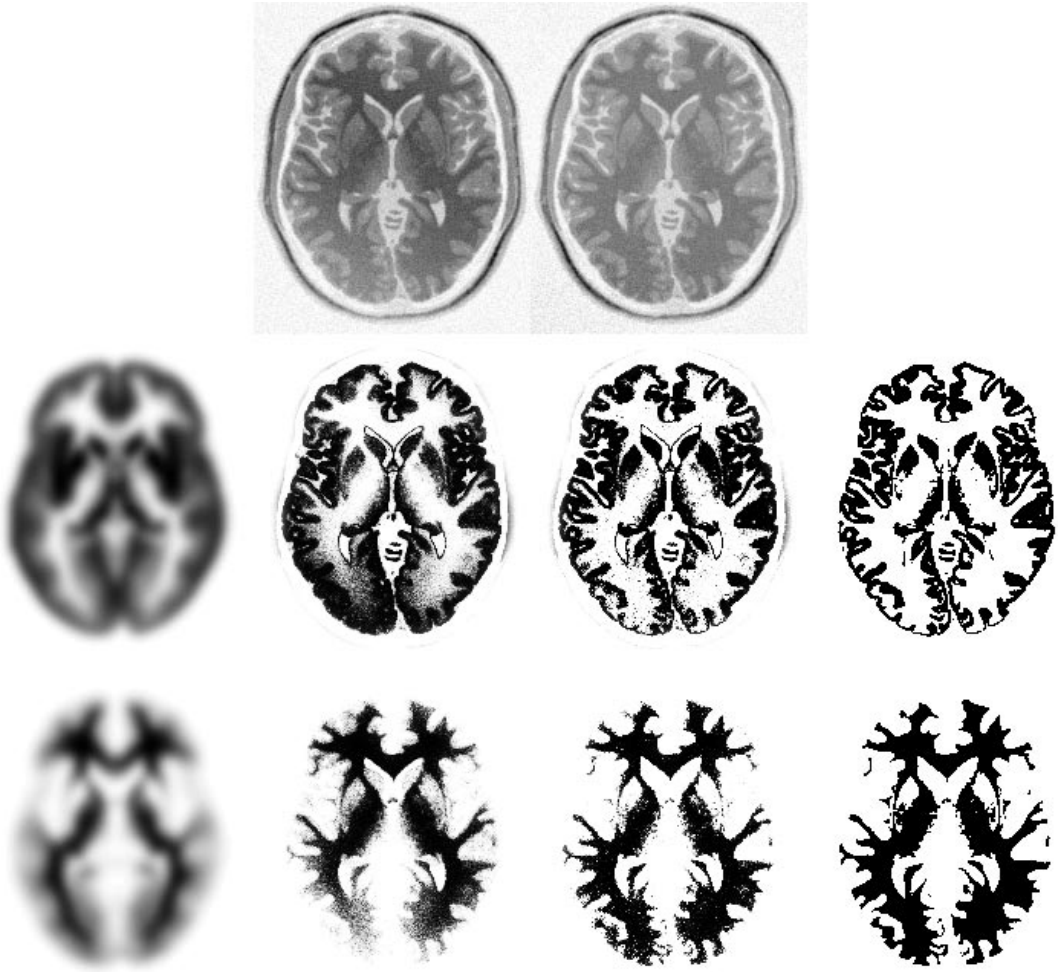

Figure 5.7: The classification of the simulated BrainWeb image. The top row shows the original simulated T1-weighted MR image with 100% non-uniformity, and the non-uniformity corrected version. From left to right, the middle row shows the *a priori* spatial distribution of grey matter used for the classification, grey matter extracted without non-uniformity correction, grey matter extracted with non-uniformity correction and the "true" distribution of grey matter (from which the simulated images were derived). The bottom row is the same as the middle, except that it shows white matter rather than grey. Without non-uniformity correction, the intensity variation causes some of the white matter in posterior areas to be classified as grey. This was also very apparent in the cerebellum because of the intensity variation in the inferior-superior direction.

with a 30^o flip angle, 18ms repeat time, 10ms echo time. The T2 and PD images were simulated by a dual echo, spin echo technique, with 90^o flip angle, 3300ms repeat time and echo times of 35 and 120ms. Three different levels of image non-uniformity were used: 0%RF - which assumes that there is no intensity variation artifact, 40%RF - that assumes a fairly typical amount of non-uniformity, and 100%RF which is more non-uniformity than would normally be expected. The simulated images were classified, both with and without sensitivity correction. Three partitions were considered in the evaluation: grey matter, white matter and other (not grey or white), and each voxel was assigned to the most likely partition. Because the data from which the simulated images were derived was available, it was possible to compare the classified images with ground truth images of grey and white matter using the κ statistic (a measure of inter-rater agreement):

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{5.15}$$

where p_o is the observed proportion of agreement, and p_e is the expected proportion of agreements by chance. If there are N observations in K categories, the observed proportional agreement is:

$$p_o = \sum_{k=1}^{K} f_{kk}/N \tag{5.16}$$

where f_{kk} is the number of agreements for the kth category. The expected proportion of agreements is given by:

$$p_e = \sum_{k=1}^{K} r_k c_k / N^2 \tag{5.17}$$

where r_k and c_k are the total number of voxels in the kth class for both the "true" and estimated partitions.

The classification of a single plane of the simulated T1 weighted BrainWeb image with the 100% non-uniformity is illustrated in Figure 5.7. It should be noted that no pre-processing to remove scalp or other non-brain tissue was performed on the image. In theory, the tissue classification method should produce slightly better results if this non-brain tissue is excluded from the computations. As the algorithm stands, a small amount of non-brain tissue remains in the grey matter partition, which has arisen from voxels that lie close to grey matter and have similar intensities.

The resulting κ statistics from classifying the different simulated images are shown in table 5.1. These results show that the non-uniformity correction made little difference to the tissue classification of images without any non-uniformity artifact. For images containing non-uniformity artifact, classifications using the correction were of about the same quality as the classifications without the artifact, and very much better than classifications without the correction. These results show the same general trends as the results presented by Van Leemput *et. al.* (1999b). The κ statistics by themselves do not really provide much intuition about the quality of the segmentation. As a guide, the segmentations shown in Figure 5.7 produced κ values of 0.85 and 0.94.

A by-product of the classification is the estimation of an intensity non-uniformity field. Figure 5.8 shows a comparison of the intensity non-uniformity present in a simulated T1 image with 100% non-uniformity (created by dividing noisless simulated images with 100% non-uniformity and no

non-uniformity) with that recovered by the classification method. A scatter-plot of "true" versus recovered non-uniformity shows a straight line, suggesting that the accuracy of the estimated non-uniformity is very good.

5.3.1 Stability With Respect to Misregistration with the Prior Probability Images

In order for the Bayesian classification to work properly, an image volume must be in register with a set of *a priori* probability images used to instate the priors. Here the effects of misregistration on the accuracy of segmentation are examined, by artificially translating (in the left-right direction) the prior probability images by different distances prior to segmenting the image. The 1mm slice thickness, 40% non-uniformity, and 3% noise simulated T1-weighted image (described above) was used for the classification, which included the non-uniformity correction. The κ statistic was computed with respect to the "true" grey and white matter for the different translations, and the results are plotted in Figure 5.9.

In addition to illustrating the effect of misregistration, the figure also gives an indication of how far a brain can deviate from the normal population of brains (that constitute the prior probability images) in order for it to be segmented adequately. Clearly, if the brain cannot be well registered with the probability images, then the segmentation will not be as accurate. This fact also has implications for severely abnormal brains, as they are more difficult to register with images that represent the prior probabilities of voxels belonging to different classes. Segmenting such abnormal brains can be a problem for the algorithm, as the prior probability images are based on normal healthy brains. The profile in Figure 5.9 depends on the smoothness or resolution of the prior probability images. By not smoothing the prior probability images, the segmentation would be optimal for normal, young and healthy brains. However, these images may need to be smoother in order to encompass more variability when patient data are to be analysed.

5.4 Discussion

The segmentation method has been found to be robust and accurate for high quality T1 weighted images, but is not beyond improvement. Currently, each voxel is assigned a probability of belonging to a particular tissue class based only on its intensity and information from the prior probability images. There is a great deal of other knowledge that could be incorporated into the classification. For example, if all a voxel's neighbours are grey matter, then there is a high probability that it should also be grey matter. Other researchers have successfully used Markov random field models to include this information in a tissue classification model (Yan & Karp, 1995; Vandermeulen *et al.*, 1996; Van Leemput *et al.*, 1999b). Another very simple prior, that can be incorporated, is the relative intensity of the different tissue types. For example, when segmenting a T1 weighted image, it is known that the white matter should have a higher intensity than the grey matter, which in turn should be more intense than the CSF. When computing the means for each cluster, this prior information could sensibly be used to bias the estimates.

In order to function properly, the classification method requires good contrast between the different tissue types. However, many central grey matter structures have image intensities that

	Single image			Multi-spectral			
	T1	T2	PD	T2/PD	T1/T2	T1/PD	T1/T2/PD
0%RF - uncorreced	0.95	0.90	0.90	0.93	0.94	0.96	0.94
0%RF - corrected	0.95	0.90	0.90	0.93	0.94	0.96	0.95
40%RF - uncorreced	0.92	0.88	0.79	0.90	0.93	0.95	0.94
40%RF - corrected	0.95	0.90	0.90	0.93	0.94	0.96	0.94
100%RF - uncorrected	0.85	0.85	0.67	0.87	0.92	0.94	0.93
100%RF - corrected	0.94	0.90	0.88	0.92	0.93	0.95	0.94

Table 5.1: This table shows the different κ statistics that were computed after classifying the simulated images.

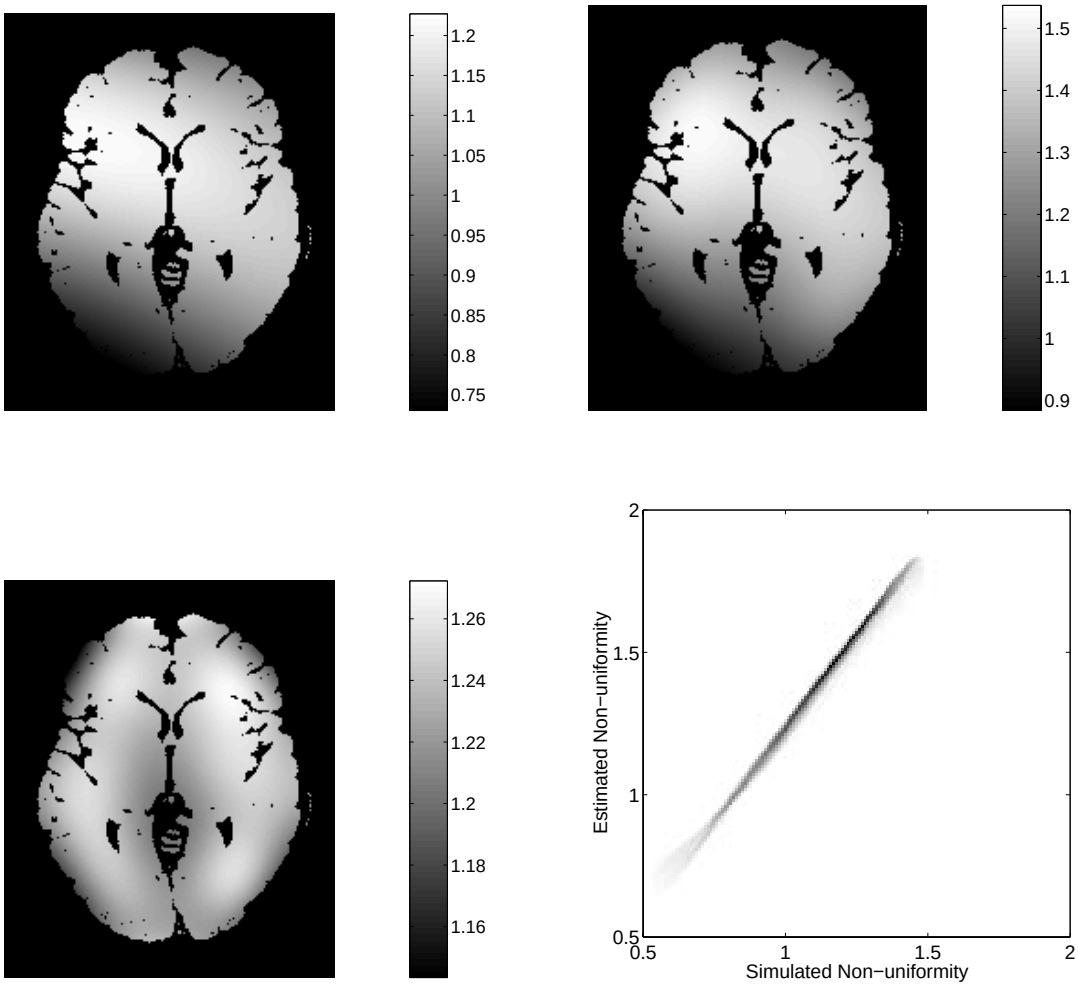

Figure 5.8: Top Left: The true intensity non-uniformity field of the simulated T1 image. Top Right: The non-uniformity recovered by the classification algorithm. Below Left: The recovered divided by the true non-uniformity. Below Right: A scatter-plot of true intensity non-uniformity versus recovered non-uniformity, derived from voxels throughout the whole volume classified as either white or grey matter. Note that the plot is a straight line, but that its gradient is not one because it is not possible to recover the absolute scaling of the field.

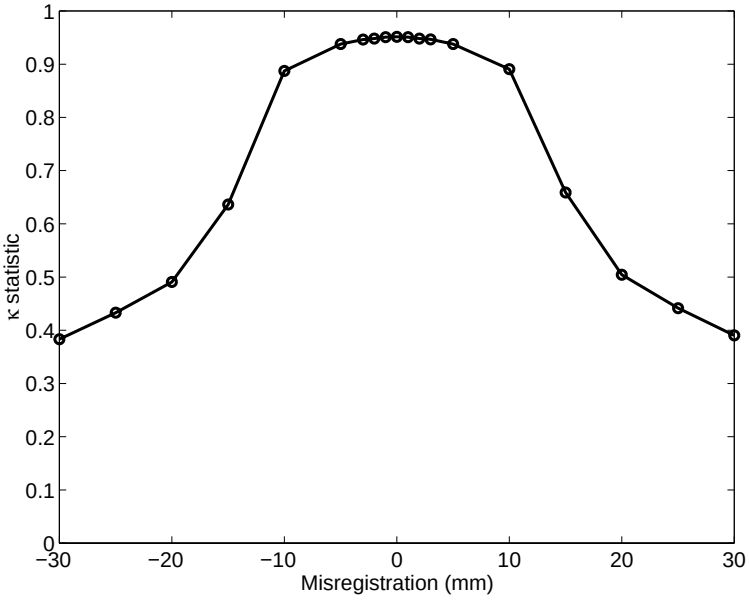

Figure 5.9: Segmentation accuracy with respect to misregistration with the *a priori* images.

are almost indistinguishable from that of white matter, so the tissue classification is not always very accurate in these regions. Another related problem is that of partial volume. Because the model assumes that all voxels contain only one tissue type, the voxels that contain a mixture of tissues may not be modelled correctly. In particular, those voxels at the interface between white matter and ventricles will often appear as grey matter. This can be seen to a small extent in Figures 5.6 and 5.7. Each voxel is assumed to be of only one tissue type, and not a combination of different tissues, so the model's assumptions are violated when voxels contain signal from more than one tissue type. This problem is greatest when the voxel dimensions are large, or if the images have been smoothed, and is illustrated using simulated data in Figure 5.10. The effect of partial volume is that it causes the distributions of the intensities to deviate from normal. Some authors have developed more complex models to describe the intensity distributions of the classes. For example, each non-Gaussian distribution could be modelled by a mixture of Gaussians. An alternative approach is to use logistic discriminant analysis to parameterise a function that best distinguishes between the classes (Bullmore *et al.*, 1995). In this approach, a training set consisting of voxels belonging to each class is first identified. From this training set, the discriminant functions are optimised so that they best distinguish between the classes, before being applied to the whole image. In theory, it may be possible to extend the method so that it uses an EM approach similar to that described in this chapter.

The incorporation of prior probability images into the clustering algorithm produces a much more robust solution when segmenting brains from a similar population to those from which the spatial priors were derived (refer to Section 5.3.1). However, for severely abnormal brains, the priors may not be representative, and this can lead to problems. For example, if a subject has very large ventricles, then CSF may appear where the priors suggest that tissue should always be WM. These CSF voxels are forced to be misclassified as WM, and the intensities of these voxels are incorporated into the computation of the WM means and variances. This results in the WM being characterised by a very broad distribution, so the algorithm is unable to distinguish it

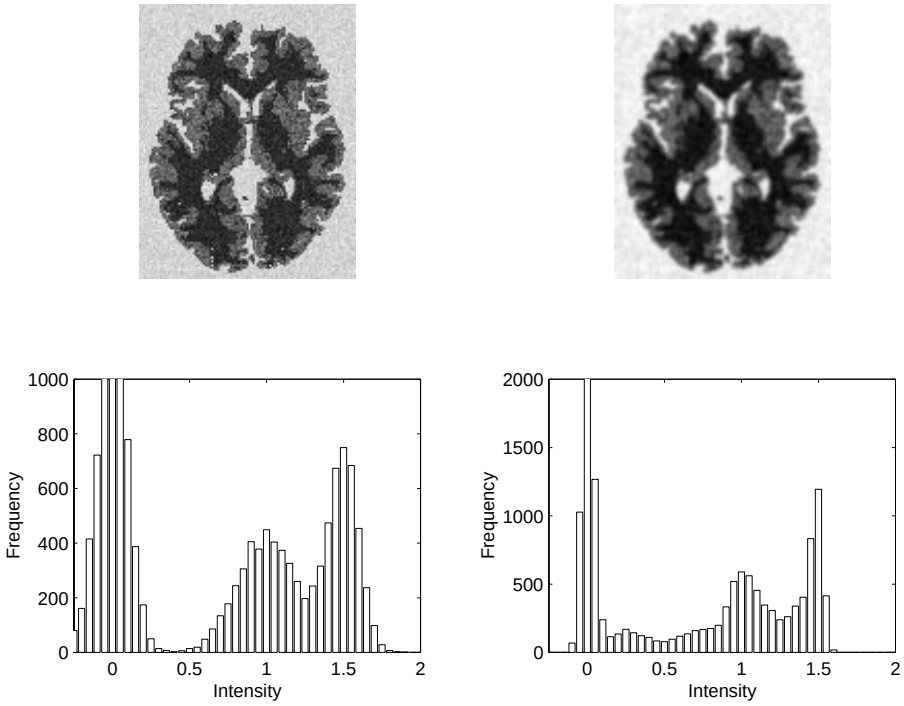

Figure 5.10: Simulated data showing the effects of partial volume on the intensity histograms. On the upper left is a simulated image consisting of three distinct clusters. The intensity histogram of this image is shown on the lower left and consists of three Gaussian distributions. The image at the top right is the simulated image after a small amount of smoothing. The corresponding intensity histogram no longer shows three distinct Gaussian distributions.

from any other tissue. For young healthy subjects, the classification is very good, but caution is required when the method is used for severely pathological brains.

MR images are normally reconstructed by taking the modulus of complex images. Normally distributed complex values are not normally distributed when the magnitude is taken. Instead, they obey a Rician distribution. This means that any clusters representing the background are not well modelled by a single Gaussian, but it makes very little difference for most of the other clusters.

The segmentation is normally run on unprocessed brain images, where non-brain tissue is not first removed. This results in a small amount of non-brain tissue being classified as brain. However, by using morphological operations on the extracted GM and WM segments, it is possible to remove most of this extra tissue. The procedure begins by eroding the extracted WM image, so that any small specs of misclassified WM are removed. This is followed by conditionally dilating the eroded WM, such that dilation can only occur where GM and WM were present in the original extracted segments. Although some non-brain structures (such as part of the sagittal sinus) may remain after this processing, most non-brain tissue is removed. Figure 5.11 shows how the GM and WM partitions can be cleaned up using this procedure, and surface rendered images of brains automatically extracted this way are shown in Figure 4.18. When applied to the segmented simulated T1 weighted image with 3% noise and no nonuniformity artifact, this cleanup procedure increased κ by 0.001.

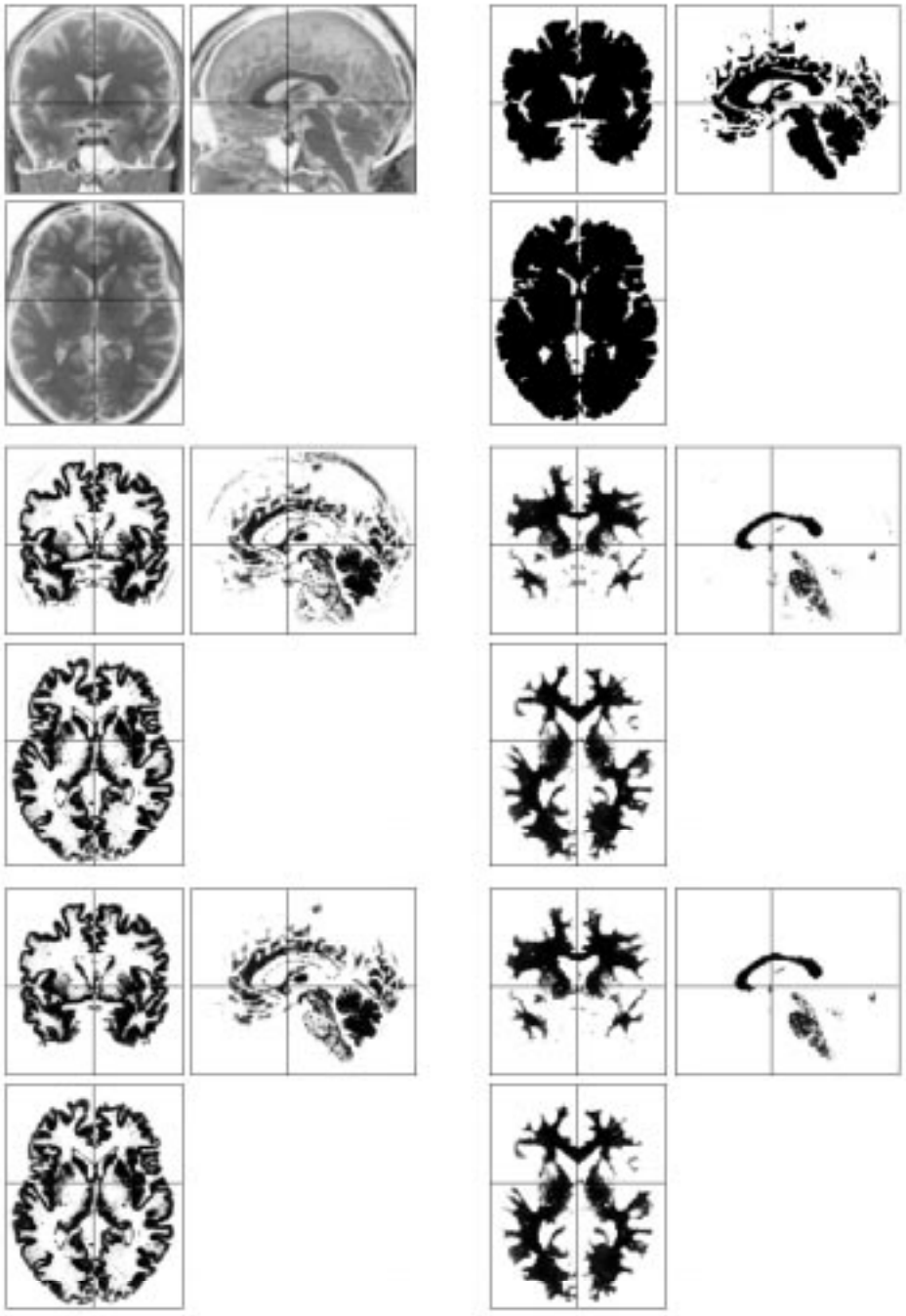

Figure 5.11: Example of automatically cleaned up segmented images. The top row shows the original T1 weighted MR image, next to an automatically generated mask of brain derived from the initial grey and white matter partitions. The second row shows the initial extracted grey and white matter. The bottom row shows the grey and white matter partitions after cleaning up by multiplying with the brain mask.