Published in final edited form as:
Neuroimage. 2015 March ; 108: 160–172. doi:10.1016/j.neuroimage.2014.12.042.

LINKS: Learning-based multi-source IntegratioN frameworK for Segmentation of infant brain images

Li Wang[a], **Yaozong Gao**[a,b], **Feng Shi**[a], **Gang Li**[a], **John H. Gilmore**[c], **Weili Lin**[d], and **Dinggang Shen**[a,e]

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

[b] Department of Computer Science, University of North Carolina at Chapel Hill, NC, USA

[c] Department of Psychiatry, University of North Carolina at Chapel Hill, NC, USA

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

[e] Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Abstract

Segmentation of infant brain MR images is challenging due to insufficient image quality, severe partial volume effect, and ongoing maturation and myelination processes. In the first year of life, the image contrast between white and gray matters of the infant brain undergoes dramatic changes. In particular, the image contrast is inverted around 6-8 months of age, and the white and gray matter tissues are isointense in both T1- and T2-weighted MR images and thus exhibit the extremely low tissue contrast, which poses significant challenges for automated segmentation. Most previous studies used multi-atlas label fusion strategy, which has the limitation of equally treating the different available image modalities and is often computationally expensive. To cope with these limitations, in this paper, we propose a novel learning-based multi-source integration framework for segmentation of infant brain images. Specifically, we employ the random forest technique to effectively integrate features from multi-source images together for tissue segmentation. Here, the multi-source images include initially only the multi-modality (T1, T2 and FA) images and later also the iteratively estimated and refined tissue probability maps of gray matter, white matter, and cerebrospinal fluid. Experimental results on 119 infants show that the proposed method achieves better performance than other state-of-the-art automated segmentation methods. Further validation was performed on the MICCAI grand challenge and the proposed method was ranked top among all competing methods. Moreover, to alleviate the possible

Corresponding author: Dinggang Shen, Professor, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, 130 Mason Farm Road, Chapel Hill, NC 27599-7513, dgshen@med.unc.edu, Phone: 919-966-3535, Fax: 919-843-2641.

anatomical errors, our method can also be combined with an anatomically-constrained multi-atlas labeling approach for further improving the segmentation accuracy.

Keywords

Infant brain images; isointense stage; random forest; multi-modality; context feature; tissue segmentation

1 Introduction

The first year of life is the most dynamic phase of the postnatal human brain development, with the rapid tissue growth and development of a wide range of cognitive and motor functions. Accurate segmentation of infant brain MR images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) in this critical phase is of great importance for studying the normal and abnormal early brain development (Gilmore et al., 2011; Hanson et al., 2013; Li et al., 2014a; Li et al., 2013a; Li et al., 2014b; Li et al., 2013b, 2014c; Li et al., 2014d; Li et al., 2014e; Lyall et al., 2014; Nie et al., 2012; Nie et al., 2014; Verma et al., 2005). However, the segmentation of infant brain MRI is challenging due to the reduced tissue contrast (Weisenfeld and Warfield, 2009), increased noise, severe partial volume effect (Xue et al., 2007), and ongoing white matter myelination (Gui et al., 2012; Weisenfeld and Warfield, 2009). In fact, there are three distinct stages in the first-year brain MR images, including (1) infantile stage (≤5 months), (2) isointense stage (6-8 months), and (3) early adult-like stage (≥9 months). The 2nd row of **Fig. 1** shows representative examples of T1 and T2 images scanned at around 6 months of age. It can be observed that the intensities of voxels in gray matter and white matter are in similar ranges (especially in the cortical regions), thus leading to the lowest image contrast in the first year and the significant difficulty for tissue segmentation. On the other hand, other two stages show a relatively good contrast in either T2-weighted MRI (at the infantile stage) or T1-weighted MRI (at the early adult-like stage), respectively, as shown in the 1st and 3rd rows of **Fig. 1**.

In the past several years, many efforts were put into neonatal brain MRI segmentation (Anbeek et al., 2008; Gui et al., 2012; Leroy et al., 2011; Merisaari et al., 2009; Wang et al., 2011; Wang et al., 2012; Wang et al., 2013c; Warfield et al., 2000; Xue et al., 2007), prompted by the increasing availability of neonatal images. Most proposed methods are atlas-based (Cocosco et al., 2003; Prastawa et al., 2005; Shi et al., 2009; Shi et al., 2011a; Song et al., 2007; Warfield et al., 2000; Weisenfeld et al., 2006; Weisenfeld and Warfield, 2009). An atlas can be generated from manual or automated segmentation of an individual image, or a group of images from different individuals (Kuklisova-Murgasova et al., 2011; Shi et al., 2011b). As an extension, multi-atlas label fusion (MALF) makes use of multiple reference atlases to compensate for the potential biases and errors introduced by using a single atlas (Aljabar et al., 2009; Heckemann et al., 2006; Lötjönen et al., 2010; Rohlfing et al., 2004; Wang et al., 2013a; Warfield et al., 2004). As briefed in Table 1, such MALF methods have recently enjoyed the increased attentions in the infant brain segmentation (Srhoj-Egekher et al., 2013; Wang et al., 2014b). However, one limitation of the current MALF methods is that they often employ a single modality image for segmentation. For

example, Wang et al. (Wang et al., 2014b) utilized 20 atlases from T2-weighted MRI for neonatal image segmentation, achieving promising results. However, to better address the challenge of segmenting isointense infant images, other modalities such as fractional anisotropy (FA) image from diffusion tensor imaging (DTI) could also be utilized to improve the segmentation of WM bundles as well as subcortical regions (Wang et al., 2014a; Yap et al., 2011), as shown in the third column of **Fig. 1**. However, the previous methods involved with multi-modality images usually consider each modality equally, which may be not optimal since certain modalities may provide superior guidance for some varying local brain regions. Another limitation of the previous methods is that they are often computationally expensive (e.g., taking hours as shown in Table 1), due to their requirement of nonlinear registrations between atlases and the target image. Moreover, the larger number of atlases used, the longer computational time is expected. This disadvantage limits the number of atlases that could be utilized by MALF. To this end, some techniques have been proposed to reduce the computational time such as employing simple linear registration, which is unfortunately often associated with compromised performance (Rousseau et al., 2011).

To address these limitations, inspired by the pioneering work (Gao and Shen, 2014; Gao et al., 2014; Zikic et al., 2013a; Zikic et al., 2014; Zikic et al., 2012), we propose a novel Learning-based multi-source IntegratioN frameworK for Segmentation of infant brain images (LINKS). The proposed framework is able to integrate information from multi-source images together for efficient tissue segmentation. Specifically, the multi-source images used in our work initially include multi-modality (T1, T2 and FA) images, and later also the iteratively estimated and refined tissue probability maps for GM, WM and CSF. As a learning-based approach, our framework consists of two stages: training and testing stages. In the training stage, we first use the classification forest (Breiman, 2001) to train a multi-class tissue classifier based on the training subjects with multiple modalities. The trained classifier provides the initial tissue probability maps for each training subject. Inspired by the auto-context model (Loog and Ginneken, 2006; Tu and Bai, 2010), the estimated tissue probability maps are further used as additional input images to train the next classifier, which combines the high-level multi-class context features from the estimated tissue probability maps with the appearance features from multi-modality images for refining tissue classification. By iteratively training the subsequent classifiers based on the updated tissue probability maps, we can finally obtain a sequence of classifiers. Similarly, in the testing stage, given a target subject, the learned classifiers are sequentially applied to iteratively refine the estimation of tissue probability maps by combining multi-modality information with the previously-estimated tissue probability maps. Compared to the previous multi-atlas label fusion methods for infant brain segmentation (Srhoj-Egekher et al., 2013; Wang et al., 2014a; Wang et al., 2014b), the proposed method allows the effective integration of multi-source information, i.e., the original multi-modality images and also the iteratively updated tissue probability maps, which are very important for optimal performance on the segmentation of infant brain images. This is achieved by automatically learning the contribution of each source through random forest with the auto-context scheme (Zikic et al., 2014; Zikic et al., 2012). In addition, in contrast to the previous multi-atlas label fusion methods, which often require nonlinear registrations between the training

subjects (used as atlases) and the target subject, our method involves only the linear registration, i.e., to ensure the same orientation for all subjects. In our case where the orientations of different subjects are similar, even a simple linear registration is skipped.

Validated on 119 infant scans collected from 0-, 3-, 6-, 9- and 12-month-old infants, the proposed method has achieved the state-of-the-art accuracy with significantly reduced runtime, compared to the previous methods. Further validation has been performed on the MICCAI grand challenge, and our method has achieved the best performance among all the competing methods. To alleviate possible anatomical errors, our method can also be combined with the anatomically-constrained multi-atlas labeling approach (Wang et al., 2014a) for further improving the segmentation accuracy.

For related work, Han (Han, 2013) first performed the traditional multi-atlas label fusion and then employed random forests to refine structure labels at “ambiguous” voxels where labels from different atlases do not fully agree. This method has been applied to the segmentation of head and neck images with promising results. However, this method still requires nonlinear registrations as in MALF. Criminisi et al. proposed a random forest-based method for efficient detection and localization of anatomical structures within CT volumes (Criminisi et al., 2009). Zikic et al. proposed a novel method based on the random forests for automatic segmentation of high-grade gliomas and their subregions from multi-channel MR images (Zikic et al., 2012). Similar work was also presented in (Zikic et al., 2013a; Zikic et al., 2013b, 2014), in which an atlas forest was introduced and iteratively employed in an auto-context scheme for efficient adult brain labeling.

2 Method

2.1 Data and Image Preprocessing

This study has been approved by Institutional Review Board (IRB) and written informed consent forms were obtained from all parents. In the training stage, for each time-point (0-, 3-, 6-, 9-, and 12-months of age), we have 10 training atlases with each having all T1, T2 and FA modality images. T1- and T2-weighted images were acquired on a Siemens head-only 3T scanners with a circular polarized head coil. During the scan, infants were asleep, unsedated, fitted with ear protection, and their heads were secured in a vacuum-fixation device. T1-weighted images were acquired with 144 sagittal slices using parameters: TR/TE=1900/4.38ms, flip angle=7°, resolution=1×1×1 mm^3. T2-weighted images were obtained with 64 axial slices: TR/TE=7380/119ms, flip angle=150° and resolution=1.25×1.25×1.95 mm^3. Diffusion weighted images consist of 60 axial slices: TR/TE=7680/82ms, resolution=2×2×2 mm^3, 42 non-collinear diffusion gradients, and b=1000s/mm^2. Seven non-diffusion-weighted reference scans were also acquired. The diffusion tensor images were reconstructed and the respective FA images were computed. Data with moderate or severe motion artifacts was discarded and a rescan was made when possible (Blumenthal et al., 2002).

For image preprocessing, T2 images onto were linearly aligned their corresponding T1 images. FA images were first linearly aligned to T2 images and then propagated to T1 images. All images were resampled into an isotropic 1×1×1 mm^3 resolution. Since those

multi-modality images were from the same subject, they shared the same brain anatomy, and thus allowed to be accurately aligned with the rigid registration. Afterwards, standard image preprocessing steps were performed before segmentation, including skull stripping (Shi et al., 2012), intensity inhomogeneity correction (Sled et al., 1998) and histogram matching, and removal of the cerebellum and brain stem by using in-house tools. To generate the manual segmentations, we first generated an initial reasonable segmentation by using a publicly available software iBEAT (Dai et al., 2013)(http://www.nitrc.org/projects/ibeat/). Then, manual editing was carefully performed by an experienced rater to correct segmentation errors by using ITK-SNAP (Yushkevich et al., 2006) (www.itksnap.org) based on T1, T2 and FA images. Since images of the isointense stage (~6 months) are the most difficult cases due to their extremely low image contrast between WM and GM in the MR images, we will mainly focus on tissue segmentation for the isointense stage. In the experiments, we will also validate our method on images scanned at other time-points of the first year of life.

2.2 Multi-source Classification with Multi-class Auto-context

In this paper, we formulate the tissue segmentation problem as a tissue classification problem. In particular, the random forest (Breiman, 2001), which are inherently suited for multi-class problems, is adopted as a multi-class classifier to produce a tissue probability map for each tissue type (i.e., WM, GM, CSF) by voxel-wise classification. The random forests can effectively handle a large number of training data with high data dimension, which allows us to explore a large number of image features to fully capture both local and contextual image information. The final segmentation is accomplished by assigning the tissue label with the largest probability at each voxel location.

As a supervised learning method, our method consists of training and testing stages. The flowchart of training stage is shown in **Fig. 2**. In the training stage, we will train a sequence of classification forests, each with the input of multi-source images/maps. For simplicity, let the *N* be the total number of the training subjects and let the multi-source images/maps $I^i_{T_1}, I^i_{T_2}, I^i_{FA}, I^i_{WM}, I^i_{GM}, \quad and \quad I^i_{CSF}\ (i{=}1,\ldots,N)$ be the T1-weighted image, T2-weighted image, FA image, tissue probability maps of WM, GM and CSF for the *i*-th training subject, respectively. In the first iteration, the classification forest takes only the multi-modality images $\boldsymbol{I}{=}\left\{I^i_{T_1}, I^i_{T_2}, I^i_{FA}, i{=}1,\ldots N\right\}$ as input, and learn the image appearance features from different modalities for voxel-wise classification. In the later iterations, the three tissue probability maps $\bar{\boldsymbol{I}}\left\{I^i_{WM}, I^i_{GM}, I^i_{CSF}, i{=}1,\ldots N\right\}$ obtained from the previous iteration will act as additional source images. Specifically, high-level multi-class context features are extracted from three tissue probability maps to assist the classification, along with multi-modality images. Since multi-class context features are informative about the nearby tissue structures for each voxel, they encode the spatial constraints into the classification, thus improving the quality of the estimated tissue probability maps, as also demonstrated in **Fig. 2**. In the following section, we will describe our adaption of random forests to the task of infant brain segmentation in details.

2.3 Random forests

We employ a random forest to determine a class label $c \in C$ for a given testing voxel $x \in \Omega$, based on its high-dimensional feature representation $f(x, \boldsymbol{I})$, where $\boldsymbol{I}$ is a set of multi-modality images. The random forest is an ensemble of decision trees, indexed by $t \in [1, T]$, where T is the total number of trees at each iteration. A decision tree consists of two types of nodes, namely internal nodes (non-leaf nodes) and leaf nodes. Each internal node stores a split (or decision) function, according to which the incoming data is sent to its left or right child node, and each leaf stores the final answer (predictor) (Criminisi et al., 2012). During training of the first iteration, each decision tree t will learn a weak class predictor $p_t(c \mid f(x, \boldsymbol{I}))$ by using the multi-modality images $\boldsymbol{I}$. The training is performed by splitting the training voxels at each internal node based on their feature representations and further assigning samples to the left and right child nodes for recursive splitting. Specifically, at each internal node, to inject the randomness for improved generalization, a sampled subset Θ of all possible features is randomly selected (Criminisi et al., 2009; Criminisi et al., 2012). Then, for each voxel x, a binary test is performed: $\theta \geq \xi$, where θ_k indicates the k-th feature of Θ and ξ is a threshold. According to the result of the test function, the training voxel x will be sent to its left or right child node. The purpose of training is to optimize both values of θ_k and ξ for each internal node by maximizing the *information gain* (Criminisi et al., 2012; Zikic et al., 2013a). Specifically, during node optimization, all variable features $\theta_K \in \Theta$ are tried one by one, in combination with many discrete values for the threshold ξ. The optimal combination of $\theta_k^* \quad and \quad \zeta^*$ and ξ^* corresponding to the maximum *information gain* is finally stored in the node for future use. The tree continues growing as more splits are made, and stops at a specified depth (D), or when satisfying the condition that a leaf node contains less than a certain number of training samples (S_{min}). Finally, by simply counting the labels of all training samples which reach each leaf node, we can associate each leaf node l with the empirical distribution over classes $p_t^l(c|f(f, \boldsymbol{I}))$. In our implementation, before training each tree, we randomly sample 10000 Haar-like features (Section 2.4) from the feature pool. During the tree training, for each node optimization, all these 10000 Haar-like features are searched with their respective randomly sampled thresholds to find the feature-threshold combination, which maximizes the information gain defined as the entropy difference before and after split. The number of thresholds randomly sampled at each node is set as 10. The threshold selection criterion is the same as that implemented in the open source library provided by Microsoft (Sherwood)[1].

During **testing**, each voxel to be classified is independently pushed through each trained tree t, by applying the learned split parameters (θ_k and ξ). Upon arriving at a leaf node l_x, the empirical distribution of the leaf node is used to determine the class probability of the testing sample x at tree t, i.e., $p_t(c|f(x, \boldsymbol{I})) = p_t^{l_x}(c|f(x, \boldsymbol{I}))$. The final probability of the testing sample x is computed as the average of the class probabilities from individual trees, i.e., $p(c|x) = \frac{1}{T}\Sigma_{t=1}^{T} p_t(c|f(x, \boldsymbol{I}))$. In this paper, by applying the trained forest in the first

[1] http://research.microsoft.com/en-us/projects/decisionforests/

iteration, each *i*-th training subject will produce three tissue probability maps $\left(\bar{\boldsymbol{I}}_{\boldsymbol{i}}=\left\{I^i_{WM}, I^i_{GM}, I^i_{CSF}\right\}\right)$, as shown in the second row of **Fig. 2**.

During the training of later iterations, the same training procedure is repeated as the first iteration. The only difference is that the tissue probability maps $\bar{\boldsymbol{I}}$ obtained from the first iteration will act as additional source images for extracting the new types of features. Then, the tissue probability maps are iteratively updated and fed into the next training iteration. Finally, a sequence of classifiers will be obtained. **Fig. 3** shows an example by applying a sequence of learned classifiers on a testing subject. As shown in **Fig. 3**, in the first iteration, three tissue probability maps are estimated with only the image appearance features obtained from multi-modality images $\boldsymbol{I}$. In the later iterations, the tissue probability maps $\bar{\boldsymbol{I}}$ estimated from the previous iteration are also fed into the next classifier for refinement. As we can see from **Fig. 3**, the tissue probability maps are gradually improved with iterations and become more and more accurate, by comparing to the ground truth shown in the last row of **Fig. 3**.

2.4 Appearance and context features

Our framework can utilize any kind of features from multi-modality and tissue probability maps, such as SIFT (Lowe, 1999), HOG (Dalal and Triggs, 2005), and LBP features (Ahonen et al., 2006), for tissue classification. In this work, we use 3D Haar-like features (Viola and Jones, 2004) due to their efficiency. Specifically, for each voxel *x*, its Haar-like features are computed as the local mean intensity of any randomly displaced cubical region R_1 (**Fig. 4**(a)), or the mean intensity difference over any two randomly displaced, asymmetric cubical regions (R_1 and R_2) (**Fig. 4**(b)), within the image patch *R* (Han, 2013):

$$f\left(x, I\right)=\frac{1}{|R_1|}\sum_{u\in R_1} I\left(u\right)-b\frac{1}{|R_2|}\sum_{v\in R_1} I\left(v\right), R_1 \quad \in \quad R, \quad R_2 \quad \in \quad R, \qquad b\in\{0,1\}$$

where *R* is the patch centered at voxel *x*, *I* is any kind of source image, and the parameter *B* $\in$ {0, 1} indicates whether one or two cubical regions are used, as shown in **Fig. 4**(a) and (b), respectively. In the intensity patch *R*, its intensities are normalized to have the unit $\ell 2$ norm (Cheng et al., 2009; Wright et al., 2010). However, for the patches from the probability maps, we did not perform any normalization. In theory, for each voxel we can determine an infinite number of such features. For simplicity, we employ 3D Haar-like features for both image appearance features and multi-class context features.

2.5 Post-processing: Imposing anatomical constraint into the segmentation

Based on the probability maps estimated by the random forest, the final segmentation of the target subject could be obtained by assigning the label with the maximal probability for each voxel. However, the classification is performed for each voxel independently, and also as noticed in our previous work (Wang et al., 2014a), the probability maps obtained by the random forest might introduce artificial anatomical errors in the final segmentation results. To deal with this possible limitation, we impose an anatomical constraint into the segmentation by using sparse representation, which has been employed in many applications (Gao et al., 2012; Shao et al., 2014; Wang et al., 2013b; Wang et al., 2014b). Specifically,

by applying the trained classification forests, each training subject *i* can obtain its corresponding *forest-based tissue probability maps* $\left(\bar{\boldsymbol{I}}_{\boldsymbol{i}}=\left\{I^i_{WM}, I^i_{GM}, I^i_{CSF}\right\}\right)$. For each voxel *x* in each *tissue probability map* of the target subject, its probability patch is represented as $\boldsymbol{M}(x) \overset{def}{=} [\boldsymbol{m}_{WM}(x); \boldsymbol{m}_{GM}(x); \boldsymbol{m}_{CSF}(x)]$ where $\boldsymbol{m}_{WM}(x), \boldsymbol{m}_{GM}(x) \quad and \quad \boldsymbol{m}_{CSF}(x)$ are the corresponding patches extracted from WM, GM and CSF probability maps, respectively. Note that $\boldsymbol{m}_{WM}(x), \boldsymbol{m}_{GM}(x) \quad and \quad \boldsymbol{m}_{CSF}(x)$ are arranged as column vectors. Note that, it is not necessary to perform any normalization on the column vectors since their range is already in [0 1]. Similarly, we can extract probability patches from the aligned probability maps of the *i*-th training subject, i.e., $\boldsymbol{M}^{\boldsymbol{i}}(y) \overset{def}{=} \left[\boldsymbol{m}^i_{WM}(y); \boldsymbol{m}^i_{GM}(y); \boldsymbol{m}^i_{CSF}(y)\right], y \in N(x) \quad where \quad N(x)$ is the neighborhood of *x*. Then we can build a dictionary $\boldsymbol{D}(x) \overset{def}{=} \left\{\boldsymbol{M}^i(y), i=1, \ldots, N, \forall y \in N(x)\right\}$ To represent the patch $\boldsymbol{M}(x)$ by the dictionaries $\boldsymbol{D}(x)$, its coefficients vector $\boldsymbol{\alpha}$ could be estimated by many coding schemes, such as sparse coding (Wright et al., 2009; Yang et al., 2009) and locality-constrained linear coding (Wang et al., 2010). Here, we employ a sparse coding scheme (Wright et al., 2009; Yang et al., 2009), which is robust to noise and outlier, to estimate the coefficient vector $\boldsymbol{\alpha}$ by minimizing a non-negative Elastic-Net problem (Zou and Hastie, 2005),

$$\min_{\substack{\alpha \\ \alpha \geq 0}} \frac{1}{2}\|\boldsymbol{M}(x) - \boldsymbol{D}(x)\alpha\|_2^2 + \lambda_1\|\alpha\|_1 + \frac{\lambda_2}{2}\|\alpha\|_2^2 \quad (2)$$

In the above Elastic-Net problem, the first term is the data fitting term based on the patch similarity, and the second term is the $\ell 2$ regularization term which is used to enforce the sparse constraint on the reconstruction coefficients $\boldsymbol{\alpha}$, and the last term is the $\ell 2$ smoothness term for enforcing similar coefficients for similar patches. Eq. (2) is a convex combination of $\ell 1$ lasso (Tibshirani, 1996) and $\ell 2$ ridge penalty, which encourages a grouping effect while keeping a similar sparsity of representation (Zou and Hastie, 2005). In our implementation, we use the LARS algorithm (Efron et al., 2004), which was implemented in the SPAMS toolbox (http://spams-devel.gforge.inria.fr), to solve the Elastic-Net problem. Each element of the sparse coefficient vector $\boldsymbol{\alpha}$, i.e., $a^i(y)$, reflects the similarity between the target patch $\boldsymbol{M}(x)$ and each patch $a^i(y)$ in the patch dictionary. Based on the assumption that similar patches should share similar tissue labels, we use the sparse coefficients $\boldsymbol{\alpha}$ to estimate the probability of the voxel *x* belonging to the *j*-th tissue, $j \in \{WM, GM, CSF\}$, i.e.,

$$P_j(x) = \sum_i \sum_{y \in N^i(x)} \alpha^i(y)\, \delta_j\left(L^i(y)\right) \quad (3)$$

where $L^i(y)$ is the segmentation label (WM, GM, or CSF) for voxel *y* in the *i*-th template image, and $\delta_l(L^i(y))$ is defined as

$$\delta_j\left(L^i(y)\right)=\begin{cases}1, L^i(y) & =j\\ 0, L^i(y) & \neq j\end{cases} \quad (4)$$

Finally, $P_j(x)$ is normalized to ensure $\Sigma_j P_j(x)=1.N^i(x)$ denotes a neighborhood around voxel *x* in the *i*-th training subject. To convert from the soft probability map to the hard segmentation, the label of the voxel is determined using the maximum a posteriori (MAP) rule.

3 Experimental Results

Due to the dynamic changes of appearance pattern in the first year of life, it is difficult to train the random forest jointly for all time-points. Therefore, we trained the random forest for each time-point separately. As mentioned in Section 2.1, in the training stage, for each time-point (0-, 3-, 6-, 9-, and 12-months of age), we have 10 training atlases with all T1, T2 and FA modality images. The validations are performed on other 119 *target* subjects consisting of 26, 22, 22, 23, and 26 subjects at 0-, 3-, 6-, 9- and 12-months of age, respectively. The manual segmentation for each subject is provided and considered as the ground truth for quantitative comparison. In the following, we will mainly focus on describing results for the 6-month images, since they are the most difficult for segmentation due to insufficient image contrast. Besides, we will also validate the proposed method on the MICCAI grand challenge (http://neobrains12.isi.uu.nl/).

In our implementation, for each tissue type, we randomly select 10000 training voxels for each class label from each training subject. Then, from the 7×7×7 patch of each training voxel, 10000 random Haar-like features are extracted from all source images: T1, T2, FA images, and three probability maps of WM, GM and CSF. In each iteration, we train 20 classification trees. We stop the tree growth at a certain depth (i.e., D=50), with a minimum number of 8 samples for each leaf node (S_{min}=8) (Zikic et al., 2013a). The selections of the parameters are based on the following cross-validation.

3.1 Impact of the parameters

Values for the parameters (e.g., the number of training subjects, the number of trees, depth of trees, and the patch size) in our proposed method were determined via leave-one-out cross-validation on all training subjects, according to the parameter settings described in (Bach et al., 2012). During parameter optimization, when optimizing a certain parameter, the other parameters were set to their own fixed values. For example, we first study the impact of the number of training subjects on segmentation accuracy in the 1st row of Fig. **5**. We conservatively set the number of tree as 30 and the maximal tree depth as 100. We further set the minimal number of samples for leaf node as 8 according to previous work (Zikic et al., 2013a). Thus, in most cases, the stopping criterion is based on the minimal number of samples in the leaf node. As expected, increasing the number of training subjects generally improves the segmentation accuracy, as the average Dice ratio increases from 0.81 (N=1) to 0.86 (N=9) for WM, from 0.83 (N=1) to 0.88 (N=9) for GM, and from 0.88 (N>=1) to 0.92 (N=9) for CSF. Also, increasing the number of training subjects seems to make the segmentations more consistent as reflected by the reduced standard deviation from 0.014

(*N*=1) to 0.009 (*N*=9) for WM, from 0.011 (*N*=1) to 0.009 (*N*=9) for GM, and from 0.010 (*N*=1) to 0.007 (*N*=9) for CSF. Though the experiment shows an increase of accuracy with the increasing number of training subjects, the segmentation performance begins to converge after *N*=9. Therefore, in this paper, we choose $N \geq 10$, which is enough to generate reasonable and accurate results. It is worth noting that the increase of the training subjects will not increase the testing time, which is different from other multi-atlas based methods (Coupé et al., 2011; Rousseau et al., 2011; Srhoj-Egekher et al., 2013; Wang et al., 2014a). The 2nd row of Fig. **5** shows the influence of the number of trees on the segmentation accuracy. We similarly find that the more the better, but also the longer it will take to do the training. In addition, note that, beyond a certain number of trees, results will stop getting significantly improved. In this paper, we finally choose 20 trees in each iteration. The 3rd row shows the impact of the maximally allowed depth of trees. In general, a low depth will be likely to under-fitting, while a high value will be likely to over-fitting. In our case, we find that the performance is gradually improved from depth of 5 to depth of 20 and keeps steady when the depth is over 20. The 4th row shows the impact of the minimally allowed number for the leaf node. The performance is steady when the number is less than 20; however, when it is larger than 50, the performance starts decreasing. This may be due to the case that the samples with different tissue labels will possibly fall into the same leaf node if a larger allowance is set, which will result in a fuzzy classification. The last row shows the influence of the patch size. The optimal patch size is related to the complexity of the anatomical structure (Coupé et al., 2011; Tong et al., 2013). Too small or too large patch size will result in poor performance. Therefore, in this paper, we select the patch size as 7×7×7.

3.2 Importance of the multi-source information

Fig. 6 shows the Dice ratios on 22 isointense subjects by sequentially applying the learned classifiers based on the multi-source. It can be seen that the Dice ratios are improved with the iterations and become stable after a few iterations (i.e., 5 iterations). Specially, in the second iteration, the Dice ratios are improved greatly due to the integration of the previously-estimated tissue probability maps for guiding classification. These results demonstrate the importance of using multi-class context features for segmentation.

We further evaluate the importances of different modalities: T1, T2 and FA. Since the multi-class context feature is important for the segmentation, as shown in **Fig. 6**, we integrate it with different combinations of three modalities for training and testing. **Fig. 7** demonstrates the Dice ratios of the proposed method with different combinations of three modalities. It is can be seen from **Fig. 7** that any combination of modalities generally produce more accurate results than any single modality, which proves that the multi-modality information is useful for guiding tissue segmentation (Wang et al., 2014a).

3.3 Comparison with existing methods

To evaluate the performance of the proposed method, we adopt the leave-one-out cross-validation. We first qualitatively make comparison with (a) the majority voting (MV), (b) our previously proposed multimodality sparse anatomical labeling (Wang et al., 2014a) on an isointense image (shown in **Fig. 3**). Specifically, majority voting method assigned the

tissue label by obtaining the most votes to each voxel based on the warped segmentations. Our previous method (Wang et al., 2014a) used a patch-based sparse representation strategy. In that work, *M* and *D* in Eq. 2 are replaced by target and training image features, and the final segmentation is calculated based on the sparse coefficients α and the training image labels. As demonstrated in (Rousseau et al., 2011), the use of nonlinear registrations to warped all the training subjects onto the target image space can produce more accurate results than the use of linear registrations. Therefore, to achieve the best performance for majority voting, and our previous method, we applied a nonlinear registration method (ANTs, http://stnava.github.io/ANTs) (Avants et al., 2011) to align all atlases to the target subject based on the multi-modality images. But, for the proposed method, we do not need any registration since all the subjects already have the same orientation. **Fig. 8** shows the segmentation results of these methods on an isointense image. The first row shows the original T1, T2, FA images and also the manual segmentation, which is regarded as the ground truth. The second row shows the segmentation results by different methods. The columns (c) and (d) show the segmentation results obtained by the proposed method *without* and *with* anatomical constraint, as described in Section 2.5. To better compare the results of different methods, the label differences (compared with the ground-truth segmentation) are also presented. The corresponding white matter surfaces obtained by different methods are shown in **Fig. 9.** Two view angles are provided in the first and the third rows, and the zoomed views are presented in the second and the fourth rows. Both the label differences and zoomed view of surfaces qualitatively demonstrate the advantage of the proposed method.

We then quantitatively evaluate the performance of different methods by employing the Dice ratios, as shown in Table 2, together with the results on other 4 time-points. We also evaluate the accuracy by measuring a modified Hausdorff distance (MHD), which is defined as the 95th-percentile Hausdorff distance. The MHD comparison is shown in Table 3. It can be clearly seen that, even for the proposed method *without* anatomical constraint (last second column), it produces a competitive accuracy at all time-points. Especially, a superior accuracy for segmenting the 6-month infant brain images is achieved, as all other methods cannot effectively utilize the multi-source information for guiding the segmentation.

3.4 Results on the NeobrainS12 MICCAI Challenge

We further tested our algorithm on three preterm born infants acquired at 40 weeks gestation corrected age, as provided by the MICCAI Grand Challenge on Neonatal Brain Segmentation (NeoBrainS12). For each infant, an axial 3D T1-weighted scan and an axial T2-weighted scan were acquired with Philips 3T MRI scanner in University Medical Center Utrecht, the Netherlands. The T1-weighted scans were acquired with the following parameters: TR=9.4 ms; TE=4.6 ms; scan time=3.44 min, FOV=180x180; reconstruction matrix=512x512; consecutive sections with thickness=2.0 mm; number of sections=50, in-plane resolution 0.35 mm x 0.35 mm. The parameters for the acquisition of T2-weighted images were: TR=6293 ms; TE=120 ms; scan time=5.40 min; FOV=180x180; reconstruction matrix=512x512; consecutive sections with thickness=2.0 mm; number of sections=50, in-plane resolution 0.35 mm x 0.35 mm. Manual (reference) segmentations were performed either by MDs who were working towards a PhD in neonatology, or by

trained medical students. The manual segmentations were then verified independently by three neonatologists, each with at least seven years of experience in reading neonatal MRI scans. In case of disagreement, the decision on segmentation was made in a consensus meeting. Two subjects with T1-weighted MRI, T2-weighted MRI, and the corresponding reference standard are available for training. A detailed description of the data and the protocol is available at http://neobrains12.isi.uu.nl.

Based on the only two available training subjects provided by the NeoBrainS12, we trained the sequence classifiers and applied on the above three testing subjects. The parameters setting is same as section 3.1. The segmentation results by the proposed method are shown in **Fig. 10**, in which we segment the infant brain into 6 classes: unmyelinated and myelinated whiter matter (WM), cortical grey matter (CGM), basal ganglia and thalami (BGT), brainstem (BS), cerebellum (CB), ventricles and cerebrospinal fluid in the extracerebral space (CSF). The Dice ratios and MHD by our method and also other competing methods (provided by the NeoBrainS12) are shown in Table 4. It can be clearly seen that our methods achieves the superior performance. Based on the overall ranking (shown in Table 5), which is calculated by the DC and MHD, our method is ranked top among all the competing methods[2].

3.5 Comparison with other methods on the MICCAI2013 SATA challenge

Besides for the infant brain segmentation, our method can be also used in other applications. For example, we straightforwardly applied it to the MICCAI2013 SATA challenge[3] for the diencephalon labeling, in which the diencephalon is labeled into 14 ROIs. This dataset consists of 35 training and 12 testing T1 MR images with a resolution of $1\times1\times1$ mm^3. The parameters setting is similar with Section 3.1, except that we randomly select 1000 samples for each ROI due to the small size of each ROI, and conservatively set the maximal tree depth D=100 and the minimal number of samples for each leaf node S_{min}=4. The segmentation results on the 12 testing subjects were submitted to the challenge evaluation system. The accuracy measured by the mean (±standard deviation) Dice ratio by our method (without post-processing) on this challenge data is 0.8426(±0.0478) and that by our method (with post-processing) is 0.8613(±0.0261), while the leading accuracy is 0.8686(±0.0237) (with details of segmentations provided in the website[4]). It can be seen that our result is still very good, only slightly short of the leading accuracy.

3.6 Computational time

The training was done on a computer cluster (2.93 GHz Intel processors, 12M L3 cache, and 48 GB memory). The average training time for one tree is around 2 hours. For each of 5 iterations, we trained 20 trees. Each tree was trained in a parallel way, thus the total training time is around 2 hours × 5 iterations = 10 hours. The average testing time is around 5 mins for a typical infant image, without including the pre-processing (Section 2.1). Note that, for the comparison methods, we also exclude the time for the pre-processing. Inspiring by

[2] http://neobrains12.isi.uu.nl/mainResults_Set1_Original.php
[3] https://masi.vuse.vanderbilt.edu/workshop2013
[4] http://masi.vuse.vanderbilt.edu/submission/leaderboard.html

recently near-real-time labeling work by Ta et al.'s work (Ta et al., 2014), we will further optimize the proposed work.

4 Discussions and Conclusion

We have presented a learning-based method to effectively integrate multi-source images and the tentatively estimated tissue probability maps for infant brain image segmentation. Specifically, we employ a random forest technique to effectively integrate features from multi-source images, including T1, T2, FA images and also the probability maps of different tissues estimated during the classification process. Experimental results on 119 infant subjects and MICCAI grand challenge show that the proposed method achieves better performance than other state-of-the-art automated segmentation methods.

Compared to the existing multi-modality sparse anatomical labeling (Wang et al., 2014a), which treats each source information equally and thus cannot effectively utilize multiple source information, our method implicitly explores the contribution of each source information by employing the random forest to learn the optimal features. Even without using the anatomical constraint, the proposed method can achieve promising results with the least computational cost. Our method can be further combined with an anatomically-constrained multi-atlas labeling approach to alleviate the possible anatomical errors.

There are many discriminative classification algorithms such as Support Vector Machines (SVM) (Burges, 1998), which have been applied successfully to many tasks. However, SVMs are inherently *binary* classifiers. In order to classify different tissues, they are often applied hierarchically or in the one-versus-all manner. Usually, several different classes have to be grouped together, which may make the classification task more complex than it should be. By contrast, the classifier employed in this paper is a random decision forest. An important advantage of random forests is that they are inherently *multi-label* classifiers, which allows us to classify different tissues simultaneously. Random forests can effectively handle a large number of training data with high feature dimensionality. In recent works (Bosch et al., 2007; Pei et al., 2007), random forests have also been shown better than SVMs in multi-class classification problems (Criminisi et al., 2009).

For the random forest, in general, a low depth will likely lead to under-fitting, while a high value will likely lead to over-fitting. In our case, we find that the performance is gradually improved from depth of 5 to depth of 20 and keeps steady when the depth is over 20. The reason why the performance is not decreased when the depth is over 50 can be summarized as follows. **a)** The use of the minimal number of leaf nodes will prevent tree growing too deep. In our experiments, although the maximal tree depth is set as 50, in most cases the tree stops growing after reaching the depth of about 25. **b)** To improve generalization, each tree is trained with a subset of training samples and also a randomly-selected subset of features, as often done in the literature (Criminisi et al., 2009; Criminisi et al., 2012). By assembling these individual trees together, the generalization power can be improved.

In our work, we trained the random forest for each time-point separately, which may be not the optimal choice. It is possible to train a single random forest jointly for all the time-points by taking temporal relationships into account, which will actually render our task as a 4D

segmentation problem. However, there are many challenges to extend 3D segmentation into 4D segmentation. First, using longitudinal constraints, the segmentation performance will highly depend on the accuracy of registering different time-point images. For the infant images, the registration itself has been proven difficult due to rapid changes in anatomy and tissue composition during the first year of age (Kuklisova-Murgasova et al., 2011). Second, it is also difficult to evaluate the segmentation consistency between different time-point images, since no ground truth is available. Considering that, in this paper, we focus on the 3D segmentation problem by considering each time-point independently. Also, the current framework can be applied to subjects with missing longitudinal data, which is unavoidable in the longitudinal studies.

Although our proposed method can produce more accurate results on the infant brain images, it still has some limitations. (1) Our proposed method requires a number of training subjects, along with their corresponding manual segmentation results. Considering that there are totally 50 training subjects for all 5 time-points, a large amount of efforts are required to achieve manual segmentations. In this paper, we performed automatic segmentations by the iBEAT (Dai et al., 2013), followed by manual editing by experts. Thus, the ground truth could be systematically biased by the iBEAT results. (2) Our current training subjects consist of only healthy subjects, which may limit the performance of our method on the pathological subjects. This particular limitation could be partially overcome by employing other imaging information such as the mean diffusivity computed from DTI. (3) In this work, we extract the same feature type, i.e., 3D Haar-like feature, from both multi-modality images and tissue probability maps, which may be not the optimal choice. We should explore other types of features as well. All the above-mentioned limitations will be investigated in our future work.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their constructive comments and suggestions. This work was supported in part by National Institutes of Health grants MH100217, MH070890, EB006733, EB008374, EB009634, AG041721, AG042599, and MH088520.

References

Ahonen T, Hadid A, Pietikainen M. Face Description with Local Binary Patterns: Application to Face Recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2006; 28:2037–2041.

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]

Anbeek P, Vincken K, Groenendaal F, Koeman A, VAN Osch M, VAN DER Grond J. Probabilistic Brain Tissue Segmentation in Neonatal Magnetic Resonance Imaging. Pediatr Res. 2008; 63:158–163. [PubMed: 18091357]

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011; 54:2033–2044. [PubMed: 20851191]

Bach F, Mairal J, Ponce J. Task-Driven Dictionary Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 34:791–804. [PubMed: 21808090]

Blumenthal JD, Zijdenbos A, Molloy E, Giedd JN. Motion Artifact in Magnetic Resonance Imaging: Implications for Automated Analysis. NeuroImage. 2002; 16:89–92. [PubMed: 11969320]

Bosch, A.; Zisserman, A.; Muoz, X. Image Classification using Random Forests and Ferns. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on; 2007. p. 1-8.

Breiman L. Random Forests. Machine Learning. 2001; 45:5–32.

Burges CC. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery. 1998; 2:121–167.

Cheng, H.; Liu, Z.; Yang, L. Sparsity induced similarity measure for label propagation. Computer Vision, 2009 IEEE 12th International Conference on; 2009. p. 317-324.

Cocosco CA, Zijdenbos AP, Evans AC. A fully automatic and robust brain MRI tissue classification method. Med Image Anal. 2003; 7:513–527. [PubMed: 14561555]

Coupé P, Manjón J, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage. 2011; 54:940–954. [PubMed: 20851199]

Criminisi A, Shotton J, Bucciarelli S. Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. MICCAI workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA). 20092009

Criminisi A, Shotton J, Konukoglu E. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. Foundations and Trends® in Computer Graphics and Vision. 2012; 7:81–227.

Dai Y, Shi F, Wang L, Wu G, Shen D. iBEAT: A Toolbox for Infant Brain Magnetic Resonance Image Processing. Neuroinformatics. 2013; 11:211–225. [PubMed: 23055044]

Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on; 2005. p. 886-893.

Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Annals of Statistics. 2004; 32:407–499.

Gao Y, Liao S, Shen D. Prostate segmentation by sparse representation based classification. Medical Physics. 2012; 39:6372–6387. [PubMed: 23039673]

Gao, Y.; Shen, D. Context-Aware Anatomical Landmark Detection: Application to Deformable Model Initialization in Prostate CT Images. In: Wu, G.; Zhang, D.; Zhou, L., editors. Machine Learning in Medical Imaging. Springer International Publishing; 2014. p. 165-173.

Gao, Y.; Wang, L.; Shao, Y.; Shen, D. Learning Distance Transform for Boundary Detection and Deformable Segmentation in CT Prostate Images. In: Wu, G.; Zhang, D.; Zhou, L., editors. Machine Learning in Medical Imaging. Springer International Publishing; 2014. p. 93-100.

Gilmore JH, Shi F, Woolson SL, Knickmeyer RC, Short SJ, Lin W, Zhu H, Hamer RM, Styner M, Shen D. Longitudinal Development of Cortical and Subcortical Gray Matter from Birth to 2 Years. Cerebral Cortex. 2011

Gui L, Lisowski R, Faundez T, Hüppi PS, Lazeyras F.o. Kocher M. Morphology-driven automatic segmentation of MR images of the neonatal brain. Med Image Anal. 2012; 16:1565–1579. [PubMed: 22921305]

Han, X. Learning-Boosted Label Fusion for Multi-atlas Auto-Segmentation. In: Wu, G.; Zhang, D.; Shen, D.; Yan, P.; Suzuki, K.; Wang, F., editors. Machine Learning in Medical Imaging. Springer International Publishing: 2013. p. 17-24.

Hanson JL, Hair N, Shen DG, Shi F, Gilmore JH, Wolfe BL, Pollak SD. Family Poverty Affects the Rate of Human Infant Brain Growth. PLoS ONE. 2013; 8:e80954. [PubMed: 24349025]

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]

Kuklisova-Murgasova M, Aljabar P, Srinivasan L, Counsell SJ, Doria V, Serag A, Gousias IS, Boardman JP, Rutherford MA, Edwards AD, Hajnal JV, Rueckert D. A dynamic 4D probabilistic atlas of the developing brain. NeuroImage. 2011; 54:2750–2763. [PubMed: 20969966]

Leroy F, Mangin J, Rousseau F, Glasel H, Hertz-Pannier L, Dubois J, Dehaene-Lambertz G. Atlas-free surface reconstruction of the cortical grey-white interface in infants. PLoS One. 2011; 6:e27128. [PubMed: 22110604]

Li G, Nie J, Wang L, Shi F, Gilmore JH, Lin W, Shen D. Measuring the dynamic longitudinal cortex development in infants by reconstruction of temporally consistent cortical surfaces. NeuroImage. 2014a; 90:266–279. [PubMed: 24374075]

Li G, Nie J, Wang L, Shi F, Lin W, Gilmore JH, Shen D. Mapping Region-Specific Longitudinal Cortical Surface Expansion from Birth to 2 Years of Age. Cerebral Cortex. 2013a; 23:2724–2733. [PubMed: 22923087]

Li G, Nie J, Wang L, Shi F, Lyall AE, Lin W, Gilmore JH, Shen D. Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age. Cerebral Cortex. 2014b; 24:1289–1300. [PubMed: 23307634]

Li, G.; Wang, L.; Shi, F.; Lin, W.; Shen, D. Multi-atlas Based Simultaneous Labeling of Longitudinal Dynamic Cortical Surfaces in Infants. In: Mori, K.; Sakuma, I.; Sato, Y.; Barillot, C.; Navab, N., editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Springer Berlin Heidelberg; 2013b. p. 58-65.

Li, G.; Wang, L.; Shi, F.; Lin, W.; Shen, D. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer International Publishing; 2014c. Constructing 4D Infant Cortical Surface Atlases Based on Dynamic Developmental Trajectories of the Cortex; p. 89-96.

Li G, Wang L, Shi F, Lyall A, Ahn M, Peng Z, Zhu H, Lin W, Gilmore J, Shen D. Cortical thickness and surface area in neonates at high risk for schizophrenia. Brain Structure and Function. 2014d:1–15. [PubMed: 23474540]

Li G, Wang L, Shi F, Lyall AE, Lin W, Gilmore JH, Shen D. Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. The Journal of Neuroscience. 2014e; 34:4228–4238. [PubMed: 24647943]

Loog M, Ginneken B. Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification. Medical Imaging, IEEE Transactions on. 2006; 25:602–611.

Lötjönen JMP, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D. Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage. 2010; 49:2352–2365. [PubMed: 19857578]

Lowe, DG. Object recognition from local scale-invariant features. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on; 1999. p. 1150-1157.

Lyall AE, Shi F, Geng X, Woolson S, Li G, Wang L, Hamer RM, Shen D, Gilmore JH. Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. Cerebral Cortex. 2014

Merisaari H, Parkkola R, Alhoniemi E, Ter?s M, Lehtonen L, Haataja L, Lapinleimu H, Nevalainen OS. Gaussian mixture model-based segmentation of MR images taken from premature infant brains. Journal of Neuroscience Methods. 2009; 182:110–122. [PubMed: 19523488]

Nie J, Li G, Wang L, Gilmore JH, Lin W, Shen D. A computational growth model for measuring dynamic cortical development in the first year of life. Cerebral Cortex. 2012; 22:2272–2284. [PubMed: 22047969]

Nie J, Li G, Wang L, Shi F, Lin W, Gilmore JH, Shen D. Longitudinal development of cortical thickness, folding, and fiber density networks in the first 2 years of life. Human Brain Mapping. 2014; 35:3726–3737. [PubMed: 24375724]

Pei, Y.; Criminisi, A.; Winn, J.; Essa, I. Tree-based Classifiers for Bilayer Video Segmentation. Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on; 2007. p. 1-8.

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, Russakoff DB, Maurer CR Jr. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. Medical Imaging, IEEE Transactions on. 2004; 23:983–994.

Rousseau F, Habas PA, Studholme C. A Supervised Patch-Based Approach for Human Brain Labeling. Medical Imaging, IEEE Transactions on. 2011; 30:1852–1862.

Shao Y, Gao Y, Guo Y, Shi Y, Yang X, Shen D. Hierarchical Lung Field Segmentation With Joint Shape and Appearance Sparse Learning. Medical Imaging, IEEE Transactions on. 2014; 33:1761–1780.

Shi F, Fan Y, Tang S, Gilmore JH, Lin W, Shen D. Neonatal Brain Image Segmentation in Longitudinal MRI Studies. NeuroImage. 2009; 49:391–400. [PubMed: 19660558]

Shi F, Shen D, Yap P, Fan Y, Cheng J, An H, Wald LL, Gerig G, Gilmore JH, Lin W. CENTS: cortical enhanced neonatal tissue segmentation. Hum Brain Mapp. 2011a; 32:382–396. [PubMed: 20690143]

Shi F, Wang L, Dai Y, Gilmore JH, Lin W, Shen D. Pediatric Brain Extraction Using Learning-based Meta-algorithm. NeuroImage. 2012; 62:1975–1986. [PubMed: 22634859]

Shi F, Yap P-T, Fan Y, Gilmore JH, Lin W, Shen D. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. NeuroImage. 2010; 51:684–693. [PubMed: 20171290]

Shi F, Yap P, Wu G, Jia H, Gilmore JH, Lin W, Shen D. Infant brain atlases from neonates to 1- and 2-year-olds. PLoS One. 2011b; 6:e18746. [PubMed: 21533194]

Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998; 17:87–97. [PubMed: 9617910]

Song Z, Awate SP, Licht DJ, Gee JC. Clinical neonatal brain MRI segmentation using adaptive nonparametric data models and intensity-based Markov priors. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv. 2007:883–890.

Srhoj-Egekher V, Benders MJNL, Viergever MA, Išgum I. Automatic neonatal brain tissue segmentation with MRI. Proc. SPIE. 2013:86691K.

Ta, V-T.; Giraud, R.; Collins, DL.; Coupé, P. Optimized PatchMatch for Near Real Time and Accurate Label Fusion. In: Golland, P.; Hata, N.; Barillot, C.; Hornegger, J.; Howe, R., editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. Springer International Publishing; 2014. p. 105-112.

Tibshirani RJ. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, Series B. 1996; 58:267–288.

Tong T, Wolz R, Coupé P, Hajnal JV, Rueckert D. Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. NeuroImage. 2013; 76:11–23. [PubMed: 23523774]

Tu Z, Bai X. Auto-Context and Its Application to High-Level Vision Tasks and 3D Brain Image Segmentation. PAMI. 2010; 32:1744–1757.

Verma R, Mori S, Shen D, Yarowsky P, Zhang J, Davatzikos C. Spatiotemporal maturation patterns of murine brain quantified by diffusion tensor MRI and deformation-based morphometry. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:6978–6983. [PubMed: 15860588]

Viola P, Jones M. Robust Real-Time Face Detection. International Journal of Computer Vision. 2004; 57:137–154.

Wang H, Suh JW, Das SR, Pluta J, Craige C, Yushkevich PA. Multi-Atlas Segmentation with Joint Label Fusion. IEEE Trans. on PAMI. 2013a; 35:611–623.

Wang J, Yang J, Yu K, Lv F, Huang TS, Gong Y. Locality-constrained Linear Coding for image classification. CVPR. 2010:3360–3367.

Wang, L.; Chen, K.; Shi, F.; Liao, S.; Li, G.; Gao, Y.; Shen, S.; Yan, J.; Lee, PM.; Chow, B.; Liu, N.; Xia, J.; Shen, D. Automated Segmentation of CBCT Image Using Spiral CT Atlases and Convex Optimization. In: Mori, K.; Sakuma, I.; Sato, Y.; Barillot, C.; Navab, N., editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Springer Berlin Heidelberg; 2013b. p. 251-258.

Wang L, Shi F, Gao Y, Li G, Gilmore JH, Lin W, Shen D. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation. NeuroImage. 2014a; 89:152–164. [PubMed: 24291615]

Wang L, Shi F, Li G, Gao Y, Lin W, Gilmore JH, Shen D. Segmentation of neonatal brain MR images using patch-driven level sets. NeuroImage. 2014b; 84:141–158. [PubMed: 23968736]

Wang L, Shi F, Lin W, Gilmore JH, Shen D. Automatic segmentation of neonatal images using convex optimization and coupled level sets. NeuroImage. 2011; 58:805–817. [PubMed: 21763443]

Wang L, Shi F, Yap P-T, Gilmore JH, Lin W, Shen D. 4D Multi-Modality Tissue Segmentation of Serial Infant Images. PLoS One. 2012; 7:e44596. [PubMed: 23049751]

Wang L, Shi F, Yap P, Lin W, Gilmore JH, Shen D. Longitudinally guided level sets for consistent tissue segmentation of neonates. Hum Brain Mapp. 2013c; 34:956–972. [PubMed: 22140029]

Warfield SK, Kaus M, Jolesz FA, Kikinis R. Adaptive, template moderated, spatially varying statistical classification. Med Image Anal. 2000; 4:43–55. [PubMed: 10972320]

Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. Medical Imaging, IEEE Transactions on. 2004; 23:903–921.

Weisenfeld NI, Mewes AUJ, Warfield SK. Segmentation of newborn brain MRI. ISBI. 2006:766–769.

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

Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust Face Recognition via Sparse Representation. IEEE Trans Pattern Anal Mach Intell. 2009; 31:210–227. [PubMed: 19110489]

Wright J, Yi M, Mairal J, Sapiro G, Huang TS, Shuicheng Y. Sparse Representation for Computer Vision and Pattern Recognition. Proceedings of the IEEE. 2010; 98:1031–1044.

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]

Yang J, Yu K, Gong Y. Linear spatial pyramid matching using sparse coding for image classification. CVPR. 2009:1794–1801. T., H.,

Yap P-T, Fan Y, Chen Y, Gilmore JH, Lin W, Shen D. Development Trends of White Matter Connectivity in the First Years of Life. PLoS ONE. 2011; 6:e24678. [PubMed: 21966364]

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]

Zikic D, Glocker B, Criminisi A. Atlas Encoding by Randomized Forests for Efficient Label Propagation. MICCAI. 2013a2013:66–73. [PubMed: 24505745]

Zikic D, Glocker B, Criminisi A. Multi-atlas label propagation with atlas encoding by randomized forests. MICCAI 2013 Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA). 2013b

Zikic D, Glocker B, Criminisi A. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. Medical Image Analysis. 2014; 18:1262–1273. [PubMed: 25042602]

Zikic, D.; Glocker, B.; Konukoglu, E.; Criminisi, A.; Demiralp, C.; Shotton, J.; Thomas, OM.; Das, T.; Jena, R.; Price, SJ. Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR. In: Ayache, N.; Delingette, H.; Golland, P.; Mori, K., editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. Springer Berlin Heidelberg: 2012. p. 369-376.

Zou H, Hastie T. Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society, Series B. 2005; 67:301–320.

Research highlight

- The proposed method effectively integrates multi-source.
- An iterative classification scheme is adopted to train sequence classifiers.
- No any nonlinear registration involved in the proposed method.
- Validation was performed on 119 infant subjects.
- The proposed method was ranked top on the MICCAI grand challenge.

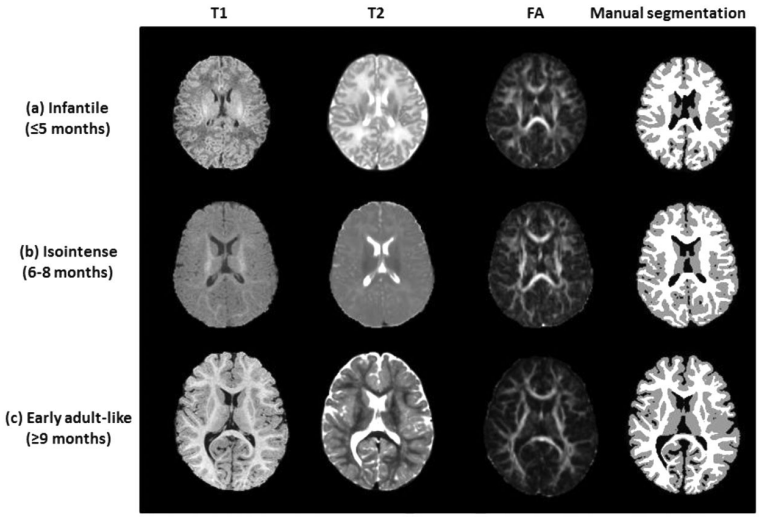

Fig. 1.

Illustration of three distinct stages in infant brain development, with each stage having quite different WM/GM contrast patterns in MR images.

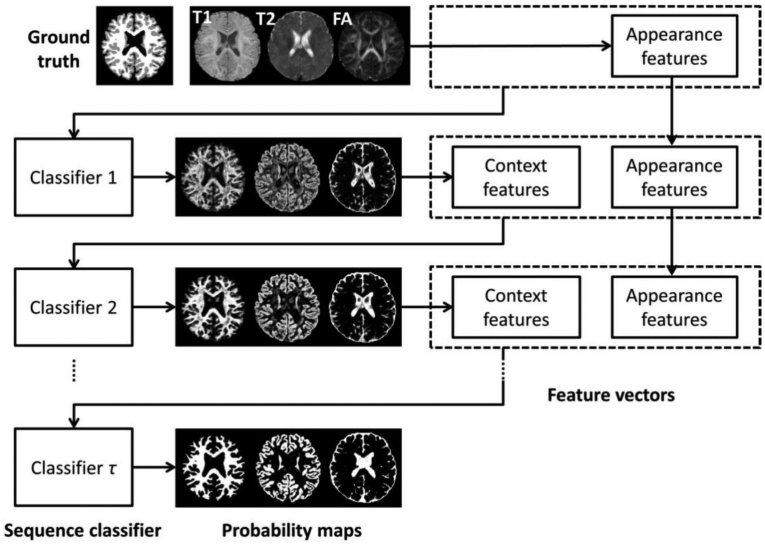

Fig. 2.

Flowchart of the training procedure for our proposed method with multi-source images, including T1, T2, and FA images, along with probability maps of WM, GM and CSF. The appearance features from multi-modality images (***I***) are used for training the first classifier, and then both appearance features and multi-class context features from three tissue probability maps ($\overline{\boldsymbol{I}}$) are employed for training the subsequent classifiers.

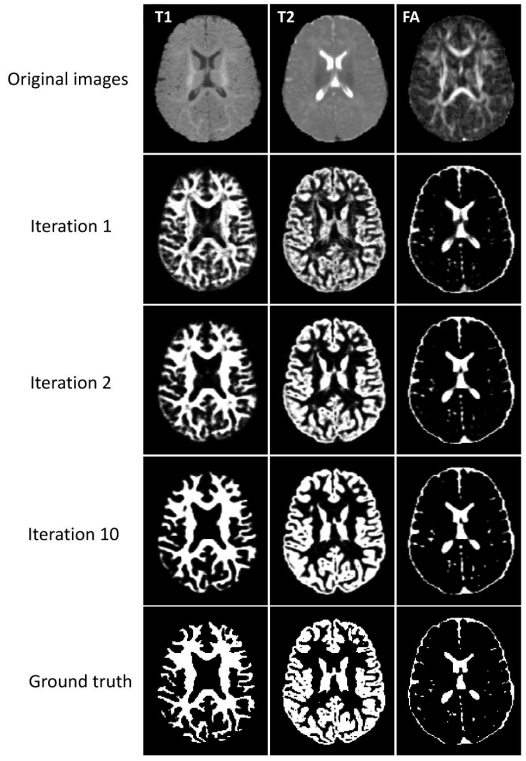

Fig. 3.

The estimated tissue probability maps by applying a sequence of learned classifiers on a target subject in the isointense stage with T1, T2 and FA modalities. The probability maps become more and more accurate and sharp with iterations. The last row shows the ground truth for comparison.

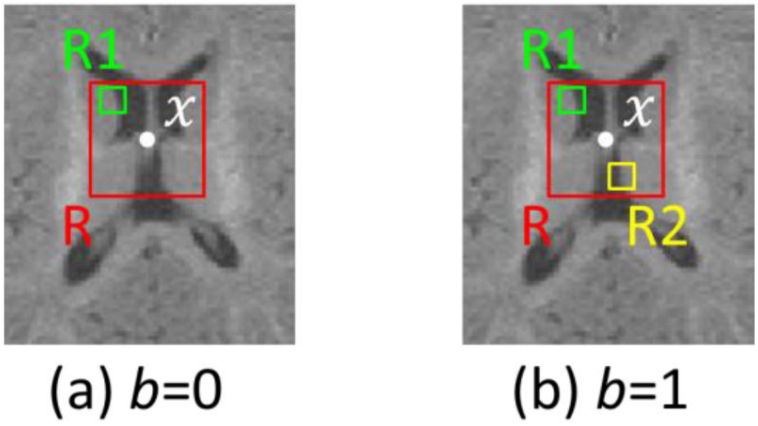

Fig. 4.

Haar-like features, a 2D illustration. The red rectangle indicated a patch *R* centered at *x*. (a) *b*=0: Haar-like features are computed as the local mean intensity of any randomly displaced cubical region within the image patch *R*. (b) *b*=1: Haar-like features are computed as the mean intensity difference over any two randomly displaced, asymmetric cubical regions (R_1 and R_1) within the image patch *R*.

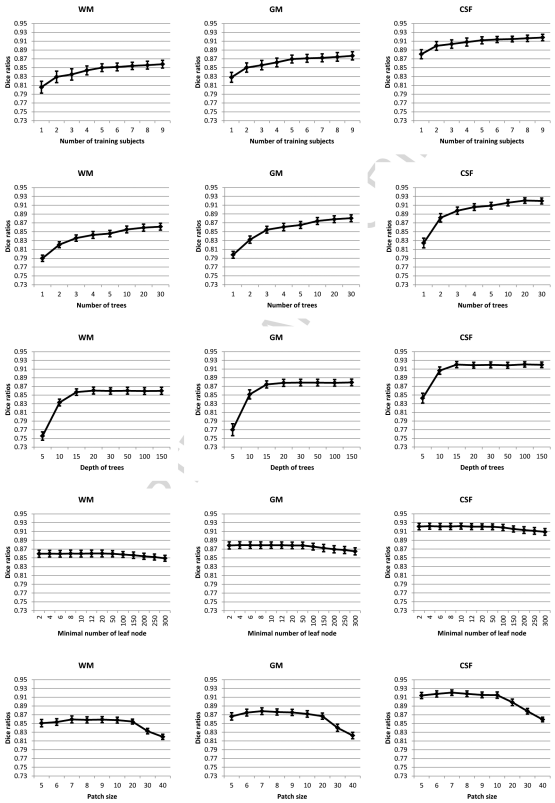

Fig. 5.

Influence of 5 different parameters: the number of training subjects (1st row), the number of trees (2nd row), the depth of each tree (3rd row), the minimally allowed number for the leaf node (4th row), and the patch size (last row).

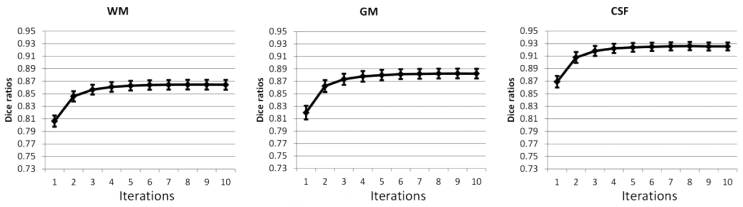

Fig. 6.

Changes of Dice ratios of WM, GM and CSF on 22 isointense subjects, with respect to the increase of iteration number.

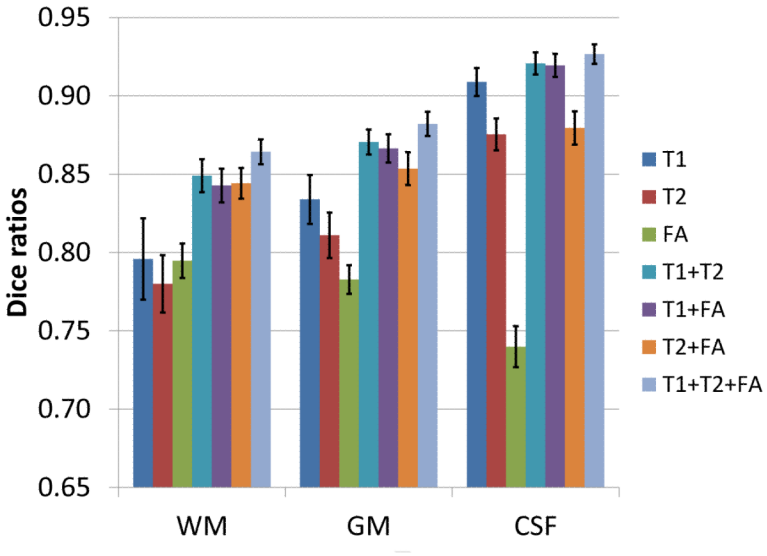

Fig. 7.

Average Dice ratios of the proposed method with respect to different combinations of 3 modalities.

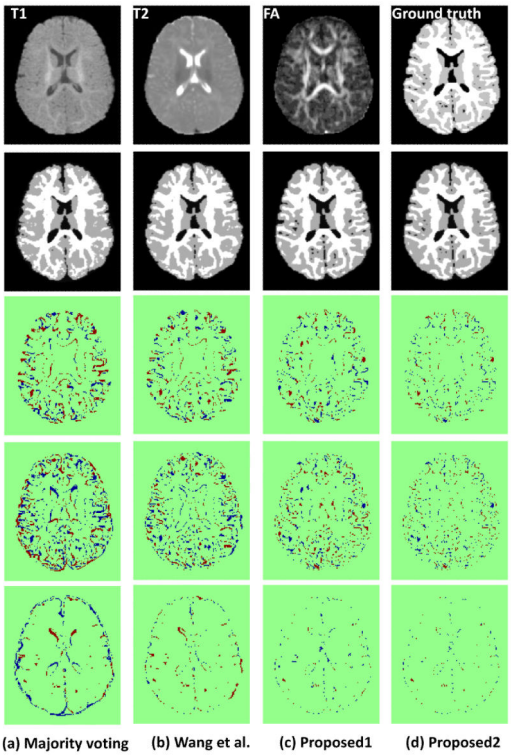

Fig. 8.

Comparison between (a) majority voting and (b) Wang et al.'s method (Wang et al., 2014a) on an isointense subject shown in **Fig. 3**. The first row shows the T1, T2, and FA images of this isointense subject. The second row shows the segmentation results of different methods. The last three rows show the label-difference maps (for WM, WM, and CSF, respectively), where the dark-red colors denote false negatives, while the dark-blue colors denote false positives. The columns (c) and (d) show the results by the proposed method *without* and *with* post-processing (i.e., using the anatomical constraint as described in Section 2.5), respectively.

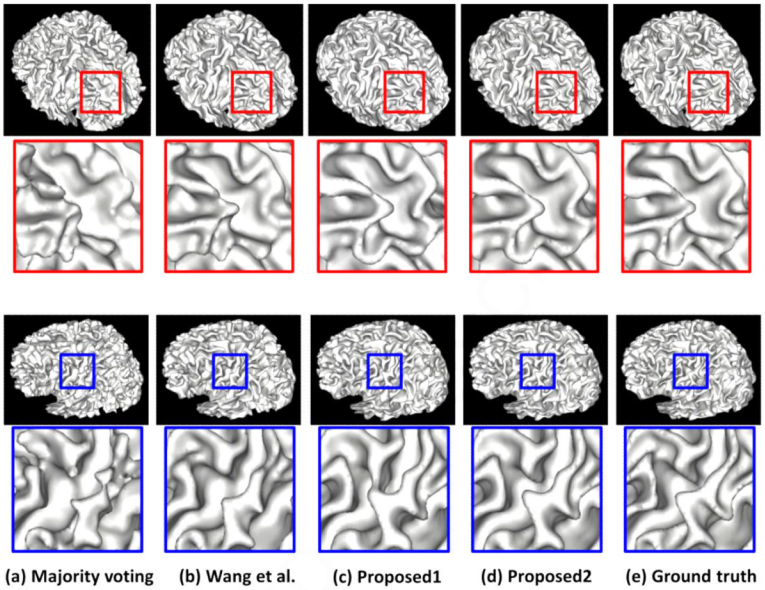

Fig. 9.

Comparison of white matter surfaces obtained with different methods (a-d), along with the ground truth shown in (e).

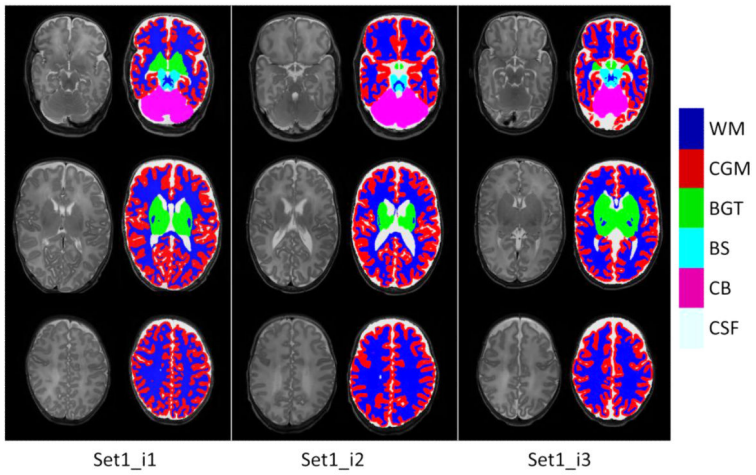

Fig. 10.

Segmentation results of the proposed method on three preterm born infants acquired at 40 weeks gestation corrected age, as provided by the MICCAI Grand Challenge on Neonatal Brain Segmentation (NeoBrainS12). Each column corresponds to three different slices of the same subject and their corresponding automatic segmentations. Abbreviations: WM - unmyelinated and myelinated whiter matter; CGM - cortical grey matter; BGT - basal ganglia and thalami; BS - brainstem; CB - cerebellum; CSF - ventricles and cerebrospinal fluid in the extracerebral space.

Table 1

A brief summary of the existing multi-atlas label fusion (MALF) methods for infant brain MR image segmentation.

Methods	Sources	Registration	Development stage at scan			Run-time
			Infantile	Isointense	Early adult-like	
(Weisenfeld and Warfield, 2009)	T1, T2	Nonlinear	✓			1h
(Shi et al., 2010)	T2	Nonlinear	✓			1h
(Srhoj-Egekher et al., 2013)	T1,T2	Nonlinear	✓			1h
(Wang et al., 2014b)	T2	Nonlinear	✓			2h
(Wang et al., 2014a)	T1, T2, FA	Nonlinear	✓	✓	✓	2h
Proposed method	**T1, T2, FA, Probability maps of WM, GM and CSF**	**Linear**	✓	✓	✓	**5m**

Table 2

Segmentation accuracies (Dice ratios in percentage) of 6 different methods on 119 infant subjects, along with information of both registration technique and runtime used by each method. Proposed1 and Proposed2 denote the proposed method *without* and *with* post-processing (i.e., using anatomical constraint as described in Section 2.5), respectively. Numbers (0, 3, 6, 9, and 12) denote months of age for the target subjects.

Method		**MV**	**Wang et al.**	**Proposed1**	**Proposed2**
Time cost		**1h**	**2h**	**5m**	**1.8h**
WM	0	81.6±0.28	90.1±0.59	91.7±0.64	92.1±0.62
	3	76.6±1.48	87.9±1.71	88.8±1.09	89.1±0.95
	6	80.1±0.83	84.2±0.78	86.4±0.79	87.9±0.68
	9	79.2±0.98	88.7±1.89	89.0±0.78	89.4±0.56
	12	82.5±1.05	91.1±1.42	91.3±0.74	91.8±0.65
GM	0	78.6±1.02	88.5±0.81	88.7±0.66	88.8±0.42
	3	77.3±1.42	87.5±0.51	88.1±1.00	88.3±0.90
	6	79.9±1.04	84.8±0.77	88.2±0.77	89.7±0.59
	9	83.6±0.69	88.4±0.54	89.5±0.49	90.3±0.54
	12	84.9±1.01	89.3±0.57	89.9±0.74	90.4±0.68
CSF	0	76.6±1.57	82.1±2.59	83.9±2.20	84.2±2.02
	3	80.6±1.55	84.6±1.10	85.1±1.52	85.4±1.49
	6	71.2±0.71	83.0±0.77	92.7±0.63	93.1±0.55
	9	68.7±1.27	82.4±2.27	83.0±1.53	83.7±1.09
	12	65.2±3.69	82.0±2.59	81.7±1.90	82.2±1.69

Table 3

Segmentation accuracies (MHD, in mm) of 6 different methods on 119 infant subjects. Proposed1 and Proposed2 denote the proposed method *without* and *with post-processing* (i.e., using anatomical constraint as described in Section 2.5), respectively. Numbers (0, 3, 6, 9, and 12) denote months of age for the target subjects. The upper part of table shows the results on WM/GM boundaries, while the bottom part shows the result on GM/CSF boundaries.

Method		MV	Wang et al.	Proposed1	Proposed2
WM / GM	0	1.84±0.14	1.25±0.21	1.13±0.20	1.02±0.20
	3	2.17±0.11	1.63±0.20	1.47±0.22	1.31±0.21
	6	2.06±0.18	1.75±0.21	1.46±0.13	1.33±0.10
	9	2.25±0.17	2.03±0.43	1.54±0.16	1.50±0.14
	12	2.08±0.25	1.34±0.21	1.28±0.25	1.20±0.24
GM / CSF	0	3.72±0.69	2.51±0.44	2.21±0.46	2.19±0.43
	3	4.35±0.53	2.49±0.33	2.26±0.44	2.21±0.38
	6	4.84±0.48	4.25±0.53	2.19±0.25	2.12±0.19
	9	4.53±0.43	2.98±0.34	2.19±0.36	2.14±0.34
	12	5.01±0.43	2.56±0.36	2.18±0.58	2.05±0.55

Table 4

Dice ratios (DC) and modified Hausdorff distance (MHD) of different methods on NeoBrainS12 Challenge data* (http://neobrains12.isi.uu.nl/mainResults_Set1_Original.php).

	WM		CGM		BGT		BS		CB		CSF	
Team Name	**DC**	**MHD**	**DC**	**MHD**	**DC**	**MHD**	**DC**	**MHD**	**DC**	**MHD**	**DC**	**MHD**
DCU	0.83	1.82	-	-	-	-	-	-	-	-	-	-
Imperial	0.89	0.70	0.84	0.73	0.91	0.8	**0.84**	1.04	0.91	0.7	0.77	1.55
Oxford	0.88	0.76	0.83	0.61	0.87	1.32	0.8	1.24	**0.92**	0.63	0.74	1.82
UCL	0.87	1.03	0.83	0.73	0.89	1.29	0.82	1.3	0.9	0.92	0.73	2.06
UPenn	0.84	1.79	0.80	1.01	0.8	4.18	0.74	1.96	0.91	0.85	0.64	2.46
UNC-IDEA	**0.92**	**0.35**	**0.86**	**0.47**	**0.92**	**0.47**	0.83	**0.9**	**0.92**	**0.5**	**0.79**	**1.18**

*The methods that used more training images than those provided by this challenge are not included.

Table 5

Overall ranks of different methods on NeoBrainS12 Challenge data (http://neobrains12.isi.uu.nl/mainResults_Set1_Original.php).

Team Name	Overall Rank	Placed	Last Update	Method Type
UNC-IDEA	**2.08**	**1**	**29-Apr-14**	**Automatic**
Imperial	3.94	2	4-Jul-12	Automatic
Oxford	4.45	3	15-Nov-12	Automatic
UCL	5.5	4	4-Jul-12	Automatic
UPenn	6	5	4-Jul-12	Automatic
DCU	7.25	6	29-Apr-14	Automatic