Electron Microscopy Image Segmentation with Graph Cuts Utilizing Estimated Symmetric Three-Dimensional Shape Prior

Huei-Fang Yang* and Yoonsuck Choe

Department of Computer Science and Engineering
Texas A&M University
College Station, TX 77843-3112
hfyang@cse.tamu.edu, choe@tamu.edu

Abstract. Understanding neural connectivity and structures in the brain requires detailed three-dimensional (3D) anatomical models, and such an understanding is essential to the study of the nervous system. However, the reconstruction of 3D models from a large set of dense nanoscale microscopy images is very challenging, due to the imperfections in staining and noise in the imaging process. To overcome this challenge, we present a 3D segmentation approach that allows segmenting densely packed neuronal structures. The proposed algorithm consists of two main parts. First, different from other methods which derive the shape prior in an offline phase, the shape prior of the objects is estimated directly by extracting medial surfaces from the data set. Second, the 3D image segmentation problem is posed as Maximum A Posteriori (MAP) estimation of Markov Random Field (MRF). First, the MAP-MRF formulation minimizes the Gibbs energy function, and then we use graph cuts to obtain the optimal solution to the energy function. The energy function consists of the estimated shape prior, the flux of the image gradients, and the gray-scale intensity. Experiments were conducted on synthetic data and nanoscale image sequences from the Serial Block Face Scanning Electron Microscopy (SBFSEM). The results show that the proposed approach provides a promising solution to EM reconstruction. We expect the reconstructed geometries to help us better analyze and understand the structure of various kinds of neurons.

1 Introduction

Understanding neural connectivity and functional structure of the brain requires detailed 3D anatomical reconstructions of neuronal models. Recent advances in high-resolution three-dimensional (3D) image acquisition instruments [1,2], Serial Block-Face Scanning Electron Microscopy (SBFSEM) [3] for example, provide sufficient resolution to identify synaptic connections and make possible the

* This work was supported in part by NIH/NINDS grant #1R01-NS54252. We would like to thank Stephen J. Smith (Stanford) for the SBFSEM data.

G. Bebis et al. (Eds.): ISVC 2010, Part II, LNCS 6454, pp. 322–331, 2010.

reconstruction of detailed 3D brain morphological neural circuits. The SBFSEM utilizes backscattering contrast and cuts slices off the surface of the block by a diamond knife, generating images with a resolution in the order of tens of nanometers. The lateral (x-y) resolution can be as small as $10 - 20$ nm/pixel, and the sectioning thickness (z-resolution) is around 30 nm. The high imaging resolution allows researchers to identify small organelles, even to trace axons and to identify synapses, thus enabling reconstruction of neural circuits. With the high image resolution, the SBFSEM data sets pose new challenges: (1) Cells in the SBFSEM image stack are densely packed, and the enormous number of cells make manual segmentation impractical, and (2) the inevitable staining noise, the incomplete boundaries, and inhomogeneous staining intensities increase the difficulty in the segmentation and the subsequent 3D reconstruction and visualization.

To reconstruct neural circuits from the SBFSEM image volumetric data, segmentation, that is partitioning an image into disjoint regions, is a fundamental step toward a fully neuronal morphological model analysis. Segmentation of the SBFSEM images amounts to delineating cell boundaries. Different approaches have been proposed in the literature for such a task. Considering segmentation as the problem of restoring noisy images, Jain *et al.* [4] proposed a supervised learning method which trained a convolutional network with $34,000$ free parameters to classify each voxel as being inside or outside a cell. Another similar approach in the machine learning paradigm was proposed by Andres *et al.* [5]. These methods require data sets with ground truth, and creation of labeled data sets is labor-intensive. Semi-automatic tracking methods utilizing level-set formulation [6] or active contours [7] have also been investigated. Computation of level-set method is expensive, and the solution can sometimes get stuck in local minima.

In this paper, we propose a 3D segmentation framework with estimated 3D symmetric shape prior. First, different from other methods which derive the shape prior in an offline phase, the shape prior of the objects is estimated directly by extracting medial surfaces of the data set. Second, the 3D image segmentation problem is posed as Maximum A Posteriori (MAP) estimation of Markov Random Field (MRF). First, the MAP-MRF formulation minimizes the Gibbs energy function, and then we use graph cuts to obtain the optimal solution to the energy function. The energy function consists of the estimated shape prior, the flux of the image gradients, and the gray-scale intensity.

2 Graph Cut Segmentation

Image segmentation is considered as a labeling problem that involves assigning image pixels a set of labels [8]. Taking an image I with the set of pixels $\mathcal{P} = \{1, 2, ..., M\}$ and the set of labels $\mathcal{L} = \{l_i, l_2, ..., l_K\}$, the goal of image segmentation is to find an optimal mapping $X : \mathcal{P} \rightarrow \mathcal{L}$. According to the random field model, the set of pixels $\mathcal{P}$ is associated with random field $X = \{X_p : p \in \mathcal{P}\}$, where each random variable X_p takes on a value from the set of labels $\mathcal{L}$. A possible labeling $x = \{X_1 = x_1, X_2 = x_2, ..., X_M = x_M\}$, with $x_p \in \mathcal{L}$, is called a

configuration of X. Each configuration represents a segmentation. Finding the optimal labeling x^* is equivalent to finding the maximum a posteriori (MAP) estimate of the underlying field given the observed image data D:

$$x^* = \text{argmax}_{x \in X} \Pr(x \mid D). \tag{1}$$

By Bayes' rule, the posterior is given by:

$$\Pr(x \mid D) \propto \Pr(D \mid x) \Pr(x), \tag{2}$$

where $\Pr(D \mid x)$ is the likelihood of D on x, and $\Pr(x)$ is the prior probability of a particular labeling x, being modeled as a Markov random field (MRF) which incorporates contextual constraints based on piecewise constancy [8]. An MRF satisfies the following two properties with respect to the neighborhood system $\mathcal{N} = \{N_p \mid p \in \mathcal{P}\}$:

$$Positivity: \quad \Pr(x)\ 0, \qquad \forall x \in \mathcal{X}, \tag{3}$$

$$Markovianity: \Pr\left(x_p \mid x_{\mathcal{P}-\{p\}}\right) = \Pr\left(x_p \mid x_{\mathcal{N}_p}\right), \forall p \in \mathcal{P}. \tag{4}$$

Furthermore, according to Hammersley-Clifford theorem [9], a random field with Markov property obeys a Gibbs distribution, which takes the following form:

$$\Pr(x) = \frac{1}{Z} \exp(-E(x)), \tag{5}$$

where Z is a normalizing constant called the partition function, and $E(x)$ is the Gibbs energy function, which is:

$$E(x) = \sum_{c \in \mathcal{C}} V_c(x_c), \tag{6}$$

where $\mathcal{C}$ is the set of cliques, and $V_c(x_c)$ is a clique potential. Taking a log likelihood of Equation 2, the MAP estimate of $\Pr(x \mid D)$ is equivalent to minimizing the energy function:

$$-\log \Pr(x \mid D) = E(x \mid D) = \sum_{p \in \mathcal{P}} V_p(x_p \mid D) + \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p} V_{pq}(x_p, x_q \mid D), \tag{7}$$

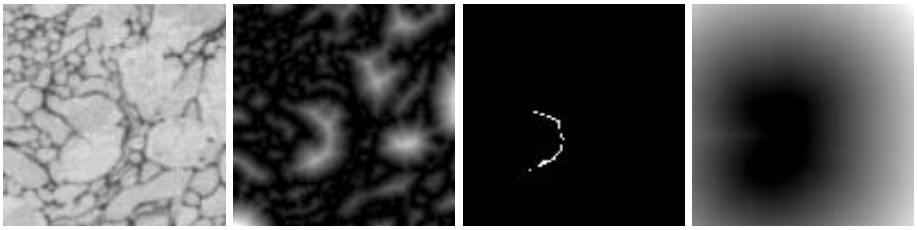
where $V_p(x_p \mid D)$ and $V_{pq}(x_p, x_q \mid D)$ are the unary and piecewise clique potentials, respectively.

Minimizing the energy function $E(x \mid D)$ is NP-hard, and the approximate solution can be obtained by graph-cuts using α-expansion algorithm [10]. The graph cuts represent an image as a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ with a set of vertices (nodes) $\mathcal{V}$ representing pixels or image regions and a set of edges $\mathcal{E}$ connecting the nodes. Each edge is associated with a nonnegative weight. The set $\mathcal{V}$ includes the nodes of the set $\mathcal{P}$ and two additional nodes, the source s and the sink t. All nodes $p \in \mathcal{V}$ are linked to the terminals s and t with weight w_{sp} and w_{pt},

respectively. Edges between the nodes and the terminals are called t-links, and edges between node p and its neighborhood q with weight w_{pq} are called n-links. The t-links and n-links model the unary and piecewise clique potentials, respectively. A cut $\mathcal{C}$ is a subset of edges $\mathcal{E}$ that separates terminals in the induced graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} - \mathcal{C} \rangle$ and thus partitions the nodes into two disjoint subsets while removing edges in the cut $\mathcal{C}$. The partitioning of a graph by a cut corresponds to a segmentation in an image. The cost of a cut, denoted as $|\mathcal{C}|$, is the sum of the edge weights in $\mathcal{C}$. Image segmentation problem then turns into finding a minimum cost cut that best partitions the graph, which can be achieved by the min-cut/max-flow algorithm [11]. One criterion of minimizing the energy function by graph cuts is that $V_{pq}(x_p, x_q)$ is submodular, that is, $V_{pq}(0,0) + V_{pq}(1,1) \leq V_{pq}(1,0) + V_{pq}(0,1)$ [10].

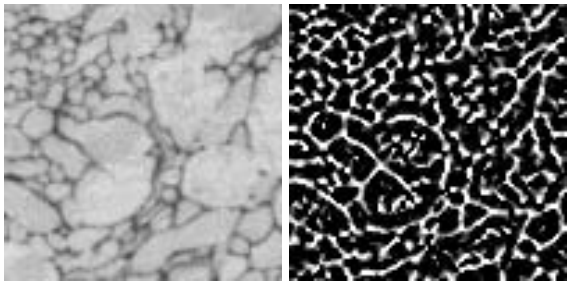
3 Symmetric Shape Prior Estimation

Anatomical structures, such as axons, dendrites, and soma, exhibit locally symmetric shapes to the medial axis which is also referred to as the skeleton and is commonly used for shape representation. The medial axis of a 3D object is generally referred to as the medial surface. Extracting medial surface approaches include distance field based methods [12], topological thinning, gradient vector flow methods [13] and others [14]. We follow the method proposed by Bouix *et al*. [12] to extract the medial surface. Gray-scale images are first converted to binary images, and a Euclidean distance function to the nearest boundary is computed at each voxel, as shown in Figure 1(b). Guiding the thinning procedure by exploiting properties of the average outward flux of the gradient vector field of a distance transform, the resulting medial surface for a particular object is shown in Figure 1(c). Finally, the estimated shape is obtained by first expanding each point in the extracted medial surface with the shortest distance to the



(a) Original image (b) Distance map (c) Skeleton (d) Shape prior

Fig. 1. Method of shape prior estimation. (a) is an image drawn from the input stack. (b) is the distance map computed from the binary image stack. (c) shows the extracted skeleton (white curves) from the distance map. (d) shows the estimated shape prior. Dark is the expanded region, and bright indicates the points outside of the expanded region which are represented by a distance function.



(a) SBFSEM image (b) Associated flux of (a)

Fig. 2. Flux of the gradient vector fields of a slice from the image stack. (a) shows part of an original gray-scale intensity image from the SBFSEM stack. (b) is the associated flux of image gradients of (a), where the foreground objects have negative flux (dark), and the background objects have positive flux (bright).

boundary. The points outside the expanded region are represented by a distance function:

$$D\,(p) = \|p - s_p\|, \tag{8}$$

where $\|p - s_p\|$ represents the Euclidean distance from p to the nearest pixel s_p in the expanded region. Shown in Figure 1(d) is the estimated shape prior for the object in Figure 1(c). The estimated shape prior is then incorporated in the unary term (Equation 11) acting as a constraint in the minimization process.

4 Definition of Unary Potential

A unary term defines the cost that a node is assigned a label x_p, that is, the corresponding cost of assigning node p to the foreground or the background. In our segmentation framework, a unary term consists of two parts: the flux of the gradient vector field and the shape prior estimated in section 3.

4.1 Flux

Flux has recently been introduced by Vasilevskiy and Siddiqi [15] into image analysis and computer vision. They incorporated flux into level-set method to segment blood vessel images. After that, flux has also been integrated into graph cuts [16] [17] to improve the segmentation accuracy. The introduction of flux into graph cuts can reduce the discretization artifacts which is a major shortcoming in graph cuts [16]. By definition, considering a vector field v defined for each point in $\mathcal{R}^3$, the total inward flux of the vector field through a given continuous hypersurface S is given by the surface integral [15]:

$$F(S) = \int_S \langle N, v \rangle \, dS, \tag{9}$$

where $\langle , \rangle$ is the Euclidean dot product, N are unit inward normals to surface element dS consistent with a given orientation. In the implementation of Equation 9, the calculation of the flux is simplified by utilizing the divergence theorem which states that the integral of the divergence of a vector field v inside a region equals to the outward flux through a bounding surface. The divergence theorem is given by:

$$\int_R \text{div v } dR = \oint_S \langle N, v \rangle \, dS, \tag{10}$$

where R is the region. For the numerical implementations, we consider the flux through a sphere in the case of 3D, and v is defined as the normalized (unit) image gradient vector field of the smoothed volume I_σ, $\frac{\bigtriangledown I_\sigma}{\|\bigtriangledown I_\sigma\|}$. Figure 2(b) shows the flux of gradient vector fields of Figure 2(a). Note that the flux is computed in 3D but only 2D case is shown here. The foreground object has negative flux (dark) whereas the background has positive flux (bright).

4.2 Incorporating Flux and Shape Prior

Combining the flux of gradient vector fields and the estimated shape prior yields a new unary term. Inspired by [17], we assigned the edge weights between node p and terminals s and t as:

$$\begin{aligned} w_{sp} &= -\min\left(0, F\left(p\right)\right), \\ w_{pt} &= \max\left(0, F\left(p\right)\right) + \alpha D\left(p\right), \end{aligned} \tag{11}$$

where $F\left(p\right)$ denotes the flux at point p, and α is a positive parameter adjusting the relative importance of the shape prior $D\left(p\right)$. In our experiments, the value of α was set to 0.2.

5 Definition of Piecewise Potential

In the SBFSEM images, the foreground and background can be discriminated by their gray-scale intensities. Compared to the background, the foreground objects usually have higher intensity values. Boundaries can thus be determined if the intensity differences between points are large. To capture the boundary discontinuity between pixels, the weight between node p and its neighbor q is defined as [18]:

$$w_{pq} = \exp\left(-\frac{\left(I_p - I_q\right)^2}{2\sigma^2}\right) \cdot \frac{1}{\|p - q\|}, \tag{12}$$

where I_p and I_q are point intensities ranging from 0 to 255, $\|p - q\|$ is the Euclidean distance between p and q, and σ is a positive parameter set to 30. Equation 12 penalizes a lot for edges with similar gray-scale intensities while it penalizes less for those with larger gray-scale differences. In other words, a cut is more likely to occur at the boundary, where the edge weights are small. For 3D images, the 6-, 18-, or 26-neighborhood system is commonly used. Here, the 6-neighborhood system was used in our experiments.

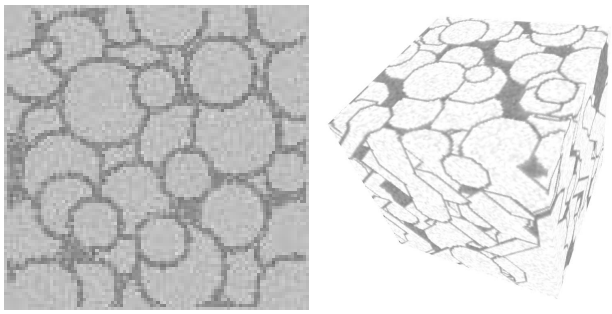
6 Experimental Results

Experiments were conducted on synthetic data sets and an SBFSEM image stack in order to evaluate the performance of the proposed approach.

6.1 Synthetic Data Sets

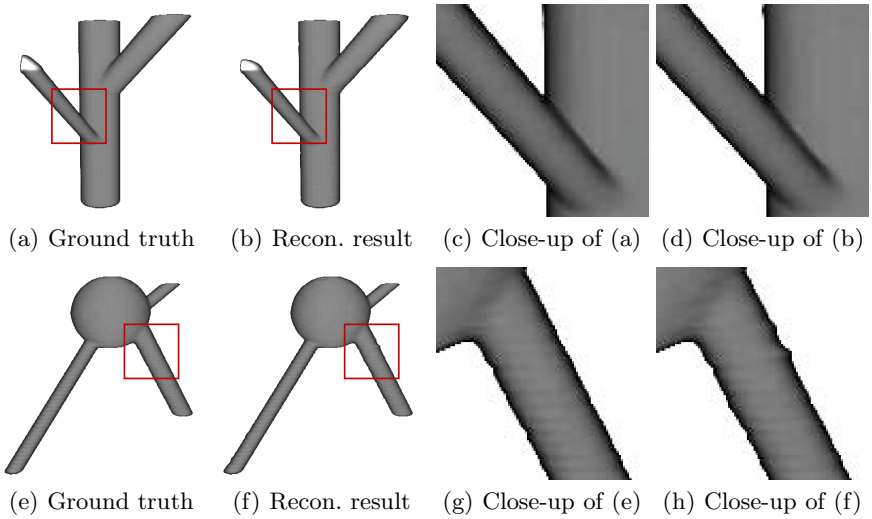
The synthetic data sets consisted of two image stacks, each of which having the size of $100 \times 100 \times 100$. Gaussian noise was added to each image slice to simulate noise during image acquisition process. Here, three different levels of Gaussian noise with standard deviation $\sigma = 0.0447, 0.0632$, and 0.0775 were added to the two synthetic data sets, thus resulting in a total of 6 image stacks. Shown in Figure 3(a) is a noisy image slice from the synthetic image stack in Figure 3(b). The reconstruction results of the two synthetic image stacks are shown in Figure 4(b) and Figure 4(f), respectively, and their ground truth is given in Figure 4(a) and Figure 4(e) accordingly. As can been seen from the close-up comparisons of the reconstruction results and the ground truth, the reconstruction results are almost identical to the ground truth with minor differences.

To quantitatively measure the performance of the proposed segmentation method, we used the F-measure, $F = \frac{2PR}{P+R}$, where P and R are the precision and recall of the segmentation results relative to the ground truth. More specifically, let Z be the set of voxels of the obtained segmentation results and G be the ground truth, then $P = \frac{|Z \cap G|}{|Z|}$ and $R = \frac{|Z \cap G|}{|G|}$, where $|\cdot|$ is the number of voxels. The average precision and recall values of the reconstruction results of the 6 synthetic image stacks were 0.9659 and 0.9853, respectively, yielding an average of F-measure being 0.9755. We also applied Dice coefficient (DC) [19] to measure the similarity between two segmentations. DC measures the overlapped



(a) An image from (b) (b) A synthetic image stack

Fig. 3. A synthetic image and a synthetic data set. (a) is a noisy image with 100×100 pixels selected from the synthetic image stack in (b). (b) shows one of the two synthetic image stacks, which contains 100 images.



(a) Ground truth (b) Recon. result (c) Close-up of (a) (d) Close-up of (b)

(e) Ground truth (f) Recon. result (g) Close-up of (e) (h) Close-up of (f)

Fig. 4. Ground truth and reconstruction results of the synthetic data sets. (a) and (e) are the ground truth of the two synthetic data sets. (b) and (f) are the reconstruction results from the image stacks in which Gaussian noise with $\sigma = 0.04477$ was added. As can be seen from the close-up comparisons of the ground truth and the reconstruction results, the reconstruction results are almost identical to the ground truth with minor differences.

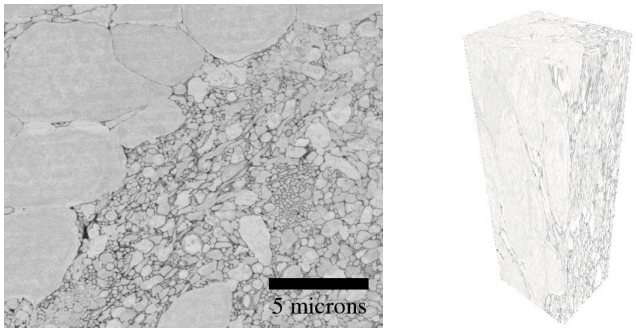
regions between the obtained segmentation results and the ground truth, defined as $\mathrm{DC} = \frac{2|Z \cap G|}{|Z|+|G|}$, where 0 indicates no overlap between two segmentations, and 1 means two segmentations are identical. The average DC value on the synthetic data sets was 0.9748, implicating the reconstruction results are almost identical to the ground truth.

The mean computation time using a Matlab implementation of the proposed approach for processing a synthetic image stack ($100 \times 100 \times 100$) on a standard PC with Core 2 Duo CPU 2.2 GHz and 2 GB memory was 20 seconds.

6.2 SBFSEM Image Stack

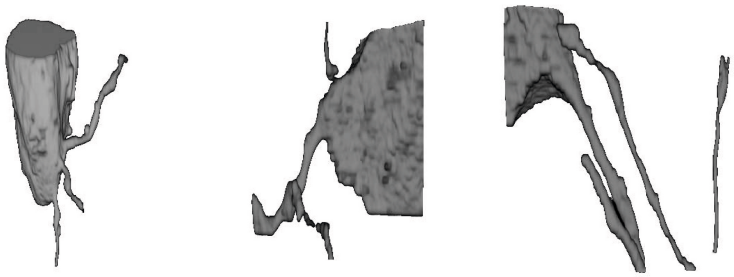
Experiments on the SBFSEM data were conducted on one image stack ($631 \times 539 \times 561$), on different parts (sub-volumes) of it. Figure 5(a) shows an EM image with 631×539 pixels from the larval zebrafish tectum volumetric data set, shown in Figure 5(b). The reconstruction results of the proposed method are shown in Figure 6, where Figure 6(a) and Figure 6(b) show parts of neurons, and Figure 6(c) shows the elongated structures.

For the validation of the proposed algorithm, we manually segmented a few neurons using TrakEM2 [20], serving as the ground truth. Again, F-measure and DC were used as the evaluation metrics. The average precision and recall values of the reconstruction results shown in Figure 6 were 0.9660 and 0.8424,



(a) An EM image from (b) (b) The SBFSEM image stack

Fig. 5. An image and a volumetric SBFSEM data set. (a) shows an EM image with 631×539 pixels from the SBFSEM stack in (b). Note that cells in the SBFSEM images are densely packed. (b) is the SBFSEM image stack of larval zebrafish optic tectum, containing 561 images.



(a) (b) (c)

Fig. 6. Reconstruction results of the proposed method. (a) and (b) show parts of neurons, and (c) shows the elongated axon structures.

respectively, and thus the average of F-measure was 0.9. The average DC value of the reconstruction results was 0.8918, showing that the proposed method can reconstruct the neuronal structures from the SBFSEM images.

7 Conclusion and Future Work

We presented a 3D segmentation method with estimated shape prior for the SBFSEM reconstruction. The shape prior was estimated directly from the data set based on the local symmetry property of anatomical structures. With the help of the shape prior along with the flux of image gradients and image gray-scale intensity, the proposed segmentation approach can reconstruct neuronal structures from densely packed EM images. Future work includes applying the method to larger 3D volumes and seeking a systematically quantitative and qualitative validation method for the SBFSEM data set.

References

1. Helmstaedter, M., Briggman, K.L., Denk, W.: 3D structural imaging of the brain with photons and electrons. Current Opinion in Neurobiology 18, 633–641 (2008)
2. Briggman, K.L., Denk, W.: Towards neural circuit reconstruction with volume electron microscopy techniques. Current Opinion in Neurobiology 16, 562–570 (2006)
3. Denk, W., Horstmann, H.: Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biology 2, e329 (2004)
4. Jain, V., Murray, J.F., Roth, F., Turaga, S., Zhigulin, V.P., Briggman, K.L., Helmstaedter, M., Denk, W., Seung, H.S.: Supervised learning of image restoration with convolutional networks. In: Proc. IEEE Int'l Conf. on Computer Vision, pp. 1–8 (2007)
5. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of sbfsem volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
6. Macke, J.H., Maack, N., Gupta, R., Denk, W., Schölkopf, B., Borst, A.: Contour-propagation algorithms for semi-automated reconstruction of neural processes. J. Neuroscience Methods 167, 349–357 (2008)
7. Jurrus, E., Hardy, M., Tasdizen, T., Fletcher, P., Koshevoy, P., Chien, C.B., Denk, W., Whitaker, R.: Axon tracking in serial block-face scanning electron microscopy. Medical Image Analysis 13, 180–188 (2009)
8. Li, S.Z.: Markov random field modeling in image analysis. Springer, New York (2001)
9. Hammersley, J.M., Clifford, P.: Markov field on finite graphs and lattices (1971)
10. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26, 147–159 (2004)
11. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23, 1222–1239 (2001)
12. Bouix, S., Siddiqi, K., Tannenbaum, A.: Flux driven automatic centerline extraction. Medical Image Analysis 9, 209–221 (2005)
13. Hassouna, M.S., Farag, A.A.: Variational curve skeletons using gradient vector flow. IEEE Trans. Pattern Anal. Mach. Intell. 31, 2257–2274 (2009)
14. Gorelick, L., Galun, M., Sharon, E., Basri, R., Brandt, A.: Shape representation and classification using the poisson equation. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1991–2005 (2006)
15. Vasilevskiy, A., Siddiqi, K.: Flux maximizing geometric flows. IEEE Trans. Pattern Anal. Mach. Intell. 24, 1565–1578 (2002)
16. Kolmogorov, V., Boykov, Y.: What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In: Proc. IEEE Int'l Conf. Computer Vision, pp. 564–571 (2005)
17. Vu, N., Manjunath, B.S.: Graph cut segmentation of neuronal structures from transmission electron micrographs. In: Proc. Int'l Conf. Image Processing, pp. 725–728 (2008)
18. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. Int'l J. Computer Vision 70, 109–131 (2006)
19. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26, 297–302 (1945)
20. Cardona, A., Saalfeld, S., Tomancak, P., Hartenstein, V.: TrakEM2: open source software for neuronal reconstruction from large serial section microscopy data. In: Proc. High Resolution Circuits Reconstruction, pp. 20–22 (2009)