

Feature Sensitive Label Fusion with Random Walker for Atlas-Based Image Segmentation

Siqi Bao^(✉) and Albert C.S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Kowloon, Hong Kong
sbao@cse.ust.hk

Abstract. In this paper, a novel label fusion method is proposed and formulated on a graph, which embraces both label priors from atlases and anatomical priors from target image. To represent a pixel in a comprehensive way, three kinds of feature vectors are generated, including intensity, gradient and structural signature. Feature Sensitive Label Prior (FSLP), which takes both the consistency and variety of different features into consideration, is proposed to gather atlas priors. As FSLP is a non-convex problem, one heuristic approach is further designed to solve it efficiently. Moreover, based on anatomical knowledge, parts of the target pixels are also employed as graph seeds to assist the label fusion process. The experiments carried out on two publicly available databases give results to demonstrate that the proposed method can obtain better segmentation quality.

1 Introduction

The human brain is a complex neural system composing many anatomical structures. To study the functional and structural properties of its subcortical regions, image segmentation is a critical step in quantitative brain image analysis and clinical diagnosis. Researchers have been focusing on developing automatic atlas-based segmentation methods which can effectively incorporate expert prior knowledge about the relationships between local intensity profiles and tissue labels. Recently, more label fusion methods based on patches [2, 9–11] have become prevalent in medical image segmentation. Generally, there are three stages in patch-based label fusion: first determine which kind of feature to be adopted as pixel representation, then distinguish candidate pixels or patches for voting and finally fuse the labels of candidate atlas nodes.

However, due to the poor contrast condition in brain Magnetic Resonance (MR) images and similar histogram profiles among adjacent structures, general label fusion methods can be adversely influenced and lead to a less satisfactory result. As suggested in [1], embracing more features, such as contextual information, can help improve the segmentation quality. As such, besides conventional intensity and gradient features, we also extract high-level property of each subcortical structure using Convolutional Neural Networks [7] as structural signature, aiming to obtain a more discriminative pixel representation.

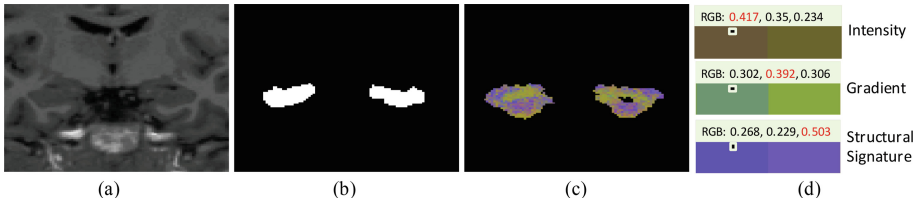


Fig. 1. Effects of feature sensitivity. **(a)** Cropped target intensity image. **(b)** Manually labeled Hippocampus. **(c)** Color image to show the effects of feature sensitivity, with RGB values standing for the feature coefficients of intensity, gradient and structural signature respectively. **(d)** 3 selected examples to illustrate their dominate features.

Rather than treating each kind of feature equally when fusing labels from atlas nodes [1], we propose to consider the feature sensitivity during label fusion and Feature Sensitive Label Prior (FSLP) is designed to capture priors from atlases. The **motivations** to introduce FSLP are two-fold. On the one hand, the contributions of distinct features are expected to be consistent during label fusion, i.e., they can reach an agreement when labeling a pixel. On the other hand, the impact of different features can change according to image conditions. For the flat regions away from structural boundaries, intensity and gradient are supposed to be more essential. As for the complex region near tissue boundaries, structural signature should play a more significant role. The experimental result with our method also justifies the initial motivation, as shown in Fig. 1.

In addition to FSLP from atlases, anatomical priors from target image are also utilized to assist our graph-based label fusion process. Based on anatomical knowledge, to label a pixel which is deep inside or outside a subcortical structure is easier, while to label one which is located around the boundary is challenging. As such, rather than updating labels for all target pixels, those far away from structural border are selected as graph seeds and their influence can be propagated to other pixels through image lattice. Unlike the graph-based labeling constructed with both atlas and target nodes [6], we further infer an equal but more concise graph to encode FSLP and anatomical prior, which only relies on target nodes. The objective energy function on the graph is formulated with Random Walker and can be solved as a discrete Dirichlet problem. To evaluate the proposed method, experiments have been carried out on two image databases and results demonstrate that our approach can obtain better performance as compared with other state-of-the-art methods.

2 Methodology

In this paper, to obtain a more discriminative representation, three kinds of features are extracted for each pixel, including intensity, gradient and structural signature. Given the demands of consistency and variety among distinct feature vectors during label fusion, a novel method FSLP is proposed to deal with this dilemma, by collecting priors from atlases with feature sensitivity. Moreover, the

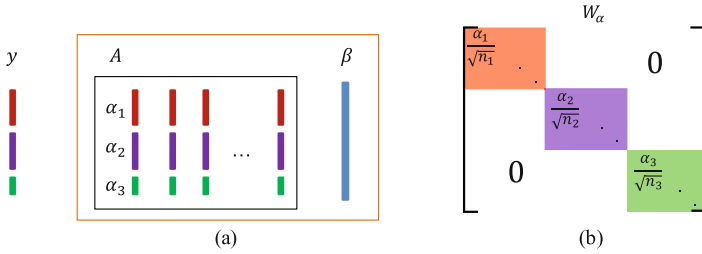


Fig. 2. (a) FSLP Illustration. y is the feature vector for the target pixel, concatenating intensity (Red), gradient (Purple) and structural signature (Green). α_i is feature coefficient and A is a dictionary constructed with atlas feature vectors. β is the vector storing reconstruction weights. (b) Feature sensitive matrix W_α . Each diagonal submatrix (Red W_1 , Purple W_2 , Green W_3) corresponds to one kind of feature vector.

pixels from target image are also selected based on anatomical knowledge, acting as anatomical prior. The whole label fusion process is modeled on an undirected graph and formulated under the framework of Random Walker.

2.1 Feature Sensitive Label Prior

FSLP captures label prior from atlases by seeking for the optimal linear combination of atlas nodes to reconstruct the feature vector of the target pixel, as illustrated in Fig. 2(a). For each considered pixel from the target image, its three kinds of features are extracted and concatenated together to formulate one augmented vector y . Its similar pixels selected from the atlases are assembled as dictionary A . Given that the confidence and significance of different features can vary considerably, the feature coefficient a_i is introduced to balance their influences. The optimal weight to reconstruct y with dictionary A is stored in vector β . The formulation of FSLP is given as follows:

$$\min_{\alpha, \beta} \frac{1}{2} |W_\alpha(y - A\beta)|_2^2 + \lambda |\alpha|_2^2, \quad s.t. \quad \sum_i \alpha_i = 1, \quad \alpha_i \geq 0. \quad (1)$$

W_α is the feature sensitive matrix, with its definition illustrated in Fig. 2(b). W_α is split into three subregions (Red W_1 , Purple W_2 and Green W_3), each corresponding to one kind of feature vector. The diagonal elements in the submatrix W_j are defined as: $\forall w_{ii} \in W_j, w_{ii} = \frac{\alpha_j}{\sqrt{n_j}}$, where n_j is the length of the j -th feature vector. Through the division between the coefficient α_j and $\sqrt{n_j}$, the normalization on various features is enforced in the feature sensitive matrix. In Eq. (1), with the regularization term on the coefficient vector α , it guarantees that no feature dominates the whole optimization procedure.

By solving Eq. (1), optimal feature coefficient α and reconstruction weight β can be obtained and label prior can be then estimated with grouped reconstruction error. However, since FSLP needs to optimize α and β simultaneously, it

turns into one non-convex problem, which may have multiple local optima and can be difficult to solve. A further solution is proposed for it.

Problem Solution. When optimizing α and β simultaneously, Eq. (1) is not a convex problem. To solve this non-convex problem efficiently, one heuristic approach is proposed in this paper by seeking optimal solutions for α and β alternately. The first step is to fix α and Eq. (1) turns into one least square problem:

$$\min_{\beta} \|W_{\alpha}(y - A\beta)\|_2^2. \quad (2)$$

This optimization problem is convex and its solution is $\hat{\beta} = (W_{\alpha}A) \setminus (W_{\alpha}y)$. With updated β , the second step is to fix it and Eq. (1) is then simplified to one quadratic programming problem:

$$\min_{\alpha} \frac{1}{2} \alpha^T A \alpha, \quad s.t. \sum_i \alpha_i = 1, \alpha_i \geq 0, \quad (3)$$

where

$$A = \begin{bmatrix} \frac{\sum f_{1j}^2}{n_1} + \lambda & 0 & 0 \\ 0 & \frac{\sum f_{2j}^2}{n_2} + \lambda & 0 \\ 0 & 0 & \frac{\sum f_{3j}^2}{n_3} + \lambda \end{bmatrix}, \quad f = y - A\beta = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}.$$

The newly introduced variables f_1 , f_2 and f_3 are vectors related to three kinds of features, with length of n_1 , n_2 and n_3 respectively. Equation (3) is also a convex problem and can be solved efficiently. The proposed heuristic algorithm iterates the above two steps until either one of the following two conditions are met: the change of α is below a threshold or iterations exceed the predefined number.

With α and β acquired, the reconstruction error using each class can be estimated as $e_F = |W_{\alpha}(y - A\beta_F)|^2$ and $e_B = |W_{\alpha}(y - A\beta_B)|^2$, where F and B refers to the foreground and background respectively. β_F and β_B refers to the weights for the foreground and background atlas nodes respectively. With the estimated reconstruction error, FSLP is encoded as edge weight on the graph during label fusion, which will be explained in next subsection.

2.2 Label Fusion with Random Walker

Besides the FSLP gathered from atlases, the anatomical knowledge from target images is also encoded in the proposed label fusion method. As mentioned in the Introduction section, to label pixels which locate deep inside or outside a sub-cortical structure is relatively easier as compared with those around structural boundary. Thanks to the location advantage, even with a rough initial label map generated by affine transformation, the labels of these pixels (far away from the object boundary) can be treated as confident results. This kind of confidence can be propagated to less confident pixels (near boundary) through image lattice, which is regarded as anatomical prior in our method. In this paper, label fusion

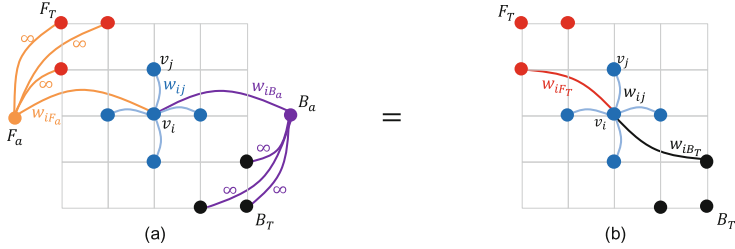


Fig. 3. Graph construction for label fusion. **(a)** Orange and Purple nodes: atlas seeds. Red and Black nodes: target foreground and background seeds. v_i is one candidate node and v_j is one of its neighbours, with w_{ij} as edge weight. FSLP is encoded to w_{iF_a} and w_{iB_a} . **(b)** An equal graph of **a**.

is formulated on an undirected graph $G = (V, E)$, where V refers to a set of nodes consisting of foreground seeds V_F , background seeds V_B and candidate nodes V_C . As both label and anatomical priors are employed in the proposed framework, two kinds of foreground seeds are included in V_F : V_{F_a} from atlases and V_{F_T} from the target image, similarly for V_B . As for V_C , it represents the set of nodes whose labels need to be determined during label fusion and these candidate nodes are selected from the target image. $E \subseteq V \times V$ is the set of edges e_{ij} connecting nodes v_i and v_j , with w_{ij} as edge weight.

Since the number and location of nodes are critical to the efficiency of segmentation algorithms, the strategy for node selection needs to be deployed carefully. As the prior from atlases has been encoded to FSLP, V_{F_a} and V_{B_a} can be represented with two terminal nodes. For the node selection in the target image, affine transformation is first utilized to generate the initial label map and then the distance between each pixel and the structural boundary can be estimated. Using pre-defined distance threshold d_T , pixels with distances larger than d_T are selected as seeds (V_{F_T} and V_{B_T}) and those with smaller values are treated as candidate nodes (V_C).

With seeds and candidate nodes settled, the graph for label fusion can be constructed with edge connections, as shown in Fig. 3(a). The Orange and Purple nodes represent the atlas seeds V_{F_a} and V_{B_a} . Red and Black nodes refer to the foreground V_{F_T} and background V_{B_T} seeds selected from the target image. The influences of target seeds can be propagated to candidate nodes through image lattice, with affinity defined using classical Gaussian function as follows:

$$\forall v_j \in \mathcal{N}(v_i), \quad w_{ij} = \exp(-\delta(I_T(v_i) - I_T(v_j))^2), \quad (4)$$

where v_i is one candidate node, $\mathcal{N}(v_i)$ refers to its 6-nearest neighbours in 3D image, $I_T(\cdot)$ is the pixel intensity value in the target image and δ is one tuning parameter. As for the atlas prior collected with FSLP, it is encoded as the edge weight between v_i and atlas seeds, with the following definition: $w_{iF_a} = e_B/(e_F + e_B)$ and $w_{iB_a} = e_F/(e_F + e_B)$.

Given that V_{F_a} and V_{F_T} are all foreground seeds, the edge weights between them are supposed to be infinity. In this case, setting up an edge between v_i and V_{F_a} is equal to appending an edge for v_i with any target foreground seed,

as illustrated in Fig. 3. In other words, the function of the atlas seeds can be replaced and FSLP can be assigned to the edges of w_{iF_T} and w_{iB_T} instead. In this way, the graph for label fusion is constructed only with target nodes and the graph complexity can be greatly reduced.

With the concise graph constructed, the label fusion process is modeled under the framework of Random Walker and the energy function is as follows [8]:

$$\begin{aligned} \min_x \quad & \sum_{v_i} [w_{iF_T}^2 (x_i - 1)^2 + w_{iB_T}^2 x_i^2] + \sum_{e_{ij}} w_{ij}^2 (x_i - x_j)^2, \\ \text{s.t.} \quad & x_{F_T} = 1, x_{B_T} = 0. \end{aligned} \quad (5)$$

x_i stands for the probability that node v_i belongs to the foreground, with the constraint $x_{F_T} = 1$ and $x_{B_T} = 0$. The first unary term considers the data fidelity of each node independently and the second binary term takes the impact between connected nodes into account. This equation can be viewed as a discrete Dirichlet problem and solved using the Laplace equation with Dirichlet conditions through Graph Analysis Toolbox [3]. With optimal solution obtained for x , the label of each node can be updated: $L(v_i) = 1$ if $x_i \geq 0.5$ and $L(v_i) = 0$ otherwise. The target label map can be updated gradually through iterations until stable.

3 Experiments

To evaluate the performance of the proposed method, experiments have been carried out on two publicly available MR brain image databases – IBSR¹ and LPBA40². The IBSR v2.0 database consists of 18 T1-weighted images with 84 manually labeled structures and the LPBA40 database includes 40 images with 56 manually labeled structures. In the experiments, each database was separated into two parts randomly, with equal number of images as training (atlas) and testing (target) data sets. Considering the intensity inconsistency, histogram matching was first conducted with the Insight Toolkit³. Then efficient pair-wise affine transformations between each target image and all atlases were estimated with FLIRT [4] provided by FSL toolbox [5]. With multiple label maps propagated by various atlases, majority voting was applied to generate the initial label map and the results were also employed as the base line during comparison.

Dice Coefficient (DC) is utilized to evaluate the quality of label fusion. The input patch size for intensity and gradient features was set to $5 \times 5 \times 5$. The size of input patch to Convolutional Neural Networks was 20×20 and the length of extracted structural signature was 18. In Eq. (3) during FSLP estimation, rather than choosing a fixed λ value, it was set to be adaptive $\lambda = \frac{1}{3} \left(\frac{\sum f_{1j}^2}{n_1} + \frac{\sum f_{2j}^2}{n_2} + \frac{\sum f_{3j}^2}{n_3} \right)$ in each iteration. The settings of the rest parameters are listed as follows: signed distance threshold $d_T = 2$ for node selection, and

¹ <http://www.nitrc.org/projects/ibsr>.

² http://www.loni.usc.edu/atlas/Atlas_Detail.php?atlas_id=12.

³ <http://www.itk.org/>.

Table 1. Experimental results on IBSR and LPBA40 database.

Database	IBSR				LPBA40			
	MV	PBL	PBAF	Our Method	MV	PBL	PBAF	Our Method
Thalamus	0.809-0.812	0.877-0.881	0.891-0.892	0.897-0.900	-	-	-	-
Caudate	0.590-0.641	0.806-0.830	0.835- 0.854	0.838-0.847	0.743-0.750	0.851-0.852	0.853-0.851	0.858-0.856
Putamen	0.772-0.762	0.814-0.832	0.834-0.860	0.886-0.893	0.787-0.797	0.841-0.844	0.866-0.863	0.866-0.867
Pallidum	0.679-0.700	0.730-0.746	0.734-0.747	0.801-0.797	-	-	-	-
Hippocampus	0.568-0.532	0.736-0.726	0.774-0.758	0.815-0.810	0.749-0.759	0.839-0.832	0.835-0.831	0.850-0.846
Amygdala	0.538-0.494	0.656-0.612	0.677-0.654	0.712-0.704	-	-	-	-
Average	0.658±0.114	0.770±0.084	0.792±0.080	0.825±0.067	0.764±0.022	0.843±0.008	0.850±0.014	0.857±0.009

the tuning parameter used in Eq. (4) was set to $\delta = 5$. As the iterative strategy is exploited to update target label map gradually in the proposed framework, to test the iterative effects, preliminary experiments have been conducted on LPBA40 database with available subcortical structures. The experimental results demonstrate the segmentation accuracy increases with the number of iterations and tends to remain stable after three iterations. Based on the test on LPBA40, for the experiments on IBSR, the iteration was set to 3.

Besides majority voting (MV), the comparison with the conventional patch-based label fusion (PBL) [9] has been made for evaluation. Considering multiple features employed in the proposed method, it was also compared with the state-of-the-art method – patch-based label fusion with augmented features (PBAF) [1]. In addition to intensity information, the spatial and context features are also appended for label fusion in PBAF. To keep consistency, the patch size for PBL and PBAF was set to $5 \times 5 \times 5$ and the size of search volume was set to $9 \times 9 \times 9$. For PBL, the key parameter is the Gaussian kernel value and it was tested from $\{1, 10, 100, 1000, 10000, 100000\}$ on two databases. Results indicate that 10 and 10000 gives the best performance on IBSR and LPBA40 respectively. For PBAF, the parameter setting of pre-selected atlas nodes amount was tested from $\{2, 16, 32, 64, 128, 256, 512\}$ and 128 produces the peak of accuracy.

The segmentation results on two databases generated with our method and compared methods are listed in Table 1, with highest DC values written in Red. There are six subcortical structures delineated in the IBSR database and each structure is separated into two parts, located in the left and right hemispheres respectively. As for LPBA40, the segmentation quality with available subcortical structures is also reported in this Table. When compared with the base line MV, our approach can create the considerable increase of 16.7% and 9.3% respectively on two databases. The proposed method can still outperform the preminent label fusion method PBAF by 3.3% and 0.7%.

4 Conclusion

In this paper, a novel framework for atlas-based image segmentation is proposed, which encodes both the label priors from atlases and anatomical prior from the

target image. Three kinds of features are employed to obtain a more discriminative pixel representation and the feature sensitivity during label fusion is taken into consideration. Besides the FSLP from atlases, the anatomical prior from the target itself is also employed for final label estimation. The label fusion process is formulated on a graph with Random Walker, with priors encoded as edge weights. Although atlas seeds involved in graph construction, an equal but simpler graph can be inferred which just relies on nodes from target image. The proposed framework has been compared with other state-of-the-art methods for comprehensive evaluation and experimental results indicate that it can obtain better label fusion quality consistently on two publicly available databases.

Acknowledgement. We would like to acknowledge the financial support by the Hong Kong Research Grants Council under Grant 16203115.

References

1. Bai, W., Shi, W., Ledig, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac mr images. *Med. Image Anal.* **19**(1), 98–109 (2015)
2. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *NeuroImage* **54**(2), 940–954 (2011)
3. Grady, L., Schwartz, E.: The graph analysis toolbox: image processing on arbitrary graphs. CAS/CNS Technical report Series no. 021 (2003)
4. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**(2), 825–841 (2002)
5. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. *NeuroImage* **62**(2), 782–790 (2012)
6. Koch, L.M., Rajchl, M., Tong, T., Passerat-Palmbach, J., Aljabar, P., Rueckert, D.: Multi-atlas segmentation as a graph labelling problem: application to partially annotated atlas data. In: Ourselin, S., Alexander, D.C., Westin, C.-F., Cardoso, M.J. (eds.) *IPMI 2015. LNCS*, vol. 9123, pp. 221–232. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-19992-4_17](https://doi.org/10.1007/978-3-319-19992-4_17)
7. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
8. Lézoray, O., Grady, L.: *Image Processing and Analysis with Graphs: Theory and Practice*. CRC Press, Boca Raton (2012)
9. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. *IEEE TMI* **30**(10), 1852–1862 (2011)
10. Ta, V.-T., Giraud, R., Collins, D.L., Coupé, P.: Optimized patchmatch for near real time and accurate label fusion. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014. LNCS*, vol. 8675, pp. 105–112. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-10443-0_14](https://doi.org/10.1007/978-3-319-10443-0_14)
11. Tong, T., Wolz, R., Coupé, P., Hajnal, J.V., Rueckert, D., Initiative, A.D.N., et al.: Segmentation of mr images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *NeuroImage* **76**, 11–23 (2013)