



Non-local statistical label fusion for multi-atlas segmentation

Andrew J. Asman^{a,*}, Bennett A. Landman^{a,b,c}

^aElectrical Engineering, Vanderbilt University, Nashville, TN 37235, USA

^bBiomedical Engineering, Vanderbilt University, Nashville, TN 37235, USA

^cRadiology and Radiological Sciences, Vanderbilt University, Nashville, TN 37235, USA

ARTICLE INFO

Article history:

Received 7 May 2012

Received in revised form 19 October 2012

Accepted 29 October 2012

Available online 29 November 2012

Keywords:

Multi-atlas segmentation

Label fusion

STAPLE

Non-local means

Rater models

ABSTRACT

Multi-atlas segmentation provides a general purpose, fully-automated approach for transferring spatial information from an existing dataset (“atlases”) to a previously unseen context (“target”) through image registration. The method to resolve voxelwise label conflicts between the registered atlases (“label fusion”) has a substantial impact on segmentation quality. Ideally, statistical fusion algorithms (e.g., STAPLE) would result in accurate segmentations as they provide a framework to elegantly integrate models of rater performance. The accuracy of statistical fusion hinges upon accurately modeling the underlying process of how raters err. Despite success on human raters, current approaches inaccurately model multi-atlas behavior as they fail to seamlessly incorporate exogenous intensity information into the estimation process. As a result, locally weighted voting algorithms represent the *de facto* standard fusion approach in clinical applications. Moreover, regardless of the approach, fusion algorithms are generally dependent upon large atlas sets and highly accurate registration as they implicitly assume that the registered atlases form a collectively unbiased representation of the target. Herein, we propose a novel statistical fusion algorithm, Non-Local STAPLE (NLS). NLS reformulates the STAPLE framework from a non-local means perspective in order to learn what label an atlas would have observed, given perfect correspondence. Through this reformulation, NLS (1) seamlessly integrates intensity into the estimation process, (2) provides a theoretically consistent model of multi-atlas observation error, and (3) largely diminishes the need for large atlas sets and very high-quality registrations. We assess the sensitivity and optimality of the approach and demonstrate significant improvement in two empirical multi-atlas experiments.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Segmentation of anatomical structures on medical images is essential for scientific inquiry into the complex relationships between biological structure and function as well as clinical diagnosis, treatment, and assessment. The long-held “gold standard” for highly robust segmentation has been through expert manual delineation (Crespo-Facorro et al., 1999; Tsang et al., 2008). Yet, manual delineation is extremely resource consuming and plagued by inter- and intra-rater variability (e.g., 10–20% by volume (Ashton et al., 2003; Joe et al., 1999)). Alternatively, fully-automated algorithms often result in robust and accurate estimations for specific classes of problems (e.g., brain-tissue classification (Cocosco et al., 2003; Van Leemput et al., 1999; Wells III et al., 1996), optic nerve segmentation (Noble and Dawant, 2011)). Unfortunately, the success of automated techniques is often dependent upon the application,

modality, and image quality (Fischl et al., 2002; Heckemann et al., 2006; Rohlfing et al., 2004a; Yeo et al., 2008).

Atlas-based segmentation methods form a middle-ground between fully-manual and fully-automatic segmentation approaches (Collins et al., 1995; Gee et al., 1993). In atlas-based models, spatial information is transferred from an existing dataset (labeled atlas) to a previously unseen context (target) through deformable registration. Proposed extensions enable the summary of multiple atlases into a common coordinate system by constructing (1) unbiased average atlases (Guimond et al., 2000; Joshi et al., 2004) and (2) target-specific atlases (Commowick et al., 2009; Ericsson et al., 2008). Yet, the accuracy of single-atlas based methods is limited due to the bias concerns and lack of correspondence to the target (Ashburner and Friston, 2005; Han and Fischl, 2007). Thus, an alternative strategy that independently utilizes multiple atlases (i.e., multi-atlas segmentation) has come to represent the *de facto* standard baseline for atlas techniques. In multi-atlas segmentation (Heckemann et al., 2006; Rohlfing et al., 2004b), multiple atlases are separately registered to the target and the voxelwise label conflicts between the registered atlases are resolved using label fusion.

* Corresponding author. Address: Vanderbilt University EECS, 2301 Vanderbilt Pl., PO Box 351679, Station B, Nashville, TN 37235-1679, USA. Tel.: +1 (615) 322 2338.

E-mail address: andrew.j.asman@vanderbilt.edu (A.J. Asman).

Perhaps surprisingly, a majority vote, the simplest fusion strategy, has been shown to result in highly robust segmentations (Aljabar et al., 2009; Heckemann et al., 2006; Rohlfing et al., 2004a; Rohlfing and Maurer, 2007). More recently, weighted voting strategies that use global (Artaechevarria et al., 2009; Chen et al., 2012), local (Isgum et al., 2009; Sabuncu et al., 2010; Wang et al., 2011), semi-local (Sabuncu et al., 2010; Wang et al., 2012), and non-local (Coupé et al., 2011) intensity similarity metrics have demonstrated consistent improvement in segmentation accuracy. Particularly for neurological applications, highly local weights have provided the most consistent results in segmentation quality (Artaechevarria et al., 2009; Sabuncu et al., 2010).

In contrast to *ad hoc* voting, statistical fusion strategies (e.g., Simultaneous Truth and Performance Level Estimation, STAPLE (Warfield et al., 2004)) directly integrate a stochastic model of rater behavior into the estimation process. Despite elegant theory and success on human raters, applications to the multi-atlas context have proven problematic (Asman and Landman, 2011a; Sabuncu et al., 2010; Wang et al., 2011, 2012). In response, a myriad of advancements to the STAPLE framework have been proposed to account for (1) spatially varying task difficulty (Asman and Landman, 2011b; Rohlfing et al., 2004b), (2) spatially varying rater performance (Asman and Landman, 2011a, 2012a; Commowick et al., 2012; Weisenfeld and Warfield, 2011), and (3) instabilities in the rater performance level parameters (Commowick and Warfield, 2010; Landman et al., 2011b). Yet, these advanced techniques remain inherently models of human observation error as they fail to directly incorporate the image intensity differences between the atlases and the target. Moreover, initial attempts to incorporate intensity into the STAPLE framework have relied upon *ad hoc* extensions that simply ignore voxels based upon *a priori* similarity measures (Cardoso et al., 2011; Weisenfeld and Warfield, 2011).

Regardless of the approach, label fusion models have consistently made an implicit assumption that the use of multiple atlases results in a voxelwise, collectively unbiased representation of the target. This assumption is manifested through the fact that nearly all fusion algorithms determine the optimal label using only *directly corresponding* intensity and label information. Ergo, multi-atlas methods are generally dependent upon highly accurate registration and the use of large numbers of atlases. We are left with several problems in multi-atlas segmentation: (1) a dependence on large-scale, high-quality registrations, (2) voting-based algorithms lack the theoretical underpinning of statistical fusion observation models and (3) statistical fusion algorithms fail to incorporate intensity information. Thus, previous approaches have failed to accurately model the stochastic process of registered atlas observation error.

Meanwhile, a relatively new framework in the field of image analysis, non-local means, has gained momentum in terms of quantifying complex image characteristics (e.g., noise structure, spatially varying correspondence). In non-local means, images are deconstructed into a collection of small volumetric patches and the similarity or correspondence between these patches is quantified to learn the underlying image structure (Buades et al., 2005). The non-local means framework has emerged in the context of image de-noising (Buades et al., 2005; Coupé et al., 2006; Kervrann et al., 2007; Liu et al., 2008; Manjón et al., 2008; Van De Ville and Kocher, 2009). However, more recent work has demonstrated the applicability of non-local means to new applications such as synthesizing image contrast (Roy et al., 2010a), in-painting (Sun and Tappen, 2011), and image segmentation (Coupé et al., 2011; Roy et al., 2010b).

Herein, we propose a novel statistical fusion algorithm (Non-Local STAPLE – NLS) that reformulates the STAPLE framework from a non-local means perspective. NLS models the registered atlases as collections of volumetric patches containing both intensity and

label information and uses the non-local criteria (Buades et al., 2005; Coupé et al., 2011) to resolve imperfect correspondence. Through this reformulation, we seamlessly integrate exogenous intensity information into the estimation process to provide a theoretically consistent model of multi-atlas observation error. NLS provides a model in which we learn which label each atlas *would have observed* given perfect correspondence with the target. This presentation is an extension and generalization of a recently published conference paper (Asman and Landman, 2012b). Herein, we provide additional examples, derivations and insights that were not part of the original conference publication.

In this manuscript, we begin by deriving the theoretical basis and the parameters for initialization and convergence governing NLS. Next, we demonstrate significant improvement over the state-of-the-art fusion algorithms on two distinct datasets: (1) computed tomography (CT) images for thyroid segmentation and (2) structural magnetic resonance (MR) images for whole-brain segmentation. For whole-brain segmentation, we demonstrate that NLS dramatically lessens the need for large-scale and highly accurate non-rigid registration. Lastly, we provide insight into the sensitivity of NLS to the various model parameters, assess the optimality of the algorithm, and provide a comparison to a direct application of non-local voting.

2. Theory

The following presentation provides the theoretical model governing NLS in the commonly used Expectation–Maximization (EM) framework (Dempster et al., 1977). For clarity and consistency, the notation closely follows the presentation of the original STAPLE algorithm (Warfield et al., 2004).

2.1. Problem definition

Consider a target gray-level image represented as a vector, $I \in \mathbb{R}^{N \times 1}$. Let $T \in L^{N \times 1}$ be the latent representation of the true target segmentation, where $L = \{0, \dots, L-1\}$ is the set of possible labels that can be assigned to a given voxel. Consider a collection of R registered atlases with associated intensity values, $A \in \mathbb{R}^{N \times R}$, and label decisions, $D \in L^{N \times R}$. Let $\theta \in \mathbb{R}^{R \times L \times L}$ parameterize the performance level of raters (registered atlases). Each element of θ , $\theta_{j's's}$, represents the probability that rater j observes label s' given that the true label is s at a given target voxel and the *corresponding* voxel on the associated atlas—i.e., $\theta_{j's's} \equiv f(D_{i'j} = s', A_j | T_i = s, I_i, \theta_{j's's})$, where i^* is the voxel on atlas j that corresponds to target voxel i . Throughout, the index variables i , i^* and i' will be used to iterate over the voxels, s and s' over the labels, and j over the registered atlases.

2.2. The non-local STAPLE algorithm

As with other statistical fusion algorithms, NLS uses EM to estimate the true (latent) segmentation based on the target intensities, atlas information, and the rater performance level parameters (see Fig. 1 for a graphical summary of NLS). In traditional EM terminology, the underlying voxelwise label probabilities represent the hidden data that we are estimating, and the performance level parameters, θ , represent the hidden model parameters that help determine the optimal solution for the target segmentation. The estimation of these parameters is accomplished by iterating between the E-step (i.e., the estimation of the voxelwise label probabilities) and the M-step (i.e., the estimation of the performance level parameters that maximize the expected value of the conditional log likelihood function). Before presenting the derivation of our EM-based approach, we define our non-local correspondence

Download English Version:

<https://daneshyari.com/en/article/10337894>

Download Persian Version:

<https://daneshyari.com/article/10337894>

[Daneshyari.com](https://daneshyari.com)