



Robust multi-atlas label propagation by deep sparse representation



Chen Zu^{a,b}, Zhengxia Wang^{a,c}, Daoqiang Zhang^b, Peipeng Liang^d, Yonghong Shi^{e,f},
Dinggang Shen^{a,g,*}, Guorong Wu^{a,*}

^a Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^b College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

^c Department of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China

^d Department of Radiology, Xuanwu Hospital, Capital Medical University, Beijing 100053, China

^e Digital Medical Research Center, School of Basic Medical Science, Fudan University, Shanghai 200032, China

^f Shanghai Key Laboratory of Medical Image Computing and Computer-Assisted Intervention, Shanghai 200032, China

^g Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

ARTICLE INFO

Article history:

Received 26 January 2016

Received in revised form

26 August 2016

Accepted 21 September 2016

Available online 21 September 2016

Keywords:

Hierarchical sparse representation

Multi-atlas segmentation

Patch-based label fusion

ABSTRACT

Recently, multi-atlas patch-based label fusion has achieved many successes in medical imaging area. The basic assumption in the current state-of-the-art approaches is that the image patch at the target image point can be represented by a patch dictionary consisting of atlas patches from registered atlas images. Therefore, the label at the target image point can be determined by fusing labels of atlas image patches with similar anatomical structures. However, such assumption on image patch representation does not always hold in label fusion since (1) the image content within the patch may be corrupted due to noise and artifact; and (2) the distribution of morphometric patterns among atlas patches might be unbalanced such that the majority patterns can dominate label fusion result over other minority patterns. The violation of the above basic assumptions could significantly undermine the label fusion accuracy. To overcome these issues, we first consider forming label-specific group for the atlas patches with the same label. Then, we alter the conventional *flat and shallow* dictionary to a deep multi-layer structure, where the top layer (*label-specific dictionaries*) consists of groups of representative atlas patches and the subsequent layers (*residual dictionaries*) hierarchically encode the patchwise residual information in different scales. Thus, the label fusion follows the representation consensus across representative dictionaries. However, the representation of target patch in each group is iteratively optimized by using the representative atlas patches in each label-specific dictionary exclusively to match the principal patterns and also using all residual patterns across groups collaboratively to overcome the issue that some groups might be absent of certain variation patterns presented in the target image patch. Promising segmentation results have been achieved in labeling hippocampus on ADNI dataset, as well as basal ganglia and brainstem structures, compared to other counterpart label fusion methods.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Automatically labeling regions of interest (ROIs) is the key step in many imaging-based studies [1–3]. Manual annotation of anatomical structures is tedious and very time-consuming, which makes it impractical in most of the current medical studies with generally a large amount of imaging data. Therefore, high-throughput and automated segmentation methods are highly desired.

Since using multiple atlases is more capable of accommodating high structural variability than using single atlas, multi-atlas segmentation has emerged as a popular automated segmentation

technique, by propagating the labels from annotated atlas images to the target image. In general, multi-atlas segmentation includes two steps, i.e., (1) Registration step, for aligning the selected atlases as well as their corresponding label images to the target image space [4–6], and (2) label fusion step, for fusing the registered label maps of the selected atlases into a consensus segmentation of the target image [1,7–14].

Among various proposed label fusion strategies, either voxel-wise [1,15] or patch-wise [9,16], they share a common assumption that similar anatomical structures should bear the same anatomical label. For example, the non-local label fusion method [9] computes patch-wise similarity between the target image patch and all possible atlas patches in a search neighbor. Intuitively, high similarity leads to large weight in label fusion. Eventually, the label with the largest weight wins for tagging the target image point. To

* Corresponding authors.

E-mail addresses: dgshen@med.unc.edu (D. Shen), grwu@med.unc.edu (G. Wu).

reduce the risk of introducing ambiguous atlas patches, sparsity constraint is recently used to enforce selection of only a very small number of atlas patches for label fusion [17,18].

It is apparent that each atlas image patch is independently represented by a set of registered atlas patches in label fusion. Although pairwise correlation between any two atlas image patches is explored in [19,20], the entire atlas patches are treated as a whole and compete each other to represent the target image patch. Consequently, there are two limitations in the current patch-based label fusion methods. *First*, image patch could be complex and might include different-scale morphometric patterns and noise. Thus, representing the whole image patches with mixed information inside is challenging. *Second*, the distribution of atlas image patches is highly complex. For certain groups of atlas image patches tagged with the same label, they might lack of diversity to accurately represent the new instance. In other words, although the related pattern of underlying structure within the image patch might be well matched, the mismatches of unrelated structures can misguide the label fusion procedure.

To solve these issues, our solution is to break down the morphometric patterns in atlas image patches into two levels: label-specific and residual patterns, which are organized into a tree-like dictionary. After that, we propose a hierarchical sparse label fusion method by efficiently representing the target image patch *layer by layer* in a *competition-collaboration* manner. Specifically, we consider that atlas image patches with the same anatomical label form a label-specific group. In the top layer, we construct a *label-specific dictionary* for each label by using the representative image patches (i.e., cluster centers) in each group. From the second layer, we continue selecting representative image patches from the remaining patches for each group. Then, we build one *residual dictionary* for each layer by combining information from all groups, where each atom is the residual pattern between the representative image patch in the current layer and its parent patch in the previous layer.

In label fusion, we represent the principal part of the target image patch with each *label-specific dictionary* separately (via sparse constraint), with the remaining part collaboratively represented by the *residual dictionaries*. In the end, the group with the least representation error wins for tagging the target image point. In this way, only the *label-specific dictionaries* compete against others to represent the target image patch and vote for the label. Atoms in the *residual dictionaries*, regardless of being with the same label or not, collaborate to express the residual between the original target image patch and the weighted average of atoms in a certain *label-specific dictionary*, in order to overcome the issue that some groups might be absent of certain variation patterns presented in the target image patch. It is worth noting that the knowledge of common variations is allowed to transfer from one group to other groups, i.e., the variation patterns in the *residual dictionary* of each layer is shared across different anatomical groups. Importantly, since we alter the conventional flat and shallow dictionary into a deep tree-like structure, we eventually break down to solve a set of small-scale patch representation problems in each layer. In light of this, our method opts for solving the large-scale sparse patch representation problem in labeling each image point.

We have evaluated the performance of segmenting hippocampus from ADNI dataset and both basal ganglia and brainstem structures from other MR images. In all experiments, our hierarchical label fusion method achieves more accurate segmentation results than the conventional non-local [9,16] and sparse label fusion methods [18].

The remainder of the paper is organized as follows. In Section 2, we present our novel label fusion by deep sparse representation. In Section 3, we evaluate its performance by comparing with

conventional patch-based methods, and we provide a brief conclusion in Section 4.

2. Method

The goal of multi-atlas label fusion is to propagate the labels from a set of registered atlas images to the target image T . Suppose we have N registered atlas images I_s ($s = 1, \dots, N$), along with their respective label maps L_s ($s = 1, \dots, N$). The conventional label fusion approaches estimate the target label f at each voxel $x \in \Omega$ of target image T in a patch-wise manner. Denote $\bar{\alpha}$ as a P -dimensional (column) vector containing the intensity values in the target image patch centered at voxel x , and matrix $\mathbf{B} = [\beta_1, \dots, \beta_i, \dots, \beta_M]$ as a dictionary of M candidate atlas image patches (arranged into column vectors), which consists of all possible atlas patches in a search neighborhood of x . Following the same column order as the matrix \mathbf{B} , $\bar{l} = [l_1, \dots, l_i, \dots, l_M]^T$ is a (column) vector of labels at the atlas patch centers, with each element $l_i \in \{-1, 1\}$ indicating either the absence or the presence of a given structure at the center of the respective atlas patch β_i . For clarity, we only focus on single structure in this paper; however, it is straightforward to extend to multiple structures.

Next, the label fusion procedure can be regarded as a patch representation problem, seeking for a linear combination of atlas patches $\mathbf{B}\bar{w}$ which can best fit the target image patch $\bar{\alpha}$. Here, \bar{w} is a M -dimensional (column) vector, where each element indicates the influence in voting for the latent label f . Sparsity constraint upon the weighing vector \bar{w} is proven a useful way to improve the label fusion accuracy [17,18,20] since encouraging more zero elements in the weighting vector can eventually reduce the risk of introducing misleading atlas patches. Thus, the objective function at each target image point x can be defined by:

$$\hat{\bar{w}} = \arg \min_{\bar{w}} \|\bar{\alpha} - \mathbf{B}\bar{w}\|_2^2 + \lambda \cdot \|\bar{w}\|, \quad (1)$$

where λ is the scalar controlling the sparsity degree. The intuition of encouraging sparsity on \bar{w} is to suppress the spurious atlas patches in the dictionary \mathbf{B} by using only a small number of good atlas patches, instead of all of them. Given the weighing vector \bar{w} , the label f at each target image point x can be determined by:

$$f = \begin{cases} 1, & \& \bar{w} \cdot \bar{l} > 0 \\ -1, & \& \bar{w} \cdot \bar{l} < 0 \end{cases} \quad (2)$$

where ' \cdot ' denotes the inner product of two vectors.

2.1. Limitation of Conventional Flat Dictionary

It is apparent that all atlas patches in the dictionary \mathbf{B} are stacked, column by column, in a single layer. Thus, solving the sparse representation problem in Eq. 1 becomes very difficult if the number of atoms in \mathbf{B} is beyond the affordable scale. Another critical issue is that each atom independently competes against other thousands of atoms. Due to possible large variations between target and atlas images and also image noise and artifacts within each patch, it is often too strict to enforce the exact and entire patch matching between two image patches.

As the toy example shown in Fig. 1(a), the target image patch is a blue box face with two ears on the top. Conventional dictionary only has one layer, consisting of two kinds of faces: box faces (#1-#4 in green) with ear(s) on either left or right side, and round faces (#5-#15 in red) with ear(s) on the top. It is clear that the shape of face is the primary pattern and specific to the label (face type) in this toy example. The variations about ears are just the external patterns, which are not related to the task of recognizing faces.

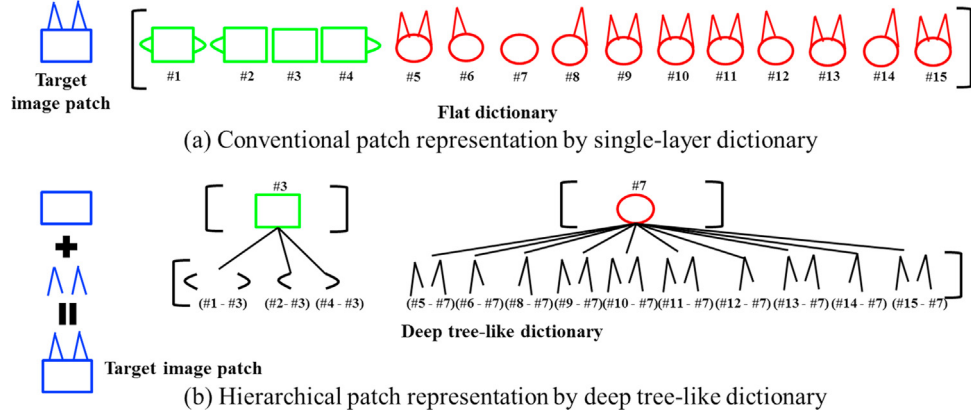


Fig. 1. The toy example of (a) conventional patch representation by single-layer dictionary and (b) our hierarchical patch representation by deep tree-like dictionary.

Unfortunately, none of the sample in the group of box faces has ear (s) on the top, while such variations of ear are abundant in the group of round faces. In the conventional patch representation scenario, it is highly possible to label the target face with round face since too many non-primary variations (the pattern of ears) presented in the dictionary may mislead the representation procedure unless being treated aside from the primary label-specific patterns.

2.2. Construction of Deep Tree-like Dictionary

In light of this, we propose to alter the flat dictionary into a deep tree-like structure. Here, we consider forming atlas patches with the same label into a group. Suppose there are R kinds of labels, denoted by $\xi_1, \dots, \xi_r, \dots, \xi_R$. Thus we can divide the atlas patches $\beta_1, \dots, \beta_i, \dots, \beta_M$ into R groups, where each group $G_r = \{\beta_i \mid i = \xi_r, i = 1, \dots, M\}$ ($r = 1, \dots, R$) only keep the atlas patches with label ξ_r . In the following, we apply the hierarchical k-means [21] to each group G_r to divide the atlas patches into several layers, where the variation patterns within the group is hierarchically encoded via the tree from majority to minority. In the beginning, we define the tree has H layers and the branching factors of the tree $\{b_1, \dots, b_h, \dots, b_{H-1}\}$ (with the last layer consisting of leaf nodes but no children), where b_h specifies the number of children of each node at each layer. Then we start hierarchical k-means from the whole group G_r . Specifically, we cluster all atlas patches of group G_r into k_r divisions based on patch appearance, where each division consists of the atlas image patches similar to a particular cluster center. Furthermore, we choose the atlas image patch closest to the cluster center to represent each cluster of group G_r . The same procedure is then recursively applied to each cluster, splitting to $b_1 \times b_2$ nodes in the second layer of tree. We repeat the above clustering procedure until $|G_r| - k_r \cdot \prod_{h=1}^{H-2} b_h$ leaf nodes are left in the bottom layer, where $|G_r|$ is the total number of image patches in the group G_r .

After we obtain the patch tree in each group, the next step is to construct the deep tree-like dictionary layer by layer. For each group G_r , we stack top-level representative image patches, column by column, and build the *label-specific dictionary* D_r in the first layer. From the second layer, we construct one *residual dictionary* E_h ($2 \leq h \leq H$) for each layer in two steps: (1) for each group G_r , we compute the voxel-wise difference between each node (image patch in the current layer) with its parent node (cluster center in the previous layer); and (2) we stack the residual patches from all groups to build the *residual-dictionary* E_h ($2 \leq h \leq H$). The residual dictionaries $\{E_h\}$ are equally shared by each *label-specific dictionary* D_r . After that, we can obtain tree-like dictionary TD_r for each label as $TD_r = \{D_r, E_2, \dots, E_H\}$. It is worth noting that our deep dictionary

allows both completion and collaboration. The label-specific patterns are not shared across different label-specific dictionary D_r in the top layer. On the contrary, they compete each other to tag the underlying target image patch. From the second layer, the variation patterns are shared for all D_r s since we consider the remaining information (after excluding the principle pattern) is not specific to any label.

It is worth noting that each atom in the *label-specific dictionary* D_r represents one of the exemplar atlas patch in group G_r . However, all tree-like dictionaries TD_r share the same residual dictionaries from E_2 to E_H , where each node only encodes the residual information w.r.t. the cluster center. The information in $\{E_2, \dots, E_H\}$ conveys the variations in different scales (from major to minor variations) as layer h increases. The toy example in Fig. 1(b) shows 2 layers. The two exemplars, one from box face (in green) and another one from round face (in red), form the *label-specific dictionary* D_1 and D_2 , respectively. The *residual dictionary* E_2 conveys various patterns of ears and is shared by D_1 and D_2 . Apparently, box face with ear(s) on the top is still box face, and vice versa. Since these variation patterns are not specific to labels (not related to box face or round face), this knowledge can be shared across different groups to represent the target image patch, as detailed next.

2.3. Hierarchical sparse patch label fusion

Since we alter the conventional flat dictionary B into a set of deep tree-like dictionary $\{TD_1, \dots, TD_R\}$, we develop the hierarchical sparse patch representation algorithm to represent target image patch $\bar{\alpha}$ via certain dictionary TD_r :

$$\left\{ \hat{\mu}_r, \hat{w}_2^r, \dots, \hat{w}_h^r, \dots, \hat{w}_H^r \right\} = \arg \min_{\{\bar{\mu}_r, \bar{w}_2^r, \dots, \bar{w}_H^r\}} \left\| (\bar{\alpha} - D_r \bar{\mu}_r) - \sum_{h=2}^H E_h \bar{w}_h^r \right\|_2^2 + \lambda_1 \|\bar{\mu}_r\|_1 + \sum_{h=2}^H \lambda_h \|\bar{w}_h^r\|_1, \quad (3)$$

where $\bar{\mu}_r$ and \bar{w}_h^r are the weighting vectors for *label-specific dictionary* D_r and E_h , respectively. The intuition behind Eq. (3) is that we first only use the *label-specific dictionary* D_r to represent the principal part of target image patch $\bar{\alpha}$. The residual part $(\bar{\alpha} - D_r \bar{\mu}_r)$ is then recursively presented by a set of *residual dictionaries* E_h ($h = 2, \dots, H$), which utilizes the variations of the observed residual patterns from all other groups.

After achieving the optimal weighting vectors $\hat{\mu}_r$ and $\{\hat{w}_2^r, \dots, \hat{w}_H^r\}$, we can compute the overall representation error ε_r regarding TD_r by $\varepsilon_r = \bar{\alpha} - D_r \hat{\mu}_r - \sum_{h=2}^H E_h \hat{w}_h^r$. The group with minimal representation error ε_r wins for tagging the target image point x

with label $f=\xi_r$. The principle behind such hierarchical sparse patch label fusion is that we only allow the *label-specific dictionaries* \mathbf{D}_r , to compete against each other to vote for the target image point x . Other variation patterns in the residual dictionaries collaborate, by sharing the variation patterns across groups, to fit the remaining part $(\bar{\alpha} - \mathbf{D}_r\bar{\mu}_r)$. As shown in Fig. 1(b), after well fitting the variations of ears by \mathbf{E}_2 , we can accurately find the target face belongs to box face although the pattern of ear variations do not exist in the box face group. But the conventional sparse representation with flat dictionary \mathbf{B} fails since it enforces the very strict whole-part matching such that the discrepancies of ears may mis-label the target face as round face.

To solve the minimization problem in Eq. (3), we resort to the Augmented Lagrange Multiplier (ALM) scheme [22,23] by converting Eq. (3) to minimize the following function:

$$\begin{aligned} & \frac{\theta}{2} \left\| \bar{\alpha} - \mathbf{D}_r\bar{\mu}_r - \sum_{h=2}^H \mathbf{E}_h\bar{w}_h^r \right\|_2^2 + \lambda_1 \|\bar{\mu}_r\|_1 + \sum_{h=2}^H \lambda_h \|\bar{w}_h^r\|_1 \\ & + \phi \left(\bar{\alpha} - \mathbf{D}_r\bar{\mu}_r - \sum_{h=2}^H \mathbf{E}_h\bar{w}_h^r \right), \end{aligned} \quad (4)$$

where ϕ is a vector of Lagrange multipliers and θ is a penalty parameter. Then, we iteratively optimize $\{\bar{\mu}_r, \bar{w}_2^r, \dots, \bar{w}_H^r\}$ by the algorithm summarized below.

Algorithm for solving Eq.4 by ALM _

Input: \mathbf{D}_r , $\bar{\alpha}$, and parameters λ_h ($h = 1, \dots, H$).

Initialization: $\bar{\mu}_r=0$, $\bar{w}_h^r=0$ ($h = 2, \dots, H$), $\xi = 1.0$, $\theta_{max} = 10^4$, $\eta=10^{-4}$, and $\sigma = 1.5$.

While not converged **do**

1. Fix others and update $\bar{\mu}_r$ by:

$$\bar{\mu}_r = \underset{\bar{\mu}_r}{\operatorname{argmin}} \left\| \bar{\alpha} - \sum_{h=2}^H \mathbf{E}_h\bar{w}_h^r \right\|_2^2 + \lambda_1 \|\bar{\mu}_r\|_1$$

2. $h=2$;

3. while $h \leq H$ do

Fix others and update \bar{w}_h^r by:

$$\bar{w}_h^r = \underset{\bar{w}_h^r}{\operatorname{argmin}} \left\| \bar{\alpha} - \mathbf{D}_r\bar{\mu}_r - \sum_{h'=2, h' \neq h}^H \mathbf{E}_{h'}\bar{w}_{h'}^r \right\|_2^2 + \lambda_h \|\bar{w}_h^r\|_1$$

$h = h+1$;

end;

4. Update the multipliers by: $\phi = \phi + \theta(\bar{\alpha} - \mathbf{D}_r\bar{\mu}_r - \sum_{h=2}^H \mathbf{E}_h\bar{w}_h^r)$

5. Update θ by $\theta = \min(\theta_{max}, \sigma \cdot \theta)$

6. Check the converge condition: $\left\| \bar{\alpha} - \mathbf{D}_r\bar{\mu}_r - \sum_{h=2}^H \mathbf{E}_h\bar{w}_h^r \right\|_2^2 < \eta$

End

Output: $\{\bar{\mu}_r, \bar{w}_2^r, \dots, \bar{w}_H^r\}$

2.4. Discussion

Tree-Guided Group Lasso [24,25] (tree-lasso) also organizes the dictionary into the tree-like structure, in order to reflect the correlations among the dictionary atoms. However, our hierarchical sparse representation method has several significant differences from the tree-lasso: (1) Our method uses the residual image patches, instead of the original ones after the first layer, while every node in tree-lasso keeps the original information. (2) In tree-lasso, the children nodes only share the information with their own parent node. However, our method allows the knowledge of variation patterns to transfer across different groups, i.e., each atom in the *label-specific dictionary* can borrow the variation patterns

derived from other groups to avoid the dilemma of mis-representation due to the lack of variation patterns presented in the target image patch. (3) In our method, we adaptively construct the tree by hierarchical k-means. In tree-lasso, the tree is usually manually constructed based on certain priori knowledge.

Notice that we sequentially solve weighting vectors $\bar{\mu}_r$ and $\{\bar{w}_2^r, \dots, \bar{w}_H^r\}$ from the roots to the leaf of the tree. Since we use sparsity constraint in each layer, it is highly possible that a large number of weights in each weighting vector are zero. For those atoms with zero weights, we do not include their children in the next layer. In this way, we dynamically form the new dictionary in transition to the next layer and keep solving the very-small-scale sparse representation problem in each layer. Therefore, the computational cost is comparable to the conventional sparse representation methods that use flat and large dictionary.

3. Experiments

In the following experiments, we compare our label fusion with the non-local [9,16] and sparse label fusion methods [18]. To label the target image, we first use FLIRT in FSL package to linearly register all atlas images onto the target image and then use diffeomorphic Demons [26] to compute the remaining local deformations. The main parameters for running diffeomorphic Demons are: 15, 10, and 5 iterations in low, middle, and high resolutions, respectively. The smoothing kernel size is 2.0. The patch size is $9 \times 9 \times 9$ for all the patch-based label fusion method. To assess label accuracy, the Dice ratio is used to measure the degree of overlap between two ROIs O_1 and O_2 as follows:

$$\text{Dice}(O_1, O_2) = 2 \times \frac{|O_1 \cap O_2|}{|O_1| + |O_2|}, \quad (5)$$

where $| \cdot |$ means the volume of the particular ROI.

3.1. Image Preprocessing and Parameter Setting

After register all atlases the to-be-segmented target image, histogram matching is performed on each registered atlas image,

Table 1

The statistics of Dice ratios, average surface distance, and computation costs in hippocampus labeling by non-local, sparse, and our patch-based methods, with affine registration. (* indicate that our proposed label fusion method achieves significant improvement over the underlying counterpart method under paired t -test with $p < 0.05$).

	Non-Local	Sparse	Our method
Dice Ratio	85.8 ± 4.1 *	86.1 ± 3.3 *	87.9 ± 2.9
Average surface distance	(0.59 ± 0.13)mm	(0.45 ± 0.09)mm	(0.40 ± 0.08)mm
Time	153 s	248 s	518 s

Table 2

The statistics of Dice ratios, average surface distance, and computation costs in hippocampus labeling by non-local, sparse, and our patch-based methods, with deformable registration. (* indicate that our proposed label fusion method achieves significant improvement over the underlying counterpart method under paired t -test with $p < 0.05$).

	Non-Local	Sparse	Our method
Dice Ratio	86.6 ± 3.5 *	87.3 ± 3.4	88.3 ± 2.6
Average surface distance	(0.43 ± 0.09)mm	(0.39 ± 0.08)mm	(0.35 ± 0.06)mm
Time	75 s	128 s	465 s

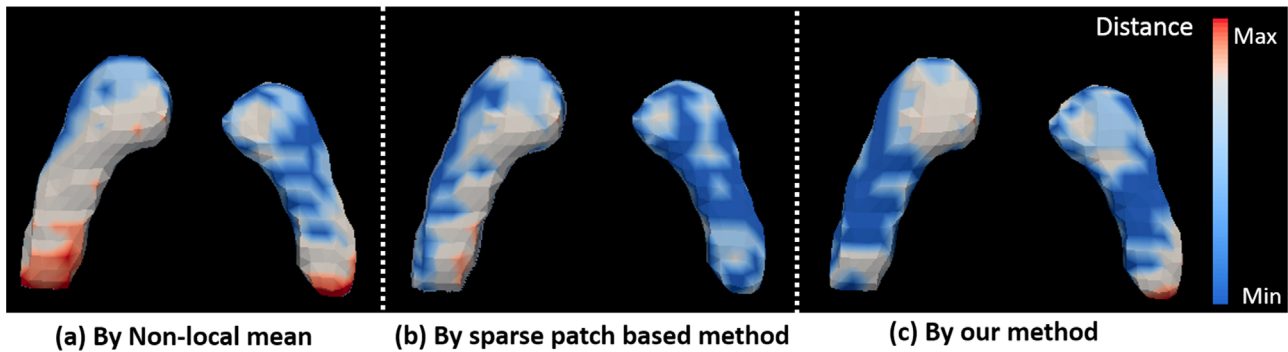


Fig. 2. The surface distance in labeling one individual subject image using Non-local mean (a), sparse (b), and our proposed label fusion methods (c). Blue and red color denote for low and large surface distance respectively.

in order to normalize the intensity range. The label fusion is independently deployed on each target image voxel. Since we have specifically evaluated the influence of patch size and search neighborhood in our previous work [20,27], we fix the patch size to $5 \times 5 \times 5$ voxel in each direction and the search neighborhood is set to $5 \times 5 \times 5 \text{mm}^3$ in linear registration scenario and $5 \times 5 \times 5 \text{mm}^3$ in non-linear registration scenario, which are found optimal in terms of labeling accuracy and computation time. As we will demonstrate in Fig. 3, four layers ($H = 4$) is optimal in balancing label fusion accuracy and computational cost. Thus, we use patch pre-selection procedure [9] to keep at least 1,000 candidate atlas patches. In building the patch tree by hierarchical k-means, we allow each group to use 30 clusters ($k_r=30$) in the top layer. The branching factors are set to $b_1=3$ and $b_2=4$, which means 90 (30×3) nodes in the second layer and 360 (90×4) nodes in the third layer. The sparsity constraint (λ in Eq. 1) for the conventional sparse patch-based label fusion method is set to 0.1. For our method, the sparse constraint in each layer is set to $\lambda_1=0.1$, $\lambda_2=0.5$, $\lambda_3=0.5$, and $\lambda_4=0.5$, respectively. It is worth noting that we will keep using the same parameter setting in the following experiment.

3.2. Experimental Result of Hippocampus Labeling

In this experiment, we randomly select 66 high resolution 3D T1-weighted MR images of elderly brains from ADNI dataset,¹ where the left and right hippocampi have been manually labeled for each brain. Specifically, these MR images were acquired on a 3.0T GE scanner in the sagittal plane using an IR-FSPGR pulse sequence, 8-channel coil, TR = 650 ms, TE = min full, flip-angle = 8° , slice thickness = 1.2 mm, resolution = 256×256 mm, and FOV = 26 cm. We evaluate label fusion performance in a leave-one-out manner. Specifically, we apply the label fusion accuracy by using affine registration and deformable registration separately. Here, we follow patch pre-selection condition in [9] to discard the less-similar image patches. Table 1 shows the mean and standard deviation of Dice ratio and surface distance on hippocampus (left and right combined) by non-local, sparse, and our hierarchical sparse label fusion method, where the atlas images are aligned to the target image by affine registration. Our method achieves the highest Dice ratio and lowest surface distance over other two counterpart methods, where the improvement of Dice ratio is significant. Similarly, as shown in Table 2, our method beat other two in the deformable registration scenario, where we obtain 1.7% and 1.0% improvements of Dice ratio over non-local and sparse patch-based methods, respectively. The typical surface distance maps in labeling one individual subject are shown in Fig. 2, where the red and blue denote for low and large surface distance. It is clear that our proposed

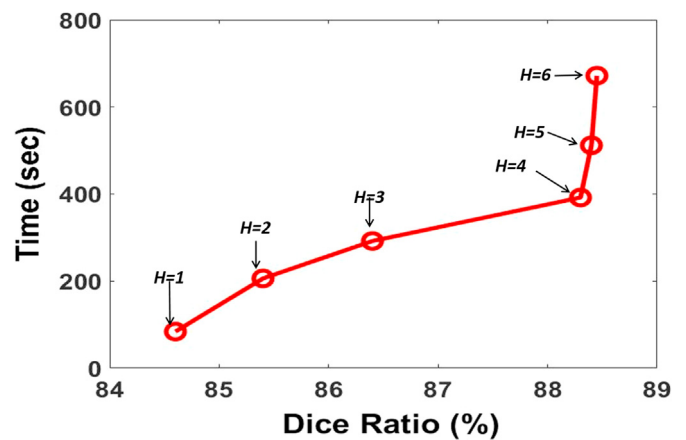


Fig. 3. The Dice ratio vs computation cost by using different number of layers in our hierarchical sparse label fusion method.

label fusion method has smallest mean surface distance (Fig. 2(c)) than non-local (Fig. 2(a)) and sparse patch-based label fusion methods (Fig. 2(b)).

We test all label fusion methods on a workstation with 8 CPU cores (@3.0 GHz) and 16 G memory. Since the label fusion procedure is independent at each target image point, we use OpenMP² to parallel the whole process. The computational cost for 3 label fusion methods by affine and deformable registrations (excluding the registration time) are shown in the bottom rows of Table 1 and Table 2, respectively.

Furthermore, we specifically evaluate the effect of tree layers in our sparse hierarchical label fusion method. Fig. 3 shows the curve of Dice ratio vs computation cost as the layer number H increases from 1 to 6. It is worth noting that only using the top layer ($H = 1$) is the degraded version of our method. Without the support from the *residual dictionaries*, it is not surprising to see that the labeling accuracy is the lowest (84.6%). As H increases, the Dice ratio is improved significantly to 88.3% at $H = 4$, which indicates that the deep tree-like structure is very useful to deal with the complex patch representation problem in label fusion. Apparently, more layers can continue improving the Dice ratio, but the performance increase is marginal at the expense of computational cost. In all experiments, we find that 4 layers are sufficient to encode the label-specific and individual variations.

3.3. Experimental Result on Segmenting Basal Ganglia Structures and Brainstem

The main components in basal ganglia include dorsal striatum

¹ www.adni-info.org

² <http://openmp.org/wp/>

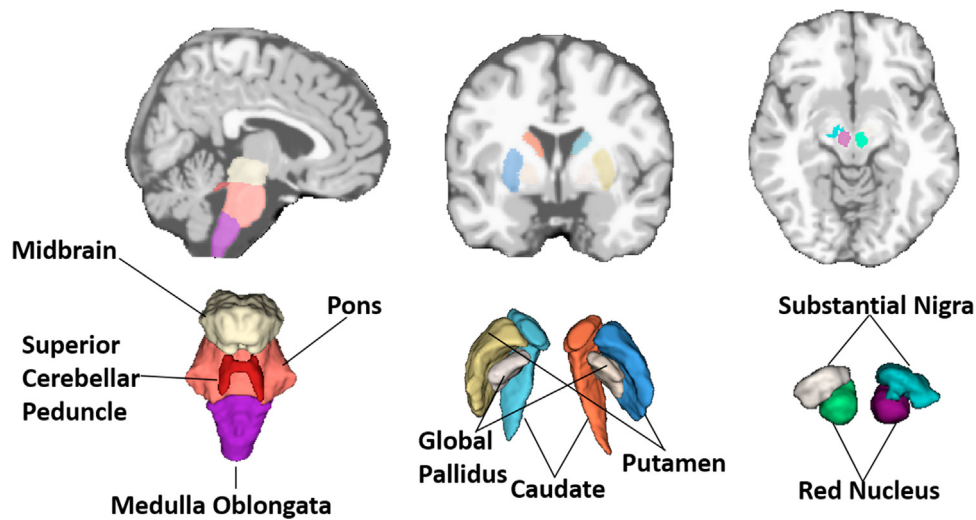


Fig. 4. Nine typical ROIs in basal ganglia and brainstem.

Table 3

The mean and standard deviation of Dice ratios in 9 ROIs by non-local, sparse, and our patch-based methods. (* indicate that our proposed label fusion method achieves significant improvement over the underlying counterpart method under paired t -test with $p < 0.05$).

	Non-Local	Sparse	Our method
Mdbrain	85.2 ± 2.6*	87.5 ± 1.8	87.7 ± 1.7
Pons	86.6 ± 3.8	88.1 ± 2.9	88.5 ± 2.8
Sup. Cere. Peduncle	65.1 ± 5.4*	67.9 ± 4.1	68.4 ± 3.8
Medulla Oblongata	84.5 ± 3.0*	86.1 ± 2.4	86.6 ± 1.6
Caudate	75.2 ± 2.9	76.1 ± 2.4	77.6 ± 1.8
Putamen	76.7 ± 1.8	77.4 ± 1.8	77.6 ± 1.5
Global Pallidus	70.0 ± 2.4	71.5 ± 2.1	71.8 ± 1.9
Substantia Nigra	53.3 ± 6.5*	55.2 ± 5.1*	57.5 ± 4.4
Red Nucleus	53.6 ± 5.9*	55.6 ± 5.5	55.9 ± 4.2

Table 4

The mean and standard deviation of surface distance in 9 ROIs by non-local, sparse, and our patch-based methods. (unit: mm).

	Non-Local	Sparse	Our method
Mdbrain	0.33 ± 0.07	0.31 ± 0.05	0.30 ± 0.06
Pons	0.41 ± 0.09	0.38 ± 0.05	0.35 ± 0.06
Sup. Cere. Peduncle	0.61 ± 0.13	0.59 ± 0.12	0.55 ± 0.10
Medulla Oblongata	0.34 ± 0.05	0.34 ± 0.05	0.34 ± 0.06
Caudate	0.36 ± 0.06	0.33 ± 0.04	0.31 ± 0.03
Putamen	0.36 ± 0.06	0.35 ± 0.06	0.34 ± 0.06
Global Pallidus	0.59 ± 0.09	0.54 ± 0.08	0.51 ± 0.07
Substantia Nigra	0.62 ± 0.11	0.60 ± 0.09	0.58 ± 0.07
Red Nucleus	0.63 ± 0.09	0.61 ± 0.09	0.59 ± 0.08

(caudate and putamen), global pallidus, substantia nigra, and red nucleus. In terms of anatomy, brainstem can be partitioned to midbrain, pons, medulla oblongata and superior cerebellar peduncle. These 9 regions, as shown in Fig. 4, are closely related with the development of Parkinson's disease. In this experiment, 3T T1-weighted MR images from 11 PD patients are used as the atlases, each of them have the above 9 ROIs manually delineated by two radiologists. The total scan time is expected to be in the range from between 20–30 min. The field of view must include the vertex, cerebellum and pons. The image size is $512 \times 512 \times 176$ and the image resolution is $0.5 \times 0.5 \times 1 \text{ mm}^3$. We evaluate the label fusion results in a leave-one-out manner. The overall Dice ratios and averaged surface distance in 9 ROIs are shown in Table 3 and Table 4, respectively. It is clear that our proposed method

has achieved the highest label fusion accuracy in all 9 ROIs.

4. Conclusion and future work

In this paper, we propose a novel hierarchical sparse representation method for multi-atlas patch-based label fusion. The main contribution of our work is that we alter the flat dictionary into a tree-like structure. Specifically, the most representative image patches within each group (label) form the top layer label-specific dictionary and the variation patterns across groups (label) are hierarchically encoded layer after layer. Since the data (candidate image patches) have been hierarchically organized into a major to minor manner, we substantially improve the representation power of sparse representation. In label fusion, we only allow the atoms in the label-specific dictionary delegating the whole group and competing against each other for voting the label. The variation patterns collected from different groups are shared across different groups and collaborate to alleviate the misrepresentation risk due to the lack of certain variation patterns in some groups. We have applied our new label fusion method to hippocampus segmentation and also the parcellation of basal ganglia and brainstem regions. Compared to the counterpart label fusion method, our proposed method has achieved more accurate labeling results.

Acknowledgment

This work is supported in part by National Institutes of Health (NIH) grants HD081467, EB006733, EB008374, EB009634, MH100217, AG041721, AG049371, AG049089, AG042599, CA140413.

References

- [1] X. Artaechevarria, A. Munoz-Barrutia, C. Ortiz-de-Solorzano, Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data, Med. Imaging, IEEE Trans. on 28 (2009) 1266–1277.
- [2] D. Devanand, G. Pradhaban, X. Liu, A. Khandji, S. De Santi, S. Segal, et al., Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease, Neurology 68 (2007) 828–836.
- [3] R.A. Heckemann, J.V. Hajnal, P. Aljabar, D. Rueckert, A. Hammers, Automatic anatomical brain MRI segmentation combining label propagation and decision fusion, NeuroImage 33 (2006) 115–126.
- [4] A. Klein, J. Andersson, B.A. Ardekani, J. Ashburner, B.B. Avants, M.C. Chiang, et al., Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration, NeuroImage 46 (2009).

- [5] D. Shen, C. Davatzikos, HAMMER: hierarchical attribute matching mechanism for elastic registration, *IEEE Trans. Med. Imaging* 21 (2002).
- [6] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration.," *NeuroImage*, vol. 45, 2009.
- [7] S.K. Warfield, K.H. Zou, W.M. Wells, Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation, *Med. Imaging, IEEE Trans. on* 23 (2004) 903–921.
- [8] M.J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, et al., STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation, *Med. Image Anal.* 17 (2013) 671–684.
- [9] P. Coupe, J.V. Manjon, V. Fonov, J. Pruessner, M. Robles, D.L. Collins, Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation, *NeuroImage* 54 (2011) 940–954.
- [10] F. Rousseau, P.A. Habas, C. Studholme, A supervised patch-based approach for human brain labeling, *IEEE Trans. Med. Imaging* 30 (2011) 1852–1862.
- [11] Y. Hao, T. Wang, X. Zhang, Y. Duan, C. Yu, T. Jiang, et al., Local Label Learning (LLL) for Subcortical Structure Segmentation: Application to Hippocampus Segmentation, *Hum. Brain Mapp.* (2013).
- [12] M. Kim, G. Wu, W. Li, L. Wang, Y.-D. Son, Z.-H. Cho, et al., Automatic hippocampus segmentation of 7.0 Tesla MR images by combining multiple atlases and auto-context models, *NeuroImage* (2013).
- [13] H. Wang, J. W. Suh, J. Pluta, M. Altinay, and P. Yushkevich, "Regression-Based Label Fusion for Multi-Atlas Segmentation," in *CVPR 2011*, 2011.
- [14] D. Zikic, B. Glocker, A. Criminisi, Atlas Encoding by Randomized Forests for Efficient Label Propagation, *Presente. MICCAI* (2013).
- [15] M.R. Sabuncu, B.T.T. Yeo, K. Van Leemput, B. Fischl, P. Golland, A Generative Model for Image Segmentation Based on Label Fusion, *Med. Imaging, IEEE Trans. on* 29 (2010) 1714–1729.
- [16] F. Rousseau, P.A. Habas, C. Studholme, A Supervised Patch-Based Approach for Human Brain Labeling, *IEEE Trans. Med. Imaging* 30 (2011) 1852–1862.
- [17] T. Tong, R. Wolz, P. Coupé, J. Hajnal, D. Rueckert, Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling, *NeuroImage* 76 (2013) 11–23.
- [18] D. Zhang, Q. Guo, G. Wu, and D. Shen, "Sparse Patch-Based Label Fusion for Multi-Atlas Segmentation," presented at the MBIA, Nice, France, 2012.
- [19] H. Wang, J.W. Suh, S.R. Das, C. Craige, P.A. Yushkevich, Multi-Atlas Segmentation with Joint Label Fusion, *IEEE Trans. Pattern Anal. Mach. Intell.* (2012).
- [20] G. Wu, Q. Wang, D. Zhang, F. Nie, H. Huang, D. Shen, A Generative Probability Model of Joint Label Fusion for Multi-Atlas Based Brain Segmentation, *Med. Image Anal.* 18 (2014) 881–890.
- [21] D. Nistér, H. Stewénius, Scalable Recognition with a Vocabulary Tree, *Presente. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2006).
- [22] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1996.
- [23] X. Jiang, J. Lai, Sparse And Dense Hybrid Representation via Dictionary Decomposition for Face Recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (2015) 1067–1079.
- [24] S. Kim, E. P. Xing, "Tree-Guided Group Lasso for Multi-Task Regression with Structured Sparsity," presented at the International Conference on Machine Learning, Haifa, Israel, 2010.
- [25] J. Liu and J. Ye, "Moreau-Yosida Regularization for Grouped Tree Structure Learning," presented at the Advances in Neural Information Processing System (NIPS 2010), Vancouver, Canada, 2010.
- [26] T. Vercauteren, X. Pennec, A. Perchant, N. Ayache, Diffeomorphic demons: efficient non-parametric image registration, *NeuroImage* 45 (2009) S61–S72.
- [27] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. Munsell, D. Shen, Hierarchical Multi-atlas Label Fusion with Multi-scale Feature Representation and Label-specific Patch Partition, *NeuroImage* 106 (2015) 34–46.

Chen Zu is a Ph.D. student at Nanjing University of Aeronautics and Astronautics, China. He received the B.S. and M.S. degrees from Nanjing University of Aeronautics and Astronautics, China, in 2010 and 2013, respectively. He is currently a visiting student with the University of North Carolina at Chapel Hill, USA, where he works with Dr. Guorong Wu. His main research interests include neuroimaging analysis, machine learning, pattern recognition, and data mining.

Zhengxia Wang received her Ph.D. Degree in computer software and theory from Chongqing University, China. She is now the visiting scholar in University of North Carolina at Chapel Hill. Her main research interests include genetic regulatory network, stability of dynamical system and application in image processing. Her current research interests mainly focus on medical image analysis and the application to the brain diseases.

Daoqiang Zhang received the B.S. and Ph.D. degrees in computer science from Nanjing University of Aeronautics and Astronautics (NUAA), China, in 1999 and 2004, respectively. In 2004, he joined the Department of Computer Science and Engineering, NUAA, as a Lecturer, where he is currently a Professor. His current research interests include machine learning, pattern recognition, data mining, and medical image analysis. He has published over 100 scientific articles in refereed international journals such as the *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *NeuroImage*, *Human Brain Mapping*, and conference proceedings such as International Joint Conferences on Artificial Intelligence, IEEE International Conference on Data Mining, and International Conference on Medical Image Computing and Computer Assisted Interventions. Dr. Zhang is a member of the Machine Learning Society of the Chinese Association of Artificial Intelligence and the Artificial Intelligence and Pattern Recognition Society of the China Computer Federation.

Peipeng Liang received the Ph.D. degree in 2009 in computer application technology from Beijing University of Technology, Beijing, China. He is currently an Associate Professor of the Department of Radiology, Xuanwu Hospital, Capital Medical University. His research interests include human brain atlas, imaging marker of neurodegenerative disease, and neural mechanism of human inductive reasoning. He has published 40 papers in international journals. He serves as EBM for several journals of reviewers for about 20 international journals.

Yonghong Shi received the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China. She is currently with the Digital Medical Research Center, School of Basic Medical Science, Fudan University, and Shanghai Key Laboratory of Medical Image Computing and Computer Assisted Intervention, as an Associate Professor. Her research interests focus on medical image computing and computer assisted diagnosis.

Dinggang Shen is a Professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering in the University of North Carolina at Chapel Hill (UNC-CH). He is currently directing the Center for Image Analysis and Informatics, the Image Display, Enhancement, and Analysis (IDEA) Lab in the Department of Radiology, and also the medical image analysis core in the BRIC. He was a tenure-track assistant professor in the University of Pennsylvania (UPenn), and a faculty member in the Johns Hopkins University. Dr. Shen's research interests include medical image analysis, computer vision, and pattern recognition. He has published more than 700 papers in the international journals and conference proceedings. He serves as an editorial board member for six international journals. He also served in the Board of Directors, The Medical Image Computing and Computer Assisted Intervention (MICCAI) Society, in 2012–2015.

Guorong Wu received the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China. He is currently with the Image Display, Enhancement, and Analysis Research Laboratory, The University of North Carolina, Chapel Hill, as an Assistant Professor. His research interests focus on fast and robust analysis of large population data, computer assisted diagnosis, and image guided radiation therapy.