Medical Image Analysis 24 (2015) 297–312

Contents lists available at ScienceDirect

Medical Image Analysis

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

Structured patch model for a unified automatic and interactive segmentation framework

Sang Hyun Park[a], Soochahn Lee[b,*], Il Dong Yun[c,*], Sang Uk Lee[a]

[a] Department of Electrical Engineering, ASRI, INMC, Seoul National University, Seoul, Republic of Korea
[b] Department of Electronic Engineering, Soonchunhyang University, Asan-si, Republic of Korea
[c] Department of Digital Information Engineering, Hankuk University of Foreign Studies, Yongin, Republic of Korea

ARTICLE INFO

Article history:
Received 24 February 2014
Revised 5 January 2015
Accepted 19 January 2015
Available online 29 January 2015

Keywords:
Structured patch model
Interactive segmentation
Adaptive prior
Markov random field
Incremental learning

ABSTRACT

We present a novel interactive segmentation framework incorporating a priori knowledge learned from training data. The knowledge is learned as a structured patch model (StPM) comprising sets of corresponding local patch priors and their pairwise spatial distribution statistics which represent the local shape and appearance along its boundary and the global shape structure, respectively. When successive user annotations are given, the StPM is appropriately adjusted in the target image and used together with the annotations to guide the segmentation. The StPM reduces the dependency on the placement and quantity of user annotations with little increase in complexity since the time-consuming StPM construction is performed offline. Furthermore, a seamless learning system can be established by directly adding the patch priors and the pairwise statistics of segmentation results to the StPM. The proposed method was evaluated on three datasets, respectively, of 2D chest CT, 3D knee MR, and 3D brain MR. The experimental results demonstrate that within an equal amount of time, the proposed interactive segmentation framework outperforms recent state-of-the-art methods in terms of accuracy, while it requires significantly less computing and editing time to obtain results with comparable accuracy.

1. Introduction

With the advancement of medical imaging technology, high quality medical images have significantly increased. Accordingly, the demand for effective techniques to analyze medical images has increased as well. Segmentation of target objects is an especially important task required to study pathological changes of organs, compare inter-subject variability, monitor disease progression and analyze clinical trials. However, manual segmentation is laborious and time-consuming. Thus, various approaches to enable efficient segmentation have been proposed. The wide variety of segmentation methods can be loosely classified as either interactive or automatic.

Interactive methods require the user to provide annotations to incrementally refine the segmentation. To update the segmentation efficiently, most interactive segmentation methods are based on low-level statistics of appearance, including live-wire (Barrett and Mortensen, 1997), region growing (Pohle and Toennies, 2001), interactive graph cut (Boykov and Funka-Lea, 2006; Shim et al., 2009a,b), random walk (Grady, 2006; Kim et al., 2008) and geodesic segmentation (Bai and Sapiro, 2007). Although these methods are fast, flexible, and facilitate intuitive editing, extremely detailed user annotations may be required for noisy images or target objects with obscure boundaries. Many different approaches are taken to overcome this problem. One is to leverage simpler user interactions such as the bounding box of the target object (Rother et al., 2004; Lempitsky et al., 2009). Another is to reduce the amount of required user annotations by using sophisticated graphs which reflect the relationships between annotated and unknown regions (Kim et al., 2010). Yet another approach is to introduce multiple categories of annotations that better represents user intention (Yang et al., 2010). Although these methods may reduce the amount of required annotations, the changes in annotation often affect the segmentation results significantly. To receive more informative user annotations on ambiguous regions, methods based on the active learning strategy have been proposed. In the method by Wang et al. (2012), the expected confidence change of superpixels is measured to inform the user about the regions where annotation is more desired. In the method of Top et al. (2011), the two-dimensional plane having the highest uncertainty among a three dimensional image space is provided to the user at each editing step for the next annotation. Although these methods

* Corresponding authors.
E-mail addresses: shpark13135@gmail.com (S.H. Park), soochahn.lee@gmail.com (S. Lee), yun@hufs.ac.kr (I.D. Yun), sanguk@ipl.snu.ac.kr (S.U. Lee).

http://dx.doi.org/10.1016/j.media.2015.01.003
1361-8415/

guide the user to provide effective annotations, a significant amount of annotations is nonetheless necessary for accurate segmentation of ambiguous regions because the methods still rely on low-level statistics, such as intensity distributions and gradients. Basically, all the aforementioned methods do not utilize a priori knowledge of the object of interest. While this enables the methods to be generalized to various target objects, it hinders use of informative cues of target objects. Moreover, this restricts the reproducibility of the methods. Therefore, the user will need to laboriously repeat similar annotations when segmenting a common target object within many similar images. Clinicians burdened in these situations, which occur very often in clinical practices, will indeed benefit from more automated methods.

On the other hand, most automatic segmentation methods take advantage of a priori knowledge of target objects learned from training data, based on the assumption that target images share the similar appearance and structure. These methods can be divided into example-based and model-based. Example-based methods (Heckemann et al., 2006; Aljabar et al., 2009; van der Lijn et al., 2008; Lotjonen et al., 2010; Coupe et al., 2011; Rousseau et al., 2011; Park et al., 2013a; Asman and Landman, 2013; Bai et al., 2013; Tong et al., 2013) search for relevant example images and their labels from the training set which are directly used to guide the segmentation of the target image. In works of Heckemann et al. (2006) and Aljabar et al. (2009), the segmentation is determined by majority voting of aligned manual segmentation labels. In works of van der Lijn et al. (2008) and Lotjonen et al. (2010), the label information is incorporated into the graph cut framework of Boykov and Funka-Lea (2006) to further deal with local variations. Unlike the methods which directly use the labels of aligned training images, in works of Coupe et al. (2011) and Rousseau et al. (2011), the labels are determined by non-local weighted voting of labels of local atlas patches according to appearance similarity. Tong et al. (2013) proposed a similar patch based label fusion method, but used sparse representation to determine the weight for fusion. Though these methods are effective with a small number of training images, highly complex registration or per-patch similarity computation is required. Thus they are not scalable to the size of training set. Model-based methods overcome the limitation of example-based methods by modeling the target object from training data (Cootes et al., 1995; Duta and Sonka, 1998; van Ginneken et al., 2002; Sukno et al., 2007; Gleason et al., 2002; Seghers et al., 2007; Ibragimov et al., 2012; Yang and Ramanan, 2011; Zhang et al., 2012). For example, in the active shape model (ASM) by Cootes et al. (1995), the average and variations of the shape are modeled by statistics of object boundary landmarks. However, these methods require a large enough training set for the model to be sufficiently generalizable, which is hard to obtain in many clinical tasks where it is common to only have a small number of images. They also require laborious tasks during training such as manual extraction of landmarks or annotation of local parts. While the example-based and model-based methods have reduced the user efforts for many clinical applications, the segmentation results can often be inaccurate due to aforementioned weaknesses, especially at ambiguous regions. Although user editing is necessary in these cases, most automatic methods cannot be easily extended to include an effective interactive editing process.

Recently, several example-based and model-based interactive methods that incorporate a priori knowledge of the target objects or images have been proposed. In works of Barnes et al. (2009) and Barnes et al. (2010), the example-based method *PatchMatch* is proposed for labeling problems by efficiently searching for image patch correspondences and propagating their manual annotations. These methods have been extended to super-resolution of cardiac MRI (Shi et al., 2013) and hippocampus segmentation (Ta et al., 2014) for medical image analyses. Nonetheless, they may not be applicable for target objects with specific shape, since spatial relationships between adjacent patches are neglected. In works of Branson et al. (2011) and Wah et al. (2011), the deformable part model (DPM) (Felzenszwalb et al., 2010) is utilized in an interactive recognition framework supporting seamless learning. However, since the DPM is based on part labels, it is not easily extended to interactive segmentation framework which relies on user annotations, often given as detailed pixelwise labels. In the work by Schwarz et al. (2008), ASM is incorporated into the interactive segmentation framework by enabling the user to edit the positions of landmark points in the determined boundary. Whenever incorrect landmark points are edited by the user, adjacent landmarks are accordingly modified by Gaussian interpolation and the whole boundary is regularized based on the ASM. In the work by Sun et al. (2013), the segmentation boundary is determined by the optimal surface finding (OSF) method, based on an initial segmentation using ASM. The user can correct errors in OSF results by marking points on the correct boundary, which are used as constraints to recompute the OSF. While these methods also incorporate interactive editing with prior information, the required user interaction of 3D point positions can be difficult to achieve with only a common 2D interface.

In this paper, we present an efficient model-based interactive framework using the structured patch model (StPM)[1] for segmentation of target objects within a large number of medical images acquired in a common environment. The proposed StPM is an *example-based-model*, comprising sets of corresponding local patch priors and their pairwise spatial distribution statistics compiled from the example images and their segmentations in the training set. When a test image is given, the optimal local patch priors are adaptively selected and localized through a global probabilistic optimization based on the user annotations, local patch similarity, and the likelihood of global structure based on the pairwise spatial distribution. Then, voxel-wise segmentation labels are computed through a global probabilistic optimization based on the selected StPM and the user annotations.

The key advantages of the proposed framework based on StPM are as follows: First, we enforce the example-based multiple patch priors, which encapsulate a wide variety of specific local instances, into a model structure. It enables the method to use the optimal examples, in terms of both local adaptiveness and global consistency, as priors for segmentation. Second, user annotations are easily incorporated into the segmentation framework. Since the StPM is compatible with all types of annotations, the user can freely insert annotations on ambiguous regions without any restrictions based the model, making efficient segmentation possible for any image. Third, since interactive segmentation is constrained by the StPM as well as user annotations, the segmentation result is robust to the quantity and placement of the annotations. Compared to the previous interactive methods, the proposed method requires fewer annotations and is more robust to their changes due to the StPM. Finally, the StPM can easily be expanded by directly incrementing the local image and segmentation patch set and the pairwise distribution with the results obtained from the proposed framework for a new test image. This incremental learning system is particularly effective when constructing the training image set since the required laborious manual annotation is significantly reduced.

We note that this paper is based on our preliminary work presented by Park et al. (2013b). The preliminary method was sensitive to initial user annotations and could not handle the drifting

[1] We use the acronym *StPM* to avoid confusion with *statistical parametric mapping (SPM)*.