Neurocomputing 114 (2013) 118–131

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Improving MRI segmentation with probabilistic GHSOM and multiobjective optimization

Andrés Ortiz [a], Juan M. Górriz [b,*], Javier Ramírez [b], Diego Salas-González [b], For the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Department of Communications Engineering, University of Málaga, Spain
[b] Department of Signal Theory, Networking and Communications, University of Granada, Spain

ARTICLE INFO

Available online 30 October 2012

Keywords:
MRI
Image segmentation
Multiobjective optimization
Self-Organizing Maps

ABSTRACT

In the last years, the improvements in Magnetic Resonance Imaging systems (MRI) provide new and additional ways to diagnose some brain disorders such as schizophrenia or the Alzheimer disease. One way to figure out these disorders from a MRI is through image segmentation. Image segmentation consist in partitioning an image into different regions. These regions determine different tissues present on the image. This results in a very interesting tool for neuroanatomical analyses. In this paper we present a segmentation method based on the Growing Hierarchical Self-Organizing Map and multiobjective-based feature selection to optimize the performance of the segmentation process. Since the features extracted from the image result crucial for the final performance of the segmentation process, optimized features are computed to maximize the performance of the segmentation process on each plane. The experiments performed on this paper use real brain scans from the Internet Brain Segmentation Repository (IBSR) and the Alzheimer Disease Neuroimaging Initiative (ADNI). Moreover, a comparison with other methods using the IBSR database shows that our method outperforms other algorithms.

1. Introduction

Nowadays, the improvement in the resolution of Magnetic Resonance Imaging (MRI) systems provides new and additional ways to diagnose some brain disorders such as schizophrenia or the Alzheimer disease. Moreover, some disorders can be diagnosed even before the manifestation of any physical symptoms. At the same time, they could help to discover the cause of brain disorders or anomalies. There is a vast amount of information contained in a MRI. On the one hand, MRI systems provide 3D volumes, composed as a set of slices. On the other hand, current MRI systems provide 16-bit depth 3D images. However, human eye is not able to distinguish more than several tens of gray levels. This way, computers can be used to exploit all the information contained in a MRI. Indeed, computer aided tools have become a very valuable tool for diagnosing some brain disorders such as the Alzheimer disease [15]. One way to figure out brain disorders from a MRI is through image segmentation. Image segmentation consists in partitioning an image into different regions which determine different tissues. This results in a very interesting tool for neuroanatomical analyses and Computer Aided Diagnosis (CAD). Image segmentation can be addressed by classifying a set of features extracted from the image. These features can be classified into two groups. The first group contains the first-order features which does not takes into account the relationship among pixels. The second group contains the second order features which are computed extracting data from the relationship among the pixels belonging to a neighbourhood [12]. This way, we extract first order and second order features to discriminate different tissues on the image. Nevertheless, not all the features are discriminant enough on all the images. Moreover, some of these extracted features could diminish the performance of the classifier. Actually, the feature extraction and selection imposes a key point on the segmentation process and will determine the overall segmentation performance. Once a set of features has been extracted from the image, a classification stage will group the pixels by means of its similarity. This way, we use the Growing Hierarchical Self-Organizing map (GHSOM) [10], a variant of the Self-Organizing Map (SOM) [16], which classifies

* Corresponding author. Tel./fax: +34 958 243271.
E-mail addresses: aortiz@ic.uma.es (A. Ortiz), gorriz@ugr.es (J.M. Górriz), javierrp@ugr.es (J. Ramírez), dsalas@ugr.es (D. Salas-González).

[1] Data used in preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.08.047

the data in an unsupervised way and allows discovering inherent hierarchies on the data. In addition, we use aprobability-based clustering scheme [22,23,21] to redefine the clusters which has been proved to be effective for improving the classification performance.

1.1. Summary and organization

After this introduction, the databases used in this work are presented. These databases contain real brain scans and are used to optimize the feature set, train the classifier and to test the overall performance of our algorithm. This section also presents the feature extraction process accomplished to compose the feature space for the classifier and the use of the GHSOM as a classifier as well as the way we have improved it with a probabilistic behaviour. This modification on the GHSOM is presented in Section 2.3. A brief background in SOM and GHSOM is also provided in this section. In addition, Section 2.4 presents the way that GHSOM can be used for image segmentation. After describing the materials and methods used in this work, Section 3 presents the experiments conducted to check the performance of our method as well as the discussion of the results. In this section, experiments using PCA and multiobjective optimization to reduce the dimensionality of the feature space are also detailed and image segmentation examples are provided. Finally, paper concludes in Section 4.

2. Materials and methods

In this section, we present three subsections which summarize the segmentation methods and the image databases used in this work to evaluate the proposed method. Segmentation references are considered as the ground truth in this work.

2.1. Databases

The performance of our proposal has been checked using real MRI images from two different sources. One of these sources is the Internet Brain Image Repository (IBSR) from the Massachusetts General Hospital [14]. This repository consist of 20 T1-weighted images from 22 to 35 years old healthy patients. Scalp and skull are already extracted in these images. Details regarding age and sex of each patient as well as the number of scans are provided in [14]. Overlap comparison metric is provided for each volume for comparing different segmentation methods. Consequently, images from the IBSR 1.0 database were used to compute the average overlap metric in order to compare to other previously proposed algorithms. The second source of images we used in this paper is the Alzheimer's Disease Neuroimaging Initiative (ADNI) [1]. This database was created to study the advance of the Alzheimer's disease, collecting a vast amount of MRI and Positron Emission Tomography (PET) images as well as blood biomarkers and cerebrospinal fluid analyses. Thus, data used in the preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (*adni.loni.ucla.edu*). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year publicprivate partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across USA and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55–90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org. We use T1-weighted MRI ADNI images in this paper to check the performance of our algorithm.

2.2. Feature extraction

Feature extraction consists in computing some properties or features from the original data set (image) which allows distinguishing among input patterns. Thus, the features computed from the image have to be descriptive enough. In fact, this imposes a key point for the classification performance. As a prior step, preprocessing the image is usually necessary to remove background noise or other parts on the image not used for the classification. This is the case of skull removal in ADNI images. Regarding background noise removal, it can be addressed building a binary mask to detect the greatest contiguous object. After multiplying the binary mask with the original image, we get the background in black. Moreover, the image is centered in order to avoid pixel losing with the sliding window described afterwards in this section. Nevertheless, the images from the IBSR database have these undesired structures and noise already removed. Thus, we are focused on brain tissue segmentation and we do not perform skull stripping or noise removal. Extracted features have to be descriptive enough to differentiate and recognize patterns present on the image. Redundant features could result on misclassification. Then, selecting the features in order to keep the most discriminant will improve the performance of the classification stage and consequently, the performance of the overall segmentation process. The statistical features used in this paper are classified into first order features and second order features. First order features are derived from the gray level of a specific pixel while higher order features are computed taking into account the spatial relationship among different pixels. An overlapped sliding square window of size 7×7 voxels is used to compute the features. In order to make use of the image resolution that window is shifted voxel-by-voxel which allows computing a feature vector for each voxel. First order computed features are

$$\text{Intensity} \quad I = i(x,y) \tag{1}$$

$$\text{Average intensity} \quad \mu = \frac{1}{\omega_x \omega_y} \sum_{x=1}^{\omega_x} \sum_{y=1}^{\omega_y} (i(x,y) - \mu) \tag{2}$$

$$\text{Intensity variance} \quad \sigma^2 = \frac{1}{\omega_x \omega_y - 1} \sum_{x=1}^{\omega_x} \sum_{y=1}^{\omega_y} (i(x,y) - \mu) \tag{3}$$

On the other hand, second order features use the relationship between the central pixel and other pixel on a 2D or 3D neighbourhood. As second order features, we use the Haralick features [12] which capture the texture, and Hu moments [13] which are invariants under rotation and scaling. Both first order and second order features are computed over each window, and

Table 1
Extracted features summary.

Feature index	Feature	Feature index	Feature
1	Intensity	13	Cluster prominence
2	Intensity mean	14	Cluster shade
3	Intensity variance	15	Dissimilarity
4	Energy	16	Variance
5	Entropy	17	Hu moment (1)
6	Contrast	18	Hu moment (2)
7	Homogeneity	19	Hu moment (3)
8	Sum Average	20	Hu moment (4)
9	Autocorrelation	21	Hu moment (5)
10	Matlab correlation	22	Hu moment (6)
11	Correlation	23	Hu moment (6)
12	Angular second moment (ASM)	24	Hu moment (7)
13	Maximum probability		

assigned to the central pixel. Therefore, each pixel will have a different feature vector. Thus, the feature vector belongs to a D-dimensional feature space if D features have been extracted. The full feature set extracted from each window is summarized in Table 1.

2.3. Background in SOM and GHSOM

SOM [16] is one of the most used artificial neural network models for unsupervised learning. The main purpose of SOM is to group the similar data instances close into a two or three dimensional lattice (output map). On the other hand, different data instances will be apart in the output map. SOM consist of a number or neurons also called units which are arranged following a previously determined lattice. During the training phase, the distance between an input vector and the weights associated to the units on the output map are calculated. Usually, the Euclidean distance is used as shown in Eq. (4). Then, the unit closer to the input vector is referred as winning unit or Best Matching Unit (BMU) and its associated weight is updated. Moreover, the weights of the units in the neighbour of the winning unit are also updated as in Eq. (5). The neighbour function defines the shape of the neighborhood and usually, a Gaussian function which shrinks in each iteration is used as shown in Eq. (6). This deals with a competitive process in which the winning neuron each iteration is called Best Matching Unit (BMU):

$$U_\omega(t) = \underset{i}{\operatorname{argmin}} \|(t) - \omega_i(t)\| \tag{4}$$

$$\omega_i(t+1) = \omega_i(t) + \alpha_i h_{U_i}(t)(x(t) - \omega_i(t)) \tag{5}$$

$$h_{U_i}(t) = e^{-\|r_U - r_i\|/2\sigma^2(t)} \tag{6}$$

In Eq. (6), r_i represents the position on the output space (2D or 3D) and $\|r_U - r_i\|$ is the distance between the winning unit and the i-neuron on the output space. On the other hand, $\sigma(t)$ controls the reduction of the Gaussian neighborhood on each iteration. $\sigma(t)$ usually takes the form of exponential decay function as in Eq. (7), where σ_0 is the initial value of σ and τ_1 is a time constant:

$$\sigma(t) = \sigma_0 e^{-n/\tau_1} \tag{7}$$

In the same way, the learning factor in Eq. (5), also diminishes in time. However, α may decay in a linear or exponential fashion. Regarding the calculation of the quality of the output map, there exist two measures. The first is the quantization error, which is a measure of the resolution of the map. This can be calculated by computing the average distance between all the BMUs and the input data vectors as shown in Eq. (8). There is another measure of the goodness of the SOM. This measure is the topographic error, which measures how the SOM preserves the topology. This error can be computed with Eq. (11), where N is the total number of input vectors and $u(\vec{x_i})$ is 1 if first and second BMU for the input vector $\vec{x_i}$ are adjacent units (0 otherwise) [16,3]. Then, the lower q_e and t_e, the better the SOM is adapted to the input patterns. Unsupervised SOM are frequently used for classification and it does not use class information in the training process. Nevertheless, the main drawback of SOM is the size of the map, which has to be selected before classification and usually determines the quality of the SOM. On the other hand, the performance of SOM with high-dimensional input data depends on the specific features, and the calculation of the clusters borders may be not optimally defined. This way, the Growing Hierarchical Self-organizing Map [10,26] is a variant of SOM which dynamically grows overcomes the limitations of the SOM and allows discovering inherent hierarchies on the data.

GHSOM is a hierarchical and non-fixed structure developed to overcome the main limitations of classical SOM [26,16]. GHSOM [10] consists of multiple layers, each of them composed of several independent SOMs. Hence, during the training process, the number of the SOM maps on each layer and the size of each SOM is determined. This provides an adaptive growing process in both, the size of the GHSOM maps and the depth of the hierarchy. The growing process is controlled by two parameters which control the depth of the hierarchy and the breadth of each map. Therefore, these two parameters are the only ones which have to be determined in advance. In order to determine how much the GHSOM grows, the *quantization error* of each unit is calculated according to Eq. (8), where C_i is the set of input vectors mapped into the unit i, x_j is the j-th input vector belonging to C_i, and ω_i is the weight associated to the unit i:

$$qe_i = \sum_{x_j \in c:i} \|\omega_i - x_j\| \tag{8}$$

$$mqe_i = \frac{1}{n_{C_i}} \sum_{x_j \in C_i} \|\omega_i - x_j\| \tag{9}$$

Initially, all the input vectors belong to C_0. This means all the inputs are used to compute the initial quantization error, qe_0. Then, quantization errors qe_i for each neuron are calculated.

This way, whenever $qe_i < \tau_2 \times qe_0$, the neuron i is expanded in a new map on the next level of the hierarchy. Each new map is trained as an independent SOM, and the BMU calculation is performed as shown in Eq. (5) by using the Euclidean distance metric. Once the new map is trained, the quantization error of each neuron on this map is computed after training. mqe_m is computed as the mean of the mqe_i values for all the units in the map m as shown in Eq. (9), and is used to check the map growth. Thus, the map m will grow $mqe_m \geq \tau_1 \times q_i$, where q_i is the quantization error of the unit i on the upper layer and τ_1 is a parameter to control the depth of the GHSOM. All the growing process has been summarized in Fig. 1. If τ_1 and τ_2 are selected in such a way the GHSOM is sightly oversized, some of the units on the GHSOM maps may remain unlabeled after training. As GHSOM is a fine-grained unsupervised classifier, it can be used for image segmentation taking advantage of its clustering properties. In addition, GHSOM can be improved including probabilistic models as shown in [21] (Fig. 2).

2.3.1. BMU calculation on GHSOM

Once the GHSOM is trained, the BMU is computed for every data sample. In the case of SOM, the BMU is calculated in the same way it is found during the training phase (i.e., computing

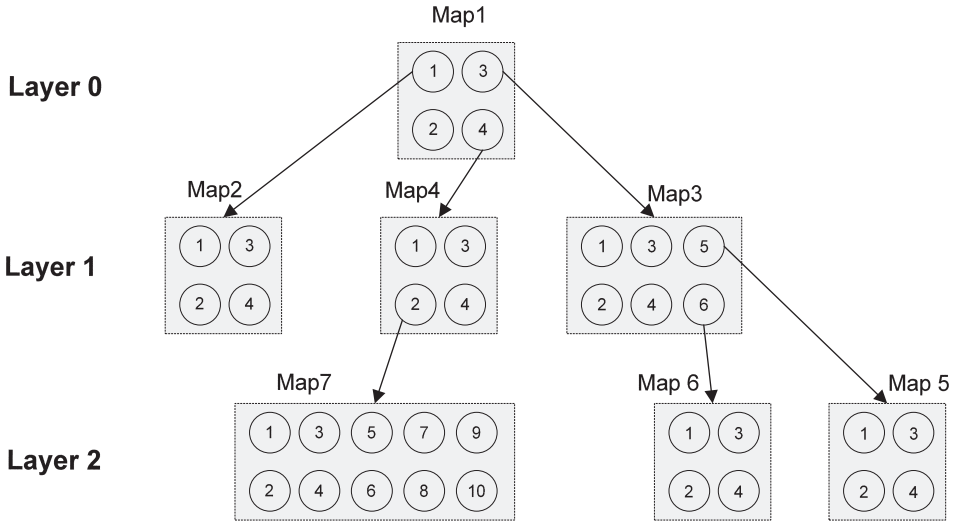


Fig. 1. GHSOM growing example.

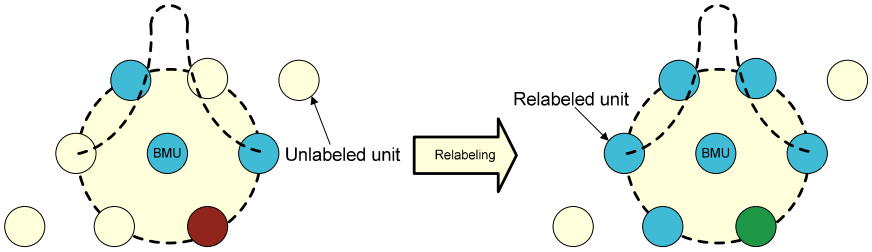


Fig. 2. GHSOM map relabeling method.

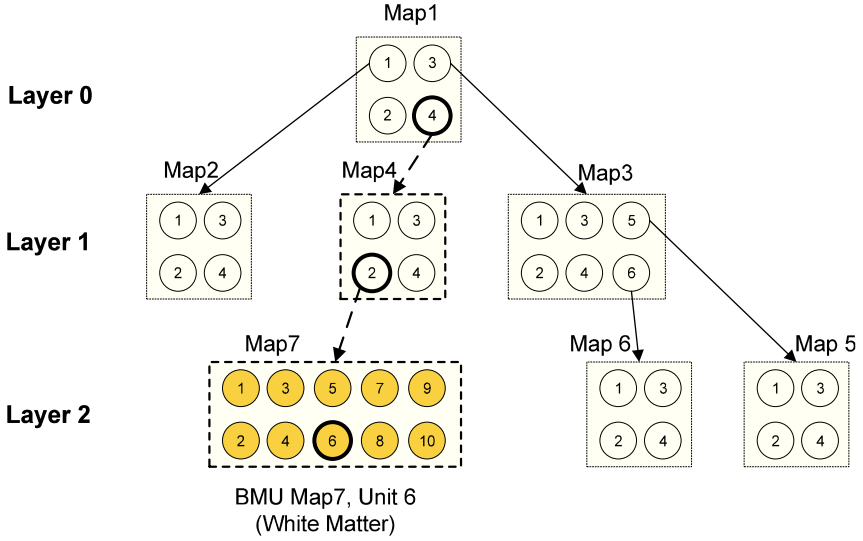


Fig. 3. GHSOM BMU calculation example.

the minimum Euclidean distance between the sample and the SOM units). However, GHSOM is composed of several SOM layers created during the training phase. This way, BMU calculation requires following all the SOM hierarchy to determine the winning unit and the map which it belongs to. Thus, an iterative algorithm has been developed as shown in Fig. 3. In this figure, an example of BMU calculation on a three-level GHSOM is shown. Let suppose we calculate the distances between an input pattern and the weight vectors of the level 0 map, and then compute the minimum of these distances. As a result, the winning neuron on map 1 is found. Since other map could be growth from this winning neuron, we have to check if the wining neuron is a parent unit. This can be accomplished with the parent vectors resulting from the GHSOM training process. If a new map arose from the wining neuron, the BMU on this map is calculated. This process is repeated until a BMU with no growing map is found. Thus, the BMU in the GHSOM is associated to a map in a level of the hierarchy. At his point, a probability-based labeling method is applied by using a 2D Gaussian kernel centered at each BMU. Then, prior probabilities are computed using a majority-voting scheme with the units inside the Gaussian

kernel, and the posterior probability is computed using the Bayes formula as shown in (12).

In Eq. (10), the Gaussian kernel used to estimate the label for unlabeled units is shown. In this equation, σ determines the width of the Gaussian kernel. In other words, it is the neighborhood taken into account for the relabeling process. On the other hand, (x,y) is the position of the BMU in the SOM grid:

$$L(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{10}$$

$$t_e = \frac{1}{N}\sum_{i=1}^{N} u(\vec{x_i}) \tag{11}$$

$$p(\omega_k|x) = \frac{p(x|\omega_k)P(\omega_k)}{p(x)} \tag{12}$$

In Eq. (12), $p(\omega_k|x)$ represents the probability that a sample vector x belongs to class ω_k, while $p(x|\omega_k)$ is the probability of the class ω_k and $p(x)$ is a normalization constant. This way, this posterior probability can be used to classify new samples. Nevertheless, in this work the posterior probabilities has been used to relabel the units which remain unlabeled during the SOM training process.

2.4. MR image segmentation with GHSOM

The image segmentation method we present consists on five steps. After the MRI is acquired ①, the features described in Section 2.2 are extracted on each slice using an overlapping and sliding window of 7×7 voxels ②. Once the feature space is composed of ③, the feature vectors are used to train a GHSOM with a portion of the feature vectors computed from the image ④ (in our experiments, usually 8% of the total feature vectors as using a higher percentage does not improves the final results significantly). After that, we search the BMU corresponding to each feature vector by using the BMU calculation and the labeling method ⑤ as described in Section 2.2. This results on pixel classification as GM, WM, CSF or Background. The image segmentation process with GHSOM is shown in Fig. 4.

The segmentation process described above, uses all the computed features. These features have to describe the image and if possible, not to content redundant information. Thus, the feature vector has to be enough different from one segment to another for the classifier. Nevertheless, not all the features provide enough different information from one voxel to another and the features have to be properly selected in order to keep only the most discriminant ones. Selecting a set of features is not straightforward since it varies from one image to another and the feature extraction process plays a decisive role in the segmentation performance. In this paper, feature selection is accomplished by multiobjective optimization as it provides a relatively fast way to find the optimal feature set. Once the GHSOM has been trained with one of the volumes on the IBSR database, we proceed to segment new images. These new images are also taken from the IBSR database in order to be able to compare with the ground truth segmentation. For segmenting a new image, the first step is to extract windows of size w as commented in Section 2.2. Then, the optimal set of features calculated in the feature selection stage are extracted from each pixel, depending on the image plane (coronal, sagital or axial). After presenting the computed feature vectors to the architecture in Fig. 1, each voxel is determined to belong to a specific tissue class (White Matter (WM), Gray Matter (GM), Cerebrospinal Fluid (CSF) or Background (BCK)).

2.5. Feature selection using multiobjective optimization

The feature extraction stage consists in computing some properties or features from the original data set (image) which allows distinguishing among input patterns. Thus, the features computed from the image are crucial for the classification performance. This way, Extracted features have to be descriptive enough to differentiate and recognize patterns present on the image. Redundant features could result on misclassification. Then, selecting the features in order to keep the most discriminant will improve the performance of the classification stage and consequently, the performance of the overall segmentation process. This way, the dimension of the feature space will be reduced from $\mathbb{R}^{24}$ (refer to Table 1) to $\mathbb{R}^{\mathbb{D}}$ ($D<24$). There are several algorithms to reduce the dimensionality of the vectors. One of the most common methods is *Principal Component Analysis* (PCA) [2]. PCA is a well-known linear technique for dimensionality reduction. This method, which is based on the covariance matrix, linearly transforms the high dimensional data into a low-dimensionality data whose variance is maximized. This is accomplished by computing the eigenvalues of the covariance matrix. Then, the eigenvectors corresponding to the largest eigenvalues are selected and they may represent original data. Thus, it is assumed that original data could be recovered from the eigenvectors. These eigenvectors are the so-called *Principal Components* (PCs). While the PCs with higher eigenvalues may be more representative of the original data, the eigenvectors with lower eigenvalues may be noise. Nevertheless, PCA starts with the assumption all the feature vectors can be expressed as a linear combination of the eigenvectors. If this is not possible, choosing a small number of PCs will result on information losing. Moreover, in this paper we use multiobjective optimization [17,8] to reduce the dimension of the feature space, since the experiments performed show better segmentation results than PCA. The target of the multiobjective optimization is to maximize the similarity between the automatically segmented image and the segmented reference from the IBSR [14]. A similarity index can be computed as the images on the IBSR database are labeled (i.e., manual segmentations are provided), in order to simultaneously measure the number of voxels being correctly classified. This way, we use the Tanimoto index as indicated in the following equation:

$$T_a(C_g,C_s) = \frac{p}{p+fn+fp} \tag{13}$$

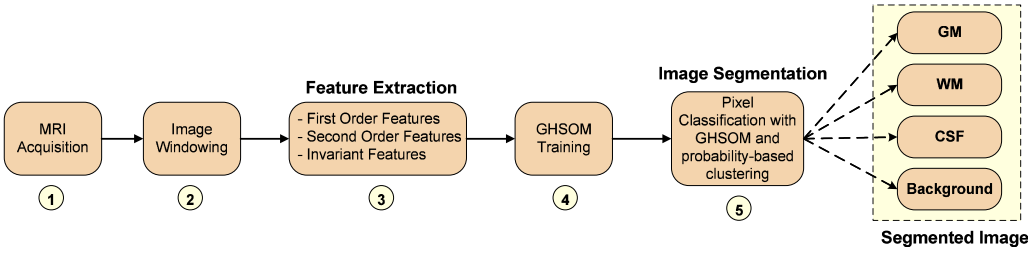


Fig. 4. GHSOM segmentation process.

where

- p is the number of voxels labeled as a and successfully classified (True Positive)
- fn is the number of voxels labeled as $a \in C_g$ but non-labeled as $a \in C_s$ (False Negative)
- fp the number of voxels non-labeled as $a \in C_g$ and labeled as $a \in C_s$ (False Positive)

Tanimoto index is similarity index related to the Jaccard index [29] and Dice index [9,30], which has been used as performance metrics in several works [14,18,6,27,11,19] among others. Moreover, Tanimoto index is used as overlapping rate in the IBSR web site for comparison among different segmentation algorithms. Thus, the Tanimoto index provides a measurement of the classification performance. This coefficient can be used to select the features which fits better for a specific segment since these selected features will maximize the similarity measurement. Maximizing the Tanimoto coefficient means maximizing three objective functions at the same time. Each of these four functions computes the Tanimoto coefficient for a tissue (WM, GM and CSF). This maximization is accomplished by the *Non-dominated Sorting Genetic Algorithm* (NSGA-2) [8], a fast elitist non-dominated sorting genetic algorithm for multiobjective optimization [8] using the objective functions shown in Table 2.

Table 2
Extracted features summary.

Objective function
$f_1 = T_a(C_g^{WM}, C_s^{WM})$
$f_2 = T_a(C_g^{GM}, C_s^{GM})$
$f_3 = T_a(C_g^{CSF}, C_s^{CSF})$

As a result of the optimization process, the Pareto front which summarizes the solutions found by NSGA-2 algorithm is found. Thus, the Pareto front contains enough information to select the features which maximizes the performance of the classifier (i.e., the Tanimoto coefficient) for each tissue class (WM, GM, CSF). Initially, the NSGA-2 algorithm generates a feature set. Then, the feature set is used to train a GHSOM by using 8% of the samples. Once the GHSOM is trained, the voxels of three different slices from a volume are classified and the Tanimoto coefficient corresponding to each tissue is computed. This classification is performed using a GHSOM with probability labeling [21] as shown in Section 2.3. The labels obtained from the manual segmentations provided by the IBSR database are used as reference for calculating the Tanimoto coefficient (ground truth). The optimization process is summarized in Fig. 5.

3. Experimental results and discussion

In this section we present the results of the experiments conducted to select the most discriminative features as well as the segmentation results using these features. Feature selection is accomplished using PCA and multiobjective optimization with the NSGA-2 algorithm. On the other hand, the effectiveness of the GHSOM classifier presented in Section 2.4 is shown. In addition, images from the IBSR and ADNI databases are segmented and shown for visual comparison. Hereafter, we refer the segmentation method which uses GHSOM and multiobjective optimization as GHSOM-MOO (GHSOM-MultiObjectiveOptimization). Feature selection by PCA is performed by computing the principal components of the feature space. Then, the feature vectors are projected to the 5, 10 and 15 principal components. This allows using only 5, 10 or 15 features instead of 24. However, it assumes a linear dependence among features. The optimization process

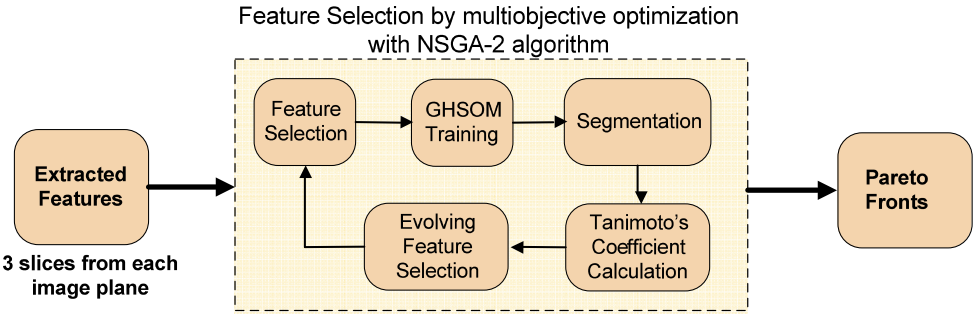


Fig. 5. Optimization process for feature selection.

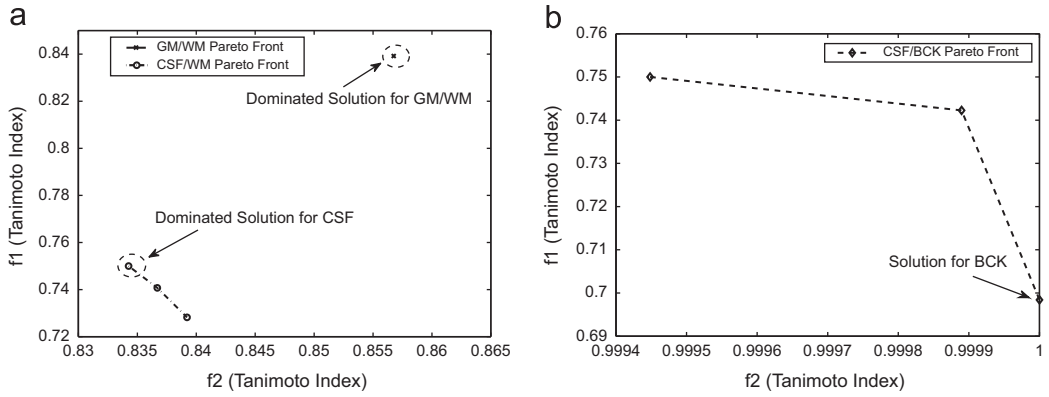


Fig. 6. (a) Pareto Fronts for GM/WM and (b) Background/CSF for axial plane.

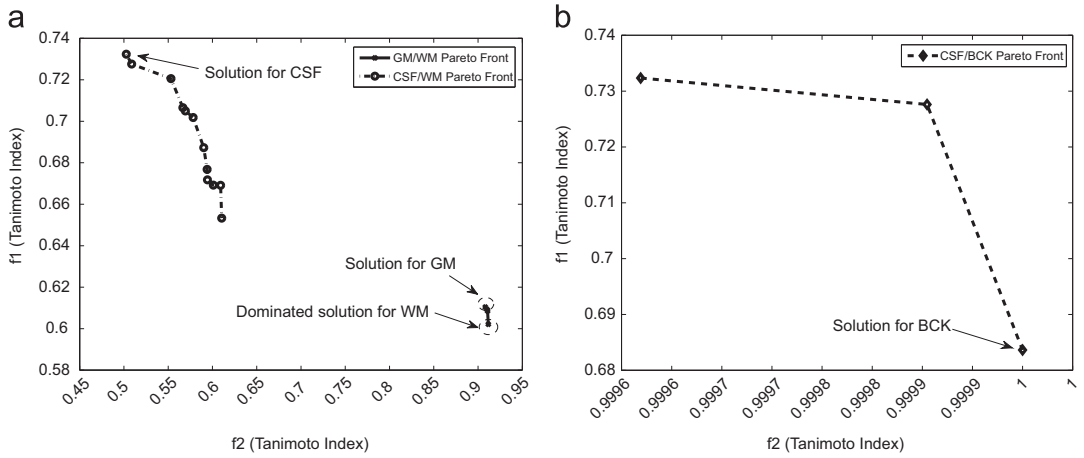


Fig. 7. (a) Pareto Fronts for GM/WM and (b) Background/CSF for coronal plane.

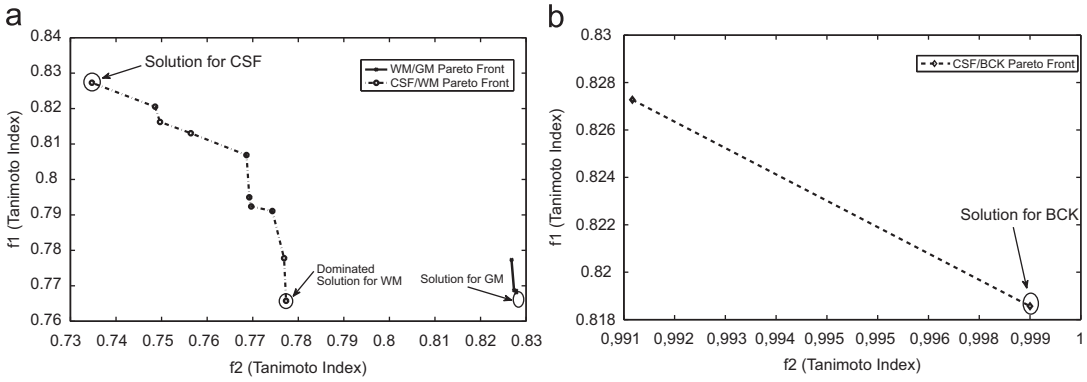


Fig. 8. (a) Pareto Fronts for GM/WM and (b) Background/CSF for sagittal plane.

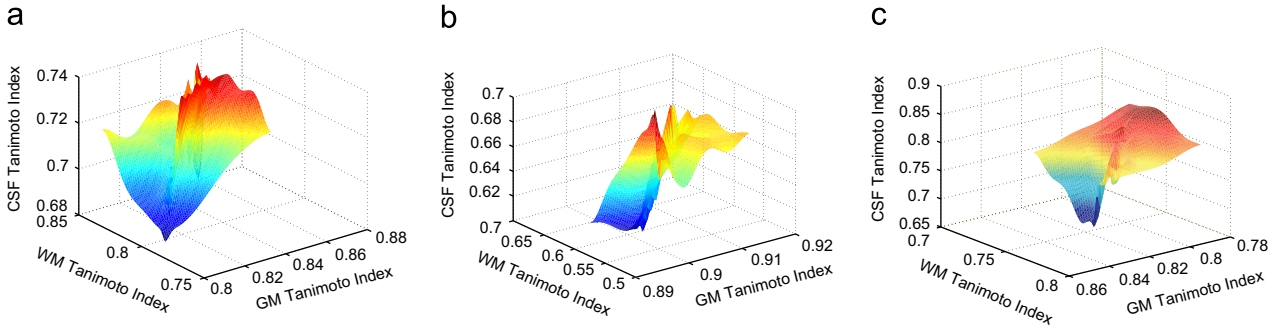


Fig. 9. Pareto Surface for (a) axial plane, (b) coronal plane and (c) sagittal plane.

which aims to maximize the overlap rate is performed for each plane separately. The goal is to find three sets of optimized features which allow maximizing the segmentation accuracy on the axial, coronal and sagittal planes respectively. As the optimization process searches for the best features which maximize the Tanimoto coefficient for WM, GM CSF and background, the objective space is 4-dimensional. Since it is not possible to plot the objective space, we show the Pareto fronts by objective function pairs. In other words, Figs. 6–8 show the Pareto fronts which represents the optimization solutions for GM/WM and CSF/WM segmentation for the axial, coronal and sagittal planes respectively.

Fig. 9 shows the resulting Pareto surfaces, which represent the solutions for GM, WM and CSF in the same figure. This summarizes the solutions found by the optimization algorithm. Each point on this surface provides a three set of features which maximizes the segmentation performance for GM, WM or CSF tissues.

The solutions found consist on the sets of features which maximize the classification performance of each tissue class. Table 3 shows these optimized feature sets for the axial, coronal and sagittal planes.

On the other hand, since we initially extracted 24 features, the optimized feature sets which contain a less number of features allow a faster segmentation. In the following section, we show the experimental results obtained by segmenting the volumes contained in the IBSR database. The performance of our method is shown by computing the average overlap metric (see Eq. (13)). The overlap metric is a method for comparing two segmentations that is more critical than comparisons using the volume [11,28] and it is a common measure to evaluate the segmentation algorithms. This is equivalent to the Tanimoto

coefficient described is Section 2.5. Moreover, since other methods have been evaluated with this metric, it allows making a performance comparison. As commented in Section 2.4, the probability labeling method [22,23] has been used at the classification step. Thus, Fig. 10 shows the ROC curves of the GHSOM labeling method. Since a probability is assigned to each data item which is labeled, the ROC curve shows the performance of this process. ROC curves are provided for the segmentation of two different layers of two different brain scans.

Table 3
Optimized features for each tissue class on each plane.

Plane	Tissue	Optimized feature indexes	Number of features
Axial	WM	1,2,3,6,11,15,16,17,18,20,21,22	12
	GM	1,2,3,6,11,15,16,17,18,20,21,22	12
	CSF	1,2,3,6,11,12,14,15,16,18,19,20,21,23,24	15
	BCK	1,2,3,6,11,14,18,20,21,23	10
Coronal	WM	1,3,9,10,11,15,17,19,22,23,24	11
	GM	1,3,9,10,11,15,18,21	8
	CSF	1,3,9,10,13,15,17,19,21,22,24	11
	BCK	1,3,9,10,13,15,17,19,20,21,23	11
Sagittal	WM	1,2,3,10,11,12,13,14,15,16,19,20,23	13
	GM	1,2,3,9,16,18,19,20,21,22,24	11
	CSF	1,2,3,12,14,15,16,17,18,19,20,23	12
	BCK	1,2,3,9,16,18,19,20,21,22,24	11

The segmentation experiments in this section have been performed by segmenting all the brains on the IBSR database and then computing the average overlap for WM, GM, CSF and Background. Since the average overlap metric values provided on the IBSR are referred to the coronal scans, we used or results from the coronal scans for comparison. However, we also provide results for axial and sagittal planes. Regarding the feature selection, the experiments on this section have been performed in two ways. First, PCA is used over the full set of features to reduce the feature space. This way, experiments with 5, 8, 10 and 15 principal components have been conducted. Second, the optimized feature sets shown in Section 2.5 are used to improve the segmentation performance. Fig. 11 shows the average overlap for the axial plane of all the brains on the IBSR. Results in Fig. 11a are obtained when using PCA for feature space reduction and Fig. 11b shows the results when multiobjective optimization is used to determine the most discriminant features.

In the same way, Fig. 12 shows the results for the coronal plane and Fig. 13 for the sagittal plane.

In Fig. 14 shows the average overlap for the axial plane of all the brains on the IBSR. Results in Fig. 14 are obtained using the optimized set of features for the three planes on the 100_23

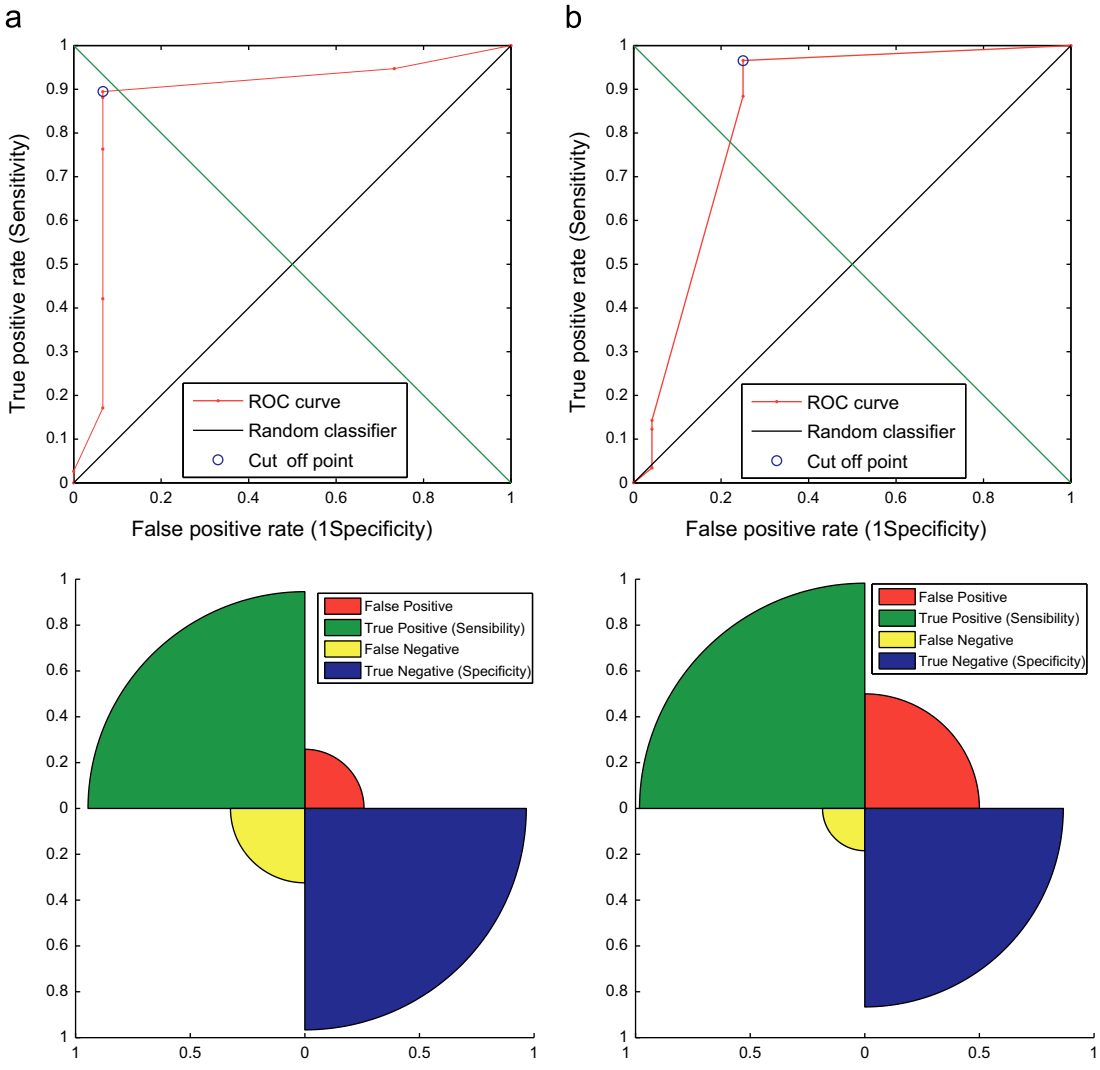


Fig. 10. ROC curves for the labeling process when segmenting the IBSR volume 100_23 with optimized feature set for CSF (above) and partest plot (below). (a) Axial plane and (b) coronal plane.

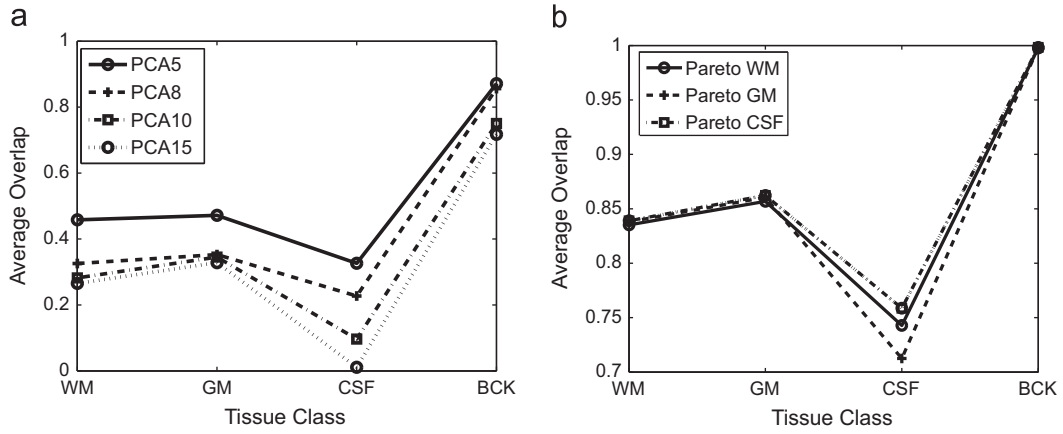


Fig. 11. Average overlap for the axial plane of all the brains on the IBSR. Feature reduction is performed with PCA (a) and multiobjective optimization (b), axial plane.

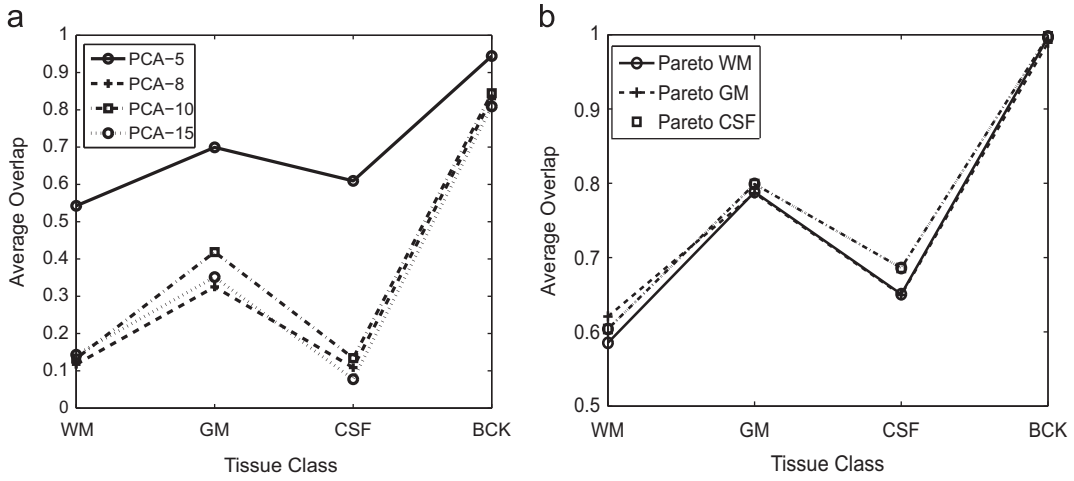


Fig. 12. Average overlap for the coronal plane of all the brains on the IBSR. Feature reduction is performed with PCA (a) and multiobjective optimization (b), coronal plane.

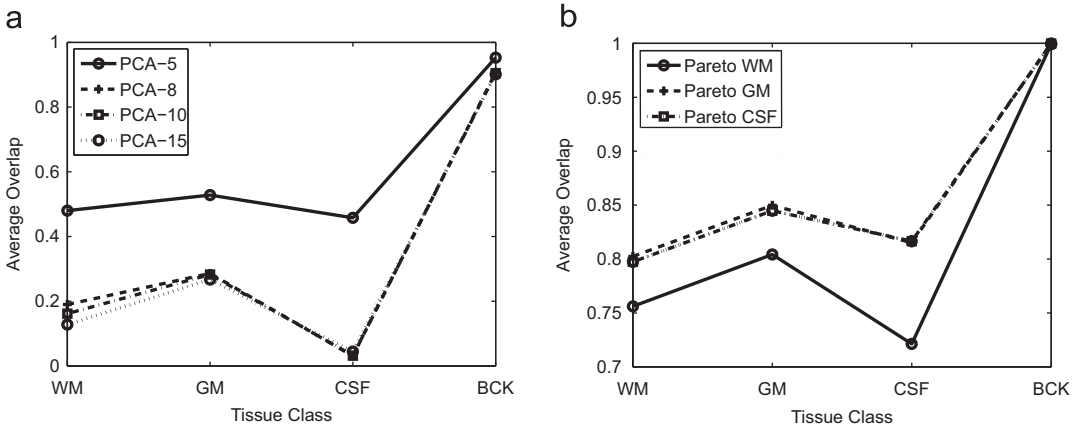


Fig. 13. Average overlap for the sagital plane of all the brains on the IBSR. Feature reduction is performed with PCA (a) and multiobjective optimization (b), sagittal plane.

volume of the IBSR database. In this figure, CGMM refers to Constrained GMM [11], MPM-MAP [18], Adaptive MAP [24], Biased MAP (BMAP) [25], Maximum a posteriori probability (MAP) [4], Tree Structure *k*-means (*tk*means) [7] and Maximum Likelihood (mlc) [5] (Figs. 15–22).

Figs 11 and 12 show segmentation examples for different brain scans from the IBSR. Ground truth is also shown for visual performance comparison.

In the experiments performed, dimensionality reduction by NSGA-2 optimization process performs better than PCA as the average overlap values provided by the features computed by NSGA-2 are higher than the ones provided by the principal components features. As a result, the feature set used cannot be effectively reduced by PCA since the features used do not depend linearly among them and cannot be expressed as a linear combination of each other.

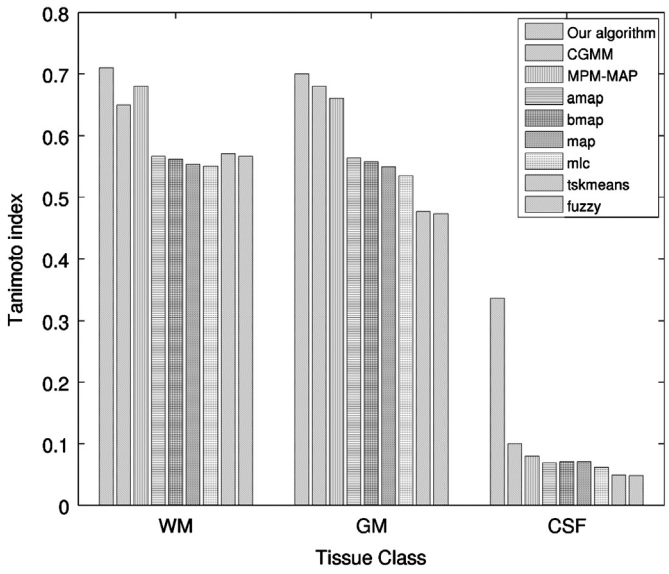


Fig. 14. Average overlap comparison for different segmentation methods.

3.1. ADNI segmentation example

In this section, segmentation of MR images from ADNI database is presented. In this case, segmentations using FAST software from the FSL library [20] for each image sequence for comparison.

4. Conclusions

In this paper, we presented a segmentation method using hybrid artificial intelligence techniques to improve the segmentation accuracy. The problem is addressed in three ways. First, the feature extraction process, which involves the use of overlapping windows to compute first order, second order and invariants. Then, we use an effective enough classifier to group the voxels belonging to a specific class according to the features computed at the feature extraction phase. This way, we use GHSOM, a more sophisticated version of SOM composed of several SOM layers of variable size. This avoids one of the classical drawbacks of the SOM, as the map size is adapted to the training data instead of keep it fixed. On the other hand, GHSOM allows discovering hierarchical structures on the input data. However, this model

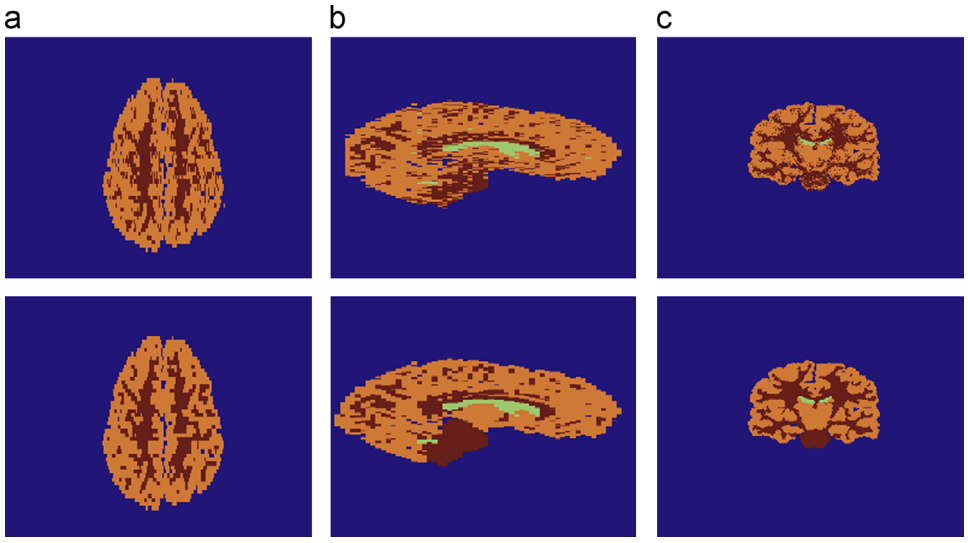


Fig. 15. Segmentation results and ground truth for the axial plane (a), coronal plane (b) and sagittal plane (c), volume 5_8, slice 128:30:166. Orange, Brown and green correspond to WM, GM and CSF respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

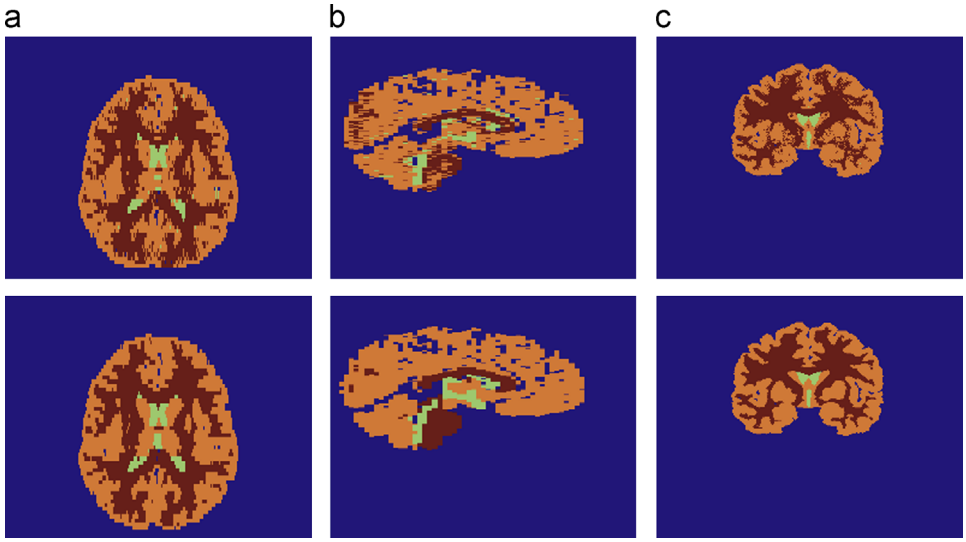


Fig. 16. Segmentation results and ground truth for the axial plane (a), coronal plane (b) and sagittal plane (c), volume 100_23, slice 128:30:166. Orange, Brown and green correspond to WM, GM and CSF respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

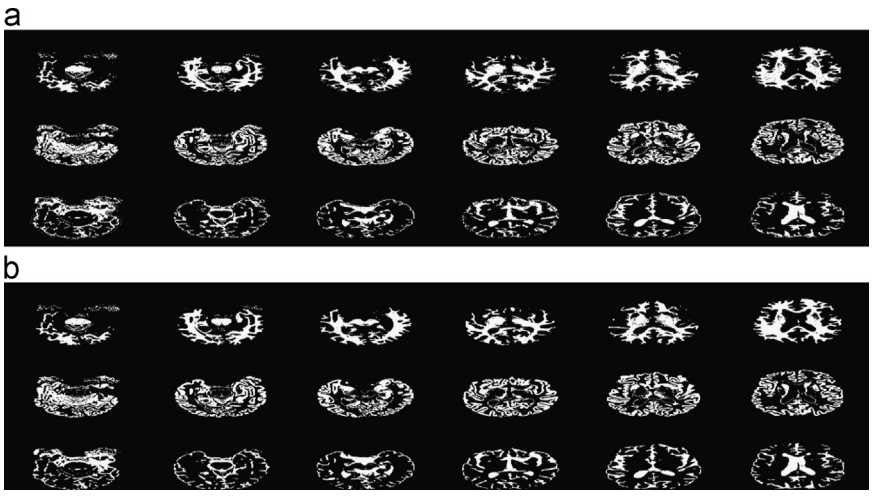


Fig. 17. Segmentation of the 002_S_0295 (Normal) ADNI volume using the GHSOM-MOO algorithm (a). FAST Segmentation is shown in (b). Slices 100, 110, 120, 130, 140 and 150 on the axial plane are shown on each column. First row corresponds to WM, second row to GM and third row to CSF.

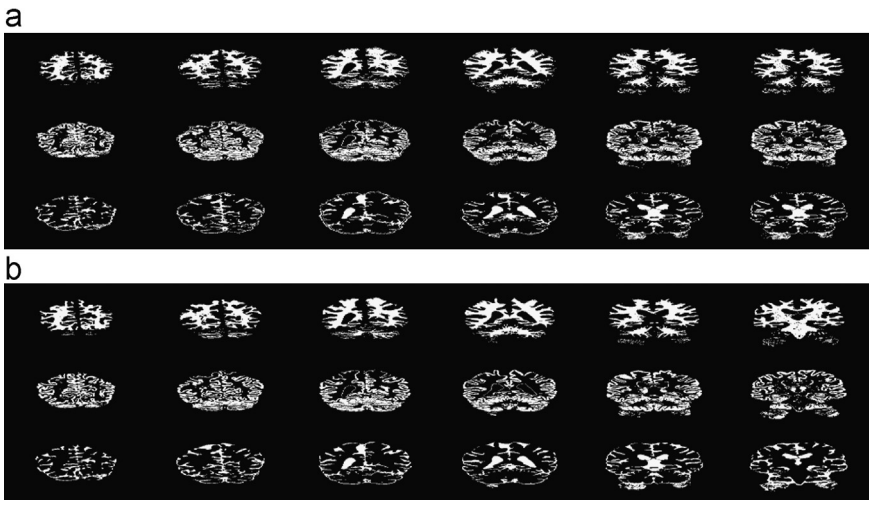


Fig. 18. Segmentation of the 002_S_0295 (Normal) ADNI volume using the GHSOM-MOO algorithm (a). FAST Segmentation is shown (b). Slices 60, 70, 80, 90, 100 and 110 on the coronal plane are shown on each column. First row corresponds to WM, second row to GM and third row to CSF.

have been improved by applying a probability-based labeling scheme for the unlabeled units. The third issue addressed on this paper concerns the technique for selecting the features which will compose the input to the classifier. Since not all the features are discriminant enough for all the image planes, a multiobjective optimization algorithm based on evolutive computation is used. Thus, the optimization algorithm figures out the features which maximize the classification performance for the three basic tissues found on a healthy brain (WM, GM and CSF). As a result, four sets of optimized features are extracted (features for GM, WM and CSF) for each image plane (axial, coronal and sagital) to maximize the performance of the classifier. The results obtained outperforms the results obtained with other segmentation methods. As commented in Section 2, both IBSR and ADNI databases contain real brain scans. Thus, all the images used in this work contain noise due to the acquisition process. In addition, the segmented images provided by the presented algorithm can be used to find patterns in AD patients such as hippocampus size or GM cortical thickness. Indeed, the algorithm presented in this paper is part of a larger study performed by the authors directed to use the tissue distribution in early AD diagnosis.

Acknowledgments

This work was partly supported by the MICINN of Spain under the TEC2008-02113 and TEC2012-34306 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Projects P07-TIC-02566, P09-TIC- 4530 and P11-TIC-7103. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company.

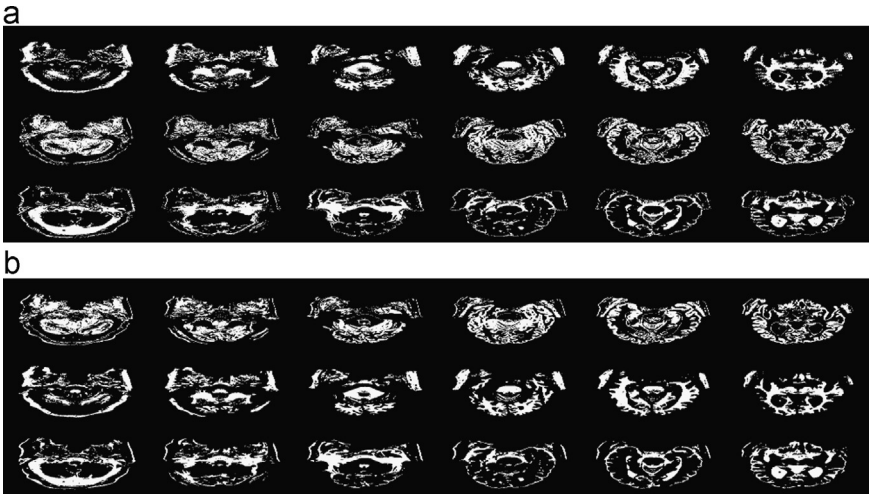


Fig. 19. Segmentation of the 002_S_0295 (MCI) ADNI volume using the GHSOM-MOO algorithm (a). FAST segmentation is shown in (b). Slices 100, 110, 120, 130, 140, and 150 on the axial plane are shown on each column. First row corresponds to WM, second row to GM and third row to CSF.

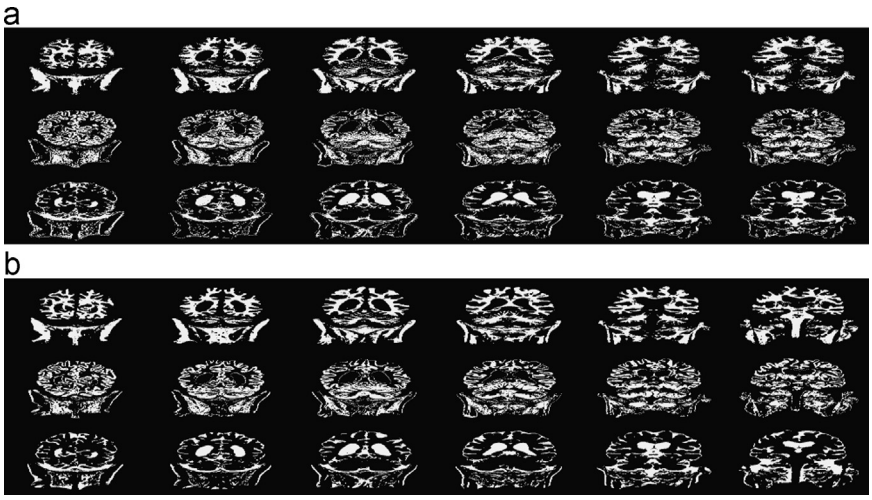


Fig. 20. Segmentation of the 002_S_0295 (MCI) ADNI volume using the GHSOM-MOO algorithm (a). FAST segmentation is shown in (b). Slices 60, 70, 80, 90, 100 and 110 on the coronal plane are shown on each column. First row corresponds to WM, second row to GM and third row to CSF.

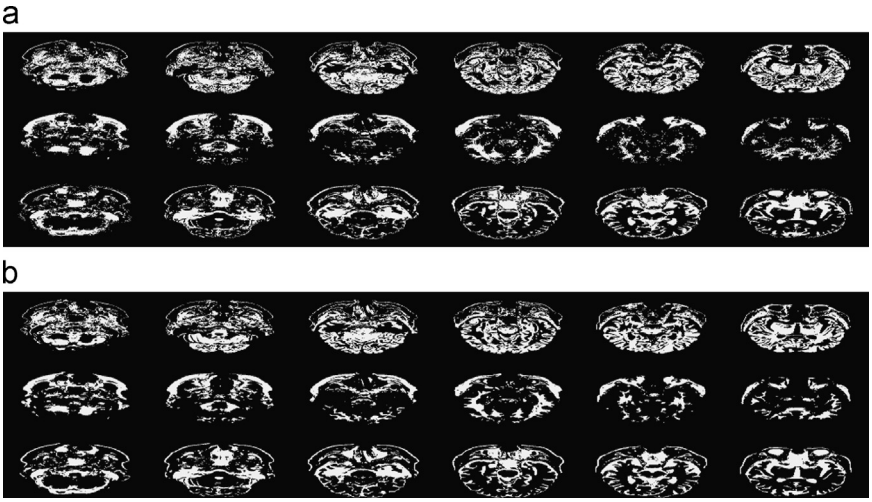


Fig. 21. Segmentation of the 002_S_0295 (AD) ADNI volume using the GHSOM-MOO algorithm (a). FAST Segmentation is shown in (b). Slices 100, 110, 120, 130, 140 and 150 on the axial plane are shown on each column. First row corresponds to WM, second row to GM and third row to CSF.

The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of

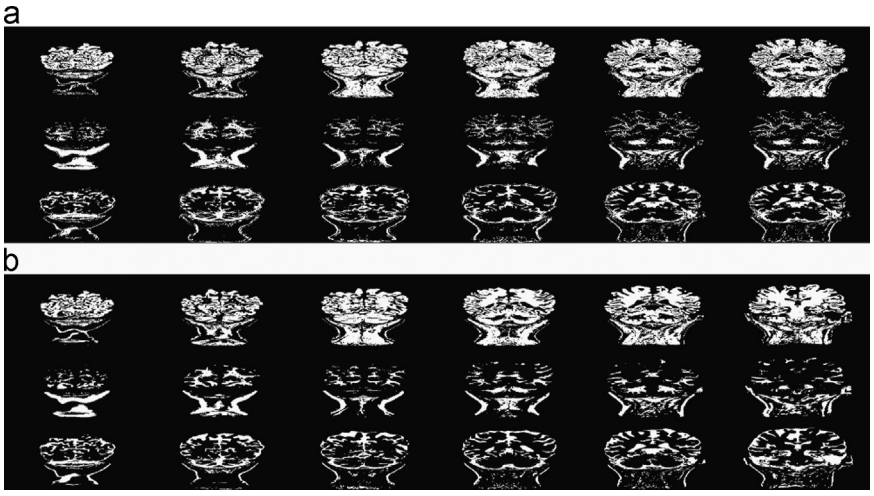


Fig. 22. Segmentation of the 002_S_0295 (AD) ADNI volume using the GHSOM-MOO algorithm (a). FAST Segmentation is shown in (b). Slices 60, 70, 80, 90, 100 and 110 on the coronal plane are shown on each column. First row corresponds to WM, second row to GM and third row to CSF.

California, San Diego. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of California, Los Angeles. This research was also supported by NIH Grants P30 AG010129, K01 AG030514, and the Dana Foundation.

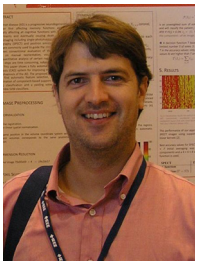
References

[1] Alzheimer Disease Neuroimaging Initiative (adni).
[2] H. Abdi, L.J. Williams, Principal component analysis, in: Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, 2010, pp. 433–459.
[3] E. Arsuaga, F. Díaz, Topology preservation in SOM, Int. J. Math. Comput. Sci. 1 (1) (2005) 19–22.
[4] M. Berthod, Z. Kato, S. Yu, J. Zerubia, Bayesian image classification using Markov random fields, Image Vis. Comput. 14 (1996) 195–285.
[5] J.C. Bezdek, L.O. Ifall, L.P. Clarke, Review of MR image segmentation techniques using pattern recognition, Med. Phys. 20 (4) (1993) 1033–1048.
[6] B. Caldairou, N. PAssat, P. Habas, C. Studholme, F Rosseau, A Non-local Fuzzy Segmentation Method: Application to Brain MRI. Technical Report, Universitè de Strasbourg, CNRS, 2010.
[7] L.P. Clarke, R.P. Veltuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, M.L. Silbiger, MRI segmentation: methods and applications, Magn. Reson. Imaging 13 (3) (1995) 343–368.
[8] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multiobjective optimisation: NSGA-II, in: Proceedings of the 6th International Conference on Parallel Problem Solving from Nature, vol. 1917, Springer-Verlag, 2000, pp. 849–858.
[9] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[10] M. Dittenbach, D. Merkl, A. Rauber, The growing hierarchical self-organizing map, in: Proceedings of the IEEE-INNS-ENNS International Neural Networks IJCNN 2000 Joint Conference, vol. 6, 2000, pp. 15–19.
[11] H. Greenspan, A. Ruf, J. Goldberger, Constrained gaussian mixture model framework for automatic segmentation of MR brain images, IEEE Trans. Med. Imaging 25 (9) (2006) 1233–1245.
[12] Robert M. Haralick, K. Shanmugam, Its'Hak, Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern 3 (6) (1973) 610–621.
[13] Ming-Kuei Hu, Visual pattern recognition by moment invariants, IEEE Trans. Inf. Theory 8 (2) (1962) 179–187.
[14] Center for Morphometric Analysis Internet Brain Segmentation Repository (IBSR). Massachusetts General Hospital, 2012.
[15] T. Kapur, W. Grimson, R. Wells, I. Kikinis, Segmentation of brain tissue from magnetic resonance images, Med. Image Anal. 1 (2) (1996) 109–127.
[16] T. Kohonen, Self-Organizing Maps, Springer, 2001.
[17] A. Konak, D.W. Coit, A.E. Smith, Multi-objective optimization using genetic algorithms: a tutorial, Reliab. Eng. Syst. Safety 91 (2006) 992–1007.
[18] J.L. Marroquin, B.C. Vemuri, S. Botello, E. Calderon, A. Fernandez-Bouzas, An accurate and efficient Bayesian method for automatic segmentation of brain MRI, IEEE Trans. Med. Imaging 21 (8) (2002) 934–945.
[19] John Melonakos, Yi Gao, Allen Tannenbaum, Tissue tracking: applications for brain MRI classification, Proceeding of the SPIE 6512 (2007) 651218.
[20] FMRIB Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Fmrib Software Library.
[21] A. Ortiz, J. Górriz, J. Ramírez, D. Salas-González, MR brain image segmentation by hierarchical growing SOM and probability clustering, Electron. Lett. 47 (10) (2011) 585–586.
[22] A. Ortiz, J. Gorriz, J. Ramirez, D. Salas-Gonzalez, MRI brain image segmentation with supervised SOM and probability-based clustering method, in: Lecture Notes in Computer Science, vol. 6686, 2011, pp. 49–58.
[23] A. Ortiz, J. Ortega, A.F. Diaz, A. Prieto, Network intrusion prevention by using hierarchical self-organizing maps and probability-based labeling, in: Proceedings of the 11th International Conference on Artificial Neural Networks Conference on Advances in Computational Intelligence, 2011.
[24] J.C. Rajapakse, J.N. Giedd, J.L. Rapoport, Statistical approach to segmentation of single-channel cerebral MR images, IEEE Trans. Med. Imaging 16 (2) (1997) 176–186.
[25] J.C. Rajapakse, F. Kruggel, Segmentation of MR images with intensity inhomogeneities, Image Vis. Comput. 16 (1998) 165–180.
[26] A. Rauber, D. Merkl, M. Dittenbach, The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data, IEEE Trans. Neural Networks 13 (6) (2002) 1331–1341.
[27] David W. Shattuck, G. Prasad, M. Mirza, K.L. Narr, A.W. Toga, Online resource for validation of brain segmentation methods, Neuroimage 45 (2009) 431–439.
[28] D.W. Shattuck, S.R. Sandor-Leathy, K.A. Schaper, D.A. Rottenberg, R.M. LEathy, Magnetic resonance image tissue classification using a partial volume model, Neuroimage 13 (2001) 856–876.
[29] Sergios Theodoridis, Konstantinos Koutroumbas, Pattern Recognition, Academic Press, 2009.
[30] A. Zijdenbos, B.M. Dawant, R.A. Margolin, A.C. Palmer, Morphometric analysis of white matter lesions in MR images: method and validation, IEEE Trans. Med. Imaging 13 (1994) 716–724.

Andres Ortiz received the M.Sc. degree in electronics in 2000, and the Ph.D. degree in 2008, both from the University of Granada. From 2000 to 2005 he was working as Systems Engineer with Telefonica, Madrid, Spain, where his work areas were high performance computing and network performance analysis. Since 2004 he has been with the Department of Communications Engineering at the University of Malaga as an Assistant Professor. His research interests include high performance networks, biomedical signal processing and artificial intelligence systems.

J.M. Górriz received a B.Sc. degree in physics (2000) and a B.Sc. in electronic engineering (2001) from the University of Granada, Spain. He then earned a Ph.D. from the University of Cádiz, Spain (2003) and a Ph.D. from the University of Granada (2006). He is currently an Associate Professor and qualified as a full Professor with the Department of Signal Theory, Networking and Communications at the University of Granada. He has coauthored more than 200 technical journals and conference papers in these areas and has served as editor for several journals and books. His present interests lie in the field of statistical signal processing and its application to speech and medical image processing.

J. Ramírez received the M.A.Sc. degree in Electronic Engineering in 1998, and the Ph.D degree in Electronic Engineering in 2001, all from the University of Granada. Since 2001, he is an Associate professor at the Department of Signal Theory Networking and Communications of the University of Granada (Spain). His research interest includes signal processing and biomedical applications including brain image processing, robust speech recognition, speech enhancement, voice activity detection, seismic signal processing and implementation of high performance digital signal processing systems. He has coauthored more than 150 technical journal and conference papers in these areas. He has served as reviewer for several international journals and conferences.



Diego Salas-González was born in Málaga, Spain, in 1980. He obtained his M.Sc. and B.Phil. degrees in physics from the University of Granada in 2003 and 2005 respectively. He was a granted national researcher from the Ministry of Education and Science of Spain from 2004 to 2008. His research interests are in statistical signal processing, Bayesian inference and applications in bioinformatics and image processing.