ARTICLE IN PRESS
JID: YCVIU [m5G;April 9, 2016;16:31]

Computer Vision and Image Understanding 000 (2016) 1–14

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Probabilistic model for 3D interactive segmentation

Tsachi Hershkovich[a,b], Tamar Shalmon[b,c], Ohad Shitrit[a,b], Nir Halay[a], Bjoern H. Menze[e], Irit Dolgopyat[d], Itamar Kahn[d], Ilan Shelef[b,c], Tammy Riklin Raviv[a,b,*]

[a] *Electrical and Computer Engineering Department, Ben-Gurion University, Beer-Sheva, Israel*
[b] *The Zlotowski Center for Neuroscience, Ben-Gurion University, Beer-Sheva, Israel*
[c] *Department of Radiology, Soroka Medical Center, Beer-Sheva, Israel*
[d] *Rappaport Faculty of Medicine, Technion, Haifa, Israel*
[e] *Department of Computer Science, TU Munchen, Munich, Germany*

ARTICLE INFO

Article history:
Received 18 February 2015
Revised 3 March 2016
Accepted 10 March 2016
Available online xxx

Keywords:
Image segmentation
User interaction
Probabilistic model
MR & CT Brain imaging
Modality fusion

ABSTRACT

Fully-automated segmentation algorithms offer fast, objective, and reproducible results for large data collections. However, these techniques cannot handle tasks that require contextual knowledge not readily available in the images alone. Thus, the supervision of an expert is necessary.

We present a generative model for image segmentation, based on a Bayesian inference. Not only does our approach support an intuitive and convenient user interaction subject to the bottom-up constraints introduced by the image intensities, it also circumvents the main limitations of a human observer—3D visualization and modality fusion. The user "dialogue" with the segmentation algorithm via several mouse clicks in regions of disagreement, is formulated as a continuous probability map, that represents the user's certainty to whether the current segmentation should be modified. Considering this probability map as the voxel-vise Bernoulli priors on the image labels allows spatial encoding of the user-provided input. The method is exemplified for the segmentation of cerebral hemorrhages (CH) in human brain CT scans; ventricles in degenerative mice brain MRIs, and tumors in multi-modal human brain MRIs and is shown to outperform three interactive, state-of-the-art segmentation methods in terms of accuracy, efficiency and user-workload.

1. Introduction

Being fundamental to medical imaging analysis, image segmentation is actively studied, and numerous approaches exist. Recent trends focus on fully automatic segmentation frameworks, which is much faster than manual annotation, less biased, and repeatable. Usually, the required workload for processing and analyzing large datasets is far behind the ability of a human rater. Moreover, the computational advancements of the machine in cases that require modality fusion or 3D visualization cannot be achieved even by an expert. Nevertheless, as the outcome of the image analysis process might have critical implications on patient recuperation prospects, the expertise of a clinician must be considered.

Interactive segmentation (IS) approaches can be classified based on the form and the type of input provided by the user as well as the underlying segmentation framework (see He et al. (2013) and Zhao and Xie (2013) and references therein). The pioneering IS work, which led to the development of the *live wire* technique or *intelligent scissors*, independently suggested by Falcao et. al. (1998) and Mortensen and Barrett (1998), is based on the image edge map. The shortest paths between the user's mouse clicks calculated by the Dijkstra algorithm form the contour of the region of interest (ROI). Here, as well as in the *united snakes* framework (Liang et al., 2006), which relies on a classical active contour framework known as *snakes* (Kass et al., 1988), the user 'plants' anchors or seed points along the desired boundary, providing guidance for the segmentation.

Mouse scribbles seem to be the most common form of user interaction. The marked regions provide information about the ROI and the background intensity distributions. A well known IS approach is the *GrabCut* technique (Rother et al., 2004), which is based on the *graph-cut* method (Boykov et al., 2001). Representing the image pixels by nodes in a graph, the graph-cut addresses a foreground-background image segmentation by solving a min-cut, max-flow problem. The user's annotated regions are assigned to either the source or the sink of the graph. In a recent paper by Nieuwenhuis and Cremers (2013), marked regions via mouse

* Corresponding author. Tel.: +97286428812; fax: +97286472949.
E-mail address: rrtammy@ee.bgu.ac.il, tammy@csail.mit.edu (T. Riklin Raviv).

http://dx.doi.org/10.1016/j.cviu.2016.03.007
1077-3142/

Please cite this article as: T. Hershkovich et al., Probabilistic model for 3D interactive segmentation, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.03.007

scribbles were used for gathering spatially varying color statistics for multi-label segmentation.

While the techniques above are very effective when prior information on the image to segment is not available, they do not handle local ambiguities resulting from overlaps between the ROI and the background intensity distributions. Moreover, often, in medical image analysis problems the ROI is well specified and image statistics can be estimated based on similar examples. The user input is required only for resolving local discrepancies based on contextual information.

In Ben-Zadok et al. (2009) and Cremers et al. (2007a) the level-set framework of Chan and Vese (2001), which implicitly models the foreground and the background intensities by two distinct normal distributions, was extended to include a spatial term provided via user interaction. This additional spatial information is provided in regions that (according to the user) are not compatible with the assumed intensity distribution model. A random walk algorithm is the foundation of the IS framework proposed in Yang et al. (2010). The method combines soft and hard input constraints to spatially guide the segmentation as well as foreground and background user strokes to learn the image statistics. Split and merge interactive operations were suggested in Paulhac et al. (2011) via region adjacency graph representation. More recently, Gao et al. proposed to use local robust statistics to describe the object features, which are learned adaptively from the strokes drawn by the user (Gao et al., 2012). Their framework allows the partition of the image into multiple regions via the simultaneous evolution of two active contours and using the 'action–reaction' principle to avoid overlaps. The idea of a concurrent evolution of a pair of level-set functions has been used in another interesting approach, based on concepts from control theory (Karasev et al., 2013). In that framework, the user's accumulated input guides the evolution of one of the contours while the evolution of the other contour is based on image intensities and smoothness term, as in Chan and Vese (2001). The two level-set functions interact with each other, leading to a closed loop behavior.

In this paper, we present a novel generative approach for interactive 3D segmentation of medical images. The proposed framework is also extended to address the extraction of a common ROI in a multi-modal image set. The key contribution is the probabilistic formulation of the user interaction. Specifically, the discrete set of 3D coordinates provided by mouse clicks in regions of disagreement is converted into a continuous probability map, that represents the user's certainty to whether the current segmentation should be modified. This continuous representation defines the voxel-vise Bernoulli priors on the image labels. The maximum a posteriori (MAP) estimate of the segmentation is therefore based on a user-provided spatial information in addition to the conventional image likelihood term and a regularization term.

The MAP problem is solved via calculus of variation, using level-set formulation, in the spirit of Riklin Raviv et al. (2010). Interaction with mouse clicks via level-set based segmentation has been suggested before by Ben-Zadok et al. (2009) and Cremers et al. (2007b). Nevertheless, there are essential differences. In Ben-Zadok et al. (2009) and Cremers et al. (2007b) binary (hard) segmentation is considered where each pixel is assigned to either the ROI or to the background. The interaction that follows the automatic segmentation is represented by a map containing clouds of positive and negative values (based on the user provided clicks) directly affecting the level-set evolution by a simple summation. In contrast to these works and to others mentioned above, the proposed framework is entirely based on probabilistic principles. First, the segmentation is fuzzy (soft) such that the value assigned to each voxel represents the likelihood that this voxel belongs to the ROI. Therefore, voxels within or nearby the ROI boundaries have maximum labeling *uncertainty*. Second, the user interactive map is constructed such that it spatially reflects the user's *certainty* that the soft labels of the current (automatic) segmentation estimate should be altered. Finally, the probability that each of the image voxels 'flips' its ROI-background assignment is determined by the user-certainty map by considering it as voxel-wise Bernoulli parameters.

The suggested probabilistic framework leads to a flexible and tolerable interaction, taking into account occasional user mistakes. Note that the user does not edit the segmentation. Instead, our model provides an elegant framework to refine segmentation by resolving voxel annotation ambiguities, through *a user-machine dialogue*. The user term in the unified cost functional is constructed such that voxels that are not within the user's regions of influence do not contribute to the evolution of the segmentation (due to the user). In other words, 'neglecting' a region is not interpreted as 'supporting' its segmentation. Therefore, regions that are incorrectly labeled by the current segmentation and are missed by the user, can be easily corrected in a subsequent interaction step. In contrast, regions that are assigned with high probability to either the ROI or the background by the current soft segmentation are less likely to alter their assignment, if marked by mistake by the user. It turns out that this mechanism, while requiring user persistency in high confidence regions (in which apparently intensity-based segmentation should work well) eventually leads to a reduction in user effort.

Our user-interactive segmentation method is exemplified on the segmentation of three different datasets including cerebral hemorrhages (CH) in human brain CT scans; ventricles in degenerative mice brain MRIs and tumors in multi-modal human brain MRIs.

We developed a GUI to allow a convenient interaction with the software. The tool was tested by our clinical collaborators who acknowledged its operating convenience and accuracy (measured subjectively by rating user satisfaction). Usually, not more than a couple of user interaction steps were needed in order to obtain almost a complete overlap with an independent, fully manual annotation.

The accuracy of the results, measured via Dice scores (Dice, 1945), the method's efficiency (which is inversely proportional to the total duration of the interactive segmentation), as well as the user's workload (indicated by the average number of required interactions) were favorably compared to those obtained by three different state-of-the-art user interactive (UI) segmentation tools, namely the *Grabcut* (Rother et al., 2004), the *TurtleSeg* (Hamarneh et al., 2005; Poon et al., 2007, 2008; Top et al., 2010, 2011) and the *ilastik* (Sommer et al., 2011). Brain tumor segmentation results of multi-modal MRI scans were favorably compared to those reported in the BRATS paper (Menze et al., 2015) suggesting IS as a means to break the limits of fully automatic segmentation approaches, when dealing with extremely challenging data.

The rest of the paper is organized as follows: In Section 2 we define the underlying probabilistic model followed by the introduction of the corresponding level-set formulation (Section 3). Implementation details are discussed in Section 4. Section 5 presents the experimental results. We summarize the paper and suggest future directions in Section 6. Finally, the GUI is described in the Appendix A.

2. Probabilistic model

2.1. Problem definition and formulation

Let $I_1, \dots, I_M$ be an ensemble of gray-level images defined on the same 3D image domain Ω. Here, we assume that $I_1, \dots, I_M$ are aligned multi-modal scans of the same subject, acquired at the same time. Our goal is to extract the common ROI denoted by $\omega \in \Omega$. Let S: $\Omega \rightarrow \{0, 1\}$ denote the corresponding, unknown

Please cite this article as: T. Hershkovich et al., Probabilistic model for 3D interactive segmentation, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.03.007