Segmentation of Cerebral MRI Scans Using a Partial Volume Model, Shading Correction, and an Anatomical Prior

Aljaž Noe[a], Stanislav Kovačič[a], James C. Gee[b]

[a] Faculty of Electrical Engineering, University of Ljubljana
Tržaška cesta 25, 1000 Ljubljana, Slovenija

[b] Department of Radiology, University of Pennsylvania
1 Silverstein, 3400 Spruce Street, Philadelphia, PA 19104, USA

ABSTRACT

A mixture-model clustering algorithm is presented for robust MRI brain image segmentation in the presence of partial volume averaging. The method uses additional classes to represent partial volume voxels of mixed tissue type in the image. Probability distributions for partial volume voxels are modeled accordingly. The image model also allows for tissue-dependent variance values and voxel neighborhood information is taken into account in the clustering formulation. Additionally we extend the image model to account for a low frequency intensity inhomogeneity that may be present in an image. This so-called shading effect is modeled as a linear combination of polynomial basis functions, and is estimated within the clustering algorithm. We also investigate the possibility of using additional anatomical prior information obtained by registering tissue class template images to the image to be segmented. The final result is the estimated fractional amount of each tissue type present within a voxel in addition to the label assigned to the voxel. A parallel implementation of the method is evaluated using synthetic and real MRI data.

Keywords: partial volume segmentation, intensity inhomogeneity, anatomical prior, MRI, brain image

1. INTRODUCTION

A fundamental operation in many applications of medical image analysis is image segmentation, the object of which is to associate with each image voxel a particular class based on its attributes, neighborhood information, or geometric characteristics of objects belonging to the class. This classification is then used by or to constrain higher-level image analysis and processing algorithms, thus robust and accurate image segmentation is a key element of many medical imaging applications.

In this work we consider the problem of segmenting magnetic resonance (MR) images, which is made difficult by the existence of partial volume (PV) averaging and intensity shading artifacts due to limited spatial resolution of the scanner and RF field inhomogeneity, respectively. To improve the quantitative precision of our segmentation, we develop a method for determining the fractional content of each tissue class for so-called partial volume voxels of mixed tissue type, taking into account shading artifacts. Of specific interest in the current work are the primary tissue constituents of the brain: gray (GM) and white matter (WM) as well as cerebrospinal fluid (CSF).

To our knowledge, two general approaches have been applied to address the problem of partial volume (PV) segmentation. A *mixel model*[1,2] assumes that every voxel in an image is a PV voxel, consisting of a mixture of pure tissue classes. The object of segmentation in this case is to determine the relative fraction of each tissue class present within every image voxel. Because of the number of parameters that must be estimated at each voxel, either multi-channel data and/or additional constraints are required to obtain the segmentation solution. A second approach[3,4] to dealing with PV voxels has been to marginalize over the variables describing the fractional portions of each *pure tissue class*. This produces an additional, new set of *partial volume classes*, with which each image voxel may be associated. However, an additional estimation step is necessary to obtain the fractional amount of the pure tissues in each voxel. In the current work, the latter method is used to adapt the maximum likelihood mixture model clustering algorithm[5–7] for segmentation of PV voxels in MR images of the brain.

Further author information: (Send correspondence to Aljaž Noe)
Aljaž Noe: aljaz.noe@uni-lj.si or noe@grasp.cis.upenn.edu
Stanislav Kovačič: stanislav.kovacic@fe.uni-lj.si
James C. Gee: gee@rad.upenn.edu

Image intensity artifacts due to RF field inhomogeneity can be either corrected before segmentation as a separate preprocessing task,[4,8,9] or during segmentation where intermediate segmentation results can be utilized. Further classification of shading correction methods is possible by differentiating between methods which model the bias field (usually of multiplicative nature) as a correction to the true intensity,[4,10,11] and methods which directly model the intensity variability as a function of voxel location in the image.[2,12] We chose the latter approach, since it is a direct extension of the existing maximum likelihood clustering algorithm.

Any additional information about the location of the tissues can improve the segmentation considerably. This prior information can be provided by registering a manually segmented image to the image that is being segmented.[7,10] This way for every voxel, some information about its tissue content is provided. However special care must be taken in these approaches as the results of a registration procedure are in general not always correct and can lead to poor segmentation results.

2. IMAGE MODEL

We generalize the image model proposed in [3,4] to account for tissue-dependent intensity variations. Experiments on MRI data show that differences in intensity variation across tissue type are not insignificant: the intensity values for CSF voxels always having the largest amount of variability, followed by GM and WM.

Let $\mathbf{I}_i = (I_{i,1}, I_{i,2}, \ldots I_{i,M})^T$ be the M-channel observation of the i-th voxel in an input image. Voxels of pure tissue class are described by a particular intensity distribution associated with the image appearance of that tissue type. Partial volume voxels, on the other hand, are represented as a linear combination of the intensity distributions associated with the K possible tissue types that can be found in those voxels:

$$\mathbf{I}_i = \sum_{k=1}^{K} t_{i,k} \mathbf{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad \sum_{k=1}^{K} t_{i,k} = 1, \quad i = 1 \ldots N, \tag{1}$$

where the voxel intensity $\mathbf{I}$ for pure tissue class k is represented as an M-element column vector of random variables, which are distributed according to a multivariate Gaussian distribution $\mathbf{N}$ with $\boldsymbol{\mu}_k = (\mu_{k,1}, \mu_{k,2}, \ldots, \mu_{k,M})^T$ the vector of mean class intensity values (M channels) for pure tissue class k, and $\boldsymbol{\Sigma}_k$ is the associated M by M covariance matrix for the M-channel observation. Term $t_{i,k}$ is the fraction of pure tissue class k that is present at the i-th voxel. There are N voxels in an image.

Note that the mean intensity values ($\boldsymbol{\mu}_k$) can also be a function of voxel location i, thus modeling shading artifacts in the MRI data. We will explain later in the paper how to model spatially dependent mean intensity values. For now let us just assume that the value of $\boldsymbol{\mu}_k$ is known for every voxel in an image. Variances ($\boldsymbol{\Sigma}_k$) on the other hand do not change with spatial location i.

2.1. Image model simplification

To determine the fractional amount of specified pure tissue classes within every image voxel, we must solve for $N \times (K-1)$ unknowns $t_{i,k}$ from N vector equations (1), one for each voxel and there are N voxels in the image. Since each vector $\mathbf{I}_i$ has M components, we have $N \times M$ equations. Assuming that the tissue class parameters ($\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$) are known, a solution can be found if $M \geq K-1$. In practice, we are interested in the three classes: CSF, GM and WM. Multi-echo images of high resolution are generally not available and even these would be partially correlated and noisy, so that the problem remains ill posed.

Additional constraints are therefore necessary and as in [3,4], we make the assumption that each partial volume voxel is a mixture of only two tissue types, which introduces negligible error in practical applications that use high-resolution T1 data. Formally, we define a number of sets $\mathcal{G}_k$ each containing indices of pure classes that are present in the k-th PV class:

$$\mathcal{G}_k = \{k_1, k_2\}, \quad k = 1 \ldots K_{PV}, \quad k_1, k_2 \in \{1 \ldots K\}, \tag{2}$$

where K_{PV} is the number of PV classes in an image while K is the total number of pure tissue classes. For voxels of pure tissue class k and PV voxels consisting of pure classes k_1 and k_2, respectively, (1) reduces to:

$$\mathbf{I}_i = \mathbf{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{3}$$

and

$$\mathbf{I}_i = t_{i,k_1} \mathbf{N}(\boldsymbol{\mu}_{k_1}, \boldsymbol{\Sigma}_{k_1}) + t_{i,k_2} \mathbf{N}(\boldsymbol{\mu}_{k_2}, \boldsymbol{\Sigma}_{k_2}), \quad t_{i,k_1} + t_{i,k_2} = 1. \tag{4}$$

3. MIXTURE MODEL CLUSTERING

To determine the parameters $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ for the pure tissue classes, an extended version of the maximum likelihood mixture model algorithm[5–7] was developed. First, appropriate probability density functions are described for the pure tissue and PV classes. Second, we introduce a Gibbs model as the weighting function to favor spatially extended classifications.[4] Finally, $P(k|\mathbf{I}_i)$ is determined and used to estimate the parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$.

3.1. Probability density functions

The intensities of voxels belonging to pure tissue class k are assumed to conform to a multivariate normal distribution: $\mathbf{I}_i = \mathbf{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. The corresponding probability density function for observing intensity $\mathbf{I}_i$ given tissue class k is therefore given by:

$$P(\mathbf{I}_i|k) = \frac{1}{\sqrt{(2\pi)^M |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\mathbf{I}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{I}_i - \boldsymbol{\mu}_k)\right). \tag{5}$$

The probability density function for PV voxels containing a mixture of pure tissue classes k_1 and k_2 is derived from (4) and involves a linear combination of two Gaussian distributions:

$$P_{PV}(\mathbf{I}_i|k_1, k_2, t) = \frac{1}{\sqrt{(2\pi)^M \left|\hat{\boldsymbol{\Sigma}}_k(t)\right|}} \exp\left(-\frac{1}{2}\big(\mathbf{I}_i - \hat{\boldsymbol{\mu}}_k(t)\big)^T \hat{\boldsymbol{\Sigma}}_k(t)^{-1}\big(\mathbf{I}_i - \hat{\boldsymbol{\mu}}_k(t)\big)\right), \tag{6}$$

$$\hat{\boldsymbol{\mu}}_k(t) = t\boldsymbol{\mu}_{k_1} + (1-t)\boldsymbol{\mu}_{k_2},$$

$$\hat{\boldsymbol{\Sigma}}_k(t) = t^2\boldsymbol{\Sigma}_{k_1} + (1-t)^2\boldsymbol{\Sigma}_{k_2}.$$

As in [3,4], we then marginalize (6) over t to obtain the probability density function for the PV classes:

$$P(\mathbf{I}_i|k) = \int_0^1 P_{PV}(\mathbf{I}_i|k_1, k_2, t)dt, \tag{7}$$

$$k = K+1\ldots K + K_{PV}, \quad k_1, k_2 \in \mathcal{G}_{k-K}, \quad k_1 \neq k_2.$$

To generalize the notation, we have numbered the PV classes from $K+1$ to $K + K_{PV}$, so that $P(\mathbf{I}_i|k)$ expresses the probability density for both pure tissue and PV classes. The integral in (7) does not have a closed form solution and must therefore be evaluated by numerical integration.

3.2. Weighting functions using neighborhood and prior anatomical information

In [5,6] the probability density function for class k is weighted by the current estimate of the voxel count for that class. This weighting is used to update the probabilities in a manner similar to that of a Bayesian prior. Here we introduce an alternative weighting function that favors segmentations which are spatially extended. Specifically, we use the slightly modified Potts model that is also applied in [4]:

$$P_i(k) = \frac{1}{Z} \exp\left(-\beta \cdot \sum_{j \in \mathcal{N}_i} \frac{\delta(k, k_j)}{d(i, j)}\right), \quad k = 1\ldots K + K_{PV}, \tag{8}$$

$$k_j = \arg\max_{k'}\big(P(\mathbf{I}_j|k')\big),$$

where

$$\delta(k_1, k_2) = \begin{cases} -2 & \text{if } k_1 = k_2 \\ +1 & \text{otherwise} \end{cases}; \quad k_1, k_2 \in \{1\ldots K + K_{PV}\}; \tag{9}$$

k_j is the current ML class estimate for the voxel at location j in the image; k is the class for which we are updating the weighting function; $\mathcal{N}_i$ is the set of D18 neighborhood voxels of voxel i; β is a parameter of the distribution, controlling the amount of influence the weighting function should exert on the solution; and Z is a normalizing constant. Function $d(i, j)$ represents the distance between voxels i and j, which limits the influence of distant neighborhood voxels.

Additional spatial information can be obtained from registered *class template images*, which contain prior class probabilities for every voxel in an image. One such template image is required for every class present in an image.

Class templates, which have been derived from MR images of a large number of subjects, were provided by the Montreal Neurological Institute.[13–15] As in [7], prior to segmentation, the class templates need to be registered to the image, which is being segmented. In this paper we do not focus on the registration procedure. We do, however assume, that the registration is not always perfect. Some limitation of the influence of class templates is therefore required, otherwise the errors in registration process will directly corrupt the segmentation solution.

We extend the spatial neighborhood based weighting function (8) to include the information from class templates:

$$P_i(k) = \frac{1}{Z} \exp\left(-\beta \cdot \sum_{j \in N_i} \frac{\delta(k, k_j) - \alpha Q_i(k)}{d(i, j)} \right), \tag{10}$$

where $0 \leq Q_i(k) \leq 1$ is the probability that the i-th voxel belongs to class k, according to the class template $Q_i(k)$. $Q_i(k)$ is assumed to be normalized so that they sum to unity over all classes. α is a constant that governs the influence of the class templates on the segmentation. Based on definition (9) of the *delta* function, we set $\alpha = 2$, to obtain an approximate equilibrium between the neighborhood and spatial class information while slightly favoring the former, which is based only on the current estimate of the class parameters. This way, even if the class template is completely wrong, the neighborhood information can accommodate this, resulting in a weighting function that still contains useful information.

3.2.1. Prior information for partial volume classes

Class templates only provide prior spatial information for pure classes. To avoid making PV classes a priori less likely, additional PV class templates need to be created. Note that class templates do not contain any specific information about location and amount of PV voxels in the image. We can however assume that PV classes are more likely to occur at locations where the prior probability for two pure classes is high. Therefore we base the prior PV information on the joint probability of the corresponding pure classes and transform it appropriately to provide the same amount of weighting as pure class templates:

$$Q_i(k) = 2\sqrt{Q_i(k_1)Q_i(k_2)}, \quad k_1, k_2 \in \mathcal{G}_{k-K}, \quad k = K + 1 \dots K + K_{PV}. \tag{11}$$

Since the class templates do not change, some parameter needs to be introduced providing information about the number of PV voxels in the image. We propose to raise the Q_i for pure classes to the power of γ and for PV classes (11) to the power of $\frac{1}{\gamma}$. This effectively widens ($\gamma > 1$) or shrinks ($\gamma < 1$) classes, for which $Q_i < 1$ (PV classes) while not affecting values close to 1 (pure classes).

3.3. Parameter estimation and intensity inhomogeneity correction

Given the probability density and weighting functions, the conditional probability $P(k|\mathbf{I}_i)$ is calculated as follows:

$$P(k|\mathbf{I}_i) = \frac{P_i(k)P(\mathbf{I}_i|k)}{\sum_{k'=1}^{K+K_{PV}} P_i(k')P(\mathbf{I}_i|k')}, \quad k = 1 \dots K + K_{PV}. \tag{12}$$

In the original maximum likelihood mixture model clustering algorithm[6] the likelihood is maximized when

$$\sum_{i=1}^{N} P(k|\mathbf{I}_i)\mathbf{\Sigma}_k^{-1}(\mathbf{I}_i - \boldsymbol{\mu}_k) = \mathbf{0}, \quad k = 1 \dots K, \tag{13}$$

which yields the following new estimates of the spatially independent $\boldsymbol{\mu}_k$ and $\mathbf{\Sigma}_k$ for each pure tissue class k:

$$
\begin{aligned}
\boldsymbol{\mu}_k &= \frac{1}{h_k} \sum_{i=1}^{N} P(k|\mathbf{I}_i) \cdot \mathbf{I}_i \\
\mathbf{\Sigma}_k &= \frac{1}{h_k} \sum_{i=1}^{N} P(k|\mathbf{I}_i) \cdot (\mathbf{I}_i - \boldsymbol{\mu}_k)(\mathbf{I}_i - \boldsymbol{\mu}_k)^T
\end{aligned}
\quad , \quad h_k = \sum_{i=1}^{N} P(k|\mathbf{I}_i), \quad k = 1 \dots K. \tag{14}
$$

To model the bias field, we let the class mean value be a smooth parametric function of voxel location i. Let us suppose that we can write this function as a linear combination of R scalar basis functions $\Phi_r(i)$, forming a vector

function $\mathbf{\Phi}(i) = (\Phi_1(i), \Phi_2(i), \ldots, \Phi_R(i))$. By proper definition of $\mathbf{\Phi}(i)$, we can represent any continuous function providing R is sufficiently large. A function representing the spatially dependent class mean values $\boldsymbol{\mu}_k(i)$ can now be written as:

$$\boldsymbol{\mu}_k(i) = \left(\mathbf{\Phi}(i)\mathbf{S}_k\right)^T, \quad \mathbf{S}_k = \begin{bmatrix} S_{k,1,1} & S_{k,1,2} & \ldots & S_{k,1,M} \\ S_{k,2,1} & S_{k,2,2} & \ldots & S_{k,2,M} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k,R,1} & S_{k,R,2} & \ldots & S_{k,R,M} \end{bmatrix}, \quad k = 1 \ldots K. \tag{15}$$

$\mathbf{S}_k$ is a matrix of or $R \times M$ parameters that define the class mean value function for the k-th pure class in the image. Practically this allows us to define a different bias field for every class on every channel of the image.

Since $\boldsymbol{\Sigma}_k$ does not depend on voxel location, we can left multiply (13) with $\boldsymbol{\Sigma}_k$, and substitute $\boldsymbol{\mu}_k$ with $\boldsymbol{\mu}_k(i)$:

$$\sum_{i=1}^{N} P(k|\mathbf{I}_i)(\mathbf{I}_i - \boldsymbol{\mu}_k(i)) = 0, \tag{16}$$

which can be approximated by N vector equations, equivalent to $N \times M$ scalar equations:

$$P(k|\mathbf{I}_i)\left(\mathbf{I}_i - (\mathbf{\Phi}(i)\mathbf{S}_k)^T\right) = 0, \quad i = 1 \ldots N. \tag{17}$$

By defining

$$\mathbf{A}_{[N \times R]} = \begin{bmatrix} \mathbf{\Phi}(1) \\ \mathbf{\Phi}(2) \\ \vdots \\ \mathbf{\Phi}(N) \end{bmatrix}, \quad \mathbf{B}_{[N \times M]} = \begin{bmatrix} \mathbf{I}_1{}^T \\ \mathbf{I}_2{}^T \\ \vdots \\ \mathbf{I}_N{}^T \end{bmatrix}, \quad \mathbf{W}_{k\,[N \times N]} = \begin{bmatrix} P(k|\mathbf{I}_1) & 0 & \ldots & 0 \\ 0 & P(k|\mathbf{I}_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P(k|\mathbf{I}_N) \end{bmatrix}, \tag{18}$$

we can rewrite (17) as

$$\mathbf{W_k}\left(\mathbf{B} - \mathbf{A}\mathbf{S}_k\right) = 0, \quad k = 1 \ldots K, \tag{19}$$

from which the solution for $\mathbf{S}_k$ can be expressed as a weighted least squares estimate:

$$\mathbf{S}_k = \left(\mathbf{A}^T \mathbf{W}_k \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{W}_k \mathbf{B}, \quad k = 1 \ldots K. \tag{20}$$

We can see that by setting $R=1$ and $\Phi_1(i)=1$, equation (20) yields the original ML class mean estimate (14) as defined in [5,6]. With the inhomogeneity model defined above, we are not limited to a specific function, nor does it matter whether the bias field is additive or multiplicative. Since we are estimating $R \times M$ shading field parameters, from $N \times M$ equations, the problem is well posed as long as $N \gg R$. A problem can, however, occur if we allow the bias field to contain frequencies that are too high. This yields an incorrect segmentation since the likelihood of the class estimates is larger, when the bias field compensates for changes in class mean values across the border of two classes. Other authors[2,12] deal with this problem by defining a model of bias field to implicitly include some constraints which limit the use of high frequencies in bias field.

Based on this limitation we choose polynomial basis functions of up to 4th order as in [9,10]. This results in $R = 34$ parameters that define the bias field. In our experience for most images polynomials of 2nd to 3rd order are adequate for bias field estimation. Polynomial basis functions of up to 2nd order are shown in Table 1.

$\Phi_1(i) = 1$	$\Phi_2(i) = x$	$\Phi_3(i) = y$	$\Phi_4(i) = z$	$\Phi_5(i) = x^2$
$\Phi_6(i) = xy$	$\Phi_7(i) = xz$	$\Phi_8(i) = y^2$	$\Phi_9(i) = yz$	$\Phi_{10}(i) = z^2$

Table 1. First and second order polynomial basis functions used for modeling the bias field. The independent variables x, y and z are normalized (to $[-1,+1]$) Cartesian image coordinates for the i-th voxel in the image.

In accordance with (15), we use these parameters to define new class mean function estimates $\boldsymbol{\mu}_k(i)$, which are then used to estimate new covariance matrices in a fashion similar to (14), except that the class mean values are now spatially varying:

$$\boldsymbol{\Sigma}_k = \frac{1}{h_k} \sum_{i=1}^{N} P(k|\mathbf{I}_i) \cdot (\mathbf{I}_i - \boldsymbol{\mu}_k(i))(\mathbf{I}_i - \boldsymbol{\mu}_k(i))^T, \quad h_k = \sum_{i=1}^{N} P(k|\mathbf{I}_i), \quad k = 1 \ldots K. \tag{21}$$

These estimates for class parameters $\boldsymbol{\mu}_k(i)$ and $\boldsymbol{\Sigma}_k$ then yield new probability density functions and the process is repeated until the voxel count in each pure tissue class does not change from one iteration to the next. This approximation to maximizing the likelihood[5,6] can also be seen as a special case of an Expectation Maximization (EM) algorithm suited for Gaussian distributions.

3.4. Initialization

Based on extensive experimentation on real and simulated MR images, we have found that the clustering algorithm can be made robust to initialization values by specifying a sufficiently large class variance. Therefore, without additional prior information about the mean and variance values of the classes, the parameters are initialized as:

$$
\begin{aligned}
\mu_{k,l} &= \frac{k}{K+1}\left(\max_i\left(I_{i,l}\right) - \min_i\left(I_{i,l}\right)\right) + \min_i\left(I_{i,l}\right) \\
\sigma_{k,l,l} &= \left(\frac{1}{K}\left(\max_i\left(I_{i,l}\right) - \min_i\left(I_{i,l}\right)\right)\right)^2, \quad k = 1\ldots K, \quad l = 1\ldots M
\end{aligned}
\tag{22}
$$

Initial mean intensity values are equally distributed between the minimum and maximum intensity values found in the image. Parameters for class mean functions which define spatial variability are set to zero – initially there is no shading correction. Diagonal elements of the covariance matrix are all set to the image intensity range divided by number of classes, whereas off-diagonal elements are set to zero.

4. PARTIAL VOLUME TISSUE CLASSIFICATION

The clustering algorithm determines $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ by iterating over the estimation of $P(k|\mathbf{I}_i)$ until convergence is achieved. Once the intensity distribution and all class parameters are known for each tissue type, the fractional portion t_{i,k_1} for a PV voxel at location i consisting of tissues k_1 and k_2 can then be obtained from (4) via maximum likelihood estimation (MLE):

$$
t_{i,k_1} = \frac{(\boldsymbol{\mu}_{k_1} - \boldsymbol{\mu}_{k_2})^T(\mathbf{I}_i - \boldsymbol{\mu}_{k_2})}{(\boldsymbol{\mu}_{k_1} - \boldsymbol{\mu}_{k_2})^T(\boldsymbol{\mu}_{k_1} - \boldsymbol{\mu}_{k_2})}.
\tag{23}
$$

To produce a segmentation without having to specify a threshold for distinguishing between partial volume and pure tissue voxels, we need to modify (23) to include the information about pure tissue classes. Using conditional class probabilities as weights, we can write:

$$
t_{i,k}^* = P(k|\mathbf{I}_i) + \sum_{k'} P(k'+K|\mathbf{I}_i)\frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k_2})^T(\mathbf{I}_i - \boldsymbol{\mu}_{k_2})}{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k_2})^T(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k_2})}, \quad t_{i,k} = \frac{t_{i,k}^*}{\sum_{k''=1}^K t_{i,k''}^*},
\tag{24}
$$
$$
k_2 \in \mathcal{G}_{k'} \wedge k_2 \neq k, \quad k = 1\ldots K,
$$

where summation index k' runs over all PV classes that contain pure class k (for which $k \in \mathcal{G}_{k'}$ is true). We must also normalize the portions of pure classes so that they sum to unity over all classes k.

5. IMPLEMENTATION

Prior to clustering and segmentation, we extract the brain parenchyma from the MR image of the head using the Brain Extraction Tool – details of the method can be found in [16].

A parallel version of the clustering algorithm was implemented by subdividing the image into a number of segments, which are then processed in separate threads, one for each processor available. All threads are synchronized at 4 time points: before and after the calculation of the weighting values (step 3) and before and after the estimation of the new class parameters (step 5). The algorithm is outlined below:

1. Initialization

   - Select number of pure tissue classes K.
   - Define PV classes, represented as sets $\mathcal{G}_k$, by specifying the corresponding combinations of pure tissue classes.

- Set initial estimates of class parameters $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ using (22).

2. Calculate the probability densities for all classes using (5) and (7) in multiple threads.

3. Calculate the weighting values in multiple threads using (10).

4. Calculate the updated probabilities using (12) for each class k in multiple threads.

5. Calculate the new estimates for the class parameters using (20), (15) and (21).

6. If the current voxel count for each pure tissue class is different from that found in the previous iteration, or the maximum allowed number of iterations have not been processed, return to step 2. We terminate the loop when the change in $\sum_{k=1}^{K} h_k$ between iterations is less than 1 or number of iterations iss 50. h_k is defined in (21).

7. Segment the image by determining the fractional amount of each tissue type for every image voxel using (24).

The algorithm was implemented in standard ANSI C code. Moreover, our multithreading support conforms to both the POSIX-PTHREADS standard and Microsoft Windows multithreading API, so the code will compile on both UNIX and Windows workstations.

6. EXPERIMENTAL RESULTS

The segmentation algorithm was evaluated using both synthetic and real data. In each of the reported experiments, β was set to 0.1 and algorithm convergence usually occurred after 10-20 iterations.

6.1. Synthetic Image

We constructed a square, 100 by 100, image and subdivided the image into 3 regions, each separated by a vertical boundary. The left and right most regions were considered pure "tissues" and their image values were drawn from normal distributions with the following mean and variance values, respectively: $\mu_1 = 70$, $\Sigma_1 = 10$ and $\mu_2 = 150$, $\Sigma_2 = 20$. The middle strip of the image, 30 pixels wide, contained partial volume pixels, which modeled a smooth linear transition between the two pure classes. The synthetic image is shown in Fig. 1.
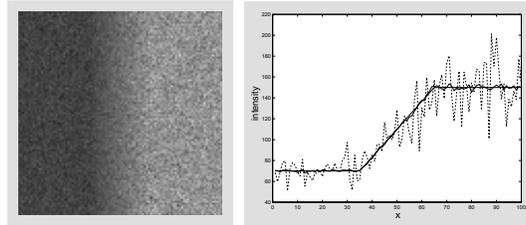
Figure 1. Synthetic data. (Left) Image to be segmented. (Right) Plotted in dashed line is the horizontal intensity profile obtained along line 50 of the image; solid thin line is the mean horizontal intensity profile averaged over all lines in the image; and in solid thick line is the ideal horizontal profile (without noise added).

Since we know the ideal segmentation t^{ideal}, the disparity between estimated and ideal t values was defined as the mean and standard deviation of the absolute difference over all pixels and classes:

$$E_{i,k} = \left| t_{i,k} - t_{i,k}^{ideal} \right|, \quad E_\mu = \frac{1}{NK} \sum_{k=1}^{K} \sum_{i=1}^{N} E_{i,k}, \quad E_\sigma = \sqrt{\frac{1}{NK-1} \sum_{k=1}^{K} \sum_{i=1}^{N} (E_{i,k} - E_\mu)^2}. \tag{25}$$

We also used the mean squared error (E_μ^2) to identify large errors in the segmentation.

$$E_\mu^2 = \frac{1}{NK} \sum_{k=1}^{K} \sum_{i=1}^{N} E_{i,k}^2 \tag{26}$$

The following are the estimated mean values and variances for the tissue classes: $\mu_1 = 70.38$, $\Sigma_1 = 10.07$; $\mu_2 = 148.28$, $\Sigma_2 = 19.25$. Mean and standard deviation of the absolute error were $E_\mu = 2.21 \times 10^{-2}$ and $E_\sigma = 4.81 \times 10^{-2}$, respectively. The segmentation results or t values for the first "tissue" class are shown in Fig. 2. The figure also shows the squared error between the ideal and estimated t values for the class – $E_\mu^2 = 5.623 \times 10^{-3}$.
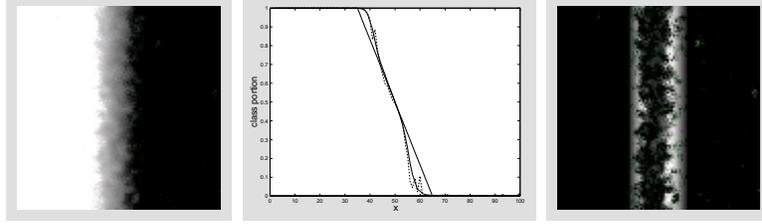

Figure 2. Segmentation results for the synthetic data. (Left) Fractional values t for the first class at each voxel plotted as an 8-bit gray-scale image with intensity $= 0$ corresponding to $t = 0.0$ and intensity $= 255$ to $t = 1.0$. (Middle) Plotted in dashed line is the horizontal intensity profile obtained along line 50 of the segmentation; solid thin line is the mean horizontal intensity profile averaged over all lines in the segmentation image; and thick line is the ideal horizontal profile. (Right) Image of point wise squared error between estimated and ideal t values for the first class

We can see that the errors occur only at the boundaries where the region with PV voxels meets the regions containing pure classes. We attribute this error largely to noise because it decreases when we reduce the amount of noise variance for the pure classes. This also explains the smaller amount of error in the segmentation of the left half of the image, where the noise variance for the first pure class was smaller.

6.1.1. Shading correction

To see and better understand the shading model, we created spatially dependent class mean values using two 2nd order polynomial functions. In the simple image used above, the difference in class intensities could easily be modeled by even linear polynomial function. We therefore needed to create a different synthetic image containing higher frequencies. We used the mirrored version of original image in x direction appended 3 times to the left, thus producing 400×100 image and applied spatially dependent class mean values. The resulting image with heavy shading is shown in Fig. 3.
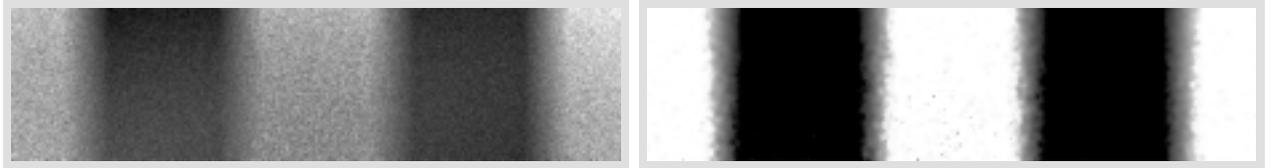
Figure 3. Shading correction. (Left) Synthetic image used for shading model evaluation. Class mean values are modeled by up to 2nd order polynomial functions – $\boldsymbol{\Phi}(i) = (1, x, y, x^2, xy, y^2)$, $\mathbf{S}_1 = (150, 10, -20, 35, -10, 10)^T$, $\mathbf{S}_2 = (70, -5, 15, -15, -17, -10)^T$. (Right) Segmentation results for first class.

Segmenting this image, we have observed great robustness to initial parameters (initial guess for bias field was zero for all classes) and accurately estimated the shading effect as long as the class mean functions could not compensate for different intensities between two classes. Error in the segmentation was $E_\mu = 1.85 \times 10^{-2}$, $E_\sigma = 4.26 \times 10^{-2}$ and the estimated mean function parameters were $\mathbf{S}_1 = (70.64, -4.81, 15.22, -15.27, -17.043, -10.77)^T$ and $\mathbf{S}_2 = (148.06, 9.58, -19.62, 34.44, -10.54, 10.52)^T$. Mean squared error $E_\mu^2 = 4.306 \times 10^{-3}$, which is of the same order as the result for the image without shading artifact.

By introducing a larger and more dynamic image, we limited the tendency for the mean function for one class to model different classes, which span across the x axis of the image. This effect can be examined by segmenting the original simple image (Fig. 1) with shading correction, which results in class mean values being very close together, while shading field models changes in voxel intensities between different classes. Segmentation results are in this case of course completely wrong and also largely depend on initial parameter estimates.

6.1.2. Prior anatomical information

Additionally we investigated the effect of class templates on the segmentation solution. We took the original true segmentation (t^{ideal}) as the pure class templates and set $\gamma = 10$ for PV classes. This extends the weighting function for PV class to almost all voxels for which $Q < 1$. The class templates are shown in Fig. 4.
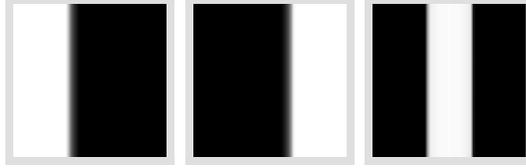
Figure 4. True class template for both pure classes (left, middle) and PV class (right), respectively, plotted as an 8-bit gray-scale image with intensity = 0 corresponding to $Q = 0.0$ and intensity = 255 to $Q = 1.0$.

We ran segmentation procedure on the synthetic image using various corrupted versions of the ideal templates. Segmentation results for different class templates are presented in Table 2. We can see that the spatial template based weighting function can have significant effect on quality of segmentation in both positive and negative sense. However, even if it is completely wrong, the segmentation procedure still finishes in correct minimum. Total error is of course larger than using no class templates at all, but the maximum possible error is limited. Wrong class templates were created from the ideal ones by changing the values of Q in such a way that classes, which don't exist at specific location had the highest spatial prior probability. This was achieved by subtracting the ideal class templates from 1. We also observed a strong correlation between the quality of class templates and the number of iterations needed for clustering.

Class template	$E_\mu \times 10^2$	$E_\sigma \times 10^2$	$E_\mu^2 \times 10^3$	$iter$
ideal	0.86	1.95	0.911	4
ideal + gaussian noise ($\sigma = 0.5$)	1.27	3.21	2.381	6
none	2.17	4.83	5.623	7
random noise (uniform, 0...1)	2.34	5.44	7.006	7
(1–ideal) + gaussian noise ($\sigma = 0.5$)	4.47	8.75	19.29	12
1–ideal	6.28	10.97	31.96	10

Table 2. Results of image segmentation using different class templates for spatial prior information. E_μ, E_σ and E_μ^2 are defined in (25) and (26). Last column shows the number of iterations needed for the clustering algorithm to converge.

6.2. Simulated T1-weighted Brain Volume

A second, more realistic synthetic dataset of an MRI head scan was created using the BrainWeb simulator.[17–20] Each simulation was a $1mm^3$ isotropic MRI volume with dimensions 181x217x181. Six datasets incorporating different amounts of noise and intensity inhomogeneity were segmented. The total mean squared error values between the ideal and estimated t values over all voxels are shown in Fig. 5 for the experiments.

noise	0%	3%	9%
0% shading	3.89	6.62	16.77
40% shading	4.15	6.89	19.86
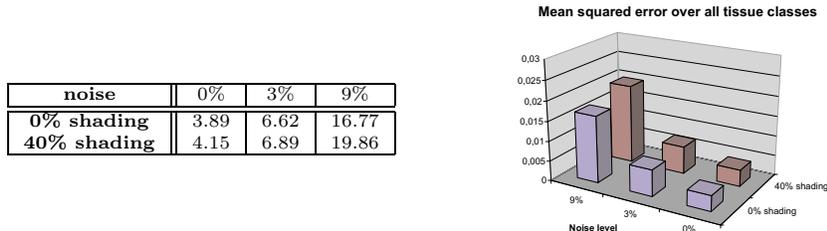
Figure 5. Total mean squared error ($E_\mu^2 \times 10^3$) for segmentation of BrainWeb simulated image. Results are shown for different amounts of noise present in the image, with and without shading artifacts.

We used polynomial functions of up to 2nd order to model shading artifacts in an image. Looking at the results in Fig. 5 and 6, we can see that there is little difference in error between images with and without shading proving that bias field was estimated correctly and estimation, using basis functions of higher order was unnecessary.
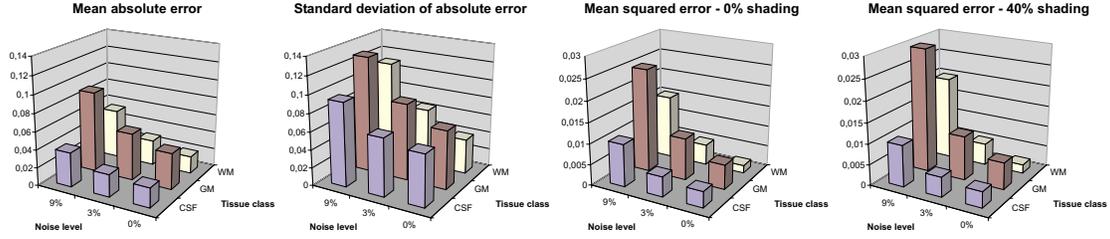
Figure 6. Errors in segmentation shown for CSF, GM and WM respectively. Mean absolute errors are plotted for images without shading artifacts.

A segmented slice of the synthetic brain volume with 9% noise level and 40% shading artifact is shown in Fig. 7. Although there appears to be minimal partial volume averaging in the results, the segmentation obtained without using PV classes (K_{PV}=0) had errors about 2 times larger and the algorithm took much longer to converge (> 50 iterations).
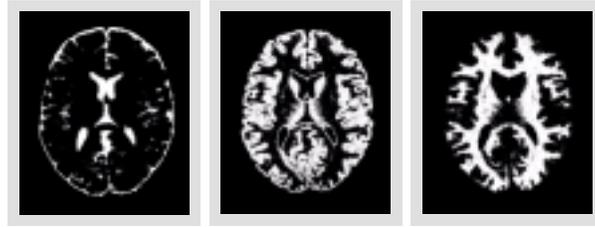
Figure 7. Partial volume segmentation of simulated brain volume, where fractional values t at each voxel are plotted as an 8-bit gray-scale image with intensity = 0 corresponding to $t = 0.0$ and intensity = 255 to $t = 1.0$. (Left) Cerebrospinal fluid. (Middle) Gray matter. (Right) White matter.

6.2.1. Using prior anatomical information on simulated MR images

We used the original reference segmentation of simulated MR images as prior spatial class information $Q_i(k)$ during the clustering procedure. This true prior information was first corrupted in various ways and segmentation results using corrupted class templates were compared. Specifically, we created 6 different class template images from the original ideal class template resulting in three images rotated 1, 5 and 15 degrees around the z (axial) axis, three images translated 1, 5 and 10 voxels. Additionally, we used the registered class templates used in the SPM segmentation algorithm.[7] These class templates have essentially the same source of origin as the BrainWeb simulated images and were downsampled to $91 \times 109 \times 91$ so that registration was an easy task.

Class template	$E_\mu^2 \times 10^3$	Class template	$E_\mu^2 \times 10^3$	Class template	$E_\mu^2 \times 10^3$
ideal	8.87	rotated 1deg	13.77	translated 1 voxels	15.47
spm prior	15.51	rotated 5deg	23.14	translated 5 voxels	26.26
none	16.77	rotated 15deg	24.82	translated 10 voxels	26.60

Table 3. Segmentation results of using prior anatomical information in the form of class templates on simulated MR images.

Image containing 9% noise and no shading artifact was used for evaluation. Parameters defining the use of prior information were set to: $\alpha = 2$ and $\gamma = 1$.

Comparing the results in table 3, we can see that the ideal prior information can significantly improve segmentation, while still limiting the error within acceptable limits when the prior information gets worse. We attribute this

to proper a balance between neighborhood and class template prior information. By changing the parameter α we can allow for larger class template prior influence, if we know that the prior information is correct. Segmentation results would improve correspondingly.

6.3. Manually Segmented Real T1 MR Images of the Brain

Twenty normal brain MRI datasets and their manual segmentations were obtained from the Center for Morphometric Analysis at Massachusetts General Hospital – these IBSR datasets are publicly available.[21] The volumes were preprocessed to extract brain parenchyma. No class templates were used during segmentation.

(a) (b) (c) (d)
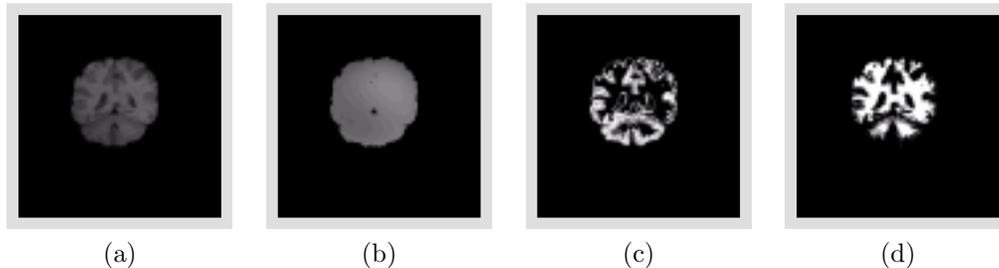
Figure 8. (a) Slice 18 of IBSR image 16_3. (b) Class mean function for GM. Segmented GM (c) and WM (d).

Since the manual segmentations for this set of images do not contain any information about fractional tissue content, we calculated a similarity index for each class by thresholding our partial volume segmentation results. Specifically, in table 4, we report the values for the Jaccard similarity $= |S_e \cap S_{ideal}|/|S_e \cup S_{ideal}|$, where S_e and S_{ideal} are the estimated and "true" sets of voxels, respectively, for a given tissue class.

Image	1_24*	100_23	11_3	110_3	111_2	112_2	12_3*	13_3	15_3*	16_3
GM	0.4620	0.8122	0.8124	0.7723	0.7745	0.7667	0.5925	0.7787	0.6229	0.6968
WM	0.4181	0.7471	0.7526	0.6862	0.7246	0.7010	0.5725	0.7219	0.5221	0.6315
Image	17_3	191_3	2_4*	202_3	205_3	4_8*	5_8*	6_10*	7_8	8_4
GM	0.6749	0.7909	0.5140	0.8150	0.8119	0.5682	0.5811	0.5941	0.7350	0.6891
WM	0.6297	0.7243	0.3907	0.7470	0.7644	0.4345	0.3977	0.3902	0.6917	0.6521

∗ Image for which intensity inhomogeneity could not be properly corrected.

Table 4. Jaccard similarity measure between estimated and "true" segmentation of IBSR datasets.

The shading effect was modeled using polynomial basis functions of up to 2nd order. However, some of the volumes exhibited strong, high frequency shading artifacts that could not be modeled by simple polynomial functions, resulting in a wrong segmentation. We have tried to increase the order of the polynomial basis functions (up to 4th order), but have observed little difference in the segmentation results. It is evident in the results in table 4 which of the volumes were the seven for which intensity inhomogeneity was still prevalent in the data. Excluding these volumes, the mean Jaccard index was 0.7639 and 0.7057 for GM and WM, respectively.

7. CONCLUSION

We have presented an algorithm for partial volume segmentation of MR images of the brain. Experimental results are comparable or superior to other published algorithms. Our method is an extension of a probabilistic clustering algorithm[5,6] to accommodate partial volume voxels and to allow class mean values to change across the image, thus modeling the shading effect. Although the convergence properties of the original technique are generally unknown, we have observed robust performance from our implementation as a function of the estimates used to initialize the class parameters and the proper use of neighborhood based prior information. Also examined were the possibilities of including the use of prior anatomic information to improve the accuracy of segmentation. Initial results of using this prior knowledge show that it can be used within the segmentation procedure in a robust way, providing useful

information if it is available. However, accurate and robust registration is required to fully utilize this prior information. We believe that using prior information is a fundamental step towards improvement of image segmentation and therefore is the focus of our future work.

REFERENCES

1. H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images - a mixel model," in *IEEE Transactions on Medical Imaging*, vol. 10, pp. 395–407, Sept. 1991.
2. L. Nocera and J. C. Gee, "Robust partial volume tissue classification of cerebral MRI scans," in *SPIE Medical Imaging*, K. M. Hanson, ed., *Proc. SPIE* **3034**, pp. 312–322, Feb. 1997.
3. D. H. Laidlaw, K. W. Flescher, and A. H. Barr, "Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms," in *IEEE Transactions on Medical Imaging*, vol. 17, pp. 74–86, Feb. 1998.
4. D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial voume model," 2000. Submitted.
5. J. A. Hartigan, *Clustering algorithms*, John Wiley & Sons, Inc., New York, 1975.
6. R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, John Wiley & Sons, Inc., New York, 1973.
7. J. Ashburner and K. Friston, "Multimodal image coregistration and partitioning - a unified framework," in *Neuroimage*, vol. 6, pp. 209–217, Oct. 1997.
8. J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," in *IEEE Transactions on Medical Imaging*, vol. 17, pp. 87–97, Feb. 1998.
9. B. Likar, M. A. Viergever, and F. Pernuš, "Retrospective correction of MR intensity inhomogeneity by information minimization," in *MICCAI'00*, S. L. Delp, A. M. DiGioia, and J. B., eds., vol. 1935 of *Lecture Notes in Computer Science*, pp. 375–384, (Pittsburgh, USA), Oct. 2000.
10. K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based bias field correction of MR images of the brain," in *IEEE Transactions on Medical Imaging*, vol. 18, pp. 885–896, Oct. 1999.
11. J. C. Rajapakse and F. Kruggel, "Segmentation of MR images with intensity inhomogeneities," in *Image and Vision Computing*, vol. 16, pp. 165–180, 1998.
12. M. X. H. Yan and J. S. Karp, "An adaptive bayesian approach to three-dimensional MR brain segmentation," in *Proc. of IEEE Med. Imaging Conf.*, (Norfolk, Virginia), Nov. 1994.
13. A. C. Evans, D. L. Collins, and B. Milner, "An MRI-based stereotactic atlas from 250 young normal subjects," in *Society for Neuroscience Abstracts*, vol. 18, p. 408, 1992.
14. A. C. Evans, D. L. Collins, D. L. Mills, E. D. Brown, R. L. Kelly, and T. M. Peters, "3D statistical neuroanatomical models from 305 MRI volumes," in *Proc. IEEE-nuclear science symposium and medical imaging conference*, pp. 1813–1817, 1993.
15. A. C. Evans, M. Kamber, D. L. Collins, and D. Macdonald, "An MRI-based probabilistic atlas of neuroanatomy," in *Magnetic resonance scanning and epilepsy*, vol. 264, pp. 263–274, 1994.
16. S. M. Smith, "Robust automated brain extraction," in *Sixth Int. Conf. on Functional Mapping of the Human Brain*, p. 625, 1998.
17. `http://www.bic.mni.mcgill.ca/brainweb/`.
18. C. A. Cocosco, V. Kollokian, R.-S. Kwan, and A. Evans, "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database," in *NeuroImage*, vol. 5, May 1997.
19. R.-S. Kwan, A. Evans, and G. B. Pike, *An Extensible MRI Simulator for Post-Processing Evaluation*, vol. 1131 of *Lecture Notes in Computer Science*, pp. 135–140. Springer-Verlag, May 1996.
20. D. L. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evans, "Design and Construction of a Realistic Digital Brain Phantom," in *IEEE Transactions on Medical Imaging*, vol. 17, pp. 463–468, June 1998.
21. `http://neuro-www.mgh.harvard.edu/cma/ibsr`.