Proceedings of the 26th Annual International Conference of the IEEE EMBS
San Francisco, CA, USA • September 1-5, 2004

Comparison Study of Clinical 3D MRI Brain Segmentation Evaluation

Ting Song[1], Elsa D. Angelini[2], Brett D. Mensh[3], Andrew Laine[1]
[1]Department of Biomedical Engineering, Columbia University, NY, USA
[2] Department of Signal and Image Processing, Ecole Nationale Supérieure des Télécommunications, Paris, France
[3] Department of Biological Psychiatry, College of Physicians and Surgeons, Columbia University, NY, USA

Abstract—**Although numerous methods to segment brain MRI for extraction of white matter, gray matter and cerebro-spinal fluid (CSF) have been proposed for the past two decades, little work has been done to evaluate and compare the performance of different segmentation methods on real clinical data sets, especially for CSF. This study focuses on the comparison of the four following methods for segmentation of cerebral brain MRI: gray levels thresholding, three-dimensional level set, fuzzy connectedness and FSL. Quantitative evaluation of segmentation accuracy was performed with comparison to manual segmentation on a database of 10 adult subjects.**

Keywords—**brain segmentation, evaluation, MRI.**

I. Introduction

Segmentation of clinical brain MRI data is an essential objective in medicine. In the context of neuro-imaging, 3D segmentation of white mater (WM), gray matter (GM), and cerebro-spinal fluid (CSF) is critical for quantitative analysis such as volume measurements, shape characterization and functional brain mapping.

An tremendous number of authors have studied this segmentation problem deriving methods that can be divided into four groups: simple thresholding methods, deformable models [1], [2], fuzzy connectedness [3], and statistical method [4], [5].

The main difficulty of automatic segmentation of clinical MRI brain images is that voxel intensities are not necessarily constant for each tissue class and the histograms of each tissue show great overlaps. Furthermore, noise artifacts in clinical data generate a fragmented segmented structure, which is not observed in simulated MRI data. Inter-subjects problem exists in clinical cases, but not in simulated data. Finally, CSF segmentation results are rarely compared with previous works.

In the context of this paper, we evaluated four segmentation frameworks for the extraction of the three cerebral brain tissues: WM, GM and CSF.

II. METHODOLOGY

Four methods have been evaluated in this study. These methods are not detailed in this paper and we refer readers to appropriate references.

1) ***Simple threshold (Method A)*****:** This is the easiest and fastest segmentation method, which is often adopted in preprocessing of medical images and pre-registration problems [6]. In order to optimize the performance of this method, we used manually labeled data to mask the MRI data, compute the histogram of the three tissues and derive the initial two threshold values that separate the tissues gray levels bins. We then used the Simplex method [7] to minimize the segmentation error. The criterion of the stability measurement maximized the Tanimoto Index [8].

2) ***Three-dimensional Level Set(Method B)***: A homo geneity-based 3D deformable model derived from the Mu mford-Shah functional was proposed by Chan and Vese [9] ,. In this framework the image is assumed to be formed by t wo regions of approximate piecewise constant intensities, o f distinct values. Given an initial curve inside the image, th e contour is deformed to minimize an energy functional tha t corresponds to the optimal portioning of the image into h omogeneous regions. Minimization of the energy functiona l is efficiently implemented with the level set method. In th is work we have implemented this method in 3D and exten ded the initial framework to four homogeneous objects, usi ng simultaneously two level set functions as described in [2].

3) ***Fuzzy Connectedness (Method C)*****:** A fuzzy object classification for a three-dimensional digital space based on fuzzy connectedness was implemented as method C [3]. This method can be used to extract fuzzy objects in the image given an initial set of seed points and prior statistics. The source code can be obtained from National Library of Medicine Insight Segmentation and Registration Toolkit (ITK) [10]. In order to optimize the threshold values which required for the construction of fuzzy maps, we also use the Simplex scheme as discussed in method A.

4) ***Hidden Markov Random Field Model and Expectation-Maximization (HMRF-EM) (Method D)*****:** This segmentation method is based on statistical classification of the image into different gray level classes. This algorithm starts with an initial estimation of the tissue class parameters and repartition in the image.

The initial segmentation is followed by a three-step expectation-maximization (EM) process which updates the class labels, tissue parameters and bias field iteratively. During the iterations, a Markov random field (MRF) maximum *a posteriori* (MAP) approach is used to estimate class labels. The bias filed is estimated by MAP approximation and the tissue parameters are estimated by maximum likelihood (ML) [4]. This segmentation method was used using the FMRIB Software Library (FSL) [11].

0-7803-8439-3/04/$20.00©2004 IEEE

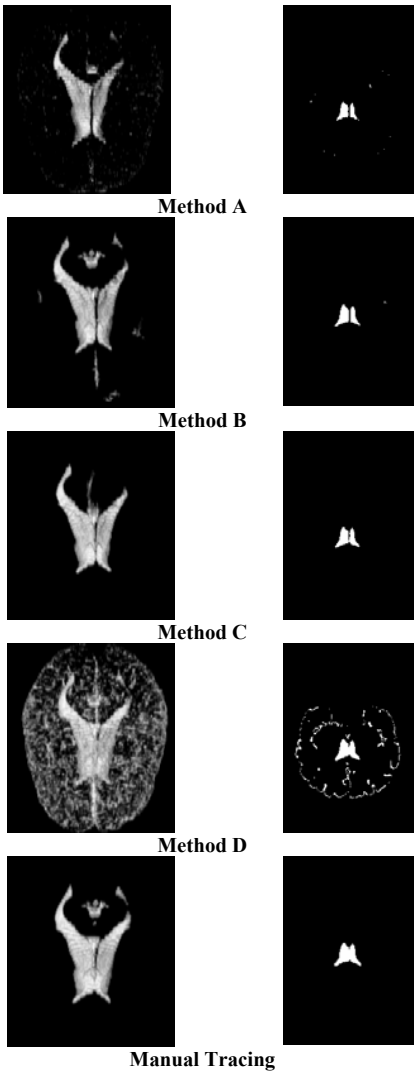


Figure 1.CSF segmentation results are compared with manually labeled data. Column one is the 3D view of the CSF, and column two is the coronal view of slice number 43.

III. RESULTS

A. Data and Evaluation Protocol

We applied our segmentation to ten T1-weighted MRI data sets acquired on healthy young volunteers. The MRI data sets were of size (256×256×73) with a 3mm slice thickness and 0.86mm in-plane resolution. These data sets have been previously labeled via a labor-intensive (40 hours per brain) manual protocol. This labeled data was used as a ground truth for evaluation of the segmentation accuracy. MRI data was pre-processed to remove all non-brain tissue, using the manually labeled data sets as binary masks. Since subcortical structure can affect the overall evaluation of the segmentation accuracy of cortical structures, we also excluded them after the segmentation process.

To be fair and objective, the evaluation protocol was designed to be independent of the segmentation algorithms. Measurements of volume, true positive and false positive voxel fractions and the Tanimoto Index (TI) [8] were calculated for gray matter, white matter and CSF for each method. Analysis of variance (ANOVA) was performed to evaluate the difference between the different methods. For each method, a three-dimensional visualization, in coronal views, of the lateral ventricles filled with CSF are shown in Figure 1.

B. Volume Evaluation

Visualization of the results showed significant differences between Method D (HMRF from FSL) and the other methods, classifying the outer border voxels as belonging to sulcal CSF versus gray matter. This classification difference can be explained by the important partial volume effect at the interface between the two tissues. A priori information and arbitrary decision rules can lead the segmentation algorithm to assign these interface voxels to a tissue type or another.

Correlation of volume measurements between the manually labeled data and the four segmentation methods was performed for each tissue type as illustrated in Figure 2. We observed in these plots that GM was over segmented in several cases by method C. In general, method A and method B displayed high correlation with the manually labeled data.

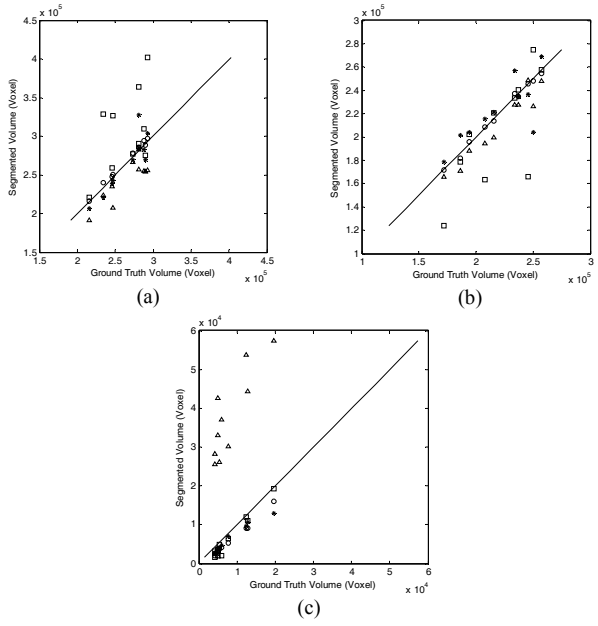


Figure 2. Correlation of volume measurements for the (a) GM, (b) WM and (c) CSF structures between the segmented volumes and the manual tracing. o is method A; * is method B; □ is method C; Δ is method D.

C. Accuracy Evaluation

Segmentation errors were measured using the recent methodology proposed by Udupa in [12] for comparison of segmentation methods. Accuracy of the object contours obtained with the proposed level set segmentation method was evaluated by comparing segmentation results to the manual ground truth. Overlap and difference between two contours were measured by counting true positive (TP) and false positive (FP) voxels. These quantities are reported as volume fractions (VF) of the ground truth delineated object volume. Error measurements for the segmentation of the three clinical cases are reported in Table I, Table II, Table III and Table IV.

In Table I, percentage measurements show excellent performance of the optimized threshold method for the GM and WM. The FN percentage for the CSF is much higher, reflecting and under segmentation of this structure, based on a wider gray scale histogram. The level set method, in Table II reports better results for the CSF segmentation, with remaining under segmentation issues. Fuzzy connectedness method results, reported in Table III, show under segmentation issues for the CSF as well as over segmentation of the WM with a high FP percentage. Finally, results for the HMRF-EM segmentation method are reported in Table IV. High FP percentage values for the CSF, correspond to an over segmentation of the sulcal CSF due to the misclassification of CSF/GM interface voxels associated to the GM by the other segmentation methods. On the other hand, the TP percentage of the CSF is higher with this method than with any other one, which illustrates the fact that the ventricles are properly segmented.

Table I: Segmentation accuracy with volume fractions of segmented voxels for the optimized thresholding method.

	GM			WM			CSF		
	FN	FP	TP	FN	FP	TP	FN	FP	TP
1	5.8	7.2	94.2	6.5	6.3	93.5	40.7	11.6	59.3
2	5.1	6.5	94.9	6.6	6.6	93.4	37.6	10.5	62.4
3	4.8	6.4	95.2	6.3	5.1	93.7	53.4	22.4	46.6
4	6.3	6.6	93.7	5.9	6.6	94.1	53.8	14.1	46.2
5	4.0	5.7	96.0	6.0	4.8	94.0	66.0	12.6	34.0
6	5.1	6.4	94.9	6.1	5.4	93.9	48.4	18.2	51.6
7	5.4	6.1	94.6	5.5	5.9	94.5	45.6	14.0	54.4
8	5.7	7.9	94.3	6.9	5.9	93.1	25.5	6.7	74.5
9	4.2	6.6	95.8	6.9	4.7	93.1	64.6	25.3	35.4
10	7.2	6.8	92.8	6.9	8.1	93.1	71.7	37.2	28.3
μ	**5.4**	**6.6**	**94.6**	**6.4**	**5.9**	**93.6**	**50.7**	**17.3**	**49.3**

Table II: Segmentation accuracy with volume fractions of segmented voxels for the level set method.

	GM			WM			CSF		
	FN	FP	TP	FN	FP	TP	FN	FP	TP
1	4.2	8.1	95.8	8.5	4.8	91.5	21.9	4.8	78.1
2	6.5	5.0	93.5	5.1	8.9	94.9	28.5	6.1	71.5
3	4.5	5.9	95.5	6.3	5.2	93.7	27.5	7.8	72.5
4	8.7	4.5	91.3	4.1	9.2	95.9	36.3	16.2	63.7
5	5.4	3.9	94.6	4.3	6.8	95.7	35.7	1.0	64.3
6	0.5	17.3	99.5	18.9	0.6	81.1	26.5	0.5	73.5
7	6.9	4.0	93.1	4.1	7.9	95.9	20.6	11.3	79.4
8	8.3	6.4	91.7	4.6	9.3	95.4	33.8	0.01	66.2
9	8.9	2.9	91.1	2.9	11.2	97.1	37.3	0.7	62.7
10	11.0	3.9	89.0	3.8	13.5	96.2	48.7	5.3	51.3
μ	**6.5**	**6.2**	**93.5**	**6.3**	**7.7**	**93.7**	**31.7**	**5.4**	**68.3**

Table III: Segmentation accuracy with volume fractions of segmented voxels for the fuzzy connectedness method.

	GM			WM			CSF		
	FN	FP	TP	FN	FP	TP	FN	FP	TP
1	8.1	45.6	91.9	33.2	0.7	66.8	17.3	3.2	82.7
2	10.7	43.0	89.3	28.8	0.6	71.2	19.1	16.2	80.9
3	7.7	36.9	92.3	6.7	7.9	93.3	67.4	1.8	32.6
4	9.2	11.4	90.8	5.8	9.9	94.2	37.9	9.7	62.1
5	7.7	9.3	92.3	6.1	8.4	93.9	70.1	7.5	29.9
6	12.0	15.5	88.0	3.8	13.6	96.2	25.3	14.3	74.7
7	11.5	16.7	88.5	22.3	1.1	77.7	27.1	10.5	72.9
8	7.4	14.9	92.6	7.5	7.8	92.5	9.1	7.5	90.9
9	8.4	48.6	91.6	9.4	5.3	90.6	33.7	10.5	66.3
10	16.6	11.5	83.4	8.9	8.5	91.1	68.9	9.0	31.1
μ	**9.9**	**25.3**	**90.1**	**13.2**	**6.4**	**86.8**	**37.6**	**9.0**	**62.4**

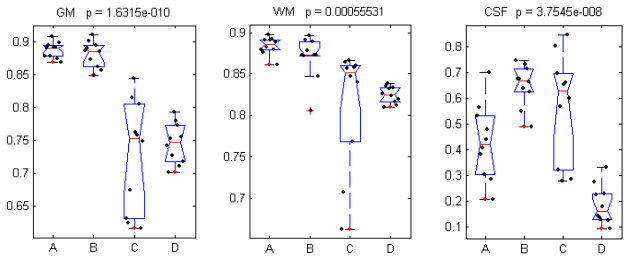
Table IV: Segmentation accuracy with volume fractions of segmented voxels for the HMRF-EM method.

	GM			WM			CSF		
	FN	FP	TP	FN	FP	TP	FN	FP	TP
1	20.2	7.6	79.8	9.0	10.2	91.0	3.8	251.7	96.2
2	24.1	8.3	75.9	12.0	8.6	88.0	2.7	335.1	97.3
3	17.8	9.2	82.2	10.8	6.7	89.2	9.6	538.1	90.4
4	20.6	9.3	79.4	10.3	7.0	89.7	9.0	571.4	91.0
5	12.6	10.2	87.4	12.6	5.0	87.4	18.5	536.3	81.5
6	12.1	12.7	87.9	14.0	4.4	86.0	11.9	396.2	88.1
7	14.8	10.2	85.2	11.9	5.7	88.1	8.7	302.0	91.3
8	21.6	10.4	78.4	11.7	8.2	88.3	2.5	194.4	97.5
9	16.0	11.3	84.0	14.0	5.9	86.0	13.7	600.9	86.3
10	21.6	9.3	78.4	11.1	8.5	88.9	19.9	777.6	80.1
μ	**18.1**	**9.9**	**81.9**	**11.7**	**7.0**	**88.3**	**10.0**	**450.4**	**90.0**

To compare directly the segmentation methods, we performed an analysis of variance (ANOVA) on a characteristic index of their performance. The Tanimoto index (TI) [8], was selected for this study. This index is a quantitative parameter used to evaluate the segmentation results and is defined as:

$$TI = \frac{TP}{1+FP} \tag{1}$$

The analysis of variance was performed for the four methods and over the 10 segmented cases, measuring the inter-method variance divided by the intra-method variance of the TI indexes. Small p values (below 0.005) indicate a significant statistical difference between the methods [13]. Distributions of the TI index over the 10 cases for each method are plotted in Figure 3.



(a) (b) (c)

Figure 3: Distribution of TI values for (a) GM, (b) WM, and (c) CSF over the four segmentation methods (A, B, C, D) for the 10 clinical cases..

Based on this comparison of the methods, all of them

report high TI values for GM and WM, with superior performances of methods A and B. On the other hand, for segmentation of the CSF structures, methods B and C show higher TI values than methods A and D, with a highest variance for method C.

IV. DISCUSSION

Significantly higher FNVF errors were observed for the CSF, corresponding to under segmentation of the ventricles, whose pixels were assigned to white matter. Very low resolution at the ventricle borders can explain in part this result. Sulcal CSF introduce high FPVF to CSF, corresponding to the over segmentation of CSF, whose pixels belong to gray matter.

On the other hand, labeling of the MRI data for the ventricle and sulcal CSF can also bear some error as localizations of its borders is difficult even for an expert performing manual tracing. In that context, Kikinis *et al*. [14] reported a variation in volumetric measurements of manual observers in the order of 15% for WM, GM and CSF.

V. CONCLUSION

This study focused on the comparison of different segmentation methods for clinical brain T1-weighted MRI. Three of the selected methods are commonly used in clinical studies, especially for evaluation of the CSF structures. The second method, based on a multi-phase three-dimensional level set deformable model, was first described for this application by our group in [2]. We compared the segmentation methods through evaluation of the accuracy the segmented volumes for WM, GM and CSF on 10 clinical cases, comparing to manual tracing. From volume visualization and evaluation, methods A and method B showed small errors and strong correlation with manual tracing. Percentage of misclassifications showed that Method C tends to over segment the GM structure in several cases while Method D tends to over segment the sulcal CSF tissue. Combining all the results, we conclude that the level set three-dimensional deformable model provides the best compromise among the four tested segmentation methods between high segmentation accuracy and low variance of performance, when compared to manual tracing. Part of our future work will focus on the elimination of the sulcal CSF which should reduce important misclassification errors for the CSF.

REFERENCES

[1] Xiaolan Zeng, Lawrence H. Staib, Robert T.Schultz, and J. S. Duncan, "Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation," *IEEE Transaction on Image Processing*, vol. 18, pp. 927-937, 1999.

[2] E. D. Angelini, T. Song, B. D. Mensh, and A. Laine, "Segmentation and quantitative evaluation of brain MRI data with a multi-phase three-dimensional implicit deformable mode," presented at SPIE International Symposium, Medical Imaging 2004, San Diego, CA USA, 2004.

[3] J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *Graphical Models and Image Processing*, vol. 58, pp. 246-261, 1996.

[4] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 45-57, 2001.

[5] J. L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas, "An accurate and efficent Bayesian method for automatic segmentation of brain MRI," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 934-945, 2002.

[6] Klaus A. Ganser, Hartmut Dickhaus, Roland Metzner, and C. R. Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Medical Image Analysis*, vol. 8, pp. 3-22, 2004.

[7] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, vol. 9, pp. 112-147, 1998.

[8] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. U.S.A: Academic Press, 1999.

[9] T. Chan and L. A. Vese, "Active contours without edges," *IEEE Transaction on Image Processing*, vol. 10, pp. 266-277, 2001.

[10] "National Libary of Medicine Insight Segmentation and Registration Toolkit (ITK) http://www.itk.org."

[11] "FMRIB Software Library (FSL) http://www.fmrib.ox.ac.uk/fsl/."

[12] J. Udupa, V. LeBlanc, H. Schmidt, C. Imielinska, P. Saha, G. Grevera, Y. Zhuge, P. Molholt, Y. Jin, and L. Currie, "A methodology for evaluating image segmentation algorithm," presented at SPIE Conference on Medical Imaging, San Diego CA, USA, 2002.

[13] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. L. Goualher, D. L. Collins, A. Evans, G. Malandain, N. Ayache, G. E. Christensen, and H. J. Johnson, "Retrospective evaluation of intersubject brain registration," *IEEE Transaction on Medical Imaging*, vol. 22, pp. 1120-1130, 2003.

[14] R. Kikinis, M. E. Shenton, G. Gerig, J. Martin, M. Anderson, D. Grevera, C. R.Guttmann, R. W. McCarley, W. Lorensen, H. Cline, and F. A. Jolesz, "Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging," *Journal of Magnetic Resonance Imaging*, vol. 2, pp. 619-629, 1992.