ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

NeuroImage

NeuroImage 18 (2003) 685–696

www.elsevier.com/locate/ynimg

Multicontext fuzzy clustering for separation of brain tissues in magnetic resonance images

Chaozhe Zhu[a] and Tianzi Jiang[a,b,*]

[a] *National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, People's Republic of China*
[b] *Department of Computer Science, University of Houston, Houston, TX 77204-3010, USA*

Received 17 May 2002; revised 3 October 2002; accepted 31 October 2002

Abstract

A local image model is proposed to eliminate the adverse impact of both artificial and inherent intensity inhomogeneities in magnetic resonance imaging on intensity-based image segmentation methods. The estimation and correction procedures for intensity inhomogeneities are no longer indispensable because the highly convoluted spatial distribution of different tissues in the brain is taken into consideration. On the basis of the local image model, multicontext fuzzy clustering (MCFC) is proposed for classifying 2D and 3D MR data into tissues of white matter, gray matter, and cerebral spinal fluid automatically. In MCFC, multiple clustering contexts are generated for each pixel, and fuzzy clustering is independently performed in each context to calculate the degree of membership of a pixel to each tissue class. To maintain the statistical reliability and spatial continuity of membership distributions, a fusion strategy is adopted to integrate the clustering outcomes from different contexts. The fusion result is taken as the final membership value of the pixel. Experimental results on both real MR images and simulated volumetric MR data show that MCFC outperforms the classic fuzzy c-means (FCM) as well as other segmentation methods that deal with intensity inhomogeneities.

Keywords: Fuzzy clustering; Cluster analysis; Information fusion; Image segmentation; Magnetic resonance imaging; Intensity inhomogeneities

1. Introduction

Magnetic resonance imaging (MRI) is an advanced, commonly used medical imaging technique. It can quantitatively provide rich information about human anatomy in two or three dimensions in a noninvasive way. In general, white matter, gray matter, and cerebral spinal fluid are three basic tissues in the brain. Brain tissue segmentation of magnetic resonance (MR) images means to specify the tissue type for each pixel or voxel in a 2D or 3D data set, respectively, on the basis of information available from both MR images and the prior knowledge of the brain (for notational simplicity, we use pixel for both 2D and 3D data). It is an important preprocessing step in many medical research and clinical applications, such as quantification of tissue volume, visualization and analysis of anatomical structures, multimodality fusion and registration, functional brain mapping, detection of pathology, surgical planning, surgical navigation, and brain substructure segmentation (Clarke et al., 1995; Suri et al., 2002). Unfortunately, intensity inhomogeneities in MR images, which can change the absolute intensity for a given tissue class in different locations, are a major obstacle to any automatic methods for MR image segmentation and make it difficult to obtain accurate segmentation results (Guillemaud and Brady, 1997; Shattuck et al., 2001; Worth et al., 1997).

Intensity inhomogeneities imply intensity variations over the same class of tissue that are not caused by random noise. Such spatial variations of the image signal result from both shading artifacts and inherent nonunifomity of tissue properties (Arnold et al., 2001).

Shading artifacts, which appear as a continuous, slowly varying shadowing effect over the whole image, are caused

* Corresponding author. National Laboratory of Pattern Recognition, Institute Automation Chinese Academy of Sciences, Beijing 100080, People's Republic of China. Fax: +86-10-62551993.
E-mail address: jiangtz@nlpr.ia.ac.cn (T. Jiang).

1053-8119/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1053-8119(03)00006-5

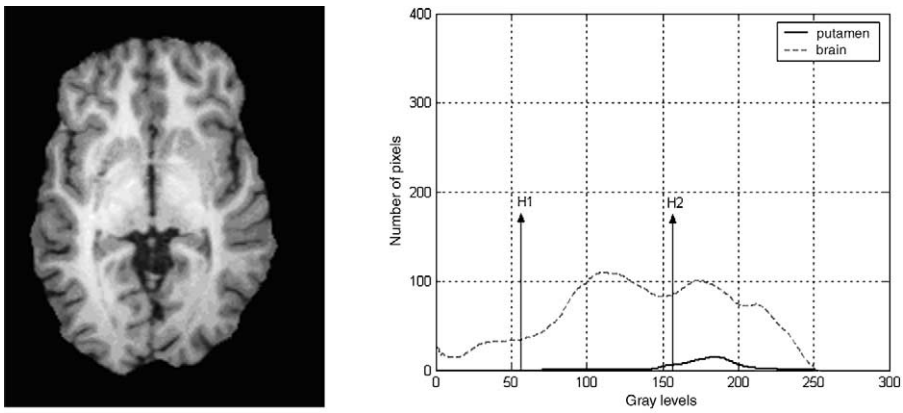


Fig. 1. Intratissue intensity variation in MR image: (a) Original image. (b) Intensity distributions of image (a).

mainly by hardware imperfection of the MRI devices, though other patient-related factors may also have an affect (Arnold et al., 2001; Guillemaud and Brady, 1997; Rajapakse and Kruggel, 1998; Simmons et al., 1994; Suri et al., 2002). Among them, RF inhomogeneity, inhomogeneous B_0 field, and gradient magnetic field nonuniformity are dominant. We refer to such a shading effect as artificial intensity variations (AIVs) because it appears in an image but not in the imaged object.

As we know, there exist quite a few spatially different substructures with different functions within each tissue class in the human brain (Tamraz and Comair, 2000). For instance, the cortex, caudate, and putamen are anatomically different substructures in the brain, but they all belong to gray matter (GM). Due to the inherent regional differences in imaging-related properties across substructures, the intensities in different substructures, even in the same tissue class, are also more or less different. The imaging-related properties that cause the inherent intensity variation include the composition, density, and magnetic properties (spin-lattice relaxation time T1, spin–spin relaxation time T2) of different tissues at different positions (Hornad, 1996). For example, white matter (WM) is brighter at the corpus callosum than in other regions, because the fiber bundles are more concentrated and coherent in direction. Moreover, intensities of caudate in T1-weighted MR images are usually higher than those of cortex as shown in Fig. 1 (Worth et al., 1997). In Fig. 1; the dashed and solid lines are the intensity distributions of the brain and putamen of the brain in Fig. 1a, respectively. H1 is the intensity boundary between cerebrospinal fluid (CSF) and GM, and H2, the intensity boundary between GM and WM. H1 and H2 are obtained from the traditional fuzzy c-means (FCM) algorithm. It is obvious that most parts of the putamen are brighter than the cortex and therefore they would be misclassified into WM when the inherent intensity variation is strong enough. Since these properties are connatural to human brain, such signal variations are also inherent in the brain. In contrast to the global, tissue class-independent artifact of AIVs, we call the inherent, tissue class-related intensity variations intratissue inherent variations (ITIVs).

Intensity inhomogeneities, including both AIVs and ITIVs have little effect on visual perception because the human visual system can correct such inhomogeneities automatically. However, they make intensity distribution within a particular tissue class flatter, and this results in overlapping intensity components among different tissues that are neighbors in the intensity histogram, as shown in Fig. 1b. That challenges the traditional image model of Gaussian mixture and impacts on the precision and reliability of automatic intensity-based segmentation methods.

To eliminate or alleviate the above-mentioned adverse impact, two kinds of methods have been proposed for correcting intensity inhomogeneities, namely, prospective and retrospective correction.

A phantom study is a typical prospective method (Axel et al., 1987; Dawant et al., 1993) in which the inhomogeneity profile of a particular MRI machine is estimated with a phantom, and then the estimated profile is applied during scanning of patients to compensate the intensity inhomogeneities. As above, intensity inhomogeneities are related to both instrument imperfection and the scanned objects. Thus, it is unrealistic to compensate different scanned patients with a fixed estimated profile. In other words, intensity inhomogeneities will vary with different machines and acquisition parameters from subject to subject and from slice to slice, which is inconsistent with the assumptions of patient independence in this method. All these drawbacks challenge the validity of prospective correction methods.

Various approaches have been proposed to compensate the intensity inhomogeneities retrospectively. Among them, a two-step method is typical, and requires a preoperation to estimate and remove the inhomogeneities, then perform segmentation on the corrected data. A direct way to approximate such inhomogeneities is with spline functions (Dawant et al., 1993) and polynomials (Meyer et al., 1995). However, the parameters of the basis functions or polynomials are estimated from manually or automatically selected

reference points, which is either time consuming or unreliable. Homomorphic filters, which are powerful tools in signal processing, have been used to remove low-frequency components of the inhomogeneities, while high-frequency details in MR images were also corrupted. The maximum entropy method (MEM) is another correction technique, and has shown great potential (Likar et al., 2000; Sled et al., 1998).

Compared with the two-step methods, several other methods perform estimation and segmentation in an iterative way so that intermediate information gained from the current segmentation can be used to improve the estimation result, which can in turn lead to more accurate segmentation. Two extended versions of the FCM method have been proposed to estimate inhomogeneities with an additive term in its object function. Pham and Prince (1999) used first- and second-order regularization terms to ensure the smoothness and continuity of the estimated bias field. Ahmed et al. (2002) used the neighborhood effect to bias the segmentation result toward piecewise-homogeneous labeling. Wells and Grimson (1996) modeled the inhomogeneities as a bias field and used the EM algorithm to perform the estimation and segmentation simultaneously in a Bayesian framework.

As discussed above, intensity inhomogeneities of AIVs have been widely addressed. ITIVs that can corrupt the intensity distribution of MR images no matter how perfect a scanner are, however, is seldom addressed in the literature. Additionally, though these methods deal with the problem in different ways, they can reach a consensus to estimate and correct the inhomogeneities. Intensity inhomogeneities estimated even from the same MR slice are quite different (Arnold et al., 2001), saying nothing of those from different scanners, individuals, and scanning times. Therefore, the biggest danger in removing intensity inhomogeneities is that some useful signals may also be removed (Arnold et al., 2001; Worth et al., 1997). These issues are addressed in this article.

The rest of this article is organized as follows. In Sections 2, we propose a novel image model to take into account both AIVs and ITIVs. By simplifying the model into its local version, the adverse impact of both AIVs and ITIVs is eliminated. Based on the local model, a new method is presented in Section 3 to classify 2D and 3D MR data effectively and precisely while avoiding the possibility of removing useful information. In Section 4, experimental results on both real 2D MR images and simulated 3D MR data are presented. A detailed discussion on the performance of the multicontext fuzzy cluster is given in Section 5. The final section is devoted to conclusions and future work.

2. Image model

2.1. Previous image models

Models of intensity inhomogeneities are the basis of retrospective correction methods. An approach to model intensity inhomogeneities has been proposed in the literature (Ahmed et al., 2002; Guillemaud and Brady, 1997; Pham and Prince, 1999; Rajapakse and Kruggel, 1998; Wells and Grimson, 1996). In this approach, intensity inhomogeneities are modeled as a continuous, slowly varying multiplicative field α over the image domain with constant true intensity v_k for each tissue class k,

$$y_i = \alpha_i x_i + n_i, \qquad x_i \in \{v_1, v_2, \ldots, v_c\}, \tag{1}$$

where y_i and x_i are the observed and true intensity at the ith pixel, respectively; n_i is the measurement noise of independent white Gaussian distribution at pixel i; N is the total number of pixels in a MR image, and C is the desired number of tissue classes (here C = 3 for WM, GM, and CSF). Model (1) is widely adopted due to its simplicity and effectiveness. It can model AIVs very well; however, it makes nearly no consideration of ITIVs.

Another approach that is not widely used assumes that the true intensity of each tissue class, v^k, is spatially varying and independent of another (Nocera and Gee, 1997). The observed intensities y_i are obtained by a model of partial volume effect (PVE),

$$y_i = \sum_{k=1}^{C} t_i^k v_i^k, \tag{2}$$

with the constraints $\forall i, \sum_{k=1}^{c} t_i^k = 1, \forall i, \forall k, t_i^k \geq 0$, where y_i, i, N, and C are the same as those in model (1). t_i^k is the signal contribution of tissue k at location i, and v_i^k is the true intensity of tissue k at location i ($i = 1,2, \ldots, N$).

The total number of true intensities increases from C in model (1) to N which indicates that the true intensities vary with location in the MR images.

2.2 A novel image model

As we discussed in Section 1, intensity inhomogeneities are common for both AIVs and ITIVs (Rajapakse and Kruggel, 1998). Moreover, AIVs and ITIVs should be modeled separately to account for their significant difference in both appearance and cause of formation. However, intensity inhomogeneities are partially modeled as either AIVs without ITIVs in model (1) or as ITIVs without AIVs in model (2). Therefore, neither model (1) nor model (2) is an objective description for intensity inhomogeneities, which limits those methods that are based on the two models.

To tackle the problem, we propose the image model

$$y_i = \alpha_i x_i + n_i, \qquad x_i \in \{v_1, v_2, \ldots, v_{NJ}\}. \tag{3}$$

In this model, y_i, x_i, n_i, i, N, and C are the same as in the previous models. AIVs are modeled with α_i as in model (1). The true intensity v_i, which varies with the location i in the brain, is used to model ITIVs. Therefore, model (3) can

properly model AIVs and ITIVs simultaneously. Moreover, it also offers an in-depth description of the formation of inhomogeneities caused by AIVs and ITIVs.

2.3. Local image model

As discussed above, model (3) is a relatively complete description of intensity inhomogeneities. However, it is difficult to estimate AIVs and ITIVs precisely. Fortunately, due to the highly convoluted 3D spatial distribution of different tissues in human brain, a local version of model (3) can be deduced to eliminate the adverse effect of intensity inhomogeneities. Therefore, estimation and correction for such inhomogeneities are no longer necessary, which guarantees no information loss during the correction process.

Clustering context is a key concept throughout this work. It is formally defined as a spatially connected subset of 2D MR images or 3D volume data. The size or capacity of a context is defined as the number of pixels in the context. We denote clustering context W of a pixel P as $W(P)$, and denote the size of $W(P)$ as $\|W(P)\|$.

On the grounds of the following three assumptions, which will be proved reasonable by experimental results in the following sections, model (3) can be simplified in its local version.

1. Bias field α_i, $1 \le i \le N$, is smooth and slowly varying.
2. Within a context, the C classes of tissues exist together and there are considerable pixels in each tissue class.
3. Within a context, all pixels of the same tissue have similar true intensities.

From the first assumption, the multiplicative field term within a clustering context of properly small size can be approximately treated as a constant field:

$$\alpha_i \approx \alpha, \ \forall i \in W. \tag{4}$$

The complex topology of the brain, which causes bends and twists of spatial distributions of tissues of WM, GM, and CSF, makes assumption (2) solid even in properly small contexts. From this assumption, it can be deduced that there are always C tissue classes in any clustering context. Hence, the determination of cluster number in a context, which is difficult in clustering analysis, is solved. With assumption (3) as well as the fixed cluster number C, the total number of true intensities in context W decreases from $\|W\|$ to C:

$$x_i \in \{v_1, v_2, \ldots, v_c\}, \ \forall i \in W. \tag{5}$$

From (4) and (5), model (3) can be simplified in the local model

$$\begin{aligned} y_i^j &= x_i^j + n_i^j \\ x_i^j &\in \{\alpha^j v_1^j, \alpha^j v_2^j, \ldots, \alpha^j v_c^j \\ &\forall i \in W_j, \quad 1 \le j \le N_w, \end{aligned} \tag{6}$$

where y_i^j is the observed intensity and x_i^j is the true intensity modulated by the bias field at pixel i in clustering context W_j; n_i^j is measurement noise at pixel i in W_j; v_k^j ($k = 1,2,\ldots,C$) is the approximate true intensity of the kth tissue in W_j; α^j is a constant in context W_j; N_w is the total number of clustering contexts in volume data; and C is the desired number of tissues.

Thanks to the local model, the essential characteristics of the ideal image model of Gaussian mixture, which was once corrupted by both AIVs and ITIVs can be recovered. This facilitates any segmentation method.

3. Methodology

3.1. Fuzzy c-means

FCM is a commonly used clustering approach (Hall et al., 1992). It is a natural generalization of the K-means algorithm allowing for soft segmentation based on fuzzy set theory. To classify a data set of N data items into C classes FCM can be formulated as a minimization problem of the objective function J with respect to the membership functions u and centroid v, where J is given by

$$J = \sum_{k=1}^{c} \sum_{j=1}^{N} u_{ki}^d a_{ki}^2, \tag{7}$$

and it is subject to

$$u_{ki} \in [0,1], \quad \sum_{k=1}^{c} u_{ki} = 1, \quad 0 < \sum_{i=1}^{N} u_{ki} < N,$$
$$1 \le i \le N, 1 \le k \le C.$$

Here m (≥ 1) is a parameter determining the amount of fuzziness of the clustering results (if $m = 1$, then the algorithm degenerates to the "crisp" K-means clustering algorithm, so we suppose $m > 1$ in the rest of the paper); $d_{ki} = \|y_i - v_k\|$ is the Euclidean distance between the observed data y_i and the class centroid v_k, and u_{ki} is the membership value reflecting the degree of similarity between y_i and v_k.

The objective function J is minimized when high membership values are assigned to pixels whose intensities are close to the centroid of its particular class, and low membership values are assigned to them when the pixel data are far from the centroid.

Taking the first derivatives of J in (7) with respect to u_{ki}, v_k we can obtain the following necessary conditions to minimize the objective function J:

$$u_{ki} = \frac{(d_{ki})^{-2/(m-1)}}{\sum_{l=1}^{c} (d_{li})^{-2/(m-1)}},$$
$$1 \le i \le N, 1 \le k \le C, \tag{8}$$

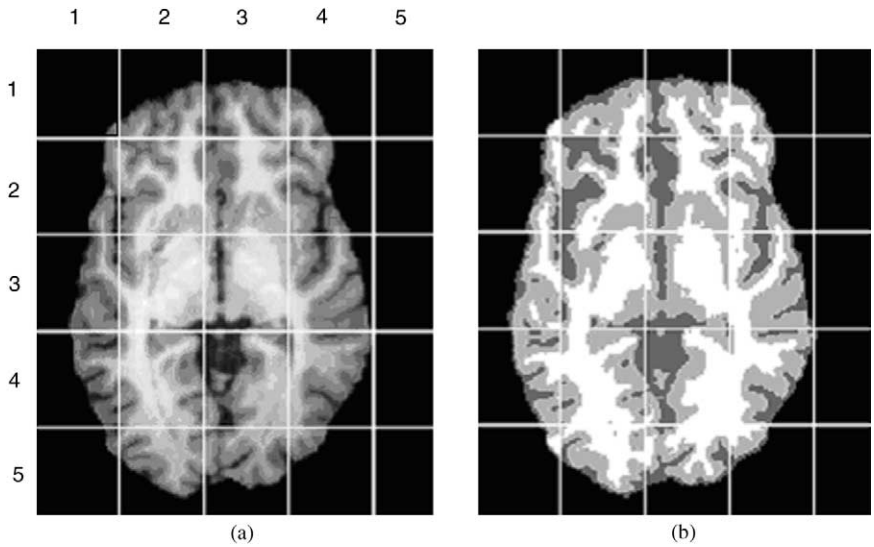


Fig. 2. MGFC hard segmentation: (a) Original T1-weighted image. (b) MGFC segmentation.

$$v_k = \frac{\sum_{i=1}^{N} u_{ki}^m y_i}{\sum_{i=1}^{N} u_{ki}^m}, \quad 1 \le k \le C. \tag{9}$$

After initialization of the centroids, u_{ki} and v_k are iteratively calculated until some stop criteria are reached. Finally, the segmentation can be obtained by the principle of maximum membership.

3.2. Multigrid fuzzy cluster (MGFC)

The nonoverlapping multigrid version of FCM is a natural and simple idea of performing FCM in a local way. One context means one grid area here, so there is only one context $W(p)$ for each pixel p in MGFC. This algorithm can be described as follows:

$$W(p) \Rightarrow u_{w(p)}^k \tag{10}$$

$$u_p^k = u_{w(p)}^k, \quad k = 1,2 \ldots C. \tag{11}$$

Though this method can avoid the disadvantage of FCM of missing spatial information, it is essentially single context. The membership value of each pixel is determined only by its single local context and no information can be exchanged among neighboring contexts. Therefore, MGFC is very sensitive to both the size of the local clustering context and the amount of data in a context, and cannot preserve the statistical reliability and spatial continuity of segmentation results. This can be illustrated with Fig. 2.

Fig. 2 shows the hard segmentation result of MGFC applied on a real MR image in Fig. 1a. The image is uniformly divided into 5×5 nonoverlapping elements as shown in Fig. 2a. FCM is carried out in each grid area independently. From Fig. 2 we can see some problems with MGFC.

First, the correctness of assumption (2) cannot hold in some contexts. For instance, in some contexts across the boundary between brain and background, only some of the pixels are brain tissues (WM, GM, and CSF) due to the removal of nonbrain tissue. In such contexts, the data involved in collecting statistics for intensity distribution are often far fewer than in the other contexts. The statistical distribution of such few samples is deviant and inconsistent with the typical distribution of the brain, which can lead to obvious misclassification. In grid (1,1) in Fig. 2b, for example, nearly all the pixels of GM are misclassified into WM. Additionally, such inconsistent intensity distributions may occur in some contexts even in the inner part of the brain. As show in context (2,2), tissues of WM and GW predominate, with little CSF, which yields a deviant intensity distribution far from that typical of the brain. To minimize the object function in (7) with such intensity distribution, FCM has to push some GM pixels of relatively low intensity into the class of CSF, which leads to misclassifications.

Second, the variation of intensity distributions of neighboring contexts would, more or less, lead to inconsistent segmentation results across the context boundaries. For example, the segmentations of GM are obviously inconsistent at the boundary of grid (3,3) and grid (3,4) in Fig. 2b.

Enlarging the size of contexts of course can make assumption (2) more reasonable by involving more sample points. However, the enlarged context would weaken the ability of eliminating the effect of AIVs and ITIVs because the larger the context, the larger the gap between assumptions (1) and (3) and the truth. Accordingly, the determination of proper size of clustering context $\|W(p)\|$ is a dilemma in MGFC, because the tissue distributions vary with the

slice orientation, position, and even the different contexts in the same slice. It hinders the use of MGFC in MR brain segmentation.

3.3. Multicontext fuzzy cluster (MCFC)

Despite the above-mentioned problems of MGFC, the assumptions for simplifying the image model are still correct in most cases of context locations, due to the complicated and convoluted structure of human brain as mentioned above. In other words, most contexts would yield good soft judgments. An intuitive idea is that multiple clustering contexts for each pixel and the adverse impact of some contexts could be submerged by all the good clustering results from other contexts. Such a consideration leads to the development of a novel method called the multicontext fuzzy cluster (MCFC) that can take advantage of the adaptation of local clustering, but also keep the classifications spatially continuous and statistically reliable.

MCFC includes two basic parts: fuzzy clustering and information fusion. The rationale of MCFC can be described as follows. In contrast to single context in MGFC, N_p clustering contexts with size $\|W\|$ are automatically generated for each pixel P. From the viewpoint of information fusion theory, the clustering contexts of point P can be regarded as information sources to determine the tissue class of the point, and the intensity distributions in contexts are the information provided by the sources. Then fuzzy clustering is independently performed in each context $W_l(P)$ ($l = 1,2, \ldots, N_p$) to calculate $u^k_{w(p)}$, the membership of P to tissues k. Here, the memberships can be regarded as soft decisions made on the information from each information source independently. In the part of information fusion, all the soft decisions $u^k_{w(P)}$ are integrated with some strategy F to bias the final membership functions toward the results, which are similar to those from most of the contexts.

For any pixel P, MCFC can be summarized as follows:

$$\text{Step 1.} \quad W_l(P) \overset{FCM}{\Longrightarrow} u^k_{w(p)}, \quad k = 1,2, \ldots, C; \quad l = 1,2, \ldots, N_p \tag{12}$$

$$\text{Step 2.} \quad u^k_p = F(u_{w1(P)k}, u^k_{w2(p)}, \ldots, u^k_{wN_p(P)}) \tag{13}$$

Here, the weighted averaging operation is taken as the fusion strategy:

$$F(u^k_{w1(p)}, u^k_{w2(p)}, \ldots, u^k_{wN_p(P)}) = \sum_{l=1}^{N_P} a_{w_l(p)} u^k_{w_l(p)} \tag{14}$$

$$a_{w_l(p)} = \frac{\exp\{-d(p, O_l(p))/\sigma\}}{\sum_{i=1}^{N_p} \exp\{-d(p, O_l(p))/\sigma\}}. \tag{15}$$

The reliability measure of an information source or the weighting coefficient, $a_{wl(p)}$ for a context, formulated as (15), is inversely proportional to the distance between p and the center of the context O. Parameter σ in (15) determines the profile of the weight function. Such a definition reflects the fact that the closer the distance between p and the center O, the more reliable the contextual information provided. As a result, the averaging of clustering outcomes from different contexts makes the final clustering result more reliable. Moreover, the weighted averaging operation can also preserve spatial continuity of the membership distributions of each tissue class.

3.4. Implementation details

To implement MCFC, the shape of contexts has to be determined in advance because it decides how to generate the clustering contexts. Theoretically, the selection of shape should satisfy the three assumptions in Section 2.3 as possible as it can, especially for the second. For simplicity, we select a rectangle for 2D MR images and the ratio of width and height of the rectangle should be the same as that of the input MR image. In a similar way, we use cuboids as a clustering context for 3D MR data. Size of a context, the only input parameter to be set for MCFC, is represented by the normalized context size, which is defined as

$$\alpha = \frac{\|w\|}{N_{\text{tissue}}}, \tag{16}$$

where N_{tissue} is the amount of tissue pixels in the input image. With this definition, the absolute size of the context can change adaptively with the size of the brain in input images. In the rest of this section, we expatiate on the implementation of MCFC on 2D images, because it is similar for 3D data.

Calculation of width and height of rectangular context

With a given normalized size of context α, equations for calculating the width (W_{w}) and height (W_{h}) of the rectangle are:

$$W_w \times W_h = \alpha \times N_{\text{tissue}}, \tag{17}$$

$$W_w / I_w = W_h / I_h, \tag{18}$$

where I_{w} and I_{h} are the width and height of the input image, respectively. By solving Eqs. (17) and (18), we have

$$W_w = \sqrt{\frac{\alpha I_w N_{\text{tissue}}}{I_h}} \tag{19}$$

$$W_h = \sqrt{\frac{\alpha I_h N_{\text{tissue}}}{I_w}} \tag{20}$$

$$N_{\text{tissue}} = N - N_{bk}, \tag{21}$$

where $N = I_{\text{w}} \times I_{\text{h}}$ is the pixel number of the whole image, and N_{bk} is the number of background pixels whose intensities are set to 0 when removing nonbrain tissues.

Generation of multiple contexts

Moving a rectangular window through the input image, step by step from left to right, and top to bottom with the step length of S_w and S_h, respectively, is the basic process for generating contexts. Once the moving is completed, the contexts required for all pixels are generated. In such a way of generating contexts, the number of clustering contexts for one pixel is determined by S_w and S_h. For example, if S_w and S_h are half of W_w and W_h, respectively, there are $2 \times 2 = 4$ contexts for each pixel. They can be rewritten as follows:

$$S_w = W_w/N_w, \tag{22}$$

$$S_h = W_h/N_h. \tag{23}$$

Then the number of contexts for each pixel is

$$N_p = N_w \times N_h. \tag{24}$$

According to our experience, $N_p = 2 \times 2 = 4$ is enough for 2D MR images and $N_p = 2 \times 2 \times 2 = 8$ is adequate for 3D MR data. During the moving process, all pixels in the rectangular window share the same context. It makes the total number of clustering contexts for an image far less than N_pN, which indicates each pixel owns N_p different clustering contexts. If $N_w = N_h = 2$, the total number of contexts for an image can be formulated as

$$NT = \left(\frac{I_w}{W_w} - 1\right) \times \left(\frac{I_h}{W_h} - 1\right). \tag{25}$$

Fuzzy clustering and weighted averaging

The fuzzy clustering as well as weighting coefficients calculating is performed along with the moving of the rectangular window. For one context, FCM algorithm can be summarized as follows:

a. Using the clustering results of $\{v_k\}_{k=1}^{C}$ from the previously generated context as the initial cluster centroids.
b. Update the membership matrix with (8).
c. Update the cluster centroids with (9).

Repeat (b) and (c) until convergence of the algorithm.

To implement the fuzzy clustering described above much more efficiently, we use a histogram-based fast FCM that is detailed in another paper to be submitted. Using the same initial cluster centroids and termination criteria, this fast algorithm yields identical clustering results with much less computational time.

As we know, initialization is very important for meaningful clustering results and reduction of computation time. In MCFC, current context makes use of the clustering results from the previous context as the initial values. Hence, the good initializations reduce the computational time and, to a certain extent, can avoid the local minimum problem in traditional FCM. After the iterative calculations of FCM terminate, weighting coefficients for each pixel in the context are calculated with (15). Then the membership values multiplied by the corresponding weighting coefficient at each pixel are accumulated to membership maps of the whole image. After all contexts are processed, the accumulated membership values at each pixel in the accumulated membership maps are divided by the context number of the corresponding pixels to obtain the weighted averaging membership values.

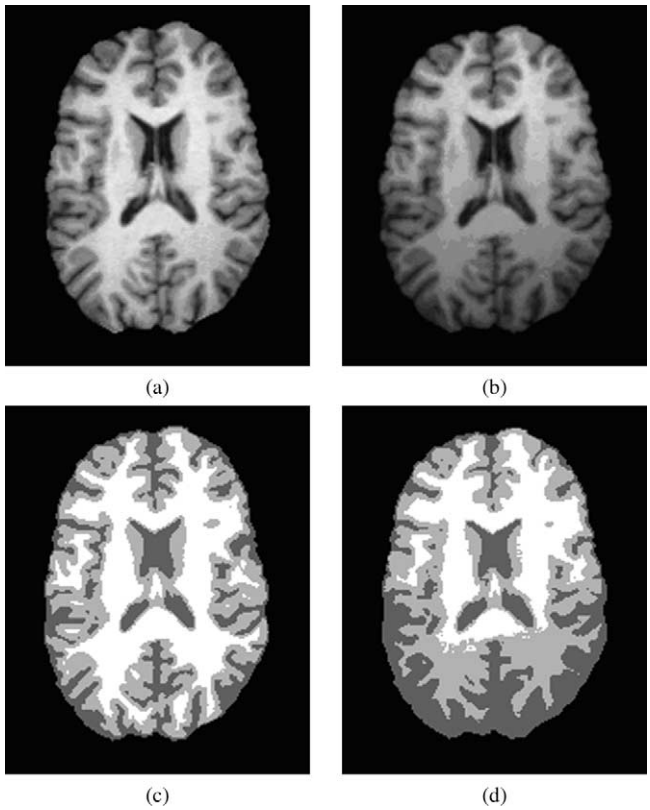


Fig. 3. MCFC and FCM hard segmentations in the case of AIV: (a) Original image. (b) Corrupted image of (a). (c) MCFC segmentation. (d) FCM segmentation.

Hard segmentation

Finally, maximum membership principle is used to obtain the hard segmentation from the averaging membership maps.

4. Results

MCFC was implemented in MATLAB on a PC with Intel Pentium III 866-MHz processor and 512M RAM. It has been tested on both real MR images and 3D simulated MRI data from the McConnell Brain Imaging Center at the Montreal Neurological Institute, McGill University (Collins et al., 1998).

4.1. Evaluation with real data

Fig. 3 shows the results from MCFC and FCM on real MR images corrupted by an obvious AIV. FCM misclassi-

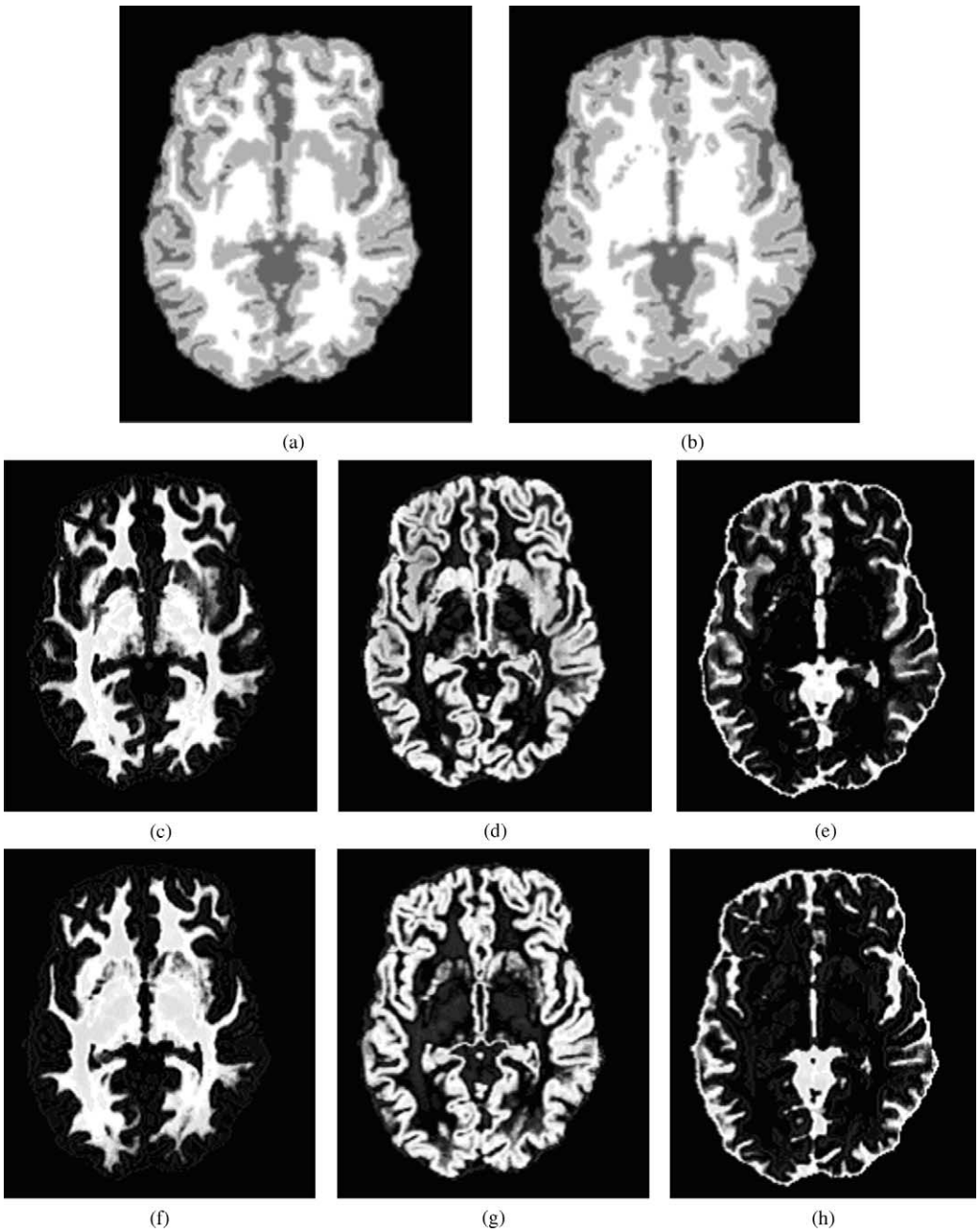


Fig. 4. FCM and MCFC hard segmentation and membership functions: (a) MCFC hard segmentation. (b) FCM hard segmentation. (c–e) Membership functions of WM GM and CSF in MCFC. (f–h) Membership functions of WM GM and CSF in FCM.

fied nearly one-third of all pixels in the image, while MCFC performing fuzzy clustering in a local way yields a much better result.

Fig. 4 shows the hard and soft segmentation results of FCM and MCFC on real MR images with ITIVs in Fig. 4. Figs. 4a and b are the hard segmentations of MCFC and FCM, respectively. Figs. 4c–e are membership functions of WM, GM, and CSF of MCFC, and Figs. 4h–f are membership functions of FCM. It is obvious that the typical intensity of putamen is higher than that of the cortex. Moreover, most intensities of putamen are closer to WM than to GM as shown in Fig. 1b. As a result, the membership function of GM at the putamen area calculated by FCM is much lower than those of other GM areas and FCM misclassified most parts of the putamen into WM with the principle of maximum membership. In contrast, MCFC can yield satisfactory result, which is more compatible with human visual perception.

We also make a comparison between MCFC and MGFC with same size of contexts. Compared with the segmentation result of MGFC in Fig. 2b, MCFC can yield a spatially continuous and reliable segmentation result as shown in Fig.

Table 1
MCR from simulated data results

Method	MCR		
	INV = 0%	INV = 20%	INV = 40%
FCM	4.020%	5.440%	9.000%
FM-AFCM	4.168%	4.322%	4.938%
MCFC	4.270%	4.200%	4.290%

4a. The inconsistent segmentations across context boundaries and obvious misclassifications due to deviant intensity distributions in Fig. 2b have been overcome by MCFC. These improvements imply that MCFC is more robust to context size and can preserve the spatial continuity of WM, GM, and CSF of each class.

4.2. Evaluation with simulated 3D data

To evaluate the performance of a clustering algorithm quantitatively, we define the misclassification error (MCR) of an algorithm as the quotient between the number of pixels misclassified by the algorithm and the total number of pixels.

We applied both MCFC and traditional FCM on three simulated T1 volume data ($217 \times 181 \times 181$, 1-mm cubic voxels, 3% noise, and 0, 20, and 40% intensity nonuniformity, respectively). The experimental results are listed in Table 1.

To compare it with another effective algorithm, called FM-AFCM, proposed by Pham and Prince (1999), the MCRs of FM-AFCM tested under the same conditions are also listed in Table 1.

From Table 1 we can see that MCFC and FM-AFCM are much more robust to increased inhomogeneities than FCM. A careful comparison between MCFC and FM-AFCM reveals that MCFC can stably preserve lower MCR than FM-AFCM as inhomogeneities increase. By comparing this table with Table 1 in Pham and Prince (1999), it can be implied that MCFC can also outperform other proposed methods such as the unsupervised EM algorithm for finite Guassian mixture models (Liang et al., 1992) and the adaptive Markov random field (AMRF) method (Goldszal et al., 1998).

Besides the quantitative evaluations, we also show the visual comparisons between traditional FCM and MCFC applied on the simulated data (3% noise and 40% intensity nonuniformity) in the following figures.

Two transverse slices drawn from the 3D segmentation results are shown in Figs. 5 and 6. Figs. 5a and 6a are the original slices from the simulated MR data. Figs. 5b and 6b are the true models of the two slices. Figs. 5c and 6c show the hard segmentation results of FCM. Figs. 5d and 6d are the hard segmentation results of MCFC.

Because of the AIVs in the simulated MR data near the

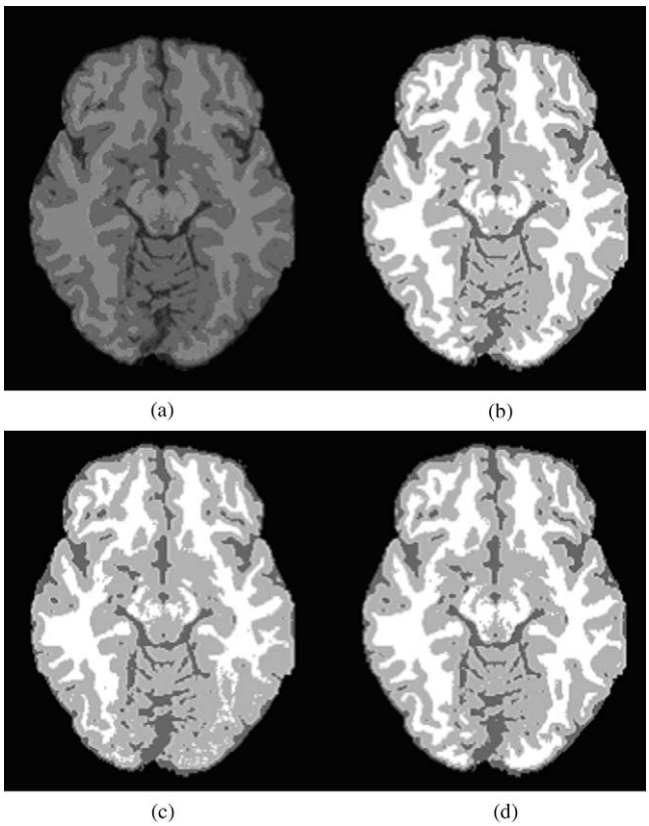


Fig. 5. FCM and MCFC hard segmentations: (a) Original image. (b) True model (c) FCM segmentation. (d) MCFC segmentation.

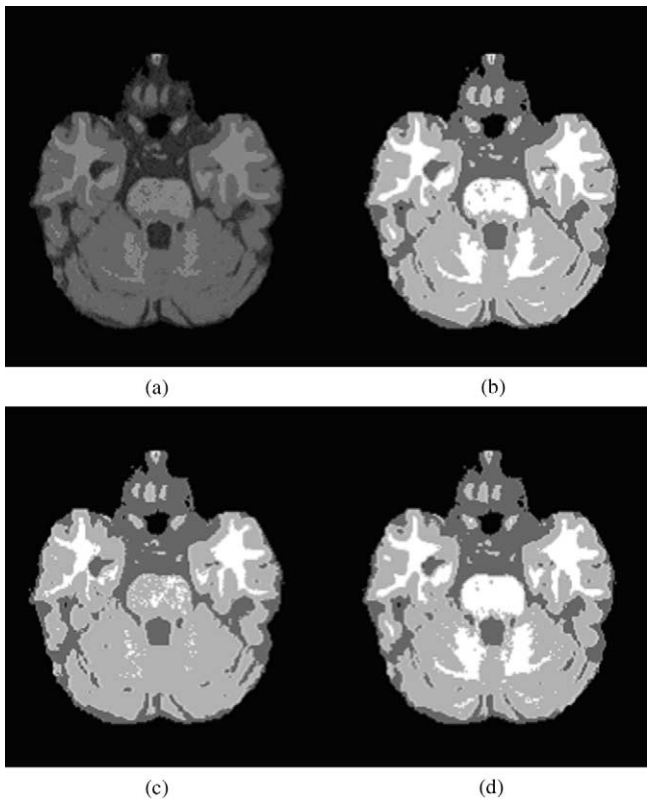
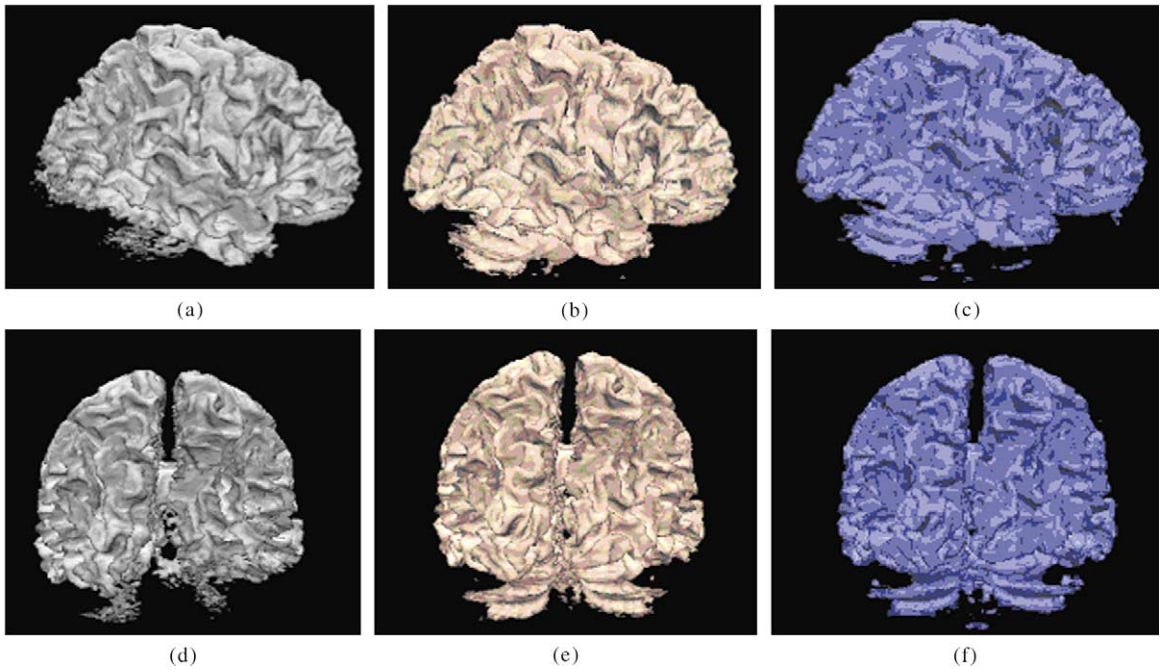
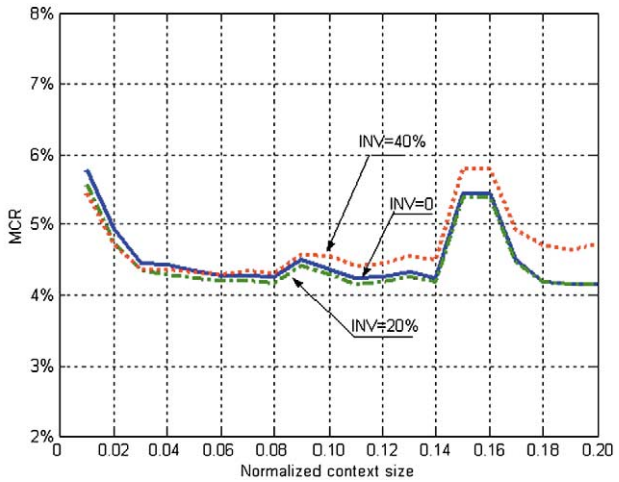


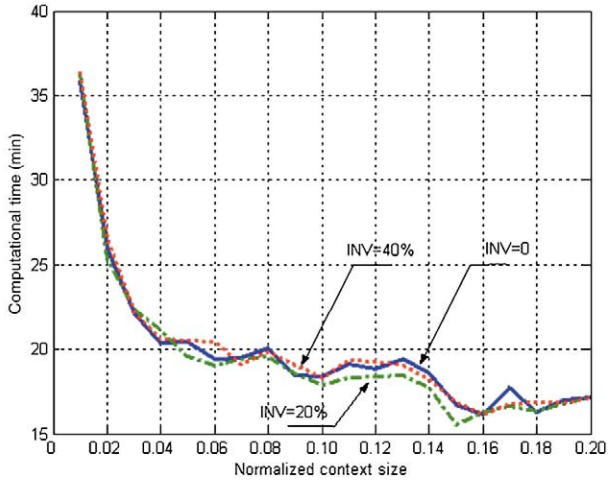
Fig. 6. FCM and MCFC hard segmentations: (a) Original image. (b) True model (c) FCM segmentation. (d) MCFC segmentation.



7



8



9

Fig. 7. Three-dimensional renderings of the segmentation results of WM: (a,d) FCM segmentation. (b,e) (b) MCFC segmentation. (e,f) True model.
Fig. 8. Relationship between MCR and size of context.
Fig. 9. Relationship between computation time and size of context.

occipital lobe and cerebellum area, the segmentation results of FCM, especially in WM, are obviously deteriorated. MCFC, however, can yield much more complete and continuous results, which are very similar with the true models.

Three-dimensional renderings of the segmentation results of WM are shown in Fig. 7. Results of FCM, results of MCFC, and the true model of WM are shown in the left, middle, and right columns, respectively. In each column, the top and bottom rows are the same results viewed from two different angles. We can see that WM in the left columns is incomplete: part of the WM near the occipital lobe and most of the cerebellar WM are missed due to the 40% artificial intensity inhomogeneities. In contrast, the WM segmented by MCFC in the right columns is much more complete, and the symmetrical structure of cerebellar WM is very clear and complete. The segmentation of MCFC is therefore much closer to the true model in the right column.

4.3. Clustering context size and MCR

In this experiment, we have also studied the relationship between size of context and segmentation accuracy. In fact, the larger the context, the more data to be clustered; the greater similarity between the intensity distribution of one context and that of the input image, the more reliable the clustering results. This is desirable for assumption (2). On the other hand, the smaller assumptions (1) and (3) require context size, to eliminate the adverse effects of AIVs and ITIVs.

So there exists a conflict in the selection of context size. MCFC, however, can offer a good solution to this dilemma.

Averaging of multiple clustering results in MCFC makes the final segmentations more insensitive both to the size of context and to the amount of data in a context because of the information exchanges among the contexts.

The relationship between MCR and context size of MCFC performed on simulated volume data is shown in Fig. 8. We see from Fig. 8 that MCFC can yield low MCR when the normalized context size is in the range 0.04–0.08 in the condition of different intensity nonuniformity variation (INV = 0%, INV = 20%, INV = 40%).

5. Discussion

5.1. Clustering strategies: global, local and monocontext, local and multicontext

FCM, as well as other global segmentation methods not corrected for intensity inhomogeneities, conduct classification based on the similarity between pixels and global cluster centroids without incorporating any spatial information. Such similarity is easily corrupted by both AIVs and ITIVs. In contrast, MCFC and MGFC conduct classification based on the local intensity difference. Therefore, spatial information is naturally incorporated, and subtle intensity differences can be detected in a context to yield much more reasonable segmentation with a smaller impact from intensity inhomogeneities.

In MGFC, each pixel has only one context for clustering. When there is even one context, in this case, assumption (2) is not valid, the misclassification in this context can never be corrected, because nothing is exchanged between the neighboring grid areas as MCFC does. This makes the whole segmentation result useless even though satisfactory results are achieved in other contexts. In practice, MGFC is too sensitive to the size of context to make any practical classification for tissue segmentation. In MCFC, however, there are multiple clustering contexts for each pixel and the adverse impact of some context could be minimized by all the good clustering results from other contexts. Accordingly, MCFC can take advantage of the adaptation of the local clustering, but it can also keep the classifications statistically reliable. Hence MCFC theoretically can outperform not only global intensity-based tissue segmentation methods but also their local and monocontext versions, especially in the presence of significant intensity inhomogeneities.

5.2. Validity

The human brain is a complicated, highly interconnected and convoluted 3D structure. It is because of the special 3D structure of the human brain that such subset clustering in MCFC can be successfully performed throughout the whole brain with constant tissue number to classify. On segmenting 2D MR brain images, some 3D information is missed. This reduces the statistical correctness of assumption (2). Hence MCFC will theoretically have better performance for 3D MR data of human brain than for 2D data.

Brains of different subjects are not identical in detail; however, subvolumes occupying approximately the same part of the different brains are statistically similar enough in the 1D intensity histogram so that MCFC can be performed on to any other normal brains with similar accuracy, leaving nearly nothing to chance in the algorithm.

5.3. Speed

Increasing the number of clustering contexts in MCFC does result in an increase in computational load. However, two advantageous factors should also be taken into account on comparing it with traditional FCM. First, each context has far fewer pixels than the whole image; therefore the computational load of a context is far lower than that of applying FCM directly to the whole image. Second, as mentioned in Section 3.4, during moving of the clustering window, current context inherits the clustering results from the previous context as its initial values. Thus, the good initializations reduce the computational time spent on a context. As a result, the final computational time of MCFC, which is approximately the product of the number of contexts and computational time per context, is moderately slower than that of FCM.

The computational time of MCFC in the experiments with simulated volumetric MRI data is plotted with the different context sizes (0.01–0.20) in Fig. 9, each for one of three INV levels (0%, 20%, and 40%). From Fig. 9, we see that computational time with any one of the three different INV levels decreases with the reduction of context size. And the computational time is between 15 and 20 min for most context size. The average computational time of FCM for the three different INV levels is approximately 15 min. Hence, MCFC is slightly slower than FCM.

Moreover, MCFC is very easy to implement in a parallel way since each context or each context packet can perform simultaneously and independently on different CPUs or different computers to reduce the computing time dramatically. For example, MCFC can be implemented in a parallel way with one master and multislave architecture. The master calculates the size of clustering context and spatially divides all the clustering contexts into packets. Therefore, each packet corresponds to a subset of the image (image block), which is the union of all the contexts in the packet. Each image block is sent from the master to its corresponding slave. After one slave receives its data, it begins to move the clustering window in the image block just as in Section 3.4 to facilitate clustering, coefficient calculation, and membership map accumulation, context by context, until the moving is completed. Once all the slaves independently finish their jobs, the accumulated membership maps corresponding to image blocks are sent back to the master, where all the membership maps of image blocks are combined to build the complete membership map for the whole image.

Finally, hard segmentation of the input image can be obtained with the principle of maximum membership.

5.4. Noise

MCFC is a framework that can embed FCM as well as any other soft segmentation methods to calculate the membership degree in each context. Therefore, any traditional noise-suppressing method, such as neighboring effects (Ahmed et al., 2002) and Markoven random field as a prior (Rajapakse and Kruggel, 1998), can be used in each context to achieve better soft segmentation even from noisy images.

6. Conclusions and future work

We have presented a theoretically simple and practically effective approach to automatic tissue segmentation of 2D or volumetric MRI data of the human brain. Experimental results have shown that our MCFC outperforms FCM as well as some other proposed methods in the presence of intensity inhomogeneities. Accordingly, MCFC would also be an effective technique available in many clinical applications.

The deviation of the intensity distribution in a context from that of the whole data set plays a key role in determining the correctness of assumption (2) and the reliability of clustering results in the context. If we can evaluate the deviations quantitatively, we can make use of them as the more effective weighting coefficients to cut down the context number for each pixel and reduce the computational time. Therefore, in the future, we plan to develop a quantitative evaluation of the deviations. Moreover, we would also like to figure out the relationship between the performance of MCFC and the shape and amount of contexts.

Acknowledgments

The authors are highly grateful to three anonymous reviewers for their significant and constructive critiques and suggestions, which improve the article very much. The authors thank Professor George Zouridakis in the Department of Computer Science, University of Houston, for his proofreading of our manuscript. We also thank the members of the Medical Imaging and Computing Group—W.L. Liu, L.F. Liu, Y. Fan, J.G. Zhang, and A. Luo—for their comments and support in this work and Y. Tang for his 3D rendering program. This work was partially supported by the Hundred Talents Programs of the Chinese Academy of Sciences, the Natural Science Foundation of China (Grants 60172056 and 697908001), State Commission of Science and Technology of China (Grant G1998030503), and the Sino-French Joint Project (LIAMA).

References

Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., Moriarty, T., 2002. A Modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans. Med. Imaging 21, 193–199.

Arnold, J.B., Liow, J.-S., Schaper, K.A., et al., 2001. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. NeuroImage 13, 931–934.

Axel, L., Costantini, J., Listerud, J., 1987. Intensity correction in surface-coil MR imaging. Am. J. Roentgenol. 148, 418–420.

Clarke, L.P., Velthuizen, R.P., Camacho, M.A., Heini, J.J., Vaidyanathan, M., Hall, L.O., Thatcher, R.W., Silbiger, M.L., 1995. MRI Segmentation: methods and applications. Magn. Reson. Imaging 13, 343–368.

Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imaging 17, 463–468.

Dawant, B.M., Zijidenbos, P., Margolin, R.A., 1993. Correction of intensity variations in MR images for computer-aided tissue classification. IEEE Trans. Med. Imaging 12, 770–781.

Goldszal, A.F., Davatzikos, C., Pham, D.L., Yah, M.X.H., Bryan, R.N., Resnick, S.M., 1998. An image processing system for qualitative and quantitative volumetric analysis of brain images. J. Comput. Assisted Tomogr 22, 827–837.

Guillemaud, R., Brady, M., 1997. Estimating the bias field of MR images. IEEE Trans. Med. Imaging 16, 238–251.

Hall, L.O., Bensaid, A.M., Clarke, L.P., Velthuizen, R.P., Silbiger, M.S., Bezdek, J.C., 1992. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. IEEE Trans. Neural Netw 3, 672–682.

Hornad, J.P. 1996. The Basics of MRI, http://www.cis.rit.edu/htbooks/mri/inside.htm.

Liang, Z., Jaszczak, R., Coleman, E., 1992. Parameter estimation of finite mixtures using the EM algorithm and information criteria with application to medical image processing. IEEE Trans. Nucl. Sci 39, 1126–1133.

Likar, B., Viergever, M.A., Pernus, F., 2000. Retrospective shading correction based on entropy minimization. J. Microsc 197, 285–295.

Meyer, C.R., Peyton, H.B., Pipe, J., 1995. Retrospective correction of intensity inhomogeneities in MRI. IEEE Trans. Med. Imaging 14, 36–41.

Nocera, L., Gee, J.C., 1997. Robust partial volume tissue classification of cerebral MRI scans. Proc. SPIE: Med. Imaging 1997 3034, 312–322.

Pham, D.L., Prince, J.L., 1999. Adaptive fuzzy segmentation of magnetic resonance images. IEEE Trans. Med. Imaging 18, 737–752.

Rajapakse, J.C., Kruggel, F., 1998. Sementation of MR images with intensity inhomogeneities. Image Vision Comput 16, 165–180.

Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13, 856–876.

Simmons, A., Tofts, P.S., Barker, G.J., Arridge, S.R., 1994. Sources of intensity nonuniformity in spin echo images at 1.5T. Magn. Res. Med 32, 121–128.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.

Suri, J.S., Setarehdan, S.K., Singh, S., 2002. Advanced Algorithmic Approaches to Medical Image Segmentation: State-of-the-Art Applications in Cardiology, Neurology, Mammography and Pathology. Spring-Verlag, London/Berlin/Heidelberg.

Tamraz, J.C., Comair, Y.G., 2000. Atlas of Regional Anatomy of the Brain Using MRI with Functional Correlations. Springer-Verlag, Berlin/Heidelberg.

Wells, W.M., Grimson, W, 1996. Adaptive segmentation of MRI data. IEEE Trans. Med. Imaging 15, 429–442.

Worth, A.J., Markis, N., Caviness, V.S., Kennedy, D.N., et al., 1997. Neuroanatomical segmentation in MRI: technological objectives. Int. J. Pattern Recog. Artific. Intell 11, 1161–1187.