Katholieke Universiteit Leuven

Departement Elektrotechniek
Afdeling ESAT/PSI
Centrum voor Beeld-en Spraakverwerking
Kardinaal Mercierlaan 94
B-3001 Heverlee - Belgium

TECHNISCH RAPPORT-TECHNICAL REPORT

Automated model-based tissue classification of MR images of the brain

Koen Van Leemput, Frederik Maes,
Dirk Vandermeulen and Paul Suetens

March 1999

Nr.KUL/ESAT/PSI/9903

Automated model-based tissue classification of MR images of the brain

Koen Van Leemput, Frederik Maes*, Dirk Vandermeulen, Paul Suetens

Katholieke Universiteit Leuven
Medical Image Computing (ESAT-Radiology)
UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
E-mail: Koen.Vanleemput@uz.kuleuven.ac.be

Abstract

We describe a fully automated method for model-based tissue classification of Magnetic Resonance (MR) images of the brain. The method interleaves classification with estimation of the model parameters, improving the classification at each iteration. The algorithm is able to segment single- and multi-spectral MR images, corrects for MR signal inhomogeneities and incorporates contextual information by means of Markov Random Fields. A digital brain atlas containing prior expectations about the spatial location of tissue classes is used to initialize the algorithm. This makes the method fully automated and therefore provides objective and reproducible segmentations. We have validated the technique on simulated as well as on real MR images of the brain.

1 Introduction

The study of many brain disorders involves the accurate segmentation of Magnetic Resonance (MR) images of the brain. In multiple sclerosis, for instance, quantification of white matter lesions is needed for assessing the effect of drug treatment, while in schizophrenia and epilepsy, volumetry of white matter, gray matter and cerebro-spinal fluid (csf) is important for morphological analysis. Since such studies typically involve a huge amount of data, manual tracing by a human expert is too time-consuming. Furthermore, human experts show significant intra- and inter-observer variability, which complicates the analysis of the resulting segmentations. Therefore, there is a need for automated methods that produce fast, reliable and reproducible segmentations.

A number of approaches have been investigated in previous research, among which intensity-based clustering methods have received a lot of attention. However, since brain and non-brain tissues often show spectral overlap in MR images, these methods generally require some post-processing in order to improve the segmentations. Furthermore, the presence of a spatially smoothly varying intensity inhomogeneity or bias field in MR images complicates matters. A number of promising methods have been presented to cope with these problems [1, 2, 3], but they still require some user interaction, making their results not fully objective and reproducible.

In this view, we present a new model-based classification method for MR images of the brain that is fully automated. Building on previous work [4, 5], we here present a method that interleaves classification with estimation of model parameters. The algorithm can segment single- or multi-spectral MR data, corrects for MR bias fields and incorporates contextual information through the concept of Markov Random Fields. The additional use of a digital brain atlas makes the method fully automated and thus an objective and reproducible tool for segmenting large amounts of data.

Section 2 describes the model-based method for automated tissue classification of MR images of the brain. The method is validated in section 3 on simulated and real MR images, and we discuss the results in section 4. Finally, section 5 summarizes the conclusions of the approach.

*Frederik Maes is Postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO-Vlaanderen, Belgium)

2 ML approach for model-based segmentation of the brain

Let the intensities in an MR image be conveniently denoted as a 1-D array $y = \{y_1, y_2, \ldots, y_n\}$ where n is the number of voxels. Let the underlying segmentation be denoted as $z = \{z_1, z_2, \ldots, z_n\}$ where z_i indicates to which tissue type voxel i belongs. Assuming that there are K tissue types, $z_i = e_k$ for some $k, 1 \leq k \leq K$ where e_k is a unit vector whose kth component is 1, all the other components being 0. The segmentation process can then be defined as an estimation problem where the underlying hidden segmentation z has to be estimated from the observed intensities y.

In this paper, we use a parameterized model for MR images of the brain to automate this estimation process. More specifically, the hidden segmentation z is modeled as the realization of a random process with some probability density function $f(z \mid \Phi_z)$ that is parameterized by the parameter set Φ_z. Furthermore, it is assumed that z has generated the observed intensities y with probability density function $f(y \mid z, \Phi_y)$ parameterized by Φ_y. As will be shown in the following, estimation of the segmentation z is straightforward once the model parameters $\Phi = \{\Phi_y, \Phi_z\}$ are known. However, estimation of these model parameters in its turn is helped by knowledge of the segmentation. Intuitively, both the segmentation and the model parameters can be estimated simultaneously by interleaving the segmentation with estimation of the model parameters.

The Expectation-Maximization (EM) algorithm [6] formalizes this intuitive approach. It estimates the maximum likelihood (ML) parameters $\hat{\Phi}$

$$\hat{\Phi} = \arg\max_{\Phi} \log f(y \mid \Phi)$$

by iteratively estimating the hidden data z based on the current parameter estimation Φ, and recalculating Φ that maximizes the likelihood of the so-called *complete* data $q = (y, z)$. More specifically, the algorithm interleaves two steps:

E-Step: find the function

$$Q(\Phi \mid \Phi^{(m)}) = E[\log f(q \mid \Phi) \mid y, \Phi^{(m)}]$$

M-Step: find

$$\Phi^{(m+1)} = \arg\max_{\Phi} Q(\Phi \mid \Phi^{(m)})$$

with m the iteration number. If during the M-step, the next estimate $\Phi^{(m)}$ is chosen only to ensure $Q(\Phi^{(m+1)} \mid \Phi^{(m)}) > Q(\Phi^{(m)} \mid \Phi^{(m)})$, then the algorithm is called a generalized EM algorithm (GEM). It has been shown that the likelihood $\log f(y \mid \Phi)$ is guaranteed to increase for EM and GEM algorithms [7]. Upon convergence, the segmentation is obtained as a byproduct as the estimated z.

A major problem when segmenting MR images with intensity-based methods is the corruption of the signal with a spatially smoothly varying intensity inhomogeneity or bias field that is caused by equipment limitations and patient-induced electrodynamic interactions [8]. Although not always visible for a human observer, such a bias field can cause serious misclassifications. Therefore, this imaging artifact must be corrected for.

Recently, we developed an automated method for correcting such bias fields in MR images of the brain, using a specific model where each voxel is randomly sampled from a parameterized intensity distribution. This approach is briefly explained in section 2.1. The model is then extended in section 2.2 by adding a regularization component using the concept of Markov-Random-Fields. Section 2.3 explains how a digital brain atlas can be added to the method in order to make it a fully automated tool for segmenting brain tissues from MR images. We describe some implementation issues in section 2.4, and we finally show how the method works on a simple example in section 2.5.

2.1 The independent model

Recently, we developed a method for automated bias field correction for MR images of the brain. Since the segmentation algorithm that is explained in section 2.2 is an extension of it, we briefly describe it here. For a more detailed explanation, the reader is referred to [4, 5].

Consider the following simple model for an MR image: the tissue type of voxel i, represented by z_i, is drawn randomly from a collection of K possible classes, each with its own known probability $P(k)$ for $1 \leq k \leq K$. That is:

$$f(z_i = e_k) = P(k)$$

Suppose furthermore that the intensity of a voxel belonging to class k is normally distributed around a certain mean μ_k with a variance σ_k, grouped in $\theta_k = \{\mu_k, \sigma_k\}$. Furthermore, suppose that the spatially smoothly varying intensity inhomogeneity or bias field can be written as a linear combination $\sum_j c_j \phi_j(x)$ of smoothly varying basis functions $\phi_j(x)$ with $1 \leq j \leq J$ where J is the number of basis functions, and x denotes the spatial position. Let $C = \{c_1, c_2, \ldots, c_J\}$ denote the bias field parameters, and $\Phi_y = \{\theta_1, \ldots, \theta_K, C\}$ the overall model parameters. The bias field in MR is usually modeled as a multiplicative effect. We therefore work on log-transformed intensities, which makes the bias additive. The probability for voxel intensity y_i in voxel i given that it belongs to class k is then

$$f(y_i \mid z_i = e_k, \Phi_y) = G_{\sigma_k}(y_i - \mu_k - \sum_j c_j \phi_j(x_i))$$

where $G_\sigma()$ denotes a zero-mean normal distribution with variance σ.

Application of the EM-algorithm that searches for the ML estimation of the model parameters Φ_y yields the following equations (see Appendix A):

$$p_{ik}^{(m+1)} \equiv [p_i^{(m+1)}]_k = \frac{f(y_i \mid z_i = e_k, \Phi_y^{(m)}) f(z_i = e_k)}{\sum_{j=1}^K f(y_i \mid z_i = e_j, \Phi_y^{(m)}) f(z_i = e_j)} \tag{1}$$

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^n p_{ik}^{(m+1)} \big(y_i - \sum_{j=1}^J c_j^{(m)} \phi_j(x_i)\big)}{\sum_{i=1}^n p_{ik}^{(m+1)}} \tag{2}$$

$$(\sigma_k^{(m+1)})^2 = \frac{\sum_{i=1}^n p_{ik}^{(m+1)} \big(y_i - \mu_k^{(m+1)} - \sum_{j=1}^J c_j^{(m)} \phi_j(x_i)\big)^2}{\sum_{i=1}^n p_{ik}^{(m+1)}} \tag{3}$$

$$\begin{bmatrix} c_1^{(m+1)} \\ c_2^{(m+1)} \\ \vdots \end{bmatrix} = \big(A^t W^{(m+1)} A\big)^{-1} A^t W^{(m+1)} r^{(m+1)} \tag{4}$$

with

$$A = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \cdots \\ \phi_1(x_2) & \phi_2(x_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad W^{(m+1)} = diag(w_i^{(m+1)})$$

$$w_i^{(m+1)} = \sum_{k=1}^K w_{ik}^{(m+1)}, \quad w_{ik}^{(m+1)} = p_{ik}^{(m+1)} / (\sigma_k^{(m+1)})^2$$

$$r^{(m+1)} = \begin{bmatrix} y_1 - \tilde{y}_1^{(m+1)} \\ y_2 - \tilde{y}_2^{(m+1)} \\ \vdots \end{bmatrix}, \quad \tilde{y}_i^{(m+1)} = \frac{\sum_{k=1}^K w_{ik}^{(m+1)} \mu_k^{(m+1)}}{\sum_{k=1}^K w_{ik}^{(m+1)}}$$

where m denotes the iteration number.

These equations can be interpreted as 3 interleaved steps: classification of the voxels (equation 1), estimation of the normal distributions (equations 2 and 3) and estimation of the bias field (equation 4). The estimation of the bias field is worth some attention. Based on the current classification and estimation of the distribution parameters, a predicted signal $\tilde{y}$ is constructed that tries to mimic the measured signal y without the bias field. Furthermore, associated with every voxel is a weight w_i that is inversely proportional to the variance of the class it belongs to. The bias estimation is the weighted least-squares fit to the difference between the measured signal y and the predicted signal $\tilde{y}$. The intensity distribution

of brain tissues is generally much sharper than that of non-brain tissues; therefore the bias field is mainly estimated from the brain and extrapolated to other regions in the image.

Previously, we have given a more detailed description of the algorithm and its practical use [4, 5]. Therefore, we direct the interested reader to these papers for more details. Suffice it here to say that the algorithm is easily extended to multi-spectral MR data by substituting the normal distributions with mean μ_k and variance σ_k to multivariate normals with mean $\vec{\mu}_k$ and covariance matrix Σ_k.

2.2 Regularization using Markov-Random-Fields

The simple independent model classifies the voxels only based on their intensity. This yields acceptable segmentation results as long as the different classes are well separated in intensity feature space, i.e. have a clearly discernible associated intensity distribution. Unfortunately, this is not always true for MR images of the brain. Where such tissues as white matter, gray matter and csf usually have a characteristic intensity, voxels surrounding the brain consist of several tissue types some of which have an MR intensity that is very similar to brain tissue. This results in erroneous classifications of regions surrounding the brain as gray matter or white matter. The problem becomes more and more stringent when only one MR channel is available. Therefore, some form of spatial information is also needed during the classification process. In [4, 5], we used a digital brain atlas that provides prior knowledge about the expected location of white matter, gray matter and csf. This reduced the amount of misclassifications already; however, it did not totally prevent some small regions to be misclassified. It is not unreasonable to assume that these small regions don't have a significant impact on the estimation of the bias field. However, for automated segmentation of brain tissues, the independent model as described in section 2.1 is not always sufficient.

We therefore incorporate general spatial and anatomical constraints, such as "a voxel surrounded by non-brain tissue cannot belong to gray matter" by introducing the concept of Markov Random Fields (MRF's) into the model. The hidden data z are assumed to be the realization of a random process where the probability that voxel i belongs to tissue type k depends on the tissue type of it's neighbors. The Hammersley-Clifford theorem states that such a random field is a Gibbs Random Field, i.e. its configurations obey a Gibbs distribution

$$f(z \mid \Phi_z) = Z(\Phi_z)^{-1} exp[-U_{mrf}(z \mid \Phi_z)]$$

where $Z(\Phi_z) = \sum_z exp[-U_{mrf}(z \mid \Phi_z)]$ is a normalization constant called the *partition function*, and $U_{mrf}(z \mid \Phi_z)$ is an energy function dependent on the MRF parameters Φ_z.

We use a simple MRF that is defined on a so-called first order neighborhood system, i.e. only the six nearest neighbors on the 3D-image lattice are used. Let $\mathcal{N}_i = \{i^a, i^p, i^l, i^r, i^i, i^s\}$ denote the neighborhood of voxel i, where i^a, i^p, i^l and i^r are its four neighbors in the plane, and i^i and i^s its two neighbors out of the plane. The voxel size in MR images is equal in the x and y-direction, but usually different in the z-direction. We therefore assume that the interaction between a voxel with its neighbors in the x-direction is the same as that with its neighbors in the y-direction, but different from that with its neighbors in the z-direction. Therefore, we use the following Ising-like energy function

$$U_{mrf}(z \mid \Phi_z) = \sum_{i=1}^{n} U_{mrf}(z_i \mid z_{\mathcal{N}_i}, \Phi_z)$$

where

$$z_{\mathcal{N}_i} = \{z_{i'} \mid i' \in \mathcal{N}_i\}$$

stands for the set of labels at the six sites neighboring i, and

$$U_{mrf}(z_i \mid z_{\mathcal{N}_i}, \Phi_z) = z_i^t G g_i + z_i^t H h_i$$

where

$$g_i = z_{i_a} + z_{i_p} + z_{i_l} + z_{i_r}$$

is a vector that counts per class k the number of neighbors of i within the plane that belong to k. Similarly,

$$h_i = z_{i_i} + z_{i_s}$$

counts per class k the number of neighbors out of the plane that belong to k. G and H are $K \times K$ matrices that together form the MRF parameters $\Phi_z = \{G, H\}$.

With the addition of the MRF, there are now two parameter sets Φ_y and Φ_z. As explained in Appendix B, the exact calculation of Φ_y is no longer practically feasible, and we use an approximation based on the so-called *mean field* theory as proposed by Zhang *et al.* [9] [10]. More specifically, equations 2, 3 and 4 remain valid, but the classification step is no longer given by equation 1 but by

$$p_{ik}^{(m+1)} \equiv [p_i^{(m+1)}]_k \simeq$$

$$\frac{f(y_i \mid z_i = e_k, \Phi_y^{(m)}) f(z_i = e_k \mid p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}{\sum_j^K f(y_i \mid z_i = e_j, \Phi_y^{(m)}) f(z_i = e_j \mid p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})} \tag{5}$$

where

$$f(z_i = e_k \mid p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}) = \frac{e^{-U_{mrf}(e_k | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}}{\sum_{j=1}^K e^{-U_{mrf}(e_j | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}}$$

The difference lies herein, that in the independent model, each voxel had the same a priori probability $f(z_i = e_k)$ to belong to class k, whereas now, this probability depends on the classification of the neighboring voxels.

The calculation of the MRF parameters $\Phi_z = \{G, H\}$ is more involved. Zhang *et al.* [9] used an approximation based on the same mean field idiom:

$$f(z \mid \Phi_z^{(m)}) \simeq \prod_i \frac{e^{-U_{mrf}(z_i | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}}{\sum_{z_i'} e^{-U_{mrf}(z_i' | p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)})}} \tag{6}$$

This results in an expression that is very similar to Besag's pseudo-likelihood approach [11] for estimation of the MRF parameters Φ_z for fully labeled data z, but now with $p^{(m)}$ from equation 5 instead of z. Unfortunately, even with this pseudo-likelihood approximation, it is not possible to derive closed-form expressions for Φ_z as it was the case for Φ_y. Instead, numerical maximization methods for solving the nonlinear equations must be used. However, these are rather time-consuming and the solution can differ depending on the initial estimate.

We therefore use a more efficient heuristic approach that is non-iterative, using a least-squares (LS) fit procedure [11]. Hereto, we define a $2K^2 \times 1$ vector θ that denotes all parameters G and H

$$\begin{aligned} \theta \;=\; & \big[\; G_{11} \ldots G_{1K} \;\big|\; G_{21} \ldots G_{2K} \;\big|\; \ldots \;\big|\; G_{K1} \ldots G_{KK} \;\big| \\ & \quad H_{11} \ldots H_{1K} \;\big|\; H_{21} \ldots H_{2K} \;\big|\; \ldots \;\big|\; H_{K1} \ldots H_{KK} \;\big]^t \end{aligned}$$

and an equally sized configuration vector v_{z_i, g_i, h_i}

$$\begin{aligned} v_{z_i,g_i,h_i} \;=\; & \big[\; z_{i1} g_i^t \quad z_{i2} g_i^t \quad \ldots \quad z_{iK} g_i^t \;\big| \\ & \quad z_{i1} h_i^t \quad z_{i2} h_i^t \quad \ldots \quad z_{iK} h_i^t \;\big]^t \end{aligned}$$

In Appendix C, we show that for any neighborhood configuration $\{g_i, h_i\}$ and for any two distinct values z_i and z_i', the following holds:

$$[v_{z_i,g_i,h_i}^t - v_{z_i',g_i,h_i}^t]\theta = \log\left(\frac{f(z_i, g_i, h_i \mid \Phi_z)}{f(z_i', g_i, h_i \mid \Phi_z)}\right) \tag{7}$$

Each distinct combination of z_i, z_i', g_i and h_i defines one such an equation. Supposing that $f(z_i, g_i, h_i \mid \Phi_z)$ can somehow be estimated, one obtains an over-determined linear system of equations that can be solved using a standard LS method. This would provide an estimation of the MRF parameters.

During iteration (m+1), we estimate $f(z_i, g_i, h_i \mid \Phi_z^{(m)})$ by constructing a histogram $H(z_i, g_i, h_i \mid \Phi_z^{(m)})$ that counts how many times a specific configuration $\{z_i, g_i, h_i\}$ occurred in the current classification $p^{(m)}$. Since $p^{(m)}$ is a *soft* classification, the contribution of a voxel i to the histogram entry $H(z_i, g_i, h_i \mid$

$\Phi_z^{(m)}$) is given by the probability that its configuration is $\{z_i, g_i, h_i\}$. Subsequently, $f(z_i, g_i, h_i \mid \Phi_z^{(m)})$ is estimated by $H(z_i, g_i, h_i \mid \Phi_z^{(m)})/n$. Solving the linear system of equations 7 then provides the estimation $\Phi_z^{(m+1)}$.

To summarize, the complete EM algorithm interleaves 4 steps: classification of the voxels (equation 5), estimation of the normal distribution parameters (equations 2 and 3), estimation of the bias field (equation 4), and estimation of the MRF parameters (equation 7). It is interesting to notice that the classification step (equation 5) shows remarkable similarity with the Iterated Conditional Modes (ICM) algorithm [11]. ICM is a local optimization algorithm that tries to find the most likely hidden data z given the data y and the parameters Φ_y and Φ_z. It iteratively updates the discrete labels z by assigning the label of each voxel to

$$z_i^{(m+1)} \leftarrow \arg\max_{z_i} \frac{f(y_i \mid z_i, \Phi_y) f(z_i \mid z_{\mathcal{N}_i}^{(m)}, \Phi_z)}{\sum_{z_i'} f(y_i \mid z_i', \Phi_y) f(z_i' \mid z_{\mathcal{N}_i}^{(m)}, \Phi_z)}$$

Comparing the above to equation 5, it can be seen that the EM algorithm is a *soft* implementation of the ICM algorithm, i.e. the classification $p^{(m+1)}$ is updated in a soft sense rather than with discrete labels. The difference with the ICM algorithm lies herein, that the parameters Φ_y and Φ_z are updated each iteration.

2.3 Initialization with a digital brain atlas

The algorithm presented above interleaves classification, distribution parameter estimation, bias field parameter estimation and MRF parameter estimation. To initialize the method, it would not be unlogical to start with the bias field parameters and the MRF parameters set to zero. If the user additionally provides some initial estimates for the distribution parameters, for instance by interactively clicking on some representative voxels for each class in the image, the algorithm can be started with the classification step (equation 5).

We use a different approach. We initialize the algorithm by providing a prior classification and from there the distribution parameters are calculated, after which a first estimation of the bias field and the MRF parameters is calculated. The prior classification is derived from a digital brain atlas, distributed with the SPM96 package [12], that contains spatially varying prior probability maps for the location of white matter, gray matter and csf as shown in figure 1. These probability maps, with an isotropic spatial resolution of 2 millimeter, were obtained by averaging binary white matter, gray matter and csf segmentations of MR brain images from a large number of subjects, after spatial normalization of all images using an affine transformation.

To apply the a priori information of the atlas, we first normalize the atlas to the space of the study image by matching a T1 template provided by SPM96 that is already co-registered with the atlas, to the study image. To this end, we use the affine multi-modality registration technique based on maximization of mutual information of corresponding voxel intensities of Maes *et al.* [13]. It has been shown [13] that this registration criterion is fairly insensitive to moderate bias fields, such that it can be applied fully automatically and reliably to uncorrected MR images. The registered a priori probability maps provided by the atlas are then reformatted and used as an initial classification. The advantage of this way of working is that interactive user intervention is avoided, which makes the results of the algorithm reproducible and objective and allows the method to be fully automated.

The atlas could also be beneficial for constraining the classification process during the subsequent iterations of the algorithm, since it can add extra spatial information in case of overlapping intensities between distinct classes. As previously mentioned, tissues surrounding the brain sometimes show intensities similar to brain tissues. Therefore, we add extra spatial information to the classification process by multiplying the prior probability $f(z_i = e_k \mid z_{\mathcal{N}_i}, \Phi_z^{(m)})$ in equation 5 in every voxel i by the prior probability of class k in the atlas. This also makes the algorithm more robust in case of very severe bias fields.

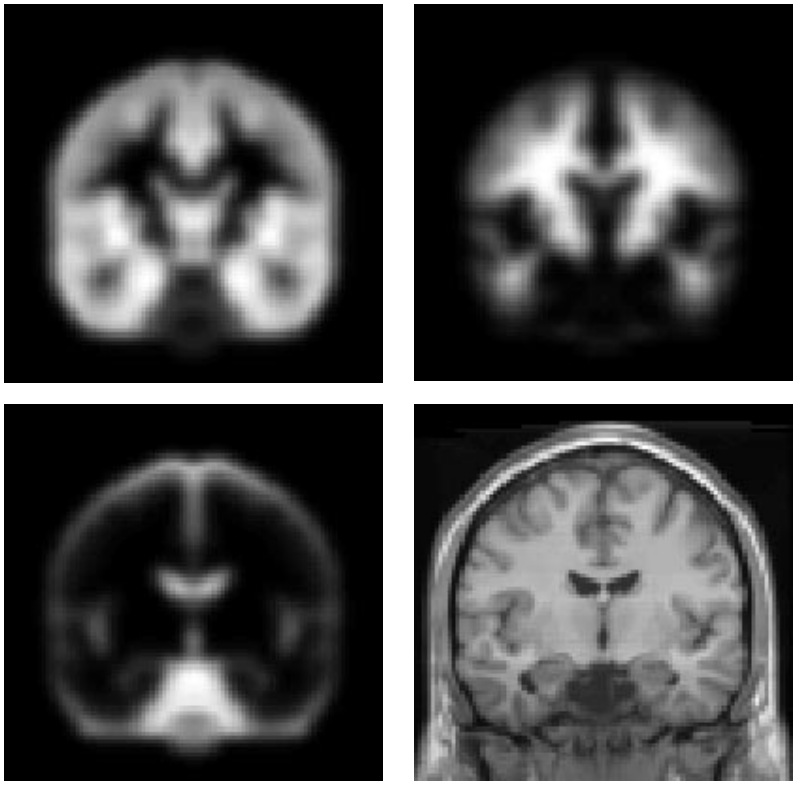


Figure 1: Digital brain atlas with spatially varying prior probability maps for gray matter, white matter and csf. High intensities indicate high a priori probabilities. The atlas also contains a T1 template image which is used to spatially register the atlas to the study image.

2.4 Practical implementation

Before the algorithm described above can be implemented as an automated tool for tissue classification of MR images of the brain, a number of practical issues must be considered. In this section, we describe our practical approach and the implementation that was used to validate the method in section 3.

We principally use 4 classes: white matter (class 1), gray matter (class 2), csf (class 3) and "other" (class 4). The a priori probability map for class 4 is obtained by subtracting the sum of the atlas probability maps for white matter, gray matter and csf from a map of ones.

The MRF parameters G and H are 4×4 matrices; the (k,j)th element describes the contribution to the energy function $U_{mrf}(z_i = e_k \mid z_{\mathcal{N}_i}, \Phi_z)$ of a neighboring voxel that belongs to class j. Direct application of this model favors configurations of z where each class is spatially clustered. More specifically, a homogeneous region of white matter would preferably be surrounded by a homogeneous region of gray matter, on its turn surrounded by csf that is finally surrounded by "other". Small regions outside the brain that are misclassified as brain tissue are not preferable, and the MRF forms an effective means to avoid such misclassifications.

However, it has been described in the literature that fine structures such as the interface between white matter and gray matter can be erased by MRF's [2]. This corresponds to our own experience: the MRF over-regularizes such subtle borders and tempts to produce nicely smooth interfaces. We therefore propose a modification that penalizes impossible combinations such as a gray matter voxel surrounded by voxels belonging to "other", while at the same time preserving edges between tissues that are known to border each other. We impose that a voxel surrounded by gray matter voxels must have the same probability to be white matter as to be gray matter. With the class numbers defined above, this can be achieved by imposing the constraint that $G_{11} = G_{21}$ and $H_{11} = H_{21}$. Similarly, we demand $G_{12} = G_{22}$ and $H_{12} = H_{22}$, which results in an equal probability for white matter and gray matter in a voxel surrounded

by white matter voxels. As a result, voxels surrounded by brain tissues have a low probability for csf and other, and a high but equal probability for white and gray matter. The actual decision between white and gray matter is therefore only based on the intensity, so that the interface between white and gray matter is unaffected by the MRF.

The same rationale applies for the interface between gray matter and csf ($G_{11} = G_{31}$ and $G_{13} = G_{33}$, the same for H), and to the interface between csf and other ($G_{33} = G_{43}$ and $G_{34} = G_{44}$, the same for H). This reduces the number of MRF parameters to be estimated by equation 7 from $2 \times 4^2 = 32$ to 20. The constraint that $G_{jk} = G_{lk}$ can easily be implemented by adding the corresponding columns in equation 7 before solving the linear system of equations.

Whereas white matter, gray matter and csf each have a characteristic intensity distribution that, to a reasonable extent, can be modeled by a single normal distribution, the other class is more difficult to model adequately. Specifically, part of the class "other" is background noise that is not affected by the bias field. We therefore extend the algorithm with an explicit model of the noise in MRI, the details of which can be found in [5]. The rest of the signal of class "other" is formed by the tissues surrounding the brain, which we model by two normal distributions since this models the non-brain tissues more accurately than a single one. To summarize, the class "other"is represented by three parameterized distributions; the prior probability $f(z_i \mid z_{\mathcal{N}_i}, \Phi_z)$ for these classes is obtained by simply dividing the prior probability for "other" equally over these three distributions.

As explained in section 2.2, we model the bias field as a linear combination $\sum_{j=1}^{J} c_j \phi_j(x)$ of spatially smoothly varying basis functions $\phi_j(x)$. The actual choice of the basis functions has still to be specified. In the literature, some authors have used splines to model a bias field [14, 15, 16], while others preferred polynomials [17, 18, 19]. We have chosen 3D polynomials of 4th degree, a basis that showed to be able to correct for most of the bias field in MR images of the brain [5]. 2D multi-slice sequence MR images, which are acquired in an interleaved way, are typically corrupted with a different constant intensity offset for each slice separately. For this kind of images, we model the bias field by a 2D polynomial of 4th degree for each slice separately, rather than one global 3D polynomial.

The EM procedure can be iterated until either the parameter estimates converge or some maximum number of iterations is reached. The ultimate stop criterion detects when the log-likelihood $\log f(y \mid \Phi)$ stops increasing significantly, since this is after all the objective function that the EM algorithm maximizes. However,

$$\log f(y \mid \Phi^{(m)}) = \log \Big[\sum_z f(y \mid z, \Phi_y^{(m)}) f(z \mid \Phi_z^{(m)}) \Big]$$

requires calculation of all the possible realizations of the MRF, which is not computationally feasible. We therefore once more call upon the mean field theory by approximating $f(z \mid \Phi_z^{(m)})$ using equation 6:

$$\begin{aligned} & \log f(y \mid \Phi^{(m)}) \\ \simeq \;& \log \textstyle\sum_z \prod_{i=1}^{n} \Big[f(y_i \mid z_i, \Phi_y^{(m)}) f(z_i \mid p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}) \Big] \\ = \;& \textstyle\sum_{i=1}^{n} \log \sum_{z_i} f(y_i \mid z_i, \Phi_y^{(m)}) f(z_i \mid p_{\mathcal{N}_i}^{(m)}, \Phi_z^{(m)}) \end{aligned} \tag{8}$$

Notice that evaluation of this objective function involves no additional computational burden, since the contribution of each voxel is calculated anyway during the classification step as the denominator of equation 5.

We have implemented the method in C-language and integrated it inside the Matlab-based [20] SPM96-software package [12]. The program is able to segment multi-spectral MR brain images and is fully automated. As a first step, we co-register and reformat all the MR channels in case of multi-spectral data, using the affine registration method based on maximization of mutual information of Maes *et al.* [13]. We then spatially normalize the atlas to the study image with the same registration program. The EM algorithm described above is then used to segment the images; voxels where the atlas indicates a zero prior probability for white matter, gray matter or csf are of no interest and are simply discarded. To speed up the computation, we update the parameters $\Phi^{(m+1)}$ only based on a limited subset of all voxels. The algorithm is stopped when the relative change of the estimation of $f(y \mid \Phi)$ between two subsequent iterations drops below 0.0001, which typically occurs after approximately 25 iterations. The overall calculation time depends on the size of the images involved; it takes less than 30 minutes to segment a single-channel image with dimensions $256 \times 256 \times 60$ on an SGI onyx 2.

2.5 Example

To clarify how the algorithm works, we here illustrate by way of an example the influence each component of the method has on the resulting segmentations. We have processed the same single-channel T1-weighted image of size $256 \times 256 \times 128$ with voxel size of $1 \times 1 \times 1.25$ millimeter (see figure 2 (a)) a number of times, each time leaving some step out of the algorithm.

To show the influence of the bias correction step, we have processed the image twice. The first time, we set the order of the bias field polynomial and the MRF parameter matrices G and H to zero, which reduces the method to the application of the independent model as described in section 2.1 where the bias field estimation step is left out. The second time, G and H were again fixed to zero, but now the order of the bias field polynomial was set to four. Figure 2 (b),(c) shows the gray matter component of the classification probability p upon convergence for both situations. Without bias correction, white matter at the top of the brain is misclassified as gray matter, whereas the result is clearly much better when the bias correction step is added to the algorithm.

We then added the MRF parameter estimation step and again ran the algorithm with bias polynomial order set to four. It can be seen from figure 2 (d) that addition of the MRF results in a better distinction between brain tissues and tissues surrounding the brain, whose intensities show similar characteristics. The MRF cleans up the segmentations of brain tissues, while preserving the detailed interface between gray and white matter, and between gray matter and csf. This is most beneficial in case of single-channel MR data, where it is often difficult to discern tissues surrounding the brain and brain tissue only based on their intensity. Figure 3 depicts a 3D volume rendering of the gray matter segmentation map when the MRF is used.

Just to show why we set some entries of G and H equal to each other as described in section 2.4, we have processed the same image once more but now without the additional constraint on G and H that classes that are known to share an interface should not favor one above the other. We started from the segmentation obtained with the constraint on G and H, re-estimated G and H from there without the constraint and performed the classification step until convergence while keeping the parameters fixed. As can be seen from figure 2 (e), the resulting segmentation now shows nicely distinct regions, but small details, such as small ridges of white matter, are lost. The MRF prior has over-regularized the segmentation and should therefore not be used in this form. By imposing the additional constraint on G and H, we only use the MRF to penalize combinations that are not possible, such as a gray matter voxel surrounded by voxels belonging to the class "other". This helps to clean up the segmentation maps of the brain tissues, but leaves fine details intact.

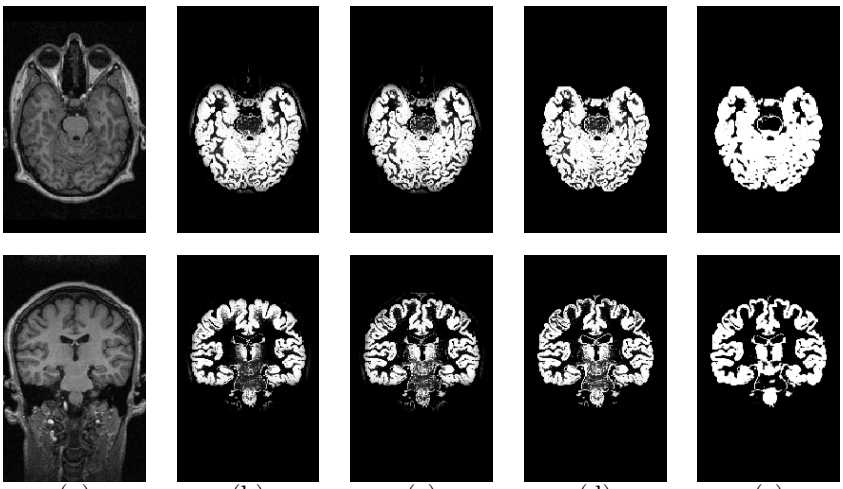
3 Validation

3.1 Simulated data

Although the algorithm should be validated on real MR data, a comprehensive validation is easier performed on simulated images, since the ground-truth is not known for in-vivo data. Furthermore, experiments with simulated data allow studying the influence of several imaging artifacts, such as noise and MR bias fields, separately.

Therefore, we have validated the EM algorithm on simulated MR images of the head (see figure 4) that were generated by the BrainWeb MR simulator [21] (see http://www.bic.mni.mcgill.ca/brainweb). We used images with an isotropic voxelsize of 1 millimeter and studied the influence of noise, field inhomogeneity and contrast (T1-, T2- or PD-weighted) by comparing the automatic segmentations with the known ground-truth. For this comparison, we made a hard final segmentation by assigning each voxel uniquely to the class where it most probably belongs to. Let V_{ae}^k denote the volume of the voxels that are assigned to a class k by both the ground truth and the automatic algorithm. Similarly, let V_a^k and V_e^k denote the volume of voxels assigned to class k by respectively the automatic algorithm and the ground-truth. We then measure the overlap between the automatic segmentation and the ground-truth for class k by $2V_{ae}^k/(V_a^k + V_e^k)$. This metric, proposed by Zijdenbos *et al.* [22], attains the value of 1 if both segmentations are in full agreement, and 0 if there is no overlap at all.

We first investigated the influence of noise by using images that were not corrupted with a bias field.



(a) (b) (c) (d) (e)

Figure 2: Example of how the different components of the algorithm work: (a) T1-weighted image; (b) gray matter segmentation without bias field correction and MRF; (c)gray matter segmentation with bias field correction but without MRF; (d) gray matter segmentation with bias field correction and MRF; (e) gray matter segmentation with bias field correction and MRF without constraints.

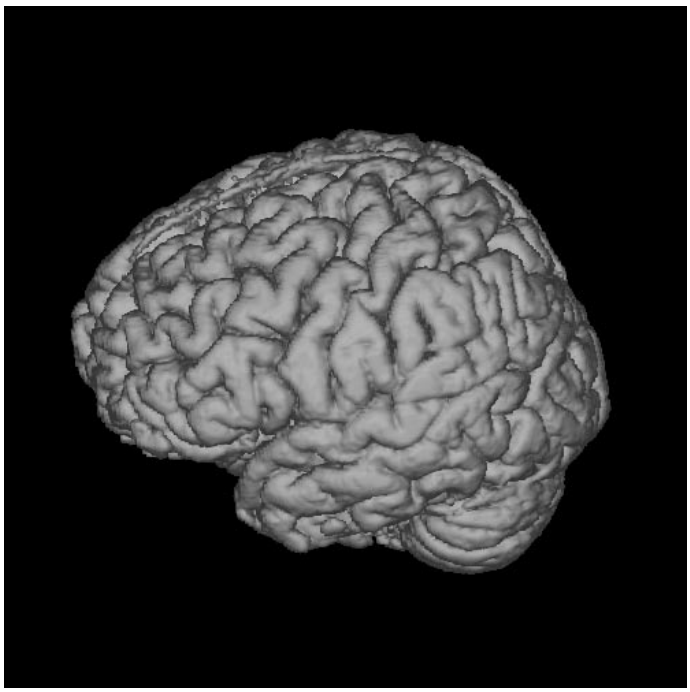


Figure 3: 3D volume rendering of the gray matter segmentation of the data of figure 2 with bias field correction and MRF

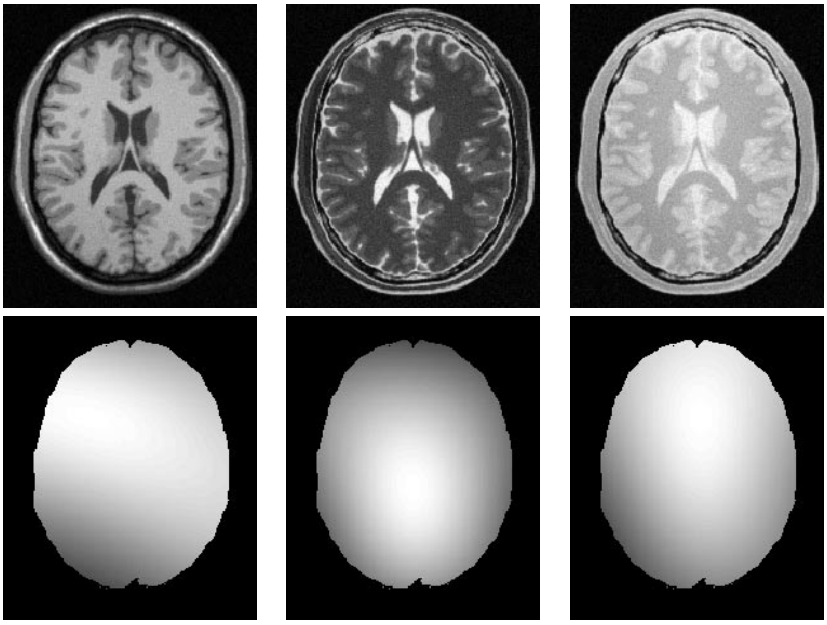


Figure 4: Top: simulated T1, T2 and PD-weighted MR images obtained from the BrainWeb simulator that were used to validate the segmentation results. Bottom: known bias fields imposed on the T1, T2 and PD-weighted images. The amplitude of the bias fields is exaggerated for visualization purposes.

We ran the EM-algorithm with the order of the bias field polynomial set to 0, that is no bias correction is performed, on T1-, T2- and PD-weighted images separately and combined, for noise levels ranging from 1% up to 9%. A noise level of 3% is considered typical, whereas 9% represents extreme conditions. The variation of the overlap metric for gray matter, white matter and total brain (both together) with the noise level is depicted in figure 5. As a general trend, the segmentation can be seen to deteriorate with increasing noise. Furthermore, the use of all three channels (T1, T2 and PD) together yielded in all cases better segmentations compared to the situation where only one channel was available for the algorithm.

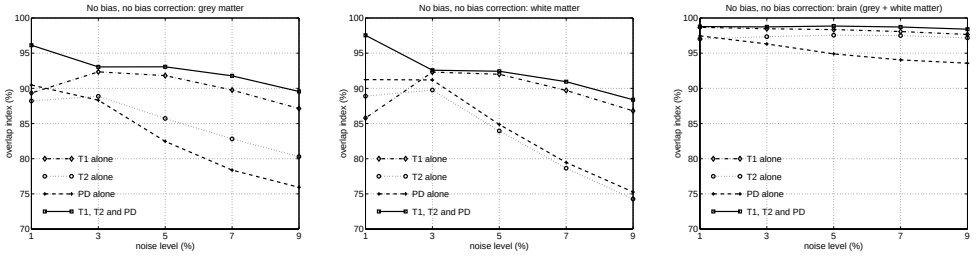


Figure 5: Overlap metric between the automatic segmentation and the known ground-truth for simulated MR images without bias field as a function of noise level for gray matter (left), white matter (middle) and total brain (right)

Surprisingly, however, the segmentation on single-channel data improves systematically when the noise level decreases from 9% to 3% but then deteriorates again for noise level 1%. Furthermore, the total brain volume is fairly accurately segmented in all cases, whereas the segmentation of gray and white matter is generally much worse. We therefore examined the classification maps more closely, and observed that both observations are caused by the same effect. Figure 6 depicts a representative slice of the T1 data set with 1% noise, and the exact location of misclassified white matter voxels in that slice for 1% and 3% noise. We concentrate on this data since the overlap metric drop from 3 to 1% was the most remarkable for white matter segmented from T1 data in figure 5.

As can be seen from figure 6, a one-voxel-thick interface between white and gray matter is consistently

misclassified in the case of 1% noise. These voxels are partial volume (PV) voxels, i.e. they have an intensity that lies somewhere in between the mean intensity of white and gray matter. In the absence of noise, the intensity variance of pure tissue voxels around their class mean is only determined by the simulated tissue characteristics. As is the case in real MR images, gray matter shows a larger variance than white matter in the simulated data. This causes the PV voxels, that do in fact not belong to neither white or gray matter but are really a mixture of both, to be classified to the class with the largest variance. Hence, these voxels are all consistently classified as gray matter. When noise is introduced, the difference in variance between white and gray matter decreases, which shifts their interface a little bit. Figure 6 shows this effect clearly: when the noise level is increased from 1% to 3%, the segmentation error is reduced. Adding more noise again deteriorates the segmentation since then pure tissue voxels start to be misclassified due to the noise.

Figure 6: Examination of misclassified voxels on the simulated data: T1 image (left), misclassified white matter voxels for 1% noise (middle), misclassified white matter voxels for 3% noise (right). The ground-truth white matter is shown in gray; misclassified white matter voxels are overlayed in bright color

To validate the performance of the algorithm on images that are corrupted by severe bias fields, we used simulated images with 40% field inhomogeneity. Figure 7 depicts the segmentations when the order of the bias field polynomial was set to 0, i.e. when no bias field correction was performed. The bias field clearly deteriorates the segmentations, especially when only one channel is available to the algorithm. When, on the contrary, the order of the bias field polynomial was set to 4, the segmentations are much better. This is shown in figure 8: comparing these results with the ones of figure 5 hardly reveals any differences, which shows that the algorithm corrects such MR field inhomogeneities well.

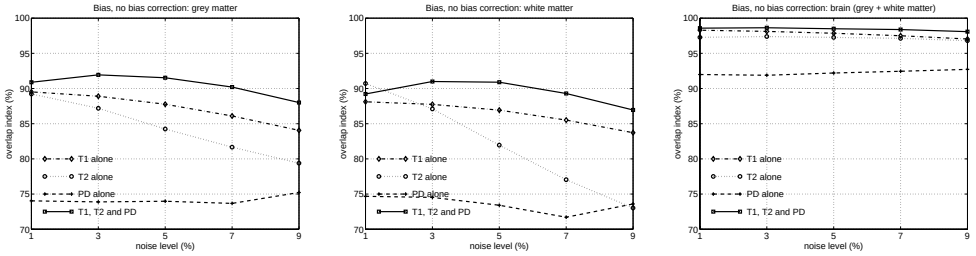


Figure 7: Overlap metric between the automatic segmentation without bias correction and the known ground-truth for simulated MR images with bias fields of 40% as a function of noise level for gray matter (left), white matter (middle) and total brain (right)

3.2 Hand-segmented data

We have validated the algorithm on real MR data as well, by comparing the automatic segmentations with manual tracings by a human expert. We analyzed 9 datasets that were acquired on a Signa 1.5T system, having 1.17×1.17 millimeter in-plane resolution and 124 1.2 millimeter thick slices acquired in the sagittal plane. These images are of the brains of children around the age of 10–16 years and

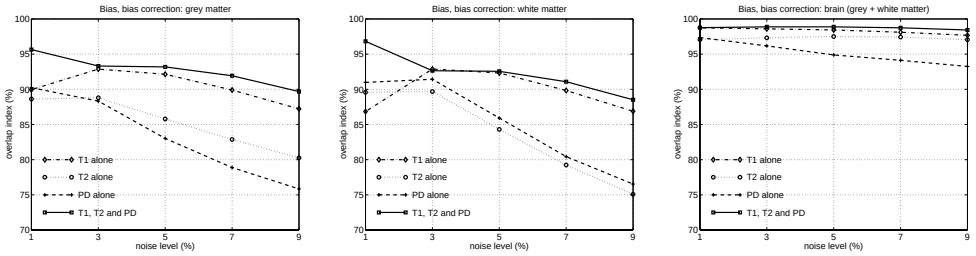


Figure 8: Overlap metric between the automatic segmentation with bias field correction and the known ground-truth for simulated MR images with bias fields of 40% as a function of noise level for gray matter (left), white matter (middle) and total brain (right)

Total brain (%)	Gray matter axial (%)	Gray matter coronal (%)
95.5	88.0	84.5
95.7	81.7	83.8
96.0	84.9	83.7
96.0	85.2	81.1
95.2	77.3	74.4
95.5	81.4	80.2
95.4	88.6	86.4
95.4	84.4	83.1
94.7	80.8	82.0

Table 1: Overlap metric between manual and automatic segmentation on 9 datasets of MR images of children. The total brain segmentation on the whole volume was compared, as well as gray matter segmentation of an axial and a coronal slice

were manually segmented by Robert T. Schultz, Child Study Center, Yale University, New Heaven, CT. For all 124 slices, the whole brain outer surface was traced, although no attempt was made to carefully delineate all sulci. For 2 slices out of each dataset, one axial and one coronal, the gray-white and gray-csf boundaries were carefully traced, trying to precisely delineate the sulci.

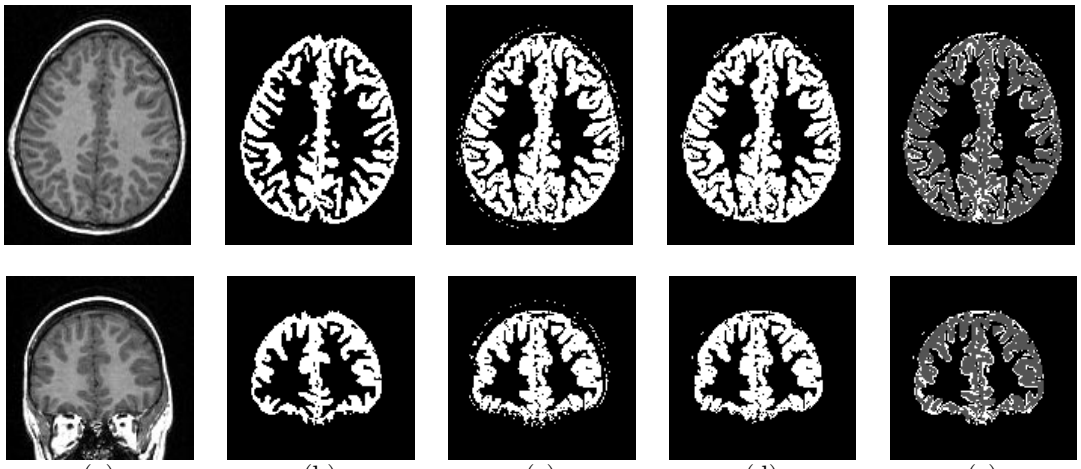
We compared the manual tracings with the automatic segmentations by calculating the overlap metric that was also used in section 3.1. Table 1 contains the result on each dataset for the total brain, and for gray matter in the coronal and axial slice that was chosen for manual delineation.

By far most of the misclassifications on the total brain are caused by the more detailed segmentation of the gray matter-csf interface by the automatic method compared to the manual tracing. As was the case with the simulated data, the gray-white matter segmentation is clearly less accurate than the total brain segmentation. Figure 9 depicts the exact location of misclassified gray matter voxels for the coronal and axial manually traced slices of a representative dataset. It can be seen that the automatic algorithm follows more accurately the gray matter-csf interface than the manual tracer. Some tissue surrounding the brain is still misclassified as gray matter, although this error is already reduced compared to the situation where no MRF prior is used. However, by far most misclassifications are due to the classification of gray-white matter partial volume voxels to gray matter by the automated method. The human observer has segmented white matter consistently as a thicker structure than the automatic algorithm.

4 Discussion

We have described and validated a fully automated model-based method for segmenting brain tissues from MR images. The algorithm interleaves voxel classification, intensity distribution parameter estimation, MR bias field correction and MRF parameter estimation until convergence is detected.

Due to the fact that the in-plane resolution is generally different than the slice thickness in MR



(a) (b) (c) (d) (e)

Figure 9: Examination of voxels that are differently classified by the human tracer than by the automated method on a representative dataset. From left to right: axial and coronal slice that was carefully manually segmented (a), corresponding manual segmentation of gray matter (b), automatic segmentation of gray matter without MRF prior (c), automatic segmentation of gray matter with MRF (d), difference between manual and automatic segmentation with MRF shown in bright color (e). The overlap metric is 84.4 % for the axial and 83.1 % for the coronal slice

images, we have modeled in-plane class transition probabilities with different parameters G than between-slice class transition probabilities parameterized by H. The in-plane resolution is typically about 1×1 millimeter; the slice thickness, however, can largely vary. We therefore re-estimate G and H for every image separately. It should be noted, however, that inclusion of the MRF parameter estimation step roughly doubles computation time compared to the situation where the MRF parameters are not calculated. H could therefore be set to zero and G could be precalculated just once on a normal brain dataset; those parameters could then be used for the segmentation of all following datasets, which would reduce calculation time. However, this approach does not make use of the full 3D nature of the MR images and can therefore be expected to yield less powerful discrimination between brain and non-brain tissues. On the contrary, if a large amount of images with equal voxel sizes have to be segmented, both G and H can be precalculated on one image and applied to all other images of the dataset. Although we have not validated this approach, we expect this to speed up the process without loss of accuracy.

A problem that showed up during the validation for both simulated and real MR data, was the partial volume (PV) effect. Whereas the model we used assumes that each voxel in the image belongs to only one single class, voxels that lie on the border between different tissue types violate this assumption. In reality, these voxels are really a mixture of tissues and every segmentation method that tries to assign them exclusively to one class is condemned to fail. The problem is especially important in images of the brain since the interface between gray and white matter is highly complex, which results in a high volume of PV voxels compared to the volume of pure tissue voxels. Misclassification of this thin interface gives therefore immediate rise to considerable segmentation errors.

Ideally, the model should be adjusted so as to model a mixture of tissue types in each voxel rather than a single tissue type only. In the literature, a number of attempts have been made in this direction. Choi *et al.* [23] introduced the concept of so-called *mixels*, but they had to introduce heuristics in order to arrive at a workable solution. More recently, Laidlaw *et al.* [24] have applied the intensity distribution of partial volume mixtures of two tissues derived by Santago and Gage [25] to the segmentation problem. However, the time figures they sketch are far too high for the method to be practically useful. Further research is therefore needed in this direction.

Often one is not only interested in segmentations of the total white and gray matter, but also in anatomical substructures of the brain. Examples include the separation of left and right hemisphere in the study of brain asymmetry associated with schizophrenia. While the segmentation of such anatomically defined structures are typically performed by matching a labeled atlas to the study image using a non-

rigid registration technique, such methods have difficulty in segmenting such highly variable structures as the white-gray matter interface. We are currently investigating the use of Thirion's demons [26] technique to subdivide the segmentation maps generated by the EM algorithm described above into anatomical substructures. Such a non-rigid registration method can also be used to generate a new atlas by averaging the segmentations of a number of normal subjects after non-rigid matching. Compared to the atlas that we use at this moment, that was generated by averaging segmentations after affine normalization, the new a priori maps for gray matter, white matter and csf would be much sharper. We expect this to lead to a better fitted brain mask and, as a result, to improved segmentations.

We are currently adapting the algorithm so that it can be applied to fully automated segmentation of Multiple Sclerosis (MS) lesions in the brain, with promising results. We are in the process of applying this technique to a large number of scans for analyzing the time evolution of MS lesions during drug treatment.

5 Conclusions

We have presented a fully-automated model-based method for tissue classification of MR images of the brain. The algorithm interleaves classification with MR bias field correction, intensity distribution estimation and estimation of Markov Random Field parameters. We use a digital brain atlas containing information about the expected a priori location of tissue types to initialize the algorithm. This yields a fully-automated method for tissue classification that produces objective and reproducible results.

The use of Markov Random Fields helps in discriminating between brain and non-brain tissues. Validation on simulated and real MR images revealed the benefit of multi-channel MR data over single-channel data, and the necessity to correct for MR bias fields. The partial volume effect needs further investigation.

Acknowledgments

This work was supported by the EC-funded BIOMORPH project 95-0845, a collaboration between the Universities of Kent and Oxford (UK), ETH Zürich (Switzerland), INRIA Sophia Antipolis (France) and KU Leuven (Belgium), by the research grant KV/E/197 (Elastische Registratie voor Medische Toepassingen) of the FWO-Vlaanderen to Dirk Vandermeulen, by a grant for research specialization from the Flemish Institute for stimulation of the scientific-technological research in the industry (IWT), by the Research Fund KU Leuven GOA/99/05 (Variability in Human Shape and Speech) and by the EC-funded BIOMED-2 BMH4-CT98-6048 (QAMRIC) project, a collaboration between the Universities of Kent and Edinburgh (UK), INRIA Sophia Antipolis (France), Georg-August-Universität Göttingen (Germany) and KU Leuven (Belgium). The authors would like to thank Robert T. Schultz and James S. Duncan from Yale University for kindly providing the hand-segmented data, and wish to acknowledge the contribution of Xiaolan Zeng.

Appendix A

We here derive the EM-algorithm for the independent model of section 2.1. The log-likelihood for the complete data $q = (y, z)$ can be written as

$$\begin{aligned} \log f(q \mid \Phi_y) &= \log f(y, z \mid \Phi_y) \\ &= \log f(y \mid z, \Phi_y) + \log f(z) \\ &= \sum_{i=1}^{n} \log f(y_i \mid z_i, \Phi_y) + \sum_{i=1}^{n} \log f(z_i) \end{aligned}$$

Since z_i is a vector with all zero components except for a single component that is unity, we have

$$\log f(y_i \mid z_i, \Phi_y) = z_i^t U(y_i \mid \Phi_y)$$

and

$$\log f(z_i) = z_i^t V$$

where $U(y_i \mid \Phi_y)$ is a vector that has as kth component $\log f(y_i \mid z_i = e_k, \Phi_y)$ and similar V has a kth component $\log f(z_i = e_k)$. Therefore, we have

$$\begin{aligned} Q(\Phi_y \mid \Phi_y^{(m)}) &= E[\log f(q \mid \Phi_y) \mid y, \Phi_y^{(m)}] \\ &= \sum_{i=1}^{n} E[z_i^t \mid y, \Phi_y^{(m)}] U(y_i \mid \Phi_y) + \\ &\quad \sum_{i=1}^{n} E[z_i^t \mid y, \Phi_y^{(m)}] V \end{aligned}$$

Define $p_i^{(m+1)} \equiv E[z_i \mid y_i, \Phi_y^{(m)}]$ as the estimation of the hidden data z_i in voxel i based on its intensity y_i and the current parameter estimate $\Phi_y^{(m)}$. It can then be seen that

$$[p_i^{(m+1)}]_k = f(z_i = e_k \mid y_i, \Phi_i^{(m)})$$

which results in equation 1 after application of Bayes' rule. Explicit maximization of $Q(\Phi_y \mid \Phi_y^{(m)})$ to Φ_y yields

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^{n} p_{ik}^{(m+1)} \left(y_i - \sum_{j=1}^{J} c_j^{(m+1)} \phi_j(x_i)\right)}{\sum_{i=1}^{n} p_{ik}^{(m+1)}} \tag{9}$$

$$(\sigma_k^{(m+1)})^2 = \frac{\sum_{i=1}^{n} p_{ik}^{(m+1)} \left(y_i - \mu_k^{(m+1)} - \sum_{j=1}^{J} c_j^{(m+1)} \phi_j(x_i)\right)^2}{\sum_{i=1}^{n} p_{ik}^{(m+1)}} \tag{10}$$

for the distribution parameters $\mu_k^{(m+1)}$ and $\sigma_k^{(m+1)}$, and equation 4 for the bias field parameters $c_j^{(m+1)}$ [4, 5]. Notice that the equations 9 and 10 and equation 4 are coupled. We therefore use the estimation $c_j^{(m)}$ of the previous iteration instead in equation 9 and 10, which explains equation 2 and equation 3. Alternatively, $\mu_k^{(m)}$ and $\sigma_k^{(m)}$ could also be used instead in equation 4. The result is a so-called generalized EM algorithm, i.e. the M-step only ensures that $Q(\Phi_y^{(m+1)} \mid \Phi_y^{(m)}) > Q(\Phi_y^{(m)} \mid \Phi_y^{(m)})$.

Appendix B

We here derive the equations for the parameters Φ_y when the MRF is added. There are now two parameter sets Φ_y and Φ_z. Denoting $\Phi = \{\Phi_y, \Phi_z\}$, the log-likelihood for the complete data $q = (y, z)$ now becomes

$$\log f(q \mid \Phi) = \sum_{i=1}^{n} \log f(y_i \mid z_i, \Phi_y) - U_{mrf}(z \mid \Phi_z) - \log Z(\Phi_z)$$

and hence

$$\begin{aligned} Q(\Phi \mid \Phi^{(m)}) &= \sum_{i=1}^{n} E[z_i^t \mid y, \Phi^{(m)}] U(y_i \mid \Phi_y) \\ &\quad -E[U_{mrf}(z \mid \Phi_z) + \log Z(\Phi_z) \mid y, \Phi^{(m)}] \end{aligned}$$

The parameters Φ_y can now be calculated in the same way as with the independent model. However, $p_i^{(m+1)} = E[z_i \mid y, \Phi^{(m)}]$ can no longer been calculated with equation 1 since the z_i are not independent and, as a result, it is difficult to obtain the exact solution. By definition,

$$\begin{aligned} E[z_i \mid y, \Phi^{(m)}] &= \sum_z z_i f(z \mid y, \Phi^{(m)}) \\ &= \frac{\sum_z z_i f(y \mid z, \Phi_y^{(m)}) f(z \mid \Phi_z^{(m)})}{f(y \mid \Phi^{(m)})} \end{aligned}$$

However, due to the interaction between the z_i's, the above would involve calculation of all the possible realizations of the MRF. Therefore, exact calculation of $p_i^{(m+1)}$ is not computationally feasible, and an approximate technique must be used.

We here adopt an approximation that was proposed by Zhang *et al.* [9] [10], based on the *mean field theory* from statistical mechanics. This mean field approach suggests an approximation to $p_i^{(m+1)}$ based on the assumption that the influence of $z_j, j \neq i$ in the calculation of $p_i^{(m+1)}$ can be approximated by the influence of $p_j^{(m)}$ from the previous iteration. This explains equation 5. Once $p^{(m+1)}$ is known, the equations for the parameters $\Phi_y^{(m+1)}$ can be derived in exactly the same way as with the independent model, and therefore, equations 2, 3 and 4 remain valid.

Appendix C

We show how equation 7 can be obtained. Using θ and v_{z_i,g_i,h_i} as defined in section 2.2, the following holds:

$$\begin{aligned} U_{mrf}(z_i \mid z_{\mathcal{N}_i}, \Phi_z) &= z_i^t G g_i + z_i^t H h_i \\ &= v_{z_i,g_i,h_i}^t \theta \end{aligned} \tag{11}$$

The conditional likelihood $f(z_i \mid z_{\mathcal{N}_i}, \Phi_z)$ is related to the potential $U_{mrf}(z_i \mid z_{\mathcal{N}_i}, \Phi_z)$ by

$$f(z_i \mid z_{\mathcal{N}_i}, \Phi_z) = \frac{e^{-U_{mrf}(z_i|z_{\mathcal{N}_i}, \Phi_z)}}{\sum_{z_i'} e^{-U_{mrf}(z_i'|z_{\mathcal{N}_i}, \Phi_z)}}$$

Furthermore,

$$f(z_i \mid z_{\mathcal{N}_i}, \Phi_z) = f(z_i \mid g_i, h_i, \Phi_z) = \frac{f(z_i, g_i, h_i \mid \Phi_z)}{f(g_i, h_i, \mid \Phi_z)}$$

so that

$$\frac{e^{-U_{mrf}(z_i|z_{\mathcal{N}_i}, \Phi_z)}}{f(z_i, g_i, h_i \mid \Phi_z)} = \frac{\sum_{z_i'} e^{-U_{mrf}(z_i'|z_{\mathcal{N}_i}, \Phi_z)}}{f(g_i, h_i \mid \Phi_z)}$$

Since the RHS of the above is independent of the value of z_i, so is the LHS and therefore

$$\frac{e^{-U_{mrf}(z_i|z_{\mathcal{N}_i}, \Phi_z)}}{f(z_i, g_i, h_i \mid \Phi_z)} = \frac{e^{-U_{mrf}(z_i'|z_{\mathcal{N}_i}, \Phi_z)}}{f(z_i', g_i, h_i \mid \Phi_z)}$$

for any two distinct values z_i and z_i'. Rearranging the above and using equation 11, we obtain equation 7.

References

[1] W.M. Wells, III, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, August 1996.

[2] K. Held, E. R. Kops, B. J. Krause, W. M. Wells III, R. Kikinis, and H. W. Müller-Gärtner. Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 16(6):878–886, december 1997.

[3] Tina Kapur, W. Eric L. Grimson, Ron Kikinis, and William M. Wells. Enhanced spatial priors for segmentation of magnetic resonance imaging. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 457–468. Springer, 1998.

[4] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automatic segmentation of brain tissues and MR bias field correction using a digital brain atlas. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 1222–1229. Springer, 1998.

[5] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automated bias field correction and tissue classification of MR images of the brain using a digital atlas. Technical Report KUL/ESAT/PSI/9901, KU Leuven, february 1999. submitted to IEEE Transactions on Medical Imaging (november 1998).

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

[7] C. F. Jeff Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

[8] John G. Sled and G. Bruce Pike. Understanding intensity non-uniformity in MRI. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 614–622. Springer, 1998.

[9] Jun Zhang. The mean-field theory in EM procedures for markov random fields. *IEEE Transactions on Signal Processing*, 40(10):2570–2583, october 1992.

[10] David A. Langan, Karl J. Molnar, James W. Modestino, and Jun Zhang. Use of the mean-field approximation in an EM-based approach to unsupervised stochastic model-based image segmentation. In *Proceedings of ICASSP'92*, volume 3, pages 57–60, San Fransisco, CA, march 1992.

[11] S.Z. Li. *Markov Random Field Modeling in Computer Vision.* Computer Science Workbench. Springer, 1995.

[12] John Ashburner, Karl Friston, Andrew Holmes, and Jean-Baptiste Poline. *Statistical Parametric Mapping.* The Wellcome Department of Cognitive Neurology, University College London.

[13] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi-modality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.

[14] Benoit M. Dawant, Alex P. Zijdenbos, and Richard A. Margolin. Correction of intensity variations in MR images for computer-aided tissue classification. *IEEE Transactions on Medical Imaging*, 12(4):770–781, December 1993.

[15] John G. Sled, Alex P. Zijdenbos, and Alan C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, February 1998.

[16] Sébastien Gilles, Michael Brady, Jérome Declerck, Jean-Philippe Thirion, and Nicholas Ayache. Bias field correction of breast MR images. In *Proc. Visualization in Biomedical Computing (VBC '96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 153–158. Springer, 1996.

[17] M. Tincher, C.R. Meyer, R. Gupta, and D.M. Williams. Polynomial modeling and reduction of RF body coil spatial inhomogeneity in MRI. *IEEE Transactions on Medical Imaging*, 12(2):361–365, June 1993.

[18] C.R. Meyer, P.H. Bland, and J. Pipe. Retrospective correction of MRI amplitude inhomogeneities. In N. Ayache, editor, *Proc. First International Conference on Computer Vision, Virtual Reality, and Robotics in Medicine (CVRMED'95)*, volume 905 of *Lecture Notes in Computer Science*, pages 513–522. Springer, Nice, France, April 1995.

[19] C. Brechbuler, G. Gerig, and G. Szekely. Compensation of spatial inhomogeneity in MRI based on a parametric bias estimate. In *Proc. Visualization in Biomedical Computing (VBC '96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 141–146. Springer, 1996.

[20] The MathWorks Inc., 24 Prime Park Way, Natick, MA, USA. *Matlab.*

[21] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans. BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4 part 2/4 S425), 1997. accessible by WWW: http://www.bic.mni.mcgill.ca/brainweb/.

[22] A. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724, december 1994.

[23] Hwan Soo Choi, David R. Haynor, and Yongmin Kim. Partial volume tissue classification of multichannel magnetic resonance images–a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407, september 1991.

[24] David H. Laidlaw, Kurt W. Fleischer, and Alan H. Barr. Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Transactions on Medical Imaging*, 17(1):74–86, february 1998.

[25] Peter Santago and H. Donald Gage. Statistical models of partial volume effect. *IEEE Transactions on Image Processing*, 4(11):1531–1540, november 1995.

[26] J.-P. Thirion. Non-rigid matching using demons. In *Computer Vision and Pattern Recognition, CVPR'96*, June 1996.