AXON SEGMENTATION IN MICROSCOPY IMAGES - A GRAPHICAL MODEL BASED APPROACH

F. Noushin Golabchi and Dana H. Brooks

Northeastern University, Electrical and Computer Engineering Department, Boston, MA

ABSTRACT

Image segmentation of very large and complex microscopy images are challenging due to variability in the images and the need for algorithms to be robust, fast and able to incorporate various types of information and constraints in the segmentation model. In this paper we propose a graphical model based image segmentation framework that combines the information in images regions with the information in their boundary in a unified probabilistic formulation.

Index Terms— model based image segmentation, probabilistic graphical models, microscopy image segmentation, axon segmentation

1. INTRODUCTION

Image analysis and segmentation of large complex microscopy images is a challenging problem. In some settings the image analysis is dominated by the need to identify and characterize a very large number of small objects of interest in images that have large amount of variability, have complicated backgrounds, and may have low contrast. Microscopy images of stained histological sections of brain or spinal cord tissue are one such example, where one seeks quantitative measurements from closely packed individual axons within a very large data set. For example, one may want to use such images to extract physiological features for comparison to other lower resolution imaging modalities, such as in studies of the physiological basis of diffusion weighted MRI [1]. The inability to efficiently identify and quantitate individual objects such as axons can lead to the use of suboptimal and subjective ROI-based methods that can be difficult to repeat or to generalize from [2].

Image segmentation algorithms for these types of data need to be fast, robust, able to handle complex images, and able to incorporate various types of information and constraints in the segmentation model. There are few reports in the literature on methods to efficiently address this problem. Deterministic approaches including clustering-based methods such as mean shift [3], active contours [4], level set-based methods [5], and graph cut-based methods [6] are often not directly applicable. They either require some image-specific manual initialization, or do not incorporate all the critical information in the images into the segmentation algorithm. Model-based probabilistic approaches have also been used in image segmentation; they formulate the segmentation problem as a stochastic optimization. The benefit of using these models is the ability to incorporate various types of information available in the data into the model. These approaches include graphical models [7], [8], such as Markov random fields (MRF), conditional random fields (CRF) and Bayesian networks (BN). Other probabilistic approaches include discriminative models without using graphical models. The benefit of using graphical models in this framework is the ability to systematically combine information in the image data as well as other prior knowledge about the data.

The histology data images of interest in this study come from human cadaver brain and spinal cord tissue. The ultimate goal is to quantify the distribution of axon density, size, eccentricity, etc. These images contain a very large number of very small homogenous objects (often only a few pixels in diameter) which include the desired axons, but also nuclei, branches of through-plane axons, and other structures that appear similar to axons based on their color and intensity, but which can be distinguished from axons by considering their shape and the structures in their local neighborhood. A sample 100×100 pixel histology image is shown in Fig. 1, where desired axon objects can be identified as small dark objects. The entire histology data consists of 128×128 images of 716×1116 pixels, i.e. the composite image to be processed contains about 10^{10} ($140{,}000\times90{,}000$) pixels, where the objects of interest are of the order of a few pixels. To address this problem, we adopt a probabilistic modeling framework, defining an axon model that represents axon pixels based on their color and intensity, the information in their local neighborhood, and the information in their boundaries. In this framework, an axon model can be thought of as a posterior class-conditional probability density function based on the above information and the image. The probability maps for class axon can then be converted to homogenous regions representing each axon.

Support for the work of FNG was provided by the Lorena Kreda Fund and the Communications and Digital Signal Processing (CDSP) Center at Northeastern University and that of DHB in part by the NIH/NCRR Center for Integrative Biomedical Computing (CIBC), 2P41. The data was provided by Dr Stephan Maier, Brigham and Women's hospital, Harvard Medical School. RR0112553-12

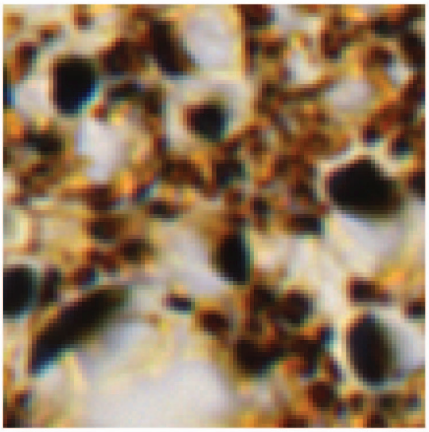


Fig. 1. Sample histology image shows axon objects as dark instances.

We report here on our development of a graphical model axon segmentation approach, motivated by [9], that unifies these two types of information - the information in axon regions and the information in their boundaries. Our axon region model incorporates intensity, color, and texture and homogeneity information available in different color channels, while our boundary model includes curvature, in addition to gradient information. We hypothesize that the curvature based information will be particularly useful in identifying structures that have intensity measurements similar to axons, but boundary shapes with low probability of being an axon boundary.

In more detail, the information available in axon regions are modeled by undirected graphical models, a CRF model, that can effectively capture noncausal relationships between different image entities, such as spatial homogeneity between different axon regions, by encouraging adjacent regions to be classified into the same label. The benefit of using CRFs [10] over other undirected graphical models such as MRFs [11] is that they directly model the posterior probability distribution of regions being axon given the region measurements, which is our desired model for an axon region. This makes the CRF a discriminative model for axons given the observations. Another benefit of using CRFs is that they assume conditional dependence between the neighboring observations using the Markovian assumptions, and incorporate relationship between neighboring observations into the model. The specific discriminative CRFs used here are based on arbitrary discriminative classifiers and are called discriminative random fields (DRF)[12].

Using the information in the boundary pixels separately, the axon boundary is modeled by DRFs as the posteriori probability distribution of boundary pixels being on a true axon boundary. The boundary information can also be thought of as the information in the edge pixels of axon regions and modeled by directed graphical models of BNs. In the BN framework the boundary model is combined with the DRF model of axon regions where the combined model systematically incorporates the two types of information, as well as the relationship between the two types of information, into one general framework, in a probabilistic manner [9].

In Sec. 2 we describe in some detail the specific modeling choices we employed for each components of our overall model, explain why those choices were made and describe how we combined them together into a single overall model. In Sec. 3, using stained images of human corpus callosum as our study example, we show axon probability maps as the output of our DRF model on image measurements, and the axon boundary classification based on applying the DRF model on the boundary measurements. In Sec. 4 we conclude and describe our next steps.

2. METHODS

In this section we first explain our CRF model for axon regions, explain how different component of the model are obtained, and relate the model to a DRF model. We then explain how the same DRF model can be applied to axon boundary nodes. Finally, we also propose a BN based model for axon boundaries and explain how they can be combined with the DRF model for axon regions.

2.1. CRF based image segmentation

A CRF model is used to segment an image into axon region (nodes) y=$\{y_i\}$, where y is the joint labeling of all regions (nodes) y_i, by modeling the posterior probability distribution of image regions (nodes) y_i, given image observation x, defined as $p(y_i|\{y_j\}_{j\neq i}, \mathbf{x})$. Conditioned on image observation x, the axon regions follow a Markovian property i.e. $p(y_i|\{y_j\}_{j\neq i}, \mathbf{x}) = p(y_i|N(y_i), \mathbf{x})$ where $N(y_i)$ represents the spatial neighborhood of y_i, so that any image region y_i depends only on its neighboring (adjacent) regions and is independent of its non-neighbor regions. The joint posterior probability distribution of region nodes y is given by:

$$P(y|x) = \frac{1}{Z}\exp\left(\sum_{i\in V} A_i(y_i, x) + \sum_{i\in V}\sum_{j\in N_i} I_{ij}(y_i, y_j, x)\right) \quad (1)$$

$$= \frac{1}{Z}\prod_{i\in V}\phi(y_i, x_i)\prod_{i\in V}\prod_{j\in N_i}\exp\left(y_i y_j \lambda^T g_{ij}(x)\right) \quad (2)$$

where $\phi(y_i, x_i)$ is the unary potential (or association potential $A_i(y_i, x)$), i.e. likelihood of label y_i given its feature vector x_i from training data or the association of node y_i to class i, $\exp\left(y_i y_j \lambda^T g_{ij}(x)\right)$ is the pairwise potential (corresponding to interaction potential $I_{ij}(y_i, y_j, x)$), i.e. adjacency between y_i, y_j, given x. $g_{ij}(x) = [1, |x_i - x_j|]^T$ is defined as the difference between feature vectors x_i, x_j, where our feature vectors include image intensity, RGB and CIELAB color information, and local neighborhood measures including standard

deviation, range, and entropy of intensity and RGB color information for each image region (node y_i). Z is the normalizing term (a detailed calculation for Z is described in [9]). Parameter λ is calculated using a maximum likelihood estimation (MLE) of the pairwise potential, and $\phi(y_i, x_i)$ is calculated from training data using a logistic regression classifier. V is the set of all image regions and N_i is the set of regions in the neighborhood of ith region.

Using the DRF framework [12] the local class posterior for region y_i being axon given its measurements x_i is defined using generalized linear models and the logistic function as the link function:

$$P(y_i = 1|\mathbf{x}) = \frac{1}{1 + \exp\left(-(\omega_0 + \omega_1^T \mathbf{x})\right)} = \sigma(\omega^T \mathbf{x}) \quad (3)$$

where $y_i = 1$ refers to region y_i being an axon region, and $y_i = -1$ a non axon region. Using (3), the probability density function of region y_i given measurements x_i is $P(y_i|\mathbf{x}) = \sigma(y_i \omega^T \mathbf{x})$ and the local association potential $A_i(y_i, x_i) = \ln(\sigma(y_i \omega^T \mathbf{x}))$. We find the ω parameters by maximum likelihood using a simplex search method.

The local interaction potential $I_{ij}(y_i, y_j, x)$ is the local posterior of two neighboring regions having same labels given their labels y_i and y_j and their measurements x_i and x_j and is defined similarly to (3):

$$P(y_i = y_j|x, y_i, y_j) = P(t_{ij} = 1|t_{ij}, x_i, x_j) = \sigma(t_{ij} \nu^T \mathbf{x})$$

where $t_{ij} = 1$ when $y_i = y_j$ and $t_{ij} = -1$ otherwise. We also find the ν parameters using maximum likelihood estimation and indeed find them jointly with the ω parameters.

2.2. DRF based axon boundary segmentation

We use the DRF framework to find the posterior probability distribution of image boundary (nodes) b_i given boundary measurements M_b where each boundary node consists of boundary pixels corresponding to an image region y_i:

$$P(b|M_b) = \frac{1}{Z} \exp\Big(\sum_{i \in W} A_i(b_i, M_b) + \quad (4)$$

$$\sum_{i \in W} \sum_{j \in W_i} I_{ij}(b_i, b_j, M_b)\Big). \quad (5)$$

We select boundary nodes using an appropriate level set of the axon region probability density function $P(y|x)$. W is the set of all boundary nodes, and W_i the set of adjacent boundary nodes to b_i. Since the level set function leads to boundary nodes that are disjoint and independent of each other, the second term in (5) will be zero. The first term is similar to a logistic regression classifier using the boundary measurements M_b and the probability of boundary node b_i being a true axon boundary given boundary measurements, given as $P(b_i|M_b) = \sigma(b_i \mu^T \mathbf{M_b})$, and the joint probability of all boundary nodes being true axon boundary is given as:

$$P(b|M_b) = \frac{1}{Z} \prod_{i \in W} \phi(b_i, M_{bi})$$

Our axon boundary measurements are divided into two category: curvature features and gradient features. We first obtain closed boundaries from a level set of axon density maps, then fit cubic splines to the boundary pixels and calculate the normalized local curvature. Curvature features are 1) the number of sign changes larger than a self-referenced threshold, 2) the curvelengths with continuous curvatures below another (negative) threshold, 3) the sum of the values in 2, and 4) the overall minimum and maximum curvature. Gradients are calculated along profiles normal to the boundary.

2.3. BN based axon boundary segmentation

In a Bayesian framework the goal is to find the state of boundary nodes b_i given not only the boundary measurements, but also the image measurements using a BN (see [9]). Assuming axon boundary nodes b_i are derived from level set of axon density maps, they are dependent on their corresponding axon region nodes y_i and their boundary measurements M_{bi}. The joint probability of all image region nodes $y = \{y_i\}$, all image boundary nodes $b = \{b_i\}$, and all image boundary measurements $M_b = \{M_{bi}\}$ is given by:

$$\begin{aligned} P(y, b, M_b) &= P(M_b|b)P(b|y)P(y) \\ &= \prod_{j=1}^{m} P(M_{b_j}|e_j)P(b_j|pa(b_j)) \prod_{i=1}^{n} P(y_i) \end{aligned}$$

where $pa(b_j)$ are the parents of b_j, i.e. image regions neighboring the boundary, and $P(y_i)$ is the prior for axon region labels, assumed to be uniform. $P(b_j|pa(b_j))$ is the probability of a true boundary given parent image regions and is learned from the data. Searching for the most probable explanation of b, y nodes given the measurements M_b results in true boundary nodes of y^* and b^*:

$$b^*, y^* = \arg\max_{b,y} P(b, y|M_b) = \arg\max_{b,y} P(b, y, M_b)$$

2.4. Unifying the DRF model of axon regions and the BN of axon boundaries

Unifying the DRF model of all axon regions y given their measurements x and and the BN model of all axon boundaries given their axon regions and boundary measurements, the joint probability of image measurement x, all axon region nodes $y = \{y_i\}$, all axon boundary nodes $b = \{b_i\}$ and all axon boundary measurements $M_b = \{M_{bi}\}$ is given by:

$$\begin{aligned} P(x, y, b, M_b) &= P(M_b|b)P(b|y)P(y|x)P(x) \quad (6) \\ &= \frac{1}{Z} \prod_{j=1}^{m} P(M_{b_j}|e_j)P(b_j|pa(b_j)) \\ &\quad \prod_{i \in V} \phi(y_i, x_i) \prod_{i \in V} \prod_{j \in N_i} \exp\left(y_i y_j \lambda^T g_{ij}(x)\right) \end{aligned}$$

3. RESULTS

The probability density map of image pixels in Fig. 1 being an axon, using the DRF framework is shown in Fig. 2.

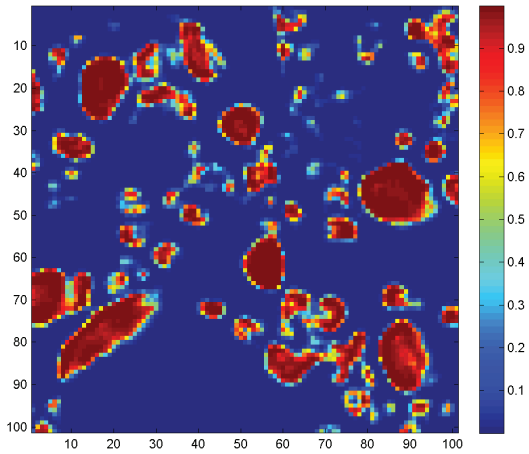


Fig. 2. Axon density map using the DRF model

The distribution of curvature boundary features show good linear separability for axon boundary labels vs. non axon boundary labels as shown in Fig. 3. These features when added to gradient features provide even greater separability among axon boundary vs. non axon boundary labels.

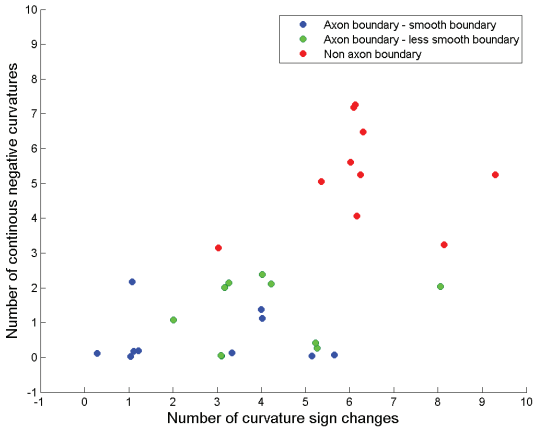


Fig. 3. Number of curvature sign changes vs. curvelengths with continuous negative curvature for axon boundary vs. non axon boundary nodes

4. CONCLUSION AND FUTURE WORK

In this paper we use DRF frameworks to find probability maps for axon regions and axon boundaries separately. The results show that both maps show good visual correspondence with the image data. However, combining the two set of data will lead to much better results since we use both models jointly, just as people do when they visually segment these types of data. Our next step is to perform true inference labeling of both BN boundary model and the joint model and compare the results with manual segmentations.

5. REFERENCES

[1] F. N. Golabchi, D. H. Brooks, W. S. Hoge, U. De Girolami, and S. E. Maier, "Pixel-based comparison of spinal cord mr diffusion anisotropy with axon packing parameters," *Magn Reson Med.*, vol. 63, no. 6, pp. 1510–1519, June 2010.

[2] O Cuisenaire, E Romero, C Veraart, and B Macq, "Automatic segmentation and measurements of axons in microscopic images," in *SPIE Medical Imaging*, Kenneth M. Hanson, Ed., May 1999, vol. 3661, pp. 920–929.

[3] D. Comaniciu and P. Meer, "Mean shift a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[4] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[5] T.F. Chan and L.A. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 266–277, Feb 2001.

[6] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888 – 905, August 2000.

[7] S. L. Lauritzen, *Graphical models*, Oxford Univ. Press, 1996.

[8] C.M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[9] L. Zhang and Q. Ji, "Image segmentation with a unified graphical model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1406–1425, August 2010.

[10] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Int'l Conf. Machine Learning*, 2001, pp. 282–289.

[11] S. Z. Li, *Markov random field modeling in image analysis*, Springer, 2001.

[12] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. of Computer Vision*, vol. 68, no. 2, pp. 179–202, 2006.