Expert Systems with Applications 41 (2014) 1492–1497

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Semi-supervised clustering for MR brain image segmentation

Nara M. Portela [a], George D.C. Cavalcanti [a,*], Tsang Ing Ren [a,b]

[a] Centro de Informática, Universidade Federal de Pernambuco, Brazil
[b] iMinds, Vision Lab, Department of Physics, University of Antwerp, Belgium

ARTICLE INFO

Keywords:
Semi-supervised learning
Gaussian mixture model
Contextual segmentation
Magnetic resonance brain multispectral image

ABSTRACT

Magnetic resonance (MR) brain image segmentation of different anatomical structures or tissue types has become a critical requirement in the diagnosis of neurological diseases. Depending on the availability of the training samples, image segmentation can be either supervised or unsupervised. While supervised learning requires a sufficient amount of labelled training data, which is expensive and time-consuming, unsupervised learning techniques suffer from the problem of local traps. Semi-supervised algorithms that includes prior knowledge into the unsupervised learning can enhance the segmentation process without the need of labelled training data. This paper proposes a method to improve the quality of MR brain tissue segmentation and to accelerate the convergence process. The proposed method is a clustering based semi-supervised classifier that does not need a set of labelled training data and uses less human expert analysis than a supervised approach. The proposed classifier labels the voxels clusters of an image slice and then uses statistics and class labels information of the resultant clusters to classify the remaining image slices by applying Gaussian Mixture Model (GMM). The experimental results show that the proposed semi-supervised approach accelerates the convergence and improves the results accuracy when comparing with the classical GMM approach.

1. Introduction

Nowadays, Magnetic resonance imaging (MRI) is extensively used for the diagnosis of neurological diseases. Currently available procedures ensure safe, painless, and non-invasive investigation of the human body and it often identifies abnormalities long before and diseases symptoms appear. Magnetic resonance imaging, in particular, is well suited for studying diseases of the nervous system due to the high spatial resolution, the high soft tissue contrast, and the multispectral characteristics of MR images with relaxation times (i.e., T1 and T2) and proton density (i.e., Pd) information.

Analysis of MRI by a trained human expert is a tedious and difficult task because the structures of interest in the image shows complex edge configurations and anatomical borders are not clearly visible most of the times. In clinical trials, the number of MR images is often so large that manual analysis by human experts is too time-consuming. Furthermore, it is not clear how an expert combines information obtained from different channels when multispectral MR data are examined. Since, the intra- and inter-observer variability associated with manual segmentations hinder the reproducibility of the results. For these reasons, automatic or semi-automatic techniques for MR brain image segmentation that can analyze large amounts of 3D multispectral MR data in a reproducible way are necessary (Suri, Wilson, & Laxminarayan, 2005; Zhang, Brady, & Smith, 2001).

A key component in image analysis, regarding quantitative measurements of the brain anatomy, is to obtain accurate segmentation of the brain image from the different anatomical structures or tissue types, especially gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Segmentation of the brain image is not only widely used for cortical surface mapping, volume measurement, tissue classification, functional and morphological adaptation assessment, and characterization of neurological disorders, but also it is a required preliminary step for many other image processing procedures, such as brain registration and voxel-based morphometry. Therefore, the accurate segmentation of the brain image has become one of the most important issues in MRI applications. Segmentation can be based on the image voxel attributes, neighborhood information, or geometric characteristics. The difficulties to obtain an accurate image segmentation arise from noise, inhomogeneities, partial volume effects and the highly convoluted geometry of the cortex.

Depending on the availability of labels for training samples, image segmentation can be either supervised or unsupervised. In general, segmentation based on supervised learning, such as neural

* Corresponding author. Address: Centro de Informática (CIn), Universidade Federal de Pernambuco (UFPE), Av. Jornalista Anibal Fernandes, Cidade Universitária, 50740-560 Recife, PE, Brazil. Tel.: +55 81 2126 8430x4346; fax: +55 81 2126 8438.

E-mail addresses: nmp@cin.ufpe.br (N.M. Portela), gdcc@cin.ufpe.br (G.D.C. Cavalcanti), tir@cin.ufpe.br (T.I. Ren).

URL: http://www.cin.ufpe.br/~viisar (G.D.C. Cavalcanti).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.046

networks (Haykin, 2008) or support vector machine (Duda, Hart, & Stork, 2001), can yield good results, but it requires a large amount of training data (labelled voxels) for every type of tissues, which is expensive and time-consuming.

In contrast, unsupervised learning methods, such as k-means (MacQueen, 1967), or mixture model based (Everitt & Hand, 1981), has well recognised advantages over supervised segmentation methods, for example few user interaction. Because almost all unsupervised techniques are in fact an optimization process that is governed by an objective function such as the total log likelihood in mixture modeling or the sum of Euclidean distance in k-means, the techniques inevitably suffer from the problem of local traps (minima or maxima). In consequence, they need to be tuned properly in order to produce satisfactory results (Qian & Li, 2004). In other words, without any prior knowledge, these methods have limited performance.

Semi-supervised algorithms that incorporates prior knowledge in the unsupervised method can improve the results of data classification without requiring a complete training data set (Chapelle, Schölkopf, & Zien, 2006; Pao, Chuang, Xu, & Fu, 2008; Song, Huang, Ma, & Hung, 2011). Recently, some techniques for semi-supervised brain image segmentation have been proposed. Examples of these techniques are the semi-supervised maximum a posteriori probability (ssMAP) (Li, Ogunbona, de Silva, & Attikiouzel, 2011) and the algorithm proposed by Song et al. (2009) to improve the segmentation results by exploring incomplete training datasets where the labelled data may be available only for a subset of tissues, that is, not every type of tissues are labelled. Zhang, Dong, Clapworthy, Zhao, and Jiao (2010) proposed a MR brain image segmentation using semi-supervised spectral clustering that improved the results of the segmented image. They supply some data pairwise constraints information, instead of labelled voxels, to the spectral clustering algorithm to obtain a better data assignment.

Here, we propose a method to improve MR brain image segmentation and accelerate the convergence process using clustering based semi-supervised classification without requiring a set of labelled training data and using less manual analysis by human expert than a supervised approach. Initially, the voxels of only one image slice are clustered and a human expert labels each resultant cluster as gray matter (GM), white matter (WM), or cerebrospinal fluid (CSF). Then, labelled information and clusters statistics measures, mean, covariance matrix, and prior probability, are used in the classification process of the others slices in the image. In comparison to other semi-supervised techniques, prior knowledge such as cluster label information and statistics measures are easily obtained without the demand of any labelled training data.

The remainder of this paper is organized as follows: a background on the Gaussian Mixture Model based clustering is presented in Section 2. The proposed method is presented in Section 3. Experimental results on the application of the proposed model to synthetic MR brain images are presented in Section 4 and conclusions are given in Section 5.

2. Clustering based on gaussian mixture model

Let X be an image slice represented by a set of voxels $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$ where each voxel is denoted by a d-dimensional random vector $\mathbf{x}^i = (x_1, \ldots, x_d)$ and n is the number of voxels in the slice. It is assumed that each region j that composes the image follows a class-conditional distribution with probability density function $p_j(\mathbf{x}|\theta_j)$, each one having its own vector of parameters θ_j, $j = 1, \ldots, k$, where k is the number of tissue classes in the image. Therefore, each voxel is drawn independently from the mixture density that describes the weighted sum of all classes taken together:

$$p(\mathbf{x}^i|\psi) = \sum_{j=1}^{k} \pi_j p_j(\mathbf{x}^i|\theta_j), \tag{1}$$

where π_j are the mixture mixing proportions that are positive and have the summazion equals to one, $p_j(\mathbf{x}|\theta_j)$ is the component density associated with region j and $\psi = \{\theta_1, \ldots, \theta_k, \pi_1, \ldots, \pi_k\}$ is the set of all mixture parameters. The mixing proportions π_j corresponds to the prior probability of any voxel belonging to the jth group.

To use mixture model-based clustering approach, it is assumed that the data to be clustered belongs to a mixture of a specified number of k groups in various proportion. Each data point is drawn independently from the mixture density in Eq. (1), where the k components correspond to the k groups. After the specification of the parametric form for each component density $p_j(\mathbf{x}|\theta_j)$, the parameters can be estimated by maximum likelihood. Once the mixture is fitted, a probabilistic clustering of the data can be obtained in terms of the fitted posterior probabilities of component membership for the data. The data clustered into k groups is obtained by assigning each data point to the component to which it has the highest estimated probability (McLachlan & Peel, 2000).

Assuming that $p_j(\mathbf{x}|\theta_j)$ are normal with parameters, $\theta_j = \{\mu_j, \Sigma_j\}$, and the variables $\mathbf{x}^i$ are independent and identically distributed (i.i.d). Let μ_j and Σ_j be the unknown mean vector and covariance matrix of the group j respectively, so that:

$$p_j(\mathbf{x}^i|\boldsymbol{\theta}_j) = \frac{1}{2\pi^{d/2}|\Sigma_j|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{x}^i - \mu_j)\Sigma_j^{-1}(\mathbf{x}^i - \mu_j)^T\right], \tag{2}$$

where $|\Sigma|$ denotes the determinant of Σ.

The step after formalize the model is fit the mixture to the data through the maximum likelihood (ML) estimation (McLachlan & Peel, 2000). The objective of ML is to obtain the parameters that maximize the joint probability density function of the available data (or the data likelihood). This can be performed using the Expectation Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). For convenience, we usually use the log-likelihood value instead of the likelihood, which is defined as:

$$\ln\ p(X|\psi) = \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{k} \pi_j p_j(\mathbf{x}^i|\theta_j)\right). \tag{3}$$

McLachlan and Peel (2000) introduced an EM algorithm for ML estimation of the parameters of the component densities for the case where $p_j(\mathbf{x}|\theta_j)$ are assumed normal. In the image clustering approach, the slice voxels are used to update the mixture parameters by alternating the following EM steps until the $\ln\ p(X|\psi)$ converges:

– Expectation-step

$$z_j(\psi^t|\mathbf{x}^i) = \frac{\pi_j^t p_j\left(\mathbf{x}^i|\theta_j^t\right)}{p(\mathbf{x}^i|\psi^t)}, \quad j = 1, \ldots, k \tag{4}$$

– Maximization-step

$$\pi_j^{t+1} = \frac{1}{n}\sum_{i=1}^{n} z_j(\psi^t|\mathbf{x}^i), \quad j = 1, \ldots, k \tag{5}$$

$$\mu_j^{t+1} = \frac{\sum_{i=1}^{n} \mathbf{x}^i z_j(\psi^t|\mathbf{x}^i)}{n\pi_j^{t+1}}, \quad j = 1, \ldots, k \tag{6}$$

and

$$\sum_{j}^{t+1} = \frac{\sum_{i=1}^{n} (\mathbf{x}^i - \mu_j^{t+1})\left(\mathbf{x}^i - \mu_j^{t+1}\right)^T z_j(\psi^t|\mathbf{x}^i)}{n\pi_j^{t+1}}, \quad j = 1, \ldots, k. \tag{7}$$

The EM algorithm starts from an initial value ψ^0. A poor choice of ψ^0 can hamper the EM convergence, and in some cases where the likelihood is unbounded on the edge of the parameter space, the sequence of parameter estimates generated by the EM may diverge if ψ^0 is chosen too close to the boundary. Another problem with mixture models is that the likelihood equation usually have multiple roots corresponding to local maxima. Therefore the EM algorithm should be applied from a wide choice of starting values in search for the local maxima, which is very time-consuming.

The parameters ψ describes the global properties of the measurements $\mathbf{x}^i$ but they do not show how to assign labels to the voxels. After the k-components mixture model is fitted to obtain the estimate of ψ, a probabilistic clustering of the n voxels $\{\mathbf{x}^1,\ldots,\mathbf{x}^n\}$ can be assigned in terms of their fitted posterior probabilities of component membership. For each $\mathbf{x}^i$, the k probabilities $p_1(\psi_1|\mathbf{x}^i),\ldots,p_k(\psi_k|\mathbf{x}^i)$, with $\psi_j=\{\mu_j,\Sigma_j,\pi_j\}$, give the estimated posterior probabilities that this observation belongs to the fist, second, ... and k^{th} components, respectively, of the mixture, $\forall i$. A hard clustering answer of these voxels can be provided by assigning each $\mathbf{x}^i$ to the component of the mixture that has the highest posterior probability. Let $\mathbf{r}^i$ be the component-label vector such that $\mathbf{r}^i_j=1$ if the i^{th} voxel belongs to the j^{th} group, and $\mathbf{r}^i_j=0$, otherwise. The prior probability of any voxel belonging to the j group is π_j, so that Prob $(\mathbf{r}^i_j=1)=\pi_j,\forall i$. The probability that the voxel $\mathbf{x}^i$ belongs to the j group is given by the class-conditional density function $p_j(\mathbf{x}^i|\theta_j)$, that is taken to be the mixture components. If $\mathbf{r}^i_j=1$, then the corresponding voxel is generated from the component density $p_j(\mathbf{x}^i|\theta_j)$; i.e.,

$$p(\mathbf{x}^i|\mathbf{r}^i_j=1)=p_j(\mathbf{x}^i|\theta_j). \tag{8}$$

By Bayes rules, the posterior probabilities of component membership are given by:

$$p_j(\psi_j|\mathbf{x}^i)=Prob\left(\mathbf{r}^i_j=1|\mathbf{x}^i,\psi_j\right)=\frac{\pi_j p_j(\mathbf{x}^i|\theta_j)}{p(x^i|\psi)}, \tag{9}$$

where the denominator $p(\mathbf{x}^i|\psi)$ is given by Eq. (1) that is a class constant value dependent of the measurement $\mathbf{x}^i$. In general, $\mathbf{x}^i$ can be used to estimate $p_j(\psi_j|\mathbf{x}^i),\forall j$ by applying Eq. (9) if the parameters of the component densities ψ are known. The Bayes classifier could be used to assign class labels to the voxels, if the parameters ψ were known,

$$j^*=\underset{j}{argmax}\, p_j(\psi_j|\mathbf{x}^i). \tag{10}$$

The voxel i is assigned to the class j^* that has the highest posterior probability.

3. Image segmentation based on semi-supervised Gaussian mixture model

The proposed approach is based on four assumptions of the available data: (1) the data are produced by a mixture model, (2) there is a one-to-one correspondence between mixture components and brain tissue classes, (3) the mixture components are multinomial distributions of individual regions present in the image, and (4) the same brain tissue classes presented in different image slices are assigned to the same distribution.

Let $\mathbf{Y}=\{X^1,\ldots,X^m\}$ be a 3D MR image represented by a set of 2D slices where m is the number of slices in the image. The proposed method is composed of two modules: a prior clustering module that associates class labels information with clusters statistics measures (mean, covariance matrix, and prior probability); and a classification module that receives prior information from the first module to produce labelled images. A block diagram summarizing the overall process is presented in Fig. 1.

In the prior clustering module (Fig. 2), an MR image slice X^g is randomly selected from $\mathbf{Y}$ and clustered by the k-means algorithm (Duda et al., 2001) into three groups. The result is a segmented image without any association between image segments and brain tissues classes: gray matter (GM), white matter (WM), cerebrospinal fluid (CSF). In the next step, a human specialist evaluates the clustering result; if the image is well segmented, the human specialist associates each segmented slice cluster, c_j, to one of the class labels, $l=\{\text{GM,WM,CSF}\}$; otherwise, the slice X^g is segmented again with different initial parameters, ψ^0. The expert interference corresponds to the supervision in the semi-supervised learning methods.

At this point, each segment of the X^g image has a class label. We assume a one-to-one correspondence between mixture model components and tissue classes, and thus use c_j to indicate the jth mixture component, as well as the j^{th} class. Given a correspondence between tissue classes and clusters of X^g, the set of mixture parameters, $\psi_j=\{\mu_j,\Sigma_j,\pi_j\}$, of each class c_j can be computed over its assigned cluster voxels, $\mathbf{x}^i,\forall j,i$,

$$\pi_j=\frac{|n_j|}{n}, \tag{11}$$

$$\mu_j=\frac{\sum_{\mathbf{x}^i\in c_j}\mathbf{x}^i}{|n_j|}, \tag{12}$$

$$\Sigma_j=\frac{\sum_{\mathbf{x}^i\in c_j}(\mathbf{x}^i-\mu_j)(\mathbf{x}^i-\mu_j)^T}{|n_j|}, \tag{13}$$

where $|n_j|$ is the number of voxels belonging to the jth class. The pairs clusters and labels, and the set of mixture parameters, ψ_j, of each class represent the prior knowledge of the data set.

In the second module (Fig. 3), the prior knowledge of the data set is used to cluster the rest of the slices. The set of remaining slices, $\mathbf{Y}^*$, is composed by $\mathbf{Y}$ excluding the slice used in the first module, X^g. When a new slice of $\mathbf{Y}^*$ is presented to the classifier, tissues classes parameters $\psi_j=\{\mu_j,\Sigma_j,\pi_j\}$, $\forall j$, computed earlier are used as the initial parameters of the GMM to smooth the EM convergence. Since the parameters are extracted from a well segmented image, it is supposed that ψ_j leads the EM convergence to the global maxima.

We also assume that the voxels in slice X^g that corresponds to a tissue class c_j and the voxels in the rest of the slices that corre-

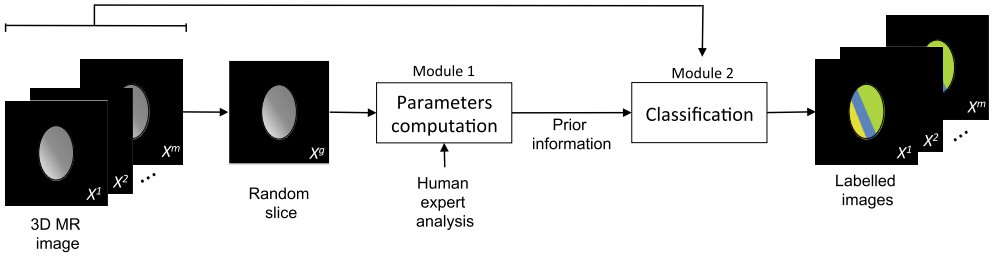

Fig. 1. Block diagram for the proposed method. Module 1 and Module 2 are detailed in Figs. 2 and 3, respectively.

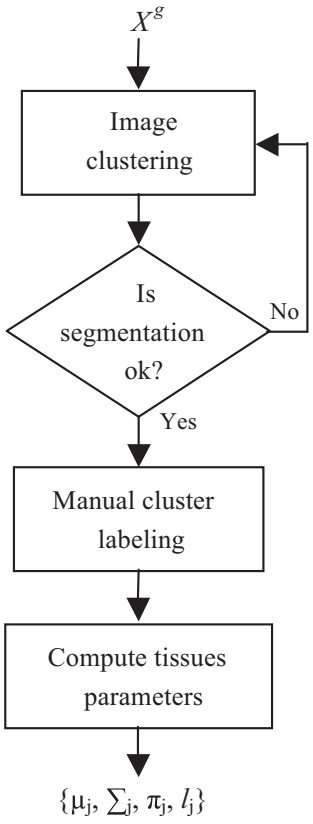

Fig. 2. First module: associating class labels information, l_j, with clusters statistics measures (mean, μ_j, covariance matrix, Σ_j, and prior probability, π_j).

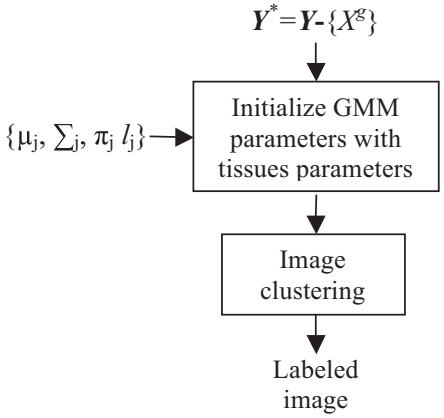

Fig. 3. Second module: using prior information (mean, μ_j, covariance matrix, Σ_j, prior probability, π_j, and class label, l_j) from the first module to produce labelled images.

sponds to c_j are generated by the same distribution. Therefore, different regions are assigned to different mixture parameters. Each mixture component from Eq. (1), $p_j(\mathbf{x}^i|\theta_j)$, corresponds to a tissue class, and receives the label assigned to the mixture parameters, ψ_j, used as the initial parameter. Therefore, after the GMM clustering is processed using the method described in Section 2, each voxel $\mathbf{x}^i$ is assigned to the labelled component of the mixture in which it has the highest posterior probability. The result is a segmented image where all voxels belonging to a cluster receives the class label of the mixture component assigned to the cluster. In other words, a new classification step is not necessary to associate the clusters from the segmentation process with each tissue class.

4. Experimental results

4.1. Image dataset

In order to quantitatively assess the performance of the proposed method, synthetic MR images from the BrainWeb simulated brain database (Cocosco, Kollokian, Kwan, Pike, & Evans, 1997) are used. The results reported in next sections are based on a synthetic normal brain multispectral image composed of proton density, T1- and T2-weighted images with 181 sagittal slices of dimension 181×217 voxels with 1 mm^3 resolution, 3% noise level, and 0% intensity non-uniformity.

Since, for the evaluation of the Gaussian density function the inverse of the covariance matrix is requires, an estimate of the covariance matrix is only useful for the classification if it is nonsingular (i.e., invertible). Among the 181 sagittal slices, we selected those slices that had at least 10 voxels belonging to each tissue class to ensure that it is be possible to estimate the parameters of the Gaussian distribution. This results in a total of 133 images. In this experiments, three classes are considered, white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). A preprocessing step is applied to separate brain from non-brain tissue before applying the image segmentation. Fig. 4 depicts the original BrainWeb images for slices number 50, 93 and 120 with 3% noise and 0% intensity non-uniformity in the left column, the same images after preprocessing step is shown in center column, and the respective ground truths is shown in right column. The images in the left column are a RGB version of the multispectral image composed of proton density, T1- and T2-weighted images.

4.2. Experimental design

We compare the proposed semi-supervised Gaussian Mixture Model (GMM^{SS}) with two others supervised Gaussian Mixture Model. The first method is a clustering procedure with random initialization parameter. The clustering result is a segmented image without labels, then a classification step is necessary to label the clusters. This method called GMM^{SUP1} is composed of 3 stages: training, clustering and clusters classification. In the training step, an image slice is randomly selected and clustered using the k-means algorithm with random parameter initialization. The resulting clusters are manually labelled as WM, GM or CSF and their statistical parameters are computed. As a consequence, each cluster is associated to a class label l_j and to a set of mixture parameters, ψ_j. In the second stage, the remaining image slices are segmented by a Gaussian Mixture Model with random parameters initialization. Next, the resultant clusters are classified by a parametric Bayesian classifier based on the parameters computed in the first step.

The second method used for comparison purpose is a supervised approach of the proposed method called GMM^{SUP2}. Since the image ground truth contains the class label of the validation data set, we use this available information to compute the set of mixture parameters, ψ, in a supervised way. In the first module described in Section 3, the clustering process is eliminated and ψ is computed directly from the labelled data using Eqs. (11)–(13). GMM^{SUP2} uses the second module as described in Section 3.

4.3. Evaluation methodology

To evaluate the classification results, the segmentations of each class j is compared with the ground truth by using the Dice Simi-

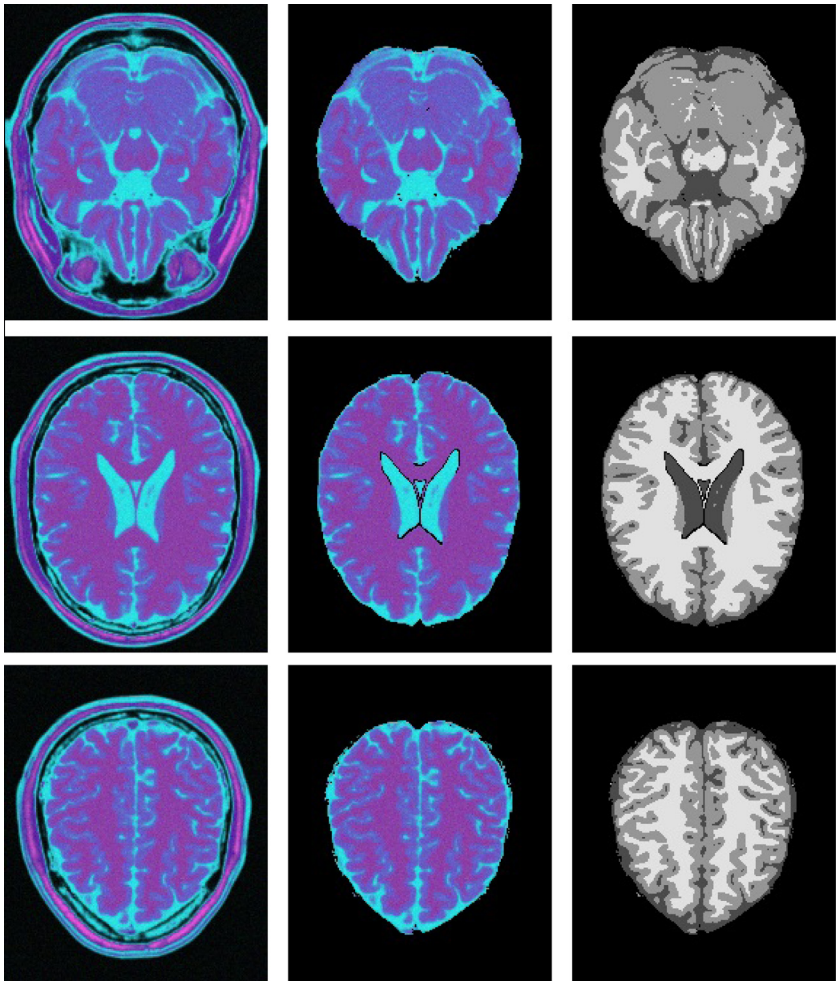

Fig. 4. (left column) Original BrainWeb images for slices number 50, 93 and 120 with 3% noise and 0% non-uniformity intensity, (center column) the same image after the preprocessing step, and (right column) the respective ground truth.

larity Index (DSI). The Dice Similarity Index S (j) (Dice, 1945) is defined as:

$$DSI = S(j) = \frac{2N_{p\cap g}(j)}{N_p(j) + N_g(j)}, \tag{14}$$

where $N_{p\cap g}(j)$ is the number of voxels classified as class j by the proposed method and the ground truth. $N_p(j)$ and $N_g(j)$ represent the number of voxels classified as class j by the proposed method and by the ground truth, respectively. The index S (j) tends to 1 if the proposed method coincides with the ground truth, and decreases towards 0 as the quality of the segmentation deteriorates. Typically, a value S (j) > 0.7 means that there is an excellent agreement between the two segmentations.

Since both implementations are sensitive to initialization, we run both methods 30 times. In every new execution, a new image is randomly selected for the first module and the remaining images are classified.

We use the pairwise t-test to compare the performance of the methods. The following convention for the p-value are used: "$\gg$" and "$\ll$" mean that the p-value is lesser than or equal to 0.01, indicating a strong evidence that a method results in a greater or minor value for the effectiveness measure than another method, respectively; "$>$" and "$<$" mean that the p-value is greater than 0.01 and lesser or equal to 0.05, indicating a weak evidence that a method results in a greater or minor value for the effectiveness measure than another method; "$\sim$" means that the p-value is greater than 0.05 indicating that it does not have significant difference when compared the performance of the two method.

4.4. Analysis of experiments

Table 1 shows the DSI values (mean and standard deviation), S (j), $\forall j$, obtained using different approaches and the class distribution average over all images. The best results are shown in bold. Table 2, which is derived form Table 1, shows the t-test results when comparing the performance of GMM^{SS} versus GMM^{SUP1} and GMM^{SS} versus GMM^{SUP2}.

Table 1
Tissue's DSI (mean ± standard deviation) for images segmentation using three techniques.

Tissue	Class distribution	Dice similarity index		
		GMM^{SUP1}	GMM^{SUP2}	GMM^{SS}
Cerebrospinal fluid	18.4%	0.860 ± 0.086	0.872 ± 0.054	**0.873 ± 0.053**
Gray matter	46.4%	0.733 ± 0.327	0.825 ± 0.210	**0.831 ± 0.203**
White matter	35.2%	0.705 ± 0.375	0.864 ± 0.224	**0.867 ± 0.221**
Average		0.766	0.854	**0.857**

Table 2
Comparison of the three different methods, GMM^{SS} × GMM^{SUP1} and GMM^{SS}× GMM^{SUP2}: t-test results.

	Tissue	Comparative method	
		GMM^{SUP1}	GMM^{SUP2}
GMM^{SS}	Cerebrospinal fluid	≫	∼
	Gray matter	≫	∼
	White matter	≫	∼

GMM^{SS} has statistically significant superior performance than GMM^{SUP1} for CSF, GM, and WM. The result can be explained by the fact that GMM clustering is highly affected by a poor choice of starting parameters and the proposed method avoid these effects by using the prior knowledge about the parameters provided by the semi-supervised approach. We can also observe that there is no statistically significant difference comparing GMM^{SUP2} with GMM^{SS} for CSF, GM, and WM. This indicates that the prior information provided by semi-supervised approach, used in the first module without any training data is similar to the information computed by a supervised approach over a set of training data.

The CPU time to process all 132 test images are obtained to compare the performance of the methods in relation to computational costs. The mean and standard deviation of the time in seconds of the proposed algorithm using an Intel Dual-Core 1.66 GHz Processor, 2G memory, and MATLAB Version 7.0, for the GMM^{SUP1}, GMM^{SUP2} and GMM^{SS} are 709.02 ± 994.72, 564.65 ± 431.23, 609.74 ± 204.37, respectively.

GMM^{SUP2} is faster than GMM^{SUP1} and GMM^{SS}. This result is already expected because GMM^{SUP1} and GMM^{SS} use clustering based semi-supervised approach to compute the set of initial mixture parameters, ψ, whereas (GMM^{SUP2}) computes ψ directly using the information from the ground truth. Thus, the time consumed by GMM^{SUP2} to compute ψ are due mainly to the direct application of the formulas shown in Eqs. (11)–(13), that is obviously smaller than the clustering convergence time considering the same data set.

However, when a training data set is not available, the GMM^{SUP2} is unable to perform. Therefore GMM^{SS} can be used since it performs better than (GMM^{SUP1}) yielding a statistically significant improvement (p-value lesser than 0.01 for a pairwise t-test). This experiment indicates that, since GMM is sensitive to initialization technique, the prior knowledge of the parameters results in a faster convergence for the cluster algorithm.

5. Conclusion

In this paper, we proposed a clustering based semi-supervised learning algorithm for human brain MR image segmentation. The proposed method avoids problem arising from poor initial choice of ψ during the mixture models parameter estimation when using the EM algorithm. We have developed an approach that performs the classification of human brain regions in MR image slices without requiring a set of labelled training data and using less manual analysis by human expert than a supervised approach.

The proposed technique is composed of 2 modules. First, an MR image slice is selected randomly and it is clustered by the k-means algorithm. Each cluster is labelled by an expert and their statistical parameters are computed. In the second module, tissues classes parameters values are used as initial parameters to Gaussian Mixture Model to cluster the remaining slices. Each mixture component represents a tissue distribution by keeping its parameter value as initial parameter. Consequently, after the GMM clustering process, the resultant clusters are already labelled. In other words, a new classification step is not necessary to associate the clusters from segmentation process with each tissue class.

The convergence of the GMM is accelerated by using a semi-supervised learning algorithm other than a unsupervised learning with the Bayesian classifier. The time consumed by the semi-supervised algorithm is smaller than the unsupervised learning with the Bayesian classifier because the clustering algorithm convergence is faster when the initial parameters are based on prior knowledge. Since GMM is sensitive to the initialization method, the prior knowledge of the parameters improves its accuracy. For future direction, we plan to include a contextual clustering approach to the proposed method using spatially variant finite mixture model.

Acknowledgment

The authors thank CAPES, CNPq (Ciência sem Fronteiras) and FACEPE for the financial support.

References

Chapelle, O., Schölkopf, B., & Zien, A. (Eds.). (2006). *Semi-Supervised Learning*. Cambridge, MA: MIT Press.

Cocosco, C. A., Kollokian, V., Kwan, R. K. S., Pike, G. B., & Evans, A. C. (1997). BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage, 5*, 425.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*, 1–38.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology, 26*, 297–302.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2 ed.). New York: Wiley.

Everitt, B., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman and Hall.

Haykin, S. (2008). *Neural Networks and Learning Machines* (3 ed.). Ny: Springer.

Li, W., Ogunbona, P., de Silva, C., & Attikiouzel, Y. (2011). Semi-supervised maximum a posteriori probability segmentation of brain tissues from dual-echo magnetic resonance scans using incomplete training data. *IET Image Processing, 5*, 222–232.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. L. Cam & J. Neyman (Eds.), *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). University of California Press.

McLachlan, G. J., & Peel, D. (2000). *Finite Mixture Models* (1 ed.). New York: John Wiley and Sons, Inc..

Pao, H. T., Chuang, S. C., Xu, Y. Y., & Fu, H. C. (2008). An EM based multiple instance learning method for image classification. *Expert Systems with Applications, 35*, 1468–1472.

Qian, T., & Li, M., (2004). Multispectral MR images segmentation using SOM network. In *International conference on computer and information technology* (pp. 155–158).

Song, E., Huang, D., Ma, G., & Hung, C. C. (2011). Semi-supervised multi-class adaboost by exploiting unlabeled data. *Expert Systems with Applications, 38*, 6720–6726.

Song, Y., Zhang, C., Lee, J., Wang, F., Xiang, S., & Zhang, D. (2009). Semi-supervised discriminative classification with application to tumorous tissues segmentation of MR brain images. *Pattern Analysis and Applications, 12*, 99–115.

Suri, J. S., Wilson, D., & Laxminarayan, S. (2005). *Handbook of Biomedical Image Analysis. Segmentation Models Part B (Topics in Biomedical Engineering International Book Series)* (vol. 2). Secaucus, NJ, USA: Springer-Verlag New York, Inc..

Zhang, X., Dong, F., Clapworthy, G., Zhao, Y., & Jiao, L. (2010). Semi-supervised tissue segmentation of 3D brain MR images. In *International Conference on Information Visualisation* (pp. 623–628).

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging, 20*, 45–57.