Difficulties of T1 brain MRI segmentation techniques

M S. Atkins[*a], K. Siu[a], B. Law[a], J. Orchard[a], W. Rosenbaum[a]
[a]School of Computing Science, Simon Fraser University

ABSTRACT

This paper looks at the difficulties that can confound published T1-weighted Magnetic Resonance Imaging (MRI) brain segmentation methods, and compares their strengths and weaknesses.

Using data from the Internet Brain Segmentation Repository (IBSR) as a "gold standard", we ran three different segmentation methods with and without correcting for intensity inhomogeneity. We then calculated the similarity index between the brain masks produced by the segmentation methods and the mask provided by the IBSR. The intensity histograms under the segmented masks were also analyzed to see if a Bi-Gaussian model could be fit onto T1 brain data.

Contrary to our initial beliefs, our study found that intensity based T1-weighted segmentation methods were comparable or even superior to, methods utilizing spatial information. All methods appear to have parameters that need adjustment depending on the data set used. Furthermore, it seems that the methods we tested for intensity inhomogeneity did not improve the segmentations due to the nature of the IBSR data set.

Keywords: brain segmentation, T1-head MRI, IBSR data

1. INTRODUCTION

We are interested in segmenting the brain in T1-weighted head scans to find the perimeter of the brain cortex, which is useful for functional MRI studies. We could not find direct comparisons of published brain segmentation techniques for T1-weighted MRI describing which method was the best. Therefore, we thought it would be beneficial to make comparisons of three methods, including our own, and observe the practical aspects of using these techniques in segmentation. It was our initial hypothesis that intensity based algorithms would be inferior to methods that use more sophisticated spatial information such as gradient (edge) information.

We used T1 MRI head scans from the Internet Brain Segmentation Repository (IBSR)[6] because these data sets came with manual segmentations that we could use as a gold standard. Two particular aspects of the data sets we studied were the intensity inhomogeneity, and the Gaussian nature of the histograms of gray/white matter tissues of the brain.

2. MATERIALS AND METHODS

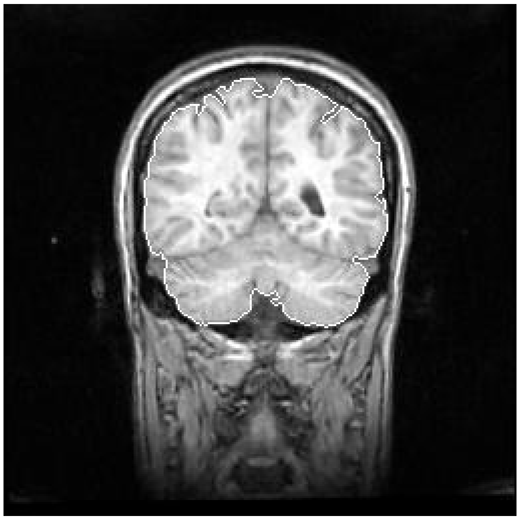
2.1 Data

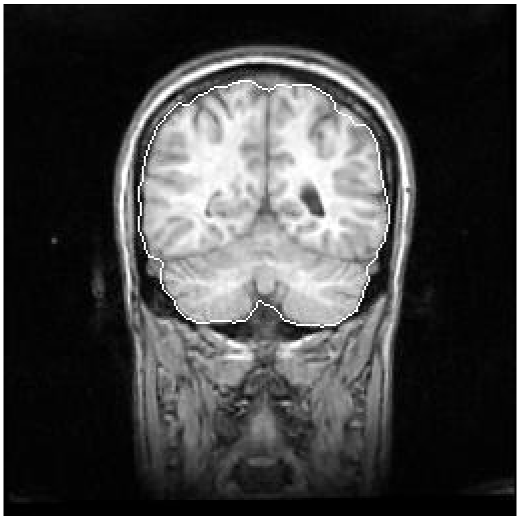
IBSR[6] provided T1-weighted coronal MRI data with gray/white/cerebral spinal fluid/other expert segmentations for twenty normal subjects. The data sets had an intensity range of $0\text{-}2^{16}$ and each slice had 256x256 pixels. The voxel dimensions are $1.17 \times 1.17 \times 3.1 mm^3$. Of these twenty data sets, ten data sets that had intensity values starting at 1024 were used for this experiment because they had a greater range of intensity values than the data sets that had minimum intensity values of 0. The ten data sets with intensities starting at 1024 had their intensity ranges shifted back to a minimum of 0 for this experiment, and the resulting intensity ranges are shown in Table 1.

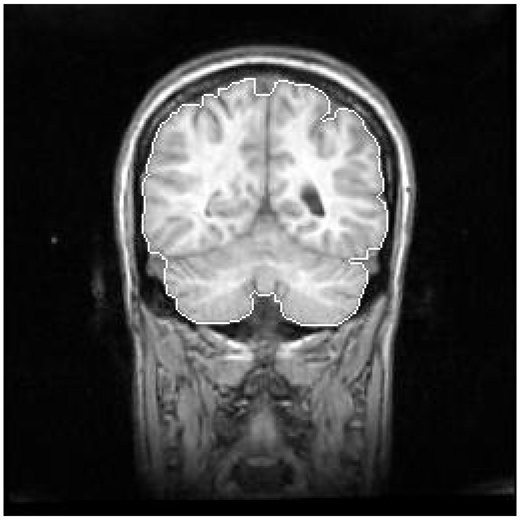
Table 1: Intensity ranges for IBSR data sets

Set Number	1_24	100_23	11_3	111_2	112_2	12_3	13_3	101_3	202_3	205_3
Max Intensity	712	871	958	822	952	956	1107	1369	1240	1740
Slices	65	63	63	63	63	63	63	63	63	63

A sample manually-segmented slice from the IBSR data set 202_3 is shown in Fig. 1a.


(a)


(b)


(c)

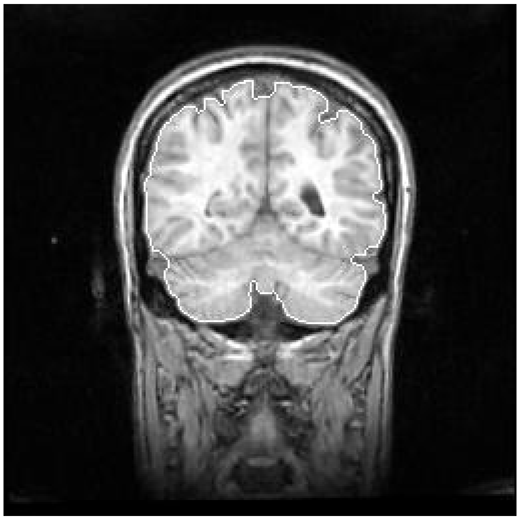

(d)

Figure 1: Segmentation of slice 21 for data set 202_3 by (a) IBSR, (b) BET, (c) Exbrain and (d) SFU Method

2.2 ImageJ

ImageJ[7] is a publicly available image processing program created in Java by researchers from the National Institute of Health. We have used ImageJ version 1.22 for basic image processing functions, and have also written Java plugins to work with the ImageJ program.

2.3 Brain Extraction Tool

The Brain Extraction Tool (BET)[4], is an automatic brain segmentation algorithm for MRI head scans. In the algorithm, a lower intensity threshold, t_2, is chosen so that 2% of the voxels have intensities that fall below it. Similarly, an upper threshold, t_{98}, is chosen so that 2% of the voxels have intensities that fall above it (hence 98% fall below it). Then, a rough brain/non-brain threshold is set to $t = (t_{98} - t_2)/10$. The rough head size, as well as the head's center of gravity, is found based on the partition defined by the threshold *t*. A tessellated sphere, centered on the center of gravity, is expanded until the vertices reach the brain's edge.

We used BET version 1.1 from the FMRIB Software Library to segment the brain from the IBSR data sets. Our initial attempts were not successful, and large regions inferior to the brain were included in the segmentation. However, it was determined that because so much of the neck was included in the head scan, the center of gravity of the volume was outside the brain, falling slightly inferior to the bottom axial slice of the brain. For the purpose of segmentation, approximately 70.2 mm (60 voxels) were removed from the bottom of the data set, leaving a cropped data set with dimensions 256x196x63. BET successfully segmented the brain in these cropped data sets.

2.4 Exbrain

The Exbrain method by Lemieux *et al*[3], is a fully automatic algorithm that segments T1-weighted MRI of the head. Exbrain uses 3D morphological operations and connected component analysis. Exbrain chooses a threshold by incrementing the threshold by unit steps until there is a significant change (three times the average volume difference of the last three thresholds) in the volume found after a set of morphological and connected component operations. This method was not available to the public so the first two steps of the algorithm was re-created in an ImageJ plugin based on the descriptions by Lemieux et al. The later steps of the Exbrain algorithm were not implemented because we achieved acceptable segmentations after the first two steps, and we believed that the later steps would not be significant for our comparisons.

We applied Exbrain to our coronal data sets with anisotropic voxels (1.17x1.17x3.1mm^3). To approximate the 3x3x3mm^3 kernel size suggested by Lemieux for the morphological kernel, we used voxel dimensions of 3x3x1. We also applied Exbrain to interpolated data sets with voxel dimensions of 3x3x3. These interpolated data sets were created by supersampling from the IBSR data sets in the x and y directions by a factor of 2.65, to produce isotropic voxels of 1.17x1.17x1.17mm^3.

2.5 SFU Method

The SFU method[1] was originally created for PD/T2 weighted axially acquired multi-spectral data sets. The SFU method first scales the data set to fit the range 0-255, with data less than or greater than three standard deviations from the mean set to the values 0 or 255 respectively. An anisotropic diffusion filter is then applied to the T2 8-bit data set. The intensity histogram of the filtered head volume is fitted with a Gaussian curve. Thresholds defined in terms of standard deviations from the mean provide the brain/non-brain classification for the segmentation. Any "blobs" (components of the current brain mask) with holes in them are filled, and then a morphological erosion is performed. The PD data set is used to obtain the head mask, and blobs with a center of mass outside a distance from the head center of mask are removed. Finally a dilation is performed to get the brain back to its original size. The ImageJ version of this algorithm is used in this research.

Enhancements were made to the ImageJ plugin version of the algorithm to handle coronal T1 data sets. The size of the bounding box for the center of mass of brain components was adjusted due to the coronal orientation, and the thresholds were adjusted to work with T1-weighted data sets.

2.6 Intensity Inhomogeneity Correction

The N3 Algorithm by Sled[8] and Standardization method by Nyul[9] were applied to the IBSR data sets to see if they would improve the masks produced by the segmentation methods. The N3 algorithm is a widely used method for correcting intensity inhomogeneity in T1-weighted MRI data. Nyul's standardization method deforms the histogram of the data set to a standard histogram. We can then obtain greater similarity of intensities in a data set of the same body regions obtained by the same protocol.

2.7 Comparison

Comparisons between the different segmentation methods were based on the manual segmentation provided by the IBSR. Voxels classified as cerebral spinal fluid (CSF), gray matter (GM) and white matter (WM) were combined to form a truth brain mask. In the case of the interpolated data set for the Exbrain algorithm, we considered every voxel that was non-zero in the interpolated mask to be a brain voxel. For a quantitative evaluation, we use the truth mask and the masks given the three segmentation algorithms to calculate a similarity index measure defined by Zijdenbos *et al*[5].

The segmentations produced by our version of the Exbrain algorithm and the segmentations provided by the IBSR have "holes" in their brain mask, while the brain masks from the SFU algorithm and BET do not. In order to get a more equitable comparison, the similarity indices were also recalculated with the holes in all of the brain masks filled.

2.8 Analysis of White Matter/Grey Matter Histograms

A Gaussian mixture model was fitted onto the pixels under the brain mask provided by the IBSR. The results were compared with the actual WM/GM histograms to see if it was feasible to use the parameters of the Gaussian model to segment the WM and GM.

3. RESULTS

3.1 Brain Extraction Tool Segmentation Results

The BET algorithm had consistent results without any manual intervention except for the cropping of all ten data sets to facilitate the center of mass calculation as described in subsection 2.2. A sample slice segmented using BET from data set 202_3 is shown in Fig. 1b.

3.2 Exbrain Segmentation Results

Using a 3x3x1 kernel size for the Exbrain algorithm on the non-interpolated data set produced some very good results for most data sets, but failed for the data sets 1_24, 12_3 and 191_3. The kernel is not really three-dimensional which may have led to the failures. Using a 3x3x3 kernel size on the interpolated data set gives a significant improvement on the three data sets mentioned previously, and a slight improvement on the other data sets. A sample slice segmented using Exbrain on the interpolated data set 202_3 is shown in Fig. 1c.

Choosing the first intensity that produces a significant change as a threshold does not always succeed. As we can see in Fig. 2, there are three significant volume changes for data set 13_3 at three different threshold levels: 413, 435 and 440. In Fig. 3 we can see that a threshold of 440 provided the best results in this case.

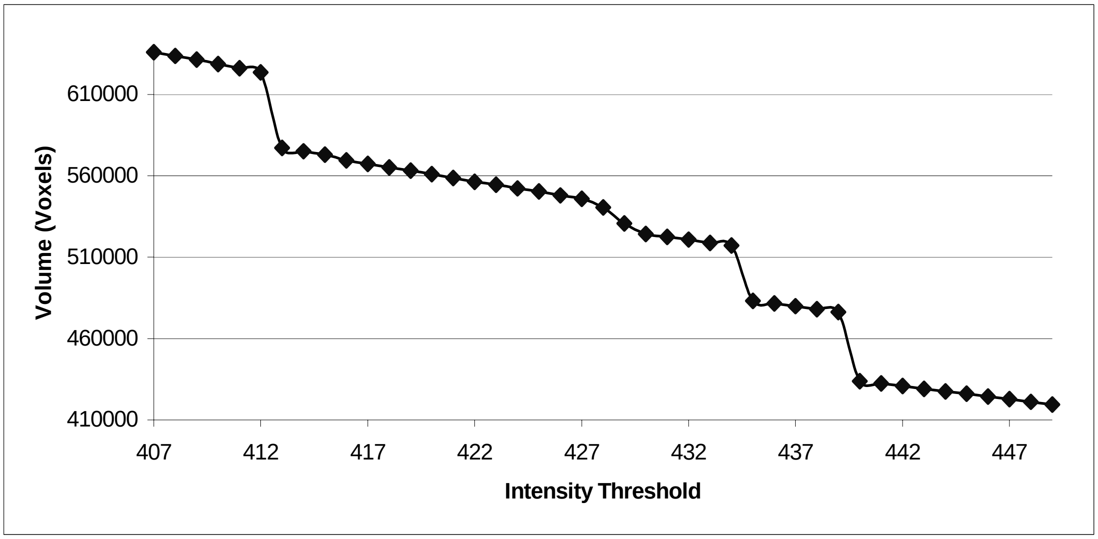
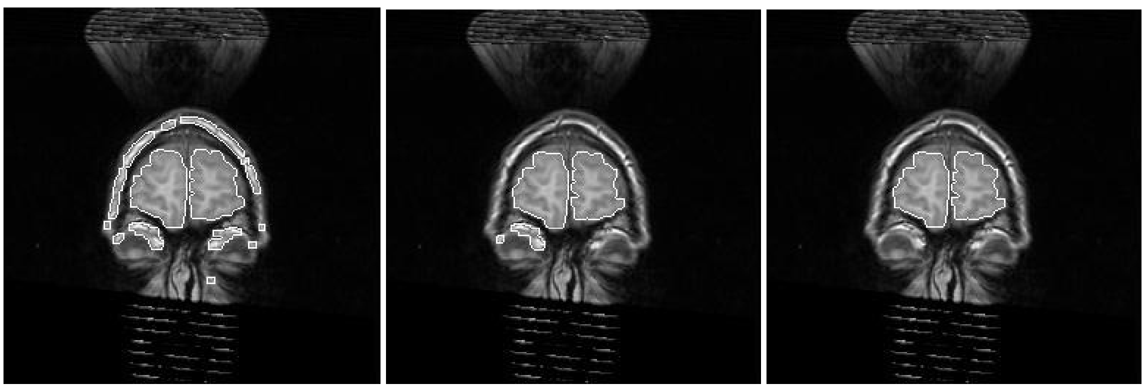


Figure 2: Volume for Different Exbrain Intensity Thresholds for a 3x3x1mm^3 kernel size on data set 13_3



(a) (b) (c)

Figure 3: Exbrain segmentations from slice 54 of data set 13_3 for thresholds of (a) 413, (b) 435 and (c) 440.

3.3 SFU Method Segmentation Results

The SFU algorithm provided comparable results to the BET algorithm, but the parameters that control threshold selection needed some adjustment. In modifying the algorithm for T1-weighted data sets, changing the upper threshold by a wide range did not result in a significantly different segmentation. A sample slice segmented from data set 202_3 using the SFU Method is shown in Fig. 1d)

3.4 Similarity Indices

Fig. 4 shows the similarity indices for all 10 data sets, using all three methods. Fig. 5 shows the similarity indices where the holes in the brain masks were all filled in.

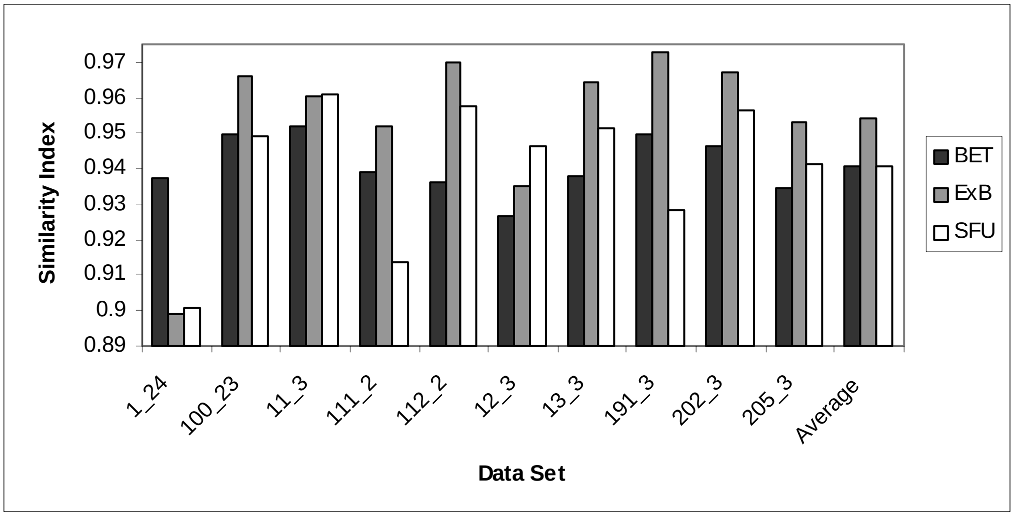


Figure 4: Similarity index for the different segmentation algorithms. Holes not filled.

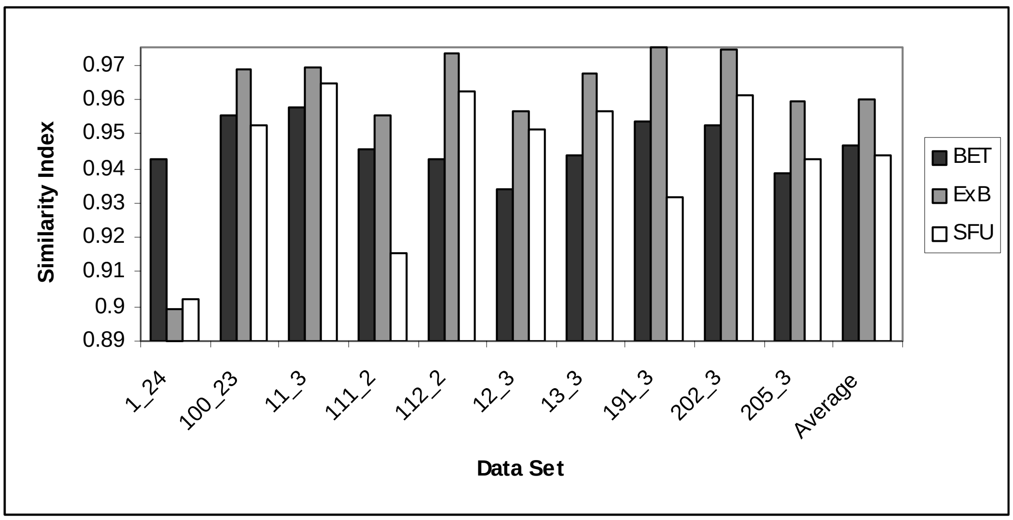


Figure 5: Similarity index for the different segmentation algorithms. Holes filled.

4. DISCUSSION

4.1 Segmentation

From Fig. 1 we can see that all three algorithms provide visually acceptable segmentations. However, it was observed throughout the IBSR data set that BET's segmentation was consistently smoother than the "truth", while the threshold-based algorithms included more fine detail. From the similarity indices, it appears that the Exbrain algorithm provides the best segmentations; however, it occasionally requires the manual selection of the appropriate threshold, as discussed in subsection 3.2. The high similarity indices of Exbrain segmentations are more extraordinary given that we have not fully implemented all the steps of the algorithm. Surprisingly the similarity indices for the Exbrain algorithm increased the most after all the holes in the head mask were filled. This may mean that the holes in the Exbrain algorithm do not correspond to the holes in the IBSR manual segmentations.

We can see from Fig. 4 that the similarity indices for the interpolated Exbrain algorithm were the highest on average, with data 1_24 being the sole exception. The average similarity indices for the BET and the SFU algorithms were quite close. Furthermore, as we can see in Fig. 5, after filling any holes in the brain mask, all of the similarity indices increased, with the similarity indices for the Exbrain algorithm increasing the most. Since we were interested only in the perimeter of the brain mask, the presence or absence of holes in the brain corresponding to “non-brain” tissue is not of concern. However, the fact that the Exbrain similarity improved so much on filling the holes does imply that Exbrain’s holes did not match the manually segmented holes.

Exbrain doesn’t use an upper threshold, and the SFU Method was relatively insensitive to changes in the upper threshold, so it seems that for threshold-based segmentations a single lower threshold may be sufficient.

4.2 Intensity Inhomogeneity Correction

Some initial tests were performed, in which intensity inhomogeneity correction of the data sets was followed by segmentation. It appears that the correction algorithms did not improve the results of the three segmentation algorithms, as there was no significant increase in the similarity index with correction. It is likely that the data sets did not have significant inhomogeneity.

4.3 Histogram Analysis

For the ten IBSR data sets, it seems that a Gaussian mixture model is a reasonable approximation for the voxel intensities under the brain mask. However, the standard deviation of the actual GM and WM histograms tended to be smaller than that of the fitted Gaussian curves (Fig. 6). Although there is considerable overlap between the intensities of GM and WM, it should be possible to use the minimum between the GM/WM peaks as a threshold for segmenting GM and WM. Additional morphological or connectivity operations may need to be performed to accurately segment the GM and WM.

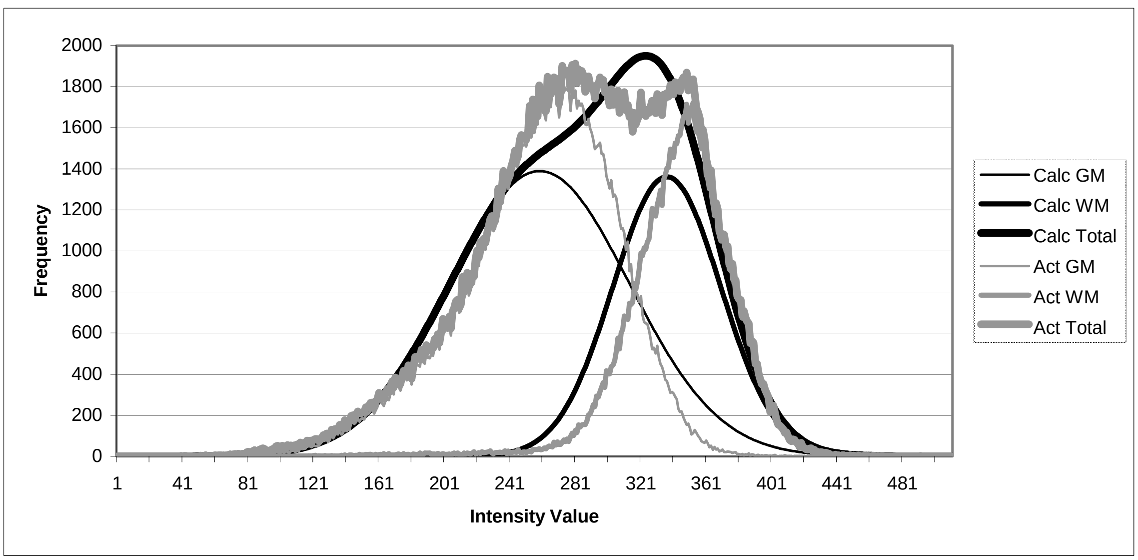


Figure 6: Histogram for dataset 1_24 under the brain mask

5. CONCLUSION AND FUTURE WORK

For the ten IBSR data sets selected for our study, the Exbrain algorithm had the highest average similarity index, even though our implementation was incomplete. Although we found that it had some parameters that need manual adjustment, this may not be true for the full implementation. On the other hand, the BET algorithm required much less manual intervention than the SFU or Exbrain methods. Furthermore, BET’s use of surface modeling attempts to match the natural shape of the brain, as opposed to the almost arbitrary nature of threshold and morphological segmentations. The similarity indices for BET are reasonable, however the perimeter of the brain may be too smooth for practical use. A hybrid method may provide the best solution.

In the future, it will be beneficial to compare the complete algorithms provided by the creators of the respective algorithms, to use alternative measures for comparing the masks produced, and to compare the algorithms on other data sets. With other data sets, correction for intensity inhomogeneity may be more important. Only three algorithms were used in this comparison and it would be interesting to include other algorithms such as atlas-based methods by Hartman *et al*[2] and the artificial intelligence based methods that are evaluated on the IBSR web site.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the National Sciences and Engineering Council of Canada.

REFERENCES

1. M.S. Atkins and B. Mackiewich, “Fully automatic segmentation of the brain in MRI,” *IEEE Transactions on Medical Imaging*, 17(1), pp. 98-107, 1998.
2. S.L. Hartmann, M.H. Parks, H.Schlack, W. Riddle, R.R. Price, P.R. Martin, and B.M. Dawant, “Automatic Computation of Brain and Cerebellum Volumes in Normal Subject and Chronic Alcoholics”, *Lecture Notes in Computer Science*. **1613**, pp. 430-435, 1999.
3. L. Lemieux, G. Hagemann, K Krakow, and F.G. Woermann, “Fast, Automatic Segmentation of the Brain in T1-weighted Volume Magnetic Resonance Image Data,” *Proc. SPIE Medical Imaging 1999: Image Processing*, K. Hanson, ed., **3661**, pp.152-160, SPIE, 1999.
4. S. Smith, "Robust automated brain extraction", In *Sixth Int.Conf. on Functional Mapping of the Human Brain*, page 625. 2000.
5. A. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation," *IEEE Transactions on Medical Imaging*, **13(**4), pp. 716-724, 1994.
6. http://neuro-www.mgh.harvard.edu/cma/ibsr/
7. http://rsb.info.nih.gov/ij/docs/index.html
8. L. Nyul, J. Udupa, and X. Zhang "New Variants of a Method of MRI Scale Standardization," *IEEE Transactions on Medical Imaging*, **19(**2), pp. 143-150, 2000.
9. J. Sled, A. Zijdenbos, and A. Evans "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI data," *IEEE Transactions on Medical Imaging*, **17(**1), pp.87-97, 1998.

* stella@cs.sfu.ca; phone 1 604 291-4288; fax 1 604 291-3045; http://www.css.sfu.ca/sites/mcl/; School of Computing Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6