

Classification of Anatomical Structures in MR Brain Images Using Fuzzy Parameters

María-Elena Algorri* and Fernando Flores-Mangas

Abstract—We present an algorithm that automatically segments and classifies the brain structures in a set of magnetic resonance (MR) brain images using expert information contained in a small subset of the image set. The algorithm is intended to do the segmentation and classification tasks mimicking the way a human expert would reason. The algorithm uses a knowledge base taken from a small subset of semiautomatically classified images that is combined with a set of fuzzy indexes that capture the experience and expectation a human expert uses during recognition tasks. The fuzzy indexes are tissue specific and spatial specific, in order to consider the biological variations in the tissues and the acquisition inhomogeneities through the image set. The brain structures are segmented and classified one at a time. For each brain structure the algorithm needs one semiautomatically classified image and makes one pass through the image set. The algorithm uses low-level image processing techniques on a pixel basis for the segmentations, then validates or corrects the segmentations, and makes the final classification decision using higher level criteria measured by the set of fuzzy indexes. We use single-echo MR images because of their high volumetric resolution; but even though we are working with only one image per brain slice, we have multiple sources of information on each pixel: absolute and relative positions in the image, gray level value, statistics of the pixel and its three-dimensional neighborhood and relation to its counterpart pixels in adjacent images. We have validated our algorithm for ease of use and precision both with clinical experts and with measurable error indexes over a Brainweb simulated MR set.

Index Terms—Biomedical image processing, image classification, image segmentation, magnetic resonance imaging.

I. INTRODUCTION

IN order to be able to use the information contained in magnetic resonance (MR) images for quantitative applications, the anatomy in the images must be segmented and classified [1]–[3]. The quantitative information produced through segmentation and classification has many applications in the clinical environment in such areas as surgical planning, radiation therapy planning, anatomical modeling, and design of medical instrumentation [4]. But there are problems associated with MR images that have to be addressed to achieve a correct segmentation of the images. The main problem of MR images, besides the presence of noise, is that the signal intensity (gray level) that represents each tissue in the image is not unique

and the boundaries between tissues do not appear as a clear change of intensity [5]. These structural variations of the imaging objects are due to thermal/electronic noise, magnetic field inhomogeneities, biological tissue variations, and partial volume effects [6]. Although these factors impose great obstacles for the automatic or semiautomatic segmentation of MR images, they do not represent a significant problem for the clinical expert. Humans do not limit their decision taking elements to low-level vision mechanisms, but use high-level pattern recognition schemes, knowledge, experience, and intuition. For these reasons, there has been a trend in recent literature to replace or complement hard mathematical models that define mechanistic systems for soft algorithms that can capture some of the basics of human reasoning and decision taking capabilities [18], [19]. Soft algorithms have been greatly privileged when dealing with imprecise and complex biological systems [23]–[29]. Zadeh [7], [8] proposes a fuzzy framework for recognition technology that has been successfully applied to medical systems. His work on fuzzy sets in medicine centers on exploiting the tolerance for imprecision, uncertainty, and partial truth to achieve robustness of the algorithms. Pham *et al.* [9] segment two-dimensional (2-D) and three-dimensional (3-D) multispectral brain MR images using an adaptive fuzzy C-means algorithm. Their algorithm allows them to classify segmented voxels into multiple classes with varying degrees of membership, accounting for the noise and partial volume averaging present in single voxels. Xu *et al.* [10] mention that fuzzy segmentations retain more information from the original image than hard segmentations. An example of a high-level recognition and interpretation task that can be captured in fuzzy segmentations is the tracking of the relative spatial arrangement of objects embedded in a complex environment proposed by Bloch [11]. Chang *et al.* [12] combine a rule-based system with a clustering method to segment brain MR images. Their rule-based system permits expert knowledge to be explicitly represented in the form of fuzzy sets and if-then rules.

Effective as they have proven to be, soft segmentation methods can still benefit from complementing themselves with low-level vision techniques and basic image measurements. This low-level visual information is also the basis on which the human expert applies his rich decision techniques. An interesting approach is to combine fuzzy logic with statistical segmentation methods like the Bayesian approach taken by Rajapakse *et al.* [13].

We propose a new algorithm that is intended to work as an expert segmenter and classifier combining low-level image processing and high-level fuzzy recognition. The algorithm uses basic image measurements, expert human input, and knowledge

Manuscript received March 10, 2003; revised January 14, 2004. This work was supported in part by CONACYT under Grant J31985A. *Asterisk indicates corresponding author.*

*M.-E. Algorri is with the Department of Digital Systems, Instituto Tecnológico Autónomo de México, Tizapán San Ángel, México D.F. 01000, México (e-mail: algorri@itam.mx).

F. Flores-Mangas was with the Instituto Tecnológico Autónomo de México. He is now with the IIMAS, Universidad Nacional Autónoma de México, México D.F. 01000, México (e-mail: fflores@alumnos.itam.mx).

Digital Object Identifier 10.1109/TBME.2004.827532

about the nature of brain MR studies. We model the different sources of information as a set of fuzzy indexes. The algorithm works on single-echo brain MR images since these images can provide smaller voxels and more precise and accurate quantitative measurements than multi-echo studies [6]. We compensate for the lack of multi-echo information by considering multiple sources of information present in an MR brain image set: gray level value of the pixels, relative position of brain tissues in an image and in the image set, image statistics, 3-D neighborhood information, and object size. When making a classification decision, we not only consider the information present in the image under consideration, but also, as a human expert would do, the contextual information of the image in the MR set. This volumetric approach allows continuity of the 3-D boundaries in segmented objects [6]. To account for the biological variations and intensity variations in the tissues along the image set, we use tissue specific and spatial specific statistics.

Our algorithm works on a pixel basis and, therefore, can better represent complex objects than algorithms based on splines or neural networks [14], [15], which are limited by the number of nodes or control points in the models. Usually, contour models such as splines, snakes, or neural networks have been privileged over pixel-based models because they will not diverge from the solution (the object boundary) as easily as pixel-based models that are not constrained by any mathematical model. Also, in a bad scenario, contour models will at least give a solution as good as the initialization (which is usually close to the final object boundary). However, the price to pay is a simplification of the biological boundaries of an object. In our algorithm, the fuzzy indexes act as a robust error controller and prevent important divergences from the real object boundaries.

The algorithm we present is aimed at producing precise quantification results of brain structures, and scientific literature has not yet proven that this can be achieved using fully automatic classifiers. Although fully automatic segmentation algorithms using different techniques have been proposed [16], [17], these are usually aimed at segmenting only particular brain structures such as the white matter or the skull or produce a segmentation of the brain images into only a limited number of tissue classes. In general, some sort of algorithm supervision or image preprocessing is a part of all algorithms that seek precise classification of the various brain structures. The input of information into the algorithm by a human expert and/or some sort of supervision is usually not an additional burden, since the human expert is expected to at least speedily inspect the results of an acquired MR study. If during this quick inspection, the human expert can use a friendly interface to click over some images containing the structures he wants to segment, then the algorithm is off to a more robust recognition process. Although segmentation algorithms are aimed at freeing the human expert from a difficult and time-consuming task, nonbatch applications can afford to involve some degree of human input. We capitalize on human input by using the information contained in a few semiautomatically classified images to guide the automatic algorithm. For most images in the study though, there is no preprocessing or human interaction involved.

An additional benefit of allowing some user interaction is that the brain MR images can be classified into multiple objects, un-

like automatic classifiers which usually end up with only three tissue classes: white matter, gray matter, and cerebrospinal fluid [20].

An important application of segmented brain MR data is the reconstruction of brain structures [10]. As a last stage of the algorithm we input the segmentation/classification results to our 3-D reconstruction tool [21], [22] to produce 3-D models of the anatomical brain structures.

II. CLASSIFICATION ALGORITHM

We use sets of MR T1 brain images. Typically, such studies contain over 100 $256 \times 256 \times 8$ images. The main idea behind our algorithm is that the user can interactively classify a reduced set of images in the set and, from these semiautomatically classified images plus knowledge about the brain MR images contained in a set of fuzzy indexes, the algorithm will automatically classify the rest of the images. The algorithm can classify multiple brain structures. For each structure the algorithm takes as input one semiautomatically classified image of the structure and will make one pass through the image set to classify it. The output of the process is a set of images where the anatomical structures of interest have been segmented (uniformly labeled with particular values) and classified (identified as belonging to a particular structure). The classification algorithm is divided in two major tasks: 1) semi-automatic classification of a selection of key images and 2) automatic classification of the MR image set using fuzzy logic. We describe each task in the following subsections.

A. Semiautomatic Classification of a Selection of Key Images

1) *Selection of the Key Images:* We must first select the “key” images in the MR image set that will later guide the automatic classification process. We need to select one image for each of the brain structures that will be classified. The only requisite for the selection of the images is that every structure of interest be represented in the key images. If one key image contains more than one structure of interest, it can be used as the reference image for more than one structure. From these key images we will extract the information to create the fuzzy indexes that will be part of the automatic classification.

The key images are filtered using a 3×3 low-pass kernel that eliminates high-frequency noise such as isolated pixels and makes the regions in the images more uniform. We classify the key images interactively one at a time using the semiautomatic algorithm described next.

2) *Semiautomatic Classification Algorithm:* For the interactive classification of the key images, we developed the window interface shown in Fig. 1. In this interface the user can quickly browse through the image set and select the key images. For every key image the user will click on the brain structure that will be segmented from it. In Fig. 1, the user has clicked over the white matter. The algorithm will then perform a segmentation of the structure using a *region growing* algorithm that is guided by the structure’s statistics (mean and standard deviation). The region growing algorithm will grow over all the pixels whose statistics satisfy the conditions defined in Section II-A3. Using a mouse interface, the user can adjust the statistical conditions that control the evolution of the region growing algorithm

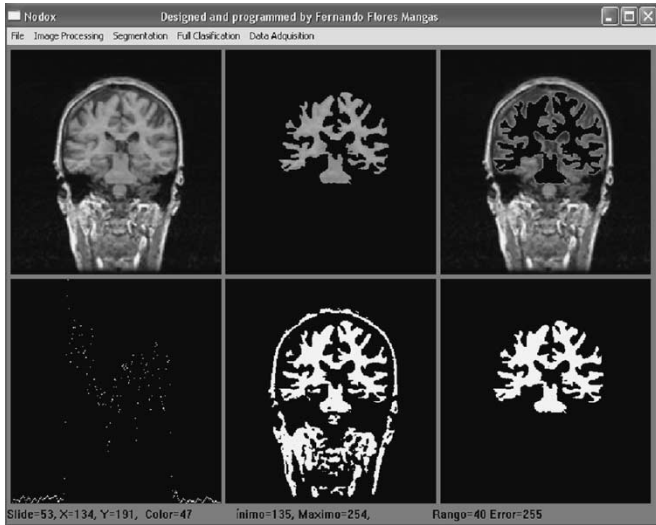


Fig. 1. User interface for the interactive classification of key images.

to avoid over- or undersegmentation of the structure. The user can visually validate his interactive segmentation using a transparency mask that allows him to superimpose the segmentation over the original image.

Once human validation of the segmentation is provided the segmented area is classified as the desired anatomical structure. We now describe in detail the segmentation procedure.

Let \mathbf{P} be the set of all the pixels in the image. Let p be one pixel in \mathbf{P} . For each p , there is an associated location in the image (i, j) and an associated gray level G_p (in the range $[0, 255]$) so that

$$p(i, j) = G_p. \quad (1)$$

Let \mathbf{A} be the set of pixels that form the area of interest or the anatomical structure that we want to segment in the image (where $\mathbf{A} \subset \mathbf{P}$).

We account for the volumetric intensity variations within each structure by allowing small variations of gray level (to within some standard deviation; see Section II-A3) in its constituent pixels, but we also consider that the gray level statistics of each structure are different (to within some level of signification; see Section II-A3) from the gray level statistics of neighboring structures. The segmentation algorithm is a simple region growing algorithm that uses a linked dynamic list to keep track of the pixels in a region and the pixels that are neighbors to the region. The algorithm starts the segmentation of a structure using as seed s the pixel that was interactively clicked on by the user when selecting the structure. Starting in pixel s the algorithm automatically finds a region \mathbf{A} such that

$$s \in \mathbf{A}. \quad (2)$$

The algorithm will grow over a region until all the pixels in \mathbf{A} are found according to a set of compliance criteria defined on \mathbf{A} (see Section II-A3). The recursive algorithm proceeds as follows: Starting at s it will evaluate the compliance criteria defined by (3) and (7) on the four-neighborhood of s as shown in Fig. 2(a). The pixels that are identified as belonging to \mathbf{A} are

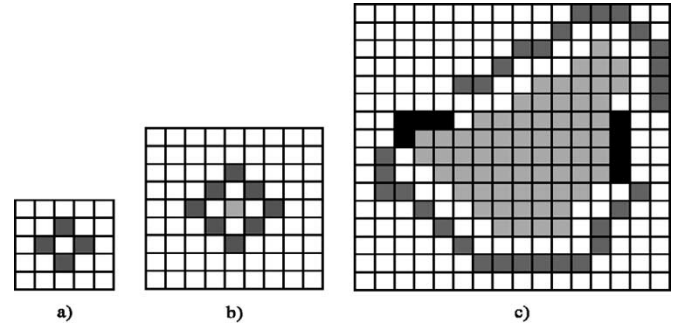


Fig. 2. (a) Seed pixel s in white and its four neighboring pixels in dark gray. (b) Second level of the recursive algorithm. Original seed pixel is labeled as belonging to \mathbf{A} and colored in light gray. Four neighboring pixels are being analyzed and their neighbors are shown in dark gray. (c) Advanced stage of the recursive algorithm with pixels belonging to \mathbf{A} shown in light gray, pixels under analysis shown in white, pixels that are still candidates for analysis shown in dark gray, and border pixels (containing no neighbors in \mathbf{A}) in black.

flagged and are no longer candidates for analysis. After a pixel is labeled as belonging to \mathbf{A} , the pixels in its four-neighborhood that are still candidates for analysis are examined [see Fig. 2(b)], and so on, until the compliance criteria fail to be met by every analyzed pixel and the algorithm stops. Fig. 2(c) shows an advanced stage of the recursive algorithm with pixels belonging to \mathbf{A} shown in light gray, pixels under analysis shown in white, pixels that are still candidates for analysis shown in dark gray, and border pixels in \mathbf{A} in black.

The compliance criteria that a pixel must meet to be part of \mathbf{A} are described next.

3) Compliance Criteria for the Segmentation Algorithm:

a) *Analysis of the average standard deviation in an image:* We assume that all pixels in \mathbf{A} contain gray level values similar to the gray level value of the seed s manually chosen by the user as belonging to the structure under classification at the beginning of the algorithm. We call the seed gray level value G_s and define the gray level value limits for pixels in \mathbf{A} as

$$\text{LowerLimit} = \begin{cases} 0, & G_s - v < 0 \\ G_s - v, & \text{else} \end{cases} \quad (3a)$$

and

$$\text{UpperLimit} = \begin{cases} 255, & G_s + v > 255 \\ G_s + v, & \text{else} \end{cases} \quad (3b)$$

where

$$v = kM \quad (4)$$

is k times the average standard deviation of the image histogram peaks (means) M and k is a constant experimentally initialized between 2 and 3, that reflects how well the histogram peaks are separated from each other. The initial value $k = 2$ provides only an initial stopping criterion for the segmentation algorithm. As described in Section II-A2, the user interactively adjusts this parameter to the specific structure statistics when visually validating the segmentation process. We define M as

$$M = \frac{1}{i} \sum_1^i \text{StdDev}(\text{HistogramPeak}(i)). \quad (5)$$

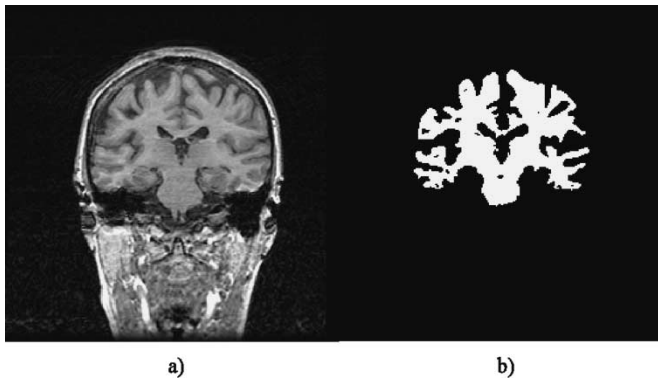


Fig. 3. Binarization of an MR image after labeling connected pixels as belonging or not to a range of gray level values.

The result of this gray level value analysis is a binarization of the image, where the connected pixels within the allowed gray level range are labeled as different from the rest of the image. Such a result for a seed pixel s inside the white matter is shown in Fig. 3.

The gray level statistics of the structure of interest are not the only stopping criterion for the region growing algorithm, we next describe an additional criterion based on the gray value statistics of the structure's neighboring pixels.

b) Gray level value differences among neighboring pixels: Let p_1, p_2, \dots, p_8 be the eight-neighborhood pixels of s . Using (1), we define an error level E

$$E = \sum_{i=1}^8 |Gp_i - G_s|. \quad (6)$$

The error E is just an accumulator of the differences in gray level between the seed pixel s and its eight surrounding neighbors. E can be interpreted as a measure of how equal or different the gray level of a pixel is from the gray level of its eight-neighborhood. We define that the accumulator E should not exceed the threshold value M (average standard deviation in the image) defined in (5). This means that the average tolerable difference between neighboring pixels to still be considered as belonging to the same anatomical structure is $M/8$.

A seed pixel s inside an anatomical structure A is, except for the case where it lies at the border of the structure, surrounded by other pixels that also belong to A . Inside A , the gray level of the pixels neighboring s is close to the gray level of s , G_s . Fig. 4 shows a sample case where s and its eight-neighborhood belong to A , the figure also shows the gray level G_{p_i} for each of the pixels.

If we calculate the coefficient E for s as shown in Fig. 4, we obtain $E_1 = 40$. Now, let us assume the situation depicted in Fig. 5 where the seed s is a boundary pixel. In this case, it is noticeable that pixels $\{p_6, p_7, p_8\} \notin A$, although we can appreciate some partial volume effects.

If we calculate E for s in Fig. 5, we get $E_2 = 200$ which is larger than $E_1 = 40$ and is typically also larger than the coefficient M defined in (5); therefore, s is flagged as a border pixel. After analyzing a pixel s , if

$$E_s < M \quad (7)$$

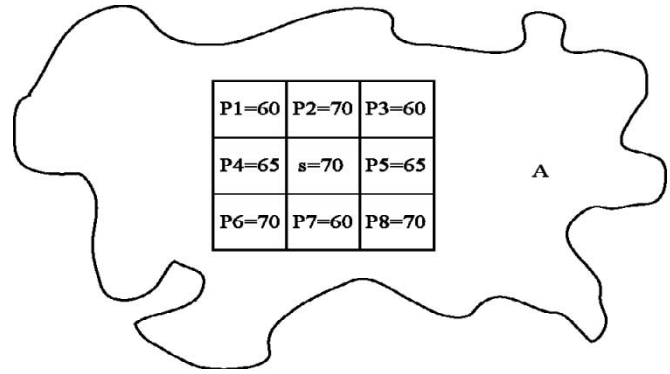


Fig. 4. Case where the seed pixel s , and its associated eight-neighborhood belong to A .

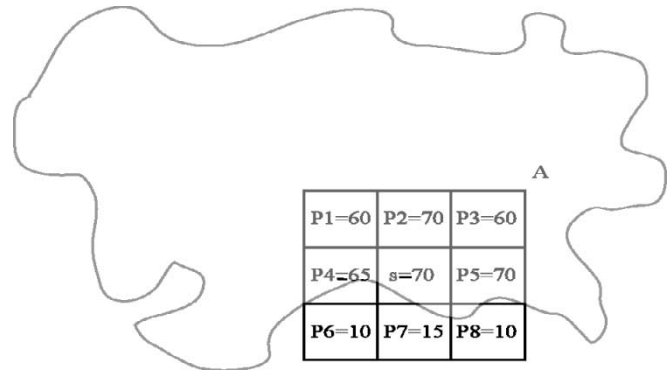


Fig. 5. Case where the seed pixel s lies on the boundary of structure A .

then $s \in A$ and s is flagged as successful (belonging to A). The procedure continues by examining all pixels in the eight-neighborhood of s that have not been flagged and flags them as part of A or as boundary pixels. The procedure terminates when all the pixels in A are flagged. When a boundary pixel is examined, its eight-neighborhood is no longer considered; instead, the algorithm proceeds to analyze the neighborhood of pixels that are marked as belonging to A .

The two compliance criteria taken together ensure that even if the differences between neighboring pixels remain low, pixels whose gray level values are significantly different from the original seed pixel s are no longer considered as belonging to A . This is the case of images that have different gray level values for different anatomical structures, but where the contrast between the structures is very little or the borders are blurred.

During the interactive classification of the key images, once the user has selected a starting seed pixel s for the region growing algorithm, he can interactively fine tune the parameters M and k for each anatomical structure of interest. The user can change the parameter values by moving the mouse in a designated window inside the user interface, thus creating the vectors $\mathbf{M} = [M_1, M_2, \dots, M_N]$, and $\mathbf{k} = [k_1, k_2, \dots, k_N]$ where N is the total number of brain structures to segment. The tuning of parameters relaxes our initial assumption that the image histogram is periodic, that is, that it is composed of equally spaced Gaussian distributions. By interactively changing \mathbf{M} and \mathbf{k} , the value of the standard deviation and the separation of the Gaussian distributions that belong to the anatomical structures of interest, the user is able to best reflect

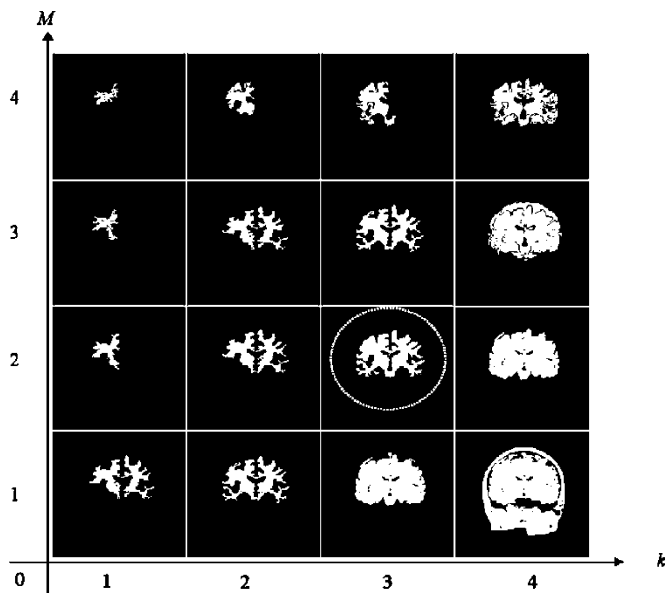


Fig. 6. Real-time segmentation results as the user interactively changes the values of the parameters M and k . Representation of the white matter is more accurate in the case of the circled image (3, 2).

the real statistics of each brain structure. As the user changes the parameter values, the results of the segmentation vary in real time, so that the user can best adjust the parameters to improve the quality of the segmentation. Fig. 6 shows the results of a segmentation of white matter as the user interactively changes the value of M (y axis) and the value of k (x axis). The best segmentation results are shown in (3, 2) (circled).

The interactive stage is over when the user has entered seed pixels for all anatomical structures of interest. After the user enters each seed pixel by giving a click of the mouse over an anatomical structure, the recursive algorithm executes itself and segments a region A . The classification of A into a particular anatomical structure is done automatically, since the user is forced by the interactive algorithm to specify an anatomical structure for each seed that is manually entered. The interactive stage also produces the vectors M and k containing the statistics of the structures of interest. The information in these parameter vectors will later be used by the automatic classification algorithm.

B. Automatic Classification of a Complete MR Study

After the user has interactively classified the key images, the algorithm can automatically classify the rest of the images. It will do so by propagating the information of the key images to neighboring images and by evaluating three fuzzy indexes that control the evolution of the classification and provide additional information about the likelihood of a pixel belonging to an anatomical structure.

The automatic classification algorithm proceeds by classifying one brain structure over the entire image set at a time. However, as will be seen later on when the stages of the algorithm are described in detail, at each pass the algorithm only evaluates the region of interest (ROI) where the structure appears in the images. The ROI for each structure is initialized from the key image segmentation and progressively adjusted

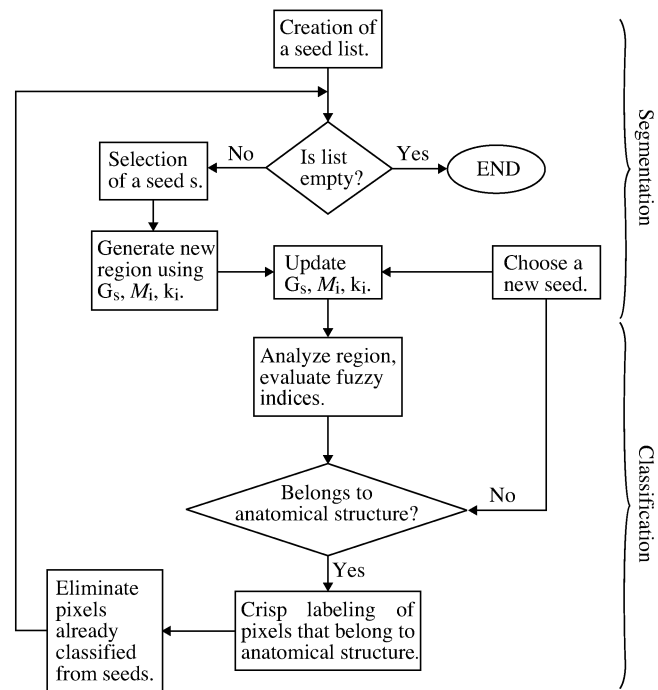


Fig. 7. Diagram of the tasks involved in the automatic classification algorithm.

as the classification proceeds through all the images. Therefore, there is no redundancy in the tasks of the algorithm, and its overall execution time is the same as it would be if all the brain structures of interest were classified simultaneously. But, by taking the approach of classifying one brain structure at a time, we are able to minimize the number of required key images to only one per structure of interest. The automatic classification algorithm will always start in the images adjacent to the key images.

The automatic classification algorithm consists of two stages: first, a *segmentation* of the images in regions (using the same region growing algorithm of the interactive segmentation), and second, a *classification* of the regions as belonging to particular anatomical structures.

To produce the segmented images, for each image in the set, the algorithm must:

- 1) create a list of candidate pixels for each region (seed list);
- 2) initialize the segmentation parameters (these are the same parameters that were used in the interactive classification);
- 3) generate the segmented regions in the images.

To produce the classified images, for each image in the set, the algorithm must:

- 1) evaluate a set of fuzzy indexes on the regions;
- 2) label the resulting pixels as belonging to a particular brain structure;
- 3) eliminate the classified pixels from the list of candidate pixels for analysis.

Fig. 7 shows a schematic diagram of the tasks involved in the automatic classifier. We describe these tasks in the next subsections.

1) *Automatic Segmentation*: The segmentation algorithm uses information from adjacent images that have already

been classified whether they are key images or automatically classified images, as well as information from the image being segmented to produce a set of parameters that will guide the region growing routine that produces the segmented regions. We will call the adjacent, already classified image that is used as segmentation reference at any particular time I_{ref} ; we will call the image under segmentation I_{seg} . As described in Section II-A2, the segmentation needs a seed pixel s that will be the origin of the region growing and two parameters. An estimate of the standard deviation of the structure's gray level values M_i , and an estimate of the width of the Gaussian curve that models the structure's distribution $v_i = k_i M_i$ to evaluate the stopping criteria (3) and (7).

The algorithm starts the segmentation of I_{seg} by selecting a seed pixel s for the brain structure that will be segmented. For this, the gradient of I_{seg} is calculated using a Sobel 3×3 operator. The gradient includes the borders between regions, high-frequency features, and noise. Since we are only interested in the pixels representing region borders, we select the pixels in the 10% upper gradient values and label these as border pixels. To close the regions, we then perform 3×3 dilation on border pixels. Each of the remaining image pixels is labeled as belonging to the same region as the pixel having the same (x, y) position in I_{ref} . Pixels in I_{seg} labeled as the structure of interest are candidates to become a seed pixel. The final seed pixel for the structure is chosen as the first pixel that is tested and complies with (3) and (7) as defined for I_{ref} , that is: 1) a pixel whose gray level value lies within the tolerable limits for the structure that it is being assigned to and 2) a pixel whose gray level value is similar to that of its eight-neighbors.

The segmentation algorithm will then produce one region in I_{seg} using the same parameter values for M_i and k_i as in I_{ref} . After the segmentation of the structure in I_{seg} is complete, we update the statistic measures M_i and k_i to reflect the intensity inhomogeneities and the biological variations along the images.

2) *Automatic Classification:* The job of the classifier is to label each of the segmented regions in I_{seg} as belonging to a particular anatomical structure. To do this, we define three fuzzy indexes, each of which, alone, is incapable of indicating whether a pixel definitely belongs or not to a particular anatomical structure, but can give us a degree of likeliness that the pixel belongs to the structure. We then define a global decision index as the weighted sum of the three fuzzy indexes. This global decision index provides the crisp decision of whether a pixel belongs or not to the structure. The indexes also act as error controllers that prevent divergences coming from the segmentation stage to be validated as anatomical structures. We define the indexes next.

a) *Fuzzy size index:* The first index measures the size of every segmented structure relative to its counterpart classified structure in I_{ref} . We expect the changes in size for every structure to vary lightly from one image to the next.

Let p_N be a pixel that does not belong to an anatomical structure in I_{ref} but does belong to the anatomical structure in I_{seg} . The number of p_N pixels for one anatomical structure in I_{seg} is P_N . Let p_M be a pixel within an anatomical structure both in I_{ref} and in I_{seg} . P_M is the number of pixels p_M . Last, let p_A be a pixel that belonged to an anatomical structure in I_{ref} but no longer belongs to the structure in I_{seg} . The number of pixels p_A

is P_A . We measure the relative difference in size of an anatomical structure between two consecutive images as

$$I_s = \frac{P_M}{P_N + P_M}. \quad (8)$$

We evaluate I_s over the segmented region in I_{seg} that is being classified. We accept the classification of a segmented region in I_{seg} as the brain structure we are interested in if $I_s > 0.7$. Notice that we do not use P_A in the definition of I_s . I_s can measure different evolutions of a brain structure through the image set. In the case where a structure is growing from I_{ref} to I_{seg} , I_s will be lower than unity. In this case, I_s will reject big size changes that could occur if the segmentation algorithm had diverged from the boundaries of the structure of interest. In the situation where a brain structure divides itself into multiple regions from I_{ref} to I_{seg} , I_s will not penalize these subdivisions that occur frequently in brain images, instead, I_s will be measured for each of the subdivided regions (from the description of the automatic segmentation we recall that the subdivided regions were all labeled as the same structure and will all be classified) and will be accepted. If we had considered P_A as part of I_s , we would consistently misclassify images where a structure subdivides into smaller regions. The last situation occurs when the structure reduces its size from I_{ref} to I_{seg} . In this case, I_s will be accepted. After I_s is accepted for a region (value > 0.7) the remaining two fuzzy indexes are evaluated over the region. If $I_s < 0.7$, we do not evaluate the remaining two fuzzy indexes and simply reject the segmented region as candidate for classification. We then return to the segmentation stage and produce a new segmented region with a different seed pixel taken from the candidate pixel list.

b) *Fuzzy indexes for gray level statistics:* We build two fuzzy indexes based on the gray level statistics of the regions: a mean index and a standard deviation index.

Mean Index: Every time we segment a region, we calculate a mean value for the gray levels of the region both in I_{ref} and I_{seg} . We are interested in the absolute difference between these two means. We build an index I_M that measures this difference

$$I_M = - \left| \frac{m_{\text{seg}} - m_{\text{ref}}}{\max(m_{\text{ref}}, m_{\text{seg}})} \right| + 1 \quad (9)$$

where m_{seg} is the mean of the region in I_{seg} and m_{ref} is the mean of the region in I_{ref} . I_M varies between $[0,1]$, with 1 meaning no difference between means and values lower than 1 meaning different means.

Standard Deviation Index: The standard deviation indicates how much the gray level values of all pixels in the region are different from the mean value. We are interested in the absolute difference between the standard deviations of the region in I_{ref} and I_{seg} . We build an index I_σ that measures this difference

$$I_\sigma = - \left| \frac{\sigma_{\text{seg}} - \sigma_{\text{ref}}}{\max(\sigma_{\text{ref}}, \sigma_{\text{seg}})} \right| + 1 \quad (10)$$

where σ_{seg} is the standard deviation of the region in I_{seg} and σ_{ref} is the standard deviation of the region in I_{ref} . I_σ varies between $[0, 1]$, with 1 meaning no difference between σ 's and values lower than 1 meaning different σ 's.

c) *Global decision index*: We combine the three fuzzy indexes just described into a global decision index. The global decision index is a joint measure of the characteristics measured by the individual indexes. It will allow the classification algorithm to take a final decision on whether a segmented region belongs or not to an anatomical structure. We define the global index as

$$I = \frac{(I_{sp} + 1.25I_M + I_\sigma)}{3.25} \quad (11)$$

where we give slightly more weight to I_M than to I_s and I_σ . This is because I_M measures statistical information present in the images, and the global index I is experimentally proven to produce a better decision when I_M is given more weight than I_s , an index that is modeled after human expectation and knowledge about the nature of the MR brain images. Our criteria for acceptance of the classification results are values of $I > 0.75$ since these values indicate a higher likelihood of the region belonging to the anatomical structure than not. If $I < 0.75$, we simply reject the classification result and choose a different seed pixel s for the region from the list of candidate pixels defined in Section II-B1. If I is accepted, we classify the region in I_{seg} as the same anatomical structure of its counterpart region in I_{ref} . Typically, if a region is well segmented, I takes values > 0.95 . The highest values of I are found when classifying images adjacent to the key images, and the values start to decrease as we classify images far away from the key images. The classification results that we present and validate in the next sections have all been obtained using only one key image for every classified brain structure. Our validation indexes compare favorably to others in literature. If the user requires a higher classification precision, it is always possible to enhance the classification results by increasing the number of key images for each classified structure.

III. RESULTS

We tested and validated our algorithm on different sources of MR images. We classified five T1 MR brain image sets and then had them qualitatively validated by a clinical expert. The sets were 256 by 256 by 8 by 110–124 images. All five studies were coronal. The images were taken from living healthy adult subjects with no brain lesions. The qualitative validation by the clinical expert was satisfactory in all five cases. In order to provide a verifiable measure of the performance of the algorithm, we also tested it on a simulated Brainweb [30]–[34] MR brain image set. The Brainweb simulated image set contained 181 images, 217 by 181 pixels, axial type, T1 mode. The parameters of the Brainweb image set used were $\sigma^2 = 0.03$, white noise, and 20% of nonuniformity, no lesions.

The algorithm was implemented in Visual C++ and executed on a Pentium IV system running at 2.4 GHz with 0.5-GB RAM and graphics accelerator. Typical classification time for one brain structure in an image set was 1 min.

Fig. 8 shows three comparative examples of the classification results of the white matter. Fig. 8(a) shows images obtained from a clinical MR study, Fig. 8(b) shows the classification results of the algorithm, Fig. 8(c) shows three images of the Brainweb simulated set, Fig. 8(d) shows the “true” classification of the

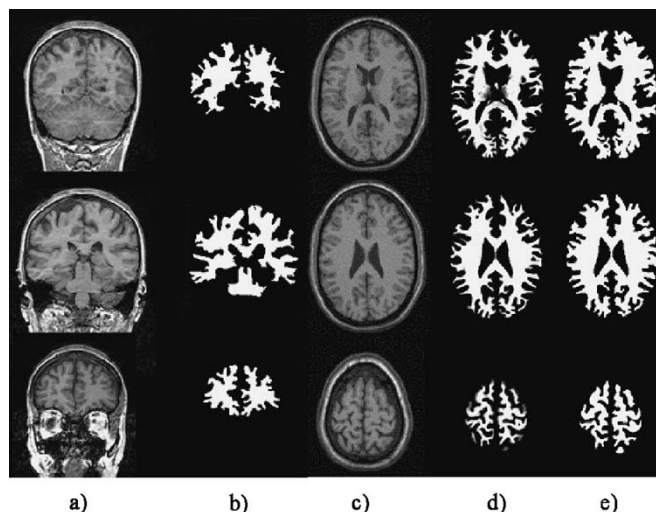


Fig. 8. (a) Clinical MR T1 images, (b) classification results of the white matter provided by our algorithm, (c) Brainweb T1 MR simulated images, (d) real classification provided by the Brainweb simulator, and (e) our classification results of the white matter in the Brainweb images.

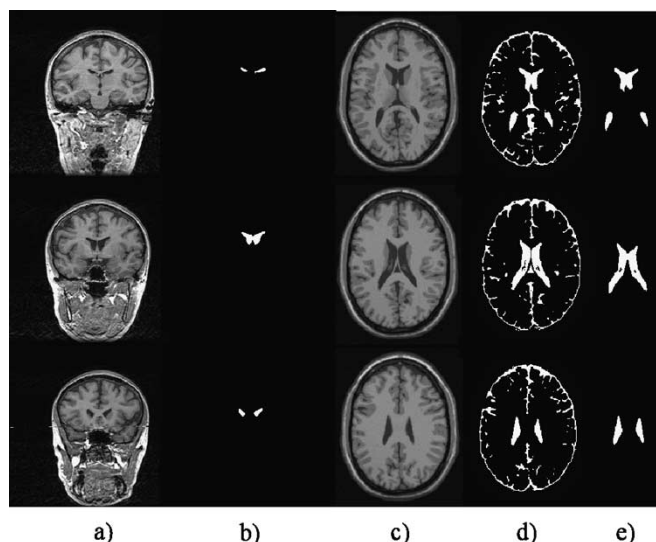


Fig. 9. (a) Clinical MR T1 images, (b) classification results of the ventricles provided by our algorithm, (c) Brainweb T1 MR simulated images, (d) real classification provided by the Brainweb simulator, and (e) our classification results of the ventricles in the Brainweb images.

white matter also provided with the Brainweb simulated images, and Fig. 8(e) shows our classification results.

Fig. 9 shows the same example of comparative results as Fig. 8, but in this case for the ventricles.

Fig. 10 shows the 3-D reconstruction of the white matter and the ventricles obtained using a general 3-D reconstruction tool, rendering and texturing.

IV. QUANTITATIVE EVALUATION

Expert validation by a clinical specialist has often been used as a basic quality control procedure for segmentation results. However, to obtain a quantitative, reproducible performance index of the classification algorithm, adequate metrics of the results must be reported on standard MR image sets. In [41], 20

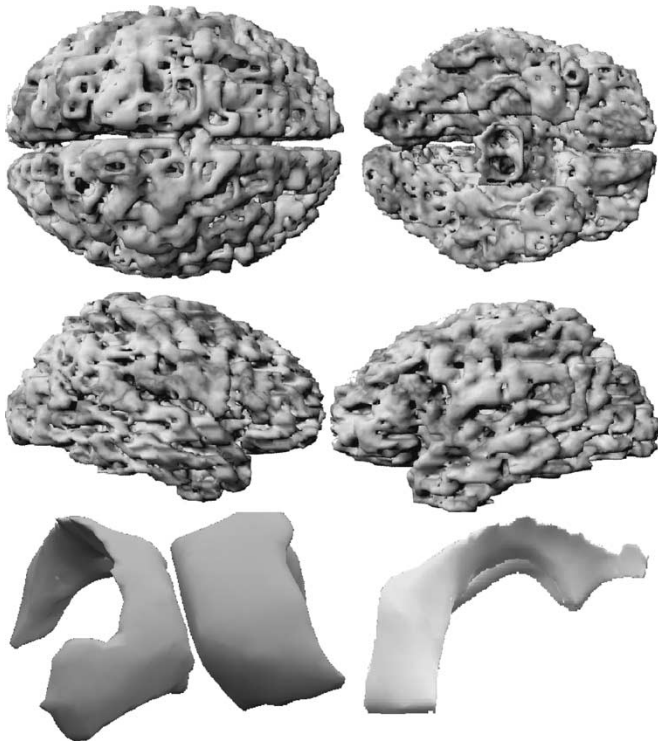


Fig. 10. Reconstruction results for a white matter and a pair of brain ventricles.

MR image sets along with their manual classifications are available to be used as standard image sets. In this paper, we carried out a quantitative evaluation of the white matter classification results obtained on the Brainweb image set. We computed three validation indexes that prevail in segmentation literature for comparison of algorithmic performance: the similarity index or overlap metric [38]–[40], [42], [43], the Tanimoto index [35], [36], [41], and the misclassification rate (MC) [9], [37]. The three indexes were obtained by comparing the image set containing the “true” classification of the white matter provided by the Brainweb simulator and the binary image set containing the results of the white matter classification provided by our algorithm. The reported indexes are the average value for the 181 image set.

1) *Similarity Index or Overlap Metric*: This index is defined as twice the number of pixels assigned to a class k by both the ground truth and the classification algorithm divided by the sum of pixels assigned to class k by the ground truth plus the pixels assigned to class k by the classification algorithm. The similarity index is sensitive to segmentation errors both in size and location, although differences in size affect the index more than differences in location. In [40], the author explains that a similarity index above 0.75 means excellent correspondence between two segmentation results. When comparing the results of our classification algorithm to the ground truth provided by the classified Brainweb image set, we got an average similarity index for the whole image set of 0.83. For individual images we got similarity indexes as high as 0.95.

2) *Tanimoto Index*: The Tanimoto index is calculated on pairs of images, where one image comes from the binary classified image set produced by the algorithm and the other comes from the true classification binary image set used as

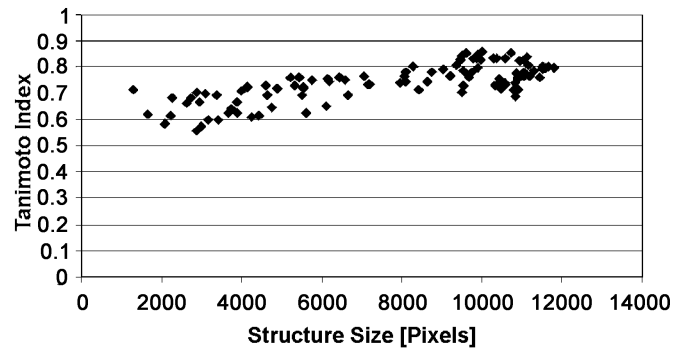


Fig. 11. Graph of the results obtained for the Tanimoto index with respect to the size in pixels of the white matter over an image (only images with more than 3% white matter are shown).

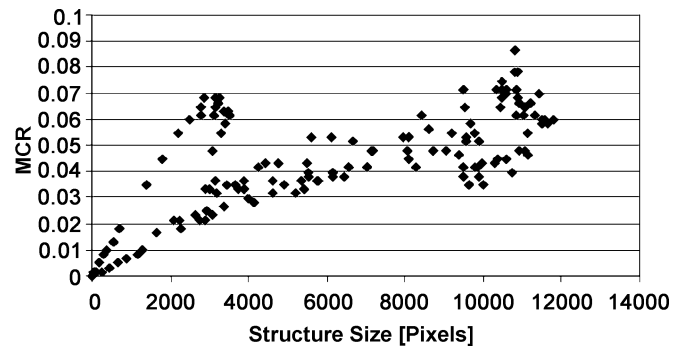


Fig. 12. Graph of the MC rate with respect to the size in pixels of the white matter in the images.

reference. The Tanimoto index is the result of dividing the cardinality of the intersection of the binary images between the cardinality of the union of the binary images. The Tanimoto index is zero when the images have no pixels in common and one when both images have all pixels in common. We calculated an accumulated Tanimoto index (average over the 181 Brainweb images) for the white matter classification of 0.722. For comparison, [41] cites Tanimoto indexes for manual segmentations of MR image sets of only 0.83. These index values further prove that even to the human expert, MR image segmentation is a difficult task. Fig. 11 shows a graph of the variation of the Tanimoto index with respect to the size in pixels of the white matter in an image.

3) *Misclassification (MC) Rate*: The MC rate is calculated as the number of false positive pixels plus the number of false negative pixels divided between the total number of pixels in the image. After computing the MC rate of the complete automatically classified white matter in the Brainweb image set, an average MC rate of 0.02 was obtained. Fig. 12 shows a graph of the MC rate with respect to the size of the white matter in the image. Obviously, as the size of the white matter is reduced in the image, the MC rate grows, and the MC rate decreases for images where the white matter covers a larger portion of the image. For comparison, [37] and [9] cite MC rate values for fuzzy classification results of 0.03 and 0.04, respectively.

V. CONCLUSION

In this paper, we present a classification algorithm for single-echo T1 MR brain images. The algorithm combines

low-level image statistics with higher level knowledge contained in semiautomatically classified key images and in a set of fuzzy indexes that capture knowledge about the basic nature of an MR brain image set. We classify the images one structure at a time in a volumetric fashion. By considering more than one image while classifying the brain structures we are able to keep track of 3-D boundary continuity as well as account for 3-D intensity and biological variations. We seek to capitalize on a variety of image aspects to mimic the way a human expert reasons when confronted with a pattern recognition task. The human expert will not take a mechanistic recognition approach, but instead will gather as much contextual information as possible and will use his own experience, expectation, and intuition. To define the set of fuzzy indexes that shape the final classifier we take into account the absolute and relative position of every pixel in the set of images, the gray level statistics of pixels, and of regions in the images. We compare size and statistics between 3-D neighborhoods and we use the user-input knowledge of the key images. Because the basic statistics that guide the segmentation of the stack of images are recalculated at every image and because we take into account the information of the segmentation results across the images, our tool succeeds in obtaining good quality results from a limited number of key images.

The issue of validation of the classification results is key and only in recent years literature has started to propose uniform reproducible validation indexes over standard MR image sets such as the Brainweb simulator and the image sets in [41]. However, we feel that much work needs to be done to propose a standard validation metric against which algorithmic performance can be compared. We believe that there is still the need in image segmentation literature to propose a robust and representative validation index. We think this index must not have the single metric form of current indexes. The new index should be the combination of a series of metrics: misclassified pixels, ratio of perimeter to area, volume preservation, higher statistical moments, topological characteristics, etc. This new metric could be read as a global index or as a series of indexes that could categorize algorithms based on different performance metrics. Current indexes are way too global and are hard to read and interpret. The methodology to measure validation indexes should also be standardized. In this paper, we have tried to give a detailed explanation of the validation of our classification algorithm so that it can be used for comparison in further algorithmic developments. The question of the highest accuracy that a classification algorithm can obtain is also important. The answer depends, among other things, of the characteristics of the image set considered as ground truth (partial volume, noise, inhomogeneities, biological variations, etc.) and which of these characteristics can be successfully taken into account by the classification algorithm. The last question can also be complemented with the question of what is considered good classification accuracy. Although we cannot give a definite answer to the accuracy questions, we think they could be better understood by obtaining standard validation indexes such as the average value of the Tanimoto index, similarity index, and MC rate over a series of manual expert classifications of the Brainweb simulated image sets. Since humans design the classification algorithms

based on their best understanding of the problem, it could be useful to know just how good they can get at the job.

REFERENCES

- [1] B. R. Gomberg, P. K. Shah, H. K. Song, S. N. Hwang, and F. W. Wehrli, "Parametric estimate of intensity inhomogeneities app to MRI," *IEEE Trans. Med. Imag.*, vol. 19, pp. 153–165, Mar. 2000.
- [2] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, T. Rohlfing, J. Goodey, and D. J. Hawkes, "Voxel similarity measures for 3D serial MR brain image registration," *IEEE Tran. Med. Imag.*, vol. 19, pp. 94–102, Feb. 2000.
- [3] J. Kershaw, W.-P. Kuan, S.-K. Jeng, and J.-H. Chen, "Application of Bayesian inference to fMRI data analysis," *IEEE Trans Med. Imag.*, vol. 18, pp. 1138–1153, Dec. 1999.
- [4] J. K. Udupa and G. T. Herman, *3D Imaging in Medicine*. Boca Raton, FL: CRC, 1991.
- [5] H. S. Choi, D. R. Haynor, and Y. Kim, "Multivariate tissue classification of MRI images for 3-D volume reconstruction," in *Proc. SPIE Medical Imaging III*, 1989, pp. 183–193.
- [6] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral MR images," *IEEE Trans. Med. Imag.*, vol. 16, pp. 176–186, Apr. 1997.
- [7] L. A. Zadeh, *Foreword in Fuzzy Logic in Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. i–iii.
- [8] —, "Foreword—Recognition technology and fuzzy logic," *IEEE Trans. Fuzzy Syst.*, vol. 9, pp. 3–4, Feb. 2001.
- [9] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, pp. 737–752, Sept. 1999.
- [10] C. Xu, D. L. Pham, M. E. Rettmann, D. N. Yu, and J. L. Prince, "Reconstruction of the human cerebral cortex from magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, pp. 467–480, June 1999.
- [11] J. L. Bloch, "Fuzzy relative position between objects in image processing: a morphological approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 657–664, July 1999.
- [12] C. Chang, H. Ying, T. A. Kent, J. Yen, L. M. Ketonen, M. L. Reynolds, and F. R. Hillman, "A new method for two-stafe hybrid fuzzy segmentation of MR images of human brains with lesions," *Int. J. Fuzzy Syst.*, vol. 4, pp. 873–882, Dec. 2002.
- [13] J. C. Rajapakse and J. Piyaratna, "Bayesian approach to segmentation of statistical parametric maps," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 1186–1194, Oct. 2001.
- [14] N. Duta and M. Sonka, "Segmentation and interpretation of MR brain images: An improved active shape model," *IEEE Trans. Med. Imag.*, vol. 17, Dec. 1998.
- [15] C. Han, W. S. Kerwin, T. S. Hatsukami, J.-N. Hwang, and C. Yuan, "Detecting objects in image sequences using rule-based control in an active contour model," *IEEE Trans. Biomed. Eng.*, vol. 50, pp. 705–710, June 2003.
- [16] M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. J. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Trans. Med. Imag.*, vol. 12, pp. 153–166, June 1993.
- [17] G. B. Aboutanos, J. Niskanen, N. Watkins, and B. M. Dawant, "Model creation and deformation for the automatic segmentation of the brain in MR images," *IEEE Trans. Biomed. Eng.*, vol. 46, Nov. 1999.
- [18] B. M. Bloch, "Fusion of numerical and structural image information in medical imaging in the framework of fuzzy sets," in *Fuzzy Syst. Medicine*: Springer Verlag, 2000, pp. 429–447.
- [19] T. Geraud, I. Bloch, and H. Maitre, "Atlas-guided recognition of cerebral structures in MRI using fusion of fuzzy structural information," in *Proc. CIMAFA'99 Symp. Artificial Intelligence*, La Havana, Cuba, 1999, pp. 99–106.
- [20] R. C. Herndon, J. L. Lancaster, J. N. Giedd, and P. T. Fox, "Quantification of white matter and gray matter volumes from three-dimensional magnetic resonance volume studies using fuzzy classifiers," *J. Magnetic Resonance Imag.*, vol. 8, pp. 1097–1105, 1998.
- [21] M. E. Algorri and F. Schmitt, "Surface reconstruction from unstructured 3D data," *Comput. Graphics Forum*, vol. 15, no. 1, pp. 47–60, 1996.
- [22] —, "Reconstructing the surface of unstructured 3D points," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1995, Proc. INRIA CVRMED, pp. 420–426.
- [23] M. Bomans, K.-H. Hoehne, U. Tiede, and M. Riemer, "3-D segmentation of MR images of the head for 3-D display," *IEEE Trans. Med. Imag.*, vol. 9, pp. 177–183, June 1990.

- [24] S. Kobashi, Y. Hata, and L. O. Hall, "Fuzzy information granulation of medical images: Blood vessel extraction from 3-D MRA images," in *Fuzzy Logic Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. 18–35.
- [25] H. D. Cheng, Y. G. Hu, D. L. Hung, and C. Y. Wu, "Breast cancer classification using fuzzy central moments," in *Fuzzy Logic Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. 36–53.
- [26] X.-S. Zhang, J. W. Huang, and F. J. Roy, "Depth of anesthesia control with fuzzy logic," in *Fuzzy Logic Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. 83–112.
- [27] M. Delgado, D. Sánchez, and M.-A. Vila, "Acquisition of fuzzy association rules from medical data," in *Fuzzy Logic Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. 286–310.
- [28] P. Felix, S. Barro, M. Lama, S. Fraga, and F. Palacios, "A fuzzy model for pattern recognition in the evolution of patients," in *Fuzzy Logic Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. 236–258.
- [29] B. Geva and D. H. Kerem, "Fuzzy clustering in medicine: Applications to electrophysiological signal processing," in *Fuzzy Logic Medicine*, S. Barro and R. Marin, Eds. Berlin, Germany: Physica-Verlag, 2002, vol. 83, pp. 139–176.
- [30] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, pp. 1085–1097, Nov. 1999.
- [31] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," *NeuroImage*, pt. 2/4, vol. 5, no. 4, p. S425, 1997.
- [32] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation," in *Lecture Notes in Computer Science*: Springer-Verlag, 1996, vol. 1131, Proc. Visualization in Biomedical Computing (VBC'96), pp. 135–140.
- [33] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, 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.
- [34] McGill Brainweb Simulator [Online]. Available: <http://www.bic.mni.mcgill.ca/brainweb/>
- [35] T. Tanimoto, "An elementary mathematical theory of classification and prediction," IBM Corp. Tech. Rep., 1958.
- [36] R. O. Richard O. Duda and P. E. Peter E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [37] W. C. Liew and H. Yan, "An adaptive fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Trans. Med. Imag.*, vol. 22, pp. 1063–1075, Sept. 2003.
- [38] P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: method and validation," *IEEE Trans. Med. Imag.*, vol. 13, pp. 716–724, Dec. 1994.
- [39] M. W. Vannier, T. K. Pilgram, C. M. Speidel, L. R. Neumann, D. L. Rickman, and L. D. Schertz, "Validation of magnetic resonance imaging (MRI) multispectral tissue classification," *Comput. Med. Imag. Graphics*, vol. 15, no. 4, pp. 217–223, 1991.
- [40] J. J. Bartko, "Measurement and reliability: Statistical thinking considerations," *Schizophrenia Bull.*, vol. 17, no. 3, pp. 483–489, 1991.
- [41] [Online]. Available: <http://www.cma.mgh.harvard.edu/isbr>
- [42] B. M. Dawant, S. L. Hartmann, J. P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel, "Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part 1, meth. and valid. on normal subjects," *IEEE Trans. Med. Imag.*, vol. 18, pp. 909–916, Oct. 1999.
- [43] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Seutens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, pp. 897–908, Oct. 1999.



María-Elena Algorri was born in Morelia, México. She received the B.S.E. degree in biomedical engineering from Universidad Iberoamericana, México City, México, in 1989, the M.S.E. degree in bioengineering from University of Washington, Seattle, in 1991, and the Ph.D. degree in signal and image processing from Ecole Nationale Supérieure de Télécommunications, Paris, France, in 1995.

From 1996 to 1998, she was a Research Engineer with Hewlett-Packard Medical Systems, Boeblingen, Germany. Since 1999, she has been a Professor in the Digital Systems Department, Instituto Tecnológico Autónomo de México, México City, where she is also head of the Neuroimaging Laboratory. Her research interests include all areas of visual computing, from computer vision to image processing and real-time interactive three-dimensional graphics.



Fernando Flores-Mangas was born in México City, México. He received the double B.S.E. degree in computer engineering and telematics engineering from Instituto Tecnológico Autónomo de México, México City, in 2002. Currently, he is pursuing the M.S. degree at the same university.

His research interests include vision, voice recognition, robotics, and artificial intelligence.