INTERACTIVE GRAPH CUT SEGMENTATION OF TOUCHING NEURONAL STRUCTURES FROM ELECTRON MICROGRAPHS

Vignesh Jagadeesh and B S Manjunath *

Center for Bio-image Informatics
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560
http://www.bioimage.ucsb.edu/

ABSTRACT

A novel interactive segmentation framework comprising of a two stage *s-t* mincut is proposed. The framework has been designed keeping in mind the need to segment touching neuronal structures in Electron Micrograph (EM) images. The first stage undersegments the image, and groups touching structures into a single class. The second stage accepts user interaction to separate touching structures. The technique introduces user feedback through a Markov Random Field formulation. Furthermore, a method for constructing interaction potentials using an edge response function is proposed. Encouraging results, and a comparison to state of the art methods is presented.

Index Terms— Graph Cuts, Interactive Segmentation, Markov Random Fields

1. INTRODUCTION

Electron Micrographs (EM) are widely used in neuroscience for morphological studies. They offer sub-cellular resolutions and are an important source of information to biologists. The basic and most important step in analyzing these images is the segmentation of individual neuronal structures. However, the segmentation of EM images [1, 2, 3] is complicated due to poorly defined image gradients, and inhomogeneous intensity distributions of foreground objects.
A grand challenge in neuroscience is to understand how neurons are wired together in the mammalian brain [4]. EM images offer resolutions at which the wiring diagrams can be accurately reconstructed. In order to tap into this rich information source, one is confronted with the problem of analyzing massive amounts of image data. As a result, manual segmentation of EM databases could be very time consuming and laborious. Hence, semi-automated solutions have been proposed to segmenting EM images. This paper presents a novel framework for interactive segmentation, that overcomes some drawbacks of previous methods. The primary contribution lies in the segmentation of multiple touching structures with a two-stage s-t mincut algorithm. The first mincut is designed to undersegment the image, while the second cut accepts user interaction to help delineate touching boundaries. In other words, segmentation of several touching structures (15-20 per image in the dataset considered) is achieved at the end of the second mincut.
The next section briefly summarizes concepts from graph cuts, distance transforms, and previous work related to the problem at hand. The third section presents the proposed framework, followed by an experimental comparison to state of the art methods. The final section concludes with a summary of our work, and possible extensions.

*We thank Dr.Robert Marc, Dr.Bryan Jones and Dr.James Anderson, affiliated with the MarcLab (John A Moran Eye Center at University of Utah) for providing images used in the experiments reported in this paper. The authors are supported by NSF OIA #0941717.

2. BACKGROUND AND PRIOR WORK

2.1. Graph Cuts

Image Segmentation can be posed as an energy minimization problem. The energy function to be minimized is parameterized by the labels ($y_p \in \{0, 1\}$) assigned to every image pixel (p). The set of labels assigned to image pixels (y) is called a labeling configuration. The objective is to find the labeling configuration that exactly minimizes the energy function defined in Equation 1.

$$E(y) = \underbrace{\sum_{p\in P} V_p(y_p)}_{\text{Unary Potentials}} + \underbrace{\sum_{p\in P, q\in N_p} V_{pq}(y_p, y_q)}_{\text{Interaction Potentials}}, \quad (1)$$

where $V_p(y_p)$ is the negative log likelihood of pixel p taking up label y_p. The term $V_{pq}(y_p, y_q)$ is usually a function of label differences between pixel p, and a pixel q in the neighborhood system N_p of p. Interaction potentials used in this work are of the form:

$$V_{pq}(y_p, y_q) = \lambda_I |y_p - y_q| \exp(-\frac{||F_p - F_q||_2^2}{2\sigma_I^2}), \quad (2)$$

where F_p and F_q are pixel level features of p and q respectively, λ_I and σ_I are parameters controlling the magnitude and smoothness of the interaction potentials. We employ graph cut based techniques for the optimization, and briefly describe the procedure.

Consider a graph, $G = (V, E)$, where $V = \{P, s, t\}$ corresponds to pixels $p \in P$ in an image, along with two special nodes, the source s and the sink t. The set $E = \{e_{pq} \cup e_{sp} \cup e_{pt} : p, q \in P\}$ consists of all edges connecting nodes on the image grid *(n-links or interaction potentials)*, and edges connecting each node on the image grid to the source and sink (s and t links), respectively. A cut in the graph is a set of directed edges $C = \{p, q; p \in S, q \in T\}$, which when removed leave no paths from source to sink. The cut with minimum sum of edge weights is called a mincut. Note that a cut partitions the graph into nodes belonging to the source S and sink sets T respectively. It can be shown that a mincut

$$c(S,T) = \sum_{p,q \in \{S,T\},(p,q)\in E} w(p,q) \quad (3)$$

on the graph G exactly finds $y^* = \arg\min_y E(y)$. $w(p,q)$ is the weight of the directed edge from p to q. We refer the reader to [5, 6, 7] for a detailed treatment of graph cuts and applications to segmentation.

Table 1. List of Variables Used

Variable	Explanation
$\mathcal{S}_{\mathcal{F}}$	Set of pixels segmented as foreground
$\mathcal{S}_{\mathcal{B}}$	Set of pixels segmented as background
$D(p, M)$	Distance Transform
$D_{gd}(p, M)$	Geodesic Distance Transform
$\mathscr{F}$	Foreground Interaction(Set of Pixels)
$\mathscr{B}$	Background Interaction(Set of Pixels)
$\mathscr{E}$	Boundary Interaction(Set of Pixels)
w_{pt}	Sink Potentials for Graph Cuts
w_{sp}	Source Potentials for Graph Cuts
w_{pq}	Interaction Potentials for Graph Cuts

Distance Transforms: For an image P, and a subset of pixels M $\{M \subset P, m_i \in M; 1 \leq i \leq |M|\ \}$,

$$D(p,M) = \begin{cases} \min(||m_i - p||_2) & p \in \{P \backslash M\} \\ 0 & p \in M \end{cases}$$

A related function is the geodesic distance transform, where distances described previously are replaced by weighted distances. The weighting function (or speed function) is defined over the image domain.

2.2. Prior Work

In interactive segmentation systems, the user gives markers that help the segmentation algorithm move towards the desired solution. The information provided by the user is treated as a gold standard, and is hardcoded into the algorithm. For example, if the user marks some pixels as foreground and background, the algorithm should be constrained to obey labels provided by the user. Boykov et al. [7], were the first to propose the idea of interactive segmentation. Vu et al. [2] presented a method ($\mathbf{M_{Vu}}$) which assumes that true foreground pixels are far away from the hardcoded background($\mathscr{B}$), while the true background pixels are far away from the hardcoded foreground ($\mathscr{F}$). They compute geodesic(weighted) distances with the image gradient magnitude as the weighting function.

They propose the following source and sink potentials after user interaction.

$$w_{sp} = \begin{cases} 1 - \exp(-D_{gd}(p, \mathscr{B})) & p \in \{\mathcal{S}_{\mathcal{F}} \backslash \mathscr{F}\} \\ K & p \in \mathscr{F} \\ 0 & p \in \{\mathcal{S}_{\mathcal{B}} \backslash \mathscr{F}\} \end{cases}$$

$$w_{pt} = \begin{cases} 1 - \exp(-D_{gd}(p, \mathscr{F})) & p \in \{\mathcal{S}_{\mathcal{B}} \backslash \mathscr{B}\} \\ K & p \in \mathscr{B} \\ 0 & p \in \{\mathcal{S}_{\mathcal{F}} \backslash \mathscr{B}\} \end{cases}$$

Grady and Funka Lea [3] presented a method the assumes that true foreground pixels are close to the hardcoded foreground ($\mathscr{F}$), while the true background pixels are close to the hardcoded background ($\mathscr{B}$). They propose two variations to the cost function, one ($\mathbf{M_{Gr1}}$) that assigns a constant confidence value (κ) to the presegmented foreground and background pixels, constrained by the user interaction.

The second cost function ($\mathbf{M_{Gr2}}$) proposed by [3] assumes localized corrections, and confidence values are modulated by their distance from the foreground and background marks by the user.

$$w_{sp} = \begin{cases} \kappa \exp(-D(p, \mathscr{F})) & p \in \{\mathcal{S}_{\mathcal{F}} \backslash \mathscr{F}\} \\ K & p \in \mathscr{F} \\ 0 & p \in \{\mathcal{S}_{\mathcal{B}} \backslash \mathscr{F}\} \end{cases}$$

$$w_{pt} = \begin{cases} \kappa \exp(-D(p, \mathscr{B})) & p \in \{\mathcal{S}_{\mathcal{B}} \backslash \mathscr{B}\} \\ K & p \in \mathscr{B} \\ 0 & p \in \{\mathcal{S}_{\mathcal{F}} \backslash \mathscr{B}\} \end{cases}$$

3. PROPOSED METHODOLOGY

A common feature in all methods discussed, is the introduction of user interaction through likelihood potentials V_p. This can be justified because previous methods aimed at correcting regional errors in segmentation through user interaction. However, the application at hand requires correction of errors at object boundaries (edge based corrections) which these methods do not achieve. In order to introduce an edge based correction factor information from user input is induced through interaction potentials. This is in keeping with

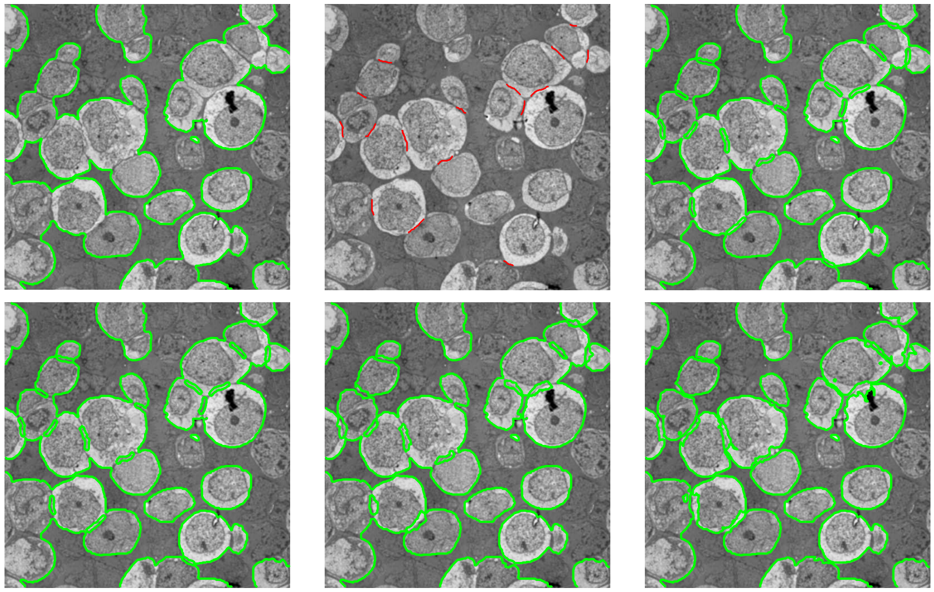

Fig. 1. (Left to Right and Top to Bottom) **(a)** First pass of graph cut **(b)** User interaction indicating boundaries of neuronal structures **(c)** Result using method $\mathbf{M_{Gr1}}$ of Grady et al. [3] **(d)** Result using method $\mathbf{M_{Gr2}}$ of Grady et al. [3] **(e)** Method $\mathbf{M_{Vu}}$ of Vu et al. [2] **(f)** Proposed Method. ***(Image Best Viewed in Color)***

the spirit of Markov Random Field formulations that introduce priors using clique potentials. The introduction of user interaction is done using an edge response function, as will be explained.

An important issue while segmenting touching structures is the delineation of their boundaries. We make this problem amenable to user interaction by dividing it into two phases. The first phase produces an undersegmentation as shown in Figure 1a. Note that undersegmentation can easily be obtained using graph cuts by increasing the effect of interaction potentials. The interaction potentials V_{pq} for the first phase uses image intensity as the feature (F_p, F_q) in Equation 2. The likelihood potentials V_p for the first *s*-*t* are learnt offline. In the second phase, the user is expected to roughly indicate the presence of a boundary between structures. Background likelihoods are increased near user interacted pixels.

We create an edge response function (Θ), formed by a smoothed gradient response. This procedure of creating a clean edge response is essential in getting rid of trivial solutions that could result due to very noisy gradients. The user interacted boundary pixels are now introduced as pixels in the edge response function with some high value K. The interaction potentials evolved from the edge response will favor creation of boundaries along user interacted marks. Refer to Figure 2 for the potentials created using the proposed modifications. Observe the large intensity difference between user interacted pixels and background pixels in the feature map F, and the sink potentials encouraging the cut to pass through user interacted pixels. The proposed formulation can be stated as:

$$F_p = \begin{cases} K & p \in \mathscr{E} \\ \Theta_p & \text{otherwise} \end{cases}$$

$$w_{sp} = \begin{cases} \kappa \exp(-\Theta_p) & p \in \{\mathcal{S}_{\mathcal{F}} \backslash \mathscr{E}\} \\ 0 & \text{otherwise} \end{cases}$$

$$w_{pt} = \begin{cases} \kappa & p \in \{\mathcal{S}_{\mathcal{B}}\} \\ 1 - \exp(-D(p, \mathscr{E})) & p \in \{\mathcal{S}_{\mathcal{F}}\} \end{cases}$$

To the best of our knowledge, the proposed method is the first of its kind in introducing user interaction through interaction potentials. Further, the technique of creating interaction potentials from smoothed edge response functions has been employed for the first time. The overall energy function being minimized in the second pass is given by:

$$\begin{aligned} E(y') &= \sum_{p \in P} w_{sp} y_p (1 - y'_p) + \sum_{p \in P} w_{pt} (1 - y_p) y'_p \\ &+ \sum_{p \in P, q \in N_p} V_{pq}(y'_p, y'_q) \end{aligned} \tag{4}$$

In the above equation, y_p is the set of labels obtained at the end of the first *s*-*t* cut. Minimization of Equation 4 yields the set of labels y'_p that separate touching structures.

4. EXPERIMENTS

The algorithm developed was tested on 2D slices of EM images. The main advantage of the proposed method is the provision of freedom, or immunity for human markings against

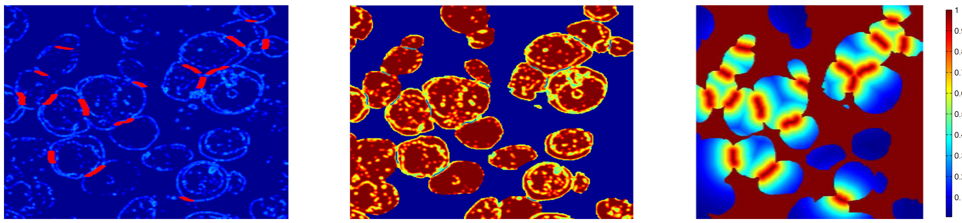

Fig. 2. Potential functions (normalized for visualization) for the second *s-t* cut. (Left to Right) **(a)** The feature map F used to compute interaction potentials **(b)** Source (Foreground Potentials) **(c)** Sink (Background Potentials). ***(Best Viewed in Color)***

minor errors. Moreover, creating interaction potentials from edge response functions provides more reliable gradient maps than existing methods. The advantage of using the proposed method is evident from the fact that around 20 structures were segmented from the single image in Figure1 with less than 10 seconds of human interaction. This procedure is much faster than performing a single *s-t* cut for segmenting neuronal structures individually. Since the segmentation is to be used on a large database, no user interactions for foreground and background were hardcoded. From the output of Figure 1(f), a simple connected components analysis would yield the contours of all neuronal structures that have been segmented. The parameters of the segmentation algorithm were fixed throughout the comparison between different methods. $\{\lambda_I, \sigma_I\}$ values for the first and second pass were $\{20, 20\}$ and $\{1, 10\}$ respectively, while κ was set to 1. The higher value of $\{\lambda_I, \sigma_I\}$ in the first pass causes the creation of larger connected components. Reduction of the values results in smaller connected components, and thus separates touching structures. Table 2 lists results obtained on different images from a 3D EM stack. The expected structures ($\mathbf{ES}$) is the number structures that would result if the algorithm correctly delineated all boundaries. The structures detected by the proposed and competing methods is listed for seven different images. A structure was considered a detection only if it could be completely isolated through a connected components analysis. For example, if the boundary separating two structures were not correctly detected, both structures were marked undetected. It can be observed that the proposed algorithm outperforms competing methods in all images. An occasional drop in performance was uniform across all methods, and can be attributed to boundaries with minimal discriminating information.

5. CONCLUSIONS

We have presented a new interactive segmentation framework that is capable of segmenting touching neuronal structures on an ensemble. Experimental results obtained show encouraging performance compared to state of the art methods. We are currently working on enhancing the existing method to work in 3-D, and in creating online methods that learn from user markings.

Table 2. Proposed method has better detection rate in comparison to existing methods

Img	$\mathbf{M_{Gr1}}$	$\mathbf{M_{Gr2}}$	$\mathbf{M_{Vu}}$	Proposed	$\mathbf{EC}$
1	1	2	4	16	16
2	2	1	3	12	14
3	1	1	4	13	16
4	4	4	8	19	19
5	-	-	10	16	23
6	2	2	6	15	21
7	-	-	4	12	14

6. REFERENCES

[1] I. Carlbom, D. Terzopoulos, and K.M. Harris, "Computer-assisted registration, segmentation, and 3D reconstruction from images of neuronal tissue sections," *IEEE TMI*, vol. 13, no. 2, pp. 351–362, 1994.

[2] N. Vu and BS Manjunath, "Graph cut segmentation of neuronal structures from transmission electron micrographs," *Proceedings of the IEEE ICIP 2008*, 2008.

[3] L. Grady and G. Funka-Lea, "An energy minimization approach to the data driven editing of presegmented images/volumes," *LNCS*, vol. 4191, pp. 888, 2006.

[4] J.R. Anderson et al., "A computational framework for ultrastructural mapping of neural circuitry," *PLoS Biology*, vol. 7, no. 3, 2009.

[5] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE TPAMI*, vol. 26, no. 2, pp. 147–159, 2004.

[6] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE TPAMI*, pp. 1124–1137, 2004.

[7] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient nd image segmentation," *IJCV*, vol. 70, no. 2, pp. 109–131, 2006.