Brain Tissue Classification of Magnetic Resonance Images Using Partial Volume Modeling

Su Ruan, Cyril Jaggi, Jing-Hao Xue, Jalal M. Fadili, Daniel Bloyet

Brain Tissue Classification of Magnetic Resonance Images Using Partial Volume Modeling

Su Ruan*, Cyril Jaggi, Jinghao Xue, Jalal Fadili, and Daniel Bloyet, *Member, IEEE*

***Abstract*—This paper presents a fully automatic three-dimensional classification of brain tissues for Magnetic Resonance (MR) images. An MR image volume may be composed of a mixture of several tissue types due to partial volume effects. Therefore, we consider that in a brain dataset there are not only the three main types of brain tissue: gray matter, white matter, and cerebro spinal fluid, called pure classes, but also mixtures, called mixclasses. A statistical model of the mixtures is proposed and studied by means of simulations. It is shown that it can be approximated by a Gaussian function under some conditions. The D'Agostino–Pearson normality test is used to assess the risk α of the approximation. In order to classify a brain into three types of brain tissue and deal with the problem of partial volume effects, the proposed algorithm uses two steps: 1) segmentation of the brain into pure and mixclasses using the mixture model; 2) reclassification of the mixclasses into the pure classes using knowledge about the obtained pure classes. Both steps use Markov random field (MRF) models. The multifractal dimension, describing the topology of the brain, is added to the MRFs to improve discrimination of the mixclasses. The algorithm is evaluated using both simulated images and real MR images with different T1-weighted acquisition sequences.**

***Index Terms*—Brain tissue, classification, Markov random fields, mixture, multifractal dimension, partial volume effects, validation.**

I. Introduction

THE ADVANTAGES of magnetic resonance imaging (MRI) over other diagnostic imaging modalities are its higher spatial resolution and its better discrimination of soft tissue. Advanced applications that use the morphological contents of MRI frequently require segmentation of the imaged volume into tissue types, i.e., gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). There are several major reasons for doing MRI brain tissue segmentation: multimodality image correlation, visualization, and quantification, and their clinical uses such as in tumor and lesion detection. Many automated or semiautomated approaches for tissue segmentation in brain images, using T1-weighted [1] or multispectral MR data [2], [3], have already been proposed. The main techniques are based on: histogram threshold determination [4], cluster analysis [5], *a priori* information about anatomy [6], and Bayesian classification [7]. An overview of segmentation methods for MRI can be found in [8]. Most of the proposed methods assume that each voxel belongs to a signal specific tissue type. However, in MR images the complexity of tissue boundaries causes many voxels to contain mixtures of tissues [9]. This is the partial volume effect problem. In this paper, we present a brain tissue segmentation which takes this problem into account.

Some researchers have worked on the issue of partial volume effects: for example, correction of the partial volume effect [10] in order to segment tumors or lesions [11], [12] in clinical applications, and modeling of the mixture distribution to segment brain tissues [13]. In [14], the authors propose a method based on unsupervised fuzzy clustering. However, the fuzzy results pose the problem of fuzzy diagnosis. Laidlaw *et al.*[15] and Santiago and Gage [16] derive a distribution of partial volume mixtures of two materials by fitting the histogram of an entire dataset or a local region. We share the idea of the histogram fit for finding the distribution. However, our mixture model is different. The details will be presented in Section III.

In this paper, we propose a new strategy for modeling the mixture probability density functions in the classification of the brain tissue. MR images are previously processed to extract the encephalon (edited brain). Based on the proposed mixture model, a partition of the encephalon is carried out in two steps by means of the Markov random fields (MRF) classification, which is widely used in medical image processing [13], [17], [18]. The first step consists of classifying the brain into three types of tissue (GM, WM, and CSF) and mixtures of these brain tissues. As the contrast between pure tissues in each mixture is very low, multifractal analysis [19], [20], which provides information about local intensity variation, is carried out to enhance the mixclasses. An additional energy component based on the shape of the local intensity pattern is incorporated into the MRF model in the second step to reclassify the mixclasses into the pure classes.

The paper is organized as follows: Section II presents the related works in mixture modeling and MRF classification in MRI. Section III describes the proposed mixture model. Section IV is devoted to the classification algorithm. The validation of the method and experimental results are presented in Section V. Finally some concluding remarks are given.

II. Related Work

A. Partial Volume Effects

The partial volume effect appears when more than one type of class or material occupies one voxel or pixel of an image. These

Manuscript received May 24, 1999; revised September 11, 2000. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was L. Clarke. *Asterisk indicates corresponding author.*

*S. Ruan is with Greyc-Ismra, Cnrs Umr 6072, 6. Bd. du Maréchal Juin, F-14050 Cedex, Caen, France (e-mail: Su.Ruan@greyc.ismra.fr).

C. Jaggi, J. Fadili, and D. Bloyet are with Greyc-Ismra, Cnrs Umr 6072, 6. Bd. du Maréchal Juin, F-14050 Cedex, Caen, France.

J. Xue is with the Department of Electronic Engineering, Tsinghua University, Beijing, China.

voxels and pixels are usually called mixels. Several models have been proposed to describe these phenomena in MRI and to find the mixture density function. The most widely used represents the intensity y_m of a mixel as follows [13], [15], [16]:

$$y_m = \sum_{k=1}^{N_t} a_k I_k + \varepsilon, \quad \forall\, k \in [1,\, N_t] \tag{1}$$

where ζ is white Gaussian noise with standard deviation σ_ε. N_t represents the number of tissues. I_k is related to the intensity value of the class k, and a_k its proportion, which is assumed to be a uniform random variable in the interval [0,1].

In the case of mixels of just two pure classes $(k = 1,\, 2)$, the probability density function of y_m is given by

$$p(y_m) = \int_0^1 \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp \frac{-[y_m - ((1-a)I_1 + aI_2)]^2}{2\sigma_\varepsilon^2}\, da \tag{2}$$

There are four major voxel intensity modifiers in MR images: random field (RF) nonuniformity, partial volume effects, additive noise, and the intrinsic heterogeneity of the tissue types. Many manuscripts and publications have attempted to model one or two of these four modifiers, assuming that the others play a minor role. In [15] and [16], I_1 and I_2 are taken as constant values (mean values); in this case, the model ignores not only RF nonuniformity, but also the heterogeneity within each type of pure class. Work carried out in [21] shows the level of variance of within the brain tissue classes is almost the same as that of the noise in T1-weighted MRI and, therefore, cannot be ignored. In [22], the authors first find normal distributions of the pure classes using regions of interest (ROIs) enclosing pure tissues; each mixture is then considered as the weighted sum of the pure classes with varying proportions of the corresponding pure component. In order to simplify the mixture model, ten different values of a are proposed in the paper, instead of using a as a continuous and random variable as in (2). We propose a more complete model for $p(y_m)$ which takes into account not only instrument noise but also the heterogeneity of each type of pure matter. This model is described in Section III.

B. MRF Classification

One obvious way of incorporating spatial correlation into a segmentation process is to use MRFs as *a priori* models [23]–[25]. The MRF has already been used in medical images. Choi *et al.* [13] used the MRF model for the classification of multichannel MRI of the brain. The MR image is considered as the mixel image, which is a MRF with respect to a given neighborhood system. In each voxel, the segmentation finds percentages of each pure tissue. In [26], the authors proposed an extension of the MRF two-dimensional model [23] to a three-dimensional model in order to segment volumic objects. Segmentation is improved by using 3-D information. Held *et al.* [17] have also proposed a 3-D MRF segmentation algorithm based on the adaptive segmentation algorithm described by Wells *et al.* [18]. The emphasis of this algorithm is on the modeling of signal inhomogeneities by *a priori* MRF.

III. The Mixel Model

A. Theory

Let y_m denote the intensity of a mixel and N_t the number of pure tissue classes present in an image. The intensity value of a mixel can be represented by the weighted sum of N_t pure tissues

$$y_m = \sum_{k=1}^{N_t} a_k y_k^p, \quad \text{with} \quad \sum_{k=1}^{N_t} a_k = 1, \quad \forall\, k \in [1,\, N_t] \tag{3}$$

where a_k is the contribution of pure class k to the mixel intensity y_m. The random variable y_k^p refers to the pure class k whose probability density results from the distributions of acquisition parameters $T_{1_k}^p$, $T_{2_k}^p$, and ρ_k^p, and from instrumental noise. The probability density function $p(y_k^p)$ may then be assumed to be asymptotically normal by application of the central limit theorem. In the case of mixels including just two pure tissue classes, the mixel intensity takes the following expression:

$$y_m = a y_1^p + (1-a) y_2^p, \qquad 0 \le a \le 1 \tag{4}$$

where a is the relative quantity of the first class. If the probability density functions of a, y_1^p and y_2^p are known, then that of y_m can be obtained. For two pure classes, which are normally distributed $(N(\mu_k,\, \sigma_k^2),\, (k = 1,\, 2))$ and statistically independent from each other, the probability density function of the mixture can be written as follows for a given a:

$$p(y_m,\, a) = \frac{1}{\sqrt{2\pi}\,\sqrt{a^2\sigma_1^2 + (1-a)^2\sigma_2^2}} \cdot \exp\left\{ -\frac{[y_m - (a\mu_1 + (1-a)\mu_2)]^2}{2[a^2\sigma_1^2 + (1-a)^2\sigma_2^2]} \right\} \tag{5}$$

$p(y_m,\, a)$ is then a Gaussian density function. However, in a volume dataset, all a values can occur with equal probability since mixels may consist of any fraction of pure tissue class on boundaries. Thus, assuming that a is a uniform random variable in the interval [0, 1] as already proposed in other publications (e.g. [13] and [16]), the marginal probability density function of the mixture is then

$$p(y_m) = \int_0^1 p(y_m,\, a)\, da \tag{6}$$

The calculation of this expression is nontrivial because of the undefined analytic formulation of the integral. Intuitively, $p(y_m)$ could not be a Gaussian function. Thus, numerical simulations described in the Section III-B have been performed in order to predict the behavior of such a probability density function.

B. Approximation of the Probability Distribution of a Mixel

Based on (4), the simulator generates mixels using two sets of Gaussian random variables representing two pure classes and one uniform random variable, yielding the histogram of the mixels. It is obvious that the intensity distribution of the mixels depends on the parameters $(\mu_k,\, \sigma_k^2)$; the effect of changing the means and the standard deviations of the pure classes can be intuitively understood as illustrated in Fig. 1. The mixture

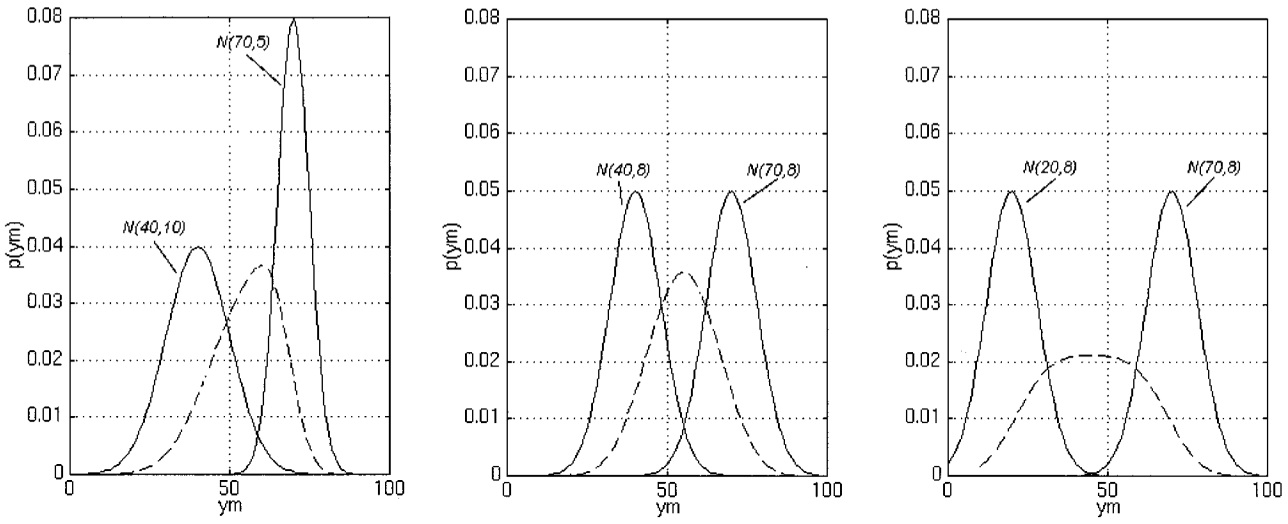


Fig. 1. Examples of mixture density function $p(y_m)$ obtained by simulations (dashed lines). The two pure classes are considered to be normally distributed (continuous lines). $p(y_m)$ depends on the difference between the means and on the standard deviation ratio of the two pure classes. In some cases, $p(y_m)$ is clearly (b) Gaussian, while in other cases it is (a) asymmetric or (c) flattened.

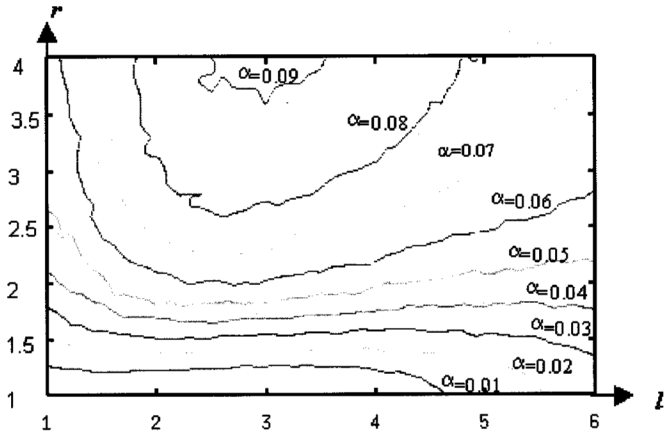


Fig. 2. Level of significance α using the D'Agostino omnibus test for normality assessment, as a function of $l = (|\mu_1 - \mu_2|)/((\sigma_1 + \sigma_2)/2)$ and $r = \sigma_1/\sigma_2$. For conditions of noise and acquisition parameters similar to real data, the level of significance α obtained on the simulated Brainweb phantom is less than 1%. This value is used as a threshold to validate the normality assumption for real MRI applications.

probability density function mainly depends on the difference between the two means and the two variances. The degree of asymmetry of $p(y_m)$increases when the ratio between the two standard deviations of pure classes increases. Moreover, the greater the difference between the means, the flatter $p(y_m)$ becomes. Despite the undefined analytic expression of $p(y_m)$, the simulations show that in some cases [Fig. 1(b)], the mixed class intensity distribution can be well approximated by a Gaussian function. The quality of the fit by a Gaussian function can be evaluated through a normality test. We have chosen the D'Agostino–Pearson K^2 omnibus test of normality [27] owing to its robustness

$$K^2 = Z^2\left(\sqrt{b_1}\right) + Z^2(b_2) \tag{7}$$

where $Z(\sqrt{b_1})$ and $Z(b_2)$ are the normal approximations to the sample estimates of skewness and kurtosis. Under the null hypothesis (validity), K^2 has a χ^2 distribution with two degrees of freedom. In fact, as illustrated in Fig. 1, the $p(y_m)$ depends on the ratio $(|\mu_1 - \mu_2|)/((\sigma_1 + \sigma_2)/2)$, noted l, and on the ratio of the two standard deviations σ_1/σ_2, noted r. Curves of α (Fig. 2), can then be drawn as a function of different values of l and r. The curve for $\alpha = 1\%$ is also reported on the map. This map shows that a Gaussian approximation can be used in a wide range of variation of l, even where the risk α is small (e.g., $\alpha = 1\%$). Thus, if the parameters of the two pure classes are known, the validation of fitting the mixture density by a Gaussian function can be performed. Let us now carry out the normality measure to the three mixel densities shown in Fig. 1. $p(y_m)$ in Fig. 1(b) can be accepted as a Gaussian function since $\alpha < 1\%$ ($l = 3.75$, $r = 1$), while the other two mixel density functions [Fig. 1(a): ($l = 4$, $r = 2$), $\alpha > 5\%$; Fig. 1(c): ($l = 6.25$, $r = 1$), $\alpha > 2\%$] are far away from a Gaussian function. In practical situations, we will show that a Gaussian approximation is almost verified for a T1-weighted acquisition sequence.

IV. Classification

A. Brain Tissue Models

We assume that a normal human brain consists of three types of tissue: WM, GM, and CSF. However, due to the spatial resolution of the acquisition system, a voxel in a MR image can be made up of a single type of tissue or of a mixture of different types. As the probability of mixing more than two types of tissue in a voxel is very low and the mixture of WM and CSF is very rare in a brain dataset, we can consider that the brain image volume will consist of the three pure classes CSF, GM, and WM, and two mixclasses: CG (a mixture of CSF and GM) and GW (a mixture of GM and WM). As pure classes exhibit heterogeneity, each pure class presents its own specific statistical distribution while additive instrument noise is spatially invariant. The three pure classes in T1-weighted MR images can be well modeled by Gaussian functions [21], as can the two mixclasses. Given an edited encephalon, the problem is then to iden-

tify the parameters of the five Gaussian functions. We used the Davidon–Fletcher–Powell method described in [28] to fit the histogram of the entire data set using the sum of five Gaussian functions as a model. Validation of the Gaussian modeling can be accomplished by the test of normality described above. Fig. 3 shows the fit of an experimental histogram using the sum of three or five Gaussian functions. The difference between the model and the histogram is only significant when three Gaussian functions are used. This means that mixels cannot be ignored.

B. MRF Classification

The MRF model has been widely used for image modeling due to its flexibility in defining the neighborhood system. Theory about MRF segmentation can be found in [23]. Using Gibbs representation for a MRF, we need only to specify the potential, expressed by a global energy function U, which can be decomposed as the sum of several functions according to the context.

The resolution of MR images (about 1 mm^3) is significantly smaller than the size of most of the brain structures we wish to study. Therefore, adjacent voxels are likely to have similar components and thus similar MRI intensities. This idea can be expressed using an MRF. As the brain is a 3-D object, a 3-D geometrical model is needed. In the first step, the global energy function U is taken as the sum of only two functions. The conditional distribution of the observation y_v at the voxel v under a given class label λ is a Gaussian distribution, as shown in Section III-A. Therefore, the first energy function U_1 which takes into account the gray levels of the voxel is described as

$$U_1(y_v/\lambda) = \log\left(\sqrt{2\pi}\sigma_\lambda\right) + \frac{(y_v - \mu_\lambda)^2}{2\sigma_\lambda^2}. \tag{8}$$

The second energy function uses the Ising model which favors local intensity homogeneity:

$$U_2(\lambda) = \beta_\lambda \sum_{(i,j,k,m,n,l)\in C} I_{\lambda(i,j,k)\neq\lambda(m,n,l)} \tag{9}$$

where C is a 3-D system of second-order neighborhood consisting of the 18 nearest neighbors, and β_λ is a weight coefficient. U_2 is minimum for homogeneous regions. Classification of the brain is equivalent to finding the optimal labeling using the maximum *a posteriori* (MAP) estimator: $\lambda = \arg\min(U_1 + U_2)$, where λ takes five possible labels: GM, WM, CSF, CG, and GW.

In the second step of classification, the two functions U_1 and U_2 are not sufficient to reclassify the mixclasses, because there is almost no contrast in these regions due to their positions close to the boundaries between pure tissue types. It is, therefore, necessary to introduce other information to deal with this problem. In fact, three types of gray level variations are typical in a brain volume: uniform regions where there is at most very little gray level variations, and convex or concave regions that correspond to hill- or valley- shaped gray level variations as shown in Fig. 4. The multifractal dimension, usually used in texture analysis [19], [20], can provide local information about

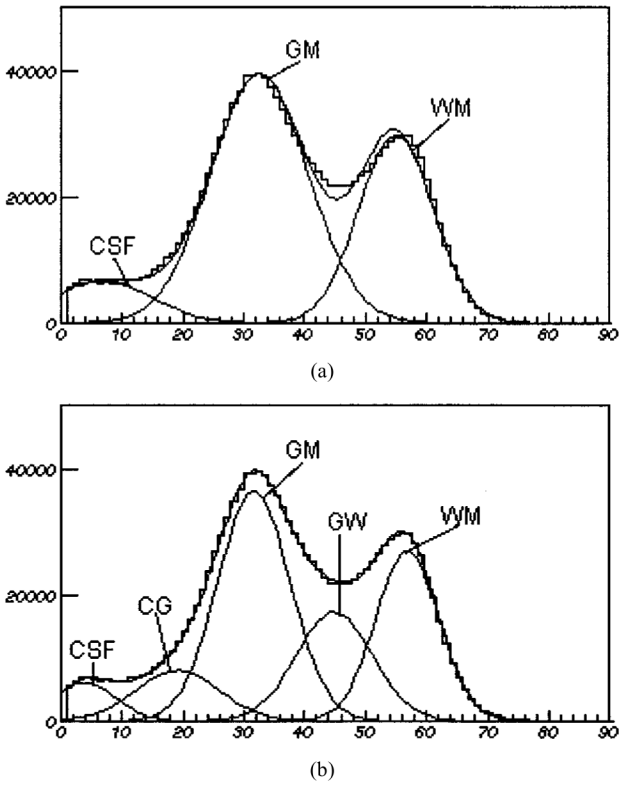


Fig. 3. Results of histogram modeling using (a) three Gaussian functions and (b) five Gaussian functions . The fit using five Gaussian functions gives a more accurate representation of the histogram [mean square error (MSE) = 10^{5} for (a) and 10^4 for (b)].

Fig. 4. Illustration of Hölder exponent α^H for different spatial variations of intensity. Knowing the value α_0^H for an homogeneous region, we have $\alpha^H > \alpha_0^H$ for a concave region and $\alpha^H < \alpha_0^H$ for a convex region.

gray level variations. Therefore, we have added a third term U_3 into the energy function, which is obtained by multifractal analysis. The multifractal dimension is locally measured by the Hölder exponent α^H, which can be estimated by the slope of $\log(Ir_v)$ at the voxel v versus $\log(r)$, where Ir_v is the sum of gray levels in a cube of size r^3 centered at the voxel v. r linearly changes from one to a given maximum value with a unit step. Let us note α_0^H the Hölder exponent within a uniform region. If $\alpha^H > \alpha_0^H$, the current voxel is in a concave region (valley), while if $\alpha^H < \alpha_0^H$, the current voxel is in a convex region (hill). It is known that, for a T1-weighted MRI brain volume, the intensities of sulci and gyri are dark and light, respectively. They can be considered as concave and convex regions, respectively (Fig. 5). Therefore, a CG mixel with $\alpha^H > \alpha_0^H$ probably belongs to the CSF, and a GW mixel with $\alpha^H < \alpha_0^H$ probably

belongs to the WM. Elsewhere, $\alpha^H = \alpha_0^H$ is mostly observed in WM. In summary, the function U_3 can be defined as follows:

$$U_3(\lambda) = \begin{cases} -\gamma F_\alpha(\alpha^H), & \text{if } \lambda \text{ is the WM label} \\ 0, & \text{if } \lambda \text{ is the GM label} \\ \gamma F_\alpha(\alpha^H), & \text{if } \lambda \text{ is the CSF label} \end{cases}$$

$$\text{with} \quad F_\alpha(\alpha^H) = \begin{cases} 1, & \text{if } \alpha^H < \alpha_0^H \\ 0, & \text{if } \alpha^H = \alpha_0^H \\ -1, & \text{if } \alpha^H > \alpha_0^H \end{cases} \tag{10}$$

where γ is a positive weighted coefficient.

Optimal labeling of the mixtures is obtained through the condition $\lambda = \arg\min(U_1 + U_2 + U_3)$, where λ corresponds to the three possible labels: GM, WM, and CSF.

The deterministic relaxation iterated conditional modes (ICMs) [29] is used for the two steps of the classification. Although the ICM usually converges to a local minimum of the energy function, this loss of optimality may be compensated for by an appropriated initial guess. In our experiments, we use U_1 for initializing labels with the parameters of distributions obtained by the histogram fit.

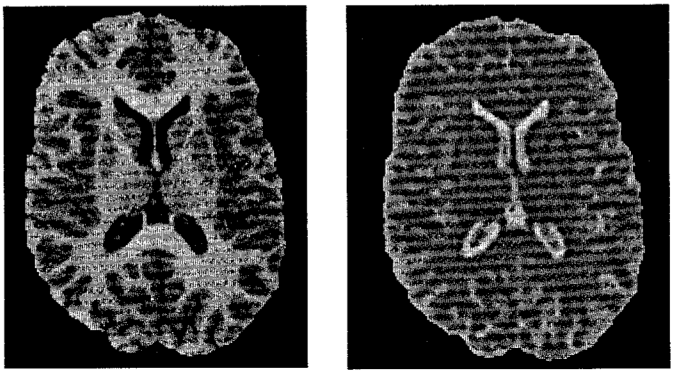
V. Validation and Results

It is impossible to obtain MR images of the human brain with known proportions of mixel components within each voxel. Therefore, a quantitative assessment of the performance of the classification method requires the use of simulated data. While testing with simulated data is necessary, it is not sufficient to validate the model. Additional testing with real data must also be completed to demonstrate the ability of the algorithm to work under real-world conditions. For this reason, we also used several real MRI datasets acquired on the same subject to test the reliability of the algorithm.

A. Simulated Data

The digital brain phantom, available on the site BrainWeb [30], [31], is used. This 3-D phantom is made up of ten volumetric data sets that define the spatial distribution for different tissues (e.g., GM, WM, CSF, etc.), where voxel intensity is proportional to the fraction of tissue within the voxel. Based on this phantom, different types of simulated image can be created. T1-weighted MRI simulated volumes with different noise levels are also available on this site, which are also used in our validation.

Each simulated volume consists of $181 \times 217 \times 181$ voxels and each voxel is $1 \times 1 \times 1$ mm^3. The noiseless volume was first segmented by the software (ATOMIA) developed in our laboratory [32] to obtain the encephalon mask where the three brain tissues are localized. The layer of glial tissue in the mask is considered as GM, since its intensity is similar to T1-weighted images. The gold standard is created using the encephalon mask in which each voxel is labeled according to the tissue type whose proportion is the highest.



(a) (b)

Fig. 5. Example of a (a) T1-weighted SPGR MR image and (b) the corresponding α^H image.

B. Performance Evaluation Criteria

Three criteria were used to assess the performance of the segmentation strategy.

1) False Positive and False Negative Ratios: The false positive ratio ξ_{fp} is defined as the number of misclassified voxels divided by the number of true voxels taken from the gold standard (N). The false negative ratio ξ_{fn} is defined as the number of falsely rejected voxels divided by N. Since ξ_{fp} and ξ_{fn} inversely vary as a function of the weighting parameters (β, γ) of the energy function U, a better quality index is given by the sum of ξ_{fp} and ξ_{fn}, noted ξ_{total}. However, this index is global, which is why we have also analyzed the errors as a function of their contents, described below.

2) Histogram of Error: Here, we propose to show the errors in terms of a histogram, which allows us to know where the errors appear with respect to the pure tissue proportions. The proportions of each class were first gathered in 10% ranges. The ratio of the number of classified voxels to the number of gold standard voxels within each range was calculated. The ideal curve is a step function (jumping at 50%). The higher the value of the histogram for proportions >50% and the lower for proportions <50%, the better the results. In this way, histograms corresponding to the three classes can be drawn for the final results.

3) Coefficient of Agreement κ: Cohen [33] has developed a coefficient of agreement denoted κ, which is a chance-corrected percent agreement with a statistical base. This coefficient allows us to measure the similarity index between two results. κ is defined as

$$\kappa = \frac{P_0 - P_c}{1 - P_c} \tag{11}$$

where P_0 is the proportion of voxels where both data sets are in agreement. P_c is the proportion of agreement due to chance. (for more detail, see [33]). $\kappa = 1$ is obtained whenever there is perfect agreement between the two data sets. $\kappa = 0$ indicates that agreement is due to chance alone. κ between zero and one reflects agreement greater than chance, whereas negative values indicate agreement less than chance.

C. Validation Using the Simulated Data

We first aimed to evaluate the estimation method of the two parameters β and γ used in our model. In the decomposition of U, each term expresses a different interaction model, each of which contributes to the global classification. The two sets of parameters β_λ (λ denotes the label) and γ weight the interactions between them. Optimal parameters are obtained when ξ_{total} is minimum. In order to find the minimum ξ_{total} and corresponding parameters, we have tested each combination of parameters by changing β_λ from zero to 0.5 with an incremental step of 0.1 and γ from zero to five with a step of 0.5. It should be pointed out that among the optimal parameter set, γ is not zero ($\gamma = 3$), which shows that U_3 improves the classification. However, these values are obtained only by simulated images (noise level $= 3\%$, RF inhomogenity $= 0$). For real images, the parameters should be adapted to each given volume.

We propose a simple method to calculate β_λ for each class as a function of U_1, and use the optimal value γ. With the smoothness assumption, the fact that a current voxel is an isolated voxel within the neighborhood system is not acceptable even if its U_1 is minimum. The minimum U_1 under the class label λ_1 is obtained when the gray level y_v of the voxel takes the mean value $\mu_{\lambda 1}$. The class label of other voxels in the neighborhood system is supposed λ_2. In order to change the class label λ_1 of this isolated voxel to λ_2, the following condition should be satisfied:

$$U_1(\mu_1/\lambda_2) + \beta_2 Ns > U_{1_{\min}}(\mu_1/\lambda_1) \tag{12}$$

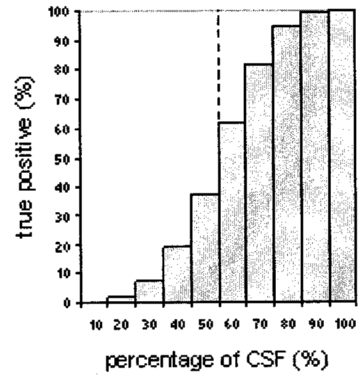
where $Ns = 18$ (18-adjacency is used). β_λ can be calculated for each class in this way. A comparison between obtained errors using optimal parameters and those calculated from (12) is shown in Table I. It is clear that classification accuracy with the calculated parameters is slightly lower than with the optimal parameters. However, the difference is acceptable. It can be seen that the errors of the CSF classification are higher than those of the other tissue types. Two reasons for this are evident: first, the CSF voxels in encephalon are mostly within the sulci, therefore, they are prone to partial volume effects; second, there are fewer CSF voxels in the encephalon than there are of the other tissue types, implying a high relative error ratio.

In order to take this last issue into account, errors are plotted in terms of a histogram. Because of the presence of mixels, the error histogram can provide an assessment the classification error as a function of the true proportion of tissue constituents. The obtained histograms for the three pure tissue classes are portrayed in Fig. 6. Compared to the ideal function (the step function shown as a dotted line), the voxels in the range of 90% to 100% are well classified. Errors can be mostly observed close to 50%. From these histograms, the results of CSF are not the worst, and are even better than that of GM. Indeed, the surface of the histogram from 50% to 100% is greater than that of GM, while the surface from 0% to 50% is less than that of GM.

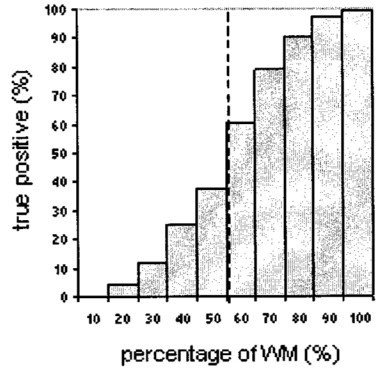
The impact of noise is analyzed using the volumes proposed in the same BrainWeb (the noise level varying from 1% to 9% and RF inhomogeneity equal to 20%). ξ_{total} and κ were calculated (Fig. 7) in comparison to the gold standard. The conclusion is not exactly the same for the two indexes. Because ξ_{total} depends on the number of voxels within the class, ξ_{total} of CSF is the highest and those of GM and WM are almost identical. Whereas, κ indicates that the GM result is the worst. This is logical because the GM has two borders: one with the CSF and the other with the WM. As a result, there is less certainty about the constitution of the mixclasses.

TABLE I
COMPARISON OF THE RESULTS OBTAINED USING OPTIMAL PARAMETERS, NOTED P1 AND USING THE PARAMETERS CALCULATED BY THE PROPOSED METHOD, P2

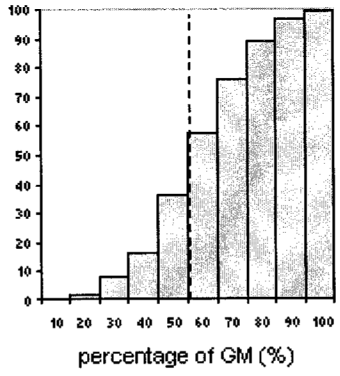
class	ξ_{fp} (%)		ξ_{fn} (%)		ξ_{total} (%)	
	P1	P2	P1	P2	P1	P2
WM	2.28	2.60	3.72	3.51	6.01	6.11
GM	3.82	4.41	2.51	2.66	6.33	7.07
CSF	3.17	3.47	4.82	8.81	7.99	12.28



(a)

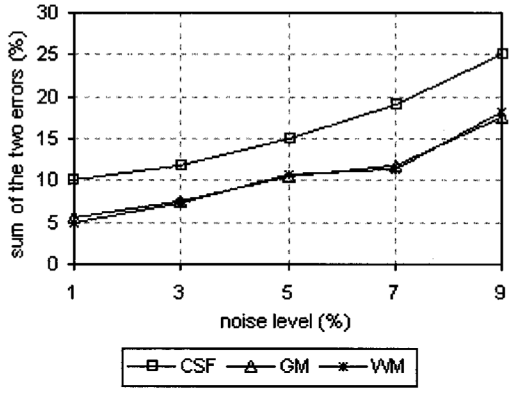


(b)

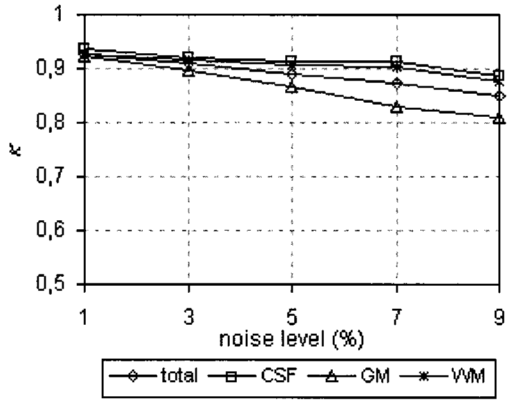


(c)

Fig. 6. Error histogram as a function of percentage of pure tissue. Results evaluated by percentages of classification in each 10% range (vertical axis) versus the pure tissue proportion in this range (horizontal axis) for (a) CSF, (b) WM, and (c) GM.



(a)



(b)

Fig. 7. Performance of the classification method using the simulated T1-weighted MRI images, under varying conditions of noise (RF inhomogeneity = 20%, slice thickness = 1 mm): (a) ξ_{total}, (b) κ.

In any case, the performance is satisfactory. We have also compared our results to those obtained by seven other methods, presented in [34] [such as artificial neural networks (ANN), K-nearest neighbors, Bayes, etc.). They used the same phantom. The results of these seven methods are obtained using a fusion of T1, T2, and PD-weighted image volumes. The author shows that the best performance is obtained by the ANN method, which derives benefit from a set of training samples. In the worst case (noise level = 9%, RF inhomogeneity = 20%), the obtained κ value is about 0.81, while our method yields a better result, $\kappa = 0.85$. In addition, our method is unsupervised and requires only T1-weighted images.

D. Results on MRI Data

For real brain MR images, a quantitative assessment is very difficult. Consequently, we propose to analyze statistical differences between the classification results, as a component of assessing the validity of the algorithm. Two series of five data sets with acquisition sequences SPGR (TE = 7 ms, TR =30 ms, $\alpha = 40°$) and IR-FGRE (TE = 2 ms, TR = 12 ms, $\alpha = 10°$) were performed on a 1.5-T GE Signa scanner for the same subject. Each dataset consists of 256× 256 × 124 voxels with a spatial resolution of 0.9375× 0.9375 × 1.2 mm^3. The signal to noise ratios (SNRs) of the two types of images are different even if they are both T1-weighted. For the comparison, the five brain data sets of each series were first realigned by the AIR package [35].

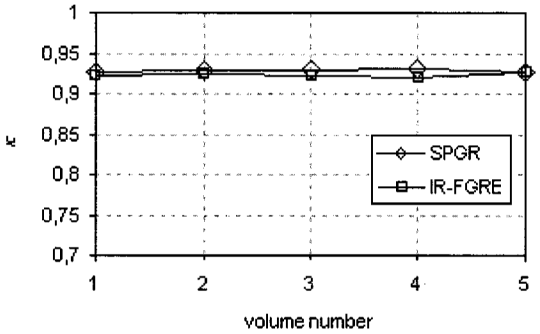
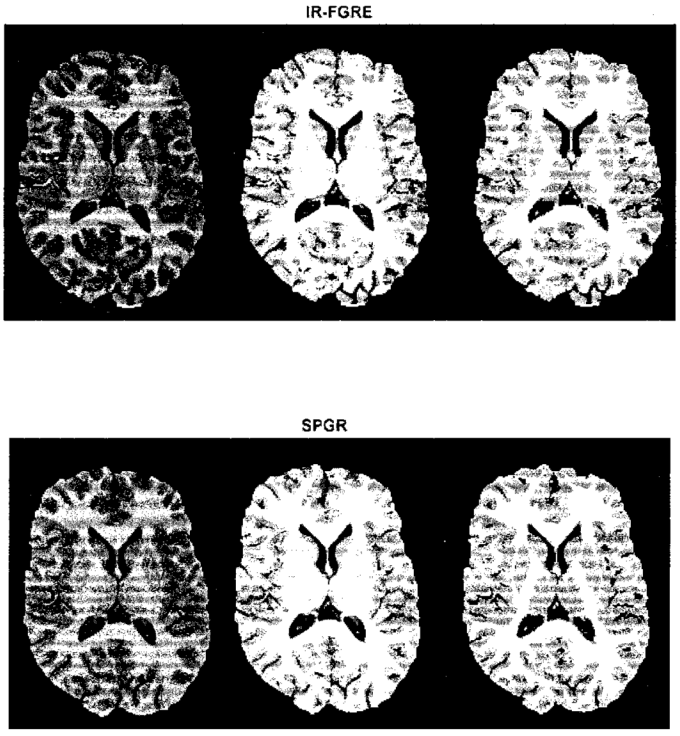


Fig. 8. The κ values for the SPGR and the IR MR sequences as a function of volume number. The reference volume taken to calculate κ is the mean volume for each sequence. The obtained values are almost constant for both acquisition sequences.



(a) (b)

Fig. 9. Results of the classification algorithm. Column (a) shows two axial MRI images (IR-FGRE and SPGR), Column (b) shows results of classification into five classes: WM (white color), GW (light gray), GM (gray), GC (dark gray), and CSF (black) classes. Column (c) shows the final results (WM, GM, and CSF).

The mean volume of each series was then calculated, whose SNR was improved in relation to that of each volume by a factor $\sqrt{5}$. All data sets were first edited by ATOMIA and then segmented by the proposed method. After modeling the histogram with five Gaussian functions, the values of the risk α for normality rejection for both sequences were calculated. They are in the range of acceptability (r is about 1.1, l is about 4.5, $\alpha < 1\%$) for both mixtures CG and GW. The results of each data set were compared to the mean volume in order to show the stability and consistency of the algorithm. κ values are shown in Fig. 8. The plots indicate that the algorithm yields almost constant errors, thereby proving the stability of the method. A visual comparison of the results is shown in Fig. 9. The results of labeling

into five classes [Fig. 9(b)] and three classes [Fig. 9(c)] are displayed for one axial slice. As expected, the obtained mixtures can be mostly observed at the pure tissue boundaries and also in subcortical locations (such as the globus pallidus and thalamus), that are known to have a high mixture of GM and WM. The final results of classification in three classes show that these regions are in general correctly segmented. As the subject is the same for both acquisition sequences, the results of SPGR and IR-FGRE are very similar. Unfortunately, we could not carry out a more precise comparison between the two sequences because of inaccuracy in the voxel dimension caused by the MR imaging system (this problem is discussed in [36]).

VI. Conclusion

A promising method is presented for segmenting WM, GM and CSF volumes using a 3-D MRF model. The algorithm is unsupervised, fully automatic, and uses only T1-weighted images. Mixtures of multiple tissue types within a voxel are taken into account in the process of classification. The Gaussian distribution of the mixel intensities is demonstrated, which allows us to simplify the image model, as well as the segmentation method. First, the two-stage algorithm identifies the pure classes and mixclasses, and then processes these mixclasses with an additional energy term in the MRF model, obtained from multifractal analysis. The results were quantified both on simulated and real MR images to demonstrate the robustness and the stability of the algorithm with respect to the noise level and to different T1-weighted MR images. Apart from the global assessment criteria which are widely used, we propose using the histograms of errors described in Section V. They provide information about misclassification in relation to the proportion of tissue types. Due to poor contrast, some inaccuracy is present in the gray nuclei region, therefore, improvement in the method based on *a priori* anatomical knowledge is needed and is actually in progress.

Acknowledgment

The authors would like to thank Prof. B. Mazoyer for providing the MRI data used in this study. They would also like to thank B. Moretti for his help, particularly in the acquisition and registration of MR volumes.

References

[1] J. C. Rajapakse, J. N. Giedd, C. DeCarli, J. W. Snell, A. McLaughlin, Y. C. Vauss, A. L. Krain, S. Hamburger, and J. L. Rapoport, "A technique for single channel MR brain tissue segmentation: Application to a pediatric sample," *Magn. Res. Imag.*, vol. 14, pp. 1053–1065, 1996.

[2] T. Taxt and A. Lundervold, "Multispectral analysis of the brain using magnetic resonance imaging," *IEEE. Trans. Med. Imag.*, vol. 13, pp. 470–481, 1994.

[3] J. C. Rajapakse, C. DeCarli, A. Mclaughlin, J. N. Giedd, A. L. Krain, S. D. Hamburger, and J. L. Rapoport, "Cerebral magnetic resonance image segmentation using data fusion," *J. Comput. Assist. Tomogr.*, vol. 20-2, pp. 206–218, 1996.

[4] H. Suzuki and J. Toriwaki, "Automatic segmentation of head MRI images by knowledge guided thresholding," *Comput. Med. Imag. Graph.*, vol. 15, pp. 233–240, 1991.

[5] A. Simmons, S. R. Arridge, G. J. Barker, A. J. Cluckie, and P. S. Tofts, "Improvement to the quality of MRI cluster analysis," *Magn. Res. Imag.*, vol. 12, pp. 1191–1204, 1994.

[6] M. Joliot and B. Mazoyer, "Three-dimensional segmentation and interpolation of magnetic resonance brain images," *IEEE Trans. Med. Imag.*, vol. 12, pp. 269–277, 1993.

[7] M. M. Chang, M. I. Sezan, A. M. Tekalp, and M. J. Berg, "Bayesian segmentation of multislice brain magnetic resonance imaging using three-dimensional Gibbsian priors," *Opt. Eng.*, vol. 35, pp. 3206–3221, 1996.

[8] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger, "Review of MRI segmentation: Methods and applications," *Magn. Res. Imag.*, vol. 13-3, pp. 343–368, 1995.

[9] J. M. Links, L. S. Beach, B. Subramana, M. A. Rubin, J. G. Hennessey, and A. L. Reiss, "Edge complexity and partial volume effects," *J. Comput. Assist. Tomogr.*, vol. 22-3, pp. 450–458, 1998.

[10] F. Bello, A. C. F. Colchester, and S. A. Roll, "A geometry- and intensity-based partial volume correction for MRI volumetric studies," *Comput. Med. Imag. Graph*, vol. 22-2, pp. 123–132, 1998.

[11] C. R. G. Guttmann, R. Kikinis, M. C. Anderson, M. Jakab, S. K. Warfield, R. J. Killiany, H. L. Weiner, and F. A. Jolesz, "Quantitative follow-up of patients with multiple sclerosis using MRI: Reproducibility," *J. Magn. Reson. Imag.*, vol. 9-4, pp. 509–518, 1999.

[12] D. J. Peck, J. P. Windham, L. L. Emery, H. Soltanian-Zadeh, D. O. Hearshen, and T. Mikkelsen, "Cerebral tumor volume calculations using planimetric and eigenimage analysis," *Med. Phys.*, vol. 23-12, pp. 2035–2042, 1996.

[13] H. S. Choi, D. 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, June 1991.

[14] F. Masulli and A. Schenone, "A fuzzy clustering based segmentation system as support to diagnosis in medical imaging," *Artif. Intell. Med.*, vol. 16-2, pp. 129–147, 1999.

[15] 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, Feb. 1998.

[16] P. Santago and H. D. Gage, "Quantification of MR images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, vol. 12, pp. 566–574, June 1993.

[17] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H.-W. Müller-Gärtner, "Markov random field segmentation of brain MR images," *IEEE. Trans. Med. Imag.*, vol. 16, pp. 878–886, Dec. 1997.

[18] 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.

[19] N. Sarkar and S. Katsuragawa, "Multifractal and generalized dimension of gray-tone digital images," *Signal Processing*, vol. 42-1, pp. 181–190, 1995.

[20] J. Levy-Vehel, P. Mignot, and J. P. Berroir, "Texture and mulitifractals: New tools for image analysis," INRIA, France, Res. Rep. , 1992.

[21] C. Jaggi, "Segmentation par méthode Markovienne de l'encéphale humain en imagerie par résonance magnétique: Théorie, mise en œuvre et évaluation," Ph.D. dissertation, University of Caen, Caen, France, 1998.

[22] A. L. Ressi, J. G. Hennessey, M. Rubin, L. Beach, M. T. Abrams, I. S. Warsofsky, A. M. C. Liu, and J. M. Links, "Reliability and validity of an algorithm for fuzzy tissue segmentation of MRI," *J. Comput. Assist. Tomogr.*, vol. 22, no. 3, pp. 471–479, 1998.

[23] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.

[24] D. K. Panjwani and G. Healey, "Markov random field models for unsupervised segmentation of textured color images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 939–954, Oct. 1995.

[25] N. Giordana and M. Pieczinski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE. Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 465–475, 1997.

[26] S. M. Choi, J. E. Lee, and M. H. Kim, "Volumetric object reconstruction using the 3D-MRF model-based segmentation," *IEEE. Trans. Med. Imag.*, vol. 16, pp. 887–892, Dec. 1997.

[27] R. B. D'Agostino, A. Belanger, and R. B. D'Agostino Jr., "A suggestion for using powerful and informative tests of normality," *American Statist.*, vol. 44, pp. 316–321, 1990.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed, ser. The Art of Scientific Computing. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[29] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. Ser. B*, vol. 36, pp. 192–242, 1974.

[30] A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans. Brain web: Online interface to a 3D MRI simulated brain database. [Online] . Available: http: //www.bic.mni.mcgill.ca/brainweb

[31] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, pp. 463–468, June 1998.

[32] B. Moretti, J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer, "Phantom-based segmentation assessement for MRI images," in *Proc. IEEE-EMBS Conf.*, Hong Kong, Oct. 1998, pp. 969–972.

[33] J. Cohen, "A coefficient of agreement for nominal scales," *Education Psychological Meas.*, vol. 20, pp. 37–46, 1960.

[34] V. Kollokian, "Performance analysis of automatic techniques for tissue classification in MRI of the humain brain," Master's thesis, Concordia Univ., Montreal, P.Q., Canada, Nov. 1996.

[35] R. P. Woods, J. C. Mazziotta, and S. R. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, vol. 17, pp. 536–546, 1993.

[36] D. L. G. Hill, C. R. Maurer, J. C. Studholme, J. M. Fitzpatrick, and D. J. Hawkes, "Correcting scaling errors in tomographic Images using a nine degree of freedom registration algorithm," *J. Comput. Assist. Tomogr.*, vol. 22-2, pp. 317–323, 1998.