Label Fusion for Multi-atlas Segmentation Based on Majority Voting

Jie Huo[1], Guanghui Wang[2(✉)], Q.M. Jonathan Wu[1], and Akilan Thangarajah[1]

[1] Department of ECE, University of Windsor, Windsor, ON N9B 3P4, Canada
[2] Department of EECS, University of Kansas, Lawrence, KS 66045, USA
ghwang@ku.edu

Abstract. Multi-atlas based segmentation is a popular approach in medical image analysis. Majority voting, as the simplest label fusion method in multi-atlas based segmentation, is a powerful tool for segmentation. In this paper, a novel majority voting-based label fusion algorithm is proposed by introducing a patch-based analysis for automatic segmentation of brain MR images. The proposed approach, by comparing the similarity between patches, avoids the over-segmentation problem of the majority fusion. The approach is successfully applied to the segmentation of hippocampus, and the experimental results demonstrate significant improvement over three state-of-the-art approaches in the literature.

Keywords: Multi-atlas segmentation · Majority voting · Label fusion

1 Introduction

Segmentation of anatomical structures in medical images is essential for scientific inquiry into the complex relationships between biological structure and clinical diagnosis, treatment and assessment. As a method of incorporating the prior knowledge and the anatomical structure similarity between target image and atlases, multi-atlas segmentation has been successfully applied in segmenting a variety of medical images, including brain, cardiac, and abdominal images [1–3]. Motivated by the observation that segmentation strongly correlates with image appearance, atlas segmentation transfers spatial information from an existing dataset (labeled atlas) to a target image via deformable registration based on image similarity. In multi-atlases segmentation, multiple atlases are separately registered to the target image, and voxelwise label conflicts between the registered atlases are resolved by using label fusion.

The label of the voxel in target image is determined by fusing the labels of corresponding voxels in each atlas. Weighted voting is the most popular method for label fusion, where the label of each atlas voxel contributes to the final result with a weight. One approach to obtaining the optimal weight is to compute the similarity between the image patch centered at the target voxel and the image

M. Kamel and A. Campilho (Eds.): ICIAR 2015, LNCS 9164, pp. 100–106, 2015.
DOI: 10.1007/978-3-319-20801-5_11

patch centered at the corresponding atlas voxel, and this forms the patch-based segmentation [4]. Other methods of considering correlated labelling errors [1] or employing sparse representation [5] are also used to determine the weight. In addition to weighted voting, another type of label fusion method is statistical label fusion, such as STAPLE [6], or non-local STAPLE [7].

Although weighted fusion and statistical fusion yield good results in segmentation of magnetic resonance (MR) image [1,2,7], the estimation of the weight and the EM estimation, which play important roles in weighted fusion and statistical fusion, is very computationally intensive. In contrast, majority voting, which is probably the simplest label fusion method, has been demonstrated to yield powerful segmentation results with less computation. Majority voting method, however, may yield over-segmentation since it does not utilize image intensity information. Patch-based method, which compares the similarity of intensity between patches, can be combined with majority voting multi-atlases segmentation to avoid such over-segmentation errors.

Motivated by this idea, we propose a novel label fusion method which combines majority voting with patch-based method to achieve automatic segmentation in brain MR images. The proposed method is successfully applied to the segmentation of hippocampus. In addition, the influences of different parameters are studied empirically, and a comparison with three closely related methods is performed to demonstrate the effectiveness of the proposed approach.

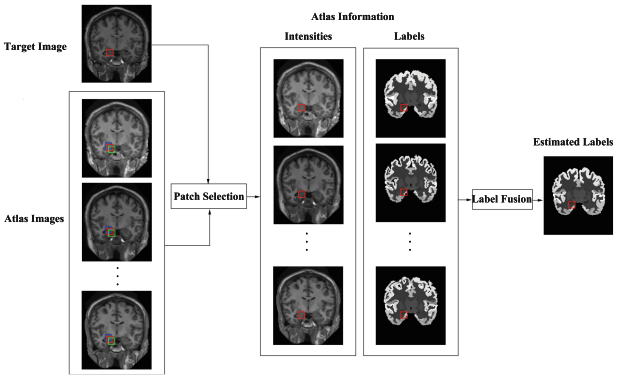

Fig. 1. Illustration of label generation for the target patch. Where red square in target image denotes the target patch; the blue, pink and green squares in atlas image denote patches in a searching window; and the best matched patch in each atlas is shown as red squares (Color figure online).

2 The Proposed Method

Consider an image $I = \{I(x)|x \in \Omega\}$, where x denotes the voxel; and $\Omega \subset \mathbb{R}^3$ denotes the lattice on which the image is defined. The goal of segmentation is

to estimate a label map L associated with the image I, in which each voxel is assigned a discrete label l. The label l takes discrete values from 1 to $\mathcal{L}$ for all the possible labels for the voxels in the image. In multi-atlas segmentation, I_T is a target image and $A_1, \cdots, A_n$ are n atlases with $A_i = (I_i, S_i)$, where I_i is the atlas image which has aligned to the target image (I_i is also called warped atlas image); and S_i is the corresponding manual segmentation of this atlas image. After combing the warped atlas images, a fused label map is generated which can be considered as the segmentation of the target image.

Figure 1 illustrates the generation of labels for the target patch of the proposed method. First, the atlases (intensity and label image) are pairwisely registered to the target image. Then, for each atlas image, a patch selection scheme is performed to choose the patch in each atlas with the highest similarity with the target patch. Finally, by applying label fusion algorithm to the patches with corresponding location of the patches in atlas images, we obtain the estimated label of each patch. The approach is applied for every voxel in the target image so as to obtain the labels for the entire target image.

2.1 Patch Selection

The performance of atlas-based segmentation can be moderately improved by applying a local searching technique [4]. Although deformable registration has been performed before label fusion, the correspondence obtained from the registration may not guarantee the maximal similarity between the patch in the target image and that in the warped atlas image. Therefore, local searching within a small neighborhood around the voxel in the warped image is performed to achieve the maximal similarity.

Summed squared distance (SSD) is used to measure the similarity between the target patch and that in the atlas image. The SSD between the patch centered at x in the target image and the patch centered at x' in the atlas image is shown below.

$$SSD(x, y) = \|I_T(\mathcal{N}(x)) - I_i(\mathcal{N}(x'))\|^2 \tag{1}$$

where $x' \in \mathcal{N}'(x)$ with $\mathcal{N}'(x)$ a local searched neighborhood. Equation (1) indicates that given a patch $I_T(\mathcal{N}(x))$ in the target image and $I_i(\mathcal{N}(x))$ in the ith atlas image, it is possible to find a patch $I_i(\mathcal{N}(x'))$ whose center belongs to the neighborhood $\mathcal{N}'(x)$. The patch centered at x^i, which is called locally searched optimal correspondence, has higher similarity with the target patch than other patches with centers inside the neighborhood $\mathcal{N}'(x)$. Thus, the locally searched optimal correspondence is

$$x^i = argmin_{x' \in \mathcal{N}'(x)}[SSD(I_T(\mathcal{N}(x)), I_i(\mathcal{N}(x')))] \tag{2}$$

where $I_i(\mathcal{N}(x'))$ is the patch in the ith atlas image centered at x' with a radius r, and $I_T(\mathcal{N}(x))$ is the target patch centered at x with a radius r. x' is the voxel in the local neighborhood $\mathcal{N}'(x)$ with a radius r_s. By calculating the SSD between the patches in the target and the atlas images, we obtain x^i, which is the location from the ith atlas with the best image matching for the location x in the target image.

2.2 Label Fusion and Validation

Majority Voting: After label fusion, n patches are selected as the candidate of voting for the target patch. The probability of that the label of x is l can be computed by counting the number of occurrence for l from x_i, $i \in 1, 2, \ldots, n$. Then, the label for x in the target image can be determined by choosing the label with the highest probability. The final label $\widehat{L}(x)$ is obtained by

$$\widehat{L}(x) = argmax_{l \in \{1,\ldots,\mathcal{L}\}} p_x(l) = argmax_{l \in \{1,\ldots,\mathcal{L}\}} \frac{1}{n} \sum_{i=1}^{n} p(l|A^i, x) \tag{3}$$

where x indexes through image voxels; $p(l|A^i, x)$ is the posterior probability that A^i votes for the label l at x. Typically, deterministic atlases have unique label for every location, which means $p(l|A^i, x) = 1$ if $S_i(x) = l$, and 0, otherwise.

Improvement on Majority Voting: The label of the center voxel of the target patch can be produced using majority voting. However, since we have chosen the most similar patch to the target patch from each atlas images based on the intensity information, these selected patches can be considered to have similar segmentation to the target patch. For each voxel in the target patch, we can find a candidate voxel from the corresponding position in each selected patch, and thus, the label of each voxel in target patch can be determined by performing (1) from its n candidate voxels. Given a three-dimensional image, for every patch with a radius r in the target image, $(2r+1)^3$ voxels within the patch will be labeled by performing the above majority voting scheme. Assuming that there are N voxels in the target image, $(2r+1)^3 \times N$ labels will be produced and each voxel in the target image have $(2r+1)^3$ candidates. Therefore, the majority voting is performed twice to generate the final label for the target voxel.

Validation: The kappa index (Dice coefficient or similarity index) [9] was computed by comparing the manual segmentations with those obtained with our method. For two binary segmentations A and B, the kappa index was computed as

$$\kappa(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{4}$$

In quantitative MR analysis, manual segmentation is usually considered as a gold standard. The segmentation quality was estimated with the Dice coefficient by comparing the expert-based segmentations with the automatic segmentations.

3 Experimental Evaluation

The proposed approach is applied to segment the hippocampus using T1-weighted MR images. The dataset used in the experiment includes 35 brain MR imaging scans obtained from the OASIS project. The manual brain segmentations of these images were produced by Neuromorphometrics, Inc., using the brain-COLOR labeling protocol. The dataset was applied in the MICCAI 2012

Multi-Atlas Labeling Challenge, where 15 subjects were used at the atlases and the remaining 20 images were used for testing.

In the experiment, we perform pairwise registered transformations between the atlas and the target images, as well as between each pair of the atlas images. The ANTs registration tool was used in this study to implement pairwise registration [10]. The antsApplyTransforms with linear interpolation was applied to generate the warped images, and the antsApplyTransforms with nearest neighbor interpolation was applied to generate the warped segmentations.

In order to improve computation efficiency, we select a region of interest (ROI) before computing. First, for every atlas image, a 3D binary image which segment the hippocampus region is generated. Then, OR operator is applied in these 3D binary images, and a new 3D binary image is obtained which fuses all the hippocampus segmentation of the atlas images. The resulted image is dilated by a $(2(r + r_s) + 1)$-dimensional cubic structuring element to produce the ROI of computation. In order to increase the robustness of image matching, instead of using the raw image intensities, we normalize the intensity vector obtained from each local image intensity patch such that the normalized vector has zero mean and a constant norm for each label fusion method.

3.1 Impact of the Size of 3D Patch and Search Volume

The proposed method has two parameters, r for the local patch radius and r_s for the local searched neighborhood. The influence of these parameters are studied by evaluating a range of values $r \in \{1, 2, 3\}$; $r_s \in \{1, 2, 3, 4, 5\}$ in the experiment. First, we studied the impact of the patch radius on segmentation accuracy. The mean dice overlap coefficient results are shown in Fig. 2 (left). Using the patch radius of $r = 1$, the algorithm performs much better than using larger patch radius. The segmentation accuracy also improves with the increase of the searched radius r_s. However, the dice overlap decreases when the searched radius $r_s > 4$. Larger searched radius improves the probability to find a similar patch with target patch, however, it also leads to an increase of mismatches. Figure 3 shows the segmentation results for different sizes of local patch and searched patch.

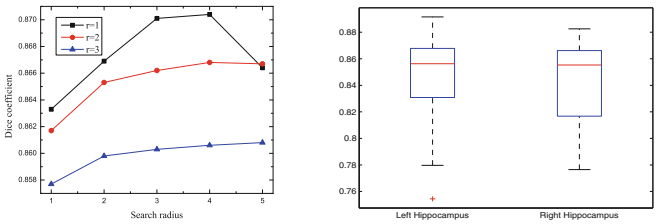

Fig. 2. (left) Hippocampus segmentation performance using different patch radius and searched patch radius. (right) The dice overlap coefficient of the left and right hippocampi.

3.2 Comparison Results in Hippocampus Segmentation

The average Dice overlap between automatic segmentation and manual segmentation for testing data is measure in the experiment. We compared our results with three popular benchmark approaches, i.e. majority voting, global weighted fusion, and STAPLE [8]. The dice overlap coefficient of the left and the right hippocampi by the proposed approach is 0.8473 ± 0.0325 and 0.8447 ± 0.0370, respectively, and the average overlap is 0.846 ± 0.03. The box plot is shown in Fig. 2 (right), where the central mark is the median, the edges of the box are the 25th and 75th percentiles. The whiskers extend to 2.7 standard deviations around the mean, and the outliers are marked individually as a '+'. As a comparison, the average Dice overlap obtained by majority voting, global weighted fusion, and STAPLE are 0.821, 0.807, and 0.836, respectively [8]. It is clear that the propose technique yields more than 1.2 % Dice overlap improvement. In addition, the results of other three approaches were obtained by conducting the experiments in a leave-one-out strategy on a data set containing 39 subjects, while our approach use only 15 subjects as atlas set. Overall, the proposed method yields better segmentation accuracy while using significantly fewer atlases than other reported methods.

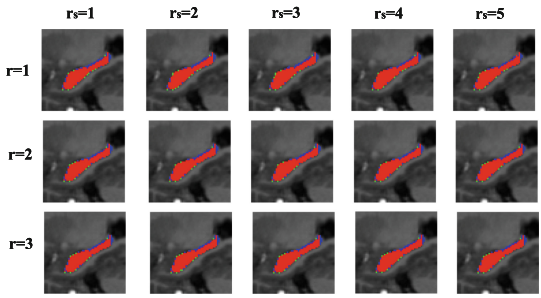

Fig. 3. Sagittal views of the segmentations produced by different patch radius and searched patch radius. Where the red region shows the overlap between the automatic and the manual segmentation; the green region is the manual segmentation; and the blue region is automatic segmentation using the proposed method (Color figure online).

4 Conclusion

In this paper, we have proposed a novel approach to automatically segment anatomical structures based on the majority voting method. A patch selection strategy is proposed to ensure that the patch in the atlas with the highest similarity to the target patch is selected as the voting candidate. The proposed approach is verified by experimental evaluations on a standard dataset. Compared with three benchmark techniques, the segmentation results are significantly improved by the proposed method.

Acknowledgment. The work is partly supported by the NSERC, Kansas NASA EPSCoR Program, and the NSFC (61273282).

References

1. Wang, H., Sun, J., Pluta, J., Craige, J., Yushkevich, P.: Multiatlas segmentation with joint label fusion. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 311–623 (2013)
2. Bai, W., et al.: A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans. Med. Imag. **32**(7), 1302–1315 (2013)
3. Wolz, R., et al.: Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Trans. Med. Imag. **32**(9), 1723–1730 (2013)
4. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Nonlocal patch-based label fusion for hippocampus segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 129–136. Springer, Heidelberg (2010)
5. Wu, G., et al.: Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition. NeroImage **103**, 34–46 (2015)
6. Asman, A.J., Landman, B.A.: Characterizing spatially varying performance to improve multi-atlas multi-label segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 85–96. Springer, Heidelberg (2011)
7. Asman, A.J., Landman, B.A.: Non-local statistical label fusion for multi-atlas segmentation. Med. Image Anal. **17**(2), 194–208 (2013)
8. Sabuncu, M.R., et al.: A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imag. **29**(10), 1714–1729 (2010)
9. Zijdenbos, A.P., et al.: Morphometric analysis of whitematter lesions in MR images: method and validation. IEEE Trans. Med. Imag. **13**, 716–724 (1994)
10. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**, 26–41 (2008)