Non-local Atlas-guided Multi-channel Forest Learning for Human Brain Labeling

Guangkai Ma[1,2], Yaozong Gao[2,3], Guorong Wu[2], Ligang Wu[1], and Dinggang Shen[2]

[1] Space Control and Inertial Technology Research Center, Harbin Institute of Technology, Harbin, China
[2] Department of Radiology and BRIC, UNC at Chapel Hill, NC, USA
[3] Department of Computer Science, UNC at Chapel Hill, NC, USA

Abstract. Labeling MR brain images into anatomically meaningful regions is important in many quantitative brain researches. In many existing label fusion methods, appearance information is widely used. Meanwhile, recent progress in computer vision suggests that the context feature is very useful in identifying an object from a complex scene. In light of this, we propose a novel learning-based label fusion method by using both low-level appearance features (computed from the target image) and high-level context features (computed from warped atlases or tentative labeling maps of the target image). In particular, we employ a multi-channel random forest to learn the nonlinear relationship between these hybrid features and the target labels (i.e., corresponding to certain anatomical structures). Moreover, to accommodate the high inter-subject variations, we further extend our learning-based label fusion to a multi-atlas scenario, i.e., we train a random forest for each atlas and then obtain the final labeling result according to the consensus of all atlases. We have comprehensively evaluated our method on both LONI-LBPA40 and IXI datasets, and achieved the highest labeling accuracy, compared to the state-of-the-art methods in the literature.

1 Introduction

Automatic labeling of MR brain images has become a hot topic in the field of medical image analysis, since quantitative brain image analysis often relies on the reliable labeling of brain images. However, due to the high complexity of brain structures, it is still a challenging task for automatic brain labeling.

Recently multi-atlas based labeling methods have achieved a great success. In these methods, a set of already-labeled MR images, namely atlases, are used to guide the labeling of new target images [3, 9]. For example, Coupé et al. [6] proposed a non-local patch-based label fusion technique by using patch-based similarity as weight to propagate the neighboring labels from the aligned atlases to the target image, for potentially overcoming errors from registration. Instead of pair-wisely estimating the patch-based similarity, Wu et al. [7] adopted sparse representation to jointly estimate all patch-based similarities between a to-be-labeled target voxel and its neighboring voxels in all the atlases. However, the

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 719–726, 2015.
DOI: 10.1007/978-3-319-24574-4_86

traditional multi-atlas based labeling techniques are still limited: the definition of patch-based similarity is often handcrafted based on the predefined features, which might not be effective for labeling all types of brain structures.

On the other hand, learning-based methods have also attracted much attention recently. In these methods, a strong classifier is typically trained for each label/ROI, based on the local appearance features. For example, Zikic et al. [2] proposed atlas forest, which encodes an atlas by learning a classification forest on it. The final labeling of a target image is achieved by averaging the labeling results from all the selected atlas forests. Tu et al. [5] adopted the probabilistic boosting tree (PBT) for labeling. To further boost the performance, an auto-context model (ACM) was also proposed to iteratively refine the labeling results. The learning-based methods often determine a target voxel's label solely based on the local image appearance, without getting clear assistance from the spatial information of labels encoded in the atlases. Accordingly, their labeling accuracy could be limited, since patches with similar local appearance could appear in different parts of the brain.

In this paper, we propose a novel atlas-guided multi-channel forest learning method for labeling multiple ROIs (Regions of Interest). Here, multi-channel means multiple representations of a target image, which include features extracted from not only the target (intensity) image but also the label maps of all aligned atlases. Instead of labeling each target voxel with only its local image appearance from the target image, we also utilize label information from the aligned atlas. To further refine the labeling result, Haar-based multi-class contexture model (HMCCM) is also proposed to iteratively construct a sequence of classification forests by updating the context features. The final labeling result is the average over all labeling results from all atlas-specific forests. Validated on both LONI-LBPA40 and IXI datasets, our proposed method consistently outperforms both traditional multi-atlas based methods and learning-based methods.

The rest of the paper is organized as follows. Section 2 describes the proposed labeling method and its application to single-ROI and multi-ROI labeling. Experiments are performed and analyzed in Section 3. Finally, discussion and conclusion are given in the last section.

2 Method

In this section, we will first present notations used in our paper. Then, we will explain the learning procedure of our atlas-guided multi-channel forest, followed by application of the learned forests to single-ROI and multi-ROI labeling. Finally, we present HMCCM to iteratively refine the labeling results.

Notations. An atlas library $\mathbf{A}$ consists of multiple atlases $\{A_i = (I_i, L_i)|i = 1, \ldots, N\}$, where I_i and L_i are the intensity image and the label image/map of the i-th atlas, and N is the total number of atlases in the library $\mathbf{A}$. Set $T = \{T_j = (H_j, B_j)|j = 1, \ldots, M\}$ represents the training set, where H_j and B_j are the intensity image and the label image/map of the j-th training sample, and M is the total number of training samples. $A_i^j = \{I_i^j, L_i^j\}, i = 1, \ldots, N, j =$

$1, \ldots, M$ denotes the intensity (I_i^j) and label (L_i^j) images of the i-th atlas after mapping to the j-th training image. Each brain ROI is assigned with a ROI/label s, $s = 1, \ldots, S$, where S is the total number of ROIs.

2.1 Atlas-guided Multi-channel Forest Learning

To increase the flexibility of our learning procedure, we will train one multi-channel random forest $F_{i,s}$ for each atlas i and each ROI s. In this way, when a new atlas is added into $\mathbf{A}$, only the new multi-channel forest needs to be trained with the new atlas, while all previously trained forests can be reused.

To label the s-th ROI, we will learn a multi-channel forest $F_{i,s}$ with each atlas, i.e., the i-th atlas. To obtain more accurate label information from atlas, registration and patch selection are performed. First, during the $F_{i,s}$ learning, we non-rigidly register the i-th atlas image I_i onto each training target image H_j, to obtain the warped atlas image I_i^j and label map L_i^j. For each sample voxel x in H_j, we first extract its appearance features from a local patch of H_j, centered at x. To reduce the registration error and further get more accurate label from the atlas, according to the similarity between local intensity patches of training image and warped atlas image, we search a nearest voxel $c_1(x)$ (with largest similarity) of x from the warped atlas image I_i^j. For efficiency, the overall intensity difference within the patch is used as the similarity measurement [6]. Then, we extract label features from the local patch of $c_1(x)$ in the aligned atlas label image L_i^j. Finally, both appearance features and label features are combined to jointly characterize the appearance and spatial label context information of each sample voxel, and use it for inferring label. Afterwards, the positive and negative samples are taken inside and outside of the s-th ROI from every training image for multi-channel forest learning, as detailed below. The flowchart shown in Fig. 1 gives an illustration for learning one multi-channel forest.

Sampling Strategy: The positive and negative samples used to train multi-channel forest for the s-th ROI are randomly sampled inside and outside the s-th ROI, respectively. Intuitively, voxels around the ROI boundary are more difficult to be correctly classified than other voxels. Therefore, more samples are drawn around the ROI boundary, as shown in the right bottom of Fig. 1, and also the numbers of positive and negative samples are kept the same.

Feature Extraction: To train multi-channel forests for the i-th atlas, as mentioned above, every training image $H_j, j = 1, \ldots, M$, will be associated with its respective aligned i-th atlas $A_i^j = \{I_i^j, L_i^j\}$. More specifically, there are $S + 1$ different channels of features extracted for: **1 channel** of local appearance features extracted from the training image (e.g., H_j), and S **channels** of local label context features extracted from the aligned i-th atlas label map (e.g., L_i^j) with respect to each of S ROIs.

Local Appearance Features. The local image appearance features extracted from a given (training) target image include: 1) patch intensities 2) outputs from the first-order difference filters (FODs), second-order difference filters (SODs),

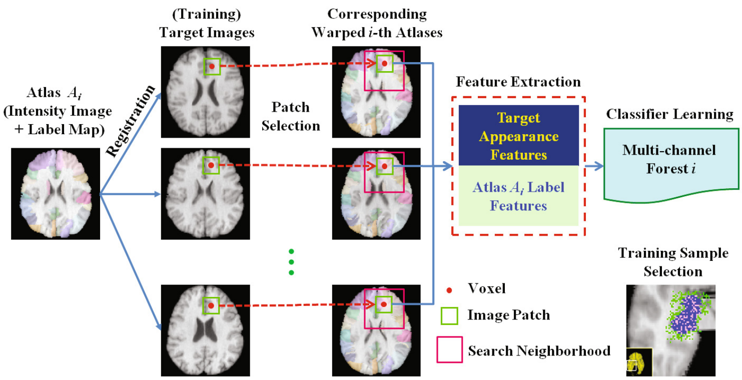

Fig. 1. The flowchart of our method for learning one multi-channel forest with the i-th atlas. An example for sample selection during the training stage is also given in the right-bottom corner, where blue points denote samples belonging to the ROI while green points denote samples belonging to the background. Note here that more samples are drawn around the ROI boundaries.

3D Hyperplan filters, 3D Sobel filters, Laplacian filters and range difference filters, and 3) the random 3D Haar-like features computed from a neighborhood. In addition, by randomly selecting different parameter values of features, the appearance features can capture rich texture information of the target image.

Local Label Context Features. To extract the label context features for each ROI, we first convert the multi-ROI atlas label map into S binary label maps, $L^j_{i,s}$, where $L^j_{i,s}$ corresponds to ROI s, with only voxels in ROI s having label 1 (positive) while all other voxels having label 0 (negative). Then, from each binary label map $L^j_{i,s}$, we uniformly and sparsely select 343 voxels within a $11 \times 11 \times 11$ neighborhood. Finally, a total of $125 \times S$ voxels are sampled, and their label values are served as local label context features.

2.2 Single-ROI and Multi-ROI Labeling

Single-ROI Labeling: To label a single ROI in a new target image, all atlases are first non-rigidly registered onto the target image. To effectively correct for inaccurate registration, we adopt the non-local strategy in the testing stage. Specifically, for a target voxel x to be labeled, we first perform a local patch search in the aligned atlas image (e.g. I^j_i) to select the top K atlas patches with similar appearance to the target patch centered at x. Here, the centers of the selected atlas patches are indexed as $c_k(x), k = 1, \ldots, K$. **1)** For each voxel $c_k(x)$, its label context features can be extracted from the binary label maps $\{L^j_{i,s}, s = 1, \ldots, S\}$. Then, these S channels of label context features and also one channel of appearance features computed from the target image can be combined as a feature representation of x. **2)** Afterwards, we apply the learned $F_{i,s}$ to estimate the label probability of x. **3)** We obtain K label probabilities

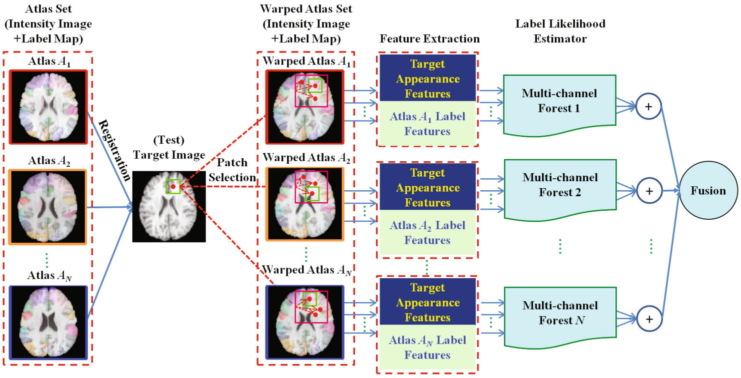

Fig. 2. A diagram for single-ROI labeling with our proposed atlas-guided multi-channel forest learning.

for K selected atlas patches, and then simply average them to obtain a final label for x. Note that, using the above step, each aligned atlas can use its own learned multi-channel forest for labeling the target image independently. Then, the labeling results from all N atlases can be further averaged to obtain the final labeling result for the target image. To increase the efficiency of voxel-wise labeling for the target image, we apply our method only to the voxels that receive votes from the warped atlas label maps. Fig. 2 gives an illustration of our single-ROI labeling method.

Multi-ROI Labeling: The extension from single-ROI labeling to multi-ROI labeling is straightforward. For each target voxel to be labeled, we first use labels of the corresponding voxels in the aligned atlases to find a set of candidate labels for this voxel. Then, we apply only the ROI classifiers responsible to those candidate labels for estimating the label probabilities of the target voxel, while all other ROI classifiers are excluded, and their corresponding label probabilities are simply set to zero. For all the ROIs, we can obtain S single-ROI labeling maps. To fuse these single-ROI labeling maps into one multi-ROI labeling map, the label of each target voxel is simply assigned by the one with the maximum probability across all different single-ROI label maps.

2.3 Haar-Based Multi-Class Contexture Model (HMCCM)

After applying our trained atlas-specific multi-channel forest to the target image, we can obtain a label probability map, which contains more relevant label context information (than the aligned atlas label map) for the target image. We, thus, further update the label context information from the newly obtained (tentative) label probability map, to learn a next multi-channel forest for refinement of labeling. By iterating this procedure, a sequence of classifiers (random forests) can be learned to iteratively improve the labeling result of the target image. In the following paragraphs, we detail the training and testing stages of HMCCM for the case of multi-ROI labeling.

Training: In the initial layer, for each atlas (e.g., the i-th atlas), we first train a set of atlas-specific multi-channel forests $\{F_{i,s}^1, s = 1, \ldots, S\}$. Then, by applying the trained $\{F_{i,s}\}$ to each training image, a set of initial label probability maps $P^1 = \{P_s^1\}$ can be obtained. In the second layer, we can extract the context information from the set of P^1, instead of binary label maps of the aligned atlas. Specifically, for each target voxel x in the target image, Haar-like features are extracted in the local patch centered at x from each P_s^1, for characterizing the multi-scale label context features. (Note that, in this study, for obtaining large-scale label context information, we adopt a large local patch). Then, we combine these updated label context features with the appearance features to re-train a next set of atlas-specific multi-channel forests $\{F_{i,s}^2\}$, which can be again used to estimate a next set of new label probability maps $P^2 = \{P_s^2\}$ for each training image. In the each of the following layers, the label context features are updated from the set of label probability maps computed in the previous layer, and then these updated features are combined with the appearance features of the target image to train a next set of atlas-specific multi-channel forests (corresponding to each ROI). Finally, after training totally O layers, we can obtain O subsequent sets of atlas-specific multi-channel forests, $\{F_{i,s}^o\}, o = 1, \ldots, O$.

Testing: For a new test H^t, each voxel is layer-wisely tested by the trained classifiers $\{F_{i,s}^o\}, o = 1, \ldots, O$. Specifically, for each atlas (e.g., the i-th atlas), we use the first layer of $\{F_{i,s}^1\}$ to obtain the initial label probability maps $P^{t,1} = \{P_s^{t,1}\}$ for H^t. In the following layer, we update the Haar-like features from the label probability maps of the previous layer as the context features. Then, these updated context features are combined with the appearance features of the test image and further input to the set of trained atlas-specific multi-channel forests of the current layer for obtaining a refined set of label probability maps of the test image. This procedure is iterated until reaching the last layer, thus, obtaining the final label probability maps for the test image (with the i-th atlas). The labeling results from all N atlases will be averaged to produce the final labeling.

3 Experimental Results

In this section, we apply our proposed method to the LONI-LPBA40 dataset [1] and IXI dataset (https://www.brain-development.org) for evaluating its performance in ROI labeling. We compared the proposed multiple atlas-guided multi-channel forest (MAMCF) and MAMCF+HCCM with two popular learning-based methods, i.e., standard random forests (SRF) [4] and auto-context model (ACM) [5]. Also, for comparison with multi-atlas based labeling methods, we apply majority voting (MV) and the conventional patch-based methods by non-local patch based labeling propagation (Nonlocal PBL) [6] and the recently proposed sparse patch-based labeling propagation (Sparse PBL) [6, 7]. To align each atlas image with the target image, affine registration is first performed by FLIRT. Then, diffeomorphic Demons is further performed for deformable registration. To quantitatively evaluate the labeling accuracy, we use the Dice Similarity Coefficient (DSC) to measure the overlap degree between automatic labeling and

Table 1. The mean and standard deviation of DSC (%) by MV, Non-local PBL, Sparse PBL, SRF, SRF+HMCCM and our method on LONI_LPBA40 and IXI datasets, respectively

Method	LONI-LPBA40	IXI
MV	78.55 ± 4.33	76.64 ± 4.56
Non-local PBL	78.58 ± 4.32	75.85 ± 4.70
Sparse PBL	80.21 ± 4.32	77.40 ± 4.52
SRF	72.48 ± 4.36	72.09 ± 4.98
SRF+ACM	73.83 ± 4.47	74.53 ± 4.49
MAMCF	81.89 ± 4.25	79.08 ± 4.41
MAMCF+HMCCM	82.56 ± 4.22	79.78 ± 4.34

manual labeling of each ROI. In the experiments, we use leave-one-out cross-validation to evaluate the performance of our method. For each test image, all other images are split into two equal parts: one used for training, and another used as an atlas images.

Parameters: For texture features (FODs, SODs etc.), parameter setting of each filter can be referred to [8]. For appearance patch size (11x11x11) and label patch size (11x11x11), we determine them by five-fold across validation on the training data. In first stage of our method, we extract total 2367 appearance features and 1331 label features for each ROI. In second stage of our method, we extract additional 500 haar-like features for each ROI from label probability maps. In the training stage, we train 20 trees for each multi-channel forest. The maximum tree depth is set to 20, and the minimum number of samples in the tree leaf node is set to 4.

LONI-LPBA40 Dataset: The dataset consists of 40 T1-weighted MRI brain images from 40 healthy volunteers, each with 54 manually labeled ROIs (excluding cerebrum and brainstem). Most of these ROIs are within the cortex. The second column of Table 1 shows the mean and standard deviation of DSC on 54 ROIs by the compared methods. Over the all 54 ROIs, the average DSCs achieved by MV, Nonlocal PBL, and Sparse PBL, SRF and SRF+ACM are $78.55\% \pm 4.33\%$, $78.58\% \pm 4.32\%$, $80.21\% \pm 4.32\%$, $72.48\% \pm 4.36\%$ and $73.83\% \pm 4.47\%$ respectively, which are lower than MAMCF ($81.89\% \pm 4.25\%$) and MAMCF+HMCCM ($82.56\% \pm 4.22\%$). In terms of average performance over all the ROIs, compared with other methods, MAMCF and MAMCF+HMCCM obtain statistically significant improvements ($p < 0.0001$) by the paired Student's t-test.

IXI Dataset: We use 30 images in the IXI dataset, which contains manual annotations of 80 structures (excluding cerebrum and brainstem). The third column of Table 1 shows the mean and standard deviation of DSC on all 80 ROIs. It can be observed that MAMCF+HMCCM ($79.78\% \pm 4.34\%$) methods are ranked top, followed by MAMCF ($79.08\% \pm 4.41\%$), Sparse PBL ($77.4\% \pm 4.52\%$), MV ($76.64\% \pm 4.56\%$), Non-local PBL ($75.85\% \pm 4.7\%$), SRF+ACM ($74.53\% \pm 4.49\%$) and SRF ($72.09\% \pm 4.98\%$). In terms of average performance over all the ROIs, compared with other methods, MAMCF and MAMCF+HMCCM obtain statistically significant improvements ($p < 0.0001$).

4 Conclusion

In this paper, we propose a novel atlas-guided multi-channel forest learning to effectively combine the advantages of both multi-atlas based labeling methods and learning-based labeling methods. Instead of labeling a target voxel based only on its own local image appearance, we also utilize label context information from the aligned atlas. A non-linear multi-channel forest is learned for automatically fusing all information. Furthermore, the Haar-based multi-class contexture model (HMCCM) is also proposed to enhance the structural and label context information of the target image. Specifically, we use Haar-like features to iteratively extract multi-scale label context information from the tentatively-estimated multi-ROI label probability maps of the target image. Our method shows more accurate labeling results than both the existing multi-atlas based labeling methods and learning-based labeling methods, on both LONI-LBPA40 and IXI datasets.

References

[1] Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39(3), 1064–1080 (2007)
[2] Zikic, D., Glocker, B., Criminisi, A.: Atlas encoding by randomized forests for efficient label propagation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 66–73. Springer, Heidelberg (2013)
[3] Wang, H., Sub, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. IEEE Trans. Pattern Analysis and Machine Intelligence 35(3), 611–623 (2013)
[4] Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
[5] Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3d brain image segmentation. IEEE Trans. Pattern Analysis and Machine Intelligence 32(10), 1744–1757 (2010)
[6] 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)
[7] Wu, G., Wang, Q., Zhang, D., Shen, D.: Robust patch-based multi-atlas labeling by joint sparsity regularization. In: MICCAI Workshop STMI (2012)
[8] Hao, Y., Wang, T., Zhang, X., Duan, Y., Yu, C., Jiang, T., Fan, Y.: Local label learning (LLL) for subcortical structure segmentation: application to hippocampus segmentation. Human Brain Mapping 35(6), 2674–2697 (2014)
[9] Gao, Y., Liao, S., Shen, D.: Prostate segmentation by sparse representation based classification. Medical Physics 39(10), 6372–6387 (2012)