Improved Skull Stripping using Graph Cuts

S. A. Sadananthan[1], W. Zheng[1], M. W. Chee[2], and V. Zagorodnov[1]

[1]Computer Engineering, Nanyang Technological University, Singapore, Singapore, [2]Cognitive Neuroscience Laboratory, Duke-NUS Graduate Medical School, Singapore, Singapore

Introduction. Removal of non-brain tissues, particularly dura, is an important step in enabling accurate measurements of brain structures. Previous approaches (2,5,6) can lead to brain loss and/or leave residual dura (4), posing a problem for subsequent cortical thickness estimation. Brain loss cannot be reversed downstream in the segmentation pipeline. Inclusion of dura mater can cause overestimation of cortical thickness. We propose a novel, fully-automatic skull stripping algorithm that is motivated by the earlier work involving intensity thresholding and removal of narrow connections using morphological processing (**Error! Reference source not found.**). Instead of morphological operations, we employed a superior graph theoretic segmentation with suitably modified edge weight assignment to enable precise removal of narrow connections and dura attachments. We wanted to test whether this approach would lead to cleaner (with fewer dura attachments) brain mask, which would be more suitable for subsequent cortical thickness estimation.

Methods. We used three data sets. Data set 1 contained 18 1.5mm T1W volumes from the Internet Brain Segmentation Repository (IBSR, http://www.cma.mgh.harvard.edu/ibsr/), data set 2 and 3 each contained 15 T1W volumes (TR 2250 ms, TE 2.6 ms, TI=2.6 ms, 1x1x1.1 mm). Images in data set 2 were of good quality, whereas data set 3 images led to various segmentation problems when processed by FreeSurfer segmentation pipeline (http://surfer.nmr.mgh.harvard.edu/fswiki/recon-all). The ground truth was defined as GM+WM (1) and the performance metrics included Jaccard similarity index (JS), false negative (FN) and false positive (FP) rates. We also evaluated an 'adjusted FP rate' that excluded from consideration dark non-brain voxels 5mm tangential to the ground truth boundary. This effectively discarded majority of CSF, brain stem and cerebellum voxels from computation of FP rate. These structures do not materially affect the quality of subsequent tissue segmentation and their preservation within the brain mask is optional.

The proposed skull stripping approach (GCUT) comprises three steps, intensity thresholding to obtain preliminary mask, removal of narrow connections using graph cuts and post-processing. The graph cuts approach (3) was used with the following graph weight assignments:

$$w_{ij} = \max_{v_i \in F}(D(v_i), D(v_j)) \left[\exp\left(k \frac{\min_{v_i \in F}(I(v_i), I(v_j)) - T}{I_{WM} - T} \right) - 1 \right]$$

where $D(v_i)$ is the distance transform value, I_{WM} is the intensity of white matter (WM) and parameter k controls the contribution of voxel intensities in deciding cut position. Such an assignment makes the weights of the edges inside narrow connections small, favoring cuts through them. Finally we performed a post-processing procedure (morphological closing) to reinstate partial volume gray matter (GM) voxels inadvertently removed by initial thresholding.

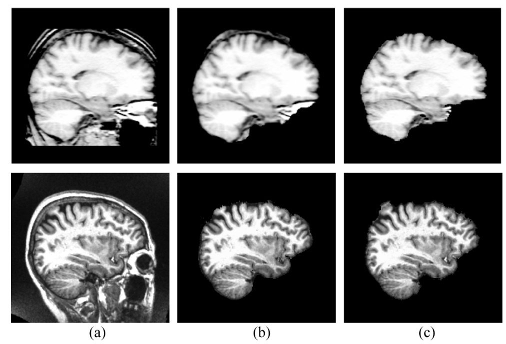

Fig. 1. Typical results exhibited by HWA (b) and GCUT (c). HWA is often confused by double boundary between skull/dura/GM, resulting in inclusion of large chunks of dura mater, rows 1. GCUT fails to cut connections where there is no noticeable separation between dura and GM, row 2.

Results. Of all tested approaches (BET (6), BSE (5) and HWA (2)), only HWA achieved acceptable brain loss (almost zero) and was further compared to GCUT. GCUT parameters were empirically fixed at $T = 0.36 I_{WM}$ and $k=2.3$ for all three data sets, HWA was used with the default parameters. Compared to HWA, our approach led to substantial decrease (10-30%) in adjusted FP rate and similar FN rate (slightly higher on data sets 1 and 2 and lower on data set 3), see Table 1. The reduction in adjusted FP rate was statistically significant for data sets 1 and 2.

Table 1. Comparison of the skull stripping approaches, *p<0.05

Metrics	Data set 1 (mean (SD))		Data set 2 (mean (SD))		Data set 3 (mean (SD))	
	HWA	GCUT	HWA	GCUT	HWA	GCUT
JS (%)	0.89* (0.03)	0.91 (0.03)	0.79* (0.02)	0.80 (0.02)	0.79 (0.02)	0.79 (0.03)
FN (%)	0.015 (0.02)	0.031 (0.04)	0.013 (0.02)	0.025 (0.03)	0.055 (0.07)	0.038 (0.04)
FP (%)	9.04* (5.43)	6.94 (4.63)	18.21* (2.10)	17.23 (1.99)	19.55 (2.61)	18.67 (3.64)
FP adj (%)	4.12* (2.27)	2.94 (2.11)	3.63* (0.54)	2.95 (0.42)	4.28 (0.79)	3.92 (1.40)

When used in FreeSurfer segmentation pipeline, despite preserving less dura, GCUT mask led to more cortical thickness overestimation on data set 3, which can be explained by the example in Fig 1. GCUT was more successful than HWA when dealing with double boundary between GM/dura/skull (top row of Fig 1), but failed to remove connections where there was no noticeable separation between the tissues (bottom row of Fig 1). However, since the two approaches tended to produce errors in different locations, we have found that intersecting the two masks led to significant decrease in cortical thickness overestimation on data set 3, from 0.06% (HWA only) to 0.03% (intersection of HWA and GCUT), completely resolving the overestimation problem in 11 out of 15 cases. Since both masks had negligible brain loss, their intersection did not lead to cortical thickness underestimation.

Conclusion. Compared to HWA, the new skull stripping approach leads to at least 10-30% reduction in dura mater without having to trade off for increased brain tissue erosion. It also doesn't rely on a shape atlas, suggesting possible deployment of this method to developmental studies and animal brains, and does not require alignment to a standard space, which may fail for strongly misaligned brains. But the main value of our approach is in using it in conjunction with HWA, for example by using a simple intersection of the two masks. This resulted in significant improvement in subsequent segmentation performance, completely resolving 11 out of 15 problematic cases in our data set.

Acknowledgements. This work is supported by grants SBIC C-012/2006 and BMRC 04/1/36/19/372 provided by A*STAR, Singapore.

References. [1] Lemieux (1999) Magnetic Resonance in Medicine,42,127-135. [2] Segonne (2004) NeuroImage,22,1060-1075. [3] Boykov (2001) ICCV,105-112. [4] Fennema-Notestine (2006) Human Brain Mapping,27,99-113. [5] Shattuck (2001) NeuroImage,13,856-876. [6] Smith (2002) Human Brain Mapping, 17, 143-155.