Published in final edited form as:
Neuroimage. 2010 June ; 51(2): 684–693. doi:10.1016/j.neuroimage.2010.02.025.

Construction of Multi-Region-Multi-Reference Atlases for Neonatal Brain MRI Segmentation

Feng Shi[1], **Pew-Thian Yap**[1], **Yong Fan**[1], **John H. Gilmore**[2], **Weili Lin**[3], and **Dinggang Shen**[1]

[1] IDEA Lab, Department of Radiology and BRIC, University of North, Carolina at Chapel Hill

[2] Department of Psychiatry, University of North Carolina at Chapel Hill

[3] MRI Lab, Department of Radiology and BRIC, University of North, Carolina at Chapel Hill

Abstract

Neonatal brain MRI segmentation is a challenging problem due to its poor image quality. Atlas-based segmentation approaches have been widely used for guiding brain tissue segmentation. Existing brain atlases are usually constructed by equally averaging pre-segmented images in a population. However, such approaches diminish local inter-subject structural variability and thus lead to lower segmentation guidance capability. To deal with this problem, we propose a multi-region-multi-reference framework for atlas-based neonatal brain segmentation. For each region of a brain parcellation, a population of spatially normalized pre-segmented images is clustered into a number of sub-populations. Each sub-population of a region represents an independent distribution from which a regional probability atlas can be generated. A selection of these regional atlases, across different sub-regions, will in the end be adaptively combined to form an overall atlas specific to the query image. Given a query image, the determination of the appropriate set of regional atlases is achieved by comparing the query image regionally with the reference, or exemplar, of each sub-population. Upon obtaining an overall atlas, an atlas-based joint registration-segmentation strategy is employed to segment the query image. Since the proposed method generates an atlas which is significant more similar to the query image than the traditional average-shape atlas, better tissue segmentation results can be expected. This is validated by applying the proposed method on a large set of neonatal brain images available in our institute. Experimental results on a randomly selected set of 10 neonatal brain images indicate that the proposed method achieves higher tissue overlap rates and lower standard deviations (SDs) in comparison with manual segmentations, i.e., 0.86 (SD 0.02) for GM, 0.83 (SD 0.03) for WM, and 0.80 (SD 0.05) for CSF. The proposed method also outperforms two other average-shape atlas based segmentation methods.

1 Introduction

In the field of brain image analysis, atlases are widely used for understanding the variations of brain anatomy (Brodmann, 1909; Shattuck et al., 2008), for driving population image registration (Joshi et al., 2004), and for embedding tissue distributions in guiding brain segmentation (Ashburner and Friston, 2005). Besides intensity, an atlas also encodes

Corresponding author: Dinggang Shen, Associate Professor, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, MRI Building, CB #7515, 106 Mason Farm Road, Chapel Hill, NC 27599, dgshen@med.unc.edu, Phone: 919-966-3535, Fax: 919-843-2641.

locations, shapes, and correlation of anatomical structures. In the context of brain tissue segmentation, an atlas generally serves as an image model which provides intensity and shape information for atlas-to-subject registration, and also tissue prior knowledge in the form of probabilistic maps of gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Atlas-based segmentation has been a standard technique especially for segmenting neonatal brain images which typically have relatively poor image quality due to their inherently low spatial resolution, insufficient tissue contrast, and ambiguous tissue intensity distributions (Prastawa et al., 2005; Shi et al., 2010).

1.1 Related work

An atlas can be generated either by manually or automatically segmenting an individual image, or by learning representative information from multiple individual subjects (Rohlfing et al., 2004a). Using an individual subject as the atlas is the most straight-forward strategy. From a population of images, this type of atlas can be randomly selected or chosen by certain criteria such as image similarity or demography. Shi et al. (Shi et al., 2010) proposed a new approach for neonatal brain segmentation by utilizing as atlas a later time-point image of the same subject. This approach takes advantage of the fact that brain gyrification remains similar during postnatal development for full-term infants. In the more general case where there is no longitudinal follow-up, using an individual subject as the atlas for segmentation will inevitably lead to bias. Therefore, the question of how to better leverage population inform for improving segmentation accuracy has attracted a lot of research interest.

Atlas construction methods can be roughly grouped into two categories: average-shape atlas methods (Prastawa et al., 2005; Song et al., 2007; Xue et al., 2007) and multi-classifier decision fusion methods (Aljabar et al., 2009; Heckemann et al., 2006; Rohlfing et al., 2004b; Warfield et al., 2004). An average-shape atlas (also called a population atlas) is constructed by averaging anatomical structures of a number of spatially aligned subjects. For example, Prastawa et al. (Prastawa et al., 2005) constructed an atlas by averaging 3 semi-automatically segmented images after affine alignment. Weisenfeld et al. (Weisenfeld et al., 2006) obtained an unbiased atlas by averaging the probability maps of 20 subjects, which were non-rigidly aligned using group-wise registration. Note, however, that these atlases were built by averaging with equal weight a group of spatially-normalized images equally. Indiscriminately averaging the images and ignoring subtle inter-subject variability will lead to a blurred atlas, thus reducing its capacity to guide segmentation, especially when fine anatomical structures in the highly folded cortical regions of neonatal images are of concern.

In multi-classifier decision fusion methods, multiple subjects in a population are selected as individual atlases to independently guide segmentation. All segmentation results from different atlases can then fused by a majority voting rule (Heckemann et al., 2006), a weighted voting (Isgum et al., 2009), or more sophisticated methods like STAPLE (Rohlfing et al., 2004b; Warfield et al., 2004). This way, the single atlas induced bias can potentially be compensated. Weisenfeld et al (Weisenfeld and Warfield, 2009), for instance, applied this method for the segmentation of newborn brain MRI. However, a major concern about this method is its high computation cost due to the need to perform multiple segmentations (Han et al., 2009).

Recently, Aljabar et al. (Aljabar et al., 2009) demonstrated that an atlas generated by choosing a subject with similar anatomical structures with the query image gives better segmentation performance than by random selection. Xue et al. (Xue et al., 2007) took a different approach which factored in age-related inter-subject variation by separating the subjects into three groups according to their gestational age, and selecting one of the subjects randomly in each group as the atlas. These studies indicate that anatomical similarity between the atlas and the query image is a crucial factor for improving

segmentation accuracy. Blezek et al. (Blezek and Miller, 2007) further suggest that a single atlas may not sufficiently characterize shape variation in a population, and if the inherent shape distribution of the population is multi-modal, multiple atlases may be necessary. To summarize, most existing atlas-based segmentation approaches have the following drawbacks. First, the brain is taken as a single entity for measuring inter-subject pairwise similarity. The limitation of this global comparison strategy is that all voxels in all subjects are indiscriminately given the same weight. Since different brain regions have very different anatomical patterns, a region-wise comparison approach may be more appropriate. This can be achieved by parcellating the brain into multiple anatomically meaningful regions. The importance of local regional atlas has been recently highlighted by some researchers (van Rikxoort et al., 2010; Wolz et al., 2010). Second, usually only a single average-shape atlas is generated from a population. A better strategy would be to construct multiple atlases which can be adapted to the query image using an adaptive weighting scheme.

1.2 Contribution of this paper

We will address these two issues by designing a method to construct, for each query image, a subject-specific atlas which is accommodated to the structural shapes of the query image. We first parcellate the average-shape atlas of a population of images into multiple regions. Then, based on intra-regional anatomical similarity, we employ an exemplar based image clustering method called *affinity propagation* (Frey and Dueck, 2007) to cluster, for each region, the images into different sub-populations (Fig. 1). Each sub-population is represented by an exemplar, and each region is represented by multiple exemplars. We call this collection of regional exemplars the *multi-region-multi-reference atlas*. The multi-region-multi-reference atlas allows the formation of a subject-specific atlas which better matches the local anatomical structures of the query image. Given a query image, one best match exemplar is selected for each region and the selected exemplars for all regions are combined to form the final subject-specific atlas. A joint registration-segmentation strategy is finally used to segment the query image. Experimental results indicate that significant segmentation accuracy improvement can be achieved.

In the following, we describe the details of our approach in the Method section. In the Experiment section, segmentation performance is compared between the proposed method and a manual rater, as well as with two average-shape atlas-based methods. The Discussion section highlights the novelty of the proposed method and concludes this paper.

2 Method

A flow chart of the proposed framework, comprising three major steps, is shown in Fig. 2. *Step 1*: All pre-segmented subjects in a given population are normalized into a common coordinate space, and are averaged equally to construct a whole brain average-shape atlas. *Step 2*: The average-shape atlas is then parcellated into multiple anatomical sub-regions. The individual subjects, residing in the common space, are then subdivided using the same parcellation, and are used to construct a set of regional exemplars and their corresponding probability atlases, which we call as a whole the multi-region-multi-reference atlas. *Step 3*: Each query subject is compared with the multi-region-multi-reference atlas to determine the most likely regional probability atlases, which are finally combined together to form a subject-specific atlas for guiding segmentation. A joint registration-segmentation strategy is then adopted for more effective segmentation of the query image. More details are given in the following subsections.

2.1 Population normalization

Due to the rapid development of neonatal brains, adult atlases or even those constructed from pediatric brains are not suitable for guiding neonatal brain segmentation (Kazemi et al., 2007). In order to construct an atlas specific for neonatal images, we use a dataset collected in our previous study, involving 68 neonatal subjects (38 males and 30 females) scanned at 1.3 (SD 0.7) months after birth. These neonatal images are segmented using a longitudinally constrained method proposed in (Shi et al., 2010), resulting in their GM, WM and CSF probability maps. It is worth noting here that the testing subjects used for evaluation are not included in this dataset.

First of all, all subjects are normalized into a common coordinate space. For normalization, the influence of the degree of regularization in registration should be taken into consideration (Yeo et al., 2008). Registration with low degrees of freedom (DOF) transform like affine transform may lead to misalignment errors, while high DOF may diminish the important local inter-subject anatomical differences due to over-registration. As the DOF changes from low to high, the resulting averaged spatial normalized images vary from being blurred to sharp. We use the following progressive registration strategy to balance the trade-off between aggressive spatial normalization and preserving anatomical variations across subjects. We use affine transform and nonlinear elastic registration sequentially to normalize the population. To avoid bias, we adopt a group-wise strategy to transform all subjects to their population center, so that all subjects can be normalized to a space of minimal transformations.

Specifically, we first align the neonatal subjects $\{I_j, i = 1,\ldots, i,$ to the standard ICBM atlas space (Ca et al., 1997) using affine transformation and average the transformed images to obtain an initial template T. Then a nonlinear registration algorithm, called HAMMER (Shen and Davatzikos, 2002), is performed to register each subject to the current template T^t, resulting in a warped subject. All the warped subjects are then averaged to obtain the updated template. By iteratively alternating between subject alignment and template generation, all images will eventually be warped into a common space. For brain parcellation, an average-shape atlas is obtained by averaging all normalized subjects with an equal weight. However, this average-shape atlas, shown in Fig. 3, is not the best choice for guiding segmentation because it is generally blurred and lacks anatomical details, especially in the cortical regions. We will describe in the next section an approach to construct a better atlas from the population for guiding the tissue segmentation of a query image.

2.2 Learning multi-region-multi-reference atlases

Shapes of local brain structures contain individual characteristics and vary significantly across subjects. In the process of constructing an atlas, these important local inter-subject differences are often diminished, resulting in blurred atlas. By parcellating the brain image into multiple anatomical regions and, for each region, estimating possible regional probability atlases, a more representative atlas can be obtained for each region.

2.2.1 Brain Parcellation—We parcellate the brain image space into multiple anatomical regions, so that each region can be treated separately and represented more precisely. This essentially breaks down the complex representation problem of a high dimensional brain image into smaller feasible representation problems. There are obviously many methods to parcellate a brain image, such as using a labeled atlas (Tzourio-Mazoyer et al., 2002), multigrid parcellation (Zhu and Jiang, 2003), and watershed segmentation (Vincent and Soille, 1991). The labeled atlas and multigrid algorithms use prior information for brain parcellation, whereas the watershed algorithm performs parcellation in a data-driven way, which does not need prior information. Specifically, in the watershed segmentation

algorithm, a flooding process is performed by computing the gradient of a given image, and then the basins would emerge along the edges so that watersheds with adjacent catchment basins can be constructed. Here we apply the watershed algorithm (Vincent and Soille, 1991) on the average-shape atlas generated in Section 2.1 to parcellate the brain into different regions. The parcellation can then be applied to all normalized individual images in the population because they are currently sitting in the same coordinate space. Before percellation, we perform an edge-preserving anisotropic diffusion (Perona and Malik, 1990) to adaptively smooth the average-shape atlas for removing noise, so that the possible over-segmentation due to image noise could be minimized. More iterations of anisotropic diffusion will result in a more homogenous image, and the parcellated regions are thus larger in size and smaller in number. The influence of the scale of the parcellated regions on the segmentation accuracy is evaluated in the Results section. Unless stated otherwise, a moderate level of smoothing is applied in this study, resulting in a total number of 76 regions. As shown in Fig. 4, the parcellation reflects the structural organization of the brain, where anatomically similar structures are grouped together. After applying the parcellation on the spatially normalized query image (as illustrated in Fig. 2), we can compare structural similarity region-wise.

2.2.2 Exemplar-based clustering—For each region, we use affinity propagation (Frey and Dueck, 2007) to cluster the population into sub-populations and, at the same time, determine their respective exemplars. In conventional clustering methods, such as the -means algorithm, the number of clusters has to be determined a priori, and the initial centers of the clusters are often randomly selected. Moreover, since different initialization parameters lead to different results, quite often the -means algorithm needs to be performed multiple times in order to obtain a stable solution. In contrast, affinity propagation initially deems all data points as potential exemplars and determines a suitable number of exemplars among them via a message passing mechanism. To use affinity propagation, we need to define the pairwise image similarity for given subjects in a certain region. Many definitions are available, such as the mean shortest distance (MSD), mutual information (MI), normalized mutual information (NMI), Kullback Leibler (KL) divergence, etc (Bhatia et al., 2007). In this paper, we use MI to obtain the pairwise similarity matrix.

An interesting property of affinity propagation is that, given a similarity matrix, it can automatically determine the most suitable number of sub-populations. Upon computing, for each region, the pairwise similarity values of all the subjects in the population, an by similarity matrix can be obtained. The diagonal elements of similarity matrix are the "preferences" which have influence on the number of exemplars (i.e., number of clusters). The higher the value of preference S, the larger the probability of the corresponding subject image to be chosen as an exemplar. In our case, all subjects are given the same probability and the preferences are initialized as the median of the input similarities. For each subject, its *responsibility* and *availability* values are computed. The "responsibility" R, sent from a data point to a candidate exemplar point, reflects the accumulated evidence of how well-suited point k is to serve as the exemplar for point i, by taking into account other potential exemplars for point i. The "availability" $A_{i,k}$, sent from a candidate exemplar point k to point i, reflects the accumulated evidence of how appropriate it would be for point i to choose point k as its exemplar, by taking into account the support from other points that point k should be an exemplar [6]. For the sake of completeness, we include the definitions for "responsibility" $R_{i,k}$ and "availability" $A_{i,k}$ below. More details can be found in (Frey and Dueck, 2007).

$R_{i,k}$ is given by:

$$R_{i,k} \leftarrow S_{i,k} - \max_{k' s.t. k' \neq k} \{A_{i,k'} + S_{i,k'}\}$$

where availabilities $A_{i,k}$ are initialized as zero. For $K = i$, this self-responsibility reflects accumulated evidence of the point k as an exemplar.

$A_{i,k}$ is defined by:

$$A_{i,k} \leftarrow min\left(0, R_{k,k} + \sum_{i' s.t. i' \notin \{i,k\}} max(0, R_{i',k})\right)$$

where the self-availability $A_{k,k}$ is updated by:

$$A_{k,k} \leftarrow \sum_{i' s.t. i' \notin k} max(0, R_{i',k})$$

$R_{i,k}$ and $A_{k,k}$ are alternatively and iteratively updated, and this stops when local decisions stay constant for a few iterations. The "representativeness" could be defined as the sum of the responsibility and availability values:

$$E_{i,k} = R_{i,k} + A_{i,k}$$

where $E_{i,k}$ can be seen as the updated similarity matrix. Subject image k is taken as exemplar if $E_{k,k}$ is larger than zero. This is illustrated in Fig 5. Fig. 5(a) is the original similarity matrix S of N =68 subjects, and Fig. 5(b) is the updated similarity matrix E. In Fig. 5(b), 4 subjects are selected as exemplars with indexes of 40, 52, 57, and 61, because all 68 subjects show high probability of choosing them as exemplars (as shown by the 4 columns with high probability values). On the other hand these 4 exemplars also show low probability of choosing other subjects as their exemplars (as indicated by the 4 rows with the low probability values). These 4 clusters are well separated as demonstrated in Fig. 5(c) by the 3D-projection views of all 68 subjects with using multidimensional scaling (MDS) (Borg and Groenen, 2005).

The left panel of Fig. 6 shows a region located in the occipital lobe (which is also used in Fig. 5 for demonstration). 68 subjects in a given population are classified into 4 clusters with 4 selected exemplars (as illustrated in the right panel of Fig. 6). For each exemplar, its representativeness values (the probability values of the corresponding column in Fig. 5(b)) are used as weights to obtain regional weighted average of the probability atlases of the subjects in the sub-population and construct the regional sub-population probability atlas (composing of GM, WM, and CSF probability maps). The outcome is the multi-region-multi-reference atlas.

2.3 Segmentation

2.3.1 Subject-specific atlas generation—To measure the image similarity between a given query image and the population, the query image should be aligned to the same coordinate space of the normalized population. The query image is warped to the population

atlas by first performing affine registration and then an intensity-based nonlinear registration (Shen, 2007).

For each region, the best match exemplar is determined using MI with respect to the available exemplars, and the regional probability atlas of the best match exemplar is taken as the most likely atlas for the particular region. This procedure is illustrated in Fig. 7. All the regional atlases are then combined to form a complete whole-brain probability atlas for GM, WM, and CSF, which serves as the subject-specific atlas for guiding the segmentation.

Notice that discontinuities may occur at between-region boundaries. This is even more pronounced when arbitrary parcellation approaches such as the multigrid method are used. We have taken three measures to reduce the effect of discontinuities. First, a data-driven watershed parcellation technique is used, so that the boundaries will not separate the regions with similar anatomy. Second, Gaussian smoothing is performed to ensure continuity of these voxels across boundaries. Third, the segmentation algorithm estimates the tissue distributions in a global manner, and is hence tolerant to certain degree of discontinuity at boundary voxels.

2.3.2 Joint registration-segmentation—After obtaining the subject-specific atlas, a joint registration-segmentation strategy is used for more effective segmentation of the neonatal brain images. The approach of combining atlas-to-subject registration and atlas-based segmentation is well studied and has proven to be more effective since they complement each other (Ashburner and Friston, 2005; Ide et al., 2008; Pohl et al., 2006). We follow a similar approach and use HAMMER (Shen and Davatzikos, 2002) for non-rigid registration. The process is briefly explained in the following paragraph. More details can be found in (Shi et al., 2010).

In a nutshell, the process involves alternating between bias correction, atlas-based tissue segmentation, and atlas-to-subject registration. The atlas represents tissue prior probabilities, modeled by a mixture of Gaussians (MOG), of each voxel in an image. Bayes rule is employed to combine these prior probabilities with the WM, GM, and CSF probabilities estimated from voxel intensities to obtain the posterior probabilities, which are eventually taken as the voxel tissue membership values. Based on the segmentation results, the registration between the atlas and the subject is refined by HAMMER (Shen and Davatzikos, 2002), which will correct the misalignment between the subject and the atlas for better subsequent segmentation accuracy. These processes are iterated until the segmentation results stop changing for a number of iterations.

3 Experimental Results

The proposed multi-region-multi-reference neonatal segmentation framework has been applied to a large set of neonatal brain images available at our institute. For the purpose of evaluation, its performance is tested on 10 neonatal subjects (6 males and 4 females) randomly selected from a dataset of over 400 neonatal subjects, with postnatal age ranging from 26 to 60 days. For these images, T1 images with 160 axial slices were obtained using a 3T head-only MR scanner with imaging parameters: TR=1900 ms, TE=4.38 ms, Flip Angle=7, acquisition matrix=256×192, and resolution=1×1×1mm^3. T2 images of 70 axial slices were obtained using parameters: TR=7380ms, TE=119ms, Flip Angle=150, acquisition matrix=256×128, and resolution=1.25×1.25×1.95 mm^3. T1 and T2 images of these 10 subjects, as well as their representative segmentation results, are shown in Fig. 8 for visual inspection. As gold standard, 2 sagittal, 3 coronal, and 3 transverse slices, taken at the interval of 20 slices, of all of these 10 T2 images, were manually segmented by an expert rater using ITK-SNAP (Yushkevich et al., 2006). Note that the central brain region was not

segmented due to extremely low tissue contrast. Results yielded by the proposed segmentation algorithm were compared with that of manual segmentation using a quantitative measure, namely Dice ratio (DR), defined as. The Dice ratio, defined as $DR = 2|A \cap B|/(|A| + |E$, is used for measuring tissue overlap rate for manual segmentation and automatic segmentation. The intersection of two segmentations and - the number of voxels having the same label in both segmentation results - is normalized by the total number of voxels with the same label in both segmentations and. DR ranges from 0 to 1. The larger DR the higher agreement is between manual and automatic segmentations.

3.1 Influence of scale of parcellated regions

An edge-preserving anisotropic diffusion (Perona and Malik, 1990) was performed on the average-shape atlas for smoothing and noise removal. The smoothed atlas was then processed by a watershed algorithm, which parcellates the brain into many non-overlapping regions. In Fig. 9(a), the number of parcellated regions is shown as a function of the iterations of anisotropic diffusion. A higher degree of smoothness dramatically decreases the number of parcellated regions.

To evaluate the influence of the scale of the parcellated regions on the segmentation accuracy, for number of regions 1, 32, 76, and 132, we construct their corresponding atlases and employ them to segment the test subjects. The means and standard deviations of Dice ratio of the GM and WM segmentations of all subjects are shown in Fig. 9(b). We can observe that parcellation of the brain can help build more localized subject-specific atlas for achieving better segmentation accuracy. When no parcellation is applied (i.e., the number of regions equals 1), we obtain the lowest Dice ratio. We also note, however, that when the number of regions is increased beyond a certain value, the improvement is marginal. For example, parcellating the brain into 132 regions has only minor improvement over 76 regions. As the number of regions increases, the number of voxels in each region becomes smaller and smaller, causing the computation of mutual information to be unreliable and very sensitive to misalignment between images. Considering the exponential increase of computation cost involved in constructing an N by N similarity matrix, we choose a moderate level of smoothing (6 iterations), resulting in a total number of 76 regions (see Fig. 4) with an average size of 24*24*24 mm^3. It took using MATLAB about 4 hours on a standard PC to compute the similarity matrix, construct the multi-region-multi-reference atlases, and assign subject-specific atlases for all 10 test subjects.

3.2 Performance evaluation

We further compare our approach with two average-shape-atlas based segmentation methods. The first method (referred to as *Population-A*) uses a one-year-old atlas, provided by Altaye et al., as guidance for segmentation (Altaye et al., 2008). This atlas was created from 76 infants of ages ranging from 9 to 15 months. The second method (referred to as *Population-B*) uses the neonatal average-shape atlas (with images in the population weighted equally), which was generated in Section 2.1 (Fig. 3). For a fair comparison, the same joint registration-segmentation strategy described in Section 2.3 was adopted in the two control methods to segment the brain images. Shown in Fig. 10(b)-(d) are the GM probability maps of *Population-A*, *Population-B*, and *Proposed method*, respectively; (f)-(h) are their segmentation results. By visually comparing the original T2 image and its segmentation results, it is obvious that the three methods in comparison yield very different segmentation results. *Population-A* yields a very coarse segmentation (Fig. 10(f)), especially in the cortical region where most WM information is lost. The subcortical region is also severely biased by the atlas. *Population-B* generates slightly better result in the subcortical region (Fig. 10(g)), but still many subtle cortical WM structure cannot be detected and

appears fragmented. The proposed method yields the best segmentation results (Fig. 10(h)) in both cortical and subcortical regions, especially in regions with fine structures.

Quantitative results are shown in Fig. 11. For GM, the mean Dice ratio is 0.86 (with a standard deviation (SD) of 0.02) for proposed method, 0.81 (SD 0.03) for *Population-A*, and 0.84 (SD 0.03) for *Population-B*. For WM, the mean Dice ratio is 0.83 (SD 0.03) for the proposed method, 0.75 (SD 0.05) for *Population-A*, and 0.78 (SD 0.05) for *Population-B*. In CSF, the Dice ratio is 0.80 (SD 0.05) for the proposed method, 0.76 (SD 0.05) for *Population-A*, and 0.77 (SD 0.05) for *Population-B*. Using the manual segmentation results as the ground truth, the proposed method yields the best average tissue overlap rates. In agreement with the visual inspection result, as shown in Fig. 10, *Population-B* (neonatal atlas) yields better performance than *Population-A* (one-year-old atlas). This may be due to the fact that the latter uses an atlas that is not specific for neonates and has larger anatomical difference. The proposed algorithm outperforms the two average-shape-atlas based algorithms, and demonstrates that the constructed subject-specific atlas, which is anatomically more similar with the query image, yields the better performance.

Segmentation accuracy in the highly folded cortical region is important for applications such as cortical surface analysis. For performance evaluation of segmentation in the cortical folding area, we generate a mask based on the manual segmentation. First, the WM in manual segmentation is eroded and only large WM bundles remain. The extracted WM bundles are then removed from the original manual segmentation, so that in the end, we are left with only WM structures in highly folded cortical regions – hereafter referred to as cortical WM (see Fig. 12b). Segmentation accuracy in the cortical WM yielded by the three methods is again evaluated in Fig. 13. It is clear that *Proposed method* yields superior performance as shown by the Dice ratio of 0.79 (SD 0.03), compared to 0.69 (SD 0.04) by *Population-A*, and 0.74 (SD 0.04) by *Population-B*.

4 Discussion

A novel multi-region-multi-reference framework for neonatal brain image segmentation is proposed. Unlike the conventional average-shape atlases which are obtained by equally weighting all images, our atlas is built adaptively by parcellating the whole brain into multiple anatomical regions. Instead of restricting to a single atlas for each brain region, multiple atlases are utilized for better representing the shape variations. This multi-region-multi-reference approach takes full advantage of the given population of pre-segmented images, and is able to generate an anatomically similar atlas for the query image, thus benefiting subsequent atlas-based segmentation. Experimental results demonstrate that our method yields the highest agreement with manual segmentations, and outperforms the two average-shape-atlas based segmentation methods. Improvement in tissue segmentation is especially crucial for researches on neonatal brain development.

Our previous work (Shi et al., 2010) has demonstrated that, by introducing a longitudinal constraint in image segmentation, improved performance can be achieved. In practice, however, not all neonatal studies have longitudinal follow-ups. The proposed method addresses this problem by leveraging the information provided by a population of pre-segmented images. This also implies that the current method is not limited to longitudinal data, but caters for a broader range of applications.

It is worth noting that, the proposed multi-region-multi-reference scheme is a flexible framework. For example, the methods used for brain parcellation, image similarity measurement, and image clustering can be replaced with other techniques. Furthermore, if the given population includes subjects with a broad range of ages, the constructed multiple

atlases in each region will learn all the representative shapes from different ages. Thus, the multi-region-multi-reference atlas makes it adaptable to a large range of datasets. In future, more thorough validations using a larger variety of datasets are needed for the current framework. Many components, such as the brain parcellation, similarity measurement, and image clustering can be further refined and optimized.

References

Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 2009;46:726–738. [PubMed: 19245840]

Altaye M, Holland SK, Wilke M, Gaser C. Infant brain probability templates for MRI segmentation and normalization. NeuroImage 2008;43:721–730. [PubMed: 18761410]

Ashburner J, Friston KJ. Unified segmentation. NeuroImage 2005;26:839–851. [PubMed: 15955494]

Bhatia, KK.; Hajnal, JV.; Hammers, A.; Rueckert, D. Similarity Metrics for Groupwise Non-rigid Registration. MICCAI 2007; Springer; 2007. p. 544-552.

Blezek DJ, Miller JV. Atlas stratification. Med Image Anal 2007;11:443–457. [PubMed: 17765003]

Borg, I.; Groenen, PJF. Modern multidimensional scaling: Theory and applications. Springer Verlag; Berlin: 2005.

Brodmann, K. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Barth Leipzig; Germany: 1909.

Ca C, Kwan K, Evans R. BrainWeb: online interface to a 3-D MRI simulated brain database. NeuroImage 1997:5.

Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007;315:972–976. [PubMed: 17218491]

Han, X.; Hibbard, L.; Willcut, V. GPU-accelerated, Gradient-free MI Deformable Registration for Atlas-based MR Brain Image Segmentation. MMBIA 2009; Miami, FL. 2009.

Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 2006;33:115–126. [PubMed: 16860573]

Ide, J.; Chen, R.; Shen, D.; Herskovits, EH. Robust Brain Registration Using Adaptive Probabilistic Atlas. MICCAI 2008; Springer; 2008. p. 1041-1049.

Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B. Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in CT scans. IEEE Trans Med Imaging 2009;28:1000–1010. [PubMed: 19131298]

Joshi S, Davis B, Jomier M, Gerig G. Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 2004;23:151–160.

Kazemi K, Moghaddam HA, Grebe R, Gondry-Jouet C, Wallois F. A neonatal atlas template for spatial normalization of whole-brain magnetic resonance images of newborns: preliminary results. NeuroImage 2007;37:463–473. [PubMed: 17560795]

Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 1990;12:629–639.

Pohl KM, Fisher J, Grimson WEL, Kikinis R, Wells WM. A Bayesian model for joint segmentation and registration. NeuroImage 2006;31:228–239. [PubMed: 16466677]

Prastawa M, Gilmore JH, Lin W, Gerig G. Automatic segmentation of MR images of the developing newborn brain. Med Image Anal 2005;9:457–466. [PubMed: 16019252]

Rohlfing T, Brandt R, Menzel R, Maurer CR Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 2004a;21:1428–1442. [PubMed: 15050568]

Rohlfing T, Russakoff DB, Maurer CR Jr. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans Med Imaging 2004b;23:983–994. [PubMed: 15338732]

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 2008;39:1064–1080. [PubMed: 18037310]

Shen D. Image registration by local histogram matching. Pattern Recognition 2007;40:1161–1172.

Shen D, Davatzikos C. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging 2002;21:1421–1439. [PubMed: 12575879]

Shi F, Fan Y, Tang S, Gilmore JH, Lin W, Shen D. Neonatal brain image segmentation in longitudinal MRI studies. NeuroImage 2010;49:391–400. [PubMed: 19660558]

Song, Z.; Awate, SP.; Licht, DJ.; Gee, JC. Clinical neonatal brain MRI segmentation using adaptive nonparametric data models and intensity-based Markov priors. MICCAI 2007; 2007. p. 883-890.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 2002;15:273–289. [PubMed: 11771995]

van Rikxoort EM, Isgum I, Arzhaeva Y, Staring M, Klein S, Viergever MA, Pluim JP, van Ginneken B. Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus. Med Image Anal 2010;14:39–49. [PubMed: 19897403]

Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based onimmersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 1991;13:583–598.

Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 2004;23:903–921. [PubMed: 15250643]

Weisenfeld, NI.; Mewes, AUJ.; Warfield, SK. Segmentation of newborn brain MRI. Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano; 2006. p. 766-769.

Weisenfeld NI, Warfield SK. Automatic segmentation of newborn brain MRI. NeuroImage 2009;47:564–572. [PubMed: 19409502]

Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D. LEAP: learning embeddings for atlas propagation. NeuroImage 2010;49:1316–1325. [PubMed: 19815080]

Xue H, Srinivasan L, Jiang S, Rutherford M, Edwards AD, Rueckert D, Hajnal JV. Automatic segmentation and reconstruction of the cortex from neonatal MRI. NeuroImage 2007;38:461–477. [PubMed: 17888685]

Yeo BTT, Sabuncu MR, Desikan R, Fischl B, Golland P. Effects of registration regularization and atlas sharpness on segmentation accuracy. Med Image Anal 2008;12:603–615. [PubMed: 18667352]

Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. NeuroImage 2006;31:1116–1128. [PubMed: 16545965]

Zhu C, Jiang T. Multicontext fuzzy clustering for separation of brain tissues in magnetic resonance images. NeuroImage 2003;18:685–696. [PubMed: 12667846]

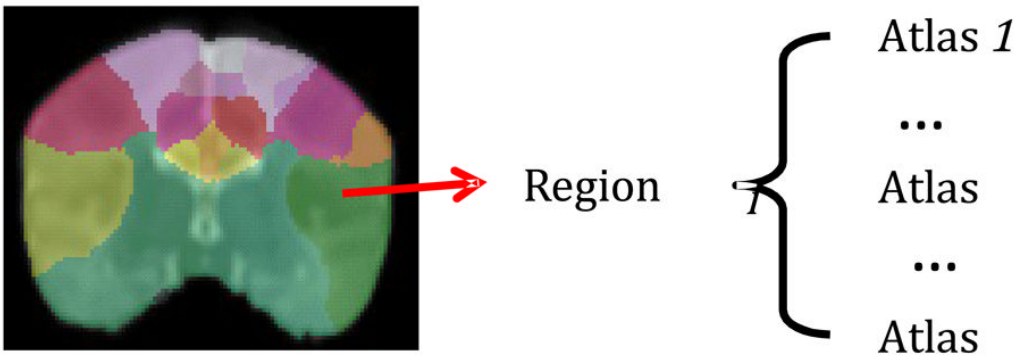

Fig. 1.
Illustration of the multi-region-multi-reference approach. For a given brain parcellation, each region will learn regional probability atlases from a population of images.

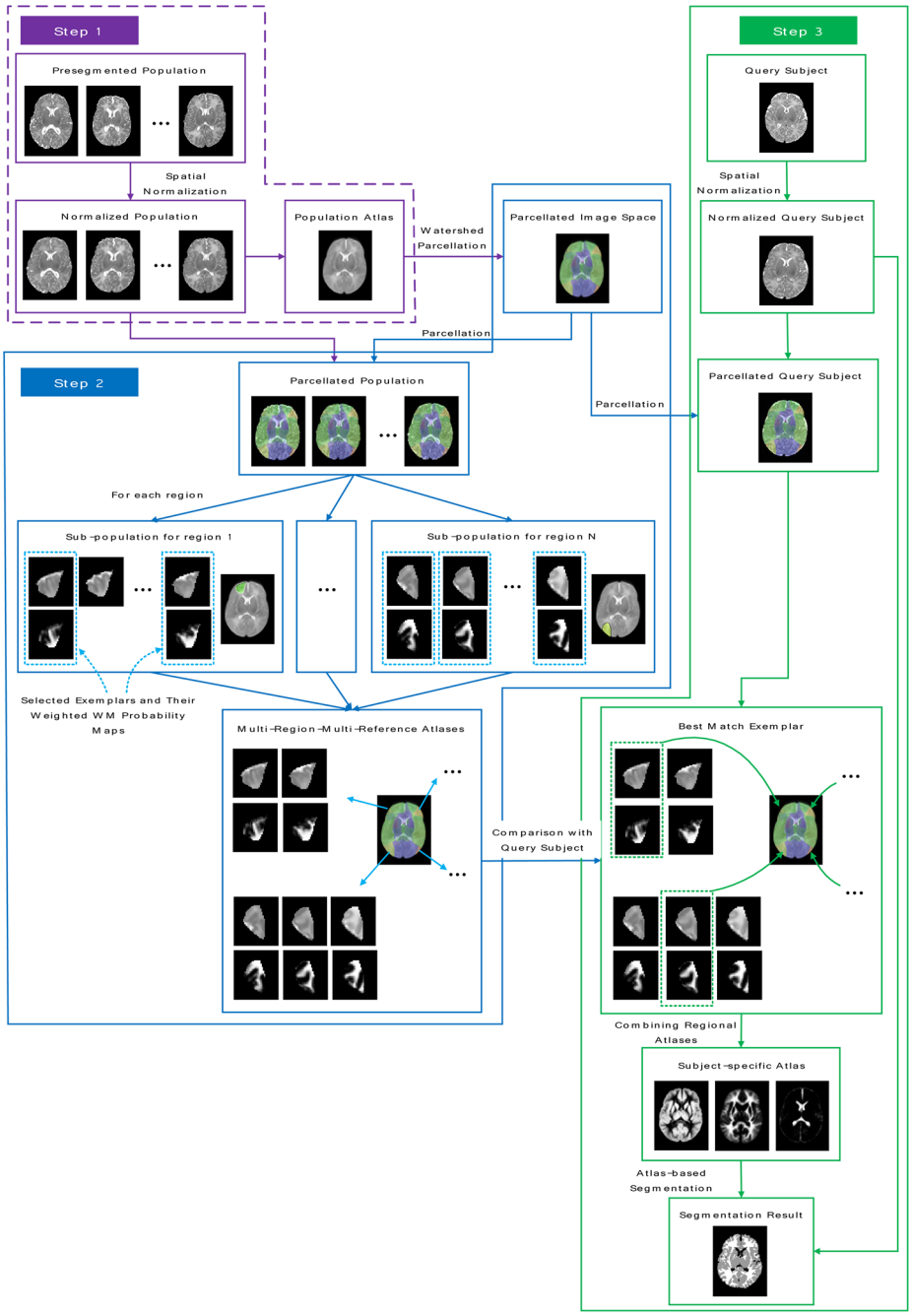

Fig. 2.
Flow chart summarizing the three main steps of the proposed method.

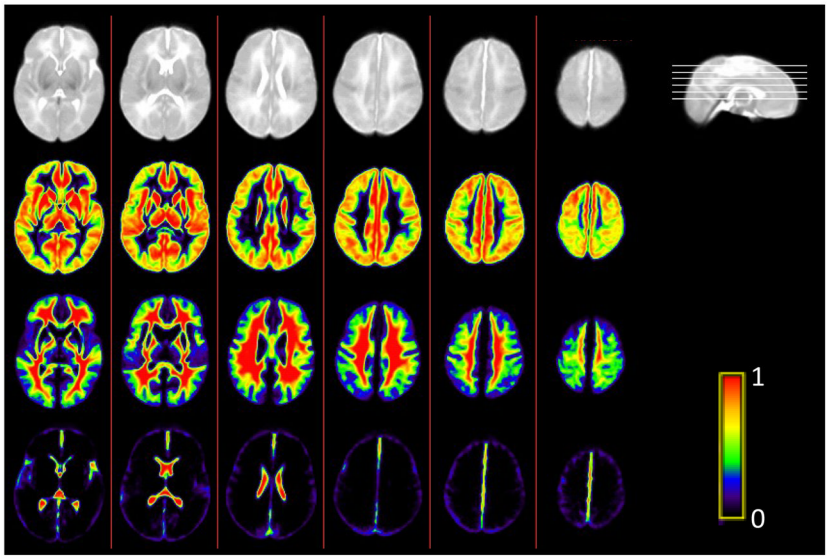

Fig. 3.
Neonatal average-shape atlas constructed with 68 subjects. From top to bottom are the averaged T2 images, and the color-coded probability maps of GM, WM, and CSF, respectively. From left to right are 6 different slices, chosen from locations shown in the sagittal view of the atlas at the top right.

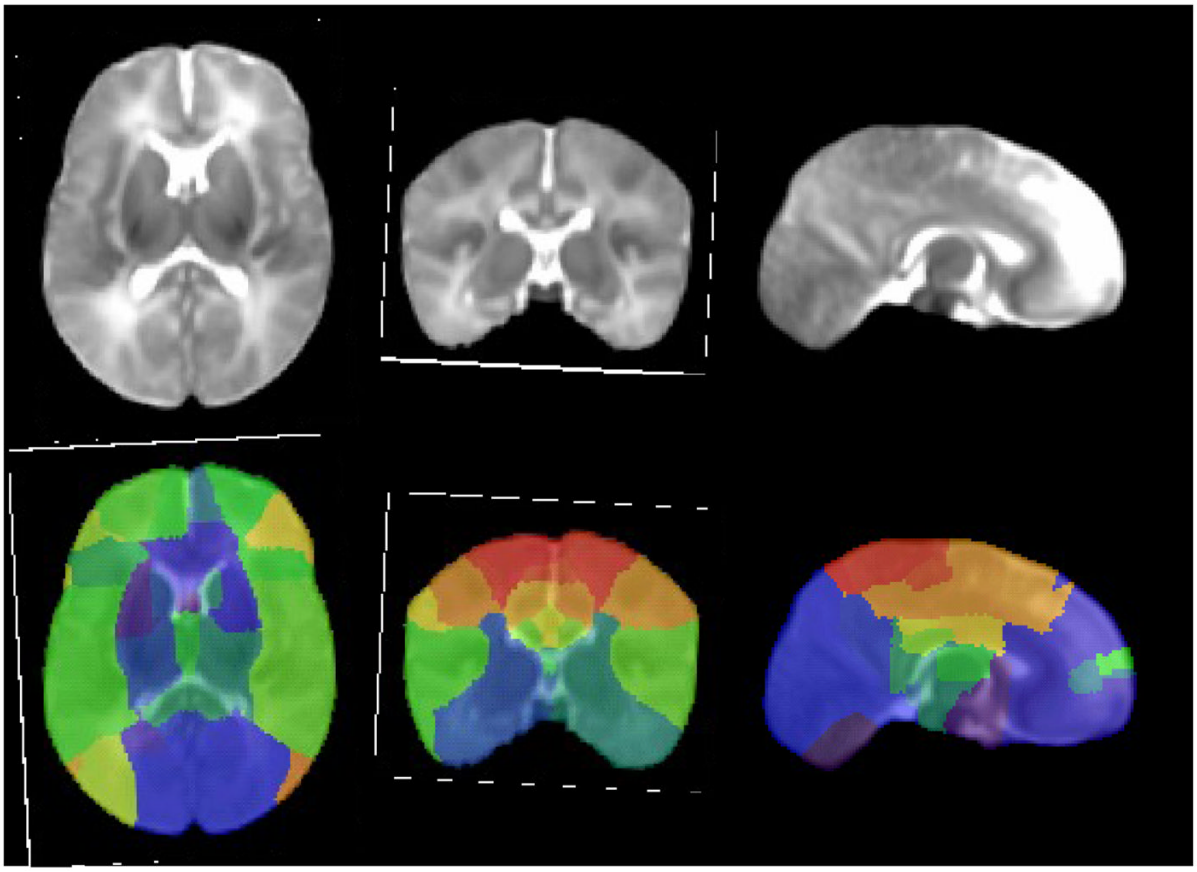

Fig. 4.
Parcellation of the whole brain into 76 regions. The top row shows the 3 orthogonal views of the average-shape atlas, and the bottom row shows their corresponding parcellations. Each color represents a brain region.

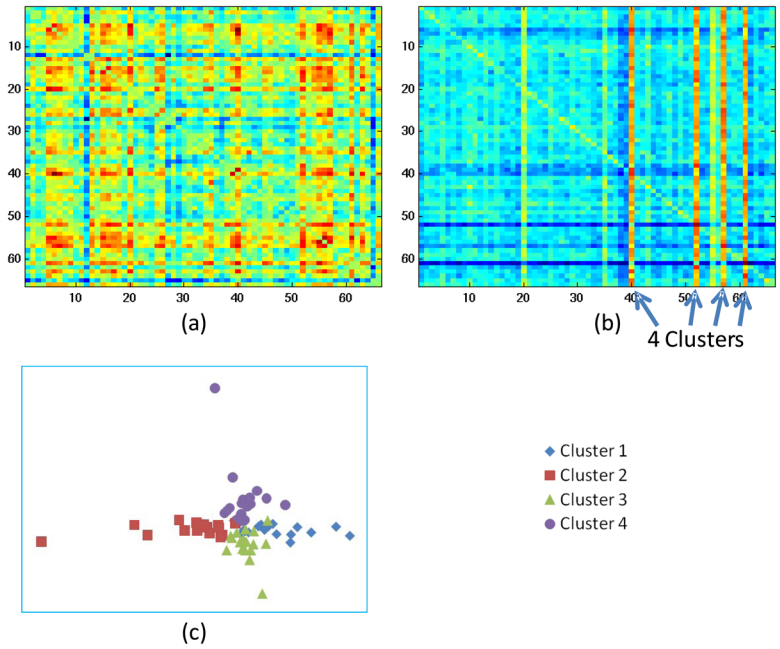

Fig. 5.

Illustration of shape-based clustering for images in a given region (shown in Fig. 6). (a) is the original pairwise similarity matrix. (b) is the updated similarity matrix, i.e., representativeness, in which 4 clusters are selected as exemplars. In (c) are the 3D projections of the 68 subjects in these 4 clusters using multidimensional scaling (MDS).

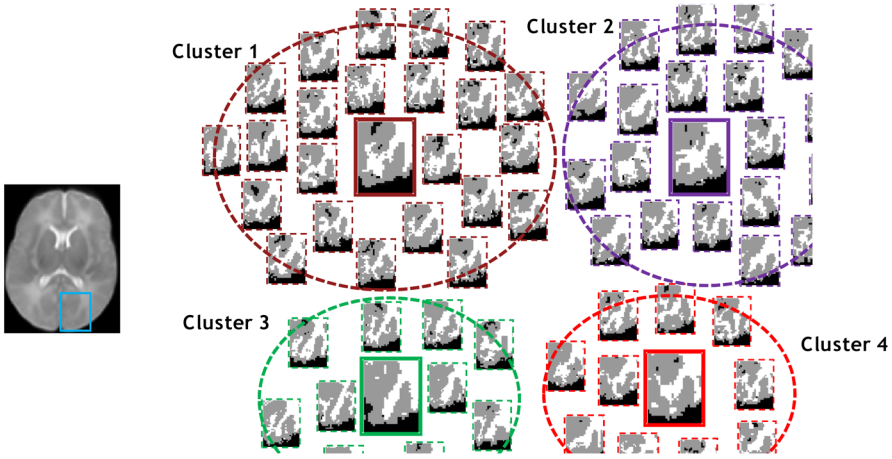

Fig. 6.

Illustration of clustering results in the given region shown in the left panel. Segmented structures are shown for better differentiation. The number of clusters is automatically determined by affinity propagation based on the shape pattern distributions. The constituent of each cluster indicates close similarity.

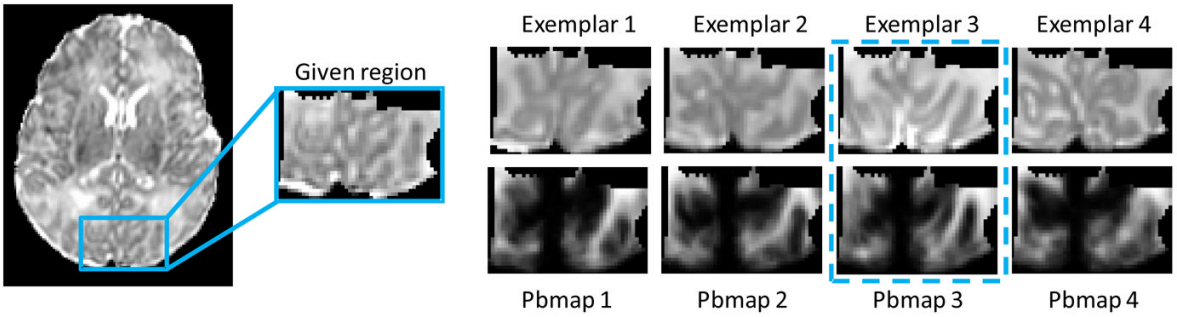

Fig. 7.

Illustration of the subject-specific atlas generation process from multi-region-multi-reference atlases. For a given region in a query image, 4 representative exemplars from the multi-region-multi-reference atlas, as well as their WM probability maps (abbreviated as Pbmap) are shown. This region of the query image is determined based on the similarity measure to be closest to exemplar 3. The probability maps for exemplar 3 will be selected for composing the final atlas for the query image.

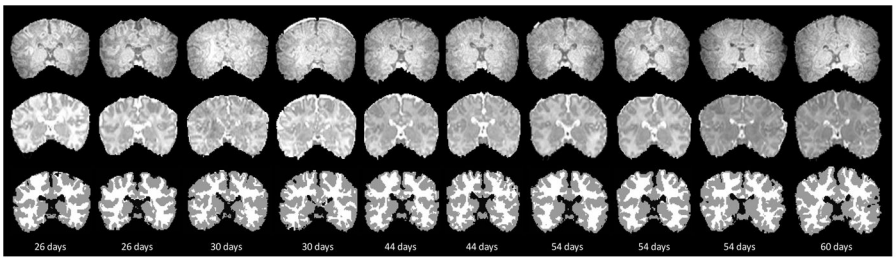

Fig. 8.
MR images of 10 neonates. The top and bottom panels contain 5 subjects each. In each panel, from top to bottom, are the T1 images, the T2 images, and the segmented images. The ages, in days, are stated at the bottom of each panel.

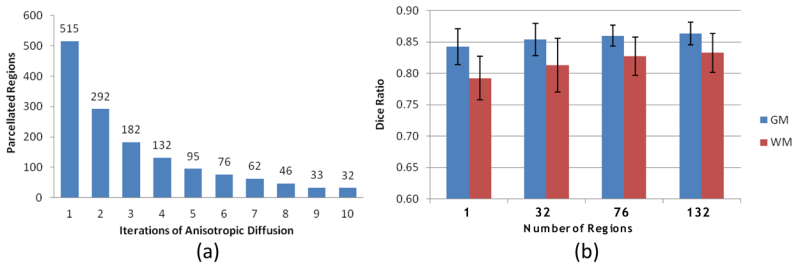

Fig. 9.

(a) Number of parcellated regions as a function of iterations of anisotropic diffusion performed on the average-shape atlas. (b) Dice ratio of GM and WM with respect to the numbers of regions: 1, 32, 76, and 132.

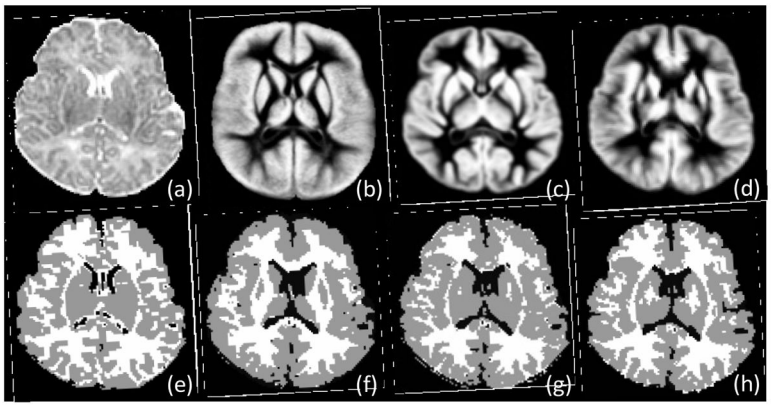

Fig. 10.

GM probability maps and segmentation results. (a) Original T2 image. (e) manual segmentation. (b)-(d) Probability maps of *Population-A*, *Population-B* and *the proposed method*. (f)-(h) The respective final segmentation results.

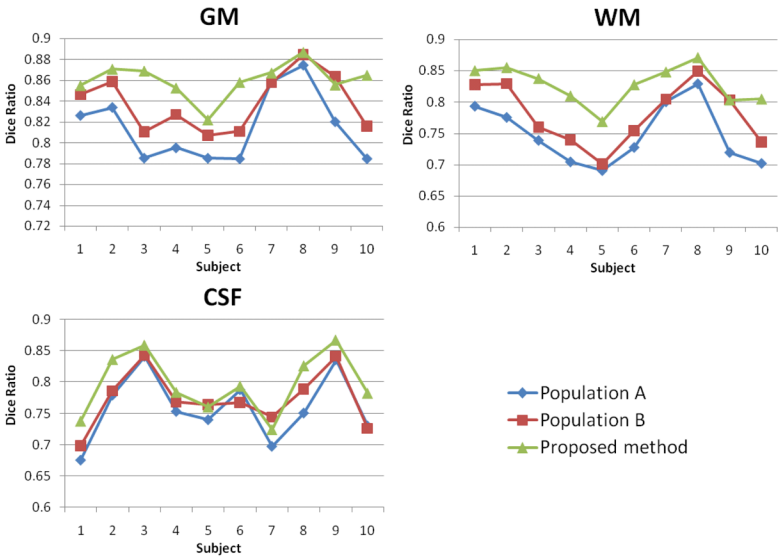

Fig. 11.

The Dice ratios for the segmentation results obtained using *Population-A*, *Population-B*, and *Proposed method*, respectively. Results for GM, WM, and CSF are all provided.

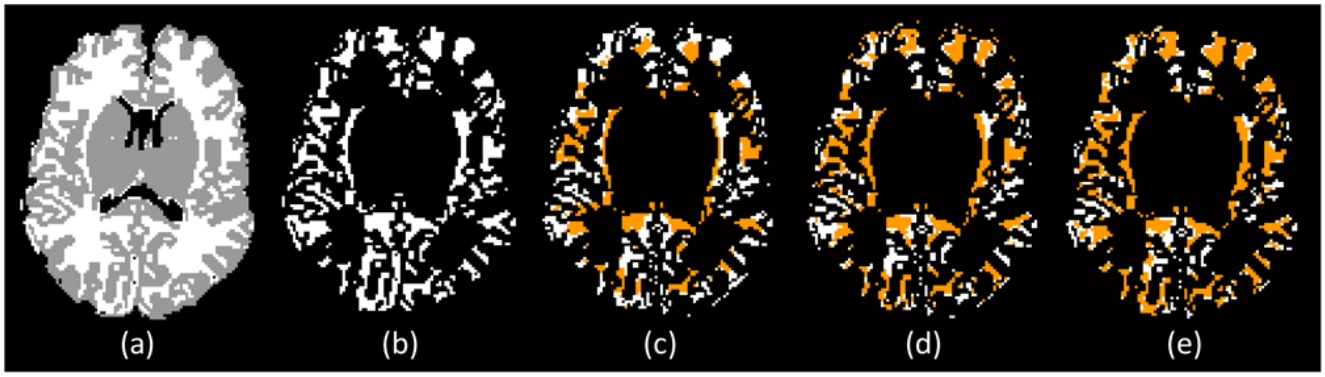

Fig. 12.
Segmentation accuracy comparison for cortical WM. (a) Manual segmentation. (b) Cortical WM mask. (c-e) Overlays cortical WM mask on segmentation results by Population-A (c), Population-B (d), and our proposed method (e). Quantitative results are provided in Fig. 13.

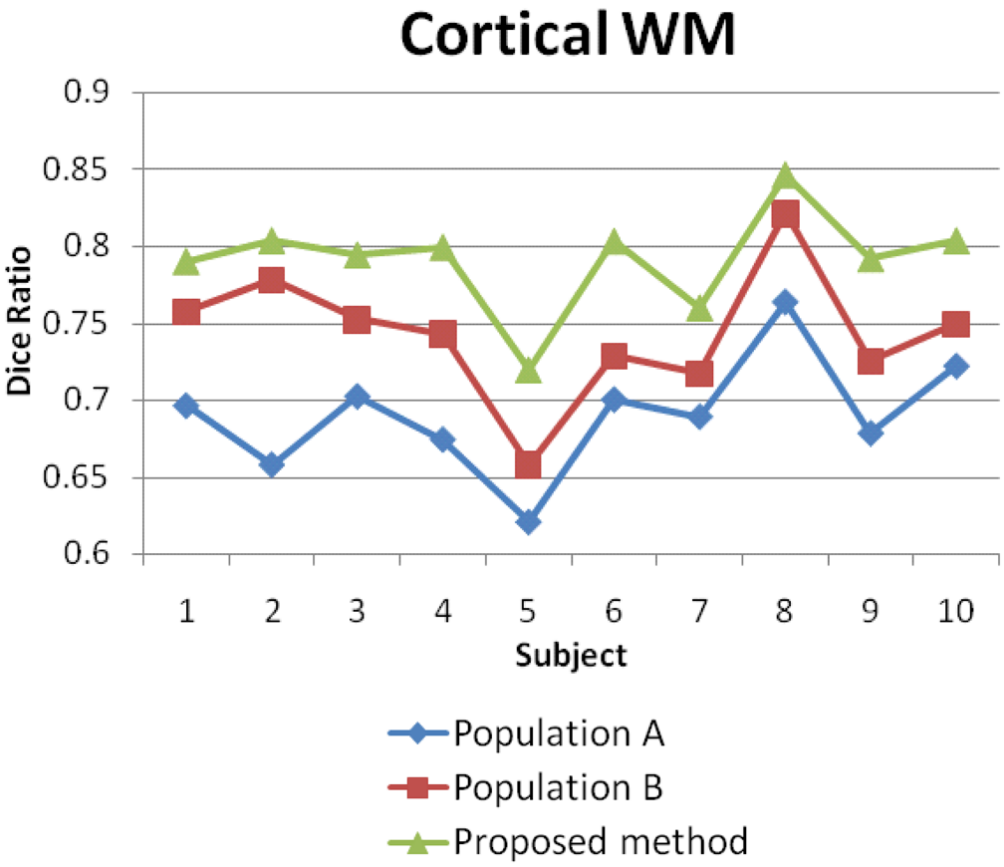

Fig. 13.
Comparison of Dice ratios for the cortical WM.