Methods 115 (2017) 119–127

Contents lists available at ScienceDirect

Methods

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

TED: A Tolerant Edit Distance for segmentation evaluation

Jan Funke [a,b,c,*], Jonas Klein [c], Francesc Moreno-Noguer [a], Albert Cardona [b], Matthew Cook [c]

[a] Institut de Robòtica i Informàtica Industrial, UPC/CSIC, Barcelona, Spain
[b] Janelia Research Campus, VA, Ashburn, United States
[c] Institute of Neuroinformatics, UZH/ETH, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form 29 December 2016
Accepted 30 December 2016
Available online 17 January 2017

Keywords:
Computer vision
Segmentation
Evaluation
Learning
Neuron segmentation
Electron microscopy

ABSTRACT

In this paper, we present a novel error measure to compare a computer-generated segmentation of images or volumes against ground truth. This measure, which we call Tolerant Edit Distance (TED), is motivated by two observations that we usually encounter in biomedical image processing: (1) Some errors, like small boundary shifts, are tolerable in practice. Which errors are tolerable is application dependent and should be explicitly expressible in the measure. (2) Non-tolerable errors have to be corrected manually. The effort needed to do so should be reflected by the error measure. Our measure is the minimal weighted sum of split and merge operations to apply to one segmentation such that it resembles another segmentation within specified tolerance bounds. This is in contrast to other commonly used measures like Rand index or variation of information, which integrate small, but tolerable, differences. Additionally, the TED provides intuitive numbers and allows the localization and classification of errors in images or volumes. We demonstrate the applicability of the TED on 3D segmentations of neurons in electron microscopy images where topological correctness is arguable more important than exact boundary locations. Furthermore, we show that the TED is not just limited to evaluation tasks. We use it as the loss function in a max-margin learning framework to find parameters of an automatic neuron segmentation algorithm. We show that training to minimize the TED, i.e., to minimize crucial errors, leads to higher segmentation accuracy compared to other learning methods.

1. Introduction

In the computer vision literature, several approaches to assess the quality of contour detection and segmentation algorithms can be found. Most of these measures have been designed to capture the intuition of what humans consider to be two similar results. In particular, these measures are supposed to be robust to certain tolerated deviations, like small shifts of contours. For the contour detection in the Berkeley segmentation dataset [1], for example, the precision and recall of detected boundary pixels within a threshold distance to the ground truth became the widely used standard [2,3]. Contour error measures are, however, not a good fit for segmentations, since small errors in the detection of a contour can lead to the split or merge of segments. Therefore, alternatives like the Variation of Information (VOI), the Rand Index [4] (RI), the probabilistic Rand index [5,6], and the segmentation covering measure [3], have been proposed.

However, these measures do not acknowledge that there are different criteria for segmentation comparison, and instead accumulate errors uniformly, even for many small differences that are irrelevant in practice. Especially in the field of biomedical image processing, we are often more interested in counting true topological errors like splits and merges of objects, instead of counting small deviations from the ground truth contours. This is in particular the case for imaging methods for which no unique "ground truth" labeling exists. In the imaging of neural tissue with Electron Microscopy (EM), for example, the preparation protocol can alter the volume of neural processes, such that it is hard to know where the true boundary was [7]. Further, the imaging resolution and data quality might just not be sufficient to clearly locate contours between objects [8], resulting in a high inter-observer variability.

1.1. Contributions

The main contribution of this paper is a novel measure to evaluate segmentations on a clearly specified tolerance criterion to address the aforementioned issues. At the core of our measure,

* Corresponding author at: Institut de Robòtica i Informàtica Industrial, UPC/CSIC, Barcelona, Spain.

E-mail addresses: jfunke@iri.upc.edu (J. Funke), klein@ini.uzh.ch (J. Klein), fmoreno@iri.upc.edu (F. Moreno-Noguer), cardonaa@janelia.hhmi.org (A. Cardona), cook@ini.uzh.ch (M. Cook).

http://dx.doi.org/10.1016/j.ymeth.2016.12.013
1046-2023/© 2016 Published by Elsevier Inc.

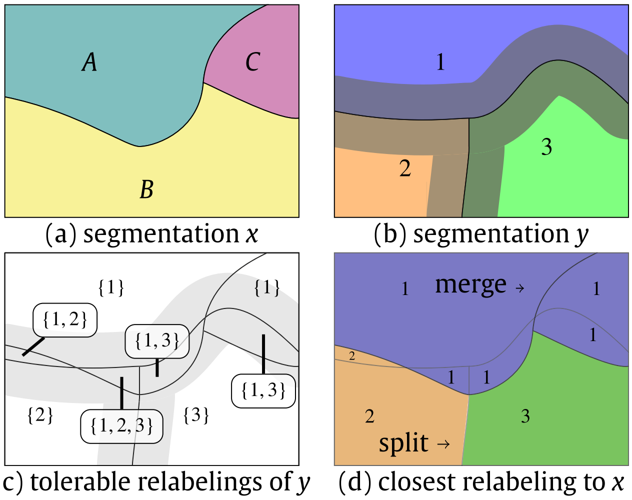

Fig. 1. Illustration of the Tolerant Edit Distance (TED) between two segmentations x and y. By tolerating boundary shifts to a certain extend, shown as shadow in (b), y is allowed to be changed to match x as closely as possible. For that, we consider regions obtained by combining x and y, illustrated in (c). For each of these regions, we enumerate a set of labels used by y that are within a threshold distance to all locations inside the region (shown in curly brackets). This threshold is the maximally allowed boundary shift. Note that in this example, the region obtained from intersecting A and 3 can change its label to 1 (or keep 3), but not to 2, since it contains points that are too far away from region 2. Regions with only one possible label are too large to be relabeled by shifting their boundary and have to keep their initial label. From all the possible ways to relabel y, the relabeling (d) minimizing the number of split and merge errors compared to x is chosen by solving an integer linear program.

which we call *Tolerant Edit Distance* (TED),[1] is an explicit tolerance criterion (*e.g.*, boundary shifts within a certain range). Using integer linear programming, we find the minimal weighted sum of split and merge operations to transform one segmentation into another, which is tolerably close to the ground truth. By setting the weights of the split and merge operations to the expected effort to perform these operations, the TED reflects the total effort needed to manually fix a segmentation. Similar to VOI and RI, our measure does not require voxels of the same object to form a connected component, and can thus be applied to volumes with missing data, known object connections via paths outside the volume, or on stitched volumes with registration artifacts. The reported numbers are intuitive (*e.g.*, time or cost effort to fix a segmentation), easy to interpret (splits and merges of objects), and errors can be localized in the volume. An illustration of the TED can be found in Fig. 1.

1.2. Application to neuron segmentation

To demonstrate the usefulness of our measure, we present our results in the context of automatic neuron segmentation from EM volumes, an active field of biomedical image processing (for recent advances, see [9–13]). We argue that especially in this field there is a need for explicit and intuitive error measures. Furthermore, we show how the TED can be used to train neuron segmentation algorithms. Our findings (based on our previous work [14]) show that training to minimize the TED leads to higher segmentation accuracy on a range of error measures, compared to other methods.

1.2.1. Evaluation

As it is the case in many biological applications, the criterion to assess the quality of a neuron segmentation depends on the biological question one would like to answer. On one hand, *skeletons* of neurons are sufficient to identify individual neurons [15], to study neuron types and their function [16], and to obtain the wiring diagram of a nervous system (the so-called *connectome*) [8]. In these cases, topological correctness is far more important than the diameter of a neural process or the exact location of its boundary (see Fig. 2 for examples). On the other hand, for biophysically realistic neuron simulation, *volumetric* information is needed to model action potential time dynamics, and to understand and simulate information processing capabilities of single neurons [17]. In this case, the segmentation should be close to the true volume of the reconstructed neurons. Only small deviations in the boundary location might still be tolerable.

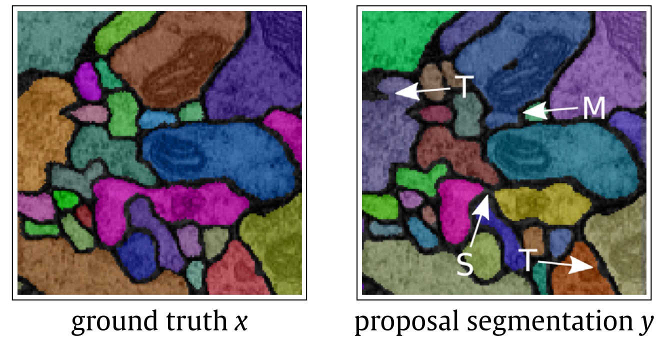

Fig. 2. Example errors made by an automatic neuron segmentation algorithm. Errors like merges (M) and splits (S) dramatically change the reconstructed topology and should be avoided. Small disagreements in the boundary location (T) are however tolerable and should be ignored during evaluation.

Currently, reporting segmentation accuracy in terms of VOI or RI is the de facto standard [11,18,10,12,13]. Less frequently used [9,19] is the *Anisotropic Edit Distance* (AED) [9] and the *Warping Error* (WE) [20]. The AED is tailored to the specific error correction steps required for anisotropic volumes (splits and merges of 2D neuron slices within a section, connections and disconnections of slices between sections). The WE aims to measure the difference between ground truth and a proposal segmentation in terms of their topological differences. As such, the WE was the first error measure for neuron segmentation that deals with the delicate question of up to which point a boundary shift is not considered to be an error. However, since the WE assumes a foreground-background segmentation where connected foreground objects represent neurons, it is only applicable to volumes in which connectedness of neurons is preserved. Furthermore, only suboptimal solutions to the WE are found using a greedy, randomized heuristic, which makes it difficult to use for evaluation purposes. Consequently, the WE has found its main application in the training of neural networks for image classification [20].

In 2 we introduce the TED as an alternative to address some of the shortcomings of existing measures. Similar to the WE, the TED is designed to ignore small deviations from the ground truth and only count true topological errors, but is computed deterministically and to global optimality and does not impose constraints on the types of volumes being compared.

1.2.2. Training

Current state-of-the-art methods for automatic neuron segmentation can broadly be divided into isotropic [11,18,12,13] and anisotropic methods [9,10,19]. Assignment models constitute the current state of the art for the segmentation of neurons from anisotropic volumes, as obtained by serial section EM [9,10]. These models enumerate and price possible assignments of candidate segments across sections of EM stacks (see Fig. 8 for an overview and 3 for details). A final segmentation is found by selecting a cost minimal and consistent subset of all assignments.

[1] Source code available at http://github.com/funkey/ted.