Published in final edited form as:

Mach Learn Multimodal Interact. 2012 ; 7588: 45–53. doi:10.1007/978-3-642-35428-1_6.

Group Sparsity Constrained Automatic Brain Label Propagation

Shu Liao, Daoqiang Zhang, Pew-Thian Yap, Guorong Wu, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina at Chapel Hill

Abstract

In this paper, we present a group sparsity constrained patch based label propagation method for multi-atlas automatic brain labeling. The proposed method formulates the label propagation process as a graph-based theoretical framework, where each voxel in the input image is linked to each candidate voxel in each atlas image by an edge in the graph. The weight of the edge is estimated based on a sparse representation framework to identify a limited number of candidate voxels whose local image patches can best represent the local image patch of each voxel in the input image. The group sparsity constraint to capture the dependency among candidate voxels with the same anatomical label is also enforced. It is shown that based on the edge weight estimated by the proposed method, the anatomical label for each voxel in the input image can be estimated more accurately by the label propagation process. Moreover, we extend our group sparsity constrained patch based label propagation framework to the reproducing kernel Hilbert space (RKHS) to capture the nonlinear similarity of patches among different voxels and construct the sparse representation in high dimensional feature space. The proposed method was evaluated on the NA0-NIREP database for automatic human brain anatomical labeling. It was also compared with several state-of-the-art multi-atlas based brain labeling algorithms. Experimental results demonstrate that our method consistently achieves the highest segmentation accuracy among all methods used for comparison.

1 Introduction

Automatic and accurate human brain labeling plays an important role in medical image analysis, computational anatomy, and pathological studies. Multi-atlas based image segmentation [1–4] is one of the main categories of methods for automatic human brain labeling. It consists of two main steps, namely the registration step, and the label fusion step. In the registration step, atlases (i.e., images with segmentation groundtruths) are registered to the image to be segmented. The corresponding segmentation groundtruth of each atlas is also registered to the image to be segmented accordingly. In the label fusion step, the registered segmentation groundtruths of atlases are mapped and fused to estimate the corresponding label map of the image to be segmented. In this paper, we mainly focus at the label fusion step.

There are many novel label fusion strategies proposed in the literature [5–9] for multi-atlas based image segmentation. For instance, the majority voting (MV) scheme assigns each voxel of the image to be segmented the label which appears most frequently at the same voxel position across all the registered atlas images. The simultaneous truth and performance level estimation (STAPLE) [9] scheme iteratively weights the contribution of each atlas based on an expectation maximization approach. Sabuncu *et al.* [7] proposed a novel probabilistic framework to capture the relationship between each atlas and the input image.

Recently, Rousseau *et al.* [4] proposed a novel label fusion method for automatic human brain labeling based on the non-local mean principle and patch based representation. More specifically, each reference voxel in the input image is linked to each voxel in atlas images based on their patch based similarity. Then, voxels in atlas images within a small neighborhood of the reference voxel are served as candidate voxels to estimate the label of the reference voxel. This method has several attractive properties: First, it does not require non-rigid image registration between atlases and input image [4]. Second, it allows one to many correspondences to select a set of good candidate voxels for label fusion. However, it also has several limitations: (1) Candidate voxels which have low patch similarity with the reference voxel will still contribute in label propagation. (2) The dependency among candidate neighboring voxels with the same anatomical label is not considered.

Therefore, we are motivated to propose a new group sparsity constrained patch based label propagation framework. Our method estimates the sparse coefficients to reconstruct the local image patch of each voxel in the input image from patches of its corresponding candidate voxels in atlas images, and then use the estimated coefficients as graph weights to perform label propagation. Candidate voxels have low patch similarity with the reference voxel in the input image can be effectively removed by the sparsity constraint. The group sparsity is also considered to capture the dependency among candidate voxels with the same anatomical label. Moreover, we extend the proposed framework to the reproducing kernel Hilbert space (RKHS) to capture the nonlinear similarity of the patch based representations. Our method was evaluated on the NA0-NIREP database and compared with several state-of-the-art label fusion methods. Experimental results show that our method achieves the highest segmentation accuracy among other methods under comparison.

2 Group Sparsity Constrained Patch Based Label Propagation

The general label propagation framework [4] can be summarized as follows: Given N atlas images I_i and their corresponding label maps L_i ($i = 1, \dots, N$) registered to the input image I_{new} , the label at each voxel position $\mathbf{x}$ in I_{new} can be estimated by Equation 1:

$$L_{new}(\mathbf{x}) = \frac{\sum_{i=1}^N \sum_{\mathbf{y} \in \mathcal{N}_i(\mathbf{x})} w_i(\mathbf{x}, \mathbf{y}) L_i(\mathbf{y})}{\sum_{i=1}^N \sum_{\mathbf{y} \in \mathcal{N}_i(\mathbf{x})} w_i(\mathbf{x}, \mathbf{y})}, \quad (1)$$

where $\mathcal{N}_i(\mathbf{x})$ denotes the neighborhood of voxel $\mathbf{x}$ in image I_i , which is defined as the $W \times W \times W$ subvolume centered at $\mathbf{x}$. $w_i(\mathbf{x}, \mathbf{y})$ is the graph weight reflecting the relationship between voxels $\mathbf{x}$ in I_{new} and $\mathbf{y}$ in I_i , which also determines the contribution of voxel $\mathbf{y}$ in I_i

during the label propagation process to estimate the anatomical label for $\mathbf{x}$. L_{new} is the estimated label probability map of I_{new} .

It should be noted that $L_i(\mathbf{y})$ is a vector of $[0, 1]^M$ denoting the proportion of each anatomical label at voxel $\mathbf{y}$, and M is the total number of anatomical labels of interest in the atlases. Therefore, $L_{new}(\mathbf{x})$ is a M dimensional vector with each element denoting the probability of each anatomical label assigned to voxel $\mathbf{x}$. Similar to [4], the final anatomical label assigned to each voxel $\mathbf{x}$ will be determined as the one with the largest probability among all the elements in $L_{new}(\mathbf{x})$.

In [4], the graph weight $w_i(\mathbf{x}, \mathbf{y})$ is estimated based on the patch based similarity between voxels $\mathbf{x}$ and $\mathbf{y}$ by Equation 2:

$$w_i(\mathbf{x}, \mathbf{y}) = \begin{cases} \Phi \left(\frac{\sum_{\mathbf{x}' \in P_{I_{new}}(\mathbf{x}), \mathbf{y}' \in P_{I_i}(\mathbf{y})} (I_{new}(\mathbf{x}') - I_i(\mathbf{y}'))^2}{2K\alpha} \right), & \text{if } \mathbf{y} \in \mathcal{N}_i(\mathbf{x}); \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Where K is the patch size, α is the smoothing parameter. $P_{I_{new}}(\mathbf{x})$ and $P_{I_i}(\mathbf{y})$ denote the $K \times K \times K$ image patches of images I_{new} and I_i centered at voxel $\mathbf{x}$ and $\mathbf{y}$, respectively. Φ is the smoothing kernel function. In this paper, the heat kernel with $\Phi(x) = e^{-x}$ is used similar to [4].

However, based on the graph weight defined by Equation 2, candidate voxels with low patch similarity to the reference voxel will still contribute in label propagation, which can lead to fuzzy label probability map and affect the segmentation accuracy as will be confirmed in our experiments. Therefore, we are motivated to estimate the graph weight $w_i(\mathbf{x}, \mathbf{y})$ based on the sparse coefficients to reconstruct the local patch of each voxel $\mathbf{x}$ in I_{new} from patches of its corresponding candidate voxels $\mathbf{y}$ in I_i such that candidate voxels with low patch similarity can be automatically removed due to the sparsity constraint. More specifically, for each voxel $\mathbf{x}$ in I_{new} , its patch can also be represented as a $K \times K \times K$ dimensional column vector, denoted as $\mathbf{f}_x$. For each voxel $\mathbf{y}$ in $\mathcal{N}_i(\mathbf{x})$, we also denote its corresponding patch as a $K \times K \times K$ dimensional column vector $\mathbf{f}_y^i$ ($i=1, \dots, N$) and organize them into a matrix $\mathbf{A}$. Then, the sparse coefficient vector β_x for voxel $\mathbf{x}$ is estimated by minimizing Equation 3:

$$E(\beta_x) = \frac{1}{2} \|\mathbf{f}_x - \mathbf{A}\beta_x\|_2^2 + \lambda \|\beta_x\|_1, \quad \beta_x \geq 0 \quad (3)$$

where $\|\cdot\|_1$ denotes the L1 norm to enforce the sparsity constraint to the reconstruction coefficient vector β_x . Minimizing Equation 3 is the constrained L1-norm optimization problem (i.e., Lasso) [10], and the optimal solution β_x^{opt} to minimize Equation 3 can be estimated by Nesterov's method [11]. Then, the graph weight $w_i(\mathbf{x}, \mathbf{y})$ can be set to the corresponding element in β_x^{opt} with respect to $\mathbf{y}$ in image I_i .

However, the graph weight estimated by Equation 3 still treats candidate voxels as independent instances during the label propagation process without encoding the prior

knowledge of whether they are belonging to the same anatomical group. As stated in [12], without considering this prior knowledge can lead to significant representation power degradation. Therefore, we also enforce the group sparsity constraint in the proposed framework. More specifically, suppose voxels $\mathbf{y}$ in $\mathcal{N}_i(\mathbf{x})$ belonging to C ($C \leq M$) different anatomical structures based on $L_i(\mathbf{y})$. Then, the patch based representation of all the voxels $\mathbf{y}$ in $\mathcal{N}_i(\mathbf{x})$ belonging to the j th ($j = 1, \dots, C$) anatomical structure form the matrix $\mathbf{A}_j$, with the corresponding reconstruction coefficient vector β_x^j . Therefore, matrix $\mathbf{A}$ in Equation 3 can

be represented as $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_C]$, and $\beta_x = \left[(\beta_x^1)^T, (\beta_x^2)^T, \dots, (\beta_x^C)^T \right]^T$. Thus, the objective energy function by incorporating the group sparsity constraint can be expressed by Equation 4:

$$E(\beta_x) = \frac{1}{2} \|\mathbf{f}_x - \mathbf{A}\beta_x\|_2^2 + \lambda \|\beta_x\|_1 + \sum_{j=1}^C \gamma_j \|\beta_x^j\|_2, \quad \beta_x \geq 0. \quad (4)$$

Where γ_j denotes the weight with respect to the j th anatomical structure, and $\|\cdot\|_2$ denotes the L2 norm.

3 Group Sparsity Constrained Patch Based Label Propagation in Kernel Space

Kernel methods have attractive property that they can capture the nonlinear relationship of patches of different voxels in the high dimensional feature space. It is shown in [13] that kernel-based sparse representation method can further improve the recognition accuracy for the application of face recognition. Therefore, we are motivated to extend the group sparsity constrained patch based label propagation framework to the reproducing kernel Hilbert space (RKHS).

More specifically, given a nonlinear mapping function ϕ which can map the original patch based representation $\mathbf{f}_x$ of each voxel $\mathbf{x}$ in the high dimensional feature space as $\phi(\mathbf{f}_x)$, the objective energy function expressed by Equation 4 for group-sparse-patch based representation is converted to Equation 5:

$$E(\beta_x) = \frac{1}{2} \|\phi(\mathbf{f}_x) - \phi(\mathbf{A})\beta_x\|_2^2 + \lambda \|\beta_x\|_1 + \sum_{j=1}^C \gamma_j \|\beta_x^j\|_2, \quad \beta_x \geq 0. \quad (5)$$

Based on the kernel theory, the product between two mapped samples $\phi(\mathbf{f}_x)$ and $\phi(\mathbf{f}_y)$ in the high dimensional feature space can be represented by a kernel function κ , where $\kappa(\mathbf{f}_x, \mathbf{f}_y) = (\phi(\mathbf{f}_x))^T (\phi(\mathbf{f}_y))$. Then, Equation 5 can be rewritten as Equation 6:

$$\begin{aligned} E(\beta_x) &= \frac{1}{2} \|\phi(\mathbf{f}_x) - \phi(\mathbf{A})\beta_x\|_2^2 + \lambda \|\beta_x\|_1 + \sum_{j=1}^C \gamma_j \|\beta_x^j\|_2. \\ &= \frac{1}{2} \kappa(\mathbf{f}_x, \mathbf{f}_x) + \frac{1}{2} (\beta_x)^T \mathbf{G}(\beta_x) - (\beta_x)^T \mathbf{V}(x) + \lambda \|\beta_x\|_1 + \sum_{j=1}^C \gamma_j \|\beta_x^j\|_2, \quad \beta_x \geq 0. \end{aligned} \quad (6)$$

Where $\mathbf{G}$ is the gram matrix with $\{\mathbf{G}\}_{ij} = \kappa(\mathbf{u}_i, \mathbf{u}_j)$, where $\mathbf{u}_i$ denotes the i th column of matrix $\mathbf{A}$. $\mathbf{V}(\mathbf{x})$ is a vector with $\{\mathbf{V}(\mathbf{x})\}_i = \kappa(\mathbf{u}_i, \mathbf{x})$. Equation 6 can also be optimized by Nesterov's method [11] besides additionally computing $\mathbf{G}$ and $\mathbf{V}(\mathbf{x})$. In this paper, the

Gaussian radial basis function kernel is adopted, with $\kappa(\mathbf{u}_i, \mathbf{u}_j) = \exp(-\tau \|\mathbf{u}_i - \mathbf{u}_j\|^2)$.

4 Experimental Results

In this section, we evaluate our method for automatic brain labeling on the NA0-NIREP database [14]. The NA0-NIREP database has 16 3D brain MR images, among which eight are scanned from male adults and eight are scanned from female adults. Each image has been manually delineated into 32 gray matter regions of interest (ROIs). The images were obtained in a General Electric Signa scanner at 1.5T, with the following protocol: SPGR/50, TR 24, TE 7, NEX 1 matrix 256×192 , FOV 24 cm. All the images were aligned by the affine transformation with the FLIRT toolkit [15] as the preprocessing step. Exemplar images of the NA0-NIREP database before registration are shown in Figure 1.

The following parameter settings were adopted for the proposed method: A small patch size $3 \times 3 \times 3$ was adopted, and the searching neighborhood size was set to $5 \times 5 \times 5$. The same patch size and searching neighborhood size were adopted for the conventional patch based label propagation method in [4] under comparison. λ was set to 10^{-4} in Equations 3, 4, 5 and 6. The kernel parameter τ was set to $1/64$ by cross validation, and the weight γ_j in Equations 4, 5, and 6 was set to 1 for each anatomical structure.

The Dice ratio was adopted as the quantitative measure. It is defined as $D(A, B) = (2|A \cap B|) / (|A| + |B|)$, where A and B denote the regions of a specific type of anatomical structure of the estimated label map and the groundtruth, and $|\cdot|$ denotes the number of voxels within an image region. Half of the images were randomly selected from the 16 images as atlases, and the remaining half of the images were served as the testing images. The experiment was repeated for 10 times, and the average Dice ratio obtained for each ROI by using the majority voting (MV) scheme, STAPLE [9], the conventional patch based label propagation (CPB) method [4], and the proposed group-sparse-patch based label propagation method in the reproducing kernel Hilbert space (GSP + RKHS) are listed in Table 1 with respect to the left and right hemispheres of brain.

It can be observed from Table 1 that our method (GSP + RKHS) achieves the highest segmentation accuracy among all the methods under comparison almost at all the ROIs, which reflects the effectiveness of the proposed method.

To visualize the segmentation accuracy of the proposed method, Figure 2 shows a typical segmentation result for the parahippocampal gyrus in the left hemisphere brain with different approaches. The segmentation groundtruth is also given in the second column in Figure 2 for reference. It can be observed from Figure 2 that the segmentation result obtained by our method is the closest to the groundtruth, which implies the high segmentation accuracy of our method.

Moreover, to further analyze the contribution of each component of the proposed method, Figure 3 shows the average Dice ratios of the 32 ROIs obtained by using the proposed method with only the global sparsity constraint defined by Equation 3, with both the global and group sparsity constraints defined by Equation 4, and with both the global and group sparsity constraints defined in the kernel space as expressed by Equation 5. It can be observed from Figure 3 that by enforcing the group sparsity constraint, the segmentation accuracy (i.e., with 74.7% overall Dice ratio) can be improved compared with using only the global sparsity constraint only (i.e., with 73.3% overall Dice ratio), which reflects the importance of incorporating the representation power of groups of input variables as stated in [12]. Moreover, by extending the group-sparse-patch based label propagation framework to the nonlinear kernel space, the segmentation accuracy can be further improved (i.e., with 76.6% overall Dice ratio).

5 Conclusion

In this paper, we propose a new patch based multi-atlas label propagation method for automatic brain labeling. Different from the conventional patch based label propagation method, our method enforces sparsity constraints in the label propagation process such that only candidate voxels whose local image patches can best represent the local image patch of the reference voxel in the input image will contribute during label propagation. More precisely, the graph weights for label propagation are estimated based on the sparse coefficient vector to reconstruct the patch of each voxel in the input image from patches of its corresponding candidate voxels in atlas images. The group sparsity constraint is also enforced to capture the dependency among candidate voxels with the same anatomical label. Moreover, our method is extended to the reproducing kernel Hilbert space (RKHS) to capture the nonlinear relationship of patches of different voxels. Our method was evaluated on the NAO-NIREP database and compared with several state-of-the-art label fusion approaches for automatic brain labeling. Experimental results demonstrate that the proposed method achieves the highest segmentation accuracy among all methods used for comparison.

References

1. Coupe P, Manjon J, Fonov V, Pruessner J, Robles M, Collins D. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*. 2011; 54:940–954. [PubMed: 20851199]
2. Asman, AJ.; Landman, BA. Characterizing Spatially Varying Performance to Improve Multi-atlas Multi-label Segmentation.. In: Székely, G.; Hahn, HK., editors. *IPMI 2011. LNCS*. Vol. 6801. Springer; Heidelberg: 2011. p. 85-96.
3. Artaechevarria X, Munoz-Barrutia A, Ortiz-de Solorzano C. Combination strategies in multi-atlas image segmentation: Application to brain mr data. *IEEE TMI*. 2009; 28:1266–1277.
4. Rousseau F, Habas P, Studholme C. A supervised patch-based approach for human brain labeling. *IEEE TMI*. 2011; 30:1852–1862.
5. Rohlfing T, Russakoff D, Maurer C. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE TMI*. 2004; 23:983–994.
6. Heckemann R, Hajnal J, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*. 2006; 33:115–126. [PubMed: 16860573]

7. Sabuncu M, Yeo B, van Leemput K, Fischl B, Golland P. A generative model for image segmentation based on label fusion. *IEEE TMI*. 2010; 29:1714–1729.
8. van Rikxoort E, Isgum I, Arzhaeva Y, Staring M, Klein S, Viergever M, Pluim J, van Ginneken B. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *MedIA*. 2010; 14:39–49.
9. Warfield S, Zou K, Wells W. Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation. *IEEE TMI*. 2004; 23:903–921.
10. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*. 1996; 58:267–288.
11. Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers; 2004.
12. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*. 2006; 68:49–67.
13. Gao, S.; Tsang, IW-H.; Chia, L-T. Kernel Sparse Representation for Image Classification and Face Recognition.. In: Daniilidis, K.; Maragos, P.; Paragios, N., editors. *ECCV 2010, Part IV*. LNCS. Vol. 6314. Springer; Heidelberg: 2010. p. 1-14.
14. Christensen, GE.; Geng, X.; Kuhl, JG.; Bruss, J.; Grabowski, TJ.; Pirwani, IA.; Vannier, MW.; Allen, JS.; Damasio, H. Introduction to the Non-rigid Image Registration Evaluation Project (NIREP).. In: Pluim, JPW.; Likar, B.; Gerritsen, FA., editors. *WBIR 2006*. LNCS. Vol. 4057. Springer; Heidelberg: 2006. p. 128-135.
15. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 2002; 17:825–841. [PubMed: 12377157]

Fig. 1.
Exemplar images from the NA0-NIREP database

Fig. 2.

Typical segmentation results for the parahippocampal gyrus anatomical structure on the left hemisphere of the brain using different label fusion approaches

Fig. 3.

The average Dice ratio for each anatomical structure obtained by the proposed method with only the global sparsity constraint, with both the global and group sparsity constraints, and with both the global and group sparsity constraints in the reproducing kernel Hilbert space, respectively, on the NA0-NIREP database.

Table 1

The average Dice ratio (in %) of different brain ROIs obtained by using different label fusion approaches with respect to the left and right hemispheres of brain on the NAO-NIREP database. The highest Dice ratio for each ROI is bolded.

Brain ROIs (left/right)	MV	STAPLE [9]	CPB [4]	GSP + RKHS
Occipital Lobe	52.3 - 54.8	59.7 - 62.8	74.9 - 79.1	79.2 - 83.7
Cingulate Gyrus	58.8 - 60.4	60.9 - 63.7	71.5 - 73.6	77.3 - 79.2
Insula Gyrus	62.4 - 66.1	63.6 - 65.7	75.3 - 78.2	81.4 - 83.6
Temporal Pole	57.4 - 62.9	60.7 - 62.8	74.3 - 78.5	79.3 - 83.5
Superior Temporal Gyrus	46.7 - 45.2	50.9 - 52.4	65.4 - 67.5	75.2 - 75.4
Infero Temporal Region	56.3 - 58.4	58.6 - 57.2	74.1 - 77.3	81.5 - 81.8
Parahippocampal Gyrus	62.7 - 66.6	60.3 - 62.4	75.3 - 77.4	78.7 - 81.2
Frontal Pole	59.2 - 60.5	57.4 - 62.0	78.4 - 81.6	79.0 - 80.8
Superior Frontal Gyrus	49.6 - 52.3	52.7 - 54.8	71.8 - 72.5	77.3 - 78.6
Middle Frontal Gyrus	51.9 - 54.2	53.5 - 56.7	68.3 - 71.4	75.3 - 77.6
Inferior Gyrus	46.9 - 45.2	48.7 - 52.4	66.4 - 68.8	73.3 - 72.5
Orbital Frontal Gyrus	58.3 - 59.0	58.5 - 61.7	78.4 - 79.5	78.2 - 80.3
Precentral Gyrus	33.6 - 35.5	36.7 - 35.9	65.4 - 67.6	70.4 - 71.8
Superior Parietal Lobule	39.6 - 44.3	48.2 - 48.8	64.0 - 68.3	71.4 - 72.7
Inferior Parietal Lobule	46.2 - 47.8	49.7 - 52.3	67.2 - 71.5	72.4 - 74.8
Postcentral Gyrus	33.5 - 34.6	39.7 - 42.5	53.1 - 58.0	59.3 - 62.9