A RIDGE SCORE WITH APPLICATION TO PIECEWISE LINEAR NEURAL RECONSTRUCTION

Erhan Bas, Deniz Erdogmus

Cognitive Systems Laboratory, ECE Department, Northeastern University, Boston, MA 02115 - USA

ABSTRACT

We proposed a novel algorithm to extract connectivity information of neuronal arbors from 3D confocal images. The method is based on the use of ridge sampling and approximation using piecewise linear segments that conform to an inequality constraint that ensures shape accuracy. An automatic neuron structure reconstruction algorithm based on kernel interpolation of intensities using multiscale local Gaussian kernels has been utilized on a sample Olfactory neuron reconstruction problem successfully.

***Index Terms*—** Manifold sampling and approximation, automated neuron reconstruction

1. INTRODUCTION

In the recent years, there has been a significant effort to extract morphology information from neuronal arbors to construct connectivity wiring diagrams. These diagrams enables neuro-scientists to design large scale simulations that helps to better understand and interpret how a neural network works and how certain neuro-degenerative diseases might occur. With recent developments in imaging techniques, large volumes of 3D high quality/resolution images have become available. However, estimation of neural wiring diagrams from 3D stacks is not straightforward. There are two major approaches. *i*) Global approaches that attain globally optimum results by checking all possible cases. In general these methods are not feasible given the size of the search space for all possible trajectories and requires lots of training to come up with a globally feasible objective function that is generalizable at least to certain types of datasets. Moreover, there is not a rigorous way to make even a small adjustment to the optimization process, which makes these methods overall not undesirable [1]. *ii*) Second category includes local approaches that constrain the search space to a local region. These methods propagate in the search space based on the decisions using local evidence [2]. This kind of approach makes it easier to develop fast and adaptive implementations, but generally not yielding reliable solution. Since decisions are subject to a small neighborhood of the search space, i.e. 26 connected voxels in 3D, they might miss the general topology due to noise or outliers. In this manuscript, we described a novel connectivity analysis method that bridges the two approaches. We use local evidence to build a reduced connectivity graph of the centerline of the whole data set that has the practicality of local approaches with a broadened view over the search space.

This work is supported by NSF under grants ECCS0929576, ECCS0934506, IIS0934509, IIS0914808, and BCS1027724. The opinions presented here are solely those of the authors and do not necessarily reflect the opinions of the funding agency.

2. RIDGE DEFINITION

A ridge is a curve that passes through the center of a high dimensional data cloud or a multivariate scalar valued function. Definition of the ridge is studied in various concepts such as in image processing and statistical data analysis [3, 4, 5]. A rigorous way of defining the ridge of a function is to check the local critical set/manifold definitions [3, 5]. Let $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ be an at least twice continuously differentiable function, $\mathbf{g}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ be its gradient and Hessian. Let $\mathbf{q}_i$ and λ_i be the i^{th} eigenvector/eigenvalue pair of the Hessian matrix of $f(\mathbf{x})$ such that $\lambda_1 \leq \cdots \leq \lambda_n$. In general, a point is on the d-dimensional principal manifold if the gradient is orthogonal to $n-d$ eigenvectors, $\mathbf{g}(\mathbf{x})^T\mathbf{q}_i(\mathbf{x}) = 0$, and the eigenvalues corresponding to these eigenvectors are all negative, $\lambda_{d+1,\ldots,\lambda_i,\ldots,\lambda_n} < 0$. Since our goal is to extract 1D centerline information, we will assume $d = 1$ in the rest of the paper.

Mentioned local conditions are generalizations of the local maxima/minima conditions; a point on the ridge is the local maximum of the function in the subspace spanned by the $n-d$ eigenvectors, $S_\perp = \text{span}(\mathbf{q}_{d+1}, \ldots, \mathbf{q}_n)$. Let $\mathbf{H}_\perp(\mathbf{x}) = \sum_{d+1}^{n}(\lambda_i\mathbf{q}_i\mathbf{q}_i^T)$ be the orthogonal component of $\mathbf{H}(\mathbf{x})$; a measure for being on the ridge can be formulated as

$$\gamma(\mathbf{x}) = \frac{\mathbf{g}(\mathbf{x})^T H_\perp(\mathbf{x})\mathbf{g}(\mathbf{x})}{\|H(\mathbf{x})\mathbf{g}(\mathbf{x})\|\|\mathbf{g}(\mathbf{x})\|} \tag{1}$$

The measure $\gamma(\mathbf{x})$ becomes 0 on the ridge. In order to explain the concept better, consider a 1D Gaussian profile ($\mu = 0$ and $\sigma = 1$) that decays orthogonal to the ridge as depicted in Fig.1(a) as shown with the blue solid curve, with its absolute value of the first derivative and the second derivative

as green dashed curve and red dots respectively. Note that, for this example, $\mathbf{H}(\mathbf{x})_{\parallel} = \mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{x})_{\perp}$ will be singular with $\lambda = 0$, since $f(\mathbf{x})$ is constant along $\mathbf{q}_1(\mathbf{x})$. Multiscale analysis using an isotropic Gaussian kernel states that optimum scale that results in maximal ridge response for a tubular structure like this is obtained for a kernel width (standard deviation-σ)that matches the radius of the tube at that location. In a dual concept for vessel segmentation, a review of methods that optimize the kernel width can be found in [6]. However these methods are just enhancement techniques and incapable of extracting high level information such as connectivity. Nevertheless, they still do provide some important geometric information about the structures such as the radius of the local tubular structure and can be used as a denoising technique at the preprocessing step. Fig. 1(b) shows the variation of the $\gamma(\mathbf{x})$ measure along the same profile without the normalization term in the denominator of Eq. (1). In general the denominator will be dominated by the constant term from the eigenvalue elongated with the ridge. Although resultant profile might be scaled, this term will not effect the location of the critical points. Notice that the described value attains non-positive values in the region where $\mathbf{H}(\mathbf{x})_{\perp}$ is negative definite around the ridge (concave), and is bounded between $-1 \leq \gamma(\mathbf{x}) \leq 0$, but not monotonic with respect to $|\mathbf{x}|$.[1] As we will discuss later, we will consider only connecting line segments between points that remain in the vicinity of the ridge, where the function is concave on a local convex region. For that purpose, we propose a more practical measure by defining the ridge of a function in terms of the inverse of the local Hessian, $\mathbf{H}^{-1}(\mathbf{x})$. For a unimodal Gaussian distribution, $\mathbf{H}^{-1}(\mathbf{x})$ is equal to the negative of local covariance, and in the rest of the paper, we will denote this quantity with $\mathbf{C}(\mathbf{x})$. One can write the orthogonal component of $\mathbf{C}(\mathbf{x})_{\perp} = \sum_{d+1}^{n}(\lambda_i^{-1}\mathbf{q}_i\mathbf{q}_i^T)$, and similar to Eq. 1 we define the membership function of a sample $\mathbf{x}$ as

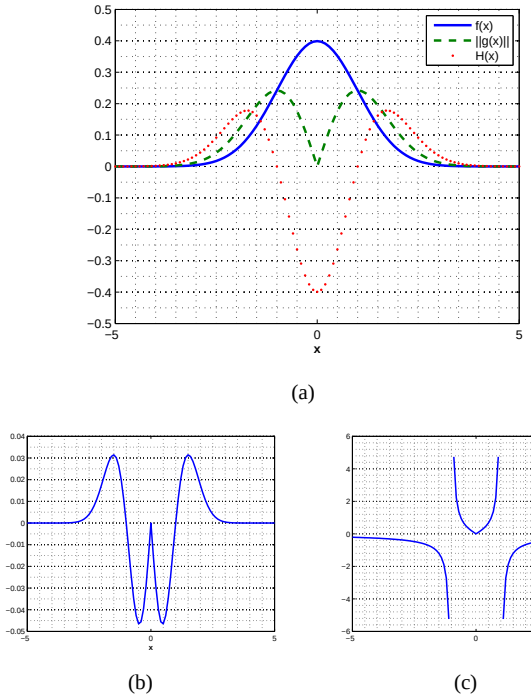
$$\eta(\mathbf{x}) = -\frac{\mathbf{g}(\mathbf{x})^T\mathbf{C}_{\perp}(\mathbf{x})\mathbf{g}(\mathbf{x})}{\|\mathbf{C}(\mathbf{x})\mathbf{g}(\mathbf{x})\|\|\mathbf{g}(\mathbf{x})\|} \qquad (2)$$

Newly proposed measure, $\eta(\mathbf{x})$, has only one local minimum at 0 (on the ridge), and it is monotonically increasing elsewhere except at the sigma, where the sign of the eigenvalues of $\mathbf{H}(\mathbf{x})$ and the $\mathbf{C}(\mathbf{x})$ change as the multiscale analysis imply. Fig. 1(c) shows the profile without the normalization term for $\eta(\mathbf{x})$.

3. CONNECTIVITY OF NEURONAL ARBORS

In order to analyze the overall connectivity of a neural structure, global methods first define a pairwise similarity between samples. Such methods, e.g. minimum spanning tree (MST)

[1]We skip the derivations due to space constraints, but all the derivations will be described in a future extended publication. Here $|\mathbf{x}|$ is used as an indicator of distance from the ridge.



(a)

(b) (c)

Fig. 1. (a) Gaussian and derivatives. (b) Numerator of (1). (c) Numerator of (2)

[7], seek to connect similar samples by minimizing an overall cost function, i.e. the sum of the all pairwise distances, in a cluster. The cost function is defined in the spatial domain, where distances are measured in terms of pixels/voxels. Here, unlike previous methods, we adopt a more statistical approach to define connectivity over spatial domain. We assume that two samples are connected iff they belong to the same ridge. So instead of searching for all possible trajectories between samples, we convert the problem to first searching for the ridges in an image, then once we obtain these ridges, checking whether the samples belong to the same ridge or not. This way, nearby samples in a neighborhood can be separated based on their underlying ridges. In fact, this method can also be used for clustering purposes. For this, we first project all the samples to the ridge by solving an ordinary differential equation using Euler integration, where the error accumulation is in the order of ϵ^2. Here ϵ is the selected step size for integration. Fig. 2(a-b) shows such cases, where the projected samples are overlayed with the Olfactory neuronal dataset [8]. Note that unlike most of the previous methods, our approach estimates the underlying ridge with subvoxel accuracy.

Once we project the points to the ridges. We define the dissimilarity score between samples (ideally 0 for a path on the ridge). Let $\wp(\mathbf{a}, \mathbf{b})$ be the dissimilarity score[2] of points $\mathbf{a}$

[2]The proposed $\wp(\mathbf{a}, \mathbf{b})$ score is a pseudometric [9], and is not always symmetric. However, it is a useful measure of dissimilarity between samples.

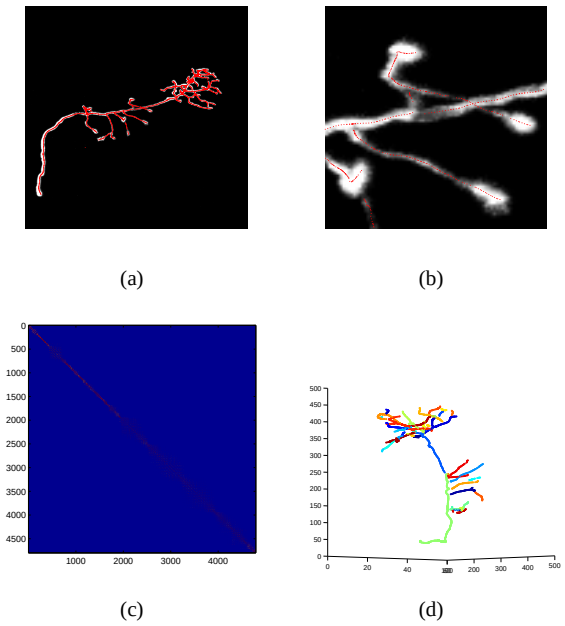


Fig. 2. (a) Olfactory Projection dataset. (b) Zoomed version. (c) Pairwise masked dissimilarity score. (d) Segmentation result.

and $\mathbf{b}$, given as the ratio of the line integral of scalar valued function, $\eta(.)$, from $\mathbf{a}$ to $\mathbf{b}$ evaluated on the curve $\mathbf{l}(t)$ and the arc length of the curve

$$\wp(\mathbf{a},\mathbf{b}) = \frac{\int_0^1 \eta(\mathbf{l}(t))[\dot{\mathbf{l}}^T(t)\dot{\mathbf{l}}(t)]^{\frac{1}{2}}dt}{L(\mathbf{a},\mathbf{b})} \tag{3}$$

here we parameterized $\mathbf{l}(t) = \mathbf{a} + t(\mathbf{b}-\mathbf{a})$ as a line with $\mathbf{l}(0) = \mathbf{a}, \mathbf{l}(1) = \mathbf{b}, \dot{\mathbf{l}}(t) = (\mathbf{b}-\mathbf{a})$, and $L = \int_0^1 [\dot{\mathbf{l}}^T(t)\dot{\mathbf{l}}(t)]^{\frac{1}{2}}dt$ is the arc length of the path.

An important concept here is the order of the eigenvalues that defines ridges. Consider a $\mathbf{T}$ shaped tubular mixture, where two ridges coincide or even intersects, then the eigenvectors that define the ridge of each tubular object will be aligned differently (in fact orthogonal to each other). In order to overcome such ambiguities, we use the eigenvalue ranking at location $\mathbf{a}$ and order the eigenvectors with their corresponding eigenvalues with respect to this initial ranking. Let $\mathbf{q}_j^{\mathbf{l}(t)}$ be the j^{th} eigenvector at $\mathbf{l}(t)$ and $\mathbf{q}_{\perp,k}^{\mathbf{a}}$ be the k^{th} eigenvector that spans $S_\perp(\mathbf{a})$ at reference $\mathbf{a}$, where $j = 1,\ldots,n$ and $k = d+1,\ldots,n$. Ranking of the k^{th} direction can be obtained as

$$\Theta_k = \arg\min_{i=1,\ldots,n} (\mathbf{q}_i^{\mathbf{l}(t)})^T q_{\perp,k}^{\mathbf{a}} \tag{4}$$

An immediate implication of the scale space theory is that, any integration on a line segment at which the cross-section of $f(\mathbf{x})$ is not concave will accumulate negative values for the portion of the function that is not concave. In fact, we consider such line segments unsuitable, since they clearly deviate too far from the ridge. For example, all the samples from a linear ridge will be connected to each other, whereas at a curvy location possible connections will be limited in length due to the rapidly changing signs of the eigenvalues along the integration path even at short distances from the ridge. In fact, this limitation is consistent with the curvilinear geometry, since the orthogonal profile shown in Fig. 1(c) is bounded with the local structure radius. Since we already map the samples to the ridges, with this concavity criterion, we also enforce all the connections to be inside a local region around the ridge such that

$$\bar{\wp}(\mathbf{a},\mathbf{b}) = \begin{cases} \wp(\mathbf{a},\mathbf{b}) & \text{if } \forall t \in [0,1], \lambda_{d+1,\ldots,n}(\mathbf{l}(t)) < 0 \\ \infty & \text{otherwise} \end{cases} \tag{5}$$

Once we obtain all local connectivities constrained to the underlying geometry defined by the concavity of the function $f(\mathbf{x})$, we can partition the structure into subsegments by thresholding the pairwise score, where each subsegment has some predefined confidence, i.e. $\mathbf{M}(\mathbf{a},\mathbf{b}) = \bar{\wp}(\mathbf{a},\mathbf{b})$, if $\bar{\wp}(\mathbf{a},\mathbf{b}) < thr$, 0 otherwise. Fig. 2(c) shows, $\mathbf{M}(\mathbf{a},\mathbf{b})$, the pairwise score matrix after thresholding. In an hierarchical setup, by varying *thr*, one can obtain a partition of the dataset based on the proposed dissimilarity score. The upper boundary of the score is 1, and is obtained by evaluating the path integral on the edge of the tubular structure. As we decrease *thr*, the initial partition will start to break down from bifurcation regions where multiple ridges are present. For example, in the previous T-shaped ridge mixture example, lets assume point $\mathbf{a}$ is on the bottom ridge, whereas $\mathbf{b}$ is positioned at the upper ridge positioned at the intersection location of the ridges. By following the initial ridge axis defined at $\mathbf{a}$ (y-direction), we have access to point $\mathbf{b}$, whereas the opposite is not true since the initial ridge direction of the top mixture is in x-direction. As we further decrease the value we start to partition individual ridges from high curvature regions. Fig. 2(d) shows the segmentation result for an Olfactory neuron using $thr = 0.1$. Note that described *thr* is uniteless, since we normalized the score values with the arc length in (3).

4. ESTIMATION OF THE FUNCTION

Unlike previous methods that uses voxel intensities directly to define the objective function, we used a more statistical approach, where we define a ridge function $f(\mathbf{x})$ in terms of position ($\mathbf{x}$) and intensity ($I(\mathbf{x})$). $f(\mathbf{x})$ can be written as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{N} I(\mathbf{x}_i) G_{\sigma_\mathbf{c}}(I(\mathbf{x}) - I(\mathbf{x}_i)) G_{\mathbf{\Sigma}_i}(\mathbf{x}-\mathbf{x}_i) \\ &= \sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x}-\mathbf{x}_i) \end{aligned} \tag{6}$$

where $w(\mathbf{x}_i) = I(\mathbf{x}_i)G_{\sigma_\mathbf{c}}(I(\mathbf{x}) - I(\mathbf{x}_i))$ is the weight for the i^{th} data sample $\mathbf{x}_i$, σ_c is the scale of the intensity values $G_{\sigma_c}(\mathbf{x}_i) = C_{\sigma_\mathbf{c}} e^{-\frac{1}{2}(I(\mathbf{x})-I(\mathbf{x}_i))^2/\sigma_c^2}$ and $\mathbf{\Sigma}_i$ is the variable kernel[3] covariance of the Gaussian kernel $G_{\mathbf{\Sigma}_i}(\mathbf{x}_i) = C_{\mathbf{\Sigma}_i} e^{-\frac{1}{2}\mathbf{x}^T {\mathbf{\Sigma}_i}^{-1}\mathbf{x}}$. The gradient and the Hessian of the KDE are:

$$\mathbf{g}(\mathbf{x}) = -\sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x}-\mathbf{x}_i) {\mathbf{\Sigma}_i}^{-1}(\mathbf{x}-\mathbf{x}_i) \quad (7)$$

$$\mathbf{H}(\mathbf{x}) = \sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x}-\mathbf{x}_i)(\mathbf{u}_i\mathbf{u}_i^T - {\mathbf{\Sigma}_i}^{-1}) \quad (8)$$

Here $\mathbf{u}_i = {\mathbf{\Sigma}_i}^{-1}(\mathbf{x}-\mathbf{x}_i)$. The covariance for each Gaussian kernel is selected as the local minimum of the maximizers of the Frangi filter responses within the neighborhood.

5. RESULTS

In order to test the proposed method, we used two 3D Olfactory confocal image stacks obtained from Drosophila fly-brain [8]. There are 60 images in one stack and 62 in the other. Each image size is 512x512 pixels. In order to estimate the initial projections we used $\epsilon = 0.01$ voxels as the step size for the numeric integration. We normalized the image stacks to [0,1] intensity range and we experimentally select $\sigma_c = 0.2$ for the intensity scale. Fig. 3 shows the 3D rendering of the neuronal reconstruction, where first column shows the manual reconstruction and second column displays our algorithms automated reconstruction. We calculated the pointwise distance between the ground-truth and the estimated centerline as 0.7437 ± 0.5163 voxels and 1.25 ± 0.9060 voxels for the first and the second datasets respectively.

6. DISCUSSION

In this paper, we proposed an automatic neuron reconstruction algorithm based on a multi-scale kernel interpolation of intensities in 3D confocal microscopy images. The algorithm identifies samples from the ridge of the intensity function and piecewise linearly connects them into a tree structure ensuring that each segment remains within closed proximity of the ridge, where proximity is defined based on the underlying ridge function. Although we start with a wholly connected graph, we end up with a highly sparse connectivity matrix as seen in Fig. 2(c). Unlike other methods, sparsity of the matrix is not due to the predefined nearest neighbors such as k or ϵ-ball. Proposed method automatically highlights the intrinsic connectivity with respect to the underlying ridges in the image. Preliminary results on Olfactory neuron reconstruction are encouraging.

[3] Assuming Gaussian kernels here for simplicity.

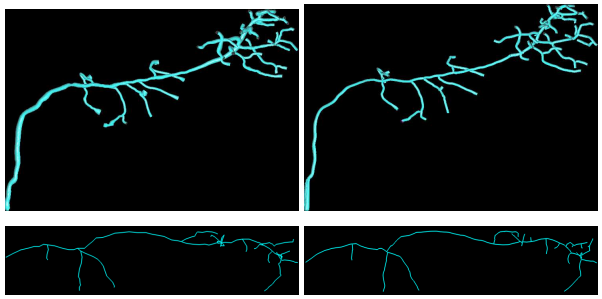


Fig. 3. (a) Ground-truth of the neuron. (b) Reconstructed neuron

7. REFERENCES

[1] S. Schmitt, J.F. Evers, C. Duch, M. Scholz, and K. Obermayer, "New methods for the computer-assisted 3-D reconstruction of neurons from confocal image stacks," *Neuroimage*, vol. 23, no. 4, pp. 1283–1298, 2004.

[2] J. Wang, X. Zhou, J. Lu, J. Lichtman, S.F. Chang, and S.T.C. Wong, "Dynamic local tracing for 3D axon curvilinear structure detection from microscopic image stack," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*. IEEE, 2007, pp. 81–84.

[3] D. Eberly, *Ridges in image and data analysis*, Kluwer Academic Pub, 1996.

[4] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.

[5] U. Ozertem, *Locally Defined Principal Curves and Surfaces*, Ph.D. thesis, Oregon Health & Science University, Department of Science & Engineering, 2008.

[6] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys*, vol. 36, no. 2, pp. 81–121, 2004.

[7] R.L. Rivest T.H. Cormen, C.E. Leiserson, *Introduction to Algorithms*, MIT Press and Mc Graw-Hill, New York, NY, USA, 1990.

[8] G.S.X.E. Jefferis, C.J. Potter, A.M. Chan, E.C. Marin, T. Rohlfing, C.R. Maurer, et al., "Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation," *Cell*, vol. 128, no. 6, pp. 1187–1203, 2007.

[9] Khelemskiĭ, *Lectures and exercises on functional analysis*.