# 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

In the same vein as the genome project that mapped the genetic structure of complex organisms such as the mouse, those pursuing the neurome are seeking a map 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 11.2%. 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 quite an 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. A typical segmentation result is shown in Fig. 1. Tree2Tree requires no seed points or manual interaction.

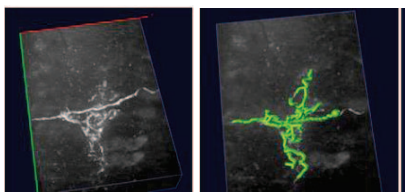


**Fig. 1:** Segmentation results.

Following segmentation, we wish to match neurons to classified neurons in our database. Additionally, we seek to quantify the morphological difference between 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 the graph theoretic segmentation of [7] 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 nodes $V = \{v_1, \dots, v_n\}$ and a set of edges $E = \{e_1, \dots, e_{n-1}\}$. 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 nodes $\{v_i, v_{i+1}\}$ is a collection of an ordered set of points $P_i = \{p_1^i, \dots, p_k^i\}$, 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:

$$parent(p_m^{\,i}) = \begin{cases} p_m^{\,i} & if\ p_m^{\,i} = v_1 \\ p_{m-1}^{\,i} & if\ p_m^{\,i} \neq v_1 \end{cases}$$

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

We can define the term *hierarchy* $(H)$ of a path-point $p_k^i \in P_i$ (where $P_i$ is the set of path points as defined in section 3.1) as follows:

$$H(p_k^{\,i}) = \begin{cases} 1 & if\ p_k^{\,i} = v_1 \\ H\left(parent(p_k^{\,i})\right) + 1 & \text{if } p_k^{\,i} \in \{v_2, , \dots, v_n\} \\ H(parent(p_k^{\,i})) & \text{otherwise} \end{cases}$$

Hence, we assign another property to every path-point $p_i$ which can now be represented by the triplet $(i, parent(p_i), H(p_i))$. Fig. 2 shows the hierarchies for a simple tree like structure.

### 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] employs 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 $\{p_i\} \in \mathbb{R}^3$.

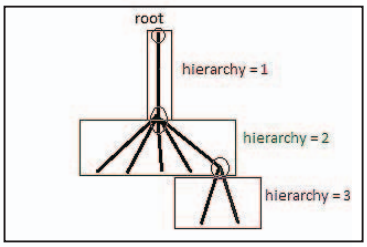



**Fig. 2:** A tree is shown, and the root and bifurcation nodes are circled. The path points enclosed by the rectangles have the same hierarchy. The path points belong to three different hierarchies – 1(red), 2(green) and 3(violet).

Hence, we can compute the shape distribution of a set of path-points $P' = \{p_1, \dots, p_n\}$ by measuring the pairwise Euclidean distance between the points $\{p_i, p_j\}$ $(1 \leq i, j \leq n)$. 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 mass 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].

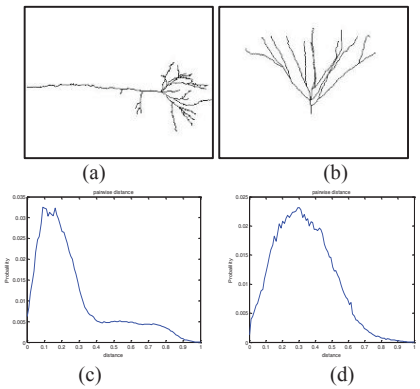



**Fig. 3:** (a) Human neuron. (b) Rat neuron (c) Normalized pairwise distance histogram of (a). (d) Normalized pairwise distance histogram of (b).

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 distributions. The Bhattacharya distance between two discrete probability distributions $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_n\}$ is given by

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

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 less, it means the distributions are dissimilar and hence, the Bhattacharya distance between them 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 may lead to spurious 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.

Our intuition in finding the degree of closeness between two neurons is as follows. In 3D Euclidean space, a neuron can be modeled as a collection of paths, with each path point belonging to a particular hierarchy. The collection of path points belonging to the same hierarchies constitutes a structure that is geometrically oriented in space. We refer to those structures as blocks. For example, in Fig. 2, all the path points enclosed by a rectangle belong to a block. As we can see, each such block is characterized by their geometric orientation and hence, by their shape distributions.

In order to compare two neurons, we examine the match of similarly populated blocks in the respective neurons. In other words, we compute the deformation energy required to register one block to another. A close match occurs if the most populated block of one neuron is well registered with the most populated block of the second neuron, and their hierarchies do not differ significantly. Once the densest block is examined for closeness, the process is repeated for the second most populated block and so on. Note that, the measure of closeness between the spatial orientations of the two blocks is obtained from our discussion of statistical shape distribution in section 3.3.

As shown in Fig. 4, the two neurons are close in match if the shape distributions of the two blocks (enclosed by rectangles) are similar. As is visually evident in Fig. 4(a), the candidate neurons differ in their geometry, whereas in Fig. 4(b), a match between the neurons is encountered.

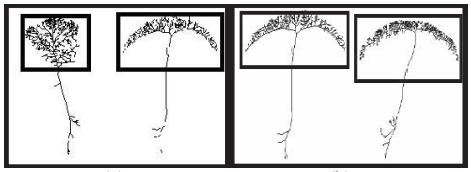



**Fig. 4:** Illustration of the blocks in the neuron structure

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 utilize the following method for neuron matching:

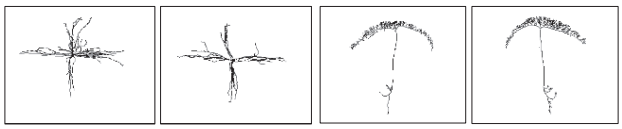
1. Image and segment a neuron *N* having *P* path-points.
2. Compute the hierarchy for all path-points for the input neuron *N*. Let $\{h_1, \dots, h_l\}$ be the distinct hierarchies of the neuron.
3. Compute the normalized hierarchy histogram $Hist(h_i)$, $(1 \le i \le l)$ such that $Hist(h_m)$ = number of path-points $p_j = (j, parent(p_j), h_m)$.
4. Let $h = \{h_1, \dots, h_p\}$ be such that $Hist(h_1) \ge Hist(h_2) \ge \dots \ge Hist(h_p)$ and $p \le n$. 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 $P_i$ $(1 \le i \le p)$, 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 $(M \ne N)$, 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 its '$p$' most populated hierarchies $g = \{g_1, \dots, g_p\}$.
   If a hierarchy $g_k$ $(g_k \in g)$ 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:

$$\tau(N, M) = \frac{d_1}{(e^{-\frac{|g_1 - h_1|}{\rho}} + \lambda)} + \dots + \frac{d_p}{(e^{-\frac{|g_p - h_p|}{\rho}} + \lambda)}$$

Note that in calculating the matching score in step 8, a higher score (thus a weaker match) is obtained if the shape distributions of two neurons are dissimilar, which is reflected by the Bhattacharya distance $(d_i)$ between the blocks. The matching score is also penalized if the hierarchies of the blocks differ considerably which is reflected in the exponential term. Here, $\rho$ is a parameter that determines the degree of penalty due to mismatch in hierarchies. Thus, a close match between two neurons *N* and *M* results in a small value of $\tau(N, M)$.

## 4. RESULTS AND DISCUSSION

We first demonstrate the results of our algorithm, anecdotally, by using it to measure the matching score between the neurons in Fig. 5, as tabulated in Table 1. In the table, the inter-class scores are much greater than the intra-class scores, reflecting the class separation.



(a) (b) (c) (d)

**Fig. 5: (a)-(b)** Neurons that belong to same class. **(c)-(d)** Neurons that belong to same class, where the class is different from that of (a) and (b).

Table 1: Separation between the neurons in Fig. 5

| Neuron | a | b | c | d |
|---|---|---|---|---|
| a | 0 | 0.1 | 92.5 | 108.9 |
| b | - | 0 | 65.6 | 141.8 |
| c | - | - | 0 | 1.7 |
| d | - | - | - | 0 |

We have also applied the combined geometric-statistical algorithm to three different classes of neurons, each class having six neurons. The dataset, as described in Table 2, is obtained from the public neuron database neuromorpho.org.

Table 2: 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 |

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 $k$ is 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 our experiment, we observe no retrieval error for the first and second closest matches. The third closest match has only two retrieval errors, where neuron 5 is matched with neuron 8 and neuron 8 is matched to 3. Hence, the retrieval error in the third closest match is 11.2%.

## 5. CONCLUSION

We have presented a similarity measure between neurons using a combination of geometric and statistical features. Preliminary experiments reveal encouraging results. To achieve a complete, robust matching, it is likely that the combination of several methods will be required. Currently, we are attempting to incorporate this tool in a content based neuron retrieval system to retrieve similar neurons from a database.

## 6. ACKNOWLEDGEMENTS

This work was supported in part by NSF under 1062433.

## 7. 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.