

# 3D NEURON POSE ESTIMATION FROM CALCIUM IMAGE STACK SEQUENCES USING NEURON PLANE MODEL

*S. Gulyanon*<sup>1</sup>, *W. D. Tracy*<sup>2</sup>, *L. He*<sup>2</sup>, and *G. Tsechpenakis*<sup>3</sup>

<sup>1</sup>Department of Computer Science, Thammasat University, Thailand;

<sup>2</sup> Department of Biology, Indiana University, USA;

<sup>3</sup>Computer and Information Science Department, Indiana University-Purdue University Indianapolis, USA;

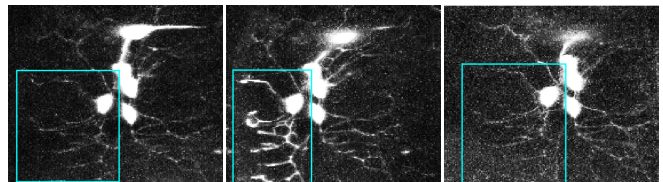
## ABSTRACT

It is known that the sensory neurons control the locomotion but we do not know how this happens. The lack of robust tools for simultaneous quantification of both morphology and neuronal activity in calcium images hinders the progress. The main issues are that these calcium volumes are noisy, have low resolution in Z-axis, contain significant deformations during movement, and suffer from the characteristics of calcium images. In this work, we developed the interactive tracking method of surfaces enclosing neurons. Tracking neurons indirectly helps handle these issues, and improve the efficiency and robustness for 3D pose estimation, while little user intervention is required to aid with dendrite detection. Our method finds the best fitting surface using our neuron plane model by incorporating the image appearance, local shape, as well as domain knowledge about neuron morphology and locomotion. We validated our results using time-lapse calcium image stacks of larval *Drosophila* sensory neuron.

**Index Terms**— Interactive pose estimation, pictorial structures, calcium images, neuron morphology

## 1. INTRODUCTION

Over decades, we know that sensory neurons control the locomotion process but we poorly understand the molecular mechanisms that determine the responses of the proprioceptive cells. Although we aware that the absence of sensory signals from these proprioceptors causes the uncoordinated and slow locomotion is uncoordinated, it is unknown exactly what these signals are [1]. To answer this question, we use high speed confocal microscopy to describe how the dendrites of proprioceptive neurons in larval *Drosophila* are deformed by the forces of crawling motions. This is possible because these sensory neurons lie beneath the epidermis of larval *Drosophila* [2]. Surprisingly, the pattern of deformation for a given neuron is unique in each direction of motion [3]; and two particular neurons (dda-D and dda-E) show strong activation level during the locomotion. To gain insight into the



**Fig. 1.** Calcium image characteristics. Three frames of the same neuron (cyan boxes) in temporal order from left to right. First and last frames show inactive neurons so their intensity is low. The middle frame shows the active neuron. Although it has high intensity, it is deforming. Image intensity is adjusted for visualization purposes.

underlying mechanisms, we must expose the link between the change in morphology and the neuronal activity of the neuron pair.

Fortunately, calcium imaging holds the key to understanding the relationship between neuron morphology and neuronal activity because its intensity is proportional to the neuronal firing level; as a result calcium images also show the location of active regions in neurons. Hence, the time-lapse calcium imagery allows us to observe both neuron morphology and neuronal activity simultaneously. Due to the large amount of image sequences required for learning the link between locomotion and sensory neuron activity, the manual approach is infeasible; and the lack of robust tools for simultaneous quantification of both morphology and neuron activity hinders the progress.

The quantification of neuronal morphology in calcium images is challenging due to four factors. First, it involves tracking neurons in 3D over noisy calcium image sequences. Second, time-lapse image stacks have low resolution in Z-axis due to the limitation of imaging system. Third, the deformation of neurons during the locomotion can be severe and creates blurry images and large displacement. Last, the most challenging problem is the characteristics of calcium images, which has low intensity when neurons are inactive so the neurons are barely visible in such cases. Neurons have high intensity when they are active but they are usually deforming in those frames as well (fig. 1). Due to these issues, previous works are incapable of recovering 3D neuron pose from noisy calcium volume sequences [3–6].

Here we developed the semi-automatic 3D neuron pose estimation method that can handle these issues and helps re-

This work is supported by NSF/DBI [#1252597]: ‘CAREER: Modeling the structure and dynamics of neuronal circuits in the *Drosophila* larvae using image analytics’ awarded to G. Tsechpenakis.

duce the operation time in quantifying both the morphology and neuron activity. It requires little user interaction, including the initial trace and the location of the farthest tips of dendrites, to help handle dendrite detection issues. Then, our method finds the best fits neuron configuration by fitting the neuron plane model surrounding the neuron pair. It incorporates the image feature, neuron local shape, and the domain knowledge about the neuron morphology and deformation during the crawling locomotion. Finally, the deformed traces are efficiently recovered from the neuron plane model.

## 2. NEURON PLANE MODEL

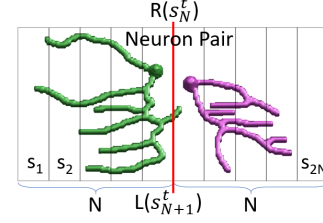
First we present the neuron plane model, which is a simplified representation of the neuron pair. It helps reduce the number of parameters required for describing the deformation of neurons over the image sequences.

The main issue of neuron tree structure is that it has high degree of freedom (DoF) and there are insufficient number of constraints to find the acceptable results. To make this problem feasible, first we assume that neuron mass is preserved and it is enclosed by the surface (the layer beneath epidermis). So we can model the deformation of the neuron through the deformation of the enclosing surface. Second, we limit the direction of crawling locomotion of samples. *Drosophila* larvae are observed by putting the samples through a tiny channel to prevent the larvae from moving into another direction except for forward or backward. As a result, we need to model the deformation only in one direction (X-axis).

With these assumptions, we define the neuron plane model enclosing the neuron pair at time  $t$  as the linear chain model of pictorial structures [7] consisting of  $2N$  equal-width stripes/parts,  $\mathbf{S}^t = \{s_i^t\}$ , for  $i = 1, \dots, 2N$  and  $t = 0, \dots, T$  (fig. 2). A set of part parameter  $\theta^t = \{\theta_i^t\}$  indicating the orientation in XZ-axis of every part at time  $t$  (we assume there are no movements in Y-axis). At  $t = 0$ ,  $\mathbf{S}^0$  is the bounding box of the initial neuron trace provided by the user,  $Y^0$ , so it is a flat surface,  $\theta_i^0 = 0$ , for  $\forall i \in [1, 2N]$ .  $L(s_i^t)$  and  $R(s_i^t)$  are the functions indicating the leftmost and rightmost locations of the stripe  $s_i^t$  respectively. To standardize the neuron plane model, the middle line of the neuron plane model  $R(s_N^t), L(s_{N+1}^t)$  lies between the two soma of a neuron pair. The centroid of soma is detected using the circular Hough transform [8].

## 3. OPTIMIZATION

Calcium image characteristics (fig. 1) makes it challenging to get reliable tracking results in every frame. To tackle this difficulty, our method requires user interaction, including the initial trace,  $Y^0$ , when the neuron pair is flat and the X-coordinate of the farthest dendrite tips at time  $t$ ,  $p_k^t$  (fig. 3). Moreover, we track the neurons indirectly by tracking the neuron-enclosing surface instead using our neuron



**Fig. 2.** Neuron plane model enclosing the neuron pair (green and magenta). The bounding box is split into  $2N$  stripes. Its middle line (red line),  $L(s_{N+1}^t)$  and  $R(s_N^t)$ , aligns with the center between soma.

plane model to improve robustness and efficiency. To convert the initial trace  $Y^0$  into the plane model, the number of parts  $N$  is picked to be just high enough to enclose the trace and set  $\theta_i^0 = 0$ .

Given the volume sequence  $V = \{V^t\}$ , we formulate the objective function as the conditional distribution over the neuron plane parameter,  $\theta^t$ , at every time step,  $\forall t \in [1, T]$ :

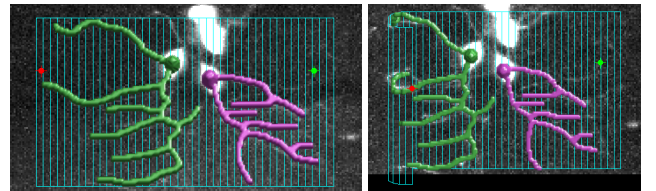
$$\begin{aligned} \hat{\theta}^t &= \arg \max_{\theta^t} P(\theta^t | V^t, \theta^{t-1}, p_k^t), \\ P(\theta^t | V^t, \theta^{t-1}, p_k^t) &\propto \exp \left\{ - \sum_{t=1}^T \left( E_x(\theta^t, V^t, p_k^t) + E_z(\theta^t, \theta^{t-1}) \right) \right\} \end{aligned} \quad (1)$$

Although using neuron plane model improves the efficiency in finding the optimal trace configuration  $Y^t$  because it has lower number of DoF, we improve its efficiency further by separating the user intervention and automatic steps. So the optimization is divided into two steps — X-optimization (Sec. 3.1) and Z-optimization (Sec. 3.2), in order to achieve low latency, crucial in interactive approach. Both steps are optimized using the simulated annealing technique [9].

### 3.1. X-Optimization

This step finds the neuron plane configuration in X-axis with respect to the user input that best match to the input volumes, while enforcing the constraints of smooth flat surface.

$$\begin{aligned} E_x(\theta^t, V^t, p_k^t) &= E_{img}(\theta^t, V^t) + E_{shape}(\theta^t) \\ &+ E_{flat}(\theta^t) + E_{con}(\theta^t, p_k^t) \end{aligned} \quad (2)$$



**Fig. 3.** User interaction required by our method. Red and green points are the X-coordinates of the farthest dendrite tips; and the traces of a neuron pair are in magenta and green color. The neuron plane model is in cyan color. (Left) at  $t = 0$ , the neuron plane model is the loose bounding box of the initial traces  $Y^0$ . (Right) at  $t = 12$ , the neuron pair is deformed with respect to the deformed neuron plane model.

**Image feature** term  $E_{img}$  selects the neuron plane configuration that maximizes the normalized cross correlation (NCC) [10] between the current frame,  $V^t$ , cropped around the neuron pair and the transformed initial frame,  $T(V^0, \theta^t)$ , with respect to the neuron plane model (see details of  $T$  in Section 3.3). Since the calcium image stacks are noisy, the maximum projection of image stacks,  $I^t$ , are preferred to make the input images more reliable.

$$E_{img}(\theta^t, V^t) = NCC(I^t, T(I^0, \theta^t)) \quad (3)$$

**Shape energy** term  $E_{shape}$  enforces the transformation smoothness over the orientation parameters of the neuron plane model. It is the local shape regularization along X-axis.

$$E_{shape}(\theta^t) = \sum_{i=1}^{2N-1} (\theta_{i+1}^t - \theta_i^t)^2 \quad (4)$$

**Flat surface energy** term  $E_{flat}$  encourages flat surface to capture the whole neurons when they start folding. This term helps us decouple the two optimization steps by making this step independent of surface depth.

$$E_{flat}(\theta^t) = \sum_{i=1}^{2N} (\theta_i^t)^2 \quad (5)$$

**Constraint energy** term  $E_{con}$  matches the dendrites tips derived from the neuron plane model with the user-given ones. It enforces the constraints provided by users.

$$E_{con}(\theta^t, p_k^t) = \sum_{k=1}^2 (p_k^t - T(p_k^0, \theta^t))^2 \quad (6)$$

### 3.2. Z-Optimization

The second step automatically tunes the neuron model configuration in Z-axis according to the input volume and our domain knowledge about the morphology and locomotion. It adjusts the neuron plane model so that it has smooth transition, low curvature, and low depth. Hence, in this step, the value of  $\theta_i^t$  is modified by reflecting across the XY-plane. Since it is an automatic step, so it is a part of the post-processing step.

$$E_z(\theta^t, \theta^{t-1}) = E_{dep}(\theta^t) + E_{curv}(\theta^t) + E_{tran}(\theta^t, \theta^{t-1}) \quad (7)$$

**Depth energy** term  $E_{dep}$  adjusts the neuron plane model to have low depth to prevent the model from expanding outside of the input volumes, since image stacks have around 5-10 slices ( $\approx 10 \mu m$ ).

$$E_{dep}(\theta^t) = \sum_{i=1}^{2N} (D(\theta_i^t))^2 \quad (8)$$

where  $D(\theta_i^t)$  is the depth of the neuron stripe/part  $i$  at time  $t$ .

**Curvature energy** term  $E_{curv}$  selects low curvature configuration to enforce shape smoothness over XZ-axes and regularize the neuron plane parameters.

$$E_{curv}(\theta^t) = \sum_{i=2}^{2N-1} (R(s_{i+1}^t) - 2L(s_i^t) + L(s_{i-1}^t))^2 \quad (9)$$

**Transition energy** term  $E_{tran}$  enforces transition smoothness between consecutive time steps to ensure consistency over sequences.

$$E_{tran}(\theta^t) = \sum_{i=1}^{2N} \{D(\theta_i^{t-1}) - D(\theta_i^t)\}^2 \quad (10)$$

### 3.3. Retrieving Deformed Traces

The neuron plane model represents the deformation of the initial trace via the uniform cubic B-spline free-form deformations (FFD) [11]. To recover the deformed neuron traces  $Y^t$ , first we need to compute the control points  $\mathbf{C}^t = \{c_k^t\}$ , which is defined by endpoints of parts.

$$c_k^t = \begin{cases} L(s_{k+1}^t), & \text{if } 0 \leq k \leq N-1 \\ R(s_k^t), & \text{if } N \leq k \leq 2N \end{cases} \quad (11)$$

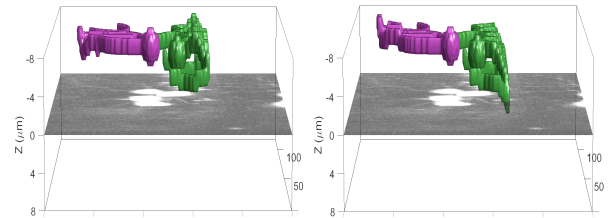
So the transformation from initial time step  $t = 0$  to the any time step  $t$ ,  $T(\mathbf{x}, \theta^t)$ , at any point  $\mathbf{x} = (x, y, z)$  is defined as,

$$T(\mathbf{x}, \theta^t) = \sum_{c_k^t \in \mathbf{C}^t} \eta(\mathbf{x}, c_k^t) \mathbf{d}_k^t \quad (12)$$

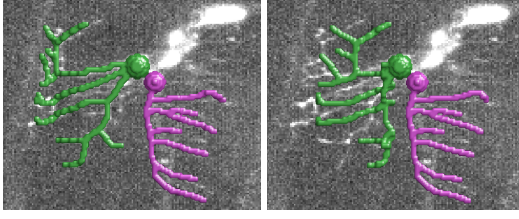
where  $\eta$  is the coefficient of the control point  $c_k^t$  with displacement  $\mathbf{d}_k^t = c_k^t - c_k^0$  at point  $\mathbf{x}$ .

## 4. EXPERIMENTS

We validated our method on the time-lapse calcium volumes dataset of larval *Drosophila* sensory neurons. Larvae expressing the genetically encoded calcium sensor GCaMP6.0F in the proprioceptors were observed in four dimensional confocal microscopy imagery. This is achievable because the neurons are found directly beneath the cuticle and epidermis which are optically transparent.



**Fig. 4.** The importance of depth energy  $E_{dep}$  that limits the depth of traces (left) within the image stack (9 slices,  $z = 0, \dots, 8$  in this example). Otherwise, the depth may be overestimated and higher than the thickness of the image stack (right).



**Fig. 5.** The importance of flat surface energy  $E_{flat}$ . It encourages the neuron plane to fold (left) instead of squeezing it to fit the input volume (right).

In this dataset, each sequence captured the neuronal displacement and deformation during the crawling locomotion over 10-50 frames (depends on the movement speed). We measure how well model fits the data using the mutual information (MI) [12] between artificial volume produced by the deformed initial traces against the input volumes. MI is defined by,  $MI(A, B) = \sum_{a \in A} \sum_{b \in B} P(a, b) \log \frac{P(a, b)}{P(a)P(b)}$ , where  $A$  is the input volume and  $B$  is the artificial volume derived from the neuron plane model. Maximizing the mutual information find the best aligned volumes. MI metric is adopted here because it was used in comparing between the model and the volume before [12].

Our method is the first for recovering 3D neuron pose from calcium image sequences. Previous works [3–5] in pose estimations are incapable of handling issues in noisy time-lapse calcium volumes during the locomotion. Table 1 emphasizes the significance of two novel components,  $E_{dep}$  and  $E_{flat}$ , in our energy function formulation or lack thereof. The depth energy is required to bound the thickness of neurons (fig. 4). Although the resolution of image stacks in Z-axis is low, the thickness of neurons can still be approximated from the number of image slices and it shows that neurons have low thickness. The flat surface energy encourages the plane to fold rather than squeeze the neurons to match the input volume (fig. 5). These two terms help decouple the two optimization steps.

Second, we show the improvement in computation compared to when these two optimization steps are combined. The reason for separating the optimization step is purely for performance. Since this is the semi-automatic method so the performance in term of efficiency is critical for providing quick response time and user interaction. Here, we achieved the average speed around 15 FPS on mid-end computer (Intel Core i7-5500U CPU 2.40 GHz DDR3 8.00GB). If we combines the two steps we end up with just around 3 FPS. The

	Our Method	$E_{dep} = 0$	$E_{flat} = 0$
Avg	0.1446	0.1380	0.1436
Std	0.0342	0.0330	0.0340

**Table 1.** The mutual information similarity results between the artificial trace volume produced from the neuron plane model against the input volumes. From left to right: our model, our model without the depth energy term, and our model without the flat surface energy term. The rows show the average and the standard deviation over the volume sequences.

efficiency gain comes from the separation of the automatic process from the user-involved process.

## 5. CONCLUSION

We presented the semi-automated 3D neuron pose estimation method via neuron plane model for calcium volume sequences. Our method incorporates the image feature, neuron local shape characteristics, and the domain knowledge about the deformation during the crawling locomotion. Our neuron plane model enclosing the neuron pair offers a simplified model. While breaking down the optimization into two steps helps reduces latency to user interactions. Our method is robust to noisy calcium volumes with severe deformations and requires only little user interaction. We validated our results using the time-lapse calcium volumes dataset of larval *Drosophila* sensory neurons.

## 6. REFERENCES

- [1] C. L. Hughes and J. B. Thomas, “A sensory feedback circuit coordinates muscle activity in *Drosophila*,” *Mol. Cell. Neurosci.*, vol. 35, no. 2, pp. 383–396, 2007.
- [2] W. B. Grueber, L. Y. Jan, and Y. N. Jan, “Tiling of the *Drosophila* epidermis by multidendritic sensory neurons,” *Development*, vol. 129, no. 12, pp. 2867–2878, 2002.
- [3] S. Gulyanov, L. He, W. D. Tracey Jr., and G. Tsechpenakis, “Neurite tracking in time-lapse calcium images using MRF-modeled pictorial structures,” *ISBI*, pp. 1564–1568, 2018.
- [4] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, “Pose search: Retrieving people using their pose,” *CVPR*, pp. 1–8, 2009.
- [5] P. Glowacki, M. A. Pinheiro, A. Mosinska, E. Türetken, D. Lebrecht, A. Holtmaat, R. Sznitman, J. Kybic, and P. Fua, “Reconstructing evolving tree structures in time lapse sequences by enforcing time-consistency,” *TPAMI*, vol. PP, no. 99, pp. 1–1, 2017.
- [6] S. Gulyanov, N. Sharifai, M. D. Kim, A. Chiba, and G. Tsechpenakis, “Neurite reconstruction from time-lapse sequences using co-segmentation,” *ISBI*, pp. 410–414, 2017.
- [7] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial structures for object recognition,” *IJCV*, vol. 61, no. 1, pp. 55–79, 2005.
- [8] R. O. Duda and P. E. Hart, “Use of the hough transformation to detect lines and curves in pictures,” *Commun. ACM*, vol. 15, no. 1, pp. 11–15, 1972.
- [9] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [10] J.-C. Yoo and T. H. Han, “Fast normalized cross-correlation,” *Circuits, Systems and Signal Processing*, vol. 28, no. 6, pp. 819, 2009.
- [11] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, “Nonrigid registration using free-form deformations: Application to breast MR images,” *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, 1999.
- [12] P. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” *IJCV*, vol. 24, no. 2, pp. 137–154, 1997.