A GEOMETRIC-STATISTICAL APPROACH TOWARD NEURON MATCHING

Suvadip Mukherjee[1], *Saurav Basu*[3], *Barry Condron*[2], *Scott T. Acton*[1]

[1]Electrical & Computer Engineering, Biomedical Engineering, University of Virginia, Charlottesville, VA USA
[2]Biology, University of Virginia, Charlottesville, VA USA
[3]Carnegie Mellon University, Pittsburgh, PA USA

ABSTRACT

Where the genome project mapped the genetic structure of complicated organisms such as the mouse, those pursuing the neurome are seeking the same for the neural anatomy. In this massive biological investigation, the tools of imaging and biological experimentation are outpacing the requisite tools in image analysis. In terms of comparing neurons, based on the geometrical structure and features within the structure, the accepted approaches are largely manual. In this paper, we propose a combined geometric-statistical approach toward automated neuron matching. We utilize the geometric information of a neuron and compute a pairwise distance histogram based on the geometric information, to find a similarity measure between neurons. The distribution function is so chosen such that it reflects the structural pattern of a set of neuronal points, and is rotationally invariant. Preliminary experiments on a set of three different classes of neurons, with six neurons in each class, provides evidence of efficacy, with the best two matches to a given query producing a retrieval error of 0% and the third best match producing an error of only 5.55%. In future work, the proposed method can be used as a component in the retrieval of similar neurons from neuronal database.

Index Terms— Biological image analysis, neural imaging, microscopy, content-based image retrieval

1. INTRODUCTION

A century after Cajal, the goal of exploring neural computation from anatomy is a quite active area. In fact, with the complete neural atlas, or *neurome*, of simple organisms such as the worm *C. Elegans* in hand, the research community is eyeing the more complex neuromes of the fly *Drosophilia* and perhaps, in the not so distant future, the neurome of the mouse.

Approaching the neurome of more complicated organisms containing hundreds of thousands and millions of neurons, success will hinge upon automated processing from the image analysis community. Essentially, two unsolved problems in automated processing stand in between the current state-of-the-art and automated morphological and connectivity analysis of the neurobiology of more advanced organisms. The first problem is one of segmentation – computing the structure of a neuron or neurite from a stack of confocal images. This problem has received significant attention [1]-[5]. Solutions to the segmentation/tracing problem have borrowed from genetic algorithms [5], statistical analysis [1], wavelet analysis [3], active contours [2] and graph theory [6].

The second barrier to full automation is the classification or matching of neurons. Although some preliminary work on this topic has been reported [7], there currently exists no demonstrated method of matching the 3D images of two neurons. This paper provides a first step to computing a cost function that could potentially be used in matching imaged neurons to those already classified or categorized. Such a match function can be employed in the retrieval of similar neurons, the quantification of environmental or genetic differences and the creation of "average" neurons in an atlas.

2. NEURON SEGMENTATION

Our vision for the automated image analysis of neuron structure and connectivity includes (1) segmentation (equivalently, *tracing*) of the neurons and (2) matching of the neuron images by way of a match measure. As this paper focuses on the latter topic of matching, we will only summarize the segmentation method applied.

We call our segmentation approach Tree2Tree [6] as it attempts to map graph theoretic trees to the dendritic tree and remaining neural structure. The segmentation is accomplished by piecing together medial trees that emerge from the coarse segmentation of the confocal stack.

As clutter and poor contrast plague such analysis, we first enhance the imagery using a Hessian-motivated technique. Our hypothesis is that neurons will be composed of tube-like structures rather than isolated blobs or sheets. At a given point $\mathbb{p}$, we have a 3D Hessian (matrix of second partial derivatives) with eigenvalues $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$. To capture tubular objects, we desire $|\lambda_1| \approx 0, |\lambda_1| \ll |\lambda_2|$ and $\lambda_2 \approx \lambda_3$. Further, the neurons are brighter than the background such that $\lambda_2, \lambda_3 \leq 0$. Therefore, we transform the intensities in the confocal stack at a given point $\mathbb{p}$ (at Gaussian scale σ) to:

$$\mathcal{N}_\sigma(\mathbb{p}) = \begin{cases} \dfrac{|\lambda_1(\mathbb{p}) - \lambda_2(\mathbb{p})|^2}{(|\lambda_1(\mathbb{p})||\lambda_2(\mathbb{p}) - \lambda_3(\mathbb{p})|) + \epsilon} & \text{when } \lambda_2(\mathbb{p}), \lambda_3(\mathbb{p}) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

For the confocal stack shown in Fig. 1(a), the corresponding Hessian enhancement via (1) is shown in Fig. 1(b). Following enhancement, the image stack is adaptively thresholded (Fig. 1(c)) and reduced to the 3D medial axis trees (Fig. 1(d)). For the skeletonization, a variant of the algorithm in [8] is applied.

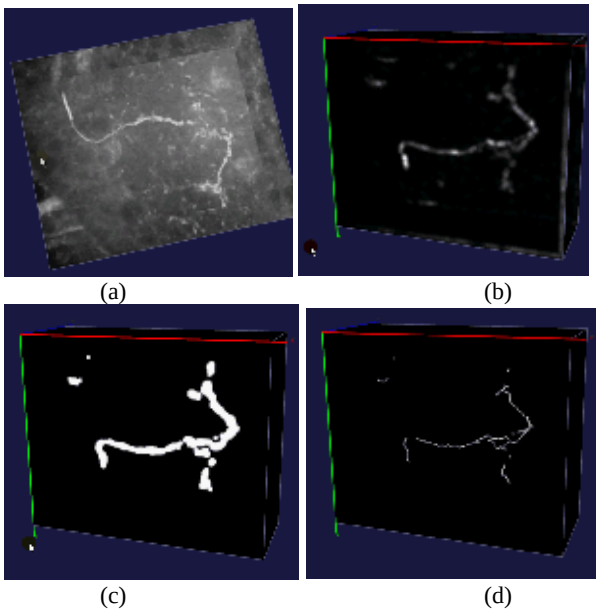

Fig. 1: The initial steps of neuron segmentation using Tree2Tree. (a) Original 3D GFP confocal stack of a *Drosophilia* ventral nerve cord (VNC) neuron. (b) After enhancement by Hessian approach. (c) After adaptive thresholding; (d) After 3D skeletonization.

The remaining steps in the segmentation begin with forming a K-nearest neighbor graph and a minimum spanning tree based of a distance measure that exploits both the distance between unconnected medial trees (skeletonized fragments of the neuron) and the angular deviation between the connected leaf tangents. Then, a filtering process [6] is applied to maximize the use of computed segments while minimizing false positive segments due to clutter. A typical segmentation result is shown in Fig. 2.

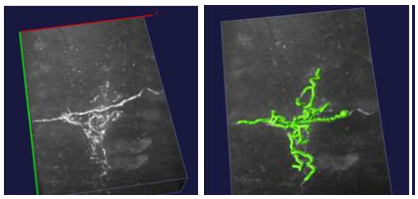

Fig. 2: *Drosophilia* VNC neuron (left) and Tree2Tree segmentation (right).

Our segmentation has the advantage of full automation requiring no seed points or manual interaction. Following segmentation, we wish to match neurons to classified neurons in our database. Moreover, we seek to quantify the morphological difference in two neurons of the same class under varied environmental or genetic scenarios.

3. NEURON MATCHING

The utility of a quantitative comparison between two neurons is enormous. The Path2Path algorithm, developed by Basu, Condron and Acton [7] generates a similarity measure between neurons based on a path matching and morphological comparison method. In this paper, we propose a geometric-statistical approach toward neuron matching.

3.1 Description of a neuronal path

As our graph theoretic segmentation exhibits, a neuron can be considered as a set of connected paths. In terms of graph theory, it is an undirected tree, with a set of vertices $V = \{v_1, \ldots, v_n\}$ and a set of edges $E = \{e_1, \ldots, e_{n-1}\}$. Here, the node v_1 is the root of the neuronal tree. Note that these vertices are the points of bifurcation of neuronal branches. In the discrete domain, an edge e_i between vertices $\{v_i, v_{i+1}\}$ is a collection of an ordered set of points $P_i = \{p_1^i, \ldots, p_k^i\}$, where $p_1^i = v_i$ and $p_k^i = v_{i+1}$. We define the term *parent* of a point $p_m \in P_i$ as follows:

$$parent(p_m^{\,i}) = \begin{cases} p_m^{\,i} & if\ p_m^{\,i} = p_1^{\,i} \\ p_{m-1}^{\,i} & if\ p_m^{\,i} \neq p_1^{\,i} \end{cases}$$

Thus, in the discrete domain, each path can be considered to be a finite collection of points $p_i = \{i, parent(p_i)\}$. In this paper, we refer to each point in a neuron path as a *path-point*, and we refer to the bifurcation points as *nodes.*

3.2 Hierarchy of a path-point

Let e_i be a path in the neuronal tree between the nodes $\{v_i, v_{i+1}\}$. Let P_i be the set of path points for the corresponding path e_i. Let $P_i = \{p_1^i, \ldots, p_k^i\}$, where, $p_1^i = v_1$ and $p_k^{\,i} = v_{i+1}$.

We can define the term *hierarchy* (H) of a path-point p_k^i as follows:

$$H(p_k^{\,i}) = \begin{cases} \rho(p_k^{\,i}) + H(p_k^{\,i}) - 1 & if\ \rho(p_k^{\,i}) = 0 \\ \rho(p_k^{\,i}) & if\ p_k^{\,i} = v_1 \\ H(p_k^{\,i}) & otherwise \end{cases}$$

Here, $\rho(p_k^{\,i})$ = number of points p_m such that $arent(p_m) = p_k^{\,i}$. Hence, we assign another property to every path-point p_i which can now be represented by the triplet $(i, parent(p_i), H(p_i))$.

3.3 Shape distributions for matching neurons

The task of developing a similarity measure for neurons is non-trivial due to large structural dissimilarity of neurons. The method in [9] uses a method of discriminating 3D objects by comparing their shape distributions. A neuron can also be perceived as a 3D object, with each path-point $\{p_i\} \in \mathbb{R}^3$.

Hence, we can compute the shape distribution of a set of path-points $P' = \{p_1, \dots, p_n\}$ by measurig the pairwise Euclidean distance between the points $\{p_i, p_j\}$ $(1 \leq i,j \leq n)$.We get a total of C_2^n distance values, which can be normalized to reduce each value in [0,1].

Let $\{x_1, \dots, x_m\}$ be the set of pairwise distances obtained by the above method $(m = C_2^n)$. We can now obtain a pairwise distance histogram of the obtained distances$\{x_1, \dots, x_m\}$. This histogram is normalized between [0,1] to create the normalized histogram of pairwise distances of the set of points $P' = \{p_1, \dots, p_n\}$. This normalized histogram can be thought of as an approximate probability distribution function of the random variable $X = \{x_1, \dots, x_m\}$, where each $x_j \in X$ denotes the pairwise Euclidean distance between any pair of points $\{p_i, p_j\} \in P'$.

The normalized histogram of the pairwise distances of the set of points provides some important information about the orientation and closeness of the points. Fig. 3(a) and 3(b) show two typical neurons from the online neuron database http://neuromorpho.org [10].

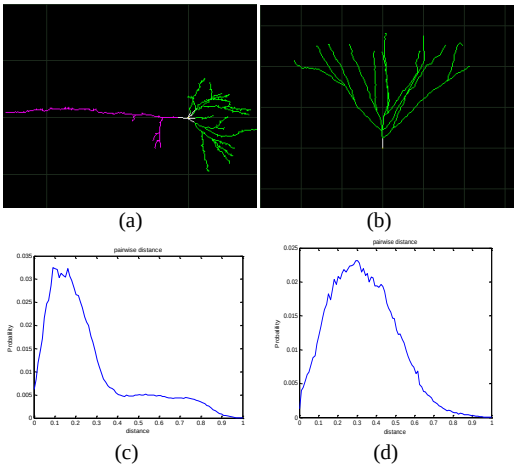

Fig. 3: (a) Human neuron. (b) Rat neuron
(c) Normalized pairwise distance histogram of (a). (d) Normalized pairwise distance histogram of (b).

Figure 3(c) and 3(d) show the shape distribution functions for the neurons 3(a) and 3(b) respectively. By visual inspection, we can see that the structural dissimilarity is reflected in the nature of the two distribution plots.

There are numerous methods proposed to measure the similarity between two probability density functions, and [11] provides a survey of such techniques. We have used the Bhattacharya distance [12] to measure the similarity between two probability distribution functions. The Bhattacharya distance between two discrete probability distributions $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_n\}$ is given by

$$d = -\ln\left(\sum_{i=1}^{n} \sqrt{p_i q_i}\right)$$

The Bhattacharya distance between P and Q is a measure of the amount of overlap between the two distributions. If the overlap between P and Q is low, it means the distributions are dissimilar and hence, the Bhattacharya distance between the two distributions is greater.

3.4 Neuron similarity

Despite the promising performance of the shape based neuronal matching technique, an approach based on global shape distribution matching often leads to erroneous results. This is mostly due to the fact that the global shape matching approach ignores the local geometrical structure of a neuron. We propose a method that combines the statistical approach mentioned in Section 3.3 with the information available from the geometrical structure of a neuron.

As mentioned in Section 3.2, any path-point p_i is represented by the unique triplet $(i, parent(p_i), H(p_i))$. Intuitively, we can think of two neurons to be of similar type if their path-point hierarchies and their orientation in space are similar. We thus devise the following method for neuron matching.

1. Image and segment a neuron *N* having *P* path-points.
2. Compute the hierarchy for all path-points for the input neuron *N*. Let $\{h_1, \dots, h_l\}$ be the distinct hierarchies of the neuron.
3. Compute the normalized hierarchy histogram $Hist(h_i)$, $(1 \leq i \leq l)$ such that $Hist(h_m)$ = number of path-points $p_j = (j, parent(p_j), h_m)$.
4. Let $h = \{h_1, \dots, h_p\}$ be such that $Hist(h_1) \leq Hist(h_2) \leq \dots \leq Hist(h_p)$ and $p \leq n$. In other words, $\{h_1, \dots, h_p\}$ is the ordered set of 'p' most populated hierarchies.
5. $P_1, .., P_p$ be set of points such that $p_j^i \in P_i$ iff. $H(p_j^i) = h_i$. For each set of points P_i $(1 \leq i \leq p)$, compute the normalized pairwise distance histogram as discussed in Sec 3.3. Denote the histograms by $\{\delta_1, \dots, \delta_p\}$.
6. Let M be a neuron in the database $(M \neq N)$, and we wish to find the similarity between N and M. Compute the pairwise distance histograms $\{\gamma_1, \dots, \gamma_p\}$ of M corresponding to the hierarchies$\{h_1, \dots, h_p\}$.
   If a hierarchy h_k $(h_k \in h)$ does not exist in M, $\gamma_k = 0$.
7. Compute the Bhattacharya distances $\{d_1, \dots, d_p\}$ between the pairwise distance histograms $\{\delta_1, \gamma_1\}, \dots, \{\delta_p, \gamma_p\}$.
8. Compute the matching score as follows:
   $\tau(N, M) = Hist(h_1)d_1 + \dots + Hist(h_p)d_p$

4. RESULTS AND DISCUSSION

We have applied the combined geometric-statistical algorithm to three different classes of neurons, each class having six neurons. The dataset, as described in Table 1, is obtained from the public neuron database neuromorpho.org.

Table 1: Neurons used in our study

Neuron	Archive	Brain Region	Cell Type
1-6	Lewis	Cerebral Cortex	Pyramidal
7-12	Stevens	Optic Nerves	Retinotectal
13-18	Cameron	Spinal Cord	Motoneuron

Table 2 shows the results obtained using the combined geometric-statistical measure described in Section 3. Here, we tabulate the results of using each 3D image as a query against the remaining neurons, with the three closest matches from the set of eighteen neurons in our database.

Table 2: Experimental Results

Neuron	Match 1	Match 2	Match 3
1	3	2	5
2	3	5	1
3	2	5	1
4	6	3	5
5	3	2	8
6	4	3	2
7	10	9	8
8	10	7	3
9	7	10	11
10	7	8	9
11	12	10	9
12	11	10	9
13	15	14	18
14	13	15	16
15	18	13	14
16	17	14	15
17	16	15	14
18	15	13	14

The first column of Table 2 shows the neuron identifiers. The second, third and fourth columns show the neuron in our database which has the minimum score with respect to the queried neuron. Formally, if i is the queried neuron ($1 \le i \le 18$), and m, n and p be the top three matches of i, then $\tau(i,m) \le \tau(i,n) \le \tau(i,p)$.

Since the set of neurons {1,…,6},{7,…,12} and {13,…,18} are obtained from three different classes (C_1,C_2 and C_3), ideally, if the closest match of a neuron $i \in C_j$ $(1 \le j \le 3)$ is k, then $k \in C_j$ $(1 \le i \le 18)$. If k be a top three match of a neuron i, then we say that our algorithm has retrieved the neuron k incorrectly if $i \in C_j$, but $k \notin C_j$.

From Table 2, we observe that no retrieval error for the first and second closest matches (Match 1 and Match 2). The third closest match (Match 3) has only one retrieval error, where neuron 5 is matched with neuron 8. Hence, the retrieval error in the third closest match is 5.55%.

5. CONCLUSION

We have presented a similarity measure between neurons using a combination of geometric and statistical features. Preliminary experiments reveal encouraging results. It is our opinion that a complete matching and comparison approach will incorporate several such methods in combination. Currently, we are attempting to incorporate this tool in a content based neuron retrieval system to retrieve similar neurons from a database.

6. REFERENCES

[1] K.A. Al-Kofahi, A. Can, S. Lasek, D. H. Szarowski, N. Dowell-Mesfin, W. Shain, J. N. Turner and B. Roysam, "Median based robust algorithms for tracing neurons from noisy confocal microscope images", *IEEE Trans. Information Technology in Biomedicine*, vol. 7, no. 4, pp 302-16, 2003.

[2] H. Cai, X. Xu, J. Lu, J. Lichtman, S. P. Yung and S. T. C. Wong, "Shape-constrained repulsive snake method to segment and track neurons in 3D microscopy images," in *Proc. 3rd IEEE Intl. Symposium on Biomedical Imaging*, pp. 538-541, 2006.

[3] A. Dima, M. Scholz and K. Obermayer, , "Automatic segmentation and skeletonization of neurons from confocal microscopy images based on the 3-D wavelet transform," *IEEE Trans. on Image Processing*, vol. 11(7), pp. 790- 801, 2002.

[4] M. L Narro, F. Yang, R. Kraft, C. Wenk, A. Efrat and L. L. Restifo, "NeuronMetrics: software for semi-automated processing of cultured neuron images", *Brain Res*., vol. 1138, pp. 57–75, 2007.

[5] A. Wolf, A. Herzog, S. Westerholz, B. Michaelis and T. Voigt, "Improving fuzzy-based axon segmentation with genetic algorithms", in *Proc. IEEE Congress on Evolutionary Computation*, pp. 1025-31, 2009.

[6] S. Basu, A. Aksel, B. Condron, S.T. Acton, "Tree2tree: neuron segmentation for generation of neuronal morphology," *Proc. IEEE International Symposium of Biomedical Imaging*, April 14-17, 2010.

[7] S. Basu, B. Condron, and S.T. Acton, "Path2Path: Hierarchical path-based analysis for neuron matching", *Proc. IEEE International Symposium of Biomedical Imaging*, pp. 996-999, 2011.

[8] S. Lobregt, P. W. Verbeek and F. C. A. Groen, "Three-dimensional skeletonization: principle and algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, pp. 75 - 77, 1980.

[9] R. Osada, T.A. Funkhouser, B. Chazelle, and D.P. Dobkin, "Matching 3D Models with Shape Distributions", in *Proc. Shape Modeling International*, 2001, pp.154-166.

[10] GA Ascoli, DE Donohue, M. Halavi (2007) NeuroMorpho.Org: a central resource for neuronal morphologies. *J Neurosci*., 27(35):9247-51.

[11] Sung-Hyuk Cha, "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions" *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, 2007.

[12] A. Bhattacharya, "On a measure of divergence between two statistical populations defined by probability distributions", *Bull. Calcutta Math. Soc*., vol. 35, pp. 99-109, 1943.