MR Image Segmentation Using a Power Transformation Approach

7 authors, including:

Hong-Ren Su

35 PUBLICATIONS   299 CITATIONS

Philip E Cheng
Academia Sinica

61 PUBLICATIONS   1,147 CITATIONS

Michelle Liou
Academia Sinica

95 PUBLICATIONS   869 CITATIONS

Arthur C Tsai
Academia Sinica

45 PUBLICATIONS   510 CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Decision making processes involved in emotional inhibition View project

Project   Heart Rate Variability View project

MR Image Segmentation Using a Power Transformation Approach

Juin-Der Lee, Hong-Ren Su, Philip E. Cheng*, Michelle Liou, John A. D. Aston, Arthur C. Tsai, Cheng-Yu Chen

Abstract—This study proposes a segmentation method for brain MR images using a distribution transformation approach. The method extends traditional Gaussian mixtures expectation-maximization segmentation to a power transformed version of mixed intensity distributions, which includes Gaussian mixtures as a special case. As MR intensities tend to exhibit non-Gaussianity due to partial volume effects, the proposed method is designed to fit non-Gaussian tissue intensity distributions. One advantage of the method is that it is intuitively appealing and computationally simple. To avoid performance degradation caused by intensity inhomogeneity, different methods for correcting bias fields were applied prior to image segmentation, and their correction effects on the segmentation results were examined in the empirical study. The partitions of brain tissues (i.e., gray and white matter) resulting from the method were validated and evaluated against manual segmentation results based on thirty-eight real T1-weighted image volumes from the Internet Brain Segmentation Repository, and eighteen simulated image volumes from BrainWeb. The Jaccard and Dice similarity indices were computed to evaluate the performance of the proposed approach relative to the expert segmentations. Empirical results suggested that the proposed segmentation method yielded higher similarity measures for both gray matter and white matter as compared with those based on the traditional segmentation using the Gaussian mixtures approach.

Index Terms—Box-Cox transformation, EM algorithm, Gaussian mixtures, statistical segmentation.

I. INTRODUCTION

IN clinical and medical studies on brain anatomical structures, a successful partition of images into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) is often an important first step. With high spatial resolution and good soft-tissue contrast, brain images acquired via the MR techniques are well suited for this purpose. For example, neurodegeneration in schizophrenic patients as compared with a control group has been investigated using MR images to measure the volume changes of gray matter in thalamus, frontal and temporal lobes as well as CSF in ventricles [1-2]. It is also of considerable interest to study regional volumes of GM and WM across different developmental stages of the human brain [3-4]. Lesion segmentation is also another important application in clinical studies [5-6]. There is a body of research in which image segmentation is still supervised by professional experts on a slice-by-slice basis which is not only labor intensive but also introduces large intra- and inter-observer variability due to unresolved partial volume effects. Therefore, it is highly desirable to apply automatic and partial-volume corrected image segmentation when conducting group comparison and longitudinal studies based on a large volume of image data.

There has been a wide range of automatic segmentation methods proposed in the literature. For example, the thresholding approach partitions brain images using deterministic values that separate tissue types according to intensity levels [7-8]. Warfield et al. have developed a brain image segmentation algorithm which integrates the techniques of K-nearest neighbors and image registration [9]. The fuzzy-based approaches generalize the K-means algorithm to allow for soft segmentations such that each voxel can be assigned to more than one class of tissue [10-14]. Fwu and Djuric have also developed a tree-structure K-means algorithm to choose initial cluster centers in the original K-means algorithm [15]. On the other hand, the adaptive mean shift methods enable the integration of intensity and spatial features [16-17]. Glass et al. have used the Kohonen Map and multi-layered back-propagation neural networks for segmentation in inversion recovery (IR) images, which use the IR pulse sequence to provide strong contrast between tissues that have different T1 relaxation times [18]. Amato et al. have described a nonparametric discriminant analysis method for multi-spectral MR image segmentation [19].

The aforementioned methods all belong to the class of non-parametric segmentation methods. Another branch of segmentation methods relies on statistical classification. These approaches assume that mixed voxel intensities reflect distinct
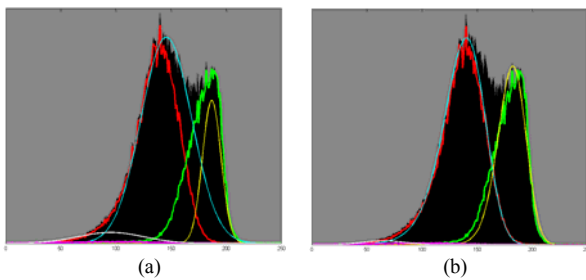
Fig. 1. Comparison of Gaussian mixtures and power-transformation mixtures on the IBSR image volume 13_3. (a) Histograms of CSF (magenta), GM (red), and WM (green) along with the estimated distributions for CSF (white), GM (cyan), and WM (yellow) using the Gaussian mixtures. (b) Histograms along with the estimated distributions using the power-transformed mixtures.

tissue groups, and individual voxels are assigned to different groups through modeling the intensity histogram as a mixture of probability distributions. For example, Ashburner et al. have proposed a modified version of the mixture model algorithm for analyzing spatial distributions of tissue groups [20-21]. Rajapakse et al. have proposed a Bayesian maximum a posteriori (MAP) model for image segmentation by incorporating Markov random fields (MRFs) as spatial priors for tissue distributions [22-23]. Zhang et al. have further developed a hidden MRF model to achieve a similar purpose [24]. Lemieux et al. have segmented T1-weighted brain images into intracranial CSF, GM and WM using morphological operations and the Gaussian mixtures [25].

The method proposed in this study falls into the class of statistical segmentation algorithms. Statistical methods have often relied on Gaussian assumptions for modeling the underlying distributions of CSF, GM and WM. However, empirical results indicate that tissue intensity distributions do not exhibit exactly a Gaussian shape. As one example, Figure 1(a) shows the tissue distributions based on manual partition of an image volume from the Internet Brain Segmentation Repository (IBSR; http://www.cma.mgh.harvard.edu/ibsr/). As indicated in the figure, the histograms of GM (red) and WM (green) are both skewed to the left. It indicates that the Gaussian mixtures fail to represent the skewed data. The GM component (cyan) is overestimated and consequently the WM component (yellow) is underestimated.

Due to limited imaging resolution, a brain voxel on the boundary of different brain tissues may consist of more than one tissue type, a phenomenon called the partial volume effect. The non-Gaussian features in MR intensity distributions mainly result from partial volume effects [26-27]. The partial volume effect is of importance in the extraction of brain tissue boundaries and quantitative measurement of brain structures, such as abnormal white matter multiple sclerosis and myelinated white matter [28-31]. Partial volume effects can be addressed by separating voxels containing multiple tissue types, called "mixels", as additional classes to pure voxel classes. The intensity of a mixel is modeled as the proportional summation of the intensities of the pure tissues within the mixel [32-37]. Cocosco et al. have presented a nonparametric method to segment brain images contaminated by partial volume effects

[38]. Bullmore has developed a logistic discriminant analysis to tackle the problem [26]. Instead of modeling mixel classes explicitly in addition to pure tissue classes, Ashburner et al. have assigned multiple Gaussians for each pure tissue class to fit the non-Gaussian intensity distribution of the pure tissue class [27].

By analogy, the method proposed in this study directly models the varying shape of tissue distributions in order to improve the final segmentation accuracy. Here we present a statistical power transformation approach to solve the problem of non-Gaussianity, without the cost of algorithmic complexity. The proposed approach extends Gaussian mixtures to power-transformed mixtures by incorporating a shape parameter, which is intrinsically useful for modeling non-Gaussian image data. In order to estimate the parameters that best fit the image intensity histogram, we suggest using a modified EM approach based on standard maximum likelihood principles.

However, statistical segmentation can be problematic due to existing within/between slice bias fields. Bias fields, also called intensity inhomogeneity (or nonuniformity), generally result from irregularities in both magnetic fields (B0) and RF excitation fields, and also from variability in the magnetization properties of brain tissues [23, 39]. The widely used models for estimating bias fields have been developed by treating the bias field as a smoothly varying multiplicative gain field. Images are simultaneously segmented while the bias field is estimated [40-41]. Other methods attempt to remove the inhomogeneity as a preprocessing step prior to actual segmentation. For example, Sled et al. have introduced a nonparametric nonuniform intensity normalization (N3) method which is independent of tissue classes and thus can be applied to the MR image before a tissue model is available [42]. As N3 addresses only the smooth intensity inhomogeneity, Su et al. have proposed a 3D wavelet-based bias correction method [43-45] which corrects for both the smooth intensity inhomogeneity and rapid intensity variations, which refers to the average intensity of one image slice being apparently higher/lower than another slice. The relative performance of these two methods, together with the proposed segmentation algorithm, will be examined.

It is often desirable to incorporate spatial priors into a tissue classification scheme. In the literature, the 3D Markov random field (MRF) has been proposed for tissue classification which assumes a Gibbs prior for the Gaussian mixtures [22-23]. In the empirical study, we will consider the Gibbs prior for the Gaussian mixtures and power transformed mixtures. In section II, we review the conventional EM algorithm with finite Gaussian mixtures. In section Ⅲ, we present the extended EM approach in detail and give a step-by-step algorithm for estimating model parameters. In section Ⅳ, we give segmentation results based on thirty-eight real image volumes from the IBSR, and eighteen simulated image volumes from BrainWeb [46-48]. The segmentation performance is summarized by the Jaccard and Dice indices for different methods. The final two sections contain discussion and

concluding remarks, followed by the appendix which details the derivation of the EM steps used in the computational algorithm.

II. GAUSSIAN MIXTURES AND EM SEGMENTATION

Let the random observation y_n be the intensity of the nth voxel in a brain region of interest. It is usually assumed that the intensity of the i-th tissue type follows a Gaussian distribution with mean value μ_i and variance σ_i^2. Let $\theta_i = (\mu_i, \sigma_i^2)$ and $\theta = \{\theta_1, \ldots, \theta_K\}$ denote the parameters of K Gaussian distributions. The Gaussian mixtures with K tissue types can be expressed as

$$p(y_n \mid \omega, \theta) = \sum_{i=1}^{K} \omega_i \cdot p(y_n \mid \theta_i),$$

where $\omega = \left\{ \omega_i \mid 0 \le w_i, \sum_i \omega_i = 1, i = 1 \ldots K \right\}$ are the weights of the tissue types, and

$$p(y_n \mid \theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{ -\frac{(y_n - \mu_i)^2}{2\sigma_i^2} \right\}. \qquad (1)$$

The unknown parameters $\{\omega = \{\omega_1, \ldots, \omega_K\}, \theta = \{\theta_1, \ldots, \theta_K\}\}$ in the finite mixture model can be estimated using the maximum likelihood estimation (MLE) method which is known to have desirable asymptotic properties if the parameter estimates are away from boundary values [49]. Let N be the total number of voxels, and $y = \{y_n \mid n = 1, \ldots, N\}$ be the set of voxel intensities. The intensities y_n are assumed to be independent, but the assumption will be relaxed when using spatial models such as the MRF model. The likelihood of the voxel intensity data y with unknown parameters is

$$\begin{aligned} L(y \mid \omega, \theta) &\equiv \prod_{n=1}^{N} p(y_n \mid \omega, \theta) \\ &= \prod_{n=1}^{N} \left\{ \sum_{i=1}^{K} \omega_i p(y_n \mid \theta_i) \right\}. \end{aligned} \qquad (2)$$

The ML estimates are obtained by solving for the parameters in the normal equations, which are derived from equating to zero the first partial derivatives of (2) with respect to the unknown parameters. For ease of exposition, hereafter, the same notations will be used for both the conditional probabilities and their estimates.

It has become a standard practice to use the EM algorithm to maximize (2). Let $p(i \mid y_n)$ be the posterior probability that the random observation y_n belongs to the i-th group. In the algorithm, the E-step updates this posterior probability given the latest estimates of ω and θ; that is, in the t-th iteration, compute

$$p(i \mid y_n, \omega^{(t)}, \theta^{(t)}) = \frac{\omega_i^{(t)} p(y_n \mid \theta_i^{(t)})}{\sum_j \omega_j^{(t)} p(y_n \mid \theta_j^{(t)})}. \qquad (3)$$

The likelihood equations (2) admit the posterior probability $p(i \mid y_n, \omega^{(t)}, \theta^{(t)})$; therefore, by inserting (3) into these equations, the M-step simply computes the unknown parameters as follows:

$$\omega_i^{(t+1)} = \frac{\sum_n p(i \mid y_n, \omega^{(t)}, \theta^{(t)})}{\sum_j \sum_n p(j \mid y_n, \omega^{(t)}, \theta^{(t)})} \qquad (4)$$

$$\mu_i^{(t+1)} = \frac{\sum_n p(i \mid y_n, \omega^{(t)}, \theta^{(t)}) \cdot y_n}{\sum_n p(i \mid y_n, \omega^{(t)}, \theta^{(t)})} \qquad (5)$$

and

$$\sigma_i^{2(t+1)} = \frac{\sum_n p(i \mid y_n, \omega^{(t)}, \theta^{(t)}) \cdot (y_n - \mu_i^{(t+1)})^2}{\sum_n p(i \mid y_n, \omega^{(t)}, \theta^{(t)})}. \qquad (6)$$

The EM algorithm simply iterates the E- and M-steps until the sequence of parameter estimates becomes stable. The segmentation is then obtained by assigning the nth voxel to the tissue type with the maximum posterior probability in (3).

It is desirable to incorporate spatial priors as contextual constraints in a tissue classification scheme to improve segmentation accuracy. The mixed Gaussian model combined with Gibbs priors, the so-called MRFs, has been proposed to take into account connectivity between neighboring tissue classes (e.g. [22-24]). Because brain voxels of the same tissue type are typically connected, MRFs are particularly suitable for modeling brain tissue topology. In addition to the independent voxel analysis given above, MRF analysis using the Gaussian mixtures will also be included in the empirical study.

III. POWER-TRANSFORMATION MIXTURES AND EM SEGMENTATION

Image segmentation methods based on mixture distributions essentially model the intensity data as a finite mixture of K tissue types. The segmentation based on the Gaussian mixtures may not identify individual tissue types accurately when the distributions are not symmetric. In applications, manual segmentations of real datasets by experts (e.g., IBSR) have indicated that the intensity distribution of each tissue type need not be symmetric. The intensity of individual tissue types may exhibit varied spreads or skewed shapes between subjects that may not be well fitted by a Gaussian distribution. Here, we present a power transformation model that can be incorporated within an EM procedure so as to increase the classification accuracy [50].

A. Power Transformation Distribution

In the literature, there are many transformations available for non-Gaussian noise (e.g., [51-52]). For conceptual simplicity, we consider power transformations. The power transformation

of a non-Gaussian positive-valued random variable is defined to be

$$x^{(\lambda)} = \begin{cases} \dfrac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log_e x, & \lambda = 0 \end{cases}$$

For appropriate λ, the transformed random variable $x^{(\lambda)}$ approximates a Gaussian shape:

$$p(x^{(\lambda)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-(x^{(\lambda)} - \mu)^2}{2\sigma^2} \right\}$$

where μ and σ^2 denote the mean and variance respectively of the transformed variable $x^{(\lambda)}$.

By variable transformation, we can express the PDF (probability density function) of x in terms of $p(x^{(\lambda)})$:

$$p(x \mid \lambda, \mu, \sigma^2) = p(x^{(\lambda)}) \cdot J = G(x^{(\lambda)} \mid \mu, \sigma^2) \cdot J \qquad (7)$$

where $G(\cdot)$ denotes the Gaussian density and $J = x^{\lambda-1}$ is the Jacobian of the transformation. The parameter λ controls the shape of distributions, that is, a distribution is symmetric ($\lambda = 1$), left-skewed ($\lambda < 1$) or right-skewed ($\lambda > 1$) depending on the values of λ. As a consequence, $p(x \mid \lambda, \mu, \sigma^2)$ in (7) is capable of fitting non-Gaussian data, while keeping the Gaussian as a special case.

B. Power Transformation Mixtures Model

Each tissue intensity distribution can be expressed as a power transformation distribution. The i-th tissue intensity distribution is described as:

$$p(y_n \mid \mu_i, \sigma_i^2, \lambda_i) = G(y_n^{(\lambda_i)} \mid \mu_i, \sigma_i^2) \cdot y_n^{(\lambda_i-1)} \qquad (8)$$

The input intensity $p(y_n)$ can be expressed as the mixtures of tissue intensity distributions:

$$p(y_n) = \sum_{i=1}^{K} \omega_i \cdot G(y_n^{(\lambda_i)} \mid \mu_i, \sigma_i^2) \cdot y_n^{(\lambda_i-1)} \qquad (9)$$

In the model setup, the unknown parameters are denoted by $\phi = \{\omega, \theta\}$, where $\omega = \{\omega_i \mid 0 \le \omega_i, \sum_i \omega_i = 1, i = 1,...,K\}$, and θ_i is redefined from equation (1) to include λ_i, thus $\theta = \{\theta_i = \{\mu_i, \sigma_i^2, \lambda_i\} \mid i = 1,\cdots,K\}$.

C. Parameter Estimation

The required parameter estimates are defined as those that maximize the likelihood function defined in (2). To increase the accuracy of the parameter estimation, we use a modified likelihood function. The posterior likelihood of ϕ given $\mathbf{y}$ is

$$L(\phi \mid \mathbf{y}) \propto \left\{ \prod_{n=}^{N} \sum_{i=1}^{K} \omega_i p(y_n^{(\lambda_i)} \mid \mu_i, \sigma_i^2) J_{n,i} \right\} \prod_{i=1}^{K} I(\omega_i \le b_i) \cdot \quad (10)$$

In applications, a constrained prior $\prod_i^K I(\omega_i \le b_i)$ is employed, where $I(E)$ is the indicator function of the event E. Reasonable values of b_i, the upper bound on the weight of the i-th tissue type, are assigned in order to reduce the chance that a particular tissue group with large ω_i dominates the likelihood. It is well known that Gaussian mixtures are affected by partial volume effects as the intensity of each individual voxel may exhibit a mixed type. The constrained prior proposed in (10) is designed to reduce the uncertainty of classifying individual voxel types. As an illustration, the implementation of (10) for the CSF can be carried out as follows. The CSF intensity on T1 images often spread across a wide range at the lower end of the intensity histogram and exhibits an overlap with the GM intensity, even though its volume content is relatively small, usually much less than ten percent of the common inner-skull brain area of interest (cf. IBSR data). A constrained prior $\omega_{CSF} \le 0.10$ can be used to improve the accuracy of segmentation between the CSF and GM.

The posterior likelihood can be maximized using the EM algorithm which treats the membership of voxels as unobserved latent values. In the t-th iteration, the E-step computes

$$p(i \mid y_n, \phi^{(t)}) = \frac{\omega_i^{(t)} G(y_n^{(\lambda_i^{(t)})} \mid \theta_i^{(t)}) \cdot y_n^{\lambda_i^{(t)}-1}}{\sum_j \omega_j^{(t)} G(y_n^{(\lambda_j^{(t)})} \mid \theta_j^{(t)}) \cdot y_n^{\lambda_j^{(t)}-1}}.$$

The M-step updates ϕ estimates using their weighted sufficient statistics and the $p(i \mid y_n, \phi^{(t)})$ value. The membership proportions are computed using

$$\omega_i^{(t+1)} = \frac{\sum_n p(i \mid y_n, \phi^{(t)})}{\sum_j \sum_n p(j \mid y_n, \phi^{(t)})}. \qquad (11)$$

In case any $\omega_i^{(t+1)}$, say $\omega_{CSF}^{(t+1)}$, falls out of the given boundary b_i, $\omega_i^{(t+1)}$ is re-estimated to satisfy the constrained prior under the condition $\sum_i \omega_i = 1$:

$$\omega_{CSF}^{(t+1)} = b_{CSF},$$

$$\omega_{i \neq CSF}^{(t+1)} = \frac{\sum_n p(i \mid y_n, \phi^{(t)})}{\sum_{j \neq CSF} \sum_n p(j \mid y_n, \phi^{(t)})} \cdot (1 - b_{CSF}).$$

The $\mu_i^{(t+1)}$ and $\sigma_i^{2(t+1)}$ estimates are computed as follows:

$$\mu_i^{(t+1)} = \frac{\sum_n p(i \mid y_n, \phi^{(t)}) y_n^{(\lambda_i^{(t)})}}{\sum_n p(i \mid y_n, \phi^{(t)})}, \qquad (12)$$

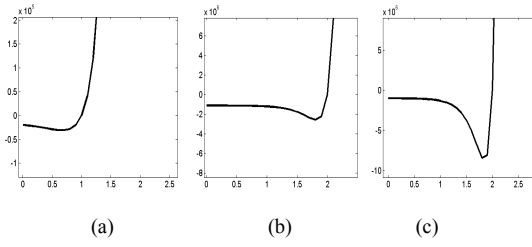<div align="center">(a)   (b)   (c)</div>

Fig. 2. Typical plots of (a) $g(\lambda_{csf})$, (b) $g(\lambda_{gm})$, and (c) $g(\lambda_{wm})$.

$$\sigma_i^{2(t+1)} = \frac{\sum_n p(i \mid y_n, \phi^{(t)})(y_n^{(\lambda_i^{(t)})} - \mu_i^{(t+1)})^2}{\sum_n p(i \mid y_n, \phi^{(t)})}. \qquad (13)$$

It has been well known that there is no closed-form solution to $\lambda_i^{(t+1)}$ [52], and the estimates for λ_i are updated by solving the following equation for $\lambda_i^{(t+1)}$, in which the other parameters are fixed at the updated values:

$$g(\lambda_i^{(t+1)}) =$$
$$\sum_n p(i \mid y_n, \phi^{(t)}) (\frac{-(y_n^{(\lambda_i^{(t+1)})} - \mu_i)}{\sigma_i^2} \cdot \frac{(\log y_n) y_n^{\lambda_i^{(t+1)}} \lambda_i + (y_n^{\lambda_i^{(t+1)}} - 1)}{\lambda_i^{2(t+1)}} + \log y_n) = 0 \qquad (14)$$

Equation (14) is a typical case of using (bounded) profile likelihood estimation with finitely many nuisance parameters [53]. Figure 2 shows typical plots of $g(\lambda_i^{(t+1)})$. As shown in the plots, $g(\lambda_i^{(t+1)})$ is a continuous single-root function whose root can be solved by the bisection root-finding method [54]. Based on empirical experiments, we used an initial range of [0, 5] for the bisection method for all of our validation data (i.e., real and simulated data).

Detailed derivations of terms in (11)-(14) can be found in the appendix. The E- and M-steps are iterated until the sequence of parameter estimates becomes stable. After the EM algorithm converges, the parameters $\{\lambda_i, w_i, \mu_i, \sigma_i^2 \mid i = 1 \cdots K\}$ which maximize the likelihood function are then used to classify image voxels. Figure 1(b) shows the fitted tissue distributions obtained from the power transformation approach to the same image data used in Figure 1(a). It is clear that the power transformed mixtures fit the skewed data distributions more accurately than the standard Gaussian mixtures.

The segmentation algorithm can be summarized as the following pseudo-code:

set K;
initialize :
$\phi^{(0)} = \{\omega^{(0)} = \{\omega_i^{(0)} \mid i = 1...K\}, \theta^{(0)} = \{\mu_i^{(0)}, \sigma_i^{2(0)}, \lambda_i^{(0)} \mid i = 1...K\}\}$.
t=0
do
  E-step:
    calculate $p(i \mid y_n, \omega^{(t)}, \theta^{(t)})$.
  M-step:
    estimate

$\phi^{(t+1)} = \{w_i^{(t+1)} \mid i = 1...K\}, \theta^{(t+1)} = \{\mu_i^{(t+1)}, \sigma_i^{2(t+1)}, \lambda_i^{(t+1)} \mid i = 1 \cdots K\}$
until (convergence of the sequence of parameters is reached)

Compute the membership probability $p(i \mid y_n)$ with the estimated parameters.
Assign each voxel y_n to the k-th class iff $k = \arg\max_i p(i \mid y_n)$.

It is also interesting to consider the spatial priors in the power-transformed mixtures. In the empirical study, we assume the Gibbs prior in the power transformed mixtures (i.e., replacing the Gaussian mixtures in Equation (6) of [22] with the power-transformed mixtures).

IV. EMPIRICAL STUDY

The IBSR has originally offered a collection of twenty real 3D T1-weighted image volumes with complete expert segmentations. These twenty volumes (i.e., 20 subjects' brain images) of coronal spoiled gradient echo MRI scans were acquired with two imaging systems. The FLASH scans were performed on a 1.5-Tesla Siemens Magnetom MR system (Iselin, NJ) with the following parameters: TR = 40 msec, TE = 8 msec, flip angle = 50 degrees, field of view = 30 cm, slice thickness = contiguous 3.1 mm, matrix = 256x256. The other 3D-CARRY scans were performed on a 1.5-Tesla General Electric Sigma MR system (Milwaukee, WI) with the following parameters: TR= 50 msec, TE = 9 msec, flip angle = 50 degrees, FOV = 24cm, slice thickness = continuous 3 mm, matrix = 256x256. These images have been partitioned into off-brain structures, CSF, GM, and WM by trained investigators using a semi-automated intensity contour mapping algorithm [55]. The off-brain voxels have been removed, and segmented CSF, GM and WM have been recorded for public reference. For the ease of exposition, we classify the twenty image volumes into either an uncontaminated group or a contaminated group depending on a preprocessing procedure using N3 [42]. The image volumes that have not shown significant differences between pre- and post-processing results are grouped into the uncontaminated dataset I, and otherwise grouped into the contaminated dataset II, including those with rapid inter-slice intensity variation that cannot be well corrected by N3.

The IBSR has newly supported an additional collection of eighteen T1-weighted FLASH MRI volumes of better quality and with slice thickness of 1.5mm, but other than expert segmentations, the scanning parameters are not given. Those additional real datasets have been corrected by the IBSR for bias fields using the CMA 'autoseg' biasfield correction routines (an interested reader may refer to the IBSR site for details). In the sequel, we will refer the bias-corrected image volumes as the preprocessed dataset III. The simulated image volumes supported by BrainWeb have adopted the International Consortium for Brain Mapping (ICBM) template with known tissue types. The central peaks of the template distributions are much higher than those of the Gaussian mixtures. The proportions of Gaussian random noise can be

generated by the users with six degrees of noise contamination (i.e., pn0, pn1, pn3, pn5, pn7 and pn9). With added Gaussian noise, the simulated tissue distributions become nearly Gaussian. On the other hand, the intensity non-uniformity (bias fields) can be specified at three different levels (i.e., rf0, rf20, and rf40). A combination of Gaussian random noise and bias fields gives eighteen simulated image volumes having voxel dimension of $1 \times 1 \times 1$ mm. We will refer to those simulated image volumes as the simulated dataset IV. Our method does not contain the process of removing non-brain voxels. For the dataset III and IV, we obtained the skull-stripped volumes by masking out non-brain voxels using their corresponding manual segmentations.

N3 has been shown to be one of the leading performers for correcting bias fields [39, 56] and has virtually become the standard method against other inhomogeneity correction methods [57]. However, N3 does not take the rapid inter-slice intensity variations into account. In the empirical study, we compare not only N3 but also a 3D wavelet-based bias correction method [43-45] which is useful for correcting not only the smooth intra-slice inhomogeneity, but also the rapid inter-slice intensity variation. Specifically, a regular 3D wavelet transformation may not work well for correcting rapid inter-slice intensity variation and smooth intra-slice inhomogeneity simultaneously because bias fields are attributable to spatial signals in lower frequencies, and inter-slice intensity variation are to those of higher frequencies. In the 3D wavelet-based method, two separate wavelet transformations have been adopted with one filter bank to correct for the inhomogeneity within 2D images, and with another filter bank along the slice direction to correct for the rapid inter-slice intensity variation.

In the intra-slice inhomogeneity correction, a complete decomposition with the wavelet transform is carried out on each slice. For example, if the slice has dimensions 128x128, then a seven level decomposition is performed. The coarsest resolution coefficient (the grand average) within each 2D slice is preserved for maintaining image scales. The next 2 levels of coarse resolutions are set to zero to remove the intra-slice inhomogeneity. The remaining levels are preserved intact.

Only one 1D wavelet transform has been performed in the inter-slice direction. This is applied to the wavelet coefficients corresponding to the grand average of each 2D wavelet decomposition. Again a full decomposition is performed on these coefficients. The top 4 levels in the decomposition corresponding to the coarsest components are this time retained (including the grand average in the inter-slice direction), and the rest of the levels are set to zero to remove the high frequency components. These correspond to fluctuations in the grand averages between slices, which are the cause of the inter-slice intensity variation. The wavelet decomposition has been applied to the log intensity scale images and these are transformed back to the original scale after bias correction and inter-slice normalization.

To evaluate the relative performance between segmentation methods, the standard Jaccard similarity indices [35] have been

calculated. The index compares the results between manual segmentations (provided by the IBSR) and automatic segmentations. Specifically, the Jaccard index, which measures the similarity between two sets, is defined as

$$\text{Jaccard Index} = \frac{\left| S_a^i \cap S_m^i \right|}{\left| S_a^i \cup S_m^i \right|},$$

where S_a^i and S_m^i denotes the voxels of the i-th tissue type classified by our automatic method and the manual method, respectively. The value of Jaccard index lies between 0, when the two sets have no common elements, and 1, when the two sets are identical. The Dice index has also been widely used for comparing between existing methods in the literature [58]. There is a one-to-one correspondence between the two similarity indices, that is, Dice Index = 2(Jaccard Index)/[1+(Jaccard Index)]. In order to compare with these published results, the Dice indices have been calculated for evaluating the eighteen simulated image volumes (provided by the BrainWeb).

In the empirical study, the tissue types of interest comprise CSF, GM and WM, and thus K was fixed at three. This could be adapted to other K values in applications, for example, by taking into account intermediate tissue types (e.g., between GM and WM) due to partial volume effects. We selected the same constraint boundaries as $b = \{b_{csf} = 0.1, b_{gm} = 0.65, b_{wm} = 0.5\}$ across datasets. In the initialization step of the EM algorithm, the tissue proportions ω were arbitrarily set to be $\omega^{(0)} = \{\omega_{csf} = 0.05, \omega_{gm} = 0.50, \omega_{wm} = 0.45\}$ for the three tissue types. Initial values for λ were selected to be $\lambda^{(0)} = \{\lambda_{csf} = 1, \lambda_{gm} = 2, \lambda_{wm} = 2\}$ due to a common trend of skewed GM and WM intensities. Image intensities of voxels were sorted according to their magnitude and partitioned into three groups according to $\omega^{(0)}$. The initial values of μ_i and σ_i^2 were simply set to be the sample mean and variance of the i-th group intensity, transformed according to $\lambda_i^{(0)}$, respectively.

A. Evaluation with the Uncontaminated Dataset I

Figure 3 shows the Jaccard indices of GM and WM segmentations of the ten uncontaminated image volumes produced by the proposed model and Gaussian mixtures. As a comparison, the figure also gives results from the MRF model with different combinations of Gaussian mixtures and power-transformed mixtures. Across the ten image volumes, the average Jaccard indices for GM and WM are 0.69 and 0.55 based on the Gaussian mixtures, and 0.82 and 0.73 based on the power-transformed mixtures, respectively.

Figure 3(b) lists Jaccard indices of WM segmentations produced by Gaussian-MRFs proposed by Rajapakse et al. [22] and power trans-MRFs, in which the Gaussian model is replaced by its power transformed counterpart. The average Jaccard index of WM segmentations is slightly improved from 0.71 to 0.72 by replacing Gaussian-MRFs with power trans-MRFs. On the other hand, the average indices of GM

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TMI-2008-0279.R1   7
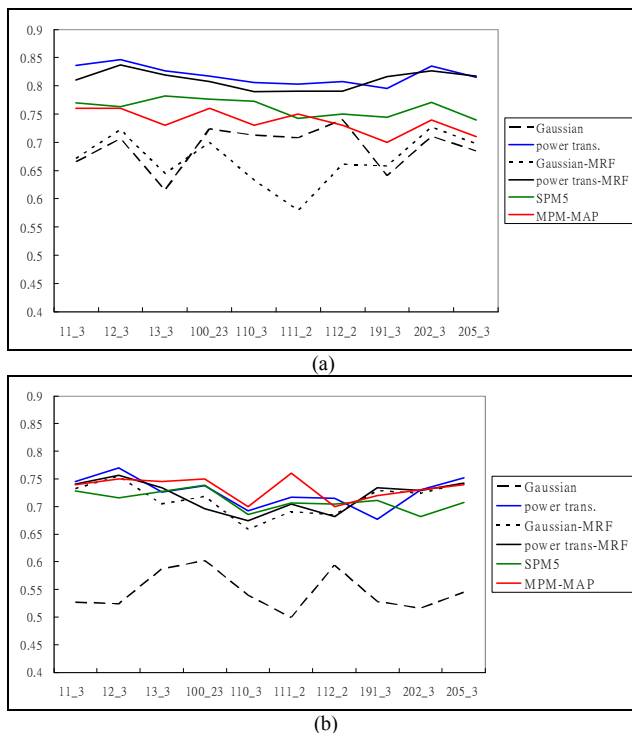
(a)

(b)

Fig. 3. (a) The Jaccard indices of GM segmentation on the uncontaminated dataset I. The average Jaccard indices are: Gaussian=0.69, power trans.=0.82, Gaussian-MRF=0.67, and power trans-MRF=0.81, SPM5=0.767, MPM-MAP=0.74. (b) The Jaccard indices of WM segmentation. The average Jaccard indices are: Gaussian=0.55, power trans.=0.73, Gaussian-MRF=0.71, and power trans- MRF=0.72, SPM5=0.71, MPM-MAP=0.73.

segmentations in 3(a) suggest that the improvement is substantial, that is, the average index for power trans-MRFs is 0.81 and for Gaussian-MRFs, is 0.67. This variation can be attributed to the different spatial complexities of GM and WM. Although for both GM and WM tissues, voxels of the same tissue type are connected to one another, GM is inherently much more tortuous than WM. Together with the fact that the Gaussian model does not accurately represents the intensity distributions, the Gaussian-MRF tends to over-smooth the segmentation. On the other hand, the better performance of the power trans-MRF over Gaussian MRF can be attributed to its accurate representation of intensity distribution. However, due to the geometric complexity of GM, the power trans-MRF does not outperform the proposed power transformation method.

In the literature, the atlas-based procedures have also been recommended for image segmentation. With those methods, partitioning of MR images into tissue types are supervised by the prior segmentation of a large number of real MR images. As a comparison, Figure 3 also shows the segmentation results using the SPM5 and MPM-MAP methods [27, 59]. Both methods register image volumes to a standardized atlas before segmentation and can be considered as modified versions of Gaussian mixtures and Gaussian MRFs, respectively, by including a prior on the spatial location of tissue types. The values of MPM-MAP in Figure 3 are based on published results [59], and those of SPM5 are based on program implementations.
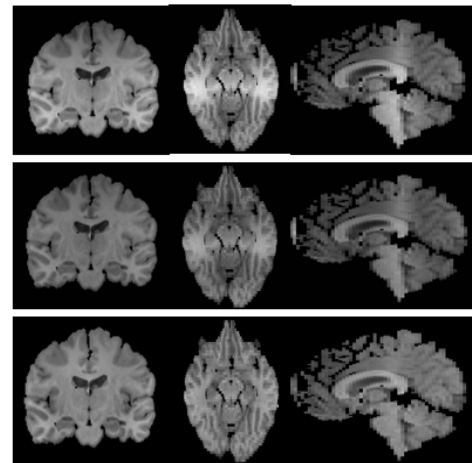

Fig. 4. The upper row: one MR slice (viewed from different directions) in the contaminated image volume 4_8. The middle row: the same slice after preprocessing using N3. The lower row: the same slice after the wavelet-based bias correction.

When implementing SPM5, the bias regularization parameter was set to "no regularization" for the Dataset I, and other parameters, to their default values. By considering a-priori information, MPM-MAP slightly improves upon the Gaussian-MRF, and SPM5 drastically improves upon the conventional Gaussian mixtures. On the average, however, the power-transformed mixtures still outperforms other competing methods in classifying GM voxels.

B. Evaluation with the Contaminated Dataset II

The ten contaminated image volumes contain either the smooth intensity inhomogeneity, or rapid inter-slice intensity variation. To keep the segmentation method simple, we did not estimate the bias field within the proposed method. Instead, we incorporated a bias correction method to remove contamination before image segmentation.

Figure 4 shows the results of applying the N3 and the wavelet-based method to the image volume 4_8. It gives representative orthogonal views of this contaminated dataset before and after preprocessing. This image volume gives a classic example of rapid inter-slice intensity variation — the slices in the posterior region of the brain are much brighter than the other slices. As the middle row shows, the rapid inter-slice intensity variation remains presented in the images processed by N3.

Our segmentation method was applied to the preprocessed images. The parameters were initialized in the same way as with the uncontaminated dataset I. The Jaccard indices of the segmented GM and WM against manual segmentation are shown in Figure 5 which also shows the ensuing Jaccard indices for GM and WM segmentation produced by Gaussian mixtures. For the wavelet based bias correction method, the average Jaccard indices achieved by Gaussian mixtures are 0.60 and 0.59 for GM and WM, respectively. The average Jaccard indices achieved by the power-transformation method are increased to 0.75 and 0.65 for GM and WM, respectively.

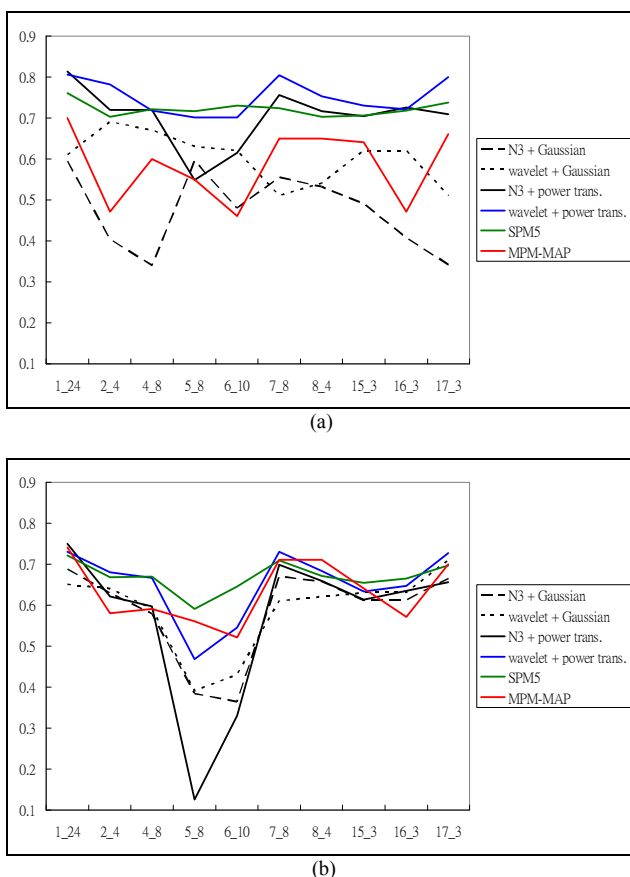Also shown in Figure 5 are Jaccard indices achieved by

(a)

Fig. 5. (a) The Jaccard indices of GM segmentation of the contaminated dataset II. The average Jaccard indices are: N3 + Gaussian= 0.47, wavelet + Gaussian = 0.60, N3 + power trans.= 0.70, wavelet + power trans.=0.75, SPM5=0.72, MPM-MAP=0.59. (b) The Jaccard indices of WM segmentation of the contaminated dataset II. The average Jaccard indices are: N3 + Gaussian= 0.58, wavelet + Gaussian = 0.59, N3 + power trans.= 0.56, wavelet + power trans.= 0.65, SPM5=0.67, MPM-MAP=0.63.

Gaussian mixtures and the power-transformed mixtures using N3 and wavelets in the preprocessing stage. The average Jaccard indices of Gaussian mixtures for GM and WM are 0.47 and 0.58, and of the power-transformation method are 0.70 and 0.56, respectively. The results of MRFs as compared with those of the power-transformed approach are consistent with the uncontaminated image volumes. The values of SPM5 (the bias regularization parameter was selected to be "medium"), and MPM-MAP are also listed for the contaminated images, and the results suggest that SPM5 is more robust to bias contamination relative to MPM-MAP.

Figure 6 shows segmentation results for the image volume 1_24. The power transformation approach took the wavelet bias-corrected image (a) as the input image, producing estimated GM probability image (b) and WM probability image (c). The final segmentation (d) was formed by comparing probability images (b) and (c) on a voxel-by-voxel basis and assigning each voxel to the tissue class with the maximum probability.

C. Evaluation with the Preprocessed Dataset III

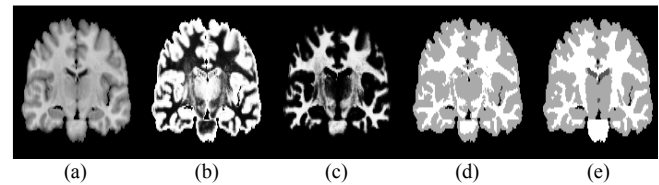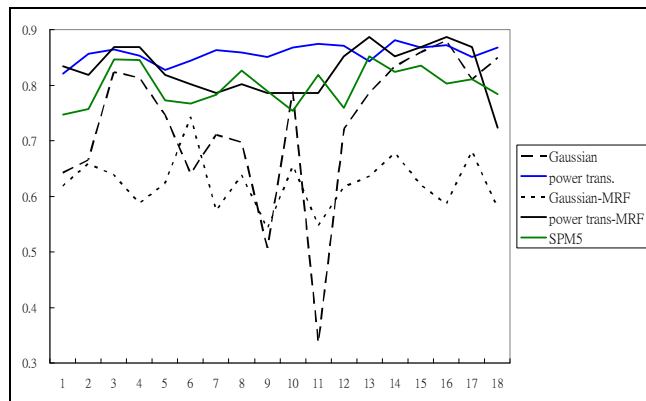Figure 7 gives the segmentation results for the eighteen

Fig. 6. The ensuing images of the image volume 1_24. (a) The bias-corrected image is the input to the power transformation approach. (b) Estimated GM probability image. (c) Estimated WM probability image. (d) Segmentation formed by comparing probability images (b) and (c) on a voxel-by-voxel basis and assigning each voxel the tissue class with the maximum probability. (e) Manual segmentation.
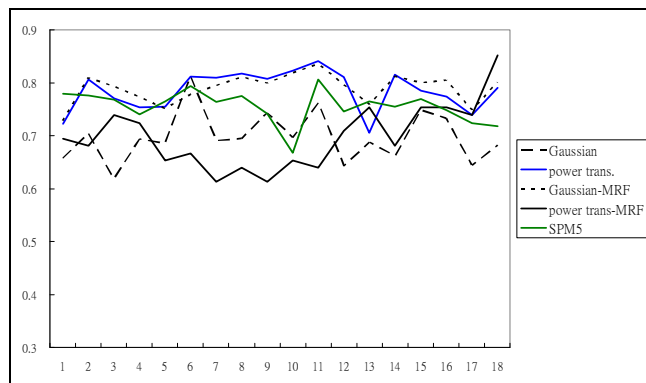
newly added image volumes from the IBSR. Because the preprocessed images have been corrected for bias, the SPM5 bias correction function was not applied to the Dataset III. On the average, the power transformation approach still outperforms the existing methods for both GM and WM segmentations, including conventional Gaussian mixtures, MRFs with either the Gaussian or non-Gaussian assumption, and the atlas-based method (i.e., SPM5). It is interesting to note that the Gaussian MRF model gives comparable results as those by the power transformation approach for the WM segmentation. However, the power-transformed MRF model gives similar results as those by the power transformation approach for the GM segmentation. This finding is consistent with the results in Figure 3. The power trans-MRF method performs less well compared with other competing methods in WM segmentation for a few image volumes in this dataset. When comparing between results in Figures 3 and 7, it is interesting to note that the average accuracy of segmentation methods is improved as the quality of image data becomes better. This improvement is more notable for the power transformation approach than for other competing methods.

D. Evaluation with the Simulated Dataset IV

As was mentioned, the template distributions are leptokurtic and become nearly Gaussian with added random noise. The bias fields in the simulated dataset were corrected using N3 before computing all methods except for SPM5. The "bias regularization" parameters in SPM5 were selected to be "none", "medium", and "heavy" when segmenting the rf0, rf20, and rf40 image volumes, respectively. In Figure 8 we observe that the Gaussian and power-transformed MRF models outperform other methods for those image volumes seriously contaminated by random noise (i.e., pn5 – pn9). The proposed power transformation approach also works for the leptokurtic distributions. It is interesting to note that the power-transformed MRF performs better than its Gaussian counterpart for those pn0 and pn1 image volumes. In the figure, the Dice values for the well-known EM approach KVL are published results in Leemput et al. [60]. The average Dice indices of the eighteen volumes for GM are: Gaussian mixtures, 0.89, power transformed mixtures, 0.93, Gaussian-MRF, 0.94, power trans-MRF, 0.94, and SPM5, 0.88. The values for KVL and MPM-MAP are published results [59], and their averages are 0.90 and 0.92, respectively, on the basis of those image volumes available in the literature. The average Dice indices

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

TMI-2008-0279.R1   9
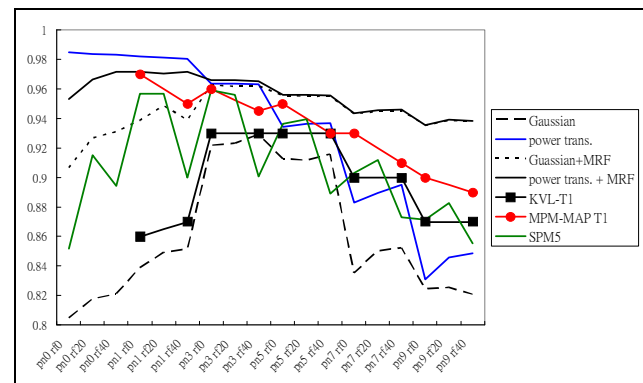
(a)

(b)

Fig. 7. Jaccard indices for the preprocessed dataset III. (a) Average Jaccard indices for GM are: Gaussian=0.73, power trans= 0.86, Gaussian-MRF= 0.62, power trans-MRF=0.83, SPM5=0.8. (b) Average Jaccard indices for WM are: Gaussian=0.70, power trans=0.79, Gaussian-MRF=0.79, power trans-MRF=0.70, SPM5=0.76.
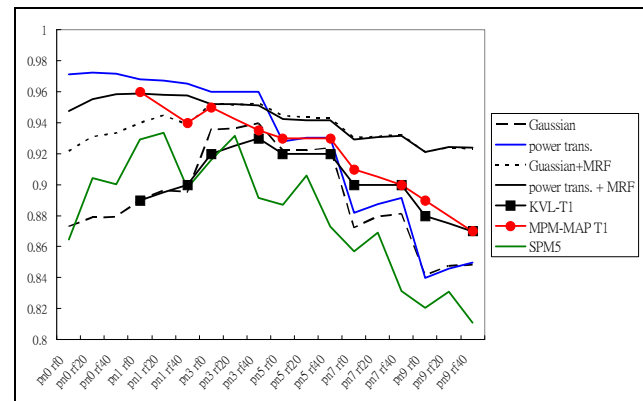
Fig. 8. Dice indices for the simulated dataset IV. (a) Average Dice indices for GM are: Gaussian=0.89, power trans=0.93, Gaussian-MRF=0.94, power trans-MRF=0.94, KVL-T1=0.9, MPM-MAP=0.92, SPM5=0.88. (b) Average Dice indices for WM are: Gaussian=0.86, power trans=0.93, Gaussian-MRF=0.94, power trans-MRF=0.96, KVL-T1=0.9, MPM-MAP=0.93, SPM5=0.91.

for WM are: Gaussian mixtures, 0.86, power transformed mixtures, 0.93, Gaussian-MRF, 0.94 and power trans-MRF, 0.96, and SPM5, 0.91. The average Dice indices for available data are 0.90 for KVL and 0.93 for MPM-MAP.

V. DISCUSSION

We proposed a power-transformation method to partitioned MR skull-stripped brain images into CSF, GM, and WM. As indicated in Figure 1, the power-transformation model fits the non-Gaussian data better than does the Gaussian mixtures, which explains the higher Jaccard indices shown in Figure 3. Across the four datasets we have considered, the power-transformation approach consistently gives higher Jaccard indices than the Gaussian mixtures.

The non-Gaussian features in MR intensity distributions are mainly due to partial volume effects. Existing MR segmentation work either ignores the partial volume effect, or explicitly segments the mixels at the cost of considerable complexity in the ensuing algorithm. The model presented here captures the asymmetry of the non-Gaussian nature of the data in a straightforward manner. The power transformation introduces a shape parameter to model the skewness of the distribution which includes Gaussian mixtures as a special case.

It has negligible effect on the kurtosis, which could also be slightly influenced by partial volume effects. However, the empirical results shown in Figure 1(b) suggest that histograms can be fitted well by taking only skewness into account, without incorporating kurtosis in the model.

Incorporating MRFs into the Gaussian mixtures can improve WM segmentation results substantially, as revealed by the dramatically increased Jaccard indices for WM in Figures 3, 7 and 8. However, the power-transformation model does not appear to benefit from incorporating the MRF constraints. MRFs compensate for lack of fit of the Gaussian intensity model to the non-Gaussian histogram. However, if MRFs and power transformations are combined together, the MRF competes with the power transformation yielding over-fitting or over-smoothing, which reduces the similarity index. Thus the overall effect did not exhibit improvement of the average accuracy except for the simulated data with random Gaussian noise (see results in Figure 8).

Our tissue segmentation method assumes that the images are not bias-contaminated. In the empirical examples, we have applied a wavelet bias correction method [43] to remove bias fields from contaminated images before applying tissue segmentation techniques. Except for the image volumes 5_8

and 6_10 in Figure 5, which are seriously contaminated by inter-slice intensity variation, the power transformation method tends to be robust to a choice between bias correction methods as compared with the Gaussian mixtures. Both N3 and the 3D wavelets give similar results when the power transformation method is used for image segmentation.

Comparing between the auto-segmented result (Figure 6(d)) and the manual segmented result (Figure 6(e)), we have noticed that a substantial part of the midbrain is misclassified to GM. This misclassification may heavily degrade performance of the WM segmentation. The misclassification has less impact on the Jaccard indices for GM which distributes in a greater region than does WM. A greater proportion of GM in the image volume explains why the Jaccard indices for GM are generally higher than those for WM, because a small amount of difference between two sets leads to minor change in the overall indices for GM.

IBSR provides the Jaccard indices of the segmentation results from several well-known segmentation methods, such as bMAP [23], Fuzzy [11], and TSK-means [15]. Comparing between the average Jaccard indices across the datasets I and II (bMAP: GM=0.55, WM=0.56, Fuzzy: GM=0.47, WM=0.56, TSK-means: GM=0.63, WM=0.42, the power-trans. method + wavelet bias correction: GM=0.79, WM=0.69, the power-trans. method + N3: GM=0.76, WM=0.65) shows that the power transformation method has outperformed other competitors in addition to those studied in detail.

In addition to being accurate and robust, the proposed method is also computationally efficient. The average computational time of our algorithm, written in MATLAB, is 2 minutes on an Intel P4 3.4 G computer. Performance of intensity-based methods, such as the proposed method, can be affected by noise other than bias fields; for example, small random noise yields scattered local errors in most segmentation procedures. Without a spatial prior, the proposed method introduced minor errors in segmentation (Figure 6(d)) as compared with results based on the manual segmentation (Figure 6(e)). Observing that brain tissues of the same type would be connected, we could reduce such random noise by applying some spatial smoothing techniques. However, as discussed earlier, using the MRF [22] may not lead to an improvement in the average accuracy due to the competing nature of the methods.

VI. CONCLUSIONS

Most statistical segmentation methods in the literature have assumed that either the intensity distribution of each tissue type is Gaussian, or the logarithmic transformation of the raw intensity is Gaussian. However, the manual segmentation results provided by the IBSR suggested that intensity distributions of brain tissues can be varyingly asymmetric and non-Gaussian. Instead of setting up additional classes to model "mixels", we proposed a power transform approach to perform automatic segmentation of brain MR images into CSF, GM and WM. It is intuitively clear that the well-known Box-Cox power

transformation model is able to provide a statistically meaningful and useful solution to this problem. The shape parameter λ can be used to extend traditional Gaussian mixture models further to encompass not only Gaussian intensity distributions but also non-Gaussian distributions. The parameters ω and θ can be estimated using the EM algorithm. We validated the approach against four real and simulated datasets of normal brains from the IBSR and BrainWeb. Any preprocessing bias-field correction method (e.g., N3 or 3D wavelet-based bias correction) can be easily incorporated into a pipeline framework for the proposed method.

Experiments on real data from the IBSR have indicated that the proposed approach achieves higher Jaccard indices compared with other methods currently in use. The power transformation approach not only preserves the simplicity of the Gaussian mixtures, but also has the potential to generalize to multivariate versions adapted for segmentation using multimodality images (e.g., PD, T1, and T2 images). Although our approach fails to benefit from the particular MRF model we applied, incorporating other MRF models or spatial techniques, such as atlas-based initialization and multiresolution analysis [61], into the proposed approach could lead to interesting alternatives. However, many automatic methods are susceptible to problems similar to the MRF approach, and these have to be resolved. These issues remain interesting research topics for the power transformation image segmentation model.

APPENDIX

Let y_n be a random voxel intensity observed from one of K brain tissue types. Suppose that intensity distribution of each tissue type can be expressed as

$$p(y_n \mid i, \theta) = G(y_n^{(\lambda_i)} \mid \mu_i, \sigma_i^2) \cdot y_n^{(\lambda_i - 1)},$$

where the parameters are denoted by $\theta = \{\mu_i, \sigma_i^2, \lambda_i \mid i = 1, ..., K\}$.

Because class membership is unobservable, we maximize the conditional expectation of the complete data log-likelihood given the observed values and the previous parameter values.

Let $Q(\phi)$ represents the expected log-likelihood, where $\phi = \{\omega, \theta\}$, and $\omega = \{\omega_i \mid 0 \le \omega_i, \sum_i \omega_i = 1, i = 1, ..., K\}$.

$$Q(\phi) = E_i[\log \prod_n p(y_n, i \mid \phi)]$$
$$= E_i[\sum_n \log p(y_n, i \mid \phi)]$$
$$= \sum_n E_i[\log p(y_n, i \mid \phi)]$$
$$= \sum_n \sum_i p(i \mid \phi, y_n) \cdot \log p(y_n, i \mid \phi)$$
$$= \sum_n \sum_i p(i \mid \phi, y_n) \cdot \log(p(y_n \mid i, \theta) \omega_i)$$

Given that $p(y_n \mid i, \theta) = G(y_n^{(\lambda_i)} \mid \mu_i, \sigma_i^2) \cdot y_n^{\lambda_i - 1}$, we then obtain

$$\log(p(y_n \mid i, \theta)\omega_i)$$
$$= \log G(y_n^{(\lambda_i)} \mid \mu_i, \sigma_i^2) + (\lambda_i - 1)\log y_n + \log \omega_i$$
$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log \sigma_i^2 - \frac{(y_n^{(\lambda_i)} - \mu_i)^2}{2\sigma_i^2} + (\lambda_i - 1)\log y_n + \log \omega_i$$

In the E-step, the posterior probability that the random observation y_n belongs to the i-th group is calculated using initialized parameters,

$$p(i \mid \phi, y_n) = \frac{p(i, y_n \mid \phi)}{p(y_n \mid \phi)} = \frac{p(y_n \mid i, \theta)\omega_i}{\sum_{i'} p(y_n \mid i', \theta)\omega_{i'}}.$$

Adding a lagrangian multiplier, we get

$$L(\phi) = Q(\phi) - \gamma(\sum_i \omega_i - 1).$$

At the M-step, we find new estimates for each parameter that maximize $L(\phi)$.

New estimate for μ_i:

$$\frac{\partial L(\phi)}{\partial \mu_i} = 0 \Rightarrow \sum_n p(i \mid \phi, y_n)\left(\frac{y_n^{(\lambda_i)} - \mu_i}{\sigma_i^2}\right) = 0 \Rightarrow \mu_i = \frac{\sum_i p(i \mid \phi, y_n) y_n^{(\lambda_i)}}{\sum_i p(i \mid \phi, y_n)}$$

New estimate for σ_i^2:

$$\frac{\partial L(\phi)}{\partial \sigma_i^2} = 0 \Rightarrow \sum_n p(i \mid \phi, y_n)\left(-\frac{1}{2}\sigma_i^{-2} + \frac{(y_n^{(\lambda_i)} - \mu_i)^2}{2}\sigma_i^{-4}\right) = 0$$
$$\Rightarrow \sigma_i^2 = \frac{\sum_n p(i \mid \phi, y_n)(y_n^{(\lambda_i)} - \mu_i)^2}{\sum_n p(i \mid \phi, y_n)}$$

New estimate for membership probability ω_i:

$$\frac{\partial L(\phi)}{\partial \omega_i} = 0 \Rightarrow \sum_n p(i \mid \phi, y_n)\omega_i^{-1} - \gamma = 0 \Rightarrow \omega_i = \frac{1}{\gamma}\sum_n p(i \mid \phi, y_n)$$

Insert ω_i under the constraint $\sum_i \omega_i = 1$,

$$\sum_i \frac{1}{\gamma}\sum_n p(i \mid \phi, y_n) = 1 \Rightarrow \gamma = \sum_i \sum_n p(i \mid \phi, y_n).$$

Thus,

$$\omega_i = \frac{\sum_n p(i \mid \phi, y_n)}{\sum_i \sum_n p(i \mid \phi, y_n)}.$$

New estimate for λ_i:

$$\frac{\partial L(\phi)}{\partial \lambda_i} = 0 \Rightarrow \sum_n p(i \mid \phi, y_n)\left(\frac{-(y_n^{(\lambda_i)} - \mu_i)}{\sigma_i^2} \cdot \frac{(\log y_n) y_n^{\lambda_i}\lambda_i + (y_n^{\lambda_i} - 1)}{\lambda_i^2} + \log y_n\right) = 0$$

These estimates become $\theta^{(t+1)}$ and $\omega^{(t+1)}$ that will be used in the next estimation step.

Now, suppose that $\omega_i \le b_i$ for $i = 1, ..., K$. If any of them, say ω_1, falls out of the boundary, we need to re-estimate ω_1 as follows:

$$\omega_1 = b_1$$

The constraint becomes $\sum_{i \ne 1}\omega_i = 1 - b_1$. Insert the modified constraint into $L(\phi)$.

$$L(\phi) = Q(\phi) - \gamma(\sum_{i \ne 1}\omega_i - (1 - b_1))$$

We then find the new estimates for ω_i that $i \ne 1$:

$$\frac{\partial L(\phi)}{\partial \omega_i} = 0 \Rightarrow \sum_n p(i \mid \phi, y_n)\omega_i^{-1} - \gamma = 0 \Rightarrow \omega_i = \frac{1}{\gamma}\sum_n p(i \mid \phi, y_n)$$

Inserting ω_i into the constraint that $\sum_{i \ne 1}\omega_i = 1 - b_1$, we solve for γ:

$$\sum_{i \ne 1}\frac{1}{\gamma}\sum_n p(i \mid \phi, y_n) = 1 - b_1 \Rightarrow \gamma = \frac{\sum_{i \ne 1}\sum_n p(i \mid \phi, y_n)}{1 - b_1}$$

Then, we obtain

$$\omega_i = \frac{\sum_n p(i \mid \phi, y_n)}{\sum_{i \ne 1}\sum_n p(i \mid \phi, y_n)} \cdot (1 - b_1).$$

This concludes the computational algorithm.

REFERENCES

[1]   R. W. McCarley, C. G. Wible, M. Frumin, Y. Hirayasu, J. J. Levitt, I. A. Fischer, and M. E. Shenton, "MRI anatomy of schizophrenia," *Biological Psychiatry*, vol. 45, pp. 1099–1119, 1999.
[2]   H. Ananth, I. Popescu, H. D. Critchley, C. D. Good, R. S. Frackowiak, and R. J. Dolan, "Cortical and Subcortical Gray Matter Abnormalities in Schizophrenia Determined Through Structural Magnetic Resonance Imaging With Optimized Volumetric Voxel-Based Morphometry," *The American Journal of Psychiatry*, vol. 159, pp. 1497–1505, 2002.
[3]   N. Iwasaki, K. Hamano, Y. Okada, Y. Horigome, J. Nakayama, T. Takeya, H. Takita, and T. Nose, "Volumetric quantification of brain development using MRI," *Neuroradiology*, vol. 39, pp. 841–846, 1997.
[4]   H. Kanemura, M. Aihara, S. Aoki, T. Araki, and S. Nakazaw, "Development of the prefrontal lobe in infants and children: a three-dimensional magnetic resonance volumetric study," *Brain and Development*, vol. 25, pp. 195–199, 2003.
[5]   K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection," *IEEE Trans. Med. Imag.*, vol. 20, pp. 677–688, 2001.
[6]   X. Descombes, F. Kruggel, G. Wollny, and H. J. Gertz, "An Object-Based Approach for Detecting Small Brain Lesions: Application to Virchow-Robin Spaces," *IEEE Trans. Med. Imag.*, vol. 23, pp. 246–255, 2004.
[7]   N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst. Man Cybernet*, vol. 9, pp. 62–66, 1979.
[8]   D. Wang and D. M. Doddrell, "A segmentation-based and partial-volume compensated method for an accurate measurement of lateral ventricular volumes on T1-weighted magnetic resonance images," *Magnetic Resonance Imaging*, vol. 19, pp. 267–272, 2001.
[9]   S. K. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis, "Adaptive, template moderated, spatially varying statistical classification," *Medical Image Analysis*, vol. 4, pp. 43–55, 2000.
[10] [10] G. B. Coleman and H. C. Andrews, "Image segmentation by clustering," *Proceedings of the IEEE*, vol. 67, pp. 773–785, 1979.
[11] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J.C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Transactions on Neural Networks*, vol. 3, pp. 672–682, 1992.
[12] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, pp. 737–752, 1999.
[13] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and

segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, pp. 193–199, 2002.

[14] A. W. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Trans. Med. Imag.*, vol. 22, pp. 1063–1075, 2003.

[15] J. K. Fwu and P. M. Djuric, "Unsupervised vector image segmentation by a tree structure–ICM algorithm," *IEEE Trans. Med. Imag.*, vol. 15, pp. 871–880, 1996.

[16] A Mayer and H. Greenspan, "Segmentation of brain MRI by adaptive mean shift," in *IEEE 3rd Intl Symposium on Biomedical Imaging: Macro to Nano*, pp. 319–322, 2006.

[17] J. R. Jimenez-Alaniz, M. Pohl-Alfaro, V. Medina-Banuelos, and O. Yanez-Suarez, "Segmenting Brain MRI using Adaptive Mean Shift," in *Proceedings of the 28th IEEE EMBS Annual Intl Conference, New York City, USA*, pp. 3114–3117, 2006.

[18] J. O. Glass, W. E. Reddick, O. Goloubeva, V. Yo, and R. G. Steen, "Hybrid artificial neural network segmentation of precise and accurate inversion recovery (PAIR) images from normal human brain," *Magnetic Resonance Imaging*, vol. 18, pp. 1245–1253, 2000.

[19] U. Amato, M. Larobina, A. Antoniadis, and B. Alfano, "Segmentation of magnetic resonance brain images through discriminant analysis," *Journal of Neuroscience Methods*, vol. 131, pp. 65–74, 2003.

[20] J. Ashburner and K. J. Friston, "Multimodal image coregistration and partitioning – a unified framework," *NeuroImage*, vol. 6, pp. 209–217, 1997.

[21] J. A. Hartigan, *Clustering algorithms*. New York: Wiley, pp.113–129, 1975.

[22] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral MR Images," *IEEE Trans. Med. Imag.*, vol. 16, pp. 176–186, 1997.

[23] J. C. Rajapakse and F. Kruggel, "Segmentation of MR images with intensity inhomogeneities," *Image Vision Computing*, vol. 16, pp. 165–180, 1998.

[24] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, pp. 45–57, 2001.

[25] L. Lemieus, A. Hammers, T. Mackinnon, and R. S.N. Liu, "Automatic segmentation of the brain and intracranial cerebrospinal fluid in T1-weighted volume MRI scans of the head, and its application to serial cerebral and intracranial volumetry," *Magnetic Resonance in Medicine*, vol. 49, pp. 872–884, 2003.

[26] E. Bullmore, M. Brammer, G. Rouleau, B. Everitt, A. Simmons, T. Sharma, S. Frangou, R. Murray, and G. Dunn, "Computerized brain tissue classification of magnetic resonance images: A new approach to the problem of partial volume artifact," *NeuroImage*, vol. 2, pp. 133–147, 1995.

[27] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.

[28] J. S. Kim, V. Singh, J. K. Lee, J. Lerch, Y. Ad-Dab'bagh, D. MacDonald, J. M. Lee, S. I. Kim, and A. C. Evans, "Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification," *NeuroImage*, vol. 27, pp. 210–221, 2005.

[29] M. A. G. Ballester, A. P. Zisserman, and Michael Brady, "Estimation of the partial volume effect in MRI," *Medical Image Analysis*, vol. 6, pp. 389–405, 2002.

[30] X. Wei, S. K. Warfield, K. H. Zou, Y. Wu, X. Li, A. Guimond, J. P. Mugler, R. R. Benson, L. Wolfson, H. L. Weiner, and C. R. Guttmann, "Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy," *J Magn Reson Imaging*, vol. 15, pp.203–209, 2002.

[31] M. Prastawa, J. H. Gilmore, W. Lin, and G. Gerig, "Automatic segmentation of MR images of the developing newborn brain," *Medical Image Analysis*, vol. 9, pp. 457–466, 2005.

[32] D. H. Laidlaw, K. W. Fleischer, and A. H. Barr, "Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms," *IEEE Trans. Med. Imag.*, vol. 17, pp. 74–86, 1998.

[33] K. V. Leemput, F. Maes, D. Vandermeulen, P. Suetens, "A unifying framework for partial volume segmentation of brain MR images," *IEEE Trans. Med. Imag.*, vol. 22, pp. 105–119, 2003.

[34] J. Tohka, A. Zijdenbos, and A. Evans, "Fast and robust parameter estimation for statistical partial volume models in brain MRI," *NeuroImage*, vol. 23, pp. 84–97, 2004.

[35] D. W. Shattuck, S. R. Sandor-Leahy, L. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *NeuroImage*, vol. 13, pp. 856–876, 2001.

[36] S. Ruan, C. Jaggi, J. Xue, J. Fadili, and D. Bloyet, "Brain tissue classification of magnetic resonance images using partial volume modeling," *IEEE Trans. Med. Imag.*, vol. 19, pp. 1179–1187, 2000.

[37] H. S. Choi, David R. Haynor, and Y. Kim, "Partial Volume Tissue Classification of Multichannel Magnetic Resonance Images – A Mixel Model," *IEEE Trans. Med. Imag.*, vol. 10, pp. 395–407, 1991.

[38] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans, "A fully automatic and robust brain MRI tissue classification method," *Medical Image Analysis*, vol. 7, pp. 513–527, 2003.

[39] J. B. Arnold, J. Liow, K. A. Schaper, J. J. Stern, J. G. Sled, D. W. Shattuck, A. J. Worth, M. S. Cohen, R. M. Leahy, J. C. Mazziotta, and D. A. Rottenberg, "Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects," *NeuroImage*, vol. 13, pp. 931–943, 2001.

[40] W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 15, pp. 429–442, 1996.

[41] R. Guillemaud and M. Brady, "Estimating the bias field of MR images," IEEE *Trans. Med. Imag.*, vol. 16, pp. 238–251, 1997.

[42] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, pp. 87–97, 1998.

[43] H-R Su, M. Liou, P. E. Cheng, J. A. D. Aston, and C-H Chuang, "MR Image Segmentation Using Wavelet Analysis Techniques," HBM 2005 Abstract [CDROM], *Neuroimage*, vol. 26, supp. 1, 2005.

[44] G Battlĕ, "A block spin construction of ondelettes, Part I: Lemarié function," *Communications in Mathematical Physics*, vol. 110, pp. 601–615, 1987.

[45] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, 1993.

[46] C. A. Cocosco, V. Kollokian, R. K. Kwan, and A. C. Evans, "Brainweb: Online interface to a 3D MRI simulated brain database," *Neuroimage*, pt. 2/4, S425, vol. 5, no. 4, 1997. (images available online: http://www.bic.mni.mcgill.ca /brainweb/).

[47] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, A.C. Evans, "Design and Construction of a Realistic Digital Brain Phantom," *IEEE Trans. Med. Imag.*, vol.17, No.3, p.463–468, 1998.

[48] R. K. S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1996, vol. 1131, Visualization in Biomedical Computing (VBC'96), pp. 135–140.

[49] M. Aitkin and D. B. Rubin, "Estimation and hypothesis testing in finite mixture models," *Journal of the Royal Statistical Society, Series B*, vol. 47, pp. 67–75, 1985.

[50] J. D. Lee, P. E. Cheng, M. Liou, and C. C. Chen, "Partition MR images by extended EM segmentation," in *Proceedings International Conference on Neural Information Processing*, Shanghai, 2001, vol. 1, pp. 430–435.

[51] In-Kwon Yeo and Richard A. Johnson, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, pp. 954–959, 2000.

[52] G E. P. Box and D. R. Cox, "An analysis of transformation," *Journal of the Royal Statistical Society, Series B*, vol. 26, pp. 211–252, 1964.

[53] McCullagh, P. and J. A. Nelder, *Generalized Linear Models, 2nd Ed.*, Chapman and Hall, 1989.

[54] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations*, Prentice-Hall, 1976.

[55] D. N. Kennedy, P. A. Filipek, and V. S. Caviness, "Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 8, pp. 1–7, 1989.

[56] B. Belaroussi, J. Milles, S. Carme, Y. M. Zhu, H. Benoit-Cattin, "Intensity non-uniformity correction in MRI: Existing methods and their validation," *Medical Image Analysis*, vol. 10, pp. 234–246, 2006.

[57] U. Vovk, F. Pernuš, and B. Likar, "A Review of Methods for Correction of Intensity Inhomogeneity in MRI," *IEEE Trans. Med. Imag.*, vol. 26, pp. 405–421, 2007.

[58] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[59] J. L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas, "An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI," *IEEE Trans. Med. Imag.*, vol. 21, pp. 934–945, 2002.

[60] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated Model-Based Tissue Classification of MR Images of the Brain," *IEEE Trans. Med. Imag.*, vol. 18, pp. 897–908, 1999.

[61] M. Zeydabadi, R.A. Zoroofi, and H. Soltanian-Zadeh, "Multiresolution automatic segmentation of T1-weighted brain MR images," in *IEEE Intl Symposium on Biomedical Imaging: Macro to Nano*, pp. 165–168, 2004.