Skull Stripping of Neonatal Brain MRI

Using Prior Shape Information with Graph Cuts

J Digit Imaging (2012) 25:802–814
DOI 10.1007/s10278-012-9460-z

Skull Stripping of Neonatal Brain MRI: Using Prior Shape Information with Graph Cuts

Dwarikanath Mahapatra

Published online: 22 February 2012

Abstract In this paper, we propose a novel technique for skull stripping of infant (neonatal) brain magnetic resonance images using prior shape information within a graph cut framework. Skull stripping plays an important role in brain image analysis and is a major challenge for neonatal brain images. Popular methods like the brain surface extractor (BSE) and brain extraction tool (BET) do not produce satisfactory results for neonatal images due to poor tissue contrast, weak boundaries between brain and non-brain regions, and low spatial resolution. Inclusion of prior shape information helps in accurate identification of brain and non-brain tissues. Prior shape information is obtained from a set of labeled training images. The probability of a pixel belonging to the brain is obtained from the prior shape mask and included in the penalty term of the cost function. An extra smoothness term is based on gradient information that helps identify the weak boundaries between the brain and non-brain region. Experimental results on real neonatal brain images show that compared to BET, BSE, and other methods, our method achieves superior segmentation performance for neonatal brain images and comparable performance for adult brain images.

Keywords Shape prior · Graph cuts · Neonatal · Brain · MRI · Segmentation · Gradient

D. Mahapatra (✉)
Department of Computer Science,
Swiss Federal Institute of Technology (ETH),
Room CAB F 61.1, Universitätstrasse 6,
8092 Zurich, Switzerland
e-mail: dmahapatra@gmail.com

Introduction

Skull stripping is an important step in brain image analysis and refers to the removal of the scalp, skull, dura, eyes, and other extraneous regions. Tissue classification, registration, volumetric analysis of the brain, and brain surface reconstruction all depend upon accurate skull stripping. Any accidental removal of brain tissues is damaging because it cannot be reversed in later processing stages. Popular skull stripping methods include region-based approaches [1, 2], boundary-based techniques [3, 4], and hybrid methods [5, 6]. In brain image analysis terminology, brain extraction refers to skull stripping, while brain segmentation refers to classification of the brain into different tissues (e.g., white matter, WM; gray matter, GM; and cerebrospinal fluid, CSF). In this paper, we propose a novel method using graph cuts that incorporates prior shape information for skull stripping of infant (neonatal) brain magnetic resonance (MR) images.

Segmentation of neonatal brain MRI is important for the study and treatment of brain injury and disorder due to prematurity. Shortly after an infant is born, neurodevelopment includes critically important maturational processes which may be measured quantitatively by brain imaging. Brain tissue volumes have been shown to change with age [7]. In neonatal brain segmentation, tissue classes apart from WM, GM, and CSF are identified to characterize brain development. In such a scenario, accurate skull stripping assumes increased significance.

Most of the skull stripping methods are designed to work with T1-weighted images as it is the most popular modality for brain MRI due to its superior contrast over other modalities like T2 or FLAIR. Most state-of-the-art skull stripping algorithms have been developed for adult brain MR images. When used on neonatal images, these algorithms do not

obtain the high segmentation accuracy of adult brain volumes. Figure 1 shows an example image of adult and neonatal brain MRI to illustrate their differences. Adult brain MRI have a well-defined boundary between the brain and skull, while in neonatal brain MRI, the brain and skull are not easily separable. Unlike neonatal volumes, in adult brain MRI, different tissues are quite clearly defined which provides more information for accurate skull stripping.

Neonatal brain extraction has some unique challenges compared to adult brain volumes. The neonatal data are characterized by poor image quality due to their inherently low spatial resolution, insufficient tissue contrast, and ambiguous tissue intensity distributions [8]. Two very popular skull stripping methods are the brain surface extractor (BSE) [1] and the brain extraction tool (BET) [3]. BSE uses a combination of anisotropic-diffusion filters, Marr-Hildreth edge detectors, and morphological operators to separate brain and non-brain tissues, but needs parameter tuning for specific images. BET initializes a spherical mesh around the center of gravity of the brain and uses a deformable model. Internal and external forces push the initial volume to the brain boundary. BET is fast and relatively insensitive to parameter settings, but can produce areas wrongly identified as the brain. 3dSkullStrip, part of the AFNI package [9], is a modified version of BET. It is adapted to avoid segmentation of eyes and ventricles, reduce leakage into the skull, and use data outside the surface to guide its evolution. The watershed algorithm (WAT) [2] is an intensity-based approach that relies on preflooding, and the basin represents the brain. The hybrid watershed algorithm (HWA) [6] which forms part of FreeSurfer software [10] exhibits greater robustness than other methods, by combining a watershed algorithm, a deformable surface, and a probabilistic atlas. The watershed algorithm makes an initial estimate of the mask by assuming connectivity of the white matter, and a statistical atlas is used to guide the evolution of a smooth surface and refine the mask.

Other approaches to skull stripping segment the brain using intensity thresholds followed by morphological operations to cut narrow connections between brain and non-brain regions [11–13]. Such operations can only remove very narrow connections (weakly connected regions). To overcome this limitation, Sadananthan et al. [14] propose a graph cut-based approach to position cuts for isolating and removing dura. The method in [15] removes narrow connections using distance transforms followed by watershed algorithm (DWAT). In [4], active contours were used to fit the brain where the curve is embedded in a higher dimensional function and locally adapted to reduce sensitivity to bias field. Zeng et al. [16] proposed a system of two level sets whose zero level curves represent the inner and outer boundaries of the cortex. Rehm et al. in [17] use a hierarchy of masks from different models to form a consensus mask for brain segmentation. A learning-based brain extraction system (ROBEX) was introduced in [18] which combines a discriminative and a generative model for brain extraction. The discriminative model is a Random Forest classifier trained to detect the brain boundary while the generative model is a point distribution model to ensure a plausible result. For a new image, the generative model is used to find

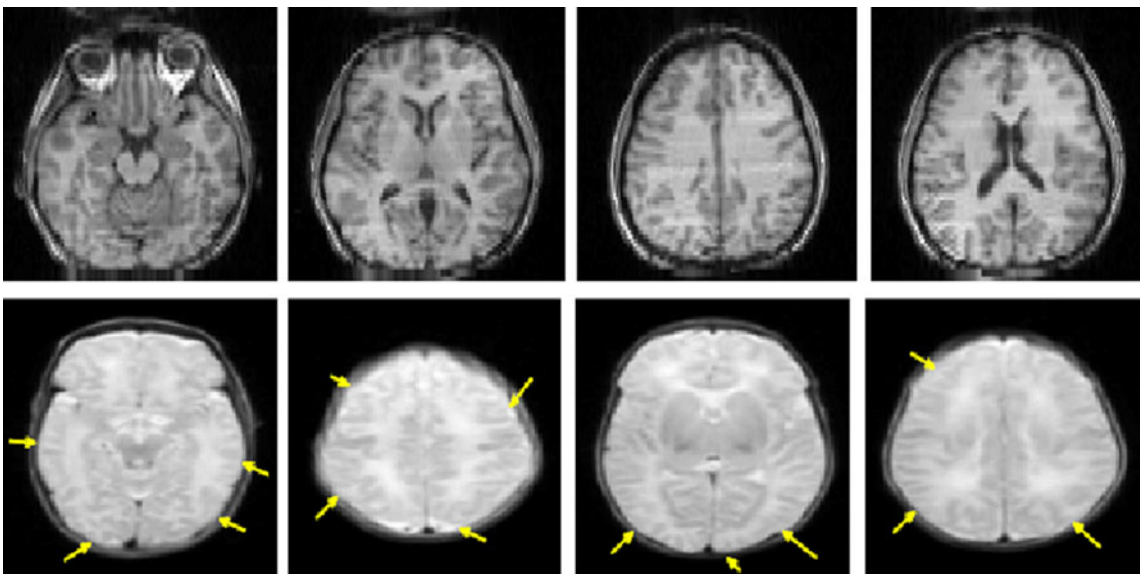

Fig. 1 Figures showing the difference between neonatal and adult brain images through different slices. The *top row* shows adult brain images while the *second row* shows neonatal images. *Yellow arrows* show those regions in neonatal images which have poor contrast, lack edge information, or ambiguous tissue information. Note that there is no one-to-one correspondence between the images in the two rows and are shown for illustration purposes

the contour with the highest likelihood according to the discriminative model. The contour is then refined using graph cuts to obtain the final segmentation. In [19], a method is proposed for segmentation of pediatric brain tumors. It combines probabilistic boosting trees (PBT) and lower-level segmentation via graph cuts. Tu et al. in [20] propose a method using PBTs for automated extraction of major sulci from brain MRI.

In this paper, we propose a method for separating the brain and non-brain from neonatal brain MR images. Graph cuts are used to segment the 3D volume and also incorporate prior shape information. The method in [14] first extracts a rough estimate of the brain and uses graph cuts to refine the segmentation. However, we use graph cuts to extract the whole brain. Prior shape information is included from labeled training data. This paper makes the following contributions: (1) graph cuts are used exclusively for brain extraction or skull stripping. (2) A simple approach is proposed to include shape information with graph cuts by constructing a prior shape atlas from manually labeled segmentations. (3) Gradient information from labeled training data is used to formulate the smoothness term which increases segmentation accuracy. A description of our method is given in "Materials and Methods." Comparative results of our algorithms with three methods, i.e., BET, BSE, and graph cut (GC) [14], are presented in "Experiments and Results." Finally, we present a brief discussion ("Discussion" section) and conclusion ("Conclusion" section).

Neonatal Brain Extraction

There are many works related to neonatal brain segmentation [21–23] which make use of conventional brain extraction tools (BSE or BET). However, neonatal brain extraction has some unique challenges not observed in adult brain MRI. As infants grow old, brain structures develop leading to less complications in brain extraction. For example, neonatal brain MRI have low tissue contrast and low image resolution. Consequently, even popular and freely available software like BET and BSE fail to accurately extract the brain. In spite of manual tuning of parameters, parts of the skull are still left connected to the brain due to the weak boundary between the brain and the skull. Neonatal MRI have very low contrast-to-noise ratio posing difficulty in segmenting regions exhibiting partial volume effect. Brain segmentation methods like [8] are able to use BSE and BET for brain extraction because their datasets consist of 1-year- and 2-year-old brain images. In older infants, brain structure is sufficiently formed for BET and BSE to give accurate segmentation results. In our datasets, the maximum age of the infants was 1 month where brain structures are not properly formed, and low-level information alone does not provide accurate extraction results.

Graph Cuts and Importance of Prior Shape Information

We choose to use graph cuts optimization because of its following advantages:

1. Graph cuts can easily find the global optima of Markov random field (MRF)-based energy function with two labels [24]. MRF-based energy functions are suitable for problems where the solution is represented as a set of labels (e.g., segmentation). MRFs enable the inclusion of context-dependent information from the pixel neighborhood and allow for a regularized solution. This ensures that neighboring pixels take similar segmentation labels and avoid isolated patches of incorrectly labeled pixels. Since our segmentation had two labels (brain and non-brain), graph cuts give a globally optimal solution in quick time.
2. Graph cuts are not sensitive to the initialization of labels. It gives a globally optimal result irrespective of the initialization and does not get trapped in local minima [24]. This provides a distinct advantage over level sets.

Many knowledge-based algorithms have been developed for neonatal brain image segmentation [8, 22, 25–28] under the guidance of an atlas encoding prior knowledge of anatomical structures, their spatial locations, shapes, and their spatial relationships. Prastawa et al. [22] generate an atlas by averaging three semi-automatic segmented neonatal brain images registered with affine transformation. Song et al. [28] built an unbiased atlas from nine out of ten neonates in a leave-one-out manner with diffeomorphic flow registration. Xue et al. in [29] use multiple age-specific atlases in an expectation-maximization framework for tissue segmentation. In MRF energy functions, contextual information is incorporated from the immediate neighborhood of a pixel. As a result, incorporating shape information in graphs is a challenging task because reliable shape information is obtained from a set of points over a larger neighborhood. The penalty term of the MRF energy is calculated for every pixel while the smoothness term considers inter-pixel interactions in the immediate neighborhood. Thus, the effectiveness of smoothness cost in including prior shape information is limited. The data penalty can be used for including prior shape information but needs a lot of training data to construct a generalized prior shape model.

The first works to use prior shape information in graph cuts were [30, 31]. In [30], the zero level set function of a shape template was used with the smoothness term to favor a segmentation close to the prior shape. Slabaugh et al. in [31] used an elliptical shape prior, under the assumption that many objects can be modeled as ellipses. They apply many iterations as a pre-initialized binary mask is updated to get the final segmentation. Vu et al. [32] use a discrete version

of shape distance functions to segment multiple objects, which can be cumbersome. A flux-maximization approach was used in [33], while in [34], the smoothness cost was modified to include star shape priors. Although there are not many works that use prior shape knowledge exclusively for brain extraction, some methods have used shape information to segment parts of the human brain like the corpus callosum [30, 35] and cerebellum [36].

Materials and Methods

Overview of the Method

A schematic of the different stages of our method is given in Fig. 2. The proposed algorithm can be divided into three stages: (1) construct the prior shape model using labeled training data; (2) determine the segmentation labels of each pixel using intensity information and graph cut optimization; and (3) using the previously obtained segmentation labels as starting point, determine the final segmentation labels using shape information and graph cuts. Each stage comprises different substages, details of which are given below.

Constructing Prior Mask

The acquired volumes were corrected for intensity inhomogeneities and affinely registered to a reference volume using the FSL toolkit (http://www.fmrib.ox.ac.uk/fsl/). An atlas prior is constructed by averaging over all the manually labeled training masks which are binary images of value 1 (brain) and 0 (non-brain). The average mask indicates the probability of a pixel belonging to the brain. Figure 3 shows different slices from a typical prior shape volume. A leave-one-out approach is used for testing 20 volumes where 19 volumes are used to construct the prior shape and the 20th volume is used for testing. All possible combinations of 19 volumes were used to construct the atlas, and the remaining volume was used for testing. Manual segmentations are available for all 20 volumes.

Intensity-Based Segmentation

MRFs are suitable for discrete labeling problems, and graph cuts can find the global optima of such MRF energy terms [24]. The energy function of a second-order MRF is given as

$$E(L) = \sum_{s \in P} D(L_s) + \lambda \sum_{(s,t) \in N} V(L_s, L_t), \quad (1)$$

where P denotes the set of pixels, L_s denotes the label of pixel s □ P, and N is the set of neighboring pixel pairs. The labels denote the segmentation class of a pixel (0 for background and 1 for object). The labels of the entire set of pixels are denoted by L. $D(L_s)$ is a unary data penalty function derived from observed data and measures how well label L_s fits pixel s. V is a pairwise interaction potential that imposes smoothness and measures the cost of assigning labels L_s and L_t to neighboring pixels s and t. λ is a weight that determines the relative contribution of the two terms. Note that both D and V consist of two terms, each incorporating intensity and shape information.

From the atlas prior, we know the areas corresponding to brain and non-brain tissues. Using this information and the original training volumes, we determine the intensity distributions of the brain and background. Since the brain area consists of different tissues, a two-component Gaussian mixture model (GMM) is used to model the intensity distribution while a single Gaussian is used for the background. Since all the volumes are normalized to have intensity values between 0 and 1, this step eliminates the need for manual selection of object and background seed points or

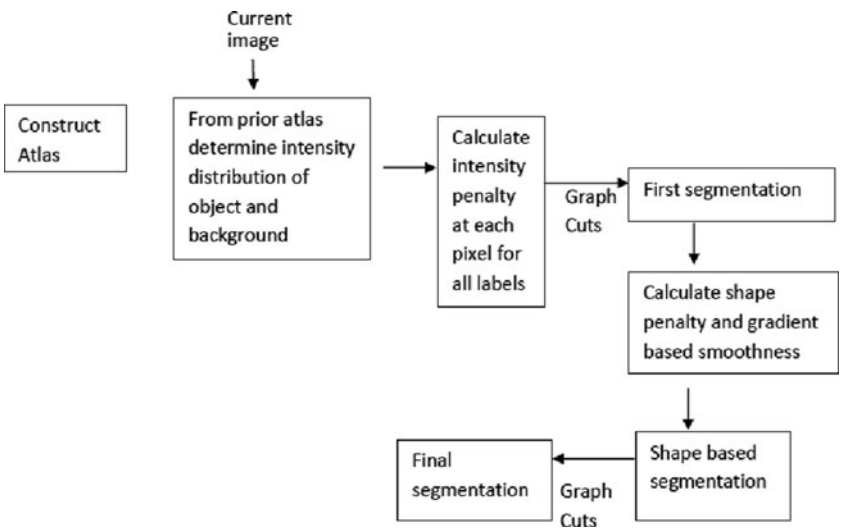

Fig. 2 Schematic diagram of steps in our shape-based segmentation method

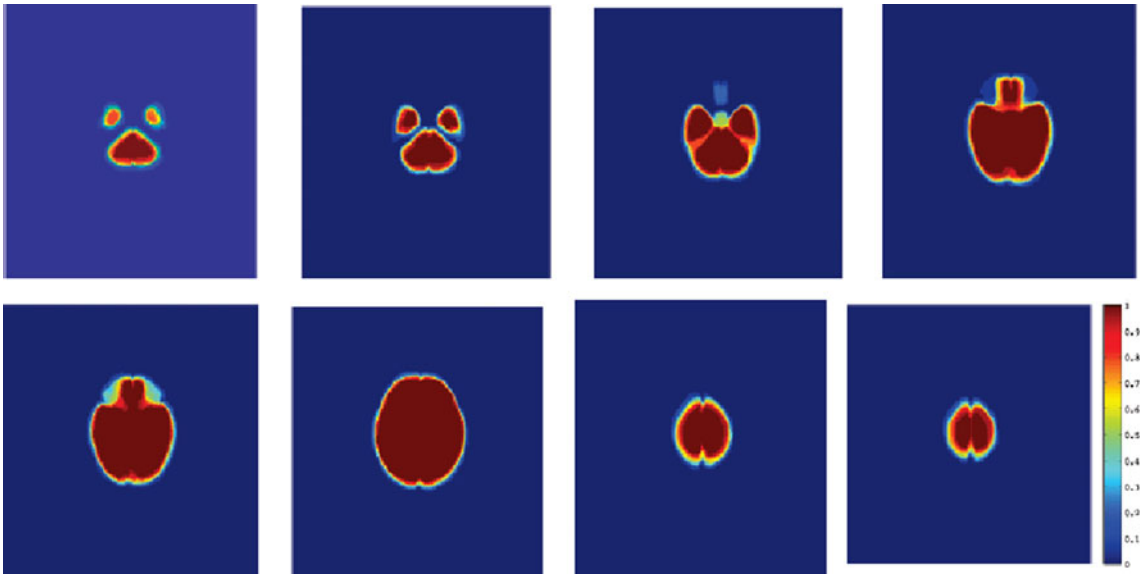

Fig. 3 Different slices from the prior volume. A higher value indicates the region is common to a greater number of training volumes. The slices are from one of the prior volumes constructed using a leave-one-out approach

patches. The penalty value, defined as the negative of the log-likelihood of its intensity, is

$$D_I(L_s) = -\log(\Pr(I_s|L_s)), \tag{2}$$

where D_I is the intensity penalty, I_s is the intensity at pixel s, L_s=1/0 (object/background) is the label, and Pr is the likelihood of I_s given the intensity distributions of L_s=1/0. Pr is obtained from the GMM modeling of intensities as described above and is a popular choice of the penalty function [24]. Note that for every pixel, there are two penalty values corresponding to the two labels. The intensity smoothness term V_I assigns a low penalty at edge points based on the intensity of neighboring pixel pairs and favors a piecewise constant segmentation result. It is defined as

$$V_I(L_s, L_t) = \left\{ \begin{matrix} e^{-\frac{(I_s - I_t)^2}{2\sigma^2}} & \cdot \frac{1}{\|s-t\|}, & L_s \neq L_t \\ 0, & & L_s = L_t \end{matrix} \right\}, \tag{3}$$

Here, σ determines the intensity difference up to which a region is considered as piecewise smooth. It is equal to the average intensity difference in a neighborhood of pixel s. $\|s-t\|$ is the Euclidean distance between s and t. Once the penalty values and interactions between different pixels have been defined, their individual segmentation labels are determined using graph cuts.

Shape Information

The prior shape (SP) incorporates shape knowledge into the penalty function. Pixels inside the brain (higher-intensity regions) have higher probability values. The penalty for object label is low (equal to zero), and the penalty for background label is high. Pixels in the lower-intensity regions (not equal to zero) have lower probability of being in brain region, and their corresponding penalty for object label is greater than zero. The penalty for pixels within the brain region of the shape prior is defined as

$$\begin{aligned} D_s(L_s = 0) &= k_1 \\ D_s(L_s = 1) &= (1 - \mathrm{SP}(s)) \times w_1 \end{aligned} \tag{4}$$

where k_1=5 is a relatively high penalty value to not favor that particular label. SP(s) is the probability value of pixel s in the shape prior SP; w_1 is a weight that determines the penalty of the pixels. Pixels within the innermost region of the shape prior will have SP(s)=1 and therefore $D_S(L_s=1)$=0 which is our objective. The penalty values for other pixels are determined by the weight w_1. Pixels having zero probability value in SP are always in the background for which the penalty values are defined as

$$\begin{aligned} Ds(L_s = 0) &= 0 \\ Ds(L_s = 1) &= k_1 \end{aligned} \tag{5}$$

Gradient Information

Another important aspect of our work is the formulation of the smoothness penalty based on gradient information. We observe that the boundary between the brain and skull is not always accurately segmented using only intensity information in the smoothness term. The gradient magnitude at the brain–skull boundary can help us distinguish the brain from

the skull. If the difference in gradient magnitude between two neighboring pixels is greater than a threshold, then they belong to different labels. Since we want the graph cut to separate pixels from different regions, we assign low weight to their connecting edges. On the other hand, a value lower than the threshold indicates the same labels for the pixel pair, and their connecting edge is given high weight. Thus, we define the smoothness cost based on gradient information as follows:

$$V_G(L_s, L_t) = \left\{ \begin{array}{c} 0.2, |g_s - g_t| \geq g_{th} \\ 5, |g_s - g_t| < g_{th} \end{array} \right\}, \quad (6)$$

Here, g_s and g_t are the gradient magnitudes of pixel s and t, respectively, g_{th} is the gradient difference threshold which determines whether two neighboring pixels are likely to take the same labels based on gradient information. To determine g_{th}, we adopt the following steps. From the labeled training masks, we extract pixel pairs lying on the brain boundary and also their immediate neighbors which lie on the skull. The difference in gradient magnitudes for these pixel pairs is calculated. This step is repeated for all training volumes where the gradient images have been normalized so that the maximum gradient magnitude is 1. The mean of the gradient differences of neighboring pixel pairs was found to be 0.15, and the maximum value was 0.25, and thus, g_{th}=0.15. To reflect the total energy function, Eq. 1 can be rewritten as

$$E(L) = \sum\nolimits_{s \in P} [w_2 D_1(L_s) + D_s(L_s)] + \lambda \sum\nolimits_{(s,t) \in N} [V_I(L_s, L_t) + V_s(L_s, L_t)], \quad (7)$$

where w_2=0.6 decides the relative contribution of intensity information to the penalty. λ=0.1 is the relative weighting between the energy terms. The above energy is optimized using graph cuts, and the output is the final segmentation labeling for each pixel.

Optimization Using Graph Cuts

A schematic of the graph structure is shown in Fig. 4, where an 8-neighborhood system is considered. Pixels are represented as nodes in a graph which consists of a set of directed edges that connect two nodes. The optimum labeling is obtained by a series of α-expansion moves that sever the edge links in such a manner that the sum of weights of severed edges is minimum. Details of graph construction and optimization can be found in [24]. The value of λ (ref. Eq. 1) is such that there is over segmentation of the brain area. This step is necessary because of the following reasons:

1. The intensity distributions estimated initially are not always accurate enough to model the entire brain volume. The segmented volume obtained from the above step is used to update the intensity distributions of the object and background.

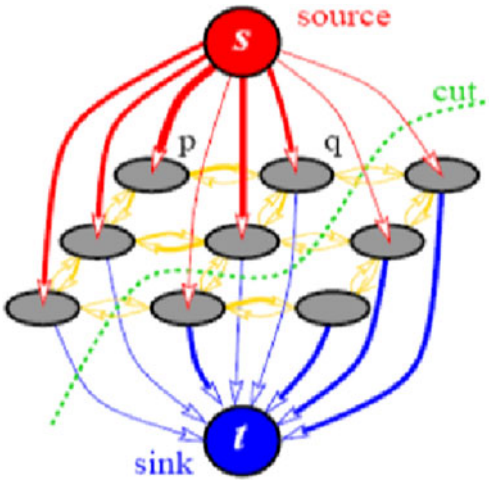

Fig. 4 Schematic of the graph structure with an 8-neighborhood system

2. Combining the updated intensity distributions with shape information, the final results are more accurate than using both shape and intensity information in the first stage.

Evaluation

The segmentations from the automated algorithms are compared with manual segmentations. Manual masks were traced by a single trained expert on all slices of the brain volumes. The expert had more than 7 years of experience in manually segmenting brain images. This manual segmentation served as the gold standard against which the performance of automated algorithms was compared. Quantitative measures for segmentation accuracy were determined according to the following metrics:

Dice metric (DM):

$$\mathrm{DM} = \frac{2|M_{\mathrm{auto}} \cap M_{\mathrm{manual}}|}{|M_{\mathrm{auto}}| + |M_{\mathrm{manual}}|} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \quad (8)$$

Jaccard index (JI):

$$\mathrm{JI} = \frac{|M_{\mathrm{auto}} \cap M_{\mathrm{manual}}|}{|M_{\mathrm{auto}} \cup M_{\mathrm{manual}}|} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \quad (9)$$

Sensitivity (Sen):

$$\mathrm{Sen} = \frac{|M_{\mathrm{auto}} \cap M_{\mathrm{manual}}|}{M_{\mathrm{manual}}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad (10)$$

Specificity (Spe):

$$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \quad (11)$$

False positive rate (Fpr):

$$\mathrm{Fpr} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{12}$$

where M_{auto} is the mask obtained from one of the automated segmentation methods, M_{manual} is the mask obtained by manual segmentation, TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative. Specificity measures the fraction of pixels correctly labeled as background, while Fpr gives an indication of how many pixels identified as brain by the automatic method were outside the manual mask.

Hausdorff distance DM gives a measure of how much the actual manual segmentation was recovered by the automatic segmentation. But the boundaries of the segmented regions may be far apart. The Hausdorff distance (HD) aims to measure the distance between the contours corresponding to different segmentations. We follow the definition of HD as given in [37]. If two curves are represented as sets of points $A=\{a_1, a_2,\ldots\}$ and $B=\{b_1, b_2,\ldots\}$, where each a_i and b_i is an ordered pair of the x and y coordinates of a point on the curve, the distance to the closest point (DCP) for a_i to the curve B is defined

$$d(a_i, B) = \min_i \left| b_j - a_i \right| \tag{13}$$

The HD is defined as the maximum of the DCPs between the two curves [38].

$$\mathrm{HD}(A, B) = \max\left(\max_i d(a_i, B), \max_j d\left(b_j, A\right)\right)_i. \tag{14}$$

Experiments and Results

We used 20 neonatal brain MR volumes for our experiments. The average age of the neonates was 22±3 days (maximum, 30 days). The data were acquired from a 3T Siemens scanner. T2 images of 60 axial slices were obtained with imaging parameters: T=7,380 ms, TE=119 ms, flip angle=150°, acquisition matrix=256×256, and resolution=0.95×0.95×1 mm^3. The volumes were made available after the above steps. The parameter values are k_1=5, w_1=2, λ=0.1, and w_2=0.6. For other datasets, different values may give optimal results. The rationale behind the choice of parameters is presented in "Influence of Parameters k_1, w_1, w_2." The image acquisition process was in accordance with the institutional guidelines and had prior approval of the appropriate ethics committee.

Segmentation Accuracy for Simulated SNR Values

Different levels of noise were added to the dataset using the *imnoise* function in MATLAB to simulate different signal-to-noise ratio (SNR) values. The original SNR value of images was in the range 35–46. Table 1 summarizes the segmentation performance at different SNR values. With the increase in added noise, DM decreases although slightly. To test the similarity between the new segmentation measures and those obtained from the original images, we performed Student's t test between the average DM values. With p>0.14, the values are statistically similar. However, the Fpr increases by 0.8, and p=0.04. Since DM is a more reliable indicator of segmentation performance, we can assume that our method is quite robust to added noise.

Comparison with Other Methods

We compared the performance of our method (graph cut using shape priors, GCSP) with four other brain extraction techniques, BET [3], BSE [1], the HWA [6] available as part of [10], and GC, the graph cut-based skull stripping method of [14]. BET is used from the MIPAV software package with the following parameters: 1,500 iterations, depth=11, image influence=0.3, and stiffness=0.3. BSE was applied to the entire brain volume using BrainSuite09 (Laboratory of Neuroimaging, UCLA, CA, USA) with the following parameters: diffusion iterations=3, diffusion constant=25, edge constant=0.064, and erosion size=1 pixel. For GC, the default parameters were used, i.e., threshold=36 and importance of intensity=2.3. Default values in [6] were used for HWA, i.e., preweight the input image using atlas information=0.82, use the basins merging atlas information=0.32, presize the preflooding height=10%, use the preweighting for the template deformation=true, use template deformation using atlas information=true, use seed points using atlas information=true. Segmentation accuracy was determined by calculating the mean and standard deviations of the measures described in "Optimization Using Graph Cuts." To test for significant difference between the results of the automatic methods, we used a paired Student's t test with p values below 0.05 considered significant.

The first row in Table 2 shows comparison of the manual segmentation with the prior shape. The aim behind this comparison was to show that the prior shape is very different from the automated segmentations (as

Table 1 Average DM and FPR values for test volumes at different SNR values

SNR values	Dice metric	FPR
Original SNR=35–46	98.9	3.8
SNR=10	98.1	4.0
SNR=5	98.0	4.5

Table 2 Average quantitative measures on neonatal images for different automatic segmentation methods. Values shown are obtained using the "leave-one-out" approach

	DM (%)	HD (mm)	FPR (%)	JI	Spe (%)	Sen (%)
Prior	90.3±0.9	6.4±3.5	21.3±3.5	82.3±0.7	91.7±1.3	98.5±0.9
$GCSP_{NG}$	97.8±1.1	1.3±0.3	4.5±2.3	95.7±0.9	97.1±1.1	96.1±1.4
GCSP	98.9±0.4	0.9±0.2	3.8±1.9	97.8±0.5	98.8±0.7	98.1±1.1
GC	93.4±1.5	3.4±1.3	11.5±4.6	87.6±1.1	93.1±1.2	92.9±1.0
BET	96.1±1.8	1.9±0.5	7.3±2.5	92.5±1.9	96.0±1.2	97.2±1.4
BSE	95.1±1.7	2.2±0.4	8.6±3.2	90.7±1.6	92.7±1.4	94.2±1.3
HWA	97.2±1.5	1.5±0.5	5.1±2.2	94.0±0.8	96.6±1.1	97.1±1.2

observed from low DM and high HD values). Thus, in no way does the prior bias the segmentations. It only serves to give an idea of which pixels are likely to be inside or outside the brain. Since the prior mask is constructed from 19 volumes, there are many regions within the mask which may correspond to non-brain areas in the new volume. $GCSP_{NG}$ denotes our method having no gradient-based smoothness term V_G. GCSP gives the best results due to the use of shape knowledge. The shape prior gives valuable information based on which the cost function was designed to achieve accurate segmentation. Availability of labeled training data allows inclusion of gradient information leading to an improvement in segmentation accuracy. This increase in segmentation accuracy can be said to be significant, as the p value between GCSP and $GCSP_{NG}$ is 0.04.

Table 3 shows the p values between the DM and HD values of different methods. The first four rows show the p values between the shape prior and different automatic methods. With p values less than 0.01, the obtained *DM* values are quite different from the prior shape. The same is also true for HD values. The p values also indicate that the segmentations of different methods are quite different, except for BSE and BET which have very similar DM and HD values.

Figures 5 and 6 show segmentation results for neonatal MRI. Figure 5 shows the results for GCSP, $GCSP_{NG}$, and GC. GC shows the least accurate segmentation as it relies only on intensity information. As a result, areas outside of the brain having similar intensity or weak boundaries are also segmented leading to lower DM values. In the absence of gradient information, $GCSP_{NG}$ exhibits oversegmentation at the boundary between the brain and skull. These areas are highlighted by yellow arrows in Fig. 5. GCSP shows improved segmentation accuracy due to the inclusion of gradient information. Figure 6 shows the segmentation results using BET, BSE, and HWA. These methods have been designed for adult brain images and do not address specific challenges of the neonatal brain MRI like weak gradients, ill-defined structures, and low contrast-to-noise ratio. Thus, their segmentation accuracy is lower than GCSP. This highlights the importance of developing methods specifically for neonatal brain images and also justifies the importance of including shape and edge information in the graph cut framework.

Table 3 p values for Dice metric and Hausdorff distance measures

	DM	HD
p (Pr-$GCSP_{NG}$)	<0.001	<0.001
p (Pr-GCSP)	<0.001	<0.001
p (Pr-BET)	<0.01	<0.01
p (Pr-BSE)	<0.01	<0.01
p (GCSP-GC)	0.01	0.02
p (GCSP-BET)	0.02	<0.01
p (GCSP-HWA)	0.03	<0.01
p (GCSP-BSE)	0.01	<0.01
p (BSE-BET)	0.2	0.34

p (GC-$GCSP_{NG}$) is the significance value between results of GC and $GCSP_{NG}$ and so on for other methods

Influence of Parameters k_1, w_1, w_2

Figure 7a shows the DM values for different values of k_1. When k_1 is closer to zero, the DM values are lower. For k_1= 1 the penalty is almost equal to the penalty for the other label (which has been set to zero). This indicates a soft constraint on the pixel labels. Graph cut being a global optimization technique assigns labels based on penalty and smoothness costs. If the penalty corresponding to the two labels has similar values, then the smoothness cost has a greater role to play in final label assignment. Thus, it is possible that some pixels may be assigned wrong labels. We observe that for $k_1 \leq 2$, the DM values are low compared to actual results. But for $k_1 \geq 3$, the DM values increase and give the best results for $k_1 > 5$ when no change in DM values is observed. w_1 determines the penalty for pixels within the brain mask, except for the innermost pixels. The value of w_1 should ensure that the penalty values are not very close to k_1, and at the same time, they are greater than zero. We find a value of $w_1 = 2$ to satisfy the above requirements and gives the best results.

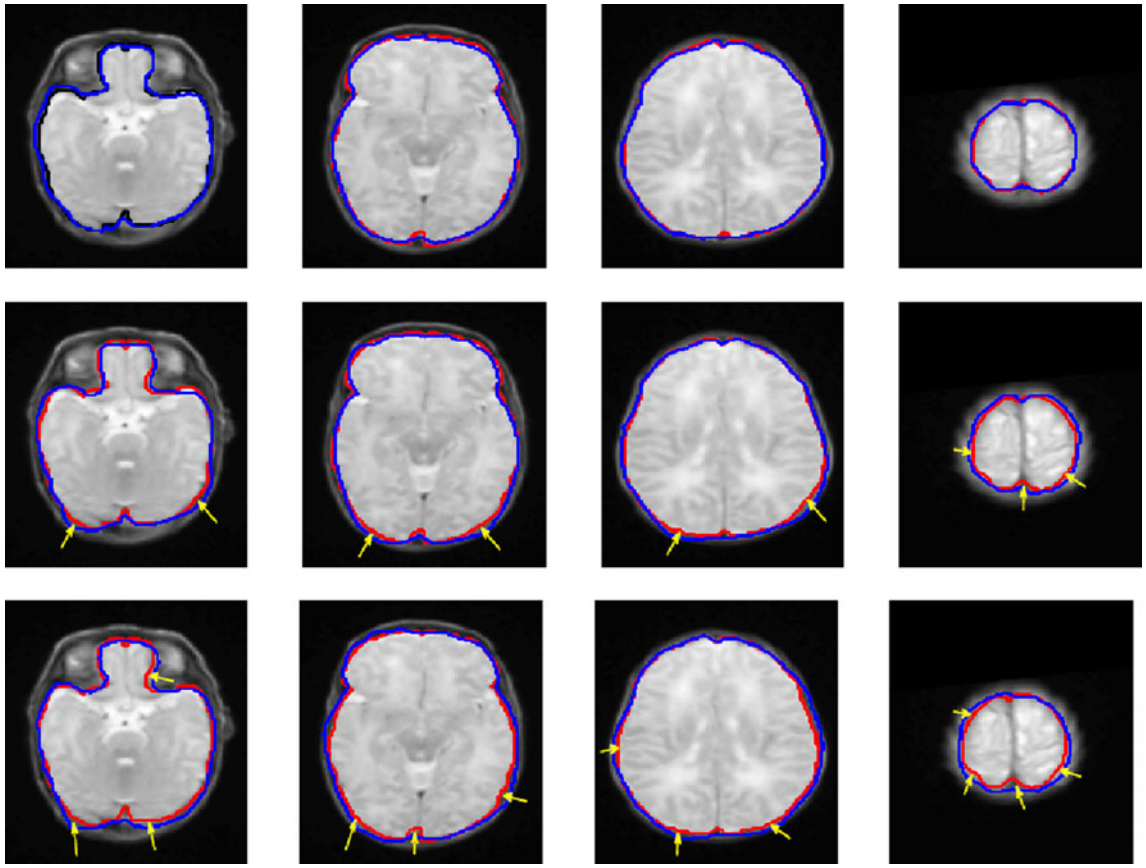

Fig. 5 Comparative segmentation results. The *first row* shows results for $GCSP$, the *second row* shows results for $GCSP_{NG}$, and the third row shows results for GC. *Yellow arrows* indicate areas of inaccurate segmentation

w_2 decides the relative contribution of D_I to the total penalty. If $w_2>0.9$, then its values are high compared to the shape penalty, and the final labels are influenced solely by the intensity penalty. On the other hand, if $w_2<0.3$, shape penalty dominates and intensity penalty has no influence on the final labels. $w_2=0.6$ is empirically chosen (using three datasets) to give the best results. As observed from Fig. 7c, the DM values start from a low value for $w_2<0.3$ (when D_S dominates), increases, and remains steady for $0.3\leq w_2\leq 0.6$, and then decreases for $w_2>0.6$ (when D_I dominates). Note that the value of w_2 will depend upon the definitions of the terms D_I and D_S. If they are defined differently, then optimal results may be obtained with different values of w_2. The values of k_1, w_1, and w_2 were individually varied (keeping the other two parameters fixed at their optimal values) and the corresponding segmentation results obtained for different slices. Figure 7d–f shows segmentation results for the slice shown in the first column of Figs. 5 and 6. Two sets of results are shown for each parameter where the values are greater and lesser than the optimal values.

Figure 8 shows the average DM values for test datasets when different numbers of training volumes are used to construct the atlas prior. When the number of volumes is less than eight, the DM values are less than 90 as the number of training volumes is not sufficient to capture the change in shape over all datasets. As we include more volumes to construct the shape prior, there is an increase in the DM. When the number of training volumes exceeds 13, the DM values do not show a large increase, indicating that a majority of the shape variations over the dataset have been incorporated with the training volumes. This observation is not limited to a specific combination of training volumes but is true for all cases.

Results for Adult Brain Images

Figure 9 shows brain extraction results for three slices of adult brain images. The dataset consists of 77 T1-weighted scans of the cross-sectional MRI dataset of the OASIS project (http://www.oasis-brains.org/). The population consists of 55 females and 22 males, aged 51.64±24.67 years. Twenty subjects were evaluated as "demented and probable Alzheimer's disease." The scans were acquired on a 1.5T Siemens scanner with a MP-RAGE sequence, TR/TE/TI/TD=9.7/4.0/20/200 ms, flip angle=10°. Sagittal slices were acquired 1.5 mm apart with

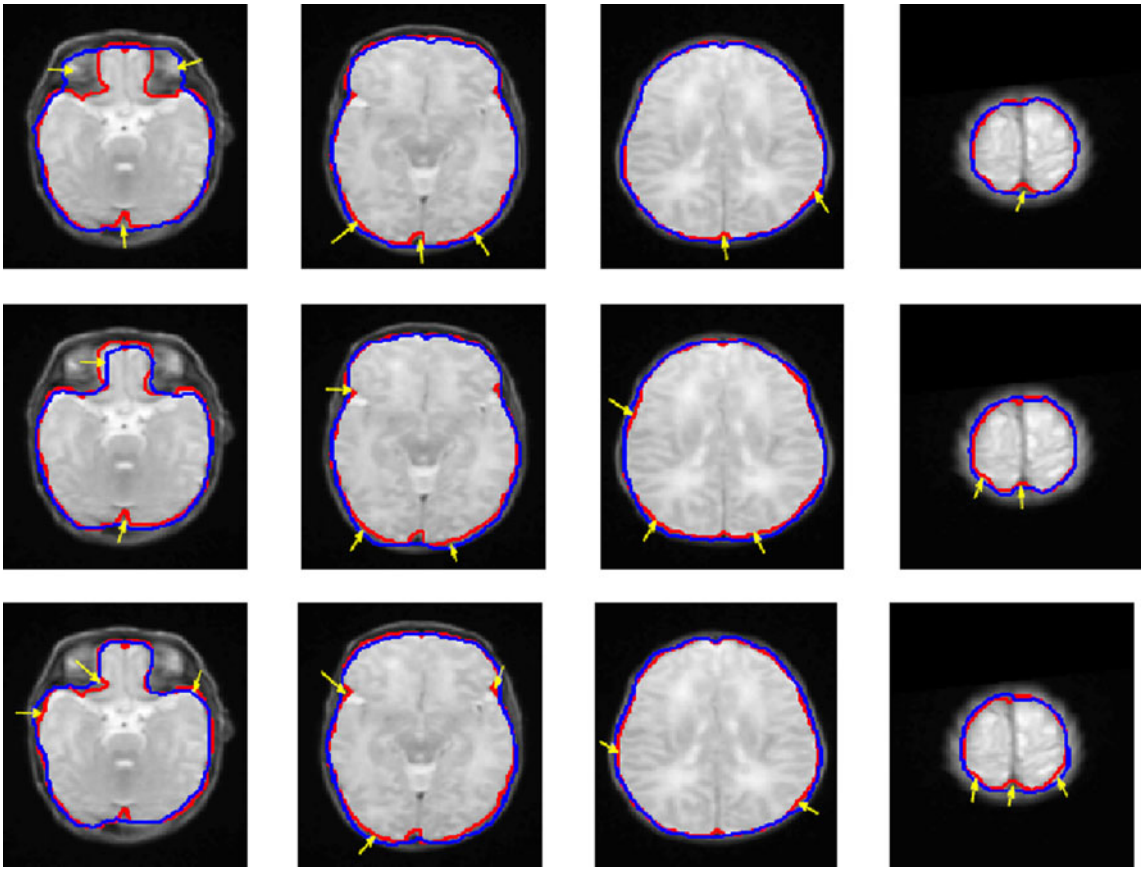

Fig. 6 Comparative segmentation results. The *first row* shows results for BET, the *second row* shows results for BSE, and the *third row* shows results for HWA. *Yellow arrows* indicate areas of inaccurate segmentation

in-plane resolution of 1 mm. The brain masks for this set were not manually delineated; instead, the brain was segmented with an in-house method based on registration to an atlas. However, the output from the method was reviewed by human experts before releasing the data, so the quality of the masks is good enough at least to test the robustness of a method. Despite this lack of exactitude, this dataset is very valuable because it includes scans from a very diverse population with a very wide age range as well as diseased brains.

Results are shown for all the six previously mentioned brain extraction algorithms. To construct the shape prior for GCSP, we use the manual segmentations of ten volumes, and these ten volumes were excluded from all experiments. Except for GCSP, all other algorithms have been designed specifically for adult brain images. As a result, all of them show very high segmentation accuracy in terms of high DM values (Table 4). In fact, there is not much to choose from their respective performances. In the case of GCSP, the segmentation accuracy is slightly lower than the other methods, but is still quite high. From Fig. 9, we see that in the first row, GCSP results in oversegmentation at certain places. However, the result for other slices is as good as the other methods. Another interesting observation is that there is not much difference in the results of GCSP and $GCSP_{NG}$ ($p>0.1$). This indicates that the use of gradient information in smoothness terms is not as critical for the adult brain as it is for neonatal brain data.

Discussion

With the help of GCSP, we were unable to separate parts of the skull in two datasets. Since the skull and brain are connected by very weak gradients, it is difficult to separate them in some slices. Although we include gradient information in the smoothness constraint, there is no significant change in segmentation accuracy for these volumes. This is due to weak edges between brain and skull which the training data are unable to account for. Thus, a more robust approach is necessary to incorporate edge information. Another limitation of our method is its reliance on training data. While our method still performs well with a reduced number of training volumes (Fig. 8), a sufficient number of labeled segmentation results may not always be available for constructing the prior shape model. Our

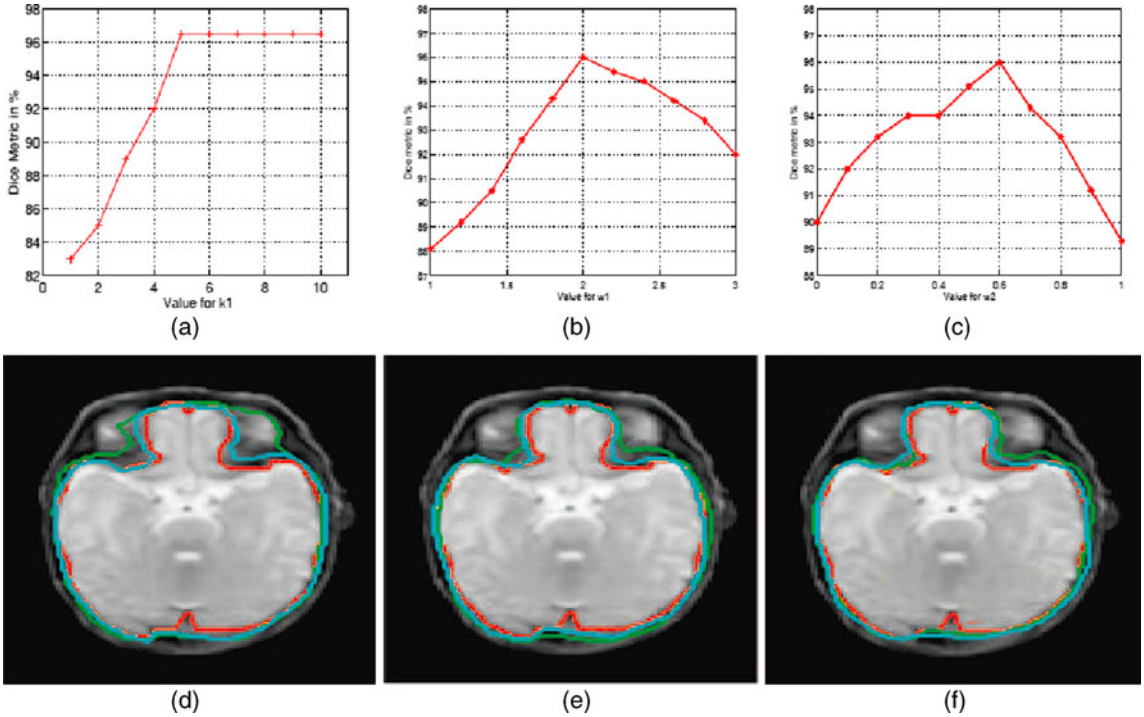

Fig. 7 Plots showing variation of DM values with different parameters: **a** k_1, **b** w_1, and **c** w_2. Segmentation results for different values of: **d** k_1, the *green contour* shows results for $k_1=1$ and the *cyan contour* for $k_1=4$; **e** w_1, the *green contour* shows results for $w_1=1{:}5$ and the *cyan contour* for $w_1=2{:}5$; **f** w_2, the *green contour* shows results for $w_2=0{:}1$ and the *cyan contour* for $w_2=0{:}8$. The slice is the same as shown in the first column of Figs. 5 and 6. The *red contour* shows the manual segmentations

datasets were acquired from the same scanner. Consequently, it was easier to build the prior shape model. We would like to improve our method to fuse data from different scanners and build a robust shape model.

For neonatal datasets, brain extraction methods designed for adult brain MRI show inferior performance than our method due to the absence of shape information (ref. Table 2). Neonatal brain MRI does not provide very significant edge information, which hampers the performance of BSE and BET. Although GC is a graph cut-based approach, it does not include shape knowledge in its formulation. One important stage of GC is the estimation of an initial mask by foreground seed identification and region growing. While foreground seed identification is usually accurate, region growing may lead to an over-segmented mask. Thus, the effectiveness of the narrow connection removal step (which separates brain from skull along the weak gradient) is reduced leading to an inaccurate mask and hence lower DM values. Amongst HWA, BET, BSE, and GC, HWA is the most robust and accurate as it combines watershed segmentation with atlas information. Information derived from the atlas is used to refine the segmentation and thus achieve good segmentation accuracy.

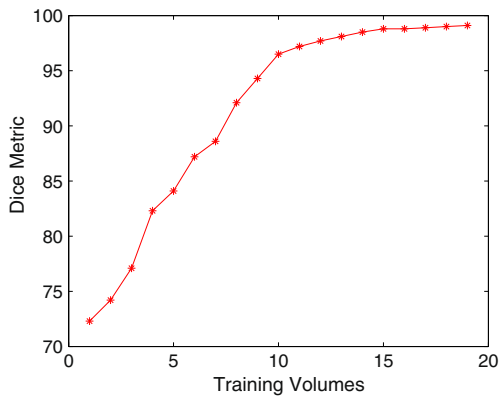

Fig. 8 DM values for varying volumes used to construct the atlas prior

The quantitative results in Table 4 show that for adult brain images, the segmentation results obtained using GCSP are nearly as good as the other methods (BSE, HWA, GC, and BET). This is despite the fact that our method is not designed specifically for adult MRI. All the other methods have been designed to specifically extract adult brain images by making use of well-defined boundaries and relationship between intensity distributions of different tissue types (which cannot be reliably obtained for neonatal volumes). In the future, we aim to adapt our algorithm for adult brain extraction by making use of the information about different tissue intensity distributions.

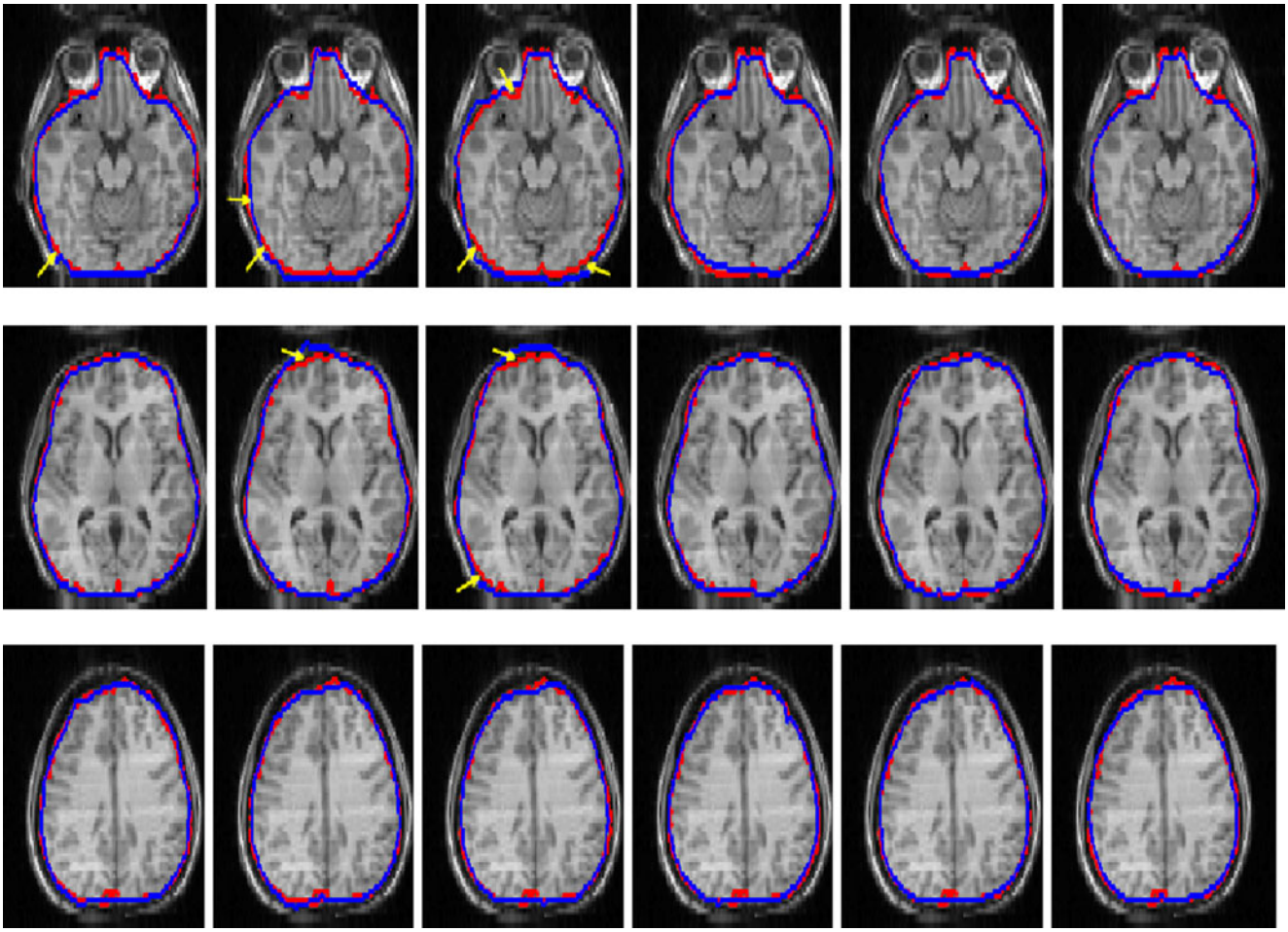

Fig. 9 Comparative segmentation results for adult brains. The *first column* shows results for GCSP; the second column shows results for $GCSP_{NG}$; the *third column* shows results for GC; the *fourth column* shows results for BET; the *fifth column* shows results for BSE; and the *sixth column* shows results for HWA

Conclusion

In this paper, we have proposed a novel shape prior segmentation method using graph cuts for skull stripping in neonatal brain MR images. The shape prior is constructed from a set of labeled training images using a leave-one-out approach. The shape prior is a probability map which gives the probability of a pixel belonging to the brain. Neonatal brain MR images pose difficulty in separating the boundary between the brain and skull, which motivated us to use gradient information in the smoothness cost for accurate identification of brain boundary. Compared to popular techniques like BET, BSE, and HWA, our method shows higher segmentation accuracy (based on popular metrics) due to the inclusion of prior shape information. In spite of being designed for neonatal brain images, our method also shows good segmentation accuracy for adult brain images as well.

Table 4 Average quantitative measures on adult brain images for different automatic segmentation methods

	DM (%)	HD (mm)	FPR (%)	JI	Spe (%)	Sen (%)
Prior	89.1±1.0	9.2±2.9	25.6±4.1	80.3±0.9	90.2±1.9	97.2±1.2
$GCSP_{NG}$	93.2±1.0	2.6±0.6	6.2±2.1	90.3±1.1	96.1±1.3	96.2±1.5
GCSP	94.1±0.7	2.2±1.0	5.7±1.6	88.9±0.9	96.5±0.9	97.3±0.6
GC	92.1±1.1	3.7±1.5	12.1±4.2	85.4±1.5	92.8±1.3	94.1±1.0
BET	95.8±1.6	1.5±2.6	3.9±1.9	91.9±2.3	96.9±1.5	97.9±1.1
BSE	94.8±1.3	1.6±0.8	4.2±2.1	90.1±1.8	94.1±1.9	96.1±1.2
HWA	96.1±1.1	1.3±0.3	3.2±1.5	92.5±1.2	97.1±1.1	97.8±0.9

Values shown are obtained using the "leave-one-out" approach

Acknowledgments The author would like to thank the Clinical Imaging Research Center (CIRC), National University Hospital (NUH), Singapore, for providing the neonatal datasets.

References

1. Shattuck, D., Sandor-Leahy, S., Schaper, K., Rottenberg, D., Leahy, R.: Magnetic resonance image tissue classification using a partial volume model. Neuroimage 13:856–876, 2001
2. Hahn H, Peitgen H: The skull stripping problem in MRI solved by a single 3D watershed transform. In: MICCAI, 2000, pp 134–143
3. Smith S.: Fast robust automated brain extraction. Hum. Brain Mapp. 17:143–155, 2002
4. Zhuang, A., Valentino, D., Toga, A.: Skull stripping magnetic resonance brain images using a model-based level set. Neuroimage 32:79–92, 2006
5. Huang A, Abugharbieh R, Ram R, Trabouslee A: MRI brain extraction with combined expectation maximization and geodesic active contours. In: IEEE Intl Symposium in Signal Processing and Information Technology, 2006
6. Segonne, F., Dale, A., Busa, E., Glesner, M., Salat, D., Hahn, H., Fischl, B.: A hybrid approach to the skull stripping problem in mri. Neuroimage 22:1060–1075, 2004
7. Huppi, P., Warfield, S., Kikinis, R., Barnes, P., Zientara, G., FA, F. J., Tsuji, M., Volpe, J.: Quantitative magnetic resonance imaging of brain development in premature and mature newborns. Ann. Neurol 43(2):224–235, 1998
8. Shi, F., Fan, Y., Tang, S., Gilmore, J., Lin, W., Shen, D.: Neonatal brain image segmentation in longitudinal mri studies. Neuroimage 49:391–400, 2010
9. Afni [online]. Available at: http://afni.nimh.nih.gov. Accessed 13 June 2011
10. Freesurfer [online]. Available at: http://surfer.nmr.mgh.harvard.edu. Accessed 15 June 2011
11. Atkins, M., Mackiewich, B.: Fully automatic segmentation of the brain in mri. IEEE Trans Med Imag 417:98–107, 1998
12. Kapur, T., Grimson, W., Wells, W., Kikinis, R.: Segmentation of brain tissue from magnetic resonance images. Med Image Anal 1 (2):109–127, 1996
13. Lemieux, L., Hag emann, G., Krakow, K., Woermann, F.: Fast, accurate and reproducible automatic segmentation of the brain in t1-weighted volume mri data. Magn.Reson. Med. 42:127–135, 1999
14. Sadananthan, S., Zheng, W., Chee, M., Zagorodnov, V.: Skull stripping using graph cuts. Neuroimage 49:225–239, 2010
15. Park, J., Keller, J.: Snakes on the watershed. IEEE Trans. Pattern Anal. Mach. Intelli. 23:1201–1205, 2001
16. Zeng, X., Staib, L., Schultz, R, Duncan, J.: Segmentation and measurement of the cortex from 3d mr images using coupled surfaces propagation. IEEE Trans. Med. Imag. 18(10):100–111, 1999
17. Rehm, K., Schaper, K., Anderson, J., Woods, R., Stoltzner, S., Rottenberg, D.: Putting our heads together a consensus approach to brain/non-brain segmentation in t1-weighted MR volumes. Neuroimage 22:1262–1270, 2004
18. Iglesias, J., Liu, C.Y., Thompson, P., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans. Med. Imag. 30(9):1617–1634, 2011
19. Wels M, Carneiro G, Aplas A, Huber M, Hornegger J, Comaniciu D: A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-d MRI. In: MICCAI, 2008, pp 67–75
20. Tu, Z., Zheng, S., Yuille, A., Reiss, A., Dutton, R., Lee, A., Galaburda, A., Dinov, I.Thompson, P., Toga, A.: Automated extraction of the cortical sulci based on a supervised learning approach. IEEE Trans. Med. Imag. 26(4):541–552, 2007
21. Shi, F., Yap, P.T., Fan, Y., Gilmore, J., Lin, W., Shen, D.: Construction of multi-regionmulti-reference atlases for neonatal brain mri segmentation. Neuroimage 51:684–693, 2010
22. Prastawa, M., Gilmore, J., Lin, W., Gerig, G.: Automatic segmentation of mr images of the developing newborn brain. Med. Image Anal. 9:457–466, 2005
23. Weisenfeld, N.,Warfield, S.: Automatic segmentation of newborn brain mri. Neuroimage 47:564–572, 2009
24. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23(11):1222–1239, 2001
25. Bazin, P., Pham, D.: Topology-preserving tissue classification of magnetic resonance brain images. IEEE Trans. Med. Imag. 26:487–496, 2007
26. Nishida, M., Makris, N., Kennedy, D., Vangel, M., Fischl, B., Krishnamoorthy, K., Caviness,V., Grant., P.: Detailed semiautomated mri based morphometry of the neonatal brain: preliminary results. NeuroImage 32:1041–1049, 2006
27. Pham, D., Prince, J.: An adaptive fuzzy c-means algorithmfor image segmentation in the presence of intensity inhomogeneities. Pattern Recogn. Lett. 20:57–68, 1999
28. Song Z, Awate S, Licht D, Gee J: Clinical neonatal brain mri segmentation using adaptive nonparametric data models and intensity-based markov priors. In: MICCAI, 2007, pp 883–890
29. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A., Rueckert, D., Hajnal, J.: Automatic segmentation and reconstruction of the cortex from neonatal mri. Neuroimage 38:461–477, 2007
30. Freedman D, Zhang T: Interactive graph cut based segmentation with shape priors. In: CVPR, 2005, pp 755–762
31. Slabaugh G, Unal G: Graph cuts segmentation using an elliptical shape prior. In: ICIP, 2005, pp 1222–1225
32. Vu N, Manjunath B: Shape prior segmentation of multiple objects with graph cuts.In: CVPR, 2008
33. Chittajallu D, Shah S, Kakadiaris I: A shape driven mrf model for the segmentation of organs in medical images. In: CVPR, 2010, pp 3233–3240
34. Veksler O: Star shape prior for graph cut segmentation. In: ECCV, 2008, pp 454–467
35. Chen, Y., Tagare, H., S. Thiruvenkadam, Huang, F., Wilson, D., Gopinath, K., Briggs, R., Geiser, E.: Using prior shapes in geometric active contours in a variational framework. Intl. J. Comp. Vis. 50(3):315–328, 2002
36. Diedrichsen, J., Balsters, J., Flavell, J., E, E.C., Ramnani, N.: A probabilistic mr atlas of the human cerebellum. NeuroImage 46 (1):39–46, 2009
37. Chalana, V., Kim, Y.: A methodology for evaluation of boundary detection algorithms on medical images. IEEE Trans. Med. Imag. 16(5):642–652, 1997
38. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the hausdorff distance. IEEE Trans. Pattern Anal. Machine Intell. 15(9):850–863, 1993