Fast and robust parameter estimation for statistical partial volume models in brain MRI

Jussi Tohka,[a,*] Alex Zijdenbos,[b] and Alan Evans[b]

[a]Digital Media Institute / Signal Processing, Tampere University of Technology, FIN-33101, Finland
[b]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

Due to the finite spatial resolution of imaging devices, a single voxel in a medical image may be composed of mixture of tissue types, an effect known as partial volume effect (PVE). Partial volume estimation, that is, the estimation of the amount of each tissue type within each voxel, has received considerable interest in recent years. Much of this work has been focused on the mixel model, a statistical model of PVE. We propose a novel trimmed minimum covariance determinant (TMCD) method for the estimation of the parameters of the mixel PVE model. In this method, each voxel is first labeled according to the most dominant tissue type. Voxels that are prone to PVE are removed from this labeled set, following which robust location estimators with high breakdown points are used to estimate the mean and the covariance of each tissue class. Comparisons between different methods for parameter estimation based on classified images as well as expectation–maximization-like (EM-like) procedure for simultaneous parameter and partial volume estimation are reported. The robust estimators based on a pruned classification as presented here are shown to perform well even if the initial classification is of poor quality. The results obtained are comparable to those obtained using the EM-like procedure, but require considerably less computation time. Segmentation results of real data based on partial volume estimation are also reported. In addition to considering the parameter estimation problem, we discuss differences between different approximations to the complete mixel model. In summary, the proposed TMCD method allows for the accurate, robust, and efficient estimation of partial volume model parameters, which is crucial to a variety of brain MRI data analysis procedures such as the accurate estimation of tissue volumes and the accurate delineation of the cortical surface.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Partial volume effect; Segmentation; Robust estimation; Mixel model

Introduction

The quantitative analysis of magnetic resonance (MR) images in the study of human brain anatomy is becoming more and more important. For example, a range of brain disorders as well as brain development and healthy aging can cause structural changes in the brain. These changes can be quantified by measuring volumes or other properties of anatomical structures of interest providing information, for example, on disease severity. Before measurements can be performed, the structures of interest must be extracted from the image data. This often includes the labeling of voxels according to their tissue type. This labeling, or classification, can be performed based on a single MR image or based on a multispectral image constructed by combining series images of the same subject acquired with different pulse sequence parameters. Typically, the tissue types of interest are white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), but also more anatomical labels can be assigned to image voxels (Collins et al., 1999; Fischl et al., 2002). However, voxel classification with three basic tissue types has itself rather direct applications such as quantifying disease burden in multiple sclerosis by estimation of the amount of brain atrophy (Collins et al., 2001). Moreover, many procedures aiming at the extraction of particular brain structures, such as cerebral cortex, can gain from the initial tissue classification (MacDonald et al., 2000; Xu et al., 1999; Zeng et al., 1999) as can partial volume correction in positron mission tomography (Rousset et al., 1998).

Because of the finite resolution of the imaging devices, a single voxel may contain several tissue types. This is known as partial volume effect (PVE). Due to PVE, the classification of a voxel reflecting the dominant tissue type (WM, GM, or CSF), does not reveal all possible information about the tissue content of that voxel. This can be problematic in small structures or highly convoluted areas of the brain. For example, algorithms aimed at extraction of the cortical surface often omit entire sulci due to the PVE between the thin ribbon of sulcal CSF and the surrounding gray matter. These problems are especially serious when pediatric subjects are considered as illustrated in Fig. 1. Estimation of the amount of each tissue type present in each voxel, that is, partial volume estimation, provides an interesting possibility to improve the accuracy of cortical surface extraction (cf. Fig. 1). Other applications that gain from modeling of PVE have been as well considered within brain MRI. Santago and Gage (1993) apply information about partial volume voxels to improve tissue quantification. González Ballester et al. (2002) study the asymmetry of temporal horns taking PVE into account, and in an earlier work (González Ballester et al., 2000), they suggest that PVE and

* Corresponding author. Digital Media Institute/Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101, Finland. Fax: +358-3-31153087.

E-mail address: jussi.tohka@tut.fi (J. Tohka).

Available online on ScienceDirect (www.sciencedirect.com).

Discrete Classification   CSF Partial Volume Estimation   Overlaid CSF PVE and Discrete Classification
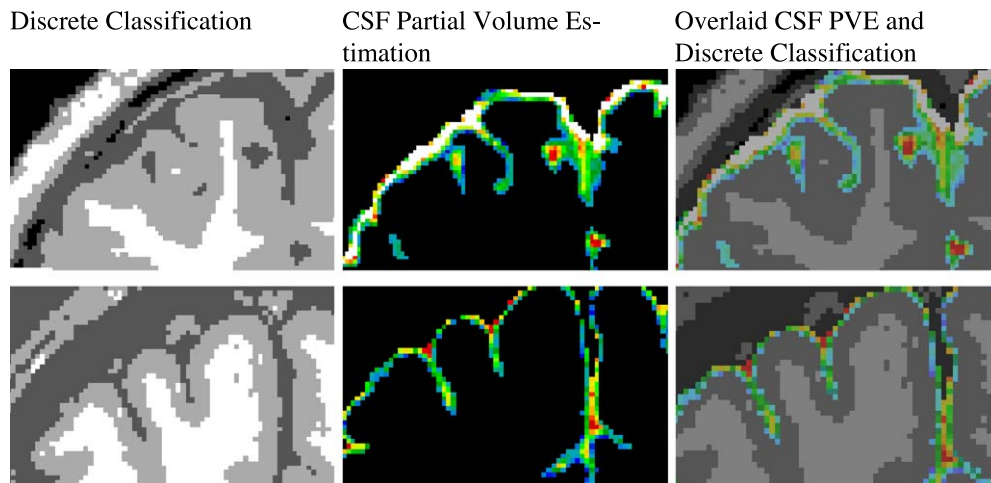
Fig. 1. PVE in the cortical region, in the top row, a pediatric subject, and in the bottom row, a middle aged subject. Pediatric brain sulci are more compacted with less (or no) intrasulcal CSF in the segmented image (see left column). However, the partial volume skeleton, estimated automatically by using the method presented (TMCD) in this study, can penetrate down into that intrasulcal CSF (see middle column) for the pediatric case (top row). Different colors correspond to different amounts of CSF within the voxels. Consequently, partial volume estimation allows, for example, for more accurate cortical surface extraction than that could be achieved by methods relying on discrete classification.

discrete sampling at boundary locations can lead to volume measurement errors in the range 20–60%.

Partial volume effect and PV estimation have been addressed in various ways in MR imaging literature. For example, Pham and Prince (1999) have proposed a fuzzy C-means algorithm. They have also studied the relationship between the fuzzy C-means objective function and statistical models of PVE in the simplified case (only two type is of tissues and single-spectral data), showing that in this case, these two approaches can be considered equivalent (Pham and Prince, 1998). However, these considerations do not extend to more realistic situations where one would have more than two tissue types and where the data would be multispectral.

Wang et al. (2001) propose to use a Bayesian classifier with a variable number of tissue classes, including classes of mixed tissue types. However, as the authors model the image histogram with a finite mixture of normal distributions and parameters (means and variances) for mixed tissue classes, which are independent of the parameters of the related pure tissue classes, there is no explicit model of the PVE. While the procedure may be reasonable for detection of voxels containing PVE, estimation of the mixing proportions can be challenging.

The most commonly used, statistically based model of PVE is the mixel model proposed by Choi et al. (1991). (A similar model was proposed earlier by Kent and Mardia, 1988, but without consideration of medical imaging applications.) This approach assumes that each intensity value in the image is a realization of a weighted sum of random variables (RVs), each of which characterizes a pure tissue type. We call these weighting factors *partial volume coefficients* (PVCs). The method involves maximum-likelihood estimation of the PVCs for each voxel that model PV fractions of pure tissue types. Some authors have studied the identification of voxels containing PVE based on the mixel or a closely related model without trying to estimate the PVCs for each voxel (Laidlaw et al., 1998; Ruan et al., 2000; Santago and Gage, 1993). Our interest in this study is in estimating PVCs and not in merely identifying voxels containing PVE.

Before statistical PV estimation can be performed, the probability density functions (pdfs) of the RVs describing pure tissue

types must be specified. In practice, this typically means that the parameters of the pdfs—usually normal distributions—need to be estimated. Unfortunately, errors in the parameter estimation often have a major impact on the quality of PV estimation. In the case of statistical "hard" classification, where each voxel is classified according to the most dominant tissue type, the parameter estimation problem differs from the one studied here as each intensity is a realization of a single RV. On the other hand, the mixel model assumes that each voxel's intensity represents a weighted sum of several RVs and aims to estimate these unknown weight parameters.

In general, there are three approaches to the parameter estimation problem: histogram analysis (Santago and Gage, 1993), simultaneous parameter, and partial volume estimation by expectation maximization (EM)-like algorithms (Noe and Gee, 2001), and estimation based on a hard segmentation of the image (Shattuck et al., 2001). These three approaches, however, each have their drawbacks. Histogram analysis requires a mixture probability density to be fit to an image histogram by parameter optimization. This involves finding the minimizer of a multimodal objective function and therefore reliability of histogram analysis for parameter estimation depends heavily on the optimization algorithm used for the fitting task. If a standard nonlinear optimization algorithm aimed at local minimization (e.g., Levenberg–Marquee algorithm used commonly for curve fitting) is used, the initialization for the algorithm has to be chosen carefully to avoid convergence to a poor local minimum. These considerations call for the use of advanced global optimization algorithms, for example, Santago and Gage (1993) propose to use the tree-annealing method (Bilbro and Snyder, 1991). The problem with global optimization methods is that they are usually far more time consuming than local optimization methods.

Like histogram analysis, expectation–maximization schemes for parameter estimation are time consuming. Besides, the use of spatial information in the form of Markov Random Fields (Besag, 1974, 1986) causes practical problems with the E-step of the EM-algorithm (Van Leemput et al., 2003). To solve this, statistical dependency between voxel labels can be ignored during the E-step

as in Noe and Gee (2001). However, this leads to an algorithm that is merely a heuristic and does not necessarily share the convergence properties of the original EM-algorithm (Dempster et al., 1977). An elegant solution to this problem based on a Monte Carlo EM-algorithm (Wei and Tanner, 1990) was recently proposed by Van Leemput et al. (2003). The algorithm was implemented only for 2D image slices and the authors reported that partial volume estimation typically requires about 20 min for a single slice of an MR image with a very fast 1.7 GHz processor. Hence, the total time consumption for an (moderately sized) image of 100 slices would be over 30 h. The authors claimed that the computation time can be significantly reduced but did not offer any figures to support this claim.

Parameter estimation based on a hard labeling can be computationally efficient due to the prior knowledge of labels of voxels that can be utilized in parameter estimation. However, due to the PVE and classification errors in the hard labeling, each class in the hard segmented image contains a large number of outliers. This fact is taken into the account in Shattuck et al. (2001), but their approach for the parameter estimation applies only to single-spectral images and assumes that each tissue type has a Gaussian noise distribution with the same variance. Also, since this method involves detecting modes of histograms, it is sensitive to noise and the estimates may not be unique.

In this study, using minimum volume ellipsoid and minimum covariance determinant estimators (Rousseeuw, 1984; Rousseeuw and Leroy, 1987), we propose routines for parameter estimation based on segmented images that are well-defined and can be used in single-spectral as well as multispectral cases. The studied estimators are robust; in other words, they tolerate deviations from the parametric form of the distribution assumed for the data. This is important because in our case, part of the data is good (correctly classified voxels of pure tissue), but the data contain also outliers (e.g. PV voxels). There are fast algorithms for computing the estimators and hence increasing the robustness of the parameter estimation does not lead to significantly increased running time for the PV estimation. We compare different techniques of parameter estimation using simulated MR data (Kwan et al., 1999) and demonstrate how the errors in the parameter estimation affect the final PV estimation results. The results are also compared against those obtained from an EM-like method, similar to Noe and Gee (2001), but with an advanced initialization technique. Furthermore, we compare our results with a fast PV estimation technique proposed by Shattuck et al. (2001) and demonstrate that our technique can yield clear improvements in the accuracy of PV estimates without considerable loss in time efficiency. As a secondary contribution, we consider differences between material and sampling noise models for PVE also on a more theoretical level (cf. Santago and Gage, 1995).

Methods

Statistical model for the partial volume effect

In this section, we state the PVE parameter estimation problem and describe the mixel model more formally. In the following, random variables (RVs) are denoted by boldface letters, while both scalars and vectors are shown in italics. Let us denote the observed image by $X = \{x_i: i = 1,\ldots, N\}$, with $x_i \in \mathbb{R}^K$, and K the number of data channels. Let the set of

possible tissue types present in the image be $\mathcal{L} = \{1,\ldots, M\}$. Moreover, $\mathbf{l}_j$ is the RV describing the tissue type j and the pdf of $\mathbf{l}_j$ is Gaussian $g(\cdot \mid \mu_j, \Sigma_j)$ with mean μ_j and covariance Σ_j. Each x_i is now a realization of the RV

$$\boldsymbol{x}_i = \sum_{j=1}^{M} w_{i_j}\boldsymbol{l}_j + \boldsymbol{a} \tag{1}$$

where $\mathbf{a}$ denotes white Gaussian noise and partial volume coefficients (PVCs) $w_{i_j} \in [0,1]$ for all i, j, and $\sum_{j=1}^{M} w_{i_j} = 1$ for all i. A partial volume context image is denoted by $W = \{w_{i_j}: i = 1,\ldots N, j = 1,\ldots, M\}$. The problem is now to find an estimate W^* for the true partial volume context image, given the observed image X. This is similar to the fuzzy classification problem, but in this case, the coefficients w_{i_j} specifically model the fraction of tissue type j present in the voxel i, as opposed to the likelihood that voxel i is of type j for example. For notational convenience, we set we $w_i = [w_{i_1},\ldots, w_{i_M}]^T$.

The model in Eq. (1) for the formation of image intensities within MRI is a simplification. Perhaps most importantly, MR images are known to contain low frequency spatial intensity variations often called RF inhomogeneity or shading artifact. The correction for this artifact can be assumed to be performed before PV estimation and there exist several methods for the task (e.g., Sled et al., 1998; Wells et al., 1996). Another somewhat controversial issue of the model is the assumed Gaussian distributions for tissue classes and for the noise component. The assumption is common within classifiers and PV estimation routines applied to brain MR images.

Since there are no methods that would allow us to estimate parameters for $\mathbf{l}_j$, $j = 1,\ldots, M$, and the parameters for the noise variable $\mathbf{a}$ at the same time, Eq. (1) needs to be simplified. Using the terminology from Santago and Gage (1995), *the material-dependent noise model* is obtained by embedding the sampling noise component $\mathbf{a}$ into RVs $\mathbf{l}_j$ describing the tissue types. Accordingly, *the sampling noise model* is obtained by ignoring variations within tissue types, that is, each tissue type is described by a deterministic but initially unknown intensity value as opposed to describing tissue types by RVs. Our attention is centered on the material-dependent noise model, because, for our topic, it can be considered to be a generalization of the sampling noise model. Indeed, the methods for the parameter estimation to be proposed are as well applicable for the sampling noise case, where only a single covariance matrix describing the sampling noise is required. This issue can be solved by considering the pooled covariance of the estimated covariance matrices for the tissue classes (Dougherty, 1990, p. 467).

Algorithm for statistical partial volume estimation

Problem statement

The maximum a posteriori (MAP) criterion is used to find W^* given the observed image, that is,

$$W^* = \arg\max_{W} p(W \mid X). \tag{2}$$

We make the assumption that the observed image intensity value x_i of voxel i depends only on the RV $\mathbf{x}_i$ related to the same voxel. In other words, we assume that the intensity at voxel i does

not depend on the tissue content of the other voxels. Hence, for the material-dependent noise model,

$$p(W\,|\,X) \propto p(W)p(X\,|\,W) = p(W)\prod_{i=1}^{N}p(x_i\,|\,w_i),\qquad(3)$$

$$p(x_i\,|\,w_i) = g\!\left(x_i\,\middle|\,\sum_{j=1}^{M}w_{ij}\mu_j, \sum_{j=1}^{M}w_{ij}^2\Sigma_j\right),\qquad(4)$$

where

$$g(x\,|\,\mu,\Sigma) = \frac{1}{(2\pi)^{K/2}\det(\Sigma)^{1/2}}\exp\!\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

denotes a multivariate Gaussian pdf with the mean μ and the covariance Σ. The expression for the likelihood term (4) is derived using the fact that a weighted sum of Gaussian RVs is a Gaussian RV with the mean and the covariance as in Eq. (4), (cf., e.g., Dougherty, 1990, Theorem 5.9). The prior probability $p(W)$ can be ignored or modeled by a Markov Random Field (MRF) (Besag, 1974, 1986; Geman and Geman, 1984). However, in MR images, labels w_i of nearby voxels are typically correlated. Ignoring these correlations would lead to loss of relevant information, and hence MRF modeling of the prior probability is preferable. This has been also confirmed empirically in some studies (Shattuck et al., 2001; Van Leemput et al., 2003).

The estimation problem (2) is challenging because both the PVCs w_i and the model parameters μ_j, Σ_j are initially unknown. Indeed, even if the model parameters would have been identified, the estimation problem (2) would be under constrained and could not be solved directly if the number of data channels $K < M - 1$ (Choi et al., 1991; Kent and Mardia, 1988). In anatomical MRI of the brain of normal subjects, three labels [white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF)] are of primary interest; therefore, at least two data channels are required to solve the estimation problem. However, often only an image from a single channel is available, and therefore the requirement for the data to be at least from two channels is quite restrictive.

An interesting way for solving the PV estimation problem even with single-spectral data has been presented in Noe and Gee (2001) and Shattuck et al. (2001) by extending ideas proposed by Santago and Gage (1993, 1995). For this method, which will be referred as *indirect*, we need only to assume that each voxel contains at most two types of tissue. The indirect method involves the identification of the tissue types present in each voxel, followed by a simplified MAP estimation of the proportion of the each tissue type per voxel. Here, the first step is called *partial volume classification* and the second step is called *PVC estimation*. An overview of the entire process, including a separate parameter estimation step, is presented in Fig. 2. Parameter estimation is discussed in the next subsection.

PV classification

In the PV classification step, a label from the set $\hat{\mathcal{L}} = \mathcal{L} \cup \{\{j, k\}: j, k \in \mathcal{L}\}$ is assigned to each voxel. Labels of type $\{j, k\}$ indicate mixed voxels of the tissue types j and k. Let us
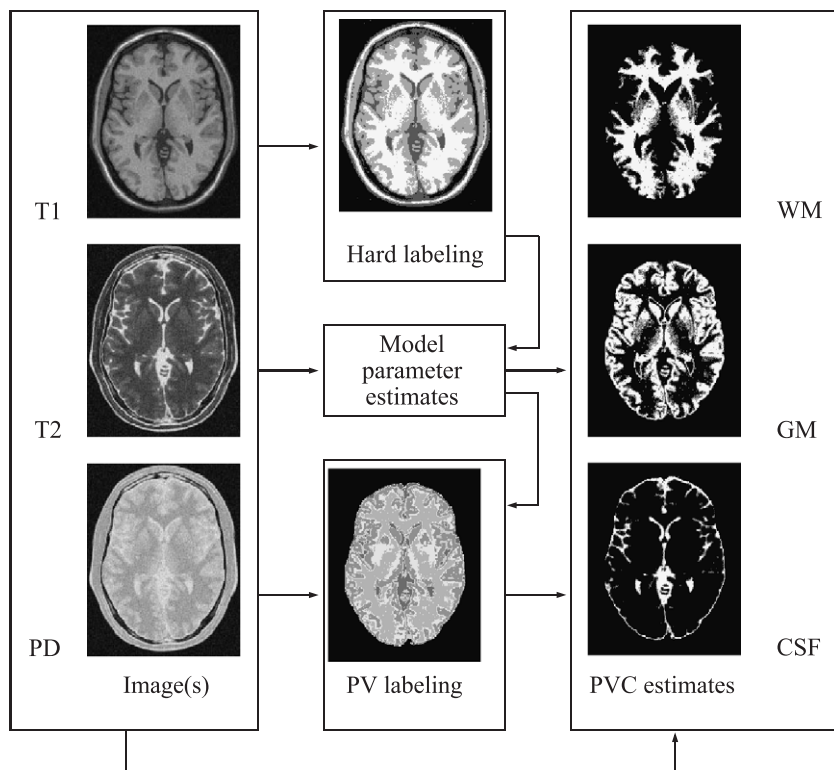

Fig. 2. Depiction of the overall process for the PV estimation.

denote a context image, consisting of the labels of each voxel by $C = \{c_i: i = 1,\ldots, N, c_i \in \hat{\mathcal{L}}\}$

For the PV classification, the probability densities for mixed tissue classes consisting of two types of tissue need to be obtained. This is achieved by marginalizing the densities over all the possible values of the PVC w (Santago and Gage, 1993, 1995). Since pure tissue RVs are Gaussian, also their weighted sums are Gaussian RVs and the marginal density is

$$p(x_i \,|\, c_i = \{j, k\}) = \int_0^1 g[x_i; \mu(w), \Sigma(w)]\mathrm{d}w, \tag{5}$$

where

$$\mu(w) = w\mu_j + (1 - w)\mu_k; \, \Sigma(w) = w^2\Sigma_j + (1 - w)^2\Sigma_k, \tag{6}$$

for the PVC $w \in [0, 1]$. The integral in Eq. (5) has no closed form solution and hence numerical integration must be used to evaluate it. Now the context image can be estimated as

$$C^* = \arg\max_C P(C \,|\, X) = \arg\max_C P(C)\prod_{i=1}^{N} p(x_i \,|\, c_i), \tag{7}$$

where the prior term $P(C)$ is modeled by an Markov Random Field (MRF). Here we use a simple Potts model suggested in Shattuck et al. (2001):

$$P(C) \propto \exp\left(\beta \sum_{i=1}^{N} \sum_{k\in\mathcal{N}_i} \frac{a_{ik}}{d(i, k)}\right), \tag{8}$$

where β is a user tunable parameter, $\mathcal{N}_i$ is the 26-neighborhood around voxel i, $d(i, k)$ is the distance between centers of voxels i and k, and

$$a_{ik} = \begin{cases} 2 & : \quad c_i = c_k \\ 1 & : \quad c_i \text{ and } c_k \text{ share a component} \\ -1 & : \quad \text{otherwise} \end{cases} \tag{9}$$

The iterated conditional modes (ICM) algorithm (Besag, 1986) is used to solve Eq. (7) locally. A well-known approach by Geman and Geman (1984) to solve the optimization problem (7) globally could also be employed, but since this method is much more time consuming than ICM, we prefer to use the latter.

PVC estimation

In the second step, PVC estimation, we estimate the proportions of each tissue type within each voxel. If $c_i^* = j$, that is labeling of voxel i indicates that it is a pure voxel, $w_{i_j}^* = 1$ and for all $k \neq j$ $w_{i_k}^* = 0$. If i is a mixed voxel, $c_i^* = \{j, k\}$, then we employ the maximum-likelihood principle:

$$
\begin{aligned}
w_{i_j}^* &= \arg\max_{w\in[0,1]} g(x_i \,|\, \mu(w), \Sigma(w)) \\
&= \arg\max_{w\in[0,1]} \ln g(x_i \,|\, \mu(w), \Sigma(w)) \\
&= \arg\min_{w\in[0,1]} [(x_i - \mu(w))^T \Sigma(w)^{-1}(x_i - \mu(w)) \\
&\quad + \ln\det(\Sigma(w))],
\end{aligned} \tag{10}
$$

where $\mu(w)$, $\Sigma(w)$ are as in Eq. (6). Furthermore, $w_{i_k}^* = 1 - w_{i_j}^*$ and all the other PVCs are zero. The maximum-likelihood PVC estimate (10) can be solved by a simple grid search, which we have found to be surprisingly efficient in our experiments.

Since Shattuck et al. (2001) and Noe and Gee (2001) adopted the sampling noise model for PVC estimation, we briefly compare Eq. (10) to the solution derivable from the sampling noise model. In that case, we would have only a single covariance matrix, which we denote by Σ_a that describes the sampling noise **a**. Otherwise, the notation is as above. The sampling noise model leads to the objective function

$$[(x_i - \mu(w))^T \Sigma_a^{-1}(x_i - \mu(w))] \tag{11}$$

to be minimized with respect to w. This minimization problem has the closed form solution

$$w_{i_j}^* = r\left(\frac{(\mu_k - x_i)^T \Sigma_a^{-1}(\mu_k - \mu_j)}{(\mu_k - \mu_j)^T \Sigma_a^{-1}(\mu_k - \mu_j)}\right), \tag{12}$$

where $r(x) = x$ when $x \in [0,1]$, $r(x) = 0$ when $x < 0$, and $r(x) = 1$ when $x > 1$. The two objective functions (Eqs. (10) and (11)) differ principally by the additional term $\ln\det(\Sigma(w))$ of Eq. (10). Multiplying Σ_a by a positive constant s does not affect the solution (12) of the sampling noise problem. On the other hand, multiplying $\Sigma(w)$ by s changes the solution of the problem (10). More precisely, if $s > 1$, more weight is put to the term $\ln \det(\Sigma(w))$, which does not depend on the data x_i. Therefore, the material-dependent noise problem (10) can be considered a regularized version of the sampling noise problem. The regularization term $\ln \det(\Sigma(w))$ is a logarithm of high degree polynomial in the multispectral case making the precise analysis of its effect challenging. However, in the single-spectral case, the minimum of $\ln \det(\Sigma(w))$ can be solved explicitly as $\Sigma_k /(\Sigma_j + \Sigma_k)$ if it is in [0, 1]. These considerations are illustrated in Fig. 3, where on the left panel, the minimum of $\ln \det(\Sigma(w))$ is shown when Σ_j is varied. It can be seen that when the variances have equal values, then the minimum of the regularization term is at $w_{i_j} = 0.5$ and hence the regularization term favors the solution $w_{i_j} = 0.5$. When Σ_j is incremented, the value of the favored solution decreases meaning that the voxel is a priori assumed to contain less of the tissue type j than the tissue type k, which has a smaller variance. On the right panel of Fig. 3, all variances, $\Sigma_j, \Sigma_k, \Sigma_a$, are assumed to have an equal value. When this value is incremented, the sampling noise solution does not change but the material noise solution tends slowly towards the value 0.5.

Parameter estimation

As already noted, the model parameters $\mu_j, \Sigma_j, j = 1,\ldots, M$ must be estimated before (or during) the PV estimation. An efficient way to do this is to first label each voxel in the image X by a hard label belonging to $\mathcal{L}$, and then to estimate the parameters based on the intensity distribution of the labeled voxels for each tissue class. In the following, we denote the set of the intensity values of voxels labeled as belonging to the class j by D_j. However, now each D_j contains values that do not represent the pure tissue type in question due to the PVE as well as to misclassified voxels. These values are called outliers. From this, it follows that the set of samples D_j contains samples that are not drawn from the distribu-
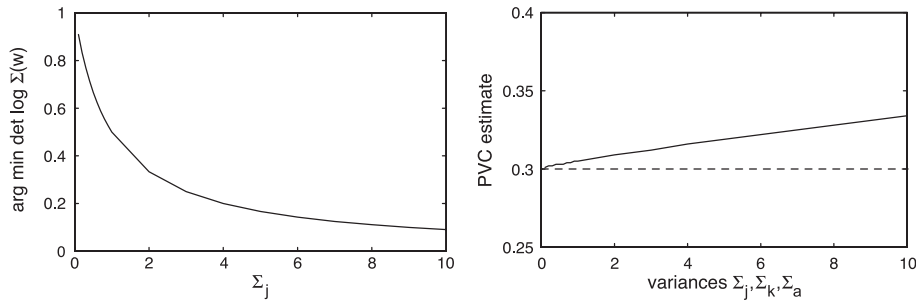
Fig. 3. Differences between the material-dependent and sampling noise models. On left, the minimum of the regularization term ln det $\Sigma(w)$ is shown when $\Sigma_k = 1$ and the value of Σ_j is varied. On right, the PVC estimates obtained by sampling noise Eq. (12) (dotted line) and by material noise Eq. (10) (solid line) are shown when the variances (Σ_a, Σ_j, Σ_{jk}) increase, $x_i = 7$, $\mu_j = 0$, $\mu_k = 10$.

tion describing the tissue type j that leads to problems when applying conventional techniques, such as the maximum-likelihood principle, for parameter estimation. Particularly, in the Gaussian case, the sample mean and covariance are poor estimators of μ_j and Σ_j when the data are contaminated by PVE.

However, we can tackle the problem by using point estimation techniques from robust statistics (Hampel et al., 1985). These robust estimators are designed to tolerate significant deviations from the assumed model while still being able to use information about the model. Especially, we apply the *minimum variance ellipsoid* (MVE) and the *minimum covariance determinant* (MCD) estimators (Rousseeuw, 1984; Rousseeuw and Leroy, 1987) for the parameter estimation task and also combine a simple outlier detection procedure with robust estimators. We will show experimentally that these techniques combined produce better parameter estimates than either of the techniques alone. We start by introducing the method for outlier detection. It is reasonable to assume that voxels lying on the boundaries between tissue types in hard-segmented image are likely to contain PVE, and that their intensity values are therefore outliers. For this reason, voxels with at least one of their 6 neighbors belonging to a different class than the voxel itself are discarded from parameter estimation. We could as well consider 26-neighborhoods instead of 6-neighborhoods, but 6-neighborhoods lead to faster computations. We will refer to the estimates computed based on this kind of reduced data set as "trimmed estimates".

Since the above procedure does not guarantee outlier-free sample sets, robust estimation techniques, namely the *minimum variance ellipsoid* (MVE) and the *minimum covariance determinant* (MCD) estimators are used to estimate parameters μ_j, Σ_j. MVE estimates the mean μ of the set of data samples $D \subset \mathrm{R}^K$ by setting (Rousseeuw and Leroy, 1987):

μ is the center of the ellipsoid with the minimal volume covering at least h points in D.

Here, h is $\lfloor \frac{|D|}{2} \rfloor + 1$, that is, we employ approximately 50% of the data samples to estimate the mean. The covariance is then the matrix defining the minimal ellipsoid around μ multiplied by $(\chi^2_{K,\ 0.5})^{-1}$, where $(\chi^2_{K,\ 0.5})$ is the median of the chi-squared distribution with K degrees of freedom. The MCD estimate is obtained by setting (Rousseeuw and Leroy, 1987):

μ is the mean of h points of D for which the determinant of the covariance matrix is minimal.

Again, h is $\lfloor \frac{|D|}{2} \rfloor + 1$. The covariance matrix is then the covariance matrix of those minimizing h points inflated or defeated to contain exactly h points multiplied by $(\chi^2_{K,\ 0.5})^{-1}$.

In the single-spectral case, both estimators can be computed exactly in $O(|D| \log |D|)$ time (Rousseeuw and Leroy, 1987). For the multispectral case, there are computationally efficient algorithms for approximating the estimators (Rousseeuw and Leroy, 1987, MVE, Rousseeuw and Van Driessen, 1999, MCD).

Besides that they can be computed efficiently, the MVE and the MCD estimators for location and scatter have several favorable properties due to which we adopted them for parameter estimation. The global reliability of a robust estimator can be measured using the concept of a breakdown point. The (finite-sample) breakdown point of an estimator is the smallest percentage of contaminated data that can cause the estimator to take on arbitrarily large aberrant values. The breakdown points of MVE and MCD estimators converge to 50% as the number of samples approaches infinity, whereas the breakdown point of the sample mean tends toward 0%. Note that 50% may be considered as the best possible breakdown point. In addition, the MVE and the MCD estimators are affine equivariant. This means that these estimators commute with the affine transformations, which implies that reparametrization of the sample space does not affect the estimate. This is clearly an important property in multispectral MRI. The combination of the affine equivariance and a high breakdown point is not a common property for robust estimators of multivariate location. For example, affine equivariant M estimators have rather low breakdown points (cf. Rousseeuw and Leroy, 1987). The MCD estimator is asymptotically normal and has a better convergence rate than the MVE-estimator (see Rousseeuw and Van Driessen, 1999), but the algorithm for the multivariate MVE is easier to implement and therefore both estimators are implemented.

Simultaneous parameter estimation and PV classification

To compare our results with other techniques, we also implemented an EM-like algorithm for simultaneous parameter estimation and PV classification. The outline of the algorithm is

(1) Initialize parameters μ_j, Σ_j for every pure voxel class;
(2) Initialize ICM by maximum-likelihood estimation;
(3) Carry out one round of ICM to maximize Eq. (7);
(4) Reestimate the parameters μ_j, Σ_j for $j = 1, \ldots, M$ based on posterior probabilities resulting from step 3;
(5) Finish if the termination condition is satisfied; else go to step 3.

This algorithm, here referred to as EM-ICM, was suggested by Besag (1986). A similar algorithm was used for PV estimation by Noe and Gee (2001), from which our EM-ICM algorithm differs only with respect to its initialization. For initialization, we incorporated an advanced routine based on statistical probability of anatomy maps (SPAMs) (Kamber et al., 1995; Kollokian, 1996). Given that images to be segmented are registered to a common stereotaxic space, SPAMs provide information on the probability of voxels to be dominantly of certain tissue type. Based on SPAMs, it is possible to automatically select a training sample for each tissue class. The training sample then consists of the intensity values of those voxels, which are very likely to be of the tissue type in question. This sample can then be used for training a classifier or for initializing clustering algorithms. Obviously, supervised classifiers trained this way are sensitive to misregistrations and unusual anatomy (Cocosco et al., 2002).

The EM-ICM algorithm is considerably more computationally expensive compared to the one suggested in the previous section for two reasons: (1) it usually takes more iterations to converge and (2) each iteration is more time consuming than a simple iteration of ICM, which has fixed model parameters. This is because each iteration involves computing new likelihoods based on updated parameter estimates and new parameter estimates in addition to the standard ICM cycle.

For example, whereas the procedure for PV estimation implemented in C for a single-spectral $181 \times 217 \times 181$ image takes about 20 min using the ideas proposed in Parameter estimation section, the EM-ICM algorithm can be expected to take slightly over 2 h, both on a 400 MHz MIPS R12000 processor (SGI®Origin® 3800 compute server).

Experiments

Simulated data

Different methods for parameter estimation and the influence of the quality of parameter estimates to PVC estimates were studied using the BrainWeb Simulated Brain Database of Montreal Neurological Institute (Cocosco et al., 1997) that is available at http://www.bic.mni.mcgill.ca/brainweb. The images in the database are generated by an MRI simulator (Kwan et al., 1999), which models the MRI data acquisition process starting from the Bloch equation. The input for simulations is a fuzzy realistic brain phantom (Collins et al., 1998) as opposed to a discrete phantom, which makes the evaluation of PV estimation algorithms possible. Examples of transverse slices of simulated T1-, T2-, and proton density (PD)-weighted images are shown in Fig. 2.

All experiments were performed using single-spectral T1-weighted as well as multispectral (T1-, T2-, and PD-weighted) images. Four image sets for both single-spectral and multispectral cases were used. Each set contained three images with varying noise levels. The voxel size of the images of the first and the second sets was $1 \times 1 \times 1$ mm. The intensity across the images of the first set was uniform, but the images in the second set were simulated with 40% intensity nonuniformity (see Cocosco et al., 1997). The third set contained images with no intensity nonuniformity but the voxel size was $1 \times 1 \times 3$ mm. The images in the fourth set were as those in the first set except that as opposed to the other sets, they were not properly registered to the stereotaxic, Talairach-based brain space used by BrainWeb. The initial classi-

Table 1
The image sets

Image set	INU (%)	Voxel size	Registration
1	0	$1 \times 1 \times 1$ mm	succeeded
2	40	$1 \times 1 \times 1$ mm	succeeded
3	0	$1 \times 1 \times 3$ mm	succeeded
4	0	$1 \times 1 \times 1$ mm	failed

The abbreviation INU stands for the strength of intensity nonuniformity field in a simulated image.

fication algorithm used here (see Implementation section) is sensitive to misregistrations and hence the fourth image set allows us to study the dependence of the parameter estimates on the quality of the initial hard classification. The image sets are summarized in Table 1.

Implementation

Preprocessing steps included the correction of the intensity inhomogeneity where necessary (Sled et al., 1998) and the classification into hard labels (WM, GM, and CSF) by a neural network classifier that was trained automatically using SPAMs (Kollokian, 1996). (The classifier used was a standard back propagation artificial neural network with one hidden layer with 10 nodes.) As described in Simultaneous parameter estimation and PV classification section, the use of SPAMs for automatic training of unsupervised classifiers is sensitive to errors in registering images to the stereotaxic space. Voxels containing primarily nonbrain tissues were labeled as background before parameter estimation. Six methods for the parameter estimation based on a hard labeling were examined. In the maximum-likelihood (ML) method, parameters for each class were the sample-mean and the sample-covariance of that class in the hard labeled image. Trimmed maximum-likelihood (TML) estimates were the sample-mean and the sample-covariance of that class in the trimmed hard labeled image. By trimming, it is meant that outliers were detected and removed for each class as described in Parameter estimation section. Minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD), respectively, estimates were obtained by applying MVE and MCD estimators to the classes in the hard labeled image. Trimmed minimum volume ellipsoid (TMVE) and trimmed minimum covariance determinant (TMCD), respectively, estimates were obtained by applying MVE and MCD estimators to the trimmed hard labeled image. Moreover, we listed parameter estimates resulting from the EM-ICM algorithm and parameter estimates used to initialize it. For the purpose of the PV classification, intensities of background voxels were assumed to have a normal distribution with the zero mean and the covariance matrix $0.1\Sigma_{CSF}$ where Σ_{CSF} is the covariance matrix of the class CSF. In reality, the image intensities in magnitude MR images are all greater than zero by definition and the intensity distribution of the background voxels is Rician. Hence, this assumption is a rough approximation of the reality and it is made purely for convenience.

During the PV classification, pure tissue classes were WM, GM, and CSF. Mixed tissue classes were CSF/background, CSF/GM, and GM/WM. Because nonbrain voxels were masked off, the background class was not required. The parameter β (see Eq. (8)) controlling the relative strength of the MRF was set to 0.1 unless otherwise mentioned. The grid search for solving the PVC estimation step (10) used an equally spaced

grid $G = \{0, 0.01, 0.02 \ldots, 1\}$ of 101 points at which the objective function in Eq. (10) was evaluated. The value of $w_{i_j} \in G$ yielding optimal value of the objective function was then selected as the estimate $w_{i_j}^*$.

Error criteria

To quantitatively evaluate location estimates μ_j, the mean of the Mahalanobis distances between the true sample mean and the estimated mean of each pure tissue class was computed. (We name this criterion "Mahalanobis error".) The true mean and the covariance matrix were computed from voxels that contained over 99% of a certain tissue type in the phantom image. This error criterion was selected due to its simplicity and its applicability to both the single and multispectral case. Direct comparison of Mahalanobis errors obtained at different noise levels would require normalization of the covariances used in calculation of errors. However, this kind of normalization would make the error criterion and its analysis more complicated.

For covariance estimates, the definition of an error criterion that would be meaningful for our problem is more difficult. In experiments, errors in location estimates usually had a greater effect to the PV estimation accuracy than errors in estimates of covariances Σ_j. For these two reasons, we only describe the quality of covariance estimates verbally and we do not attempt to evaluate it quantitatively.

For estimates of PVCs, the following error criterion is used

$$E_{\mathrm{PVE}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} |w_{i_j}^* - w_{i_j}|, \qquad (13)$$

where $w_{i_j}^*$ is the estimated PVC, w_{i_j} is the true PVC, M is the number of pure tissue types, and N is the number of voxels in the brain volume. Here, $M = 3$. Note that the value of this criterion does not depend on the number of the pure tissue types in the image, because we assume that each voxel contains at most two types of tissue. The range of this error criterion is from 0 to 2. We have selected this criterion closely related to mean absolute error (MAE) because we are interested in the average accuracy that can be obtained in PV estimation. Error criteria that are based on mean squared error (MSE) could be as well employed, but they put more weight to gross errors, hence evaluating more robustness than accuracy.

To establish a link between Mahalanobis errors in the model parameter estimates and the error criterion for the PVC estimates, a simulation was performed. For the simulation, we generated intensities of WM/GM mixed voxels with varying mixing proportions. Statistical models for the required tissue classes were obtained from the simulated images in the same manner as when computing the error criterion for parameter estimates. That is, parameters for the densities were computed based on voxels that contained over 99% of a certain tissue type (WM or GM) in the phantom image. We also randomly generated a set of erroneous model parameter estimates yielding a fixed Mahalanobis error. Subsequently, PVCs were estimated using Eq. (10) based on the simulated intensity values and the erroneous location estimates. Errors in covariance estimates were not simulated, because this simplifies the interpretation of the simulation results. A more detailed description of the simulation can be found in the Appendix A.

As can be seen in Fig. 4, the relationship between the PVC and Mahalanobis errors is approximately linear in most cases.
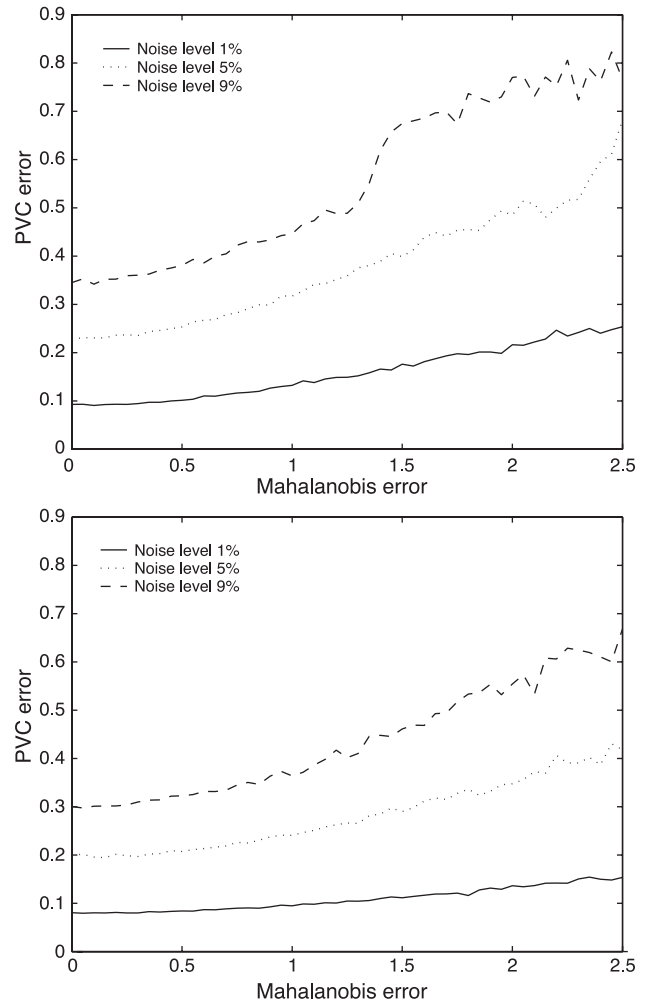

Fig. 4. The influence on the errors in parameter estimates to the errors in PVC estimates. The top and bottom panels show results for, respectively, the single-spectral (T1-weighted only) and multispectral (T1-, T2-, and proton density-weighted) cases.

Only the curve corresponding to the highest level of noise in the single-spectral case seems to be nonlinear. Based on this curve, Mahalanobis error values above 1.35 are worse than the values below 1.35. However, experiments with the simulated MR images rarely gave as large as 1.35 Mahalanobis error values with the highest noise level. Also, it can be noticed in Fig. 4 that the PVC error increases more rapidly relative to Mahalanobis error in the single-spectral case than in the multispectral case.

IBSR data set

Simulated images allow us to study in detail each step of the proposed procedure. To examine the practicality of the proposed method, it is tested on real data sets. With real data, quantitative valuation of the PV estimation is restricted to evaluation of the quality of the subsequent hard segmentations. For this, the segmentation from the PV estimation with TMCD parameter estimates was evaluated with the internet brain segmentation repository (IBSR) data set. The IBSR data set comprises of the 20 normal MR brain data sets and their manual segmentations

that were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://neuro-www.mgh.harvard.edu/cma/ibsr. These images have been used to evaluate the quality of several (automatic) tissue classification methods. The quality measure used is the Tanimoto coefficient between the tissue class in the expert segmentation and in the automatic segmentation (cf., e.g., Shattuck et al., 2001).

The preprocessing of the images consisted of correction of interslice intensity nonuniformity that was done (see Zijdenbos et al. (1995)) before registering them with stereotaxic (Talairach) space. After stereotaxic registration, the N3 algorithm (Sled et al., 1998) was used to correct for the remaining intensity nonuniformity. Images were classified for parameter estimation in the same way as simulated images. A hard segmentation was generated from PVC estimates by selecting the class with the largest PVC for each voxel as its label. The dimensions for the images were $256 \times 256 \times 64$.

Results

Simulated data

Parameter estimation

Results for the single-spectral case are presented in Table 2. Robust estimators combined with trimming were the most reliable among the segmentation-based parameter estimation techniques. They were also better than parameter estimates that were used to initialize EM-ICM and usually better than the final results of EM-ICM. For example, with image set 1 and 5% noise, TMCD produced a Mahalanobis error of 0.06, whereas the error resulting from MCD was 0.19 and the error resulting from EM-ICM was 0.11. Only with the image set 4, where the initial classifications were not accurate, the EM-ICM algorithm yielded better results than image-based parameter estimation techniques. Without trimming of the initial classification, robust estimators did well only with the lowest level of noise. The most extreme case occurred with image set 3 and the highest noise level, where the errors produced by robust estimators were 10 times higher than those

Table 2
Means of Mahalanobis distances between true and estimated mean intensities of tissue types of T1-weighted data

Image set	Noise level (%)	ML	MVE	MCD	TML	TMVE	TMCD	INIT	EM-ICM
1	1	1.16	0.11	0.13	0.20	0.21	0.19	0.27	0.11
	5	0.46	0.18	0.19	0.06	0.07	0.06	0.13	0.09
	9	0.28	0.55	0.39	0.17	0.04	0.05	0.06	0.14
2	1	0.99	0.07	0.07	0.11	0.14	0.16	0.26	0.09
	5	0.48	0.16	0.17	0.09	0.07	0.07	0.14	0.10
	9	0.28	0.48	0.39	0.15	0.06	0.05	0.23	0.19
3	1	2.32	0.30	0.13	0.50	0.14	0.13	0.70	0.20
	5	0.66	0.44	0.38	0.13	0.10	0.05	0.22	0.22
	9	0.31	0.64	0.51	0.19	0.05	0.04	0.13	0.25
4	1	3.03	0.65	0.54	1.51	0.58	0.61	7.58	0.11
	5	1.38	0.39	0.45	0.84	0.39	0.46	2.84	0.10
	9	0.92	1.37	1.26	0.58	0.27	0.37	1.62	0.15

Names for different methods for the parameter estimation are given in Implementation section. INIT is the initialization for the EM-ICM algorithm.

Table 3
Means of Mahalanobis distances between true and estimated mean intensities of tissue types of multispectral data (cf. Table 2)

Image set	Noise (%)	ML	MVE	MCD	TML	TMVE	TMCD	INIT	EM-ICM
1	1	1.57	0.52	0.33	0.32	0.40	0.37	0.52	0.30
	5	0.54	0.36	0.41	0.25	0.36	0.23	0.26	0.11
	9	0.49	0.37	0.44	0.33	0.24	0.31	0.14	0.13
2	1	1.36	0.44	0.32	0.27	0.49	0.30	0.60	0.19
	5	0.54	0.58	0.44	0.23	0.39	0.22	0.30	0.17
	9	0.51	0.41	0.48	0.33	0.25	0.33	0.23	0.20
3	1	3.26	2.49	2.30	0.61	0.43	0.25	1.05	1.25
	5	1.09	1.05	0.94	0.22	0.24	0.18	0.36	0.29
	9	0.63	0.67	0.61	0.19	0.21	0.18	0.19	0.26
4	1	2.26	0.52	0.30	0.76	0.43	0.37	9.93	5.67
	5	2.71	2.09	1.99	2.16	1.98	1.81	3.93	0.11
	9	0.91	0.92	0.87	0.37	0.36	0.34	2.38	0.13

produced by trimmed robust estimators. The performance of TMVE and TMCD did not differ much in the single-spectral case. The best variance estimates were achieved using the trimmed robust estimators, while the trimmed ML estimator typically underestimated the variances and nontrimmed estimators overestimated them, although over/underestimation was always consistent.

In Table 2, error values typically decrease when the noise level increases. This is because, as was explained in Error criteria section, the values of the Mahalanobis error at different noise levels are not necessarily comparable.

Results for the multispectral case are presented in Table 3. These results are similar to those with single-spectral data, except that TML estimator was almost as good as TMCD estimator and even better than TMVE estimator. With the image set 1 and the noise level of 5%, for example, the error of TML was 0.25, the error of the TMVE was 0.36, and the error of TMCD was 0.23. It remains to be studied whether poor results of the TMVE estimator as compared to the single-spectral case were due to the estimator itself or to the approximative algorithm used to compute it. The failure of all image-based estimators with the image set 4 and a 5% noise level was due to very poor initial classifications (about two-thirds of GM voxels were labeled as WM), causing also robust estimators to break down. The quality of initial classifications was better with the other noise levels. With all the image sets, the covariance estimates obtained by trimmed robust estimators were better than the ones of the TML estimator, and the effect of this can be seen in the quality of PVC estimates (see Table 5).

The EM-ICM algorithm typically produced the best estimates when the voxel size was 1 mm^3 (i.e., image sets 1, 2, and 4). However, with the image set 4 and the lowest noise level, EM-ICM failed in the parameter estimation (error 5.67) and this can be seen also in the quality of the corresponding PVC estimates (error 0.783). This failure was probably because the EM-ICM algorithm failed to converge to a strong maximum due to a poor initialization based on the image that was not properly registered to the stereotaxic space. Especially, the initial covariance estimates greatly overestimated the (scaling of) covariance matrices, which could explain why EM-ICM failed particularly with the lowest noise level. The application of the EM-ICM algorithm did not always improve the parameter estimates when compared the parameter estimates used for its initialization. Indeed, there are no results

Table 4
Errors (13) in PV estimation with T1-weighted data (cf. Table 2)

Image set	Noise (%)	ML	TML	TMVE	TMCD	EM-ICM
1	1	0.108	0.083	0.084	0.084	0.078
	5	0.157	0.156	0.156	0.155	0.156
	9	0.249	0.259	0.247	0.248	0.252
2	1	0.112	0.089	0.089	0.089	0.086
	5	0.160	0.158	0.157	0.157	0.157
	9	0.252	0.262	0.252	0.253	0.256
3	1	0.156	0.106	0.089	0.089	0.088
	5	0.206	0.195	0.195	0.195	0.196
	9	0.316	0.317	0.306	0.308	0.318
4	1	0.160	0.107	0.082	0.083	0.079
	5	0.257	0.207	0.162	0.168	0.156
	9	0.364	0.324	0.276	0.286	0.255

concerning the convergence properties of the EM-ICM algorithm (cf. Besag, 1986). Furthermore, the convergence to a local minimum (or maximum) may actually draw estimates further from the global minimum (or maximum).

Partial volume coefficients

Errors in the PV estimation are listed in Tables 4 and 5. In Fig. 5, PV estimation errors obtained by applying different parameter estimation methods are compared with some of the image sets. Improvements in the quality of the model parameter estimates improved PVC estimates in most cases. In that respect, these results were consistent with the simulation described in Error criteria section. However, generally the errors in PVC estimates obtained with simulated images were lower than those obtained in the simulation of Error criteria section assuming the same value of the Mahalanobis error. This is not surprising since in the simulated images some voxels are classified as pure voxels that yield a PVC error value of zero in the case of a correct classification. Similar reasoning could be used to explain why the improvements to an already small Mahalanobis error led rarely to a considerably improved PVC estimation. An example of this phenomenon can be seen in Table 4 by comparing the results by TMCD (Mahalanobis error 0.05, PVC error 0.195) and EM-ICM (Mahalanobis error 0.22, PVC error 0.196) with the image set 3 and the noise level of 5%.

PVC estimates obtained with multispectral images were considerably better than those obtained with single-spectral

Table 5
Errors (13) in PV estimation with multispectral data (cf. Table 2)

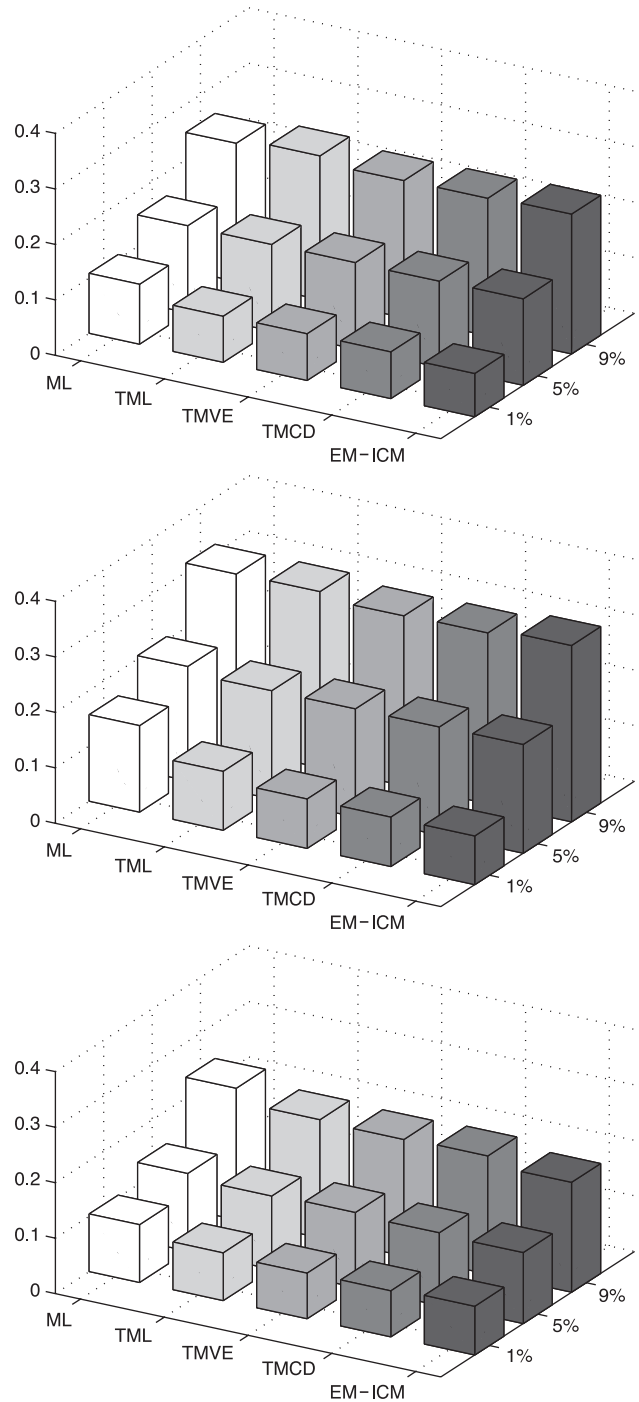| Image set	Noise (%)	ML	TML	TMVE	TMCD	EM-ICM
1	1	0.104	0.086	0.082	0.083	0.088
	5	0.140	0.132	0.135	0.131	0.128
	9	0.237	0.213	0.210	0.213	0.198
2	1	0.111	0.087	0.086	0.084	0.081
	5	0.147	0.137	0.138	0.137	0.134
	9	0.243	0.217	0.223	0.216	0.201
3	1	0.153	0.104	0.084	0.085	0.099
	5	0.202	0.169	0.165	0.164	0.163
	9	0.271	0.257	0.264	0.256	0.252
4	1	0.130	0.095	0.083	0.082	0.783
	5	0.642	0.538	0.485	0.443	0.128
	9	0.285	0.245	0.235	0.241	0.199

Fig. 5. Errors in PV estimation with different parameter estimation methods and noise levels. Form top, the sets 1 and 3 of T1-weighted images and the set 1 of multispectral images.

images (see Fig. 6). Moreover, it can be seen from PV estimation results that gain from multispectral images increased with the noise level. The differences in the errors in PVC estimates between single and multispectral cases increased with the noise level. However, it should be noted that in this case, multispectral images were perfectly registered with each other, which is rarely the case with real data. In the multispectral image set 4, better PVC estimates with the noise level of 9%
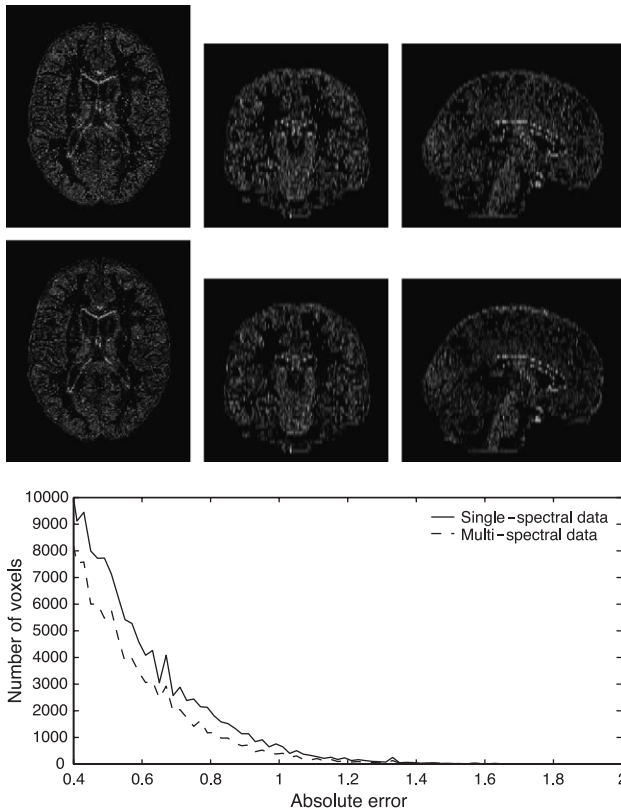
Fig. 6. Voxel-wise absolute errors in PVC estimates obtained with single-spectral (top) and multispectral (center) data. In the bottom, histograms of voxel-wise absolute errors are shown. The noise level is 5% and the voxel size is 1 × 1 × 3 mm. The parameter estimates were obtained with TMCD estimator.

compared to those with the noise level of 5% were due to particularly poor initial classification when the noise level was 5%.

Shattuck et al. (2001) have evaluated their very fast PV estimation routine using BrainWeb images but using an error criterion different from ours. Therefore, to compare our results with theirs, we computed errors of TMCD-based PVC estimates
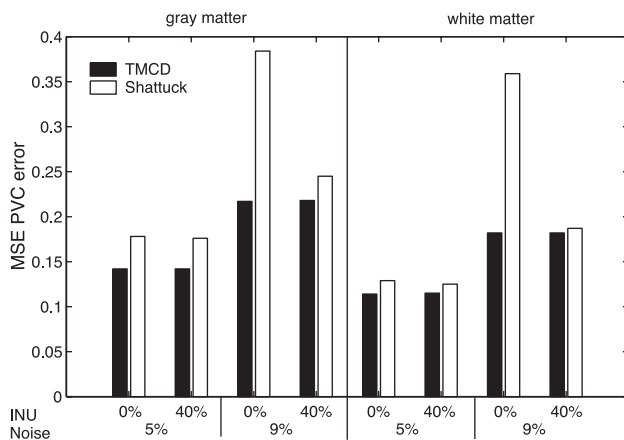

Fig. 7. MSE error values as in Eq. (14) with BrainWeb images. PVCs estimated with the TMCD method are compared to the ones published in Shattuck et al. (2001).

Table 6
Misclassification rates in the brain area in percent

	Noise (%)	ORIG	ML	TML	TMVE	TMCD	EM-ICM
Image set 4	1	18.8	4.5	2.1	1.9	1.9	1.9
	5	24.3	8.6	6.8	6.0	6.1	5.9
	9	31.6	17.0	15.1	13.6	14.0	13.4

ORIG denotes the initial hard classification. Otherwise, abbreviations are as in Table 2.

also using the criterion they applied. The criterion is defined as (cf. Eq. (13) for notation)

$$E_{MSE_j} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |w_{i_j}^* - w_{i_j}|^2},$$ (14)

for each tissue class j. Particularly, we compared results with single-spectral images from sets 1 and 2. (In Shattuck et al., 2001, no results with multispectral data or images corresponding our sets 3 and 4 were available.) Results are shown in Fig. 7. As can be seen invariably, the TMCD-based method was better of the two. Particularly, the PVC estimates for the gray matter were considerably better with the TMCD estimated parameters. For the white matter, differences were smaller. Although the methods differ in all their components, they are both based on a division of the problem in smaller subproblems. Hence, these results are a good indication that our parameter estimation techniques are a real improvement as compared to previous similar approaches. The model by which the data were simulated is the material-dependent noise model, but with very similar covariance for each class. Hence, the results of the PVC estimation step are similar with both models (cf. PVC estimation section).

Comparison with the initial hard segmentation

With image set 4, we also list the misclassification rates of the initial hard labeling and the hard labeling derived from PV estimation in Tables 6 and 7. For this, the PV estimations were hardened by labeling each voxel by the most dominant tissue type. Almost all PV estimation techniques improved the results of the original labeling (see Fig. 8 for an example). The only exception was EM-ICM in the multispectral case with the lowest noise level. In this case, the poor classification result was due to poor parameter estimates (cf. Parameter estimation section).

Computation times

The entire procedure to estimate PVCs with a single-spectral 181 × 217 × 181 image took on average about 20 min using parameter estimation based on hard classification. Most of the time (over 10 min) was spent to the PV classification step. Image

Table 7
Misclassification rates in the brain area in percent with multispectral data

	Noise (%)	ORIG	ML	TML	TMVE	TMCD	EM-ICM
Image set 4	1	8.2	3.2	2.9	2.6	2.6	37.8
	5	38.2	32.4	25.6	19.7	18.9	4.5
	9	21.7	11.4	9.9	9.8	9.6	9.0

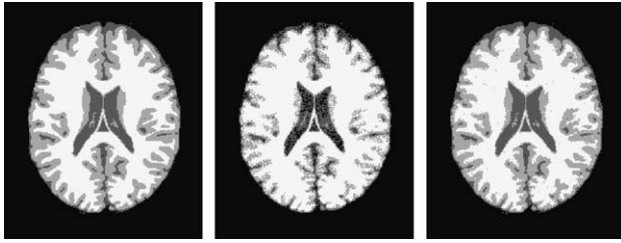ORIG denotes the original hard classification. Otherwise, abbreviations are as in Table 2.

Fig. 8. The improvement of the hard classification results by using PV model. Left: the ground truth classification. Center: the initial classification of the misregistered image with 5% of noise. Right: the hardened partial volume estimation with TMCD parameters of the same image.

preprocessing (hard classification and nonuniformity correction) typically took 3 min. Parameter estimation and PV estimation were fast, requiring about 1 min each. In the multispectral case, the total computation time was approximately 40 min. PV estimation with EM-ICM algorithm took 2 h in the single-spectral case and 6 h in the multispectral case. All computation times are on a 400 MHz MIPS R12000 processor (SGI®Origin® 3800 compute server). The algorithms were implemented in C. These running times, particularly for the PV classification step, could be probably improved by paying more attention on the efficiency of the implementation.

Results with the IBSR data set

The Tanimoto coefficients (TCs) of segmentations obtained from PV estimations using the TMCD parameter estimates can be found in Table 8. The mean value of the Tanimoto coefficient was 0.6686 for the WM and 0.6760 for the GM. We experimented the procedure also with a higher value (0.5) for the parameter β controlling the strength of the MRF prior. The average TC values in this case were 0.6637 (WM) and 0.6807 (GM).

We selected to evaluate our algorithm with this data set due to its frequent use in the evaluation of MR segmentation algorithms. Although our interest in PVE is not in improving hard segmentations by modeling of it, it is interesting to compare the results of our method to those obtained with algorithms for hard segmentation. Rajapakse and Krugge (1998) compared several algorithms and the best TC values were obtained by *Adaptive* MAP method, and they were 0.567 for the WM and 0.564 for the GM. A more recent method by Marroquin et al. (2002) achieved TCs 0.683 for the WM and 0.662 for the GM. The average running time for this algorithm was 19.2 min with the IBSR data set, which is similar to our methods, but no PV estimation can be obtained with this method.

Discussion

In this paper, various methods for the parameter estimation for a statistical model of the partial volume effect have been studied. It

has been shown that it is possible to estimate parameters in a reliable and fast way based on the initial hard labeling of the image before the actual partial volume estimation procedure. For this, outliers of each class in the classified image were eliminated using a simple morphological rule, and thereafter, parameters for model were computed by using robust estimators: MVE and MCD. Parameters estimated by MCD were of better quality than those relying on sample means and covariances as estimates. It was also observed that the trimming of the initial classification is necessary even when using robust estimators. The parameter estimates based on the hard classification were almost as good as the ones obtained by the EM-ICM algorithm for simultaneous PV classification and parameter estimation. The EM-ICM algorithm is, however, more time consuming. Another downside of the EM-ICM algorithm for parameter estimation is that it is essentially a local optimization algorithm and hence its results depend on its initialization. In our experiments, EM-ICM failed completely in parameter estimation and subsequently in PV estimation with one of the multispectral images. In that case, the parameter estimates that were used to initialize the EM-ICM algorithm were of particularly poor quality due to the failed stereotaxic registration.

The influence of improved parameter estimates on the PV estimation was studied. As expected, better parameter estimates produced better estimates of partial volume coefficients. Furthermore, we examined the effect of increased error in parameter estimation to the error in PV estimation through a simplified simulation. The simulation showed that the relationship between Mahalanobis errors in parameter estimation and errors in PV estimation is approximately linear for a fixed level of noise.

We have studied our parameter estimation procedures in conjunction with the material-dependent noise model, an indirect PV estimation algorithm and a particular method for preliminary hard classification. But in fact, the proposed parameter estimation methods (TMVE and TMCD) can be applied more generally, and hence other choices could have been made as well. First, location parameter estimates for the sampling noise model can be obtained in the exactly same manner as for the material-dependent noise model. However, only a single covariance matrix describing the sampling noise is required. This issue can be solved by considering the pooled covariance of the estimated covariance for the tissue classes (Dougherty, 1990, p. 467). Secondly, as parameter estimation is a separate step of the whole procedure, it could be applied also with direct PV estimation algorithms. Furthermore, the initial hard classification and correction of intensity nonuniformity can be performed with other methods than those applied in this study. The state of the art algorithms for these tasks (such as Marroquin et al., 2002; Van Leemput et al., 1999; Zhang et al., 2001) are however more time consuming than the methods applied in this study. Moreover, the robustness in parameter estimation compensates also for classification errors as was shown with misregistered Brainweb images.

Table 8
IBSR data set results

Image	100_23	110_3	112_2	11_3	12_3	13_3	111_2	15_3	16_3	17_3
GM	0.6783	0.7363	0.7063	0.7381	0.8044	0.7716	0.7765	0.6076	0.6261	0.6026
WM	0.6844	0.6841	0.6632	0.7300	0.7393	0.7550	0.7409	0.6009	0.6275	0.6609
Image	191_3	1_24	202_3	205_3	2_4	4_8	5_8	6_10	7_8	8_4
GM	0.6896	0.7106	0.7968	0.7426	0.4535	0.5876	0.6581	0.5632	0.6447	0.6261
WM	0.7399	0.7328	0.7516	0.7519	0.5399	0.5625	0.6006	0.4851	0.6632	0.6591

To conclude, we have presented fast and robust methods to estimate parameters for the mixel model. These robustified parameter estimates were shown to improve the performance of PV estimation as compared standard estimates and heuristics used in Shattuck et al. (2001). Moreover, parameter estimation based on segmented image was demonstrated to be much faster than competing EM style algorithms (Noe and Gee, 2001; Van Leemput et al., 2003). Also, as demonstrated with the IBSR data set, our fast PV estimation routine produced hard segmentations of a similar quality than a state of the art algorithm for this task (Marroquin et al., 2002). In summary, we recommend the use of the TMCD estimator based on hard labeling for fast and reliable parameter estimation for a statistical PVE model.

Acknowledgments

Appendix A

This appendix describes the simulation used to link the Mahalanobis errors in parameter estimation and the absolute errors in PVC estimation. Only mixtures of the white matter and the gray matter tissue types were considered. The intensities were generated by using the material-dependent noise model and assuming parameters as in GM and WM classes of the simulated images. The mixing proportions were 0.1, 0.3, 0.5, 0.7, and 0.9. The number of samples generated per each mixing proportion were 1000.

Different Mahalanobis errors evaluated were 0, 0.05, 0.1,..., 2.5. One hundred different location parameter estimates were generated for each fixed value of Mahalanobis error. The location parameter estimates were generated by first randomly selecting the direction of the error for both WM and GM tissue classes. In the single-spectral case, this is easy, since only the choice between positive or negative direction has to be made. This is a bit more complicated in the multispectral case, where random points lying on the unit sphere have to be drawn (see Marsaglia, 1972). After the selection of directions of the error, the proportion of the total Mahalanobis error rising from the estimate of the mean of WM class was randomly drawn from the interval [0.2, 0.8]. Thereafter, the exact covariance and each erroneous location estimate was used to estimate the true PVC-coefficient from each 5000 samples of intensity values. The results of the simulation are shown in Fig. 4. We also performed a simulation where Mahalanobis errors were assumed to be equally distributed between tissue types, but results were similar to the more general experiment described above.

References

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc., Ser. B 36, 192–236.

Besag, J., 1986. On the statistical analysis of dirty pictures. J. R. Stat. Soc., Ser. B 48 (3), 259–302.

Bilbro, G., Snyder, W.E., 1991. Optimization of functions with many minima. IEEE Trans. Syst. Man Cybern. 21, 840–849.

Choi, H.S., Haynor, D.R., Kim, Y., 1991. Partial volume tissue classification of multichannel magnetic resonance images—A mixel model. IEEE Trans. Med. Imag. 10 (3), 395–407.

Cocosco, C., Kollokian, V., Kwan, R.-S., Evans, A., 1997. Brainweb: online interface to a 3D MRI simulated brain database. Proc. of 3rd Int. Conf. on Functional Mapping of the Human Brain. NeuroImage, vol. 5. (no.4, part 2/4, Suppl.), S425.

Cocosco, C., Zijdenbos, A., Evans, A., 2002. Automatic generation of training data for brain tissue classification from MRI. Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI02), LNCS 2488. Springer-verlag, Berlin/Heidelberg, pp. 516–523.

Collins, L., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imag. 17 (3), 463–468.

Collins, D.L., Zijdenbos, A.P., Baar, W.F.C., Evans, A.C., 1999. ANIMAL+INSECT: improved cortical structure segmentation. Proc. of the 16th International Conference on Information Processing in Medical Imaging (IPMI). LNCS, vol. 1613, pp. 210–223.

Collins, D.L., Montagnat, J., Zijdenbos, A.P., Evans, A.C., Arnold, D.L., 2001. Automated estimation of brain volume in multiple sclerosis with BICCR. Proc. of the Annual Symposium on Information Processing in Medical Imaging (IPMI01). LNCS, vol. 2082, pp. 141–147.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Ser. B Methodol. 39 (1), 1–39.

Dougherty, E.R., 1990. Probability and Statistics for the Engineering, Computing and Physical Sciences. Prentice-Hall, New Jersey.

Fischl, B., Salat, D., Busa, E., Albert, M., Dietrich, M., Haselgrov, C., van der Kouwe, A., Killany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions and the Bayesin restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6 (6), 721–741.

González Ballester, M.A., Zisserman, A., Brady, M., 2000. Segmentation and measurement of brain structures in MRI including confidence bounds. Med. Image Anal. 4 (3), 189–200.

González Ballester, M.A., Zisserman, A., Brady, M., 2002. Estimation of the partial volume effect in MRI. Med. Image Anal. 6 (4), 389–405.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1985. Robust Statistics. An Approach Based on Influence Functions. Wiley, New York.

Kamber, M., Shinghal, R., Collins, D.L., Francis, G.S., Evans, A.C., 1995. Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. IEEE Trans. Med. Imag. 14 (3), 442–453 (Sep.).

Kent, J.T., Mardia, K.V., 1988. Spatial classification using fuzzy membership models. IEEE Trans. Pattern Anal. Mach. Intell. 10 (5), 659–671.

Kollokian, V., 1996. Performance analysis of automatic techniques for tissue classification in magnetic resonance images of the human brain. Master's thesis, Dept. of Computer Science, Concordia University, Montreal, Canada.

Kwan, R.-S., Evans, A., Pik, G., 1999. MRI simulation-based evaluation of image-processing and classification methods. IEEE Trans. Med. Imag. 18 (11), 1085–1097.

Laidlaw, D.H., Fleischer, K.W., Barr, A.H., 1998. Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms. IEEE Trans. Med. Imag. 17 (1), 74–86.

MacDonald, D., Kabani, N., Avis, D., Evans, A., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. NeuroImage 12 (3), 340–356.

Marroquin, J., Vemuri, B., Botello, S., Calderon, F., Fernandez-Bouzas, A., 2002. An accurate and efficient Bayesian method for automatic segmentation of brain MRI. IEEE Trans. Med. Imag. 21 (8), 934–945.

Marsaglia, G., 1972. Choosing a point from the surface of a sphere. Ann. Math. Stat. 43 (2), 645–646.

Noe, A., Gee, J., 2001. Partial volume segmentation of cerebral MRI scans with mixture model clustering. Proc. of Information Processing in Medical Imaging, 17th International Conference. LNCS 2082 IPMI. Springer-verlag, Berlin/Heidelberg, pp. 423–430.

Pham, D.L., Prince, J., 1998. Partial volume estimation and the fuzzy c-means algorithm. Proc.-Int. Conf. Image Proc. (ICIP98), vol. III, pp. 819–822.

Pham, D.L., Prince, J., 1999. Adaptive fuzzy segmentation of magnetic resonance images. IEEE Trans. Med. Imag. 18 (9), 737–752.

Rajapakse, J., Krugge, F., 1998. Segmentation of MR images with intensity inhomogeneities. Image Vis. Comput. 16, 165–180.

Rousseeuw, P.J., 1984. Least median of squares regression. J. Am. Stat. Assoc. 79, 871–880.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley-Interscience, New York.

Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41 (3), 212–223.

Rousset, O., Ma, Y., Evans, A., 1998. Correction for partial volume effects in pet: principle and validation. J. Nucl. Med. 39, 904–911.

Ruan, S., Jaggi, C., Xue, J., Fadili, J., Bloyet, D., 2000. Brain tissue classification of magnetic resonance images using partial volume modeling. IEEE Trans. Med. Imag. 19 (12), 1179–1187.

Santago, P., Gage, H.D., 1993 (September). Quantification of MR brain images by mixture density and partial volume modeling. IEEE Trans. Med. Imag. 12 (3), 566–574.

Santago, P., Gage, H.D., 1995 (November). Statistical models of partial volume effect. IEEE Trans. Image Process. 4 (11), 1531–1540.

Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K., Rottenberg, D., Leahy, R., 2001. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13 (5), 856–876.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Trans. Med. Imag. 17 (1), 87–97.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. IEEE Trans. Med. Imag. 18 (10), 897–908.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 2003. A unifying framework for partial volume segmentation of brain MR images. IEEE Trans. Med. Imag. 22 (1), 105–119.

Wang, Y., Adah, T., Xuan, J., Szabo, Z., 2001 (June). Magnetic resonance image analysis by information theoretic criteria and stochastic site models. IEEE Trans. Inf. Technol. Biomed. 5 (2), 150–158.

Wei, G., Tanner, M., 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. J. Am. Stat. Assoc. 85, 699–704.

Wells III, W., Grimson, W., Kikinis, R., Jolesz, F.A. 1996. Adaptive segmentation of MRI data. IEEE Trans. Med. Imag. 15 (4), 429–442.

Xu, C., Pham, D., Rettman, M., Yu, D., Prince, J., 1999 (June). Reconstruction of the human cerebral cortex from magnetic resonance images. IEEE Trans. Med. Imag. 18 (6), 467–480.

Zeng, X., Staib, L., Schultz, R., Duncan, J., 1999 (October). Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation. IEEE Trans. Med. Imag. 18 (10), 927–937.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden random markov field model and the expectation–maximization algorithm. IEEE Trans. Med. Imag. 20 (1), 45–57.

Zijdenbos, A., Dawant, B.M., Margolin, R.A., 1995. Intensity correction and its effect on measurement variability in the computer-aided analysis of MRI. Proc. of the 9th Internation Symposium and Exhibition on Computer Assisted Radiology (CAR). Springer-verlag, Berlin/Heidelberg, pp. 216–221.