Robust Deformable-Surface-Based Skull-Stripping for Large-Scale Studies

Yaping Wang[1,2], Jingxin Nie[2], Pew-Thian Yap[2], Feng Shi[2], Lei Guo[1], and Dinggang Shen[2,*]

[1] School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi Province, China
[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA
`dgshen@med.unc.edu`

Abstract. Skull-stripping refers to the separation of brain tissue from non-brain tissue, such as the scalp, skull, and dura. In large-scale studies involving a significant number of subjects, a fully automatic method is highly desirable, since manual skull-stripping requires tremendous human effort and can be inconsistent even after sufficient training. We propose in this paper a robust and effective method that is capable of skull-stripping a large number of images accurately with minimal dependence on the parameter setting. The key of our method involves an initial skull-stripping by co-registration of an atlas, followed by a refinement phase with a surface deformation scheme that is guided by prior information obtained from a set of real brain images. Evaluation based on a total of 831 images, consisting of normal controls (NC) and patients with mild cognitive impairment (MCI) or Alzheimer's Disease (AD), from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database indicates that our method performs favorably at a consistent overall overlap rate of approximately 98% when compared with expert results. The software package will be made available to the public to facilitate neuroimaging studies.

1 Introduction

Automated skull-stripping is an indispensible tool for large-scale multi-site studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [1], where thousands of MRI scans need to be processed. An accurate and highly automated algorithm is hence highly desirable, preferably avoiding any form of human intervention that can possibly cause bias and inter-expert differences. The inaccuracy of skull-stripping may result in error in cortical thickness estimation, volumetric analysis, brain atrophy estimation, and registration between structural and functional images (fMRI and PET). For instance, the unremoved dura can result in an overestimation of cortical thickness [2], whereas the unintended removal of brain tissue may lead to underestimation of cortical atrophy. These errors cannot be recovered in subsequent processing stages. The existing skull-stripping methods can generally be divided into three categories: 1) morphology-based methods, 2) surface-based methods, and 3) meta methods.

---

* Corresponding author.

Morphology-based methods extract the brain through a series of thresholding and morphological operations [3, 4]. In [4] for example, a heuristic approach utilizing thresholding, morphological opening, and connected component analysis is devised to deal with specific problems encountered in skull-stripping. Another popular method is the Brain Surface Extractor (BSE) [5], which removes non-brain tissue using a combination of anisotropic diffusion filtering, edge detection, and mathematical morphology. Also, a 3D watershed transform, modified by combining pre-flooding to avoid over-segmentation, is proposed in [6]. A graph cuts based method is introduced in [7], utilizing intensity thresholding followed by removal of narrow connection using a graph theoretic image segmentation technique. *While some published approaches are effective, morphology-based methods generally involve some degree of user interaction, and are also sensitive to the scanning parameters and intensity inhomogeneity.*

Surface-based methods rely mainly on the image gradient information to locate the brain surface, which is modeled by an active contour. In [8], a tessellated ellipsoidal template is deformed to the inner surface of the skull, and is then iteratively deformed using forces derived from image gradients and a curvature constraint. Brain Extraction Tool (BET) [9] uses a deformable model that evolves a surface to fit the brain boundary by application of a set of locally adaptive forces. These forces account for surface smoothness and voxel intensity changes in the surface vicinity. Hybrid Watershed Algorithm (HWA) [10] combines watershed algorithm developed in [6] with a deformable-surface model, using surface curvature and statistics of the distance of the surface to the center of gravity (COG) to detect and correct inaccuracies in brain extraction. The above-described methods are generally more robust and less sensitive to image artifacts, and require less human interaction. *However, these methods require that the initial active contour should be close enough to the targeted surface to avoid local minima; and failure to meet this requirement can cause unpredictable results.*

Meta methods combine various skull-stripping methods for achieving the best outcome [11, 12]. However, they rely heavily on the component methods and do not solve the fundamental problems pertaining to these methods.

Due to the presence of imaging artifacts, anatomical variability, and varying contrast properties, most of these techniques, however, do not give satisfactory results over a wide range of scan types and neuroanatomies without some form of manual intervention. In this paper, we focus on a surface-based method, since it gives us the flexibility of incorporating both surface geometry and atlas-based prior information into the skull-stripping process. For a good initial extraction of the brain, we utilize FLIRT [13] and Demons-based registration [14] to warp an atlas, and hence its brain mask, to the targeted image for initial skull-stripping. We then employ a probability map, incorporating prior information gathered from a population of real brain images with manual extractions, to guide surface evolution for refinement of the skull-stripping result. This allows us to guide the skull-stripping process using realistic shape and topological constraints to weight the forces in driving surface deformation, thus giving us more accurate and consistent results over a large number of images.

2   Method

The skull-stripping method consists of the following steps: 1) Initialization by warping an atlas to the target image, where the brain mask for the atlas is used for

approximate skull-stripping, and 2) Refinement of the former result with a deformable-surface, guided by a brain probability map generated from a set of real images. The details are given in the following.

2.1   Initialization

A good initialization of the deformable surface is important for avoiding local minima and suboptimal solutions. To facilitate more accurate parameter estimation for better positioning of the initial active contour, FLIRT [13] and the Demons registration algorithm [14] are employed to help mask out non-brain voxels. Demons registration works especially well in the image regions with clear intensity changes. In a T1-weighted MR image, the brain boundary appears as a surface with low cerebrospinal fluid (CSF) level intensity, compared with the GM and the skull, thus providing good contrast for alignment. We first register the ICBM template (with skull) [15] to the target image and then use the accompanying brain mask, which is dilated for avoiding over-skull-stripping, as an initial skull-stripping, which will be further refined using a deformable-surface-based approach as described below.

2.2   Deformable-Surface-Based Skull-Stripping

2.2.1   Brain Probability Map

A probability map, indicating the likelihood of whether a voxel belongs to a part of the brain tissue, is obtained from a set of manually skull-stripped images. Specifically, we utilize the ICBM template as the reference image, onto which a set of images (with skulls) are aligned first linearly (FLIRT) [13] and then non-linearly (Demons) [14]. The manually delineated brain masks of the images are warped to the template space using the same set of transformations. From the aligned brain masks, we can evaluate for each voxel the probability that it belongs to the brain by simply computing the fraction of brain masks that consider this voxel a part of the brain. That is, for each point (i, j, k), the probability is $p(i, j, k) = \sum_{n=1}^{N} B_n(i, j, k) / N$, where B_n represents the n-th brain mask out of a total of N. From Fig. 1, we can see that the probability map has a value of 1 for the majority of the brain, except the boundary where the voxel memberships are ambiguous. This probability map is warped onto the target image to guide skull-stripping using a deformable surface. The probability map helps greatly restrict the extent of deformation of the surface, thus minimizing the chance of it falling in a less desirable region.

Fig. 1. Brain probability map overlaid on the brain image, shown in three different views

2.2.2 Surface Deformation

Following the work of Smith [9], the implementation[1] of the evolution equation is accomplished with a parametric active contour. The brain boundary is modeled by a surface tessellated using connected triangles. The surface deformation is driven by three main forces: 1) force from the smoothness constraint, 2) force from the intensity gradient in the surface vicinity, and 3) force from the brain probability map.

The force pertaining to the smoothness constraint consists of two parts:

$$\mathbf{F_1} = c_1 \mathbf{u_t} + c_2 \mathbf{u_n}, \tag{1}$$

where $\mathbf{u_t}$ and $\mathbf{u_n}$ are the tangential and normal components of $\mathbf{u}$. $\mathbf{u}$ is the difference vector between the positions of the current vertex and its mean neighboring vertices, i.e.,

$$\mathbf{u} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_m - \mathbf{x}_0, \ \mathbf{u_n} = (\mathbf{u} \cdot \mathbf{n})\mathbf{n}, \ \mathbf{u_t} = \mathbf{u} - \mathbf{u_n}, \tag{2}$$

where $\mathbf{n}$ is the surface outward normal at current vertex $\mathbf{x}_0$, M is the number of neighboring vertices, and $\mathbf{x}_m$ is the m-th neighboring vertex. The sole role of the tangent force $\mathbf{u_t}$ is to keep all vertices in the surface equally spaced, i.e., moving them only within the surface. The normal force $\mathbf{u_n}$ is a curvature-reducing force to ensure smoothness of the surface during the evolution process. Usually c_1 is set as 0.5. To ensure that the surface is sufficiently smooth and meanwhile avoids underestimation of curvature at parts of the surface, c_2 is a nonlinear function adaptive to local surface geometry. For this purpose, we need first to determine the local radius of curvature r: $r = (d^2)/(2|\mathbf{u_n}|)$, where d is the mean distance from vertex to neighboring vertices over the whole surface. Weight c_2 is then defined using a sigmoid function:

$$c_2 = 0.5(1 + \tanh(l_2 * (1/r - l_1))) \tag{3}$$

with $l_1 = 0.5(1/r_{MIN} + 1/r_{MAX})$ and $l_2 = 6/(1/r_{MIN} - 1/r_{MAX})$. Here, the values for r_{MIN} and r_{MAX} are empirically optimized for typical geometries in the human brain and are set as 3.33 mm and 10 mm, respectively. High local mean surface curvature is hence penalized to achieve surface smoothing.

The second force is derived from the image intensity information:

$$\mathbf{F_2} = \frac{2(I_{MIN} - G_1)}{I_{MAX} - G_2} \mathbf{n}, \tag{4}$$

where I_{MIN} and I_{MAX} are the minimum and maximum of local intensities quantified along the normal direction pointing inward from the current vertex:

$$I_{MIN} = \max\big(G_2, \min\big(G_m, I(0), I(1), \dots, I(S_1)\big)\big);$$

$$I_{MAX} = \min\big(G_m, \max\big(G_s, I(0), I(1), \dots, I(S_2)\big)\big).$$

The search range pertaining to the minimum and maximum intensities are S_1 and S_2, respectively. Typically, S_1 is set as 20 mm and $S_2 = S_1/2$. Here the image intensity minimum G_2, the median intensity G_m, and the threshold G_s that separates brain and non-brain matter are approximately estimated according to the cumulative intensity

---

[1] The proposed method is a modification of the original BET algorithm.

histogram of the brain image. They're used to avoid voxels that are too dark or too bright. G_1 in Equation (4) is a locally estimated intensity threshold which is used to distinguish between brain and non-brain:

$$G_1 = (I_{MAX} - G_2) * f + G_2. \tag{5}$$

Parameter f is called the fractional intensity threshold and falls between the range of 0 and 1.

The third force is derived from the brain probability map obtained in **Section 2.2.1**:

$$\mathbf{F}_3 = (p - 0.5)\mathbf{n}, \tag{6}$$

where $p \in [0,1]$ is the probability value from the map. At a point where the surface is close to the brain boundary, $p \approx 0.5$, and $\mathbf{F}_3$ is small; when $p > 0.5$, the point is likely within the brain, and an outward force is exerted; when $p < 0.5$, a inward force is exerted. This restricts the searching of the brain boundary in this range and drives it to find the true boundary. $\mathbf{F}_3$ is directly added to $\mathbf{F}_2$ with a certain weight c_4, which we set as 1 here. Both of them act in the direction of $\mathbf{n}$. Thus, for each vertex i at iteration t, the update equation is:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + [\mathbf{F}_1 + c_3(\mathbf{F}_2 + c_4\mathbf{F}_3)], \tag{7}$$

where $c_3 = 0.05 * d$.

3 Experimental Results

3.1 Dataset

The proposed algorithm was evaluated based on the ADNI dataset. Subjects were recruited from over 50 sites across the U.S. and Canada. MRI scans were acquired using GE, Philips, and Siemens scanners. The baseline T1-weighted MRI scans of 831 subjects (230 NC, 403 MCI, 198 AD; ages range: 55~90 years) were used in this work. We randomly selected images of 25 subjects from each group for generating the respective probability maps; the rest of the images were used as testing data for evaluation of all considered methods. For consistency, all images were resampled to be of dimensions 256×256×256 and resolution 1×1×1 mm³. Nonparametric nonuniform intensity normalization (N3) [16] was performed for correcting the intensity inhomogeneity.

3.2 Qualitative Evaluation

The following 5 methods were evaluated: 1) BET, 2) BET-B: BET with bias field correction and neck cleanup (option "-B"), 3) BSE, 4) HWA, and 5) the proposed method. Let P and Q represent the extracted brain image and the manually skull-stripped image, respectively. We measured their similarity using the Dice ratio: $D(P,Q) = 2|P \cap Q|/(|P| + |Q|)$. For each method compared, except our method, we determined for each image the result that gave the best Dice ratio by grid search over a range of parameter values, compared with manual skull-stripping results by an expert. The following is the range of values considered for the parameters (in

MATLAB notation): BET, fractional intensity threshold = [0.1:0.05:0.8]; BET-B, fractional intensity threshold = [0.1:0.05:0.8], option "-B" turned on; BSE, diffusion constant = [5:5:60], edge detection constant = [0.3:0.05:0.8], diffusion iterations = [3:1:5]; HWA, the default parameters with or without "-less", "-more", "-atlas", "-less -atlas" or "-more -atlas" were used. For the proposed method, a fractional intensity threshold value of 0.6 was used for all images. Typical results given by these 5 methods are shown in Fig. 2. For the image of Fig. 2, the best parameters selected are listed below: BET, fractional intensity threshold = 0.75; BET-B, fractional intensity threshold = 0.4, option "-B" turned on; BSE, diffusion constant = 5, edge detection constant = 0.7, diffusion iterations = 3; HWA, the default parameters were used; the proposed method, fractional intensity threshold = 0.6. The arrows in Fig. 2 highlight several problematic areas. BET typically over skull-strips the anterior frontal cortex, anterior temporal cortex, posterior occipital cortex, and the cerebellar areas. BSE typically leaves some dura unremoved in these regions. HWA retains most of the brain tissue, but leaves some dura intact. The proposed method successfully overcomes all these problems.
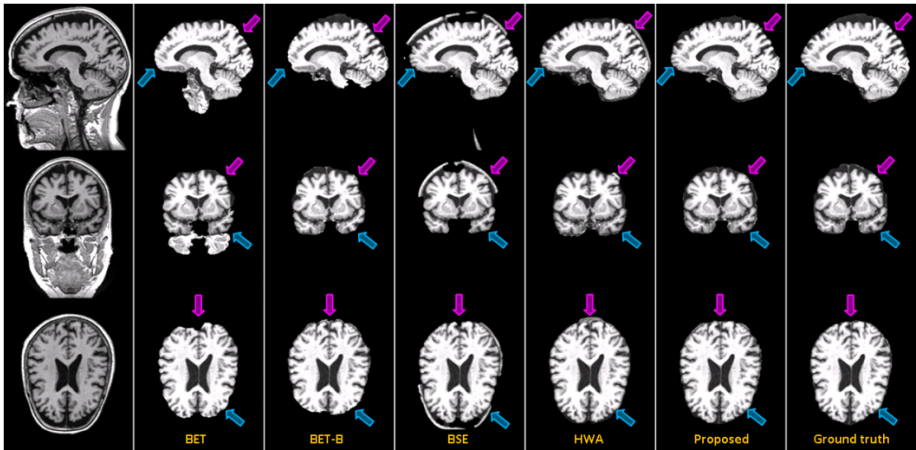

Fig. 2. Typical skull-stripping results of existing methods, compared with the proposed method

3.3 Quantitative Evaluation

Similar to the qualitative evaluation, we compared the proposed method with BET, BET-B, BSE and HWA. And for each method compared, except our method, the optimal parameters for each image were selected as described in Section 3.2. For the proposed method, a fractional intensity threshold value of 0.6 was used for all images as described in Section 3.2. As can be appreciated from Fig. 3, the proposed method yields consistently the best results when compared with all other methods, despite the fact that one single set of parameters was used for all images, whereas optimized parameters were used for other methods. Fig. 4 shows that the performance of the proposed method is insensitive to different fractional intensity threshold values. Of note, on a 2.8 GHz machine, the proposed method took approximately 2 minutes for

registration using FLIRT and Demons, and an extra 1 minute for skull-stripping an image. Other methods typically took 60-70 seconds, except for BET when used with option "-B", which typically took 20-30 minutes.
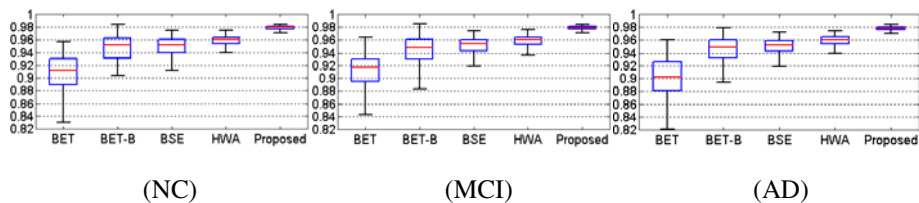

(NC)   (MCI)   (AD)

Fig. 3. Distributions of Dice ratios for different methods and different subject groups
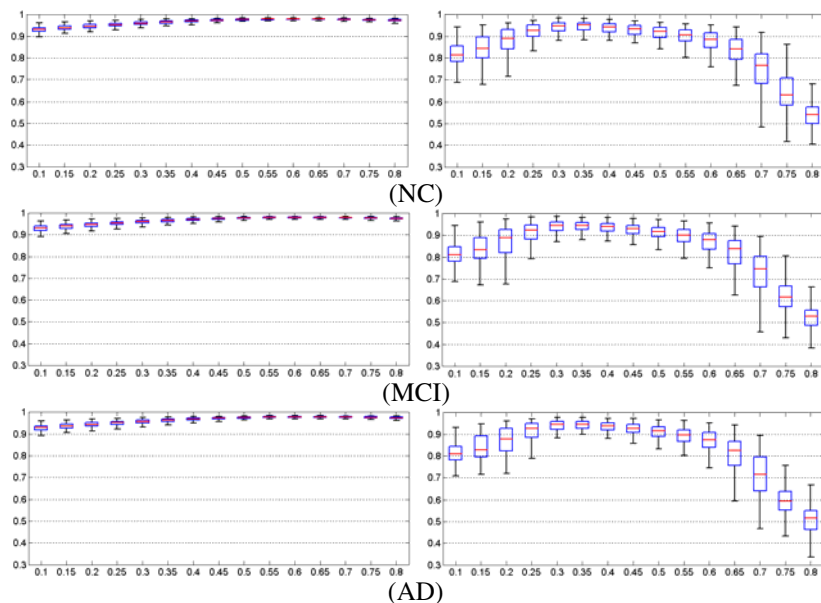

(NC)

(MCI)

(AD)

Fig. 4. Distributions of Dice ratios over different fractional intensity threshold values for the proposed method (left) and BET-B (right). The proposed method is robust to parameter variation.

4 Conclusion

We have proposed in this paper a robust and effective skull-stripping method that is capable of skull-stripping a large number of images automatically without human intervention. Our method combines an atlas-based approach, for coarse skull-stripping, with a deformable-surface-based approach, for more localized refinement. Experimental results indicate that our method yields highly consistent results using

only a set of fixed parameters. The software package will be released to the public for facilitating neuroimaging studies.

References

1. Mueller, S.G., Weiner, M., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. N. Am. 15, 869–877 (2005)
2. van der Kouwe, A.J., Benner, T., Salat, D.H., Fischl, B.: Brain morphometry with multiecho MPRAGE. NeuroImage 40, 559–569 (2008)
3. Höhne, K.H., Hanson, W.A.: Interactive 3D segmentation of MRI and CT volumes using morphological operations. J. Comput. Assist. Tomogr. 16, 285–294 (1992)
4. Lemieux, L., Hagemann, G., Krakow, K., Woermann, F.G.: Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. Magn. Reson. Med. 42, 127–135 (1999)
5. Shattuck, D.W., Sandor-Leahy, S., Schaper, K.A., Rottenberg, D.A., Leahy, R.M.: Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13, 856–876 (2001)
6. Hahn, H.K., Peitgen, H.-O.: The Skull Stripping Problem in MRI Solved by a Single 3D Watershed Transform. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 134–143. Springer, Heidelberg (2000)
7. Sadananthan, S.A., Zheng, W., Chee, W.L., Zagorodnov, V.: Skull stripping using graph cuts. NeuroImage 49, 225–239 (2010)
8. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9, 179–194 (1999)
9. Smith, S.M.: Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155 (2002)
10. Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B.: A hybrid approach to the skull stripping problem in MRI. NeuroImage 22, 1060–1075 (2004)
11. Rex, D.E., Shattuck, D.W., Woods, R.P., Narr, K.L., Luders, E., Rehm, K., Stoltzner, S.E., Rottenberg, D.A., Toga, A.W.: A meta-algorithm for brain extraction in MRI. NeuroImage 23, 625–637 (2004)
12. Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S., ADNI: Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. NeuroImage 55, 1091–1108 (2011)
13. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Medical Image Analysis 5, 143–156 (2001)
14. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric Diffeomorphic Image Registration with the Demons Algorithm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 319–326. Springer, Heidelberg (2007)
15. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C.: Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 22, 324–333 (1998)
16. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97 (1998)