Robust Patch-Based Multi-Atlas Labeling by Joint Sparsity Regularization

Guorong Wu1, Qian Wang1,2, Daoqiang1,3, and Dinggang Shen1

1 Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA
{grwu,dgshen}@med.unc.edu
2 Department of Computer Science, University of North Carolina at Chapel Hill, USA
qianwang@email.unc.edu
3 Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China
dqzhang@nuaa.edu.cn

Abstract. Automated labeling of anatomical structures on MR brain images is widely investigated in many neuroscience and clinic studies to quantitatively measure either individual or group structural/functional difference. To address the issue of registration errors that affect multi-atlas based labeling, patch-based label fusion methods have been recently proposed by inspecting the similarity of each patch in the subject with all possible local patches from multiple atlases. Although the patch-based methods take the advantage of one-to-many correspondences to look for the correct matches for each subject point, it also takes the risk of including misleading local patches with different labels, which might undermine the final label fusion result. To alleviate this issue, we introduce the concept of sparsity in non-local label fusion by considering only a small number of patches from multiple atlases that can best represent the subject patch. Also, note that the procedure of label fusion in the previous methods is often performed independently for each point, thus unable to guarantee the spatial consistency in labeling results. To address this issue, we propose a novel patch-based method with $l_{2,1}$-norm regularization to enforce joint sparsity during the label fusion. Promising results on NIREP-NA0 and LONI-LPBA40 datasets have demonstrated much better labeling performance of our proposed method, compared to other state-of-the-art multi-atlas labeling methods.

1 Introduction

With the advent of magnetic resonance (MR) imaging technique, image analysis on MR images plays a very important role in quantitatively measuring the structure difference between either individuals or groups. In many neuroscience and clinic studies, some regions-of-interest (ROIs), e.g., hippocampus, in the human brain are specifically investigated due to their relatedness with brain diseases such as dementia. Consequently, automation of accurate labeling and measurement of anatomical structures becomes significantly important in those studies to deal with huge volume of clinical
data. However, the automatic labeling still remains as a challenging problem due to the high complexity of brain structures and variability between individual brains.

Multi-atlas based labeling methods have achieved a great success in the last decade, which can be classified into two types, i.e., registration-based and patched-based methods. Registration-based multi-atlas labeling methods [1, 2] usually use the deformable registration to establish the correspondence between each atlas with known labels and the subject to be labeled. After that, the procedure of label fusion is applied on each point of the underlying subject, based on the context of all warped atlas images. As pointed in [3, 4], however, the performance of registration-based labeling method is limited by the registration accuracy.

Recently, patch-based labeling methods, which need no explicit registration between atlases and subject images, have emerged as an important type of multi-atlas based labeling method [3-6]. The main assumption of this approach is that, if patches of the input subject image are locally similar to the patches of atlases, they should have a similar label [4]. Specifically, a weighted graph is computed on each subject point, where the weights on the links in the graph are the function of the similarity between the underlying subject patch and all local patches from the atlases. Then, the weighted label fusion is performed to determine the label for each subject point by considering all labels from all atlases and their associated weights in the graph.

However, there are two limitations related to the current patch-based methods. First, the graph on each point is built upon all possible local patches from all atlases, as long as they have sufficiently high patch similarities. However, some mismatches might also lead to high similarity values, since the neighborhood size for calculating patch similarity is often too small to ensure correct correspondence matching. Therefore, it might undermine the label fusion result due to the lack of reasonable constraint in building the graph. Second, the procedure of label fusion is usually performed independently on each subject point, without considering that the constructed graphs and the propagated labels should be consistent in the spatial domain of subject image.

Inspired by the discriminative power of sparse representation [7, 8], we propose a novel patch-based multi-atlas labeling method, by introducing the sparsity into the construction of weighted graph for each subject point. Specifically, we formulate the graph construction as a regression problem which seeks for a sparse representation of the underlying subject patch with regard to a large number of local patches in the multiple atlases. In this way, we are able to reduce the risk of taking incorrect or ambiguous patches and finally consider only a small number of well-matched patches. Furthermore, we notice that the neighboring points with similar local image appearance should have similar sparsity patterns. Thus, we first partition the whole image domain into a set of overlapping groups according the complexity of image appearance by using Octree [9]. Then, the sparse representations of all points in each group are simultaneously estimated by using $l_{2,1}$-norm to enforce the joint sparsity, which requires that (1) the weighting vector on each point should be as sparse as possible, and (2) the weighting vectors of all points in the group should be as consistent as possible.

Our proposed patch-based multi-atlas labeling method has been evaluated on NIREP-NA0 [10] and LONI-LPBA40 [11] datasets, both with manually-labeled
ROIs. In the experiments, we compared our method with recently published patched-based labeling methods (PBL)\cite{3,4}, indicating that our method can achieve significant improvement over these state-of-the-art methods.

2 Methods

2.1 Motivation

In general, the goal of multi-atlas based labeling is to determine the label for each point $x$ in the subject image $I$, by using $N$ atlases $T_s$, along with their label maps $L_s$ ($s = 1, ..., N$). For patch-based labeling\cite{3,4}, a weighted graph is constructed on each subject point $x \in I$, where each link describes the contribution (called as weight $w(x, y_s)$) of each atlas point $y_s \in T_s$ in determining the label for subject point $x$. Usually the weight $w(x, y_s)$ is defined based on the heat kernel as:

$$w(x, y_s) = \begin{cases} \exp \left( -\frac{|p_I(x) - p_{T_s}(y_s)|^2}{\sigma^2} \right), & y_s \in \theta(x), \\ 0, & \text{otherwise} \end{cases}$$

(1)

where $\sigma$ denotes the kernel width. $p_I(x)$ and $p_{T_s}(y_s)$ are the column vectors of image intensities in local patches of input image $I$ and atlas $T_s$, with $x$ and $y_s$ as their centers respectively. To save the computational time, only the points within a search neighborhood $\theta(x)$ in the atlases will be considered. Then the label of subject point $x$ can be determined by following Eq. 2:

$$L_I(x) = \frac{\sum_{s=1}^{N} \sum_{y_s \in T_s} w(x, y_s) \cdot L_s(y_s)}{\sum_{s=1}^{N} \sum_{y_s \in T_s} w(x, y_s)}$$

(2)

Based on Eq. 2, it is obvious that all candidate points $y_s \in \theta(x)$ will be considered in the label fusion, regardless of their high or low contributions. This one-to-many point matching strategy used in the patch-based methods has the advantage of looking for the right corresponding patches in each atlas domain, which might be missed by one-to-one point matching in registration-based methods. However, it also takes a risk of incorporating many incorrect matches with considerably high similarity, although some pre-selection criteria based on the patch intensities\cite{3} have been used to filter out some irrelevant patches. Fig. 1 demonstrates our motivation related to this issue. In conventional patch-based method\cite{3}, each subject point $x$ in the red box of Fig. 1 examines all possible local patches from different atlases (blue boxes in Fig. 1) to calculate the weights on the graph in the fashion of non-local mean (NLM). However, this procedure is independently performed on each subject point $x$, resulting each column of the matrix shown in Fig. 1(a) displaying the weighting vector of particular $x$ and each row of matrix representing the weights of the same atlas patch across all subject points in the red box. Here, we apply a deformable registration method to identify the correspondences of subject points in the red box w.r.t. each atlas, which is designated by pink dashed boxes in Fig. 1. Then we sort the row order of matrix by putting row vectors associated with the atlas points in the pink boxes (good matches) in the top of matrix and the remaining (non-optimal matches) in the bottom. It is worth noting that deformable registration is not required in our patch-based labeling.
method (It is only used here for demo. All atlases in the experiment are only linearly registered with the subject). As shown in Fig. 1(a), a lot of non-optimal matched points (below the pink line) also have significant weights by conventional patch-based method, thus misleading the label fusion in Eq. 2. Apparently, one good solution to alleviate this issue is to keep the large weights to the good matches while suppress the distraction from non-optimal or incorrect matches, as described below.

Fig. 1. Illustration of graph construction on subject point x. The graph weights computed with non-local mean (NLM), sparsity-only based constraint (l1-norm), and joint sparsity constraint (l2,1-norm) are displayed in (a)-(c), respectively.

2.2 Graph Construction with Joint Sparsity

Inspired by the recent successes of sparse representation in face recognition and multi-task learning, we propose to use the sparsity on weighting vector \( \mathbf{w}(x) = \{w(x, y_s)\}_{y_s \in n(x), s = 1, ..., N} \) as a constraint to construct the graph and estimate its weights. Intuitively, the sparsity constraint on \( \mathbf{w}(x) \) encourages to use only a small number of similar patches from atlases to efficiently represent the local patch \( \mathbf{p}_i(x) \) in the subject image. In this way, the risk of incorporating the misleading patches (as shown in Fig. 1(a)) can be greatly alleviated.

Moreover, it is reasonable to require that the neighboring points should have similar sparse representations, since they generally have very similar local appearance. Hence, in our patch-based labeling method, we go one step further to apply not only the sparsity on a single point, but also the joint sparsity in the spatial neighborhood. Specifically, the weights on the graphs for a group of subject points will be simultaneously estimated to ensure that (1) the weighting vector on each subject point is the sparse representation of all possible patches in the atlases, and (2) the weighting vectors across all points in the group are consistent.

Hereafter, we improve the conventional patch-based labeling method from the way of non-local mean to the regression problem with the joint sparsity constraint. For all the points in a particular group \( \mathcal{G}_t \), we first assemble local patches of all subject points into a matrix \( \mathbf{Y} = [\mathbf{p}_i(x)] (x \in \mathcal{G}_t) \). Similarly, we arrange intensity of each atlas patch into a column vector \( \mathbf{p}_{r_s}(y_s) \) and further assemble all column vectors into another matrix \( \mathbf{A} \). Then, we seek for the sparse representation of all local patches in the group \( \mathcal{G}_t \) by minimizing the fitting error \( \| \mathbf{AW} - \mathbf{Y} \|_F^2 \), where each column in matrix \( \mathbf{W} \) is the weighting vector \( \mathbf{w}(x) \) on the particular point \( x \in \mathcal{G}_t \). It is worth noting that \( \mathbf{W} = \ldots \)
\[ [w_{ij}] \] is a \( m \times n \) matrix, with \( m \) and \( n \) denoting the number of candidate patches in the atlases and the number of subject points in the group \( G_t \), respectively. To enforce the joint sparsity, \( l_{2,1} \)-norm [12] is applied upon \( W \) to construct the graphs in the entire group \( G_t \) as:

\[
\| W \|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} w_{ij}^2 }.
\] (3)

Apparently, \( l_{2,1} \)-norm encourages all column vectors in \( W \) share the similar sparsity pattern.

After the whole image domain \( \Omega \) has been partitioned into a number of spatially-connected groups (i.e., \( \bigcup_t G_t = \Omega \)), we optimize the following objective function to get the graph weights for all points in for the group \( G_t \):

\[
\min_{W} \| A W - Y \|^2_{2} + \lambda \| W \|_{2,1}, \text{s.t.}, w_{ij} > 0,
\] (4)

where \( \lambda \) is a scaling parameter to control the strength of \( l_{2,1} \)-norm regularization. To solve this optimization problem, we use the \( l_{2,1} \)-regularized Euclidian projection method in [12, 13].

The matrix shown in Fig. 1(c) displays jointly estimated weighting vectors for all subject points in the red box of Fig. 1 with \( l_{2,1} \)-norm constraint. It is clear that each weighting vector (i.e., each column of the matrix in Fig. 1(c)) appears sparse enough to exclude the non-optimal matches (i.e., significant weights are assigned only to the candidate patches above the pink line in Fig. 1(c)). More importantly, the weighting vectors share a similar sparsity pattern. It is also worth noting that our objective function in Eq. 4 will be degenerated to \( l_{1} \)-norm based Lasso problem [8], when each point is regarded as a group. As shown in Fig. 1(b), the sparsity patterns of all points in the red box are not consistent, although each weighting vector is sparse enough.

2.3 Spatially Consistent Patch-based Multi-Atlas Labeling

In our spatially consistent patch-based labeling method (SCPBL), we first use Octree technique [9] to partition the whole brain into a number of non-overlapping groups (as shown in the left side of Fig. 2), each with similar image appearance. Here, for each block, we use its intensity entropy as the complexity measurement to decide whether it is necessary to further divide it into smaller sub-blocks. After completing the partition for the whole image, we optimize the objective function in Eq. 4 to jointly estimate the weighting vectors \( W \) (as displayed in the middle of Fig. 2) for all points in each final block. To make sure that the labeling results are spatially smooth between two groups, we further expand the blocks by 1-2 pixels to obtain the overlapped groups in optimizing Eq. 4. For label fusion, we generally follow Eq. 2 to get the label for each point \( x \). However, since we focus on binary segmentation (i.e., two classes with label 0 and 1) in this paper, the final label \( \Gamma(x) \) on point \( x \) needs to be binarized as follows:

\[
\Gamma(x) = \begin{cases} 
1, & L_t(x) > 0.5 \\
0, & \text{otherwise} 
\end{cases}
\] (5)
where \( L_1(x) \) is calculated from Eq. 2. The right side of Fig. 2 demonstrates the labeling result of left superior temporal gyrus on one subject from NIREP-NA0 database by our patch-based labeling method.

3 Experimental Results

\[
A \approx W Y \quad \text{Labeling result}
\]

![Fig. 2. Illustration of our spatially consistent patch-based labeling method with \( l_{2,1}\)-norm. First, subject image is partitioned into a number of groups by Octree (left). For each group, we optimize the energy function with \( l_{2,1}\)-norm regularization to simultaneously estimate the graph weights for all points in the same group (middle). After repeating this procedure on every group in the image domain, the labeling result on each point can be determined according to the weighted label fusion procedure in Eq. 5 (right).](image)

In this section, we apply our spatially consistent patch-based labeling method (SCPBL) in a multi-atlas framework to 16 NIREP-NA0 (with 32 manually labeled ROIs) and 40 LONI-LPBA40 (with 54 manually labeled ROIs) brain images. To evaluate the labeling performance of our SCPBL method, it is compared with the recently developed PBL methods [3, 4]. In order to show the advantage of using our new strategy of joint sparsity, we further make a degenerated version of our method by considering each point as an independent group and then replacing \( l_{2,1}\)-norm with \( l_1\)-norm in Eq. 4. We call this degenerated method as sparse-only patch-based method (SPBL).

For those three methods under comparison, i.e., PBL, SPBL, and SCPBL, the patch size is fixed to \( 3 \times 3 \times 3 \text{mm}^3 \) and the search neighborhood (\( \theta \) in Eq. 1) is fixed to \( 5 \times 5 \times 5 \text{mm}^3 \). In each experiment below, we report the Dice overlap ratio between ground truths and estimated labeling results in a leave-one-out cross-validation fashion. For example, we use 15 images among 16 images in NIREP dataset as the atlases to label the rest image in each leave-one-out case. For each ROI, we set the label of each atlas point as 1, if it belongs to the underlying ROI and 0 otherwise.

3.1 Experiment Result on NIREP-NA0 Dataset

The NIREP-NA0 dataset consists of 16 MR images of 8 normal male adults and 8 normal female adults, each with 32 manually-delineated gray matter ROIs. The image size is \( 256 \times 300 \times 256 \) with voxel dimension \( 0.7 \times 0.7 \times 0.7 \text{mm}^3 \).

The obtained Dice ratios of all 32 ROIs are 73.37±3.25% by PBL, 74.58±2.87% by SPBL, and 76.33±2.25% by our SCPBL method, respectively. The average and standard deviation of Dice ratios in 8 pairs (left and right) of selected ROIs (note that
here we select 8 common ROIs in both NIREP-NA0 and LONI-LPBA40 dataset) is displayed in Fig. 3(a) by PBL in blue, SPBL in green, and our SCPBL in red, respectively. It is clear that our method achieves the best labeling accuracy in most ROIs.

Fig. 4 shows the labeling result of left superior temporal gyrus (top row) and left postcentral gyrus (bottom row) by PBL (b), SPBL (c), and SCPBL (d). Comparing to the ground truth (manually-delineated ROIs) displayed in (a), our proposed method achieves the best labeling results in both accuracy and spatial smoothness.

3.2 Experiment Result on LONI-LPBA40 Dataset

LPBA40 dataset consists of 40 brain images, each with 54 manually labeled ROIs (excluding cerebrum and brainstem), most of which are within the cortex. The average and standard deviations of Dice ratio of all 54 ROIs are 75.06±2.35% by PBL, 76.46±1.96% by SPBL, and 78.04±1.34% by our SCPBL method. Also, the average and standard deviation of Dice ratios in 8 pairs of selected ROIs are displayed in Fig. 3(b) by PBL in blue, SPBL in green, and our SCPBL in red, where our method outperforms in most of ROIs.

3.3 Computation Time

Large portion of computation time in our method is used for $l_{2,1}$-norm based optimization. Taking a typical ROI (left occipital lobe with ~60,000 points) of one NIREP-NA0 image as an example, the conventional PBL takes 10 minutes to complete the labeling. With the same parameters (patch size and search range) and computational environment (Quad-core CPU@2.4GHz with 64GB memory), our SPBL and SCPBL methods take 28 minutes and 45 minutes, respectively, since the $l_{2,1}$-norm based optimization is much more complex than the one with $l_1$-norm regularization.
Fig. 4. The labeling results of left superior temporal gyrus (top row) and left postcentral gyrus (bottom row). The manually-labeled ROIs (ground truth) are shown in (a). The labeling results by PBL, SPBL, and SCPBL are displayed in (b)-(d), respectively.

4 Conclusions

In this paper, we propose a novel spatially consistent patch-based method (SCPBL) for multi-atlas labeling. To achieve the accurate and consistent labeling result, we simultaneously construct graphs and estimate their weights for a group of points, by using $l_{2,1}$-norm upon the graph weights to enforce the joint sparsity. Our proposed segmentation method has been evaluated on both NIREP-NA0 and LONI-LPBA40 datasets, achieving very promising results compared to the recently developed patch-based labeling methods.

References