

Novel statistical approach for segmentation of brain magnetic resonance imaging using an improved expectation maximization algorithm

YONG YANG¹, SHUYING HUANG²

¹School of Information Management, Jiangxi University of Finance and Economics,
Nanchang 330013, P.R. China

²School of Electronics, Jiangxi University of Finance and Economics,
Nanchang 330013, P.R. China

In this paper, an improved expectation maximization (EM) algorithm called statistical histogram based expectation maximization (SHEM) algorithm is presented. The algorithm is put forward to overcome the drawback of standard EM algorithm, which is extremely computationally expensive for calculating the maximum likelihood (ML) parameters in the statistical segmentation. Combining the SHEM algorithm and the connected threshold region-growing algorithm that is used to provide *a priori* knowledge, a novel statistical approach for segmentation of brain magnetic resonance (MR) image data is thus proposed. The performance of our SHEM based method is compared with those of the EM based method and the commonly applied fuzzy *C*-means (FCM) segmentation. Experimental results show the proposed approach to be effective, robust and significantly faster than the conventional EM based method.

Keywords: expectation maximization, maximum likelihood, magnetic resonance imaging (MRI), fuzzy *C*-means (FCM).

1. Introduction

High spatial resolution and excellent discrimination of soft tissues are the advantages that make magnetic resonance imaging (MRI) offer more accurate anatomical information than other imaging modalities [1, 2]. So, MR images are widely used not only for detecting tissue deformities such as cancers and injuries, but also for studying brain pathology [3]. In order to offer useful and accurate clinical information, the segmentation and recognition algorithms of MR images are becoming an important subject of the study on medical image processing. Many methods have been reported for segmentation of MRI in the literature. CLARKE *et al.* [4] gave an early survey of MRI segmentation, and divided the techniques into the following groups: threshold-based segmentation, statistical methods and region growing methods.

RAJAPAKSE *et al.* [5] proposed a more exact summarization: the available methods for MR image segmentation can be categorized into classical, statistical, fuzzy, and neural network techniques. In this paper, we take the recently most frequently used statistical approach as has been done by many others for MR image segmentation [1–3, 5–7].

Most previously reported statistical approaches used the expectation maximization (EM) [8] algorithm to compute the maximum likelihood (ML) estimation of the segmentation parameters. However, since the EM is an iterative algorithm, it always meets the problem of slow convergence [6] or painfully slow computing [9], which influences its practical clinical applications. To overcome the above problem, an improved EM algorithm named statistical histogram based expectation maximization (SHEM) algorithm is presented in this paper. Using this SHEM algorithm, a novel statistical method is then proposed for segmentation of the single-channel brain MR image data instead of the conventional EM based method. The method involves three steps. Firstly, after pre-processing the image with the curvature anisotropic diffusion filter, the background (BG) and brain masks of the image are obtained by applying a combination approach of thresholding with morphology. Secondly, the connected threshold region growing technique is employed to get the preliminary results of white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) on a brain MRI. Finally, the previous results serve as *a priori* knowledge for the improved EM algorithm to segment the brain MRI.

The paper is organized as follows. Section 2 describes our methods for getting the preliminary results of the brain MRI. The standard EM algorithm is briefly reviewed in Section 3 and the SHEM algorithm is described in Section 4. The experimental results are discussed in Section 5. Finally, some conclusions are drawn in Section 6.

2. Methods of obtaining preliminary results

2.1. Curvature anisotropic diffusion filtering

Acquired medical images are often degraded by various types of artifacts resulting in the lowering of signal-to-noise ratio (SNR) or contrast-to-noise ratio (CNR) [10]. The small SNR or the low CNR make the detection of the anatomical structure difficult. Moreover, image artifacts may affect many image processing tasks such as segmentation, registration, and visual rendition, which are crucial in many applications. Therefore, noise reduction is very important in many imaging applications. The conventional filtering techniques such as mean and median filtering, along with reducing the noise, often blur important structures such as boundaries and detailed structures. Many authors have used the nonlinear anisotropic diffusion [10] and curvature anisotropic diffusion [11] for medical images; the results show that these methods can preserve boundary sharpness and fine details while suppressing noise and enhancing SNR or CNR.

The modified curvature diffusion equation (MCDE) is employed in this subsection. This method has been proved more aggressive than ordinary anisotropic diffusion at

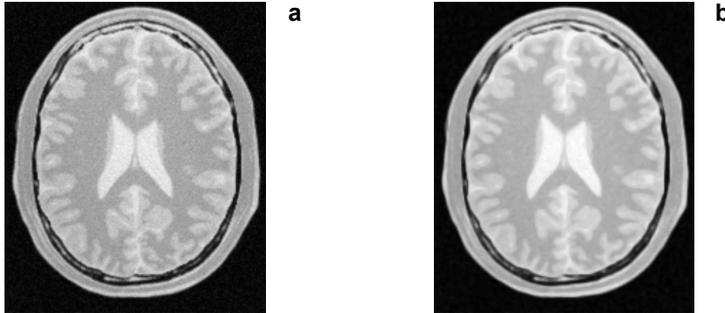


Fig. 1. Results of applying the MCDE filter; original PD-weighted image (a), image after filtering (b).

enhancing and preserving edges and detailed structures. The MCDE equation is given as:

$$f_t = |\nabla f| \nabla \cdot c |\nabla f| \frac{\nabla f}{|\nabla f|} \quad (1)$$

where $f = f(x, y, t)$ and $f(x, y, 0) = I(x, y)$ – the input image, c is called conductance function and is a monotonically decreasing function containing a free parameter k , which determines the contrast of edges that will significantly affect the smoothing. Figure 1 illustrates the effect of this filter on a MRI proton density (PD) weighted image of the brain from the digital brain phantom [12]. In this example, the filter was run with a time step of 0.125, 5 iterations and a conductance value of 1.0. Figure 1b shows how homogeneous regions are smoothed and edges are preserved.

2.2. Generating background and brain masks

After the image has been filtered, an initial segmentation into foreground/background is achieved using simple intensity thresholding, thus a BG mask binary image is produced. Then we segment the inside of the brain from non-brain tissues and remove small connections between the brain and surrounding tissues. The segmentation

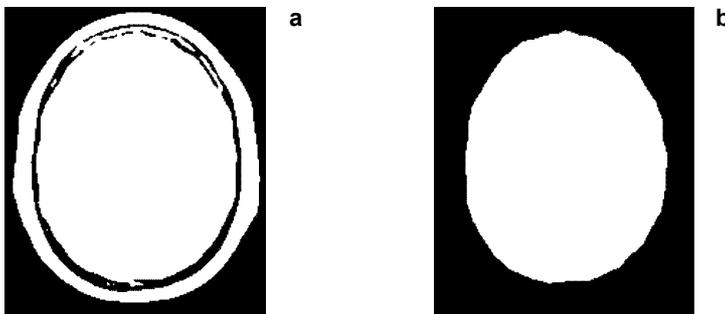


Fig. 2. Cerebral region mask produced using automatic thresholding and morphology; BG mask (a), brain mask (b).

method used here is thresholding with morphology. The brain region is almost always connected, often via fairly thin strands of bright pixels to non-brain tissues such as the bone, cortex, and the fat tissues. Hence, before getting the binary brain mask, it must be disconnected from the non-brain bright tissues. This is normally achieved by morphological filtering, the bright regions in the binary image are being eroded away until any links between brain and non-brain are eliminated, the largest single region is then chosen, and it is next dilated to the same extent as the erosion, hopefully resulting in an accurate brain mask.

We use an automated method as described in [3] to find the binary threshold; the result of BG mask is shown in Fig. 2a. A morphological erosion operation is then applied with a 5×5 rectangular structural element. After erosion, a labeling algorithm is used to find the largest single region. Finally, binary dilation with the same 5×5 kernel as for erosion, is performed on the remaining region to make it close to the original size. The final brain mask is shown in Fig. 2b.

2.3. Connected threshold region growing

Region growing (often called flood-fill) algorithms have been proven to be an effective approach for image segmentation. The basic approach of a region-growing algorithm is to start from a seed region (typically one or more pixels) that is considered to be inside the object to be segmented. The pixels neighboring this region are evaluated to determine whether they should also be considered to be the part of the object by homogeneity criteria. If so, they are added to the region and the process continues as long as new pixels are being added to the region.

A simple region growing method is introduced here, namely connected threshold. The criterion used by the connected threshold is based on an interval of intensity values. Values of lower and upper threshold should be provided. The region-growing algorithm includes those pixels whose intensities are inside the interval,

$$I(x, y) \in [\text{lower}, \text{upper}] \quad (2)$$

The problem is the definition of these two intervals. We do this by statistical evaluation about the gray value distribution. Then, we can easily segment the major

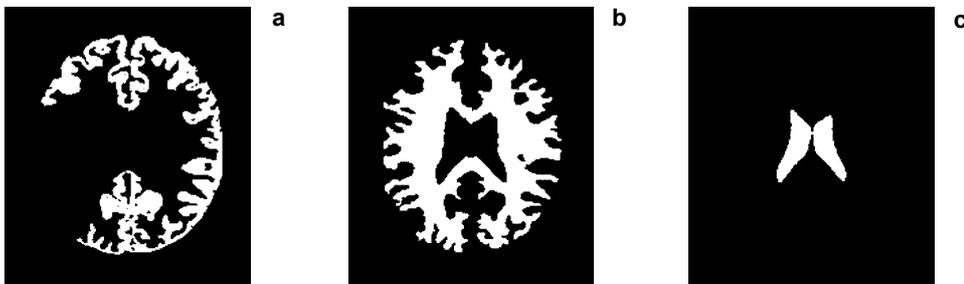


Fig. 3. Results of the connected threshold region growing; GM mask (a), WM mask (b), CSF mask (c).

anatomical structures by providing seeds in the appropriate locations. After binary thresholding, we get three clusters' masks. The results of the experiment are shown in Fig. 3. From Figs. 3a and 3c, we notice that the GM and CSF are not being completely segmented. This illustrates the vulnerability of the region growing methods. However, these incomplete segmentation masks can be used as *a priori* ones for the following SHEM or EM algorithm.

3. Standard expectation maximization algorithm

The standard EM algorithm is a general algorithm for ML estimation where the data are “incomplete” or the likelihood function involves latent variables. This algorithm [8] considers the observed variables y as the “incomplete data”, x as the “hidden data” and the couple (x, y) as the “complete data” characterized by the joint distribution $p(x, y|\phi)$, where ϕ is a parameter vector to be estimated. The purpose is to find $\hat{\phi}$ which maximizes the likelihood of observed data $p(y|\phi)$

$$\hat{\phi} = \arg \max \log p(y|\phi) \quad (3)$$

The EM algorithm starts with randomly assigning values or some prior knowledge to all the parameters to be estimated. It then iteratively alternates between two steps, called the expectation step (*i.e.*, the *E*-step) and the maximization step (*i.e.*, the *M*-step), respectively. In the *E*-step, given the current settings of parameters and our incomplete data, the expected log joint likelihood for the complete data (*i.e.*, *Q*-function) is computed, where the expectation is taken with respect to the computed conditional distribution of the “hidden data”. In the *M*-step, all the parameters are re-estimated by maximizing the *Q*-function. Once we have a new generation of parameter values, we can repeat the *E*-step and another *M*-step. This process continues until the likelihood converges, *i.e.*, reaching local maxima. The procedure is iterative and repeats the two following steps until convergence: *E*-step – find the function

$$Q(\phi|\phi^{(b)}) = E \left[\log p(X, Y|\phi) \middle| y, \phi^{(b)} \right] \quad (4)$$

and *M*-step – find

$$\phi^{(b+1)} = \arg \max Q(\phi|\phi^{(b)}) \quad (5)$$

with b being the iteration number.

4. Statistical histogram based expectation maximization algorithm

As proposed in the publications (*e.g.*, [1, 5]), we assume here that the brain image intensity corresponding to a tissue can be well modeled as a multivariate Gaussian distribution,

$$p(x_t | \theta_k) = (2\pi)^{-M/2} |\Sigma_k|^{-1/2} \cdot \exp \left[-\frac{1}{2} (x_t - \mu_k)^T \Sigma_k^{-1} (x_t - \mu_k) \right] \quad (6)$$

where $\theta_k = (\mu_k, \Sigma_k)$ is the vector of parameter associated with each type of tissue (or class) k , μ_k is the mean vector, and Σ_k is the covariance (positive definite symmetric) matrix associated with class k ($1 \leq k \leq K$, where K is the number of classes), M is the number of channels or spectra in the image, and T denotes matrix transpose. In this paper, we consider only a single MR image of the object; such an image is referred to as single-channel image (*i.e.*, $M = 1$). The model of (6) can then take the form:

$$p(x_t | \theta_k) = \frac{1}{\sqrt{2\pi} \sigma_k} \exp \left[-\frac{1}{2\sigma_k^2} (x_t - \mu_k)^2 \right] \quad (7)$$

where σ_k is the standard deviation of class k . The brain is a mixture of different tissues, which are assumed mutually independent. With these assumptions, the likelihood of the image data can be written as:

$$L(\phi) = \prod_{t=1}^n \sum_{k=1}^K w_k p(x_t | \theta_k) \quad (8)$$

where $\phi = \{\mu_k, \sigma_k^2, w_k\}$ for $k = 1, 2, \dots, K$, n is the total number of the image pixels and w_k is the proportion of each tissue component, where $\sum_{k=1}^K w_k = 1$ and $w_k \geq 0$. The log-likelihood can then be expressed by:

$$\log L(\phi) = \sum_{t=1}^n \log \sum_{k=1}^K w_k p(x_t | \theta_k) \quad (9)$$

Many numerical techniques have been proposed to perform the ML estimation of the above class parameters, among which EM algorithm is the most commonly used method as many authors have reported [5–8]. The EM algorithm used above is based on the intensity of the image, which counts the parameters pixel-by-pixel, and as a result, the convergence of the iteration is slow, with more computational time being needed. In this section, we use the statistical histogram of the image to overcome the problems.

Define the non-negative integrate set $G = \{L_{\min}, L_{\min+1}, \dots, L_{\max}\}$ as gray level, where L_{\min} is the minimum gray level, L_{\max} is the maximum gray level, so the grayscale is $L_{\max} - L_{\min}$. For image size $U \times V$, at point (u, v) , $f(u, v)$ is the gray level with $0 \leq u \leq U - 1$, $0 \leq v \leq V - 1$. We use $\text{His}(g)$ to denote the number of pixels having gray level g , $g \in G$. The statistical histogram function is as follows:

$$\text{His}(g) = \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} \delta(f(u, v) - g) \quad (10)$$

where $g = \{L_{\min}, L_{\min+1}, \dots, L_{\max}\}$, $\delta(0) = 1$ and $\delta(g \neq 0) = 0$.

Let i be the intensity of the pixel with $L_{\min} \leq i \leq L_{\max}$, and all the pixels of the k -th tissue cluster have a mean intensity μ_k , variance σ_k^2 , and proportional ratio w_k . The K mixed Gaussian distribution can be written as:

$$p(i|\phi) = \sum_{k=1}^K w_k p(i|\theta_k) \quad (11)$$

where $\sum_{k=1}^K w_k = 1$ and

$$p(i|\theta_k) = \frac{1}{\sqrt{2\pi} \sigma_k} \exp\left[-\frac{1}{2\sigma_k^2}(i - \mu_k)^2\right] \quad (12)$$

The above parameters can be obtained by equating the first partial derivatives of Eq. (9) with respect to unknown parameters to zero. With the statistical histogram, the SHEM algorithm can then be expressed by:

1. The E -step

$$\psi_{ik}^{(b+1)} = \frac{p(i|\theta_k^{(b)}) \cdot w_k^{(b)}}{\sum_{k=1}^K p(i|\theta_k^{(b)}) \cdot w_k^{(b)}} \quad (13)$$

where ψ_{ik} is the posterior probability that intensity i belongs to class k .

2. The M -step. The second step updates the unknown parameters with the statistical histogram $\text{His}(i)$

$$T_k^{(b+1)} = \sum_{i=L_{\min}}^{L_{\max}} \psi_{ik}^{(b+1)} \cdot \text{His}(i) \quad (14)$$

$$w_k^{(b+1)} = \frac{T_k^{(b+1)}}{U \cdot V} \quad (15)$$

$$\mu_k^{(b+1)} = \frac{\sum_{i=L_{\min}}^{L_{\max}} \psi_{ik}^{(b+1)} \cdot i \cdot \text{His}(i)}{T_k^{(b+1)}} \quad (16)$$

$$\left[\sigma_k^{(b+1)} \right]^2 = \frac{\sum_{i=L_{\min}}^{L_{\max}} \psi_{ik}^{(b+1)} \cdot \left(i - \mu_k^{(b+1)} \right)^2 \cdot \text{His}(i)}{T_k^{(b+1)}} \quad (17)$$

where b is the iteration number. The procedure of the SHEM algorithm is the following:

1. Initialize $\phi^{(0)}$ randomly or according to some prior knowledge of where might be the optimal parameters;
2. Iteratively improve the estimate of ϕ by alternating between the E - and M -step;
3. The iteration will stop when the number of iteration $> N$ or the convergence is stable.

5. Experimental results

In this section, both standard EM and SHEM algorithms are used for the segmentation of the filtered PD-weighted MRI. The segmentation was implemented in VC++6.0 language on a PC. We attempt to segment the MRI into four clusters (GM, WM, CSF, and BG), and apply the corresponding previous segmentation mask results to compute the initial value of $\phi^{(0)}$. The algorithm is terminated after the convergence has been stable. The final results are gained by extracting the brain region from the results of SHEM algorithm. The segmentation results are shown in Fig. 4. Compared with the connected threshold region growing results in Figs. 3a and 3c, GM and CSF are extracted completely and accurately, as shown in Figs. 4a and 4c.

In the experiment, we see both EM and SHEM to have spent the same time performing iterations to accomplish the segmentation process and get the same results; however, the proposed SHEM algorithm consumes less time than the standard EM algorithm. In this example, the standard EM algorithm spends 0.64 seconds in each iteration, while the corresponding time for SHEM algorithm is nearly 0.004 seconds. In this case, the SHEM algorithm converged approximately 160 times faster than the EM algorithm. Because most brain MR scans consist of more than 100 2D slices, the proposed SHEM algorithm can save significantly large amount of computational time. Therefore, the clinical information can be provided more quickly than that of standard EM algorithm.

To test the performance of the proposed approach, we compared the method with the popular FCM algorithm [13, 14] to segment the MRI. The algorithm is implemented with the total number classes $C = 4$ and the weighting exponent $m = 2$. After convergence, the maximum membership segmentation is applied to each pixel of the image. The GM, WM and CSF results of FCM segmentation are shown in Fig. 5. Compared to the corresponding results in Fig. 4, it can be seen by FCM algorithm that CSF is over-segmented, WM is under-segmented, and GM on the top and bottom is somewhat under-segmented while in some other place over-segmented.

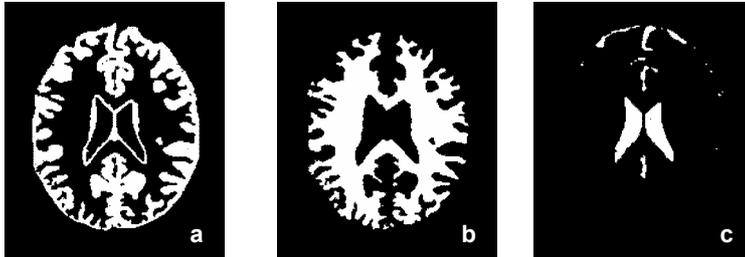


Fig. 4. Results of SHEM algorithm – GM (a), WM (b), CSF (c).

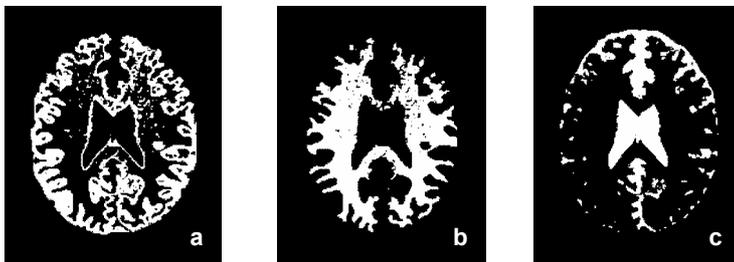


Fig. 5. Results of FCM segmentation – GM (a), WM (b), CSF (c).

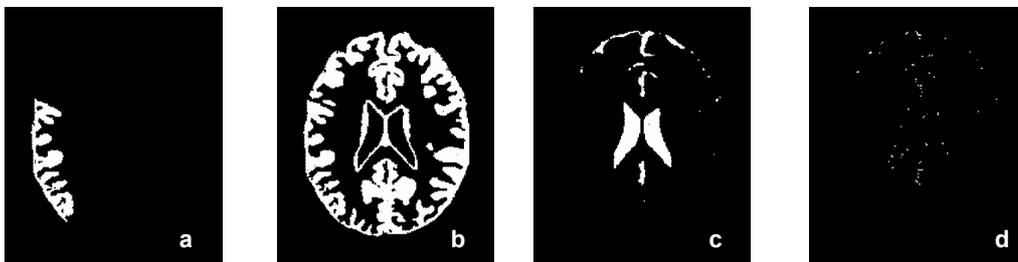


Fig. 6. Results with different seed points – GM mask (a), GM (b), CSF (c), difference of GM or CSF (d).

In addition, in the experiment, we find that when FCM algorithm has been badly initialized, it converges to wrong maximas and cannot get the four true clusters. However, it is noted that even if the initialization quits differently, the proposed method still can get identical results. When different groups of seed points are tested, the same masks are always obtained except for the GM mask. The final results of GM and CSF are a little different, while WM is identical compared to the corresponding results of Fig. 4. The difference is computed by subtracting the GM and CSF from the corresponding components in Fig. 4. Through subtraction, it is noted that the difference of GM and CSF is either very small or none, which means that they are misclassified to each other so that the total number of the brain regions is unchanged. Series of different seed points are tested and here just a group of results are listed

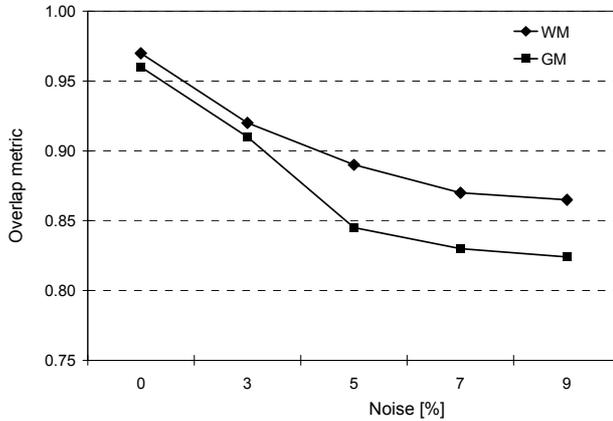


Fig. 7. Overlap metrics with different noise levels for SHEM based segmentation on T1-weighted MR phantoms.

in Fig. 6. In this example, the result of WM is identical to that of Fig. 4b. Figure 6a shows the different GM mask compared to Fig. 3a. Figures 6b and 6c show the results of GM and CSF, respectively, and Fig. 6d shows the difference of GM or CSF. The results given here do show the proposed method to be less sensitive to initialization and robust.

Finally, to further quantitatively evaluate the performance of the algorithm, our method is realized to segment the digital MR phantoms [12] with different noise levels. There are many advantages of using digital phantoms rather than real image data for validating segmentation methods. These advantages include prior knowledge of the true tissue type and control over image parameters such as modality, slice thickness, noise and intensity inhomogeneities. Here, in our experiments, we use the high-resolution T1-weighted MR phantoms with slice thickness of 1 mm, no intensity inhomogeneities and 0–9% noises. To measure the segmentation accuracy, the overlap metric is utilized as the criteria [15]. The overlap metric is a measure for comparing two segmentations, which is defined for a given class assignment as the sum of the number of pixels, so that both have the class assignment in each segmentation divided by the sum of pixels where either segmentation has the class assignment. Larger metric means more similar results. Figure 7 gives the overlap metrics of WM and GW. As the level of noise increases, the overlap metric of our algorithm gradually degrades, which is because no spatial information is incorporated into the algorithm. However, it is important to note that at 0% noise level, the overlap metrics of both WM and GM are higher than 0.95; even at 3% noise level, the corresponding over-metrics are still higher than 0.90. The results presented here can prove that our method is effective and allows us to obtain correct segmentation results at low noise level.

6. Conclusions

In this paper, a novel statistical approach based on an improved EM algorithm called SHEM algorithm for segmentation of the single-channel brain MRI is proposed and tested. After a preliminary processing, the four (GM, WM, CSF, BG) cluster masks are extracted, which serve as *a priori* knowledge for the SHEM algorithm. The tissue regions of the image are satisfactorily segmented in this way, which demonstrates that the method is effective. We compared our results with those of standard EM algorithm and FCM segmentation. The SHEM algorithm produces identical results as the EM algorithm with faster convergence. The SHEM based statistical segmentation outperformed the FCM segmentation in both the effectivity and robustness to initialization. Future work will focus on combining the spatial context into the algorithm in order to improve its robustness to noise and compensating for the intensity inhomogeneities during segmentation.

References

- [1] WELLS W.M., GRIMSON W.E.L., KIKINS R., JOLESZ F.A., *Adaptive segmentation of MRI data*, IEEE Transactions on Medical Imaging **15**(4), 1996, pp. 429–42.
- [2] HELD K., KOPS E.R., KRAUSE B.J., WELLS W.M., KIKINS R., MULLER-GARTNER H.W., *Markov random field segmentation of brain MR images*, IEEE Transactions on Medical Imaging **16**(6), 1997, pp. 878–86.
- [3] ATKINS M.S., MACKIEWICH B.T., *Fully automatic segmentation of the brain in MRI*, IEEE Transactions on Medical Imaging **17**(1), 1998, pp. 98–107.
- [4] CLARKE L.P., VELTHUIZEN R.P., CAMACHO M.A., HEINE J.J., VAIDYANATHAN M., HALL L.O., THATCHER R.W., SILBINGER M.L., *MRI segmentation: methods and applications*, Magnetic Resonance Imaging **13**(3), 1995, pp. 343–68.
- [5] RAJAPAKSE J.C., GIEDD J.N., RAPOPORT J.L., *Statistical approach to segmentation of single-channel cerebral MR images*, IEEE Transactions on Medical Imaging **16**(2), 1997, pp. 176–186.
- [6] HASHIMOTO A., KUDO H., *Ordered-subsets EM algorithm for image segmentation with application to brain MRI*, [In] *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, Vol. 3, 2000, pp. 18/118–21.
- [7] LIANG Z., MACFALL J.R., HARRINGTON D.P., *Parameter estimation and tissue segmentation from multispectral MR images*, IEEE Transactions on Medical Imaging **13**(3), 1994, pp. 441–9.
- [8] DEMPSTER A.P., LAIRD N.M., RUBIN D.B., *Maximum-likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society **39**, 1977, pp. 1–38.
- [9] LAFERTE J.M., HEITZ F., PEREZ P., *A multiresolution EM algorithm for unsupervised image classification*, [In] *Proceedings of the 13th International Conference on Pattern Recognition*, Vol. 2, 1996, pp. 849–53.
- [10] GERIG G., KÜBLER O., KIKINS R., JOLESZ F.A., *Nonlinear anisotropic filtering of MRI data*, IEEE Transactions on Medical Imaging **11**(2), 1992, pp. 221–32.
- [11] WHITAKER R.T., XUE X.W., *Variable-conductance, level-set curvature for image denoising*, [In] *Proceedings of the International Conference on Image Processing*, Vol. 3, 2001, pp. 142–5.
- [12] COLLINS D.L., ZIJDENBOS A.P., KOLLOKIAN V., SLED J.G., KABANI N.J., HOLMES C.J., EVANS A.C., *Design and construction of a realistic digital brain phantom*, IEEE Transactions on Medical Imaging **17**(3), 1998, pp. 463–8.

- [13] BEZDEK J., *Pattern Recognition with Fuzzy Objective Functions Algorithms*, Plenum Press, New York 1981.
- [14] HALL L., BENSALD A.M., CLARKE L., VELTHUIZEN R.P., SILBINGER M.S., BEZDEK J.C., *A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain*, IEEE Transactions on Neural Networks **3**(5), 1992, pp. 672–82.
- [15] <http://www.bic.mni.mcgill.ca/brainweb/>

Received March 14, 2005