

Integration of Sparse Multi-modality Representation and Geometrical Constraint for Isointense Infant Brain Segmentation

Li Wang¹, Feng Shi¹, Gang Li¹, Weili Lin¹, John H. Gilmore², and Dinggang Shen¹

¹ Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

² Department of Psychiatry, University of North Carolina at Chapel Hill, NC, USA

Abstract. Segmentation of infant brain MR images is challenging due to insufficient image quality, severe partial volume effect, and ongoing maturation and myelination process. During the first year of life, the signal contrast between white matter (WM) and gray matter (GM) in MR images undergoes inverse changes. In particular, the inversion of WM/GM signal contrast appears around 6-8 months of age, where brain tissues appear isointense and hence exhibit extremely low tissue contrast, posing significant challenges for automated segmentation. In this paper, we propose a novel segmentation method to address the above-mentioned challenge based on the sparse representation of the complementary tissue distribution information from T1, T2 and diffusion-weighted images. Specifically, we first derive an initial segmentation from a library of aligned multi-modality images with ground-truth segmentations by using sparse representation in a patch-based fashion. The segmentation is further refined by the integration of the geometrical constraint information. The proposed method was evaluated on 22 6-month-old training subjects using leave-one-out cross-validation, as well as 10 additional infant testing subjects, showing superior results in comparison to other state-of-the-art methods.

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 tissue segmentation of infant brain MR images into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) in this stage is of great importance in studying the normal and abnormal early brain development. It is well-known that segmentation of infant brain MRI is considerably more difficult than that of the adult, due to the reduced tissue contrast [1], increased noise, severe partial volume effect [2], and ongoing WM myelination [1, 3] in the infant images. Actually, there are three distinct stages in the first year brain MR images, with each stage having quite different white-gray matter contrast patterns (in chronological order) [4]: (1) the infantile stage (≤ 5 months), in which the GM shows a higher signal intensity than the WM in T1 images; (2) the isointense stage (6-12 months), in which the signal intensity of the WM is increasing during the development due to the myelination and

maturation process; in this stage, the GM has the lowest signal differentiation with the WM in both T1 and T2 images; (3) the early adult-like stage (>12 months), where the GM intensity is much lower than that of the WM in T1 images, and this pattern is similar with that of the adult MR images. As an illustration, the first two images in the first row of Fig. 1 show examples of T1 and T2 images around 6 months. It can be observed that the WM and GM exhibit almost the same intensity level (especially in the cortical regions), resulting in the lowest image contrast and hence significant difficulties for tissue segmentation.

Although many methods have been proposed for infant brain image segmentation, most of them focused either on segmentation of the neonatal images (≤ 3 months) or infant images (>12 months) using a single T1 or T2 modality [2, 3, 5, 6], which demonstrates a relatively good contrast between the WM and GM. Few studies have addressed the difficulties in segmentation of the isointense infant images. Shi et al. [7] first proposed a 4D joint registration and segmentation framework for the segmentation of infant MR images in the first year of life. In this method, longitudinal images in both infantile and early adult-like stages were used to guide the segmentation of images in the isointense stage. A similar strategy was later adopted in [8]. The major limitation of these methods is that they fully depend on the availability of longitudinal datasets [9]. Due to the fact that the majority of infant images are single time-point, a standalone method working for cross-sectional single-time point image is mostly desired. Kim et al. [9] proposed an adaptive prior and spatial temporal intensity change estimation to overcome the low contrast. However, their work was only evaluated on the images acquired around 12 months. Moreover, none of these methods takes advantages from the geometrical information that the WM/GM and GM/CSF surfaces should be free of geometrical defects. For example, WM surfaces obtained by these methods are normally discontinuous, corrupted by holes (or handles, which are topologically equivalent) and “sharp breaks” in cortical gyri, which could be improved by imposing the geometrical constraint [10].

Motivated by the fact that many classes of signals, such as audio and images, have naturally sparse representations with respect to each other, sparse representation has been widely and successfully used in many fields, i.e., for visual tracking, compressive sensing, image de-noising, and face recognition [11, 12]. In this paper, we propose to employ the sparse representation technique for effective utilization of multi-modality information to address the isointense infant brain segmentation. Multi-modality information comes from T1, T2 and fractional anisotropy (FA) images (the first row of Fig. 1), which provide rich information of major WM bundles [13], to deal with the problem of insufficient tissue contrast [4]. Specifically, we first construct a library consisting of a set of multi-modality images from the training subjects and their corresponding ground-truth segmentations. Then we employ a patch-based method [14] to represent each patch of the testing multi-modality images by using a sparse set of library patches. The initial segmentation is thus obtained based on the majority label of library patches. By utilizing the geometrical constraint, the initial segmentation will be iteratively refined with further consideration of the patch similarities between the segmented testing image and the ground-truth segmentation in the library images.

2 Method

This study has been proved by institute IRB and the written informed consent forms were obtained from parents. A total of 22 healthy infant subjects (12 males/10 females) were recruited, and scanned at 27 ± 0.9 postnatal weeks. T1, T2 and FA images were acquired for each subject. T2 and FA images were then linearly aligned onto their corresponding T1 images. Image preprocessing includes resampling to $1 \times 1 \times 1 \text{ mm}^3$, bias correction, skull stripping, and cerebellum removal. To generate the ground truth segmentation, we took a practical approach by first generating an initial segmentation using a publicly available software iBEAT (www.nitrc.org/projects/ibeat). Manual editing was then performed by experienced raters to correct segmentation errors and geometric defects by using ITK-SNAP (www.itksnap.org) with the help of surface rendering, e.g., filling the holes.

2.1 Deriving Initial Segmentation from the Library by Sparse Representation

To segment a testing image $\mathbf{I} = \{I_{T1}, I_{T2}, I_{FA}\}$, N template images $\mathbf{I}^i = \{I_{T1}^i, I_{T2}^i, I_{FA}^i\}$ and their corresponding segmentation maps L^i ($i = 1, \dots, N$) are first nonlinearly aligned onto the space of the testing image using Diffeomorphic Demons [15], based on T1 images. Then, for each voxel x in each modality image of the testing image $\mathbf{I}$, its intensity patch (taken from $w \times w \times w$ neighborhood) can be represented as a $w \times w \times w$ dimensional column vector. By taking the T1 image as an example, the T1 intensity patch can be denoted as $\mathbf{m}_{T1}(x)$. Furthermore, its patch dictionary can be adaptively built from all N aligned templates as follows. First, let $\mathcal{N}^i(x)$ be the neighborhood of voxel x in the i -th template image I_{T1}^i , with the neighborhood size as $w_p \times w_p \times w_p$. Then, for each voxel $y \in \mathcal{N}^i(x)$, we can obtain its corresponding patch from the i -th template, i.e., a $w \times w \times w$ dimensional column vector $\mathbf{m}_{T1}^i(y)$. By gathering all these patches from $w_p \times w_p \times w_p$ neighborhoods of all N aligned templates, we can build a dictionary matrix $\mathbf{D}_{T1}$, where each patch is represented by a column vector. In the same manner, we can also extract T2 intensity patch $\mathbf{m}_{T2}(x)$ and FA intensity patch $\mathbf{m}_{FA}(x)$ and further build their respective dictionary matrices $\mathbf{D}_{T2}$ and $\mathbf{D}_{FA}$. Let $\mathbf{M}(x) = [\mathbf{m}_{T1}(x); \mathbf{m}_{T2}(x); \mathbf{m}_{FA}(x)]$ be the testing multi-modality patch and $\mathbf{M}^i(y) = [\mathbf{m}_{T1}^i(y); \mathbf{m}_{T2}^i(y); \mathbf{m}_{FA}^i(y)]$ be the template multi-modality patch in the dictionary. To represent the patch $\mathbf{M}(x)$ by the dictionaries $\mathbf{D}_{T1}$, $\mathbf{D}_{T2}$ and $\mathbf{D}_{FA}$, its coefficients vector α could be estimated by many coding schemes, such as sparse coding [11, 16] and locality-constrained linear coding [17]. Here, we employ sparse coding scheme [11, 16], which is robust to the noise and outlier, to estimate the coefficient vector α by minimizing a non-negative Elastic-Net problem [18],

$$\min_{\alpha \geq 0} \sum_k \|\mathbf{D}_k \alpha - \mathbf{m}_k(x)\|_2^2 + \lambda_1 \|\alpha\|_1 + \lambda_2 \|\alpha\|_2^2 \quad (1)$$

where $k \in \{T1, T2, FA\}$. In the above Elastic-Net problem, the first term is the data fitting term based on the intensity patch similarity, and the second term is the ℓ_1 regularization term which is used to enforce the sparsity constraint on the reconstruction

coefficients α , and the last term is the ℓ_2 smoothness term to enforce the coefficients to be similar for the similar patches. Each element of the sparse coefficient vector α , i.e., $\alpha^i(y)$, reflects the similarity between the target patch $M(x)$ and the patch $M^i(y)$ in the patch dictionary. Based on the assumption that similar patches should share similar labels, we use the sparse coefficients α to estimate the probability belonging to the j -th tissue, i.e., $P_j(x) = \sum_i \sum_{y \in \mathcal{N}^i(x)} \alpha^i(y) \delta_j(L^i(y))$, where $L^i(y)$ is the segmentation label (WM, GM, or CSF) for voxel y in the i -th template image, and $\delta_j(L^i(y))=1$ if $L^i(y) = j$; otherwise $\delta_j(L^i(y))=0$. Finally, $P_j(x)$ is normalized to ensure $\sum_j P_j(x) = 1$. The second row of Fig. 1 shows an example of the estimated probability maps for a testing image, with the original T1, T2 and FA images shown in the first row. To convert from the soft probability map to the hard segmentation, the label of the voxel x is determined using the *maximum a posteriori* (MAP) rule.

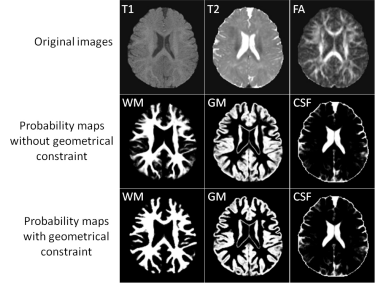


Fig. 1. Tissue probability maps estimated by the proposed method without and with the geometrical constraint

2.2 Imposing Geometrical Constraints into the Segmentation

The tissue probability maps derived in the Section 2.2 are purely based on the intensity patch similarity using the sparse representation technique. However, due to the low tissue contrast, the reliability of the patch similarity could be limited, which may result in considerable artificial geometrical errors in the tissue probability maps. A typical example is shown in Fig. 4(a), where we can observe many undesired holes (green rectangles), incorrect connections (red rectangles), and inaccurate segmentations (blue rectangles). In this section, we further address these problems by considering the geometrical constraint. As ground-truth segmentation results of template images in the library are almost free of the geometrical errors, we could expect the combination of these segmentation results will largely reduce the possible geometrical errors. Specifically, we can extract the patch $m_{seg}(x)$ from the tentative segmentation result of the testing image and also construct the segmentation patch dictionary D_{seg} from all the aligned segmented images in the library. Based on Eq. (1), we further incorporate the geometrical constraint to derive the tissue probability maps:

$$\min_{\alpha \geq 0} \sum_k \|D_k \alpha - m_k(x)\|_2^2 + v \|D_{seg} \alpha - m_{seg}(x)\|_2^2 + \lambda_1 \|\alpha\|_1 + \lambda_2 \|\alpha\|_2^2 \quad (2)$$

where $k \in \{T1, T2, FA\}$ and v is the weight parameters. In the same way, we can use the derived sparse coefficient vector α to estimate new tissue probabilities, which will be iteratively refined by using Eq. (2) until converged. An example of the probabilities derived with the geometrical constraint is shown in the third row of Fig. 1. Compared with the probability maps without the geometrical constraint (the second row), the new probability maps are more accurate.

3 Experimental Results and Analysis

The parameters used in this paper were determined via cross validation on a set of training images. We finally chose the following parameters for all experiments below: the weight for ℓ_1 -norm term $\lambda_1 = 0.2$, the weight for ℓ_2 -norm term $\lambda_2 = 0.01$, the patch size $w = 5$, the neighborhood size $w_p = 5$, and the weight for the geometrical constraint term $\nu = 1$.

Leave-One-Out Cross-Validation. To evaluate the performance of the proposed method, we adopted the leave-one-out cross-validation. Fig. 2(a) demonstrates the segmentation results of different methods for one typical subject. We choose to compare with the coupled level sets (CLS) method [19] provided by publicly available software iBEAT, in which multi-modality images from T1, T2 and FA were also employed. We also make comparison with the majority voting (MV) and conventional patch-based (CPB) method [14]. There are 7 different combinations for three modalities. The following rows only show the results by the proposed method with three representative combinations of three modalities (Eq. (1)). The last row shows the result by the proposed method on multi-modality images with the geometrical constraint (Eq. (2)). To better compare the results by different methods, the label differences compared with the ground-truth segmentation were also presented, which qualitatively demonstrate the advantage of the proposed method. We then quantitatively evaluate the performance of different methods by employing Dice ratio. The average Dice ratios of different methods on 22 subjects are shown Fig. 2(b). Besides the Dice ratio, we also measure the mean surface distance error between the generated WM/GM (GM/CSF) surfaces and the ground-truth surfaces, which are plotted in Fig. 2(b) and further demonstrate the accuracy of the proposed method. It is worth noting that any combination of these different modalities generally produce more accurate results than any single modality in terms of both Dice ratios and surface distance errors.

Results on 10 New Testing Subjects with Manual Segmentations. Instead of using the leave-one-out cross-validation fashion, we further validated our proposed method on 10 additional subjects, which were not included in the library. The manual segmentations by experts were referred to as our golden standard. The Dice ratios and surface distance errors on 10 subjects by different methods are shown in the Fig. 3, which again demonstrates the advantage of our proposed method.

Importance of the Geometrical Constraint. To further demonstrate the benefit of incorporating the geometrical constraint into the proposed method, we take the WM/GM surfaces as an example to compare the results by the proposed method without and with the geometrical constraint in Fig. 4. Fig. 4(a) shows the result without the geometrical constraint. It can be observed that there are many geometrical defects such as incorrect connections indicated by the red rectangle, inaccurate “zigzag” segmentations indicated by the blue rectangle and holes indicated by the green rectangle. The intermediate and final results by the geometrical constraint are shown in Fig. 4(b) and (c). It can be observed the incorrect connections and inaccurate “zigzag” segmentations are gradually corrected. Although the proposed method cannot guarantee the

topological correctness of the final WM/GM (GM/CSF) surface, the topological errors are largely reduced. By referring to the ground-truth segmentation shown in Fig. 4(d), the result with the geometrical constraint is much more accurate and reasonable than the result without the geometrical constraint, which can also be demonstrated by the quantitative evaluation results with the Dice ratios and surface distance errors as shown in Fig. 2(b).

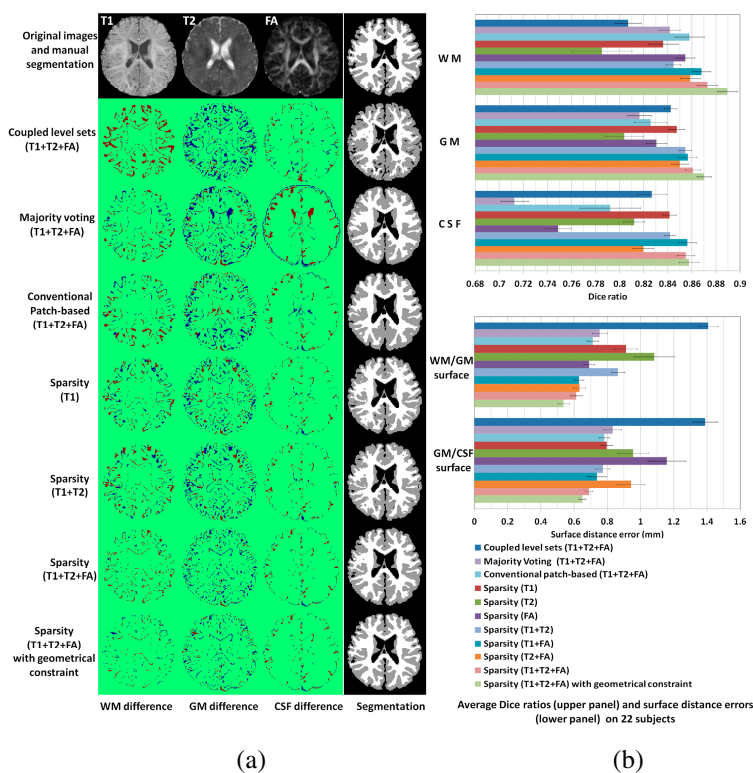


Fig. 2. (a) Comparison with the coupled level sets method [19], majority voting, conventional patch-based method [14] on T1+T2+FA images and the proposed sparsity method with different combinations of 3 modalities. In each label difference map, dark red colors indicate false negatives and the dark blue colors indicate false positives. (b) Average Dice ratios and surface distance errors on 22 subjects are shown in the right panel.

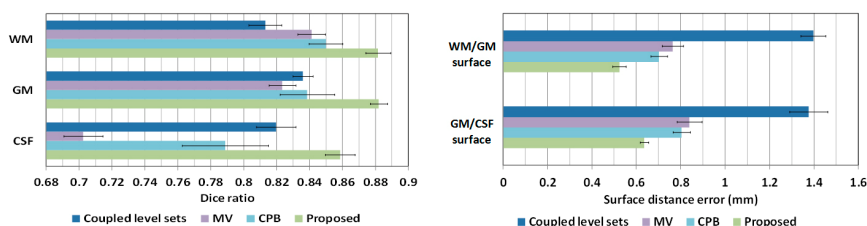


Fig. 3. The Dice ratios and surface distance errors on 10 subjects

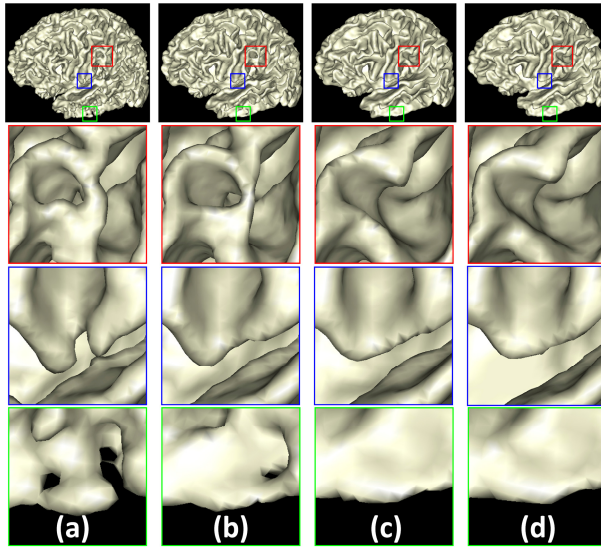


Fig. 4. Importance of using the geometrical constraint. From (a) to (c) shows the surface evolution from the initial stage to the final stage with the geometrical constraint. (d) is ground truth.

4 Discussion and Conclusion

In this paper, we have proposed a novel patch-based method for isointense infant brain MR image segmentation by utilizing the sparse multi-modality information. The segmentation is initially obtained based on the intensity patch similarity and then further refined with the geometrical constraint. The proposed method has been extensively evaluated on 22 training subjects using leave-one-out cross-validation, and also on 10 additional testing subjects. It is worth noting that our framework can also be directly applied to the segmentation of images in infantile and adult-like stages, for obtaining higher Dice ratios (compared with the isointense stage) due to their better contrast.

FA images provide rich information of major fiber bundles, especially in the sub-cortical regions where GM and WM are hardly distinguishable in the T1/T2 images. Therefore, FA images play a more important role in the WM/GM differentiation than GM/CSF differentiation, as demonstrated in Fig. 2.

In our experiment, we found that increasing the number of templates would generally improve the segmentation accuracy. However, more templates would also bring in more computational cost. In our test, when the number reached 20, the improvement rate of segmentation accuracies converged.

In our current method, the contributions of different modalities are equally weighted. In the future, we will further investigate to assign different weights to different modalities in different brain regions and validate on more datasets. In addition, our current library consists of only healthy subjects; therefore, it may not work well in pathological subjects. This will be our future work as well.

References

1. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. *Neuroimage* 47, 564–572 (2009)
2. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., et al.: Automatic segmentation and reconstruction of the cortex from neonatal MRI. *Neuroimage* (2007)
3. Gui, L., Lisowski, R., Faundez, T., Hüppi, P.S., et al.: Morphology-driven automatic segmentation of MR images of the neonatal brain. *Med. Image. Anal.* 16, 1565–1579 (2012)
4. Paus, T., Collins, D.L., Evans, A.C., Leonard, G., et al.: Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin* 54, 255–266 (2001)
5. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of MR images of the developing newborn brain. *Med. Image Anal.* 9, 457–466 (2005)
6. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. *Med. Image Anal.* 4, 43–55 (2000)
7. Shi, F., Yap, P.-T., Gilmore, J.H., Lin, W., et al.: Spatial-temporal constraint for segmentation of serial infant brain MR images. *MIAR* (2010)
8. Wang, L., Shi, F., Yap, P.-T., Gilmore, J.H., et al.: 4D Multi-Modality Tissue Segmentation of Serial Infant Images. *PLoS ONE* 7, e44596 (2012)
9. Kim, S.H., Fonov, V.S., Dietrich, C., Vachet, C., et al.: Adaptive prior probability and spatial temporal intensity change estimation for segmentation of the one-year-old human brain. *Journal of Neuroscience Methods* 212, 43–55 (2013)
10. Segonne, F., Pacheco, J., Fischl, B.: Geometrically Accurate Topology-Correction of Cortical Surfaces Using Nonseparating Loops. *IEEE Trans. Med. Imaging* 26, 518–529 (2007)
11. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., et al.: Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227 (2009)
12. Tong, T., Wolz, R., Hajnal, J.V., Rueckert, D.: Segmentation of brain MR images via sparse patch representation. In: *MICCAI Workshop on Sparsity Techniques in Medical Imaging (STMI)* (2012)
13. Liu, T., Li, H., Wong, K., Tarokh, A., et al.: Brain tissue segmentation based on DTI data. *Neuroimage* 38, 114–123 (2007)
14. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* 54, 940–954 (2011)
15. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. *Neuroimage* 45, S61–S72 (2009)
16. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: *CVPR*, pp. 1794–1801 (2009)
17. Wang, J., Yang, J., Yu, K., Lv, F., et al.: Locality-constrained Linear Coding for image classification. In: *CVPR*, pp. 3360–3367 (2010)
18. Zou, H., Hastie, T.: Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* 67, 301–320 (2005)
19. Wang, L., Shi, F., Lin, W., Gilmore, J.H., et al.: Automatic segmentation of neonatal images using convex optimization and coupled level sets. *Neuroimage* 58, 805–817 (2011)