Artificial Intelligence in Medicine 64 (2015) 117–129

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

A label fusion method using conditional random fields with higher-order potentials: Application to hippocampal segmentation

Carlos Platero*, M. Carmen Tobar

Health Science Technology Group, Technical University of Madrid, Ronda de Valencia 3, 28012 Madrid, Spain

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 21 January 2015
Accepted 26 April 2015

Keywords:
Atlas-based segmentation
Image registration
Label fusion
Graph cuts
Global optimization
Hippocampal segmentation
Magnetic resonance imaging

ABSTRACT

Objective: The objective of this study is to develop a probabilistic modeling framework for segmenting structures of interest from a collection of atlases. We present a label fusion method that is based on minimizing an energy function using graph-cut techniques.
Methods and materials: We use a conditional random field (CRF) model that allows us to efficiently incorporate shape, appearance and context information. This model is characterized by a pseudo-Boolean function defined on unary, pairwise and higher-order potentials. Given a subset of registered atlases in the target image for a particular region of interest (ROI), we first derive an appearance-shape model from these registered atlases. The unary potentials combine an appearance model based on multiple features with a label prior using a weighted voting method. The pairwise terms are defined from a Finsler metric that minimizes the surface of separation between voxels whose labels are different. The higher-order potentials used in our framework are based on the robust P^n model proposed by Kohli et al. The higher-order potentials enforce label consistency in cliques; hence, the proposed method can be viewed as an approach to integrate high-level information with images based on low-level features. To evaluate the performance and the robustness of the proposed label fusion method, we employ two available databases of T1-weighted (T1W) magnetic resonance (MR) images of human brains. We compare our approach with other label fusion methods in the automatic hippocampal segmentation from T1W-MR images.
Results: Our label fusion method yields mean Dice coefficients of 0.829 and 0.790 for the two databases used with mean times of approximately 80 and 160 s, respectively.
Conclusions: We introduce a new label fusion method based on a CRF model and on ROIs. The CRF model is characterized by a pseudo-Boolean function defined on unary, pairwise and higher-order potentials. The proposed Boolean function is representable by graphs. A globally optimal binary labeling is found using a st-mincut algorithm in each ROI. We show that the proposed approach is very competitive with respect to recently reported methods.

1. Introduction

The automatic segmentation of subcortical structures in human brain magnetic resonance (MR) images plays a crucial role in clinical practice. The extraction of biomarkers from MR images is directed at variations in subcortical shapes and their volume measurements [1,2]. This task is very important but difficult to perform, even by hand. Neuroanatomists often develop and use complicated protocols in guiding the manual delineation process [3,4]. Specifically, hippocampal segmentation is an important tool for studying neurodegenerative diseases. Hippocampal volume and shape measures are commonly used as biomarkers for Alzheimer's disease, epilepsy and schizophrenia, among other [5]. This structure is difficult to segment because of its small size, high variability and low contrast in MR images.

Many approaches have been proposed, and most segmentation methods use deformable surfaces or atlas-based techniques. Deformable methods tend to explicitly use learned shape variation as a priori information in segmentation. Various methods for representing shapes and their relationships with the appearances have been proposed, such as region growing [6], level-set within a Bayesian framework [7], probabilistic boosting tree [8] or using active shape-appearance models [9]. Atlas-based segmentation has become a standard technique for identifying structures from brain MR images. The atlas-based methods have been demonstrated to outperform other algorithms [10]. The atlas-based approaches are generally based on non-rigid registrations. In the context of this study, an atlas is an image in one modality with its respective

* Corresponding author. Tel.: +34 913366878; fax: +34 913367709.
E-mail address: carlos.platero@upm.es (C. Platero).

http://dx.doi.org/10.1016/j.artmed.2015.04.005
0933-3657/

labeling (typically generated by manual segmentation) [11]. The atlas-based methods allow a priori knowledge about the appearance and the shape of the anatomical structures to be introduced in a relatively simple way: only a registration method and a number of pre-segmented data sets are required. Barnes et al. [12] proposed registering the most similar atlas from an atlas set to segment the hippocampus. However, segmentations with a single atlas are intrinsically biased toward the shape and the appearance of a subject. Several studies have shown that approaches that incorporate the properties of a group of atlases outperform those that use a single atlas [11,13–16]. The primary benefit of the multi-atlas segmentation approach is that the effect of the errors associated with any single atlas propagation can be reduced in the process of combination. The transferred atlases are used to construct a model for segmenting the target image. This process is often called label fusion. Therefore, there are two steps in the multi-atlas based segmentation: (1) image registration and (2) label fusion. We focus on label fusion in this paper.

The label fusion methods have been classified into two categories: global weighted voting and local weighted voting. Most existing label fusion methods are based on global weighted voting, such as majority voting (MV) [13], STAPLE [17] and weighted voting (WV) [18], which are widely used in medical image segmentation. In these approaches, each atlas contributes to the resulting segmentation with the same weight for all of its voxels. It is very sensitive to the registration errors because it does not take into account the relevance of each sample. Recent works have shown that local weighted voting methods outperform global weighted voting methods [18–20].

Two approaches can be used to take into account the information of each voxel: (i) the atlases are registered non-rigidly into the target image [11,15,16] and (ii) the atlases are aligned in the target image and a patch-based label fusion method is applied [20–24]. The first approach has the advantage of forcing the resulting segmentation to have a similar global shape to those of expert-labeled structures in the atlases. There is a one-to-one mapping between the target image and each atlas. The label fusion methods of this approach generally calculate the labeling associated with the target image via maximum a posteriori (MAP) estimation [15,19,25]. In contrast to fusing label maps using non-rigid registrations, the second approach is based on the nonlocal mean principle [26]. This second approach increases the number of samples considered during the labeling estimation. The typical assumption of one-to-one mapping in non-rigid registration-based techniques is relaxed through the use of local search windows. However, the labeling is local and independent, without global constraints. In this article, we focus on the label fusion methods that use non-rigid registrations. We leave the patch-based labeling methods for future studies.

Approach

The new segmentation algorithms attempt to integrate high-level information with image-based low-level features. At the low level, the appearance of an image patch leads to ambiguities in its labels. For example, in the case of the hippocampal head, the appearance of this structure is convoluted and blends with the amygdala. To overcome these ambiguities, it is necessary to incorporate extra information, such as a priori shape information and contextual information. In medical images, context plays a very important role because the anatomical structures are mostly constrained to relatively fixed positions. From the Bayesian perspective, context information is carried in the joint multivariate statistics in the posterior probability, which is often decomposed into likelihood and prior. In image processing, likelihood and prior often correspond to appearance and shape, respectively.

Brain images present different structures of interest to be segmented. A region-wise approach is more appropriate [27], which can be achieved by dividing the image into multiple anatomically meaningful regions [28]. Therefore, our task is to segment a given 3D target image into K anatomical structures, where K is fixed. We assume that it is possible to define a region of interest (ROI) such that its voxels only belong to a k-structure or to the background. Partitioning the problem into ROIs improves the results of registrations and segmentations. Indeed, the multi-atlas approaches have greater accuracy when the registrations are only made near the object of interest and not in the entire image [28]. Furthermore, these approaches convert the complex multi-label problem into feasible binary segmentation problems. For each ROI, a segmentation is denoted as S_k, and the optimal solution S_k^* can be obtained by the following Bayesian framework:

$$S_k^* = \arg\max_{S_k} p(I_k|S_k;\Theta_k)p(S_k;\Theta_k)$$

where I_k is the target image in the k-ROI and $p(I_k|S_k;\Theta_k)$ and $p(S_k;\Theta_k)$ define the image likelihood and the shape prior of the ROI, respectively. The classification model parameters for the k-ROI are denoted by Θ_k. In general, either a generative or discriminative model is used for the image likelihood, whereas the shape models use the transferred labels from the registered atlases combined with simple geometric constraints. In terms of appearance, generative models have explicit model parameters and are able to capture the global variability, but they often have simplified assumptions, which limits their ability to model inhomogeneous patterns. By contrast, in a discriminative appearance model, there are no explicit parameters to estimate. Discriminative models are able to combine many of the local statistics, which are insensitive to complex and inhomogeneous texture patterns [8]. However, these models have difficulty in taking the regional information into account. For this reason, discriminative appearance models are combined with shape prior models [8,29,30].

Considerable research effort has been devoted to developing efficient algorithms for estimating the MAP solution. The simplest approach is to treat the segmentation as independent voxels and apply the standard classification algorithms. This approach is straightforward, but it loses the important interdependency information. The other extreme of the solution is to treat each instance of L as a single label and estimate its posterior probability. This implementation is infeasible because the space of the output labels grows exponentially with the size of the image. Conditional random fields (CRF) [31] have been widely used to model the correlations of the structured labels. The use of a CRF allows appearance, shape and context to be incorporated in a single unified model, although there are also other alternative approaches [29,32,33].

However, CRF models are typically defined on the basis of a fixed neighborhood structure and make unrealistic conditional independence assumptions, thereby limiting their modeling capabilities. Segmentations using only unary-pairwise potentials tend to over-smooth, and they also have difficulties in capturing global shapes. To overcome these drawbacks, CRF models can be improved through the use of higher-order potentials defined on sets of voxels or cliques [34]. In this study, a CRF model is used for fusing the registered atlases, and this model is characterized by an energy function defined on unary, pairwise and higher-order potentials. The unary potentials of the CRF model are defined as the negative log of the likelihood of a label being assigned to a voxel. It is computed from an appearance model and a label prior. The pairwise edge potentials have the form of a spatial regularizer that minimizes the surface of separation between two different labels [35]. The conventional unary and pairwise cues are coupled with higher-order potentials that are defined on voxel sets generated using *textons* (a texton is a label given to a voxel that describes the local texture