Confidence-Guided Sequential Label Fusion for Multi-atlas Based Segmentation

Daoqiang Zhang[1,2], Guorong Wu[1], Hongjun Jia[1], and Dinggang Shen[1]

[1] Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599
[2] Dept. of Computer Science and Engineering, Nanjing
University of Aeronautics and Astronautics, Nanjing 210016, China
{zhangd,grwu,jiahj,dgshen}@med.unc.edu

Abstract. Label fusion is a key step in multi-atlas based segmentation, which combines labels from multiple atlases to make the final decision. However, most of the current label fusion methods consider each voxel equally and independently during label fusion. In our point of view, however, different voxels act different roles in the way that some voxels might have much higher confidence in label determination than others, i.e., because of their better alignment across all registered atlases. In light of this, we propose a sequential label fusion framework for multi-atlas based image segmentation by hierarchically using the voxels with high confidence to guide the labeling procedure of other challenging voxels (whose registration results among deformed atlases are not good enough) to afford more accurate label fusion. Specifically, we first measure the corresponding labeling confidence for each voxel based on the *k*-nearest-neighbor rule, and then perform label fusion sequentially according to the estimated labeling confidence on each voxel. In particular, for each label fusion process, we use not only the propagated labels from atlases, but also the estimated labels from the neighboring voxels with higher labeling confidence. We demonstrate the advantage of our method by deploying it to the two popular label fusion algorithms, i.e., majority voting and local weighted voting. Experimental results show that our sequential label fusion method can consistently improve the performance of both algorithms in terms of segmentation/labeling accuracy.

1 Introduction

In computational anatomy, the accurate image segmentation and labeling is a critical step for many clinical studies, such as pathology detection and brain parcellation. Although a lot of automatic image segmentation methods have been investigated, it is still a hot topic in medical image analysis.

Recently, multi-atlas based segmentation methods have shown great success in segmenting brain into anatomical structures [1-7]. A typical multi-atlas based segmentation procedure contains two major steps: 1) image registration step to register each atlas image to the target image and warp the corresponding label image by following the same estimated deformation field; and 2) label fusion step to combine the multiple propagated labels from different atlases to obtain the final labels

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part III, LNCS 6893, pp. 643–650, 2011.

of the target image by some heuristics. The current methods usually complete this procedure by independently performing the two steps, without considering the relationship between them.

A number of label fusion strategies have been proposed for multi-atlas based segmentation in the literature. Among them, majority voting (MV) is probably the simplest one and has been widely used in medical image segmentation. In MV, the candidate segmentations from each atlas are equally weighted and the label with largest agreement from all atlases is assigned as the final label. A natural extension of MV is to improve from simple averaging to adaptive weighted averaging. In [4], various weighting strategies are categorized into two groups, i.e., global weighted voting and local weighted voting, and it has been shown that the local weighted method outperforms the global solution when segmenting high-contrast brain structures. On the other hand, instead of using all atlases, using a selected subset of atlases usually results in improved performance [5]. The recent trend is to use the more advanced learning based methods to further improve performances of label fusion. In [6], a probabilistic label fusion method is proposed to explicitly model the relationship between the atlas and the target image. To avoid the possible registration error, a non-local label fusion method based on the patch-based strategy is proposed in [7], which is widely used in the machine learning and computer vision community.

One common limitation of existing methods is that the label fusion procedure treats each voxel independently and equally. Although it simplifies the labeling procedure, the coherent spatial correlations between neighboring voxels are ignored, which are actually very useful to achieve the accurate and robust segmentation. Moreover, due to the huge anatomical variations in the population, the registration accuracy varies not only across different subjects but also at different locations of the same subject. As a result, some voxels may have more reliable estimation on labels because of the more accurate alignment. In light of this, we should first perform label fusion on voxels with higher labeling confidence, and then use the estimated (usually more reliable) labels to guide the label fusion of the neighboring voxels with lower labeling confidence, in a sequential way.

Based on this observation, in this paper, we present a novel sequential label fusion framework for multi-atlas based segmentation. First, a novel criterion on labeling confidence is defined which considers not only the similarity but also the matching consistency of the two underlying local patches. Base on the labeling confidence on each voxel, the label procedure is sequentially performed from the voxels with high confidence to the voxels with low confidences. Thus, our method contains two subsequent steps, i.e., 1) labeling confidence estimation, and 2) sequential label fusion. Specifically, we first estimate the labeling confidence for each voxel based on the *k*-nearest neighbor (*k*-NN) rule [8], and then perform label fusion sequentially according to the measured confidence. When labeling each voxel, we use not only the propagated labels from the atlases, but also the already estimated labels from the neighboring voxels with higher labeling confidence. To the best of our knowledge, the proposed idea on sequential label fusion is new in the multi-atlas based segmentation. Our framework is general and can be integrated with any other label fusion methods. As confirmed in the experiment on NIREP dataset [9], our method is able to significantly improve the labeling accuracy, in comparison with the two popular label fusion algorithms: majority voting (MV) and local weighted voting (LWV).

2 Methods

In this section, we will present the sequential label fusion framework which consists of two steps: 1) labeling confidence estimation, and 2) sequential label fusion. We will first introduce how to estimate the registration confidence through *k*-NN searching in Section 2.1, and then detail the flow chart of the sequential label fusion framework in Section 2.2.

2.1 Labeling Confidence Estimation

Here, we use T to denote the target image to be labeled, and assume that all atlases and their associated label images have been already registered with the target image T, represented as $\{(A_i, L_i)| i=1,\ldots,M\}$. Thus, the target image T and all atlases are considered in the same domain Ω, and the procedure of label fusion on each voxel p in target image T is performed among a stack of $\{A_i(p), L_i(p)\}$ for each voxel p ($p \in \Omega$).

To measure the importance of each voxel in label fusion, we propose a novel criterion, called *labeling confidence*, which is based on local image appearance and matching consistency. Given voxel p in T, the calculation of its labeling confidence with respect to atlas A_i is performed in two steps:

(1) *Forward Matching:* We first compute the patch difference between *fixed voxel p* in T and *moving voxel o* in the forward search neighborhood (dented as $N_f(p)$ and displayed as the red circle in Fig. 1) in atlas A_i, and then perform a *k*-NN search to find voxel p's k nearest neighbors q_s^i (displayed as blue squares in the atlases, Fig.1). We denote $Q_i(p) = \{q_s^i \mid s = 1,...,k\}$ as the set of those k neighbors.

(2) *Backward Matching*: For each voxel q_s^i in $Q_i(p)$, we perform another backward matching as follows. We first compute the patch difference between fixed voxel o and the moving voxels p' in the backward search neighborhood (dented as $N_b(q_s^i)$ and displayed as the blue circles in Fig. 1) in the T, and then perform a *k*-NN search to find voxel q_s^i 's k nearest neighbors $u_s^i(t)$ (displayed as green triangles in T in Fig.1). Similarly, we denote $P_s^i(p) = \{u_s^i(t) \mid t = 1,...,k\}$ as the set of those k neighbors.

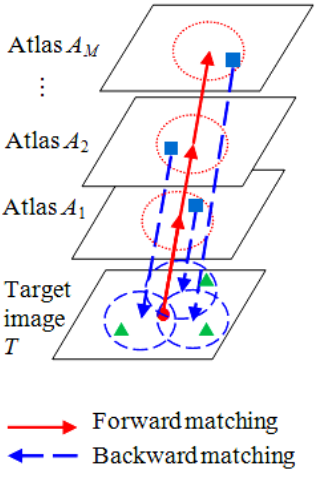

Fig. 1. Illustration on forward and backward matching for labeling confidence estimation with k=1

Then, the labeling confidence of the voxel p w.r.t. atlas A_i is computed as:

$$C_i(p) = \frac{1}{k^2}\sum_{s=1}^{k}\sum_{t=1}^{k}\exp\left(-\delta_1^{-1}(p)D\left(p,q_s^i\right)\right)\exp\left(-\delta_2^{-1}(p)\left\|p-u_s^i(t)\right\|_2^2\right), \quad (1)$$

with

$$\delta_1(p) = \min_{q_s^i \in Q_i(p)} D\left(p,q_s^i\right)+\varepsilon, \text{ and } \delta_2(p) = \min_{u_s^i(t)\in P_s^i(p)}\left\|p-u_s^i(t)\right\|_2^2+\varepsilon$$

used for normalization, where ε is a small constant. In Eq. 1, the first term after the sum symbols is the normalized patch difference (based on local image appearance) under the distance function D, while the last term measures the matching consistency. Here, the intuition is that for good registration the distance between p and $u_s^i(t)$, weighted by the local appearance similarity, are required to be as small as possible after forward and backward matching. In this way, both image appearance and the consistency are embedding in the measurement of labeling confidence.

The confidence degree estimated in Eq. 1 for each atlas A_i will be used as a weighting map for weighted voting in the following section. Moreover, we can average the confidence degree $C_i(p)$ from individual atlas A_i to measure overall confidence degree at each target image voxel p, i.e., $C(p)=(1/M)\sum_{i=1}^{M}C_i(p)$, which will be used to guide the sequential label fusion in next subsection. After that, we call $C(p)$ as *labeling confidence*.

2.2 Sequential Label Fusion

The estimated registration confidence map $C=\{C(p)|p\in\Omega\}$ reflects the registration confidences of different voxels in the target image T. Thus, based on the confidence map, we present a sequential label fusion framework, as shown in Fig. 2. For simplicity, we focus on the binary segmentation of anatomical structures, with label '1' indicating the structure of interest and '0' for the others. We first generate the initial voxels set V for the target image T according to the union of all L_is. Then, we start from the voxel p in V with the highest labeling confidence of $C(p)$ and compute its soft label (i.e., segmentation probability) by

$$L(p) = \frac{1}{Z_p}\left[l(p\mid L_1,...,L_M)+\sum\nolimits_{p'\in NH(p)}L(p')\exp\left(-\frac{1}{\sqrt{3}r}\left\|p-p'\right\|_2^2\right)\right]. \quad (2)$$

Here, $l(p|L_1,\ldots,L_M)$ denotes the propagated soft label from atlases for voxel p which can be obtained by using any label fusion algorithms, $NH(P)$ is the set of the neighboring voxels in the neighborhood $N(p)$ which satisfies $C(p')>C(p)$, and Z_p is a normalizing term to make $L(p)\in[0,1]$. It is obvious that the procedure of label fusion on the voxel of lower confidence will be guided by the neighboring voxels which have higher confidence and are thus able to correctly identify their labels with low risk (see the second term in the square bracket in Eq. 2). This process is iterated until all voxels in the volume have their labels determined. Finally, the final label of voxel p is computed as: sign($L(p)$-0.5).

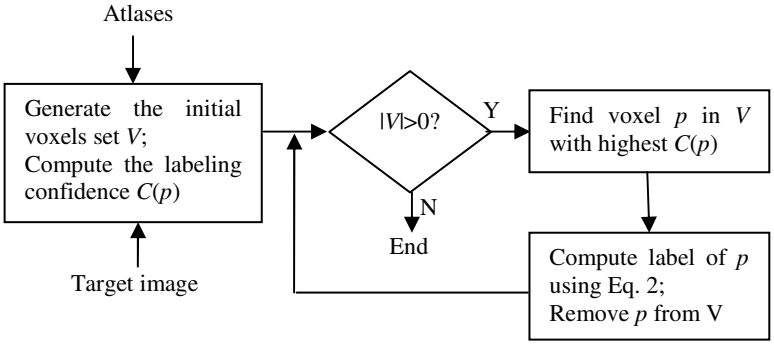

Fig. 2. Flow chart of the proposed sequential label fusion framework

3 Experiments

In this section, we evaluate the performance of the proposed sequential label fusion method. We integrate our sequential label fusion framework with the two most widely used algorithms, i.e., majority voting (MV) and local weighted voting (LWV). In LWV, the weights are based on the mean square distance on the local regions ($r \times r \times r$) between target image and atlas [4]. We denote the new MV and LWV algorithms equipped with our sequential label fusion as SC-MV and SC-LWV, respectively. In order to show the advantage of sequential labeling strategy, we also compare with two other variants which only use the confidence maps C_i (i=1,…,M) as weights but still treat each voxel independently. We denote these two methods as C-MV and C-LWV, respectively. In the following experiment, we perform all these six algorithms on the NIREP database [9].

The NIREP database [9] consists of 16 T1-weighted MR image (8 normal males and 8 females) with 32 manually delineated regions of interest (ROIs). For each of the ROIs, a Leave-One-Out (LOO) cross-validation is performed to test the segmentation performance, and the averaged Dice overlap measures are reported. Specifically, at each LOO fold, all other 15 subjects are used as the atlases and aligned onto the remaining image (used as target image) for guiding the segmentation. In our method, the sizes of forward and backward search neighborhood N_f and N_b are set to 5, and a local patch of size 3 is used to compute the Euclidean distance $D(p, q)$ between voxels p and q. k is set to 3, which means 3 candidates with the smallest patch difference are selected in k-NN search. For other method (MV and LWV), we use their optimal parameters.

Fig. 3 plots the confidence map and five intermediate segmentation results of SC-LWV. In Fig. 3, we show the overall labeling confidence map, along with the segmentation results at the five stages that top 20%, 40%, 60%, 80% and all pixels (from column (b)~(f)) are labeled with the guidance of the confidence map C (column (a)). In this way, Fig. 3 validates that high-confidence pixels should guide the labeling procedure on low-confidence pixels, in order to increase the label accuracy, as proposed in our method.

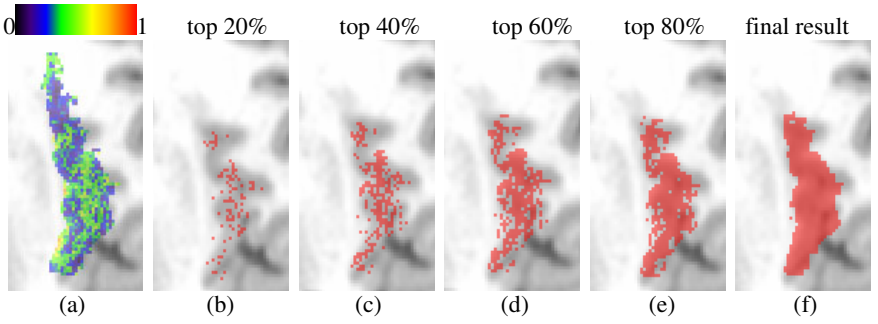

Fig. 3. Plots of the confidence map (a) and five intermediate segmentation results (b-f) using SC-LWV on *Left insula gyrus* segmentation

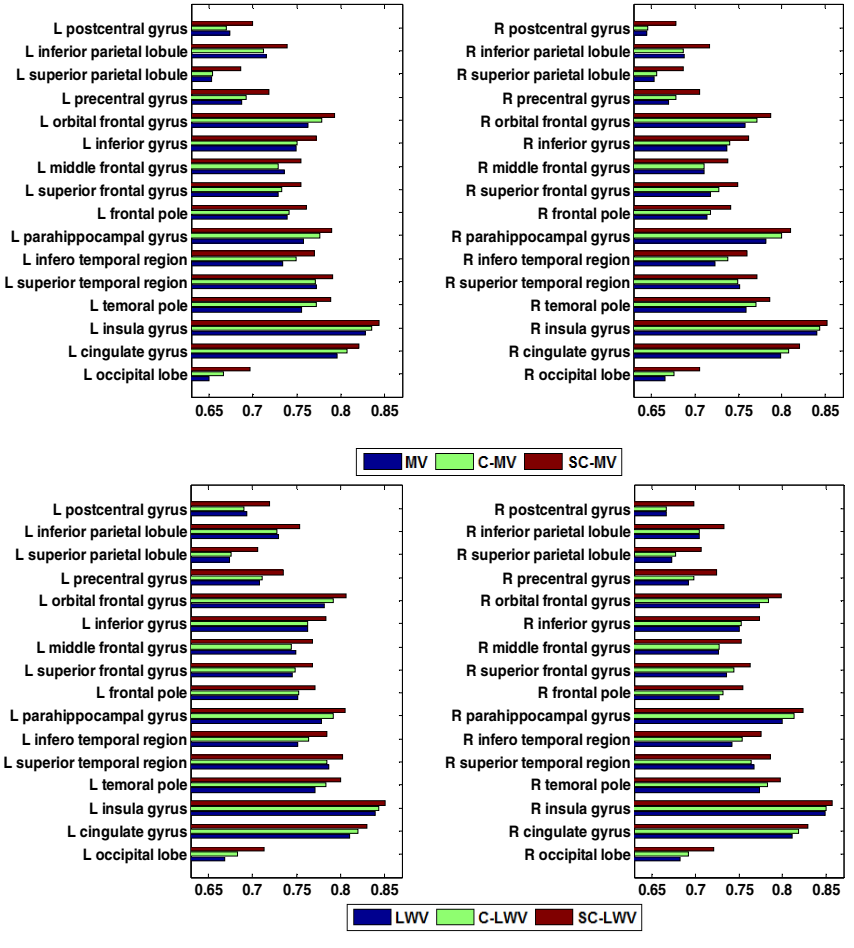

Fig. 4. Segmentation results on different brain structures measured by Dice overlap with six different label fusion algorithms

Fig. 4 shows the segmentation results on different brain structures (ROIs) using the six multi-atlas label fusion based segmentation algorithms. Here, the performance is measured with the Dice overlap, i.e., $Dice(S_a, S_b) = 2|S_a \cap S_b|/(|S_a| + |S_b|)$, where $\cap$ indicates the overlapping voxels between the two segmentations, and $|S_a|$ indicates the number of voxels of the corresponding segmentation.

As we can see from Fig. 4, SC-MV and SC-LWV consistently improve the performance of MV and LWV, respectively, on all ROIs. For example, on 'L occipital lobe' ROI, SC-MV improves the overlap ratio from 0.650 (by MV) to 0.697, and SC-LWV improves the overlap ration from 0.669 (by LWM) to 0.713. Fig. 4 also indicates that, in most cases, C-MV and C-LWV outperform MV and LWV respectively, but they are inferior to both SC-MV and SC-LWV in all cases. This demonstrates the importance of using the confidence-guided sequential labeling for ROI segmentation. Moreover, Fig. 4 shows that the LWV-based methods (LWV, C-LWV and SC-LWV) usually outperform the corresponding MV-based methods (MV, C-MV and SC-MV), which is consistent with previous studies [2, 4].

Finally, to have a summary on the segmentation accuracy on all 32 ROIs, we give the box plot for the results of the six algorithms, as shown in Fig. 5. It is observable that SC-MV and SC-LWV significantly improve the performance of the segmentation results by MV and LWV, respectively, while C-MV and C-LWV only slightly improve over the baseline methods (MV and LWV). Specifically, the averaged Dice overlap of MV, C-MV, SC-MV, LWV, C-LWV and SC-LWV are 0.730, 0.736, 0.758, 0.746, 0.751 and 0.772, respectively. Furthermore, the significant tests using paired t-test show that results of the proposed SC-MV and SC-LWV are significantly better than those of MV and LWV, respectively, at the 95% significance level. It validates the efficacy of the proposed sequential label fusion method.

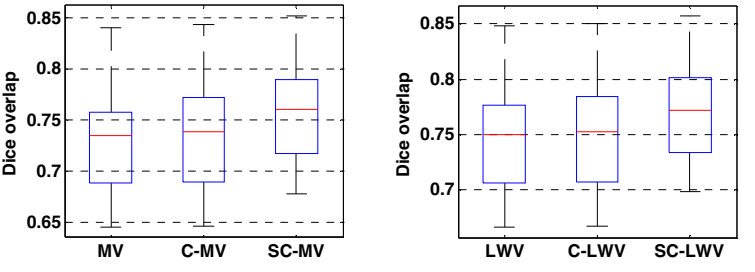

Fig. 5. Averaged Dice overlap using six different label fusion algorithms

4 Conclusion

We have presented a new label fusion method for multi-atlas based segmentation. In contrast to most existing label fusion methods which equally treat each voxel during label propagation, the proposed method considers the dependency among neighboring voxels for sequential labeling the voxels with confidence from high to low. To

achieve the sequential labeling, we define the labeling confidence which embeds not only the patch similarity but also the matching consistency to resolve the anatomy uncertainty in label fusion. Our method can be easily integrated with the current label fusion methods to significantly improve the label accuracy. Further work includes investigation of different estimations of label confidence, the effect of the sizes of local patch and neighborhood in label fusion, and more comparison with existing label fusion methods, e.g., STAPLE [10].

Acknowledgments. This work was supported in part by NIH grants EB006733, EB008374, EB009634 and MH088520, and also by National Science Foundation of China under grant No. 60875030.

References

1. Lotjonen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 49, 2352–2365 (2010)
2. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. IEEE Trans. Med.l Imag. 28, 1000–1010 (2009)
3. Langerak, T.R., van der Heide, U.A., Kotte, A.N., Viergever, M.A., van Vulpen, M., Pluim, J.P.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans. Med.l Imag. 29, 2000–2008 (2010)
4. Artaechevarria, X., Munoz-Barrutia, A., de Solorzano, C.O.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. IEEE Trans. Med.l Imag. 28, 1266–1277 (2009)
5. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage 46, 726–738 (2009)
6. Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE Trans. Med.l Imag. 29, 1714–1729 (2010)
7. Coupe, P., Manjon, J.V., Fonov, V., Pruessner, J.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage 54, 940–954 (2011)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley and Sons, Inc., Chichester (2000)
9. Christensen, G., Geng, X., Kuhl, J., Bruss, J., Grabowski, T., Pirwani, I., Vannier, M., Allen, J., Damasio, H.: Introduction to the non-rigid Image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
10. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans. Med.l Imag. 23, 903–921 (2004)