Medical Image Analysis 36 (2017) 162–171

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Progressive multi-atlas label fusion by dictionary evolution

Yantao Song[a,b], Guorong Wu[b], Khosro Bahrami[b], Quansen Sun[a], Dinggang Shen[b,c,*]

[a] School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China
[b] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA
[c] Department of Brain and Cognitive Engineering, Korea University, Seoul, 02841, Republic of Korea

ARTICLE INFO

Article history:
Received 29 February 2016
Revised 8 October 2016
Accepted 18 November 2016
Available online 24 November 2016

Keywords:
Brain MRI
Sparse representation
Hippocampus
Label fusion
Multi-atlas

ABSTRACT

Accurate segmentation of anatomical structures in medical images is important in recent imaging based studies. In the past years, multi-atlas patch-based label fusion methods have achieved a great success in medical image segmentation. In these methods, the appearance of each input image patch is first represented by an *atlas patch dictionary* (in the image domain), and then the latent label of the input image patch is predicted by applying the estimated representation coefficients to the corresponding anatomical labels of the atlas patches in the *atlas label dictionary* (in the label domain). However, due to the generally large gap between the patch appearance in the *image domain* and the patch structure in the *label domain*, the estimated (patch) representation coefficients from the image domain may not be optimal for the final label fusion, thus reducing the labeling accuracy. To address this issue, we propose a novel label fusion framework to seek for the suitable label fusion weights by progressively constructing a dynamic dictionary in a layer-by-layer manner, where the intermediate dictionaries act as a sequence of guidance to steer the transition of (patch) representation coefficients from the image domain to the label domain. Our proposed multi-layer label fusion framework is flexible enough to be applied to the existing labeling methods for improving their label fusion performance, i.e., by extending their single-layer *static* dictionary to the multi-layer *dynamic* dictionary. The experimental results show that our proposed progressive label fusion method achieves more accurate hippocampal segmentation results for the ADNI dataset, compared to the counterpart methods using only the single-layer *static* dictionary.

1. Introduction

Magnetic resonance imaging (MRI) is an advanced medical imaging technique, which plays an essential role in neuroscience research and clinical studies. However, due to the large amount of MRI data produced every day, it is time-consuming and expensive to process medical images manually. Therefore, automated and accurate segmentation is in high demand in existing imaging-based studies, in order to *either* discover group differences between individual subjects *or* quantify subtle changes over time. For instance, the hippocampus is known as an important structure related to Alzheimer's disease, epilepsy, and schizophrenia (Devanand et al., 2007; Dickerson et al., 2001; Holland et al., 2012; Van Leemput et al., 2009). Therefore, automated and accurate segmentation of the hippocampus is critical.

However, since anatomical structures (i.e., hippocampus) vary significantly across individuals, the prior knowledge of shape and appearance learned from a certain template is often not sufficient for guiding the segmenta, multi-atlas based segmentation methods have been recently developed and achieved great success by letting the target labels on the target image follow the consensus of labels of multiple atlases with similar local image appearance. Generally, with more atlases, higher segmentation accuracy can be achieved by reducing the variations between the target and atlas images.

To do the segmentation, followed by registering atlas images to the target image, the latent anatomical label on each target image point can be determined by a certain label fusion strategy, such as majority voting (MV) (Heckemann et al., 2006; Rohlfing et al., 2005). Majority voting is a classical label fusion method, which simply chooses the label with the highest vote as the final label. To improve the labeling accuracy, local weighted voting (LWV) was also proposed by replacing the *hard* voting (considering only the label information) with *soft* voting which is proportional to the patch-wise appearance similarity (Sabuncu et al., 2010).

Apparently, the above *point-wise* label fusion strategies are highly dependent on the accuracy of image registration. To address the potential issue of inaccurate registration, many *patch-based* label fusion methods have been proposed in recent years

* Corresponding author.
E-mail addresses: yantaosong@hotmail.com (Y. Song), dgshen@med.unc.edu (D. Shen).

http://dx.doi.org/10.1016/j.media.2016.11.005
1361-8415/

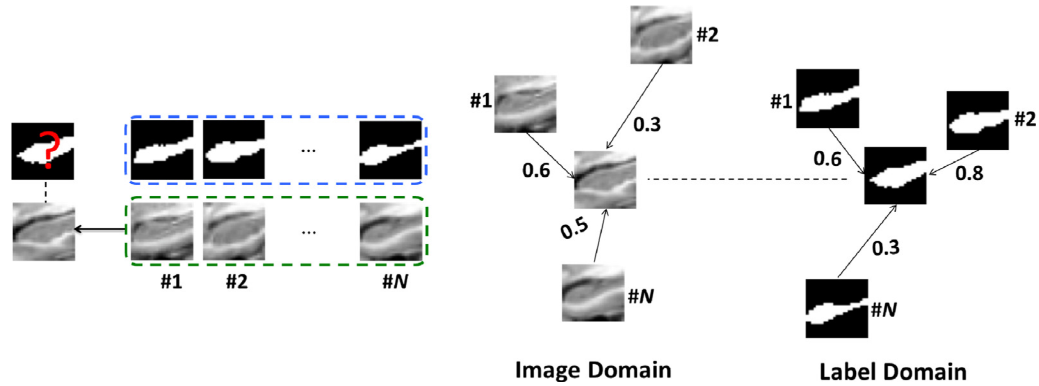

Fig. 1. Demonstration of the significant gap of representation profiles computed in the image domain using image appearance and in the label domain using label information.

(Artaechevarria et al., 2009; Song et al., 2015; Wang et al., 2011; Yan et al., 2013; Zhang et al., 2011). In these methods, the main assumption is that, if two image patches have similar appearance, they should bear the same anatomical label. The typical patch-based label fusion methods include *nonlocal patch-based labeling* (NPBL) (Coupé et al., 2011; Rousseau et al., 2011) and *sparse patch-based labeling* (SPBL) (Tong et al., 2013; Zhang et al., 2012a, 2012c). Note that all patch-based label fusion methods collect candidate atlas patches in a search neighborhood across all registered atlas images. The weights used for label fusion in NPBL are proportional to the decayed patch-wise similarities penalized by the exponential function. Inspired by the discriminative power of sparse representation (Tibshirani, 2011; Zhang et al., 2012b, 2012a), the SPBL method has been proposed to introduce sparsity into the optimization of the weighting vector at each image point. Since the sparsity constraint enforces many zero elements in the weighting vector, the SPBL method can reduce the risk of taking incorrect or ambiguous patches and finally use only a small number of well-matched patches for labeling. More advanced methods can be found in Wang et al. (2013) and Wu et al. (2014), where the pairwise dependency between atlas patches is further modeled to avoid the repeating label fusion error by similar atlas patches.

In all of the above state-of-the-art methods, the weights are exclusively optimized in terms of patch-wise image appearance. The computed weighting vector is regarded as an *appearance representation profile* and then directly used to determine the (binary) labels for the target image. Despite its simplicity and effectiveness, there is no evidence showing that such weights are domain-invariant, i.e., the optimized weights derived from the best image patch presentation may not be necessarily optimal for label fusion. Fig. 1 demonstrates the significant gap of representation profiles estimated in the *image domain* using appearance information (left) and in the *label domain* using label information (which is assumed to be known for the unseen target image). It is clear that there is no guarantee for the current state-of-the-art label fusion methods to achieve the optimal labeling results by directly applying the *appearance-based* representation profile for label fusion.

To address this issue, we propose a novel label propagation framework to progressively convert the *representation profile* from the image domain to the *optimal weighting vector* (for label fusion) in the label domain by constructing a set of intermediate dictionaries to bridge the image domain and the label domain. Such intermediate dictionaries provide a sequence of guidance to steer the estimation of the *appearance representation profile* to the *optimal weighting vector* for label fusion.

Specifically, in the training stage for each target image patch, the initial-layer dictionary consists of the original atlas image patches (in the image domain), similar to the most of traditional label fusion methods. Since each atlas image patch has its corresponding label patch, it is straightforward to build the label patch dictionary (in the label domain) by arranging the corresponding label patches with the same order as the original atlas image patches in the initial-layer dictionary. To remedy the large transitions from the image domain to the label domain, we first apply a label fusion technique (e.g., NPBL or SPBL) to obtain the *representation profile* for each atlas image patch in the initial-layer dictionary, while regarding all other instances in the initial-layer dictionary as the atlas image patches. Then, we compute a *label probability patch* by applying the obtained *representation profile* to the respective atlas label patches. By repeating the above leave-one-out label fusion procedure to all the patches in the initial-layer dictionary, we can construct the first-layer intermediate dictionary. Similarly, we can construct the subsequent intermediate dictionaries, as shown in Fig. 2. In the end, we can construct a sequence of intermediate dictionaries, where the label probability patches become sharper and sharper, close to the binary shape of the corresponding atlas label patches.

In the testing stage, given the learned multi-layer dictionary at each target image location, the final weights for voting the label are also estimated in a progressive way. Specifically, starting from the initial layer, we gradually refine the label fusion weights by alternating the following two steps. *First*, we compute the *representation profile* of the target image patch by using the patch dictionary in the current layer. *Second*, we refine the label probability map within the target image patch by applying the latest *representation profile* to the binary atlas label patches, and then use the obtained new probability patch as the new target image in the next layer of the label estimation. In this way, we can obtain more and more accurate weights to determine the anatomical label for the original target image, with the guidance of the intermediate dictionary at each layer.

The contributions of our proposed method include: (1) since we harness the multi-layer dictionary to remedy the gap between patch appearances and anatomical labels, our label fusion essentially seeks the best label fusion weights, instead of just patch-wise representation; (2) the sequence of built intermediate dictionaries allows us using *not only* appearance features *but also* structural context information (Tu and Bai, 2010) to significantly improve the robustness in patch representation; (3) our proposed progressive patch representation by a multi-layer dictionary is general enough to be integrated with many conventional patch-based segmentation methods for improving their performances. Our proposed method has been evaluated in the segmentation of the hippocampus from elderly brain MR images in the ADNI dataset. More accurate segmentation results have been achieved, compared to the state-of-the-art methods, i.e., NPBL and SPBL.