NeuroImage 106 (2015) 34–46

Contents lists available at ScienceDirect

NeuroImage

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

Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition☆

Guorong Wu [a], Minjeong Kim [a], Gerard Sanroma [a], Qian Wang [b], Brent C. Munsell [c], Dinggang Shen [a,d,*], The Alzheimer's Disease Neuroimaging Initiative

[a] BRIC and Department of Radiology, University of NC, Chapel Hill, USA
[b] Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China
[c] Computer Science Department, College of Charleston, Charleston, SC 29424, USA
[d] Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

A R T I C L E I N F O

Article history:
Received 6 May 2014
Revised 5 November 2014
Accepted 12 November 2014
Available online 20 November 2014

Keywords:
Patch-based labeling
Multi-atlas based segmentation
Multi-scale feature representation
Label-specific patch partition
Sparse representation

A B S T R A C T

Multi-atlas patch-based label fusion methods have been successfully used to improve segmentation accuracy in many important medical image analysis applications. In general, to achieve label fusion a single target image is first registered to several atlas images. After registration a label is assigned to each target point in the target image by determining the similarity between the underlying target image patch (centered at the target point) and the aligned image patch in each atlas image. To achieve the highest level of accuracy during the label fusion process it's critical for the chosen patch similarity measurement to accurately capture the tissue/shape appearance of the anatomical structure. One major limitation of existing state-of-the-art label fusion methods is that they often apply a fixed size image patch throughout the entire label fusion procedure. Doing so may severely affect the fidelity of the patch similarity measurement, which in turn may not adequately capture complex tissue appearance patterns expressed by the anatomical structure. To address this limitation, we advance state-of-the-art by adding three new label fusion contributions: First, each image patch is now characterized by a multi-scale feature representation that encodes both local and semi-local image information. Doing so will increase the accuracy of the patch-based similarity measurement. Second, to limit the possibility of the patch-based similarity measurement being wrongly guided by the presence of multiple anatomical structures in the same image patch, each atlas image patch is further partitioned into a set of label-specific partial image patches according to the existing labels. Since image information has now been semantically divided into different patterns, these new label-specific atlas patches make the label fusion process more specific and flexible. Lastly, in order to correct target points that are mislabeled during label fusion, a hierarchical approach is used to improve the label fusion results. In particular, a coarse-to-fine iterative label fusion approach is used that gradually reduces the patch size. To evaluate the accuracy of our label fusion approach, the proposed method was used to segment the hippocampus in the ADNI dataset and 7.0 T MR images, sub-cortical regions in LONI LBPA40 dataset, mid-brain regions in SATA dataset from MICCAI 2013 segmentation challenge, and a set of key internal gray matter structures in IXI dataset. In all experiments, the segmentation results of the proposed hierarchical label fusion method with multi-scale feature representations and label-specific atlas patches are more accurate than several well-known state-of-the-art label fusion methods.

Introduction

Many medical image analysis studies require an accurate segmentation of anatomical structures in order to measure structural differences across individuals or between groups (Aljabar et al., 2009; Hsu et al., 2002). For example, in connectome applications multiple brain regions, in hundreds of brain MR images, need to be automatically identified before constructing a brain connectivity network (Liu et al., 2012; Liu and Ye, 2010) that describes network architecture of the human brain. Therefore, to improve segmentation accuracy the development of automatic ROI (region of interest) labeling methods has seen increased

☆ Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). As such, the investigators with the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

* Corresponding author at: BRIC and Department of Radiology, University of North Carolina, Chapel Hill, USA. Fax: +1 919 843 2641.
E-mail address: dgshen@med.unc.edu (D. Shen).

http://dx.doi.org/10.1016/j.neuroimage.2014.11.025
1053-8119/

attention in the medical imaging field over the last several years (Aljabar et al., 2009; Coupé et al., 2011; Rousseau et al., 2011; Tong et al., 2012; Wang et al., 2011a; Wang et al., 2011b; Warfield et al., 2004; Wu et al., 2014).

Multiple atlases with manually identified labels have proven to be very useful when used to detect and label ROIs in the target image that may show high structural variations in the population. The basic assumption behind multi-atlas based segmentation is that the target image point should bear the same label as the atlas image point if the local tissue shape or appearance is very similar. All atlas images are required to be registered to a target image before label fusion. To alleviate possible registration errors, patch-based label fusion (Coupé et al., 2011; Rousseau et al., 2011) seeks multiple correspondence candidates using patchwise similarity measurements between the target image patch and the atlas image patches within a certain voxel neighborhood. Intuitively, if the calculated similarity measurement between a target image patch and a particular atlas image patch is very high, then the atlas label assigned to the target point is the correct one.

To accurately assess image patch similarity, the identification and selection of ideal image patches are key components of patch-based label fusion methods. Most state-of-the-art methods simply use fixed size patches throughout the entire label fusion procedure. For example, $7 \times 7 \times 7$ or $9 \times 9 \times 9$ cubic patches are usually used in the literature (Coupé et al., 2011; Rousseau et al., 2011; Tong et al., 2012; Wang et al., 2011a). In order to make the label fusion robust to noise, image patches are required to be sufficiently large enough to capture the intended image content. However, using a large image patch may create additional problems when labeling small anatomical structures, e.g. the patchwise similarity measurement could be dominated by other larger anatomical structures surrounding the smaller one in the image patch. In short, methods that use fixed-size patches lack discriminative power to characterize complex appearance patterns in the medical imaging data.

During the last decade, many efforts have been made to improve the discrimination ability of image patches during label fusion. For instance, sparse dictionary learning is used in Tong et al. (2013) to find the best feature representations prior to label fusion. Moreover, in Wang et al. (2011a) and Wu et al. (2014) dependencies among atlas image patches have been investigated to improve labeling accuracy by iteratively inspecting incorrectly labeled patches that show similar labeling error patterns. However, these state-of-the-art approaches use patches with fixed size and therefore still suffer from this limitation.

In this paper, we address the above limitations by developing hierarchical and high-level feature representations to adequately describe image patches. We propose the following three contributions: *First*, a layer-wise multi-scale feature representation adaptively encodes image features at different scales for each image point in the image patch. In the proposed approach, feature representations near the center of the patch provide more detailed (fine-scale) shape or appearance information, whereas feature representations near the edge of the patch provide less detailed (coarse-scale) shape or appearance information. *Second*, it's very common that the structure to be segmented, e.g. the hippocampus, is surrounded by other anatomical structures in the image patch. In such cases it becomes very difficult to correctly recognize the intended structure from the surrounding ones and mislabeling is likely to occur. In computer vision, object recognition algorithms address this limitation by attempting to separate the foreground pattern from background clutter (Li et al., 2010). In light of this research, a novel label-specific patch partition technique is proposed that splits each atlas patch into a set of new complementary label-specific (or structure-specific) image patches. To handle the increased number of label-specific image patches after the proposed patch splitting strategy a group sparsity constraint is included. As a result, the discriminative power of each label-specific image patch is enhanced because it only contains the image information of the corresponding anatomical structure. To the best of our knowledge, this type of representation is rarely exploited in label fusion. *Third*, because existing label fusion methods typically use a fixed patch size, and label the entire target image in one pass, they are not given a chance to correct possible errors. To overcome this limitation the proposed method uses an iterative label-fusion procedure. Specifically, larger image patches are used in the beginning to increase the search range, however at each iteration the labeling result is evaluated and the size of the image patch is gradually reduced. To ensure that spurious artifacts do not dominate the proposed label-fusion method, a sparsity constraint is included that only allows a small number of atlas patches to participate in the label fusion process.

It should be noted that this paper is an extension of our previous work in Wu and Shen (2014). However, there are several differences, specifically: a group sparsity constraint is used instead of a weighting vector sparsity constraint, a more comprehensive validation of each contribution (i.e., multi-scale feature representation, label-specific patch partition, and iterative label fusion), and additional datasets are used to evaluate the performance of the proposed label fusion method.

Performance of the proposed label fusion method is compared to existing state-of-the-art patch-based labeling methods (Coupé et al., 2011; Rousseau et al., 2011) using several different datasets. Specifically, the datasets used to evaluate the proposed method are the MICCAI 2013 segmentation challenge dataset (Landman and Warfield, 2012) with 14 manually labeled ROIs in the mid-brain, the LONI LBPA40 dataset (Shattuck et al., 2008) with 54 manually labeled ROIs at subcortical regions, and the IXI dataset with 83 manually labeled ROIs (Hammers et al., 2003; Hammers et al., 2007). Finally, we also include hippocampus segmentation experiments using the ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset and 7.0 T MR images (Cho et al., 2010). For each dataset the proposed method achieves a more accurate labeling result.

The remainder of the paper is organized as follows: In the Method section we present our novel generative probability model for label fusion, in the Experiments section we evaluate its performance by comparing it with conventional patch-based methods, and in the Discussion section we provide a brief conclusion.

Method

Given the target image $\boldsymbol{T}$, the goal of label fusion is to automatically determine the label map $\boldsymbol{L_T}$ for the target image. We first register each atlas image, as well as the label maps, onto the target image space. We use $\boldsymbol{I} = \{\boldsymbol{I_s} | \boldsymbol{s} = 1, \ldots, \boldsymbol{N}\}$ and $\boldsymbol{L} = \{\boldsymbol{L_s} | \boldsymbol{s} = 1, \ldots, \boldsymbol{N}\}$ to denote the $\boldsymbol{N}$ registered atlases and label maps, respectively. For each target image point $\boldsymbol{x}$ ($\boldsymbol{x} \in \boldsymbol{T}$), all the atlas patches[*] within a certain search neighborhood $\boldsymbol{n}(\boldsymbol{x})$, denoted as $\vec{\boldsymbol{\beta}}_{\boldsymbol{s},\boldsymbol{y}}$ ($\vec{\boldsymbol{\beta}}_{\boldsymbol{s},\boldsymbol{y}} \subset \boldsymbol{I_s}, \boldsymbol{y} \in \boldsymbol{n}(\boldsymbol{x})$), are used to compute the patchwise similarities w.r.t. the target image patch $\vec{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{x}}$ ($\vec{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{x}} \subset \boldsymbol{T}$). We arrange each patch, $\vec{\boldsymbol{\beta}}_{\boldsymbol{s},\boldsymbol{y}}$ and $\vec{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{x}}$, into a column vector. We use the tuple $\boldsymbol{b} = (\boldsymbol{s}, \boldsymbol{y})$ to denote both the atlas image index $\boldsymbol{s}$ and the location of the patch center point $\boldsymbol{y}$, respectively. Thus, each atlas image patch $\vec{\boldsymbol{\beta}}_{\boldsymbol{s},\boldsymbol{y}}$ can now be simplified to $\boldsymbol{\beta}_{\boldsymbol{b}}$ ($\boldsymbol{b} = 1, \ldots, \boldsymbol{Q}$), where $\boldsymbol{Q} = \boldsymbol{N} \times |\boldsymbol{n}(\boldsymbol{x})|$ is the total number of atlas image patches which are used to label the center point of the target image patch $\vec{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{x}}$. For clarity, we use only $\vec{\boldsymbol{\alpha}}$ to denote the underlying target image patch by dropping off the subscripts in $\vec{\boldsymbol{\alpha}}_{\boldsymbol{T},\boldsymbol{x}}$.

Label fusion methods such as non-local averaging (Coupé et al., 2011; Rousseau et al., 2011), can be used to calculate the weighting vector $\vec{w} = [w_b]_{b=1,\ldots,Q}$ for all atlas patches, each of which is denoted by $\vec{\beta}_b$. As we will explain in the Label-specific Atlas Patch Partition section, we adopt the sparsity constraint (Liu et al., 2009a,b; Tibshirani, 1996) in our method by regarding the label fusion procedure as the problem of

[*] Some label fusion methods use patch pre-selection to discard the less similar patches.