Attribute Similarity and Mutual-Saliency Weighting for Registration and Label Fusion

Yangming Ou, Jimit Doshi, Guray Erus, and Christos Davatzikos

Section of Biomedical Image Analysis (SBIA)
Department of Radiology, University of Pennsylvania

Abstract. Multi-atlas segmentation relies on two major components: image registration to propagate segmentation labels, and label fusion to combine multiple labels into one at each voxel. In this paper, we propose to drive both components by an attribute-based similarity metric and a mutual-saliency-based reliability metric. The fundamental idea is to improve registration and label fusion by looking for corresponding voxels that are similar (as measured by their Gabor attributes), and more importantly, reliably similar (as measured by the mutual-saliency of their matching) between atlas and target images. We apply this pipeline to segment 140 structures in brain MRI of 15 training subjects and 20 testing subjects in MICCAI Challenge 2012.

1 Introduction

Multi-atlas segmentation has gained increasing interest in recent years [1–5]. One premise in this approach is that it allows using a priori knowledge, as encoded in atlas segmentations, to infer segmentation in target image via atlas-to-target image registration. Another premise is that it allows different atlases to correct each other’s errors in a process often known as label fusion. The fused segmentation has shown remarkable improvement over single-atlas-based segmentation in various brain, cardiac and prostate structures.

Despite exciting research in recent years, both image registration and label fusion are not without challenges. In registration, a fundamental question is how to find reliable correspondences across images. This is especially important when there exist considerable structural difference between atlas and target images.

In label fusion, recent studies have obtained improved accuracy by assigning higher weights to atlases that are more similar to target at local voxel level [2, 3]. But a fundamental question is how to properly measure similarity between atlas and target at voxel level. Researchers have used correlation or intensity difference to imply voxel similarities [2, 3]. Ideally there can be a more robust similarity measure incorporating richer geometric context of each voxel. In addition, we hypothesize that a proper reliability measure (i.e., whether an atlas voxel and a target voxel are reliably matched) will further improve label fusion accuracy too. Here matching between two voxels are said reliable if they are similar to each other and meanwhile not similar to anything else in the neighborhood [6]. This is a higher level of confidence in the matching and label inheritance.
In this paper, we propose to improve both registration and label fusion by attribute-based similarity and matching reliability metrics. The idea is the following. When registering atlas to target, we rely more on those regions, compared to other regions, that can establish more reliable matching. When fusing labels, we assign higher confidence/weight to those atlases, compared to other atlases, that are more reliably similar to the target at each voxel. All experiments are done using 15 training brain MR images and 20 testing MR images in the MICCAI 2012 Multi-Atlas Segmentation Challenge.

2 Methods

In this section, we first introduce attribute-based similarity metric and mutual-saliency-based matching-reliability metric (Sec. 2.1). Then we describe their use in guiding registration (Sec. 2.2) and guiding label fusion (Sec. 2.3).

2.1 Definition of Attribute Similarity and Mutual-Saliency

These two concepts were proposed in our recent paper [6]. For the completeness of this paper, below is a brief description. First of all, we represent each voxel \( x \) by geometric context around this voxel, in a \( d \)-dimensional multi-scale and multi-orientation Gabor attribute vector \( A(x) \). This attribute representation has rendered each voxel more distinctive than intensity information alone [6]. Then, we shall say that two voxels \( x \) and \( y \) in two images are similar, if we observe small difference in their attribute representations, i.e., \( \text{sim}(x, y) = \frac{1}{1+\frac{1}{d}\|A(x)-A(y)\|^2} \).

A pair of voxels \( x, y \) in two images is said mutually-salient, if they are similar to each other and meanwhile less similar to any other voxels in the neighborhood. As shown in Fig. 1, similarity map between \( x \) and all voxels in the vicinity of \( y \) exhibit a delta-shape distribution peaking at \( y \). What this means is that the matching between those two voxels are reliable, because no other voxel in the neighborhood of \( y \) can replace it with same high similarity to \( x \).

Mathematically, mutual-saliency, \( \text{ms}(x, y) \), is approximated by dividing the mean similarity between voxel \( x \) and all voxels in the core neighborhood (CN) of \( y \), by the mean similarity between voxel \( x \) and all voxels in the peripheral neighborhood (PN) of \( y \), where CN and PN are defined in accordance to the scale where Gabor attributes are extracted (see [6] for more details), i.e.,

\[
\text{ms}(x, y) = \frac{1}{|\text{CN}(y)|} \sum_{w \in \text{CN}(y)} \text{sim}(x, w) \frac{|\text{PN}(y)|}{|\text{CN}(y)|} \sum_{w \in \text{PN}(y)} \text{sim}(x, w).
\]
Fig. 2 shows a typical set of similarity and mutual-saliency maps from an atlas-to-target registration. Matching in cortical regions observes lower similarity than matching in deep brain structures (as shown in similarity map), and lower reliability (as shown in mutual-saliency map). Contrary is the matching in ventricle and peri-ventricle white matter regions. So we would have more confidence in following the warped segmentation labels in latter regions.

Fig. 2. A typical set of similarity map and mutual-saliency map resulted from registration from an atlas to the target image.

2.2 Registration

The above defined similarity and mutual-saliency are used to modulate registration, as implemented in the DRAMMS software [6]. Specifically, DRAMMS seeks a non-rigid transformation $T$, based on free form deformation (FFD) model [7], that minimizes the mutual-saliency-weighted attribute differences over target image domain $\Omega \subset \mathbb{R}^3$, 

$$
\text{arg max}_T \text{ Energy}(T) = \int_{u \in \Omega} \text{ms}(T^{-1}(u), u) \cdot \frac{1}{2} \| A(T^{-1}(u)) - A(u) \|^2 \text{d}u \quad (1)
$$

In essence, voxels are matched by their geometric context other than intensity. And, the whole registration is mainly driven by regions/voxels that can reliably match across images.

2.3 Label Fusion

DRAMMS registration maps segmentation regions from different atlases into the same target image. Now we need label fusion to combine those multiple labels at each voxel. Assuming that $N$ atlases, indexed by $n$, have been each registered to the same target image via a deformation $T_n$. A voxel $u$ in the target image space $\Omega$ will tentatively have $N$ segmentation labels propagated from all those $N$ atlases, denoted as $\{\text{label}(T_n^{-1}(u))\}_{n=1}^N$. To fuse them into a single segmentation label, we use a similarity and mutual-saliency weighted voting strategy. Specifically, we first calculate the probability of this voxel having each of all $L$ segmentation labels $\{1, 2, \ldots, L\}$, i.e., $\forall l \in 1, 2, \ldots, L$

$$
\Pr(\text{label}(u) = l) = \frac{\sum_{n} \text{sim}(T_n^{-1}(u), u) \cdot \text{ms}(T_n^{-1}(u), u) \cdot 1(\text{label}(T_n^{-1}(u)) = l)}{\sum_{n} \text{sim}(T_n^{-1}(u), u) \cdot \text{ms}(T_n^{-1}(u), u)} \quad (2)
$$

Then, we assign the most likely label $l^*$ to this voxel $u$, i.e., $\text{label}(u) = l^*$, such that $l^* = \text{arg max}_l \Pr(\text{label}(u) = l)$. In extreme cases, if $\text{sim}(\cdot, \cdot) \equiv 1$ and $\text{ms}(\cdot, \cdot) \equiv 1$, we end up with the classic majority voting, as all atlases are equally trusted at each voxel.
3 Results

Fig. 3 shows leave-one-out results in training dataset (15 subjects from OASIS dataset). We compared the proposed \((\text{sim} \times \text{ms})\)-weighted voting mechanism with classic majority voting for label fusion. We have several observations:

1) \((\text{sim} \times \text{ms})\)-weighted voting always improves majority voting.

2) Brain ROI segmentation accuracy is sensitive to initial skull-stripping. A perfect skull-stripping (using ground-truth) helps improve segmenting brain structures. When ground-truth skull-stripping is not used, segmentation accuracy is similar between no skull-stripping and automatic skull-stripping (multi-atlas segmentation with DRAMMS registration and majority voting).

3) Accuracy is also a bit sensitive to how the Dice scores in all 140 brain ROIs are averaged. The ROI-volume-weighted average Dice score (in left figure) is less sensitive to skull-stripping accuracy than the direct average Dice score.

Fig. 3. Leave-one-out results in training dataset (15 subjects).

References


