A SCALE-SPACE APPROACH FOR 3D NEURONAL TRACES ANALYSIS

Minh Son Phan[1], Katherine Matho[2], Jean Livet[2], Emmanuel Beaurepaire[1], and Anatole Chessel[1]

[1]Laboratory for Optics and Biosciences, CNRS, INSERM, Ecole Polytechnique, IP Paris, Palaiseau, France
[2]Sorbonne Université, Institut de la Vison, INSERM, CNRS, Paris, France

June 1, 2020

ABSTRACT

The advent of large-scale microscopy along with advances in automated image analysis algorithms is currently revolutionizing neuroscience. These approaches result in rapidly increasing libraries of neuron reconstructions requiring innovative computational methods to draw biological insight from. Here, we propose a framework from differential geometry based on scale-space representation to extract a quantitative structural readout of neurite traces seen as tridimensional (3D) curves within the anatomical space. We define and propose algorithms to compute a multiscale hierarchical decomposition of traced neurites according to their intrinsic dimensionality, from which we deduce a local 3D scale, i.e. the scale in microns at which the curve is fully 3D as opposed to being embedded in a 2D plane or a 1D line. We applied our scale-space analysis to recently published data including zebrafish whole brain traces to demonstrate the importance of the computed local 3D scale for description and comparison at the single arbor levels and as a local spatialized information characterizing axons populations at the whole brain level. The use of this broadly applicable approach is facilitated through an open-source implementation in Python available through GeNePy3D, a quantitative geometry library.

Keywords Computational neuroanatomy · scale space · quantitative geometry · python

1 Introduction

Throughout the evolution of the field of neuroscience, neuroanatomy has played a key role via the analysis of neuronal arbor traces, either at single cell or gross projection levels. The NeuroMorpho.Org database (Ascoli, 2006; Ascoli *et al.*, 2007), which collects and indexes neuronal tracing data, currently hosts more than one hundred thousand arbors of diverse neurons from various animal species. Thanks to technological advances in microscopy, through block-face electron microscopy (Helmstaedter *et al.*, 2013) and large-scale fluorescence microscopy (Abdeladim *et al.*, 2019), we are now able to image even larger tissue volumes with ever increasing resolution and contrast modalities. Crucially, the coming of age of computer vision through the advent of deep learning is offering ways to automate the extraction of neurite traces (Magliaro *et al.*, 2019), a process that is both tedious and time consuming to do manually. As a result, we are about to experience an exponential increase in the amount of neuronal traces extracted in diverse species, brain regions, developmental stages and experimental conditions to answer key questions across neuroscience (Meinertzhagen, 2018). Methods from quantitative and computational geometry will play a major role in handling and analyzing the

growing body of neuronal reconstruction data, in its full tridimensional (3D) complexity, a requisite to linking neuronal structure with development and function of brain circuits and systems.

A vast array of geometric algorithmic methods and corresponding software has been proposed and used to process neuronal reconstructions and answer biological questions (Bates *et al.*, 2020; Cuntz *et al.*, 2010). Indeed, to address the issue of neuronal cell type classification, morphological information has played a prominent role, now extended with recent advances in single cell transcriptomic approaches (Bates *et al.*, 2019). Morphological features for constructing neuron ontologies, described for instance in the Petilla convention (Petilla Interneuron Nomenclature Group *et al.*, 2008), have been computed and used in machine learning methods (Mihaljević *et al.*, 2018). These features have also been used to address more targeted questions, and across neuroscience to compare morphologies across conditions. Moreover, as the neuronal arbor is the result of complex underlying developmental processes, such tools have been used to analyze mechanisms controlling the geometry of neurites in developing tissue (Santos *et al.*, 2018). In the context of large scale traces across whole brain structures or whole brains, examples include using morphological measurements for reconstruction proofreading (Schneider-Mizell *et al.*, 2016) or probing changes in neuronal structure and connectivity through larval growth (Gerhard *et al.*, 2017). Yet, the tools used in all these studies to date fail to provide a fine analysis of the geometrical properties of the neuronal trajectories across scales beyond global metrics like tortuosity, the ratio of curvilinear and Euclidean distance along the path.

In contrast, differential geometry has been very successfully used in pattern recognition and classical computer vision (Sapiro, 2006; Cao, 2003). In particular, the concept of *scale-space* representation has led to very thorough theoretical developments and practical applications. The key idea is that from an original signal one can compute a family of signals that estimate the original at various scale, retaining information only from that scale and enabling multi-scale analysis (Lindeberg, 1990). On 2D curves, the Mean Curvature Motion has been proposed as a well-defined scale-space of curves and has been broadly applied (Cao, 2003). Comparatively little has been proposed on 3D curves because such data is not as common as 2D curves whose extension is usually 3D surface scale-space (Digne *et al.*, 2011). Inherent mathematical difficulty is another reason for this gap. Nevertheless, application of curvature and torsion scale spaces in 3D curve have been reported (Mokhtarian, 1988).

Here, we develop a complete framework for a scale space analysis of 3D curves and use it to analyze neuronal arbors. It allows us to compute the local 3D scale which is the size in micrometers at which the curve locally needs three dimensions to be described. This scale is quite small when the trace is very straight, and larger when the trace has a complex and convoluted pattern over long distances. We propose examples and applications on several published datasets, demonstrating the purpose of this information as a metric for comparing conditions and performing machine learning classifications, to extract large scale morphological features at the single axon level from large scale studies and as a region level descriptor for whole brain analysis. Implementation of the algorithms along with an array of geometry routines and functions is available in the GeNePy3D Python library (https://msphan.gitlab.io/genepy3d). Code to reproduce all figures in this study, exemplifying its use, are available at https://gitlab.com/msphan/multiscale-intrinsic-dim.

2 Results

2.1 Decomposition of neuronal traces into local intrinsic dimensions

We introduce a new form of analysis to investigate the spatial complexity of neuronal arbors. We decompose each neuronal arbor into branches, and each branch into sequences of local curve sections, to classify them according to their intrinsic dimensions, i.e. according to the smallest dimension that accurately describes them. A section is thus intrinsically 1D (line), 2D (plane) or 3D. In addition, the decomposition is hierarchical by capturing the fact that the 1D line is basically embedded in the 2D plane (Figure 1A). The composition of these intrinsic dimensions provides a local shape description of each neuronal branch.

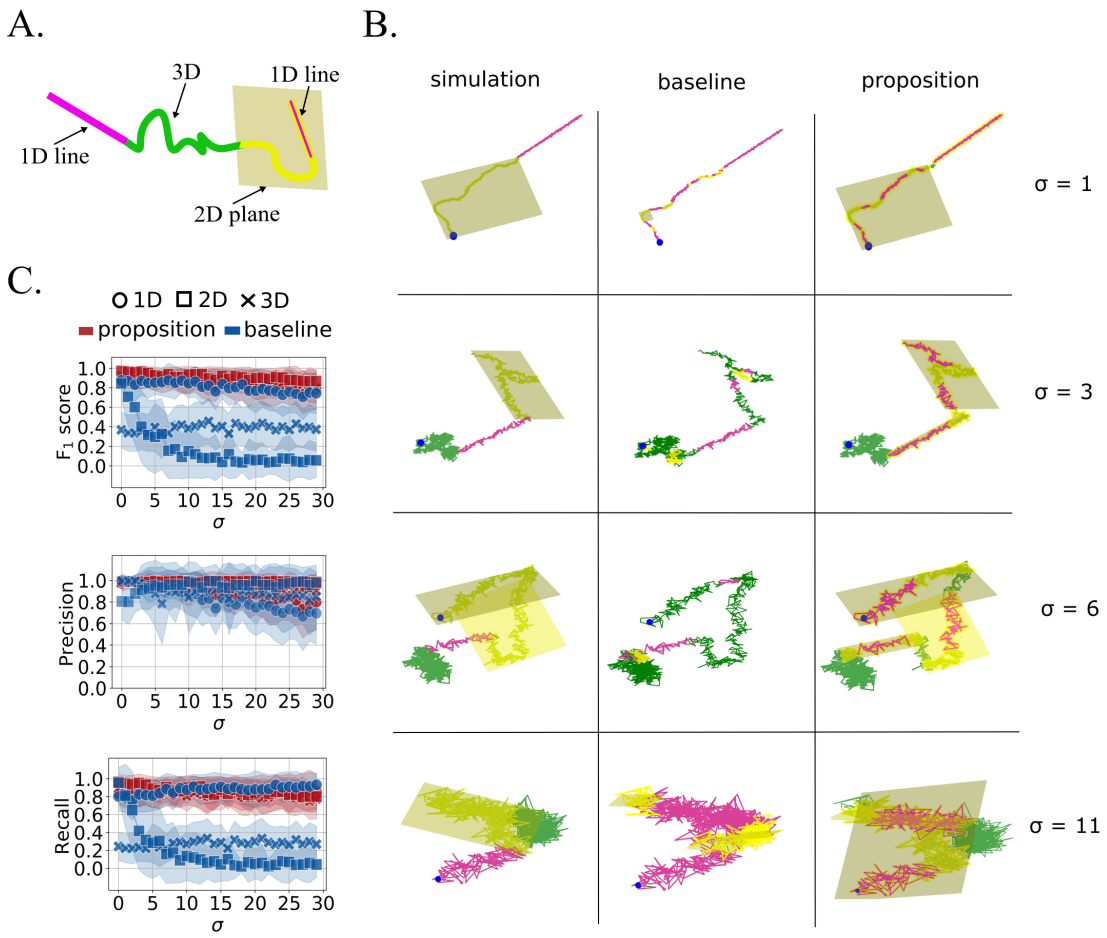

Figure 1: Decomposition of 3D trace into local intrinsic dimensions and validation on simulated data. (A) Example of decomposing a 3D simulated trace into fragments of one, two and three dimensions. (B) Examples of dimension decomposition of proposed method at the right column and a baseline method (Yang *et al.*, 2016; Ma *et al.*, 2017) at the middle column on the simulated traces at the left column with different noise levels (measured in micrometer). (C) F1 score, precision and recall of the proposed method and the baseline w.r.t. noise levels.

First we validate the proposed decomposition on simulated 3D traces, and compare it to a baseline method (Yang *et al.*, 2016; Ma *et al.*, 2017). We vary the noise level, σ, between 1 and $30\mu m$. Examples of simulated traces are shown in Figure 1B in the first column with $\sigma = 1, 3, 6$ and $11\mu m$ where we observe that some regions on the plane (e.g. arc) were entirely deformed at $\sigma = 11\mu m$. The corresponding predictions are shown in the middle column for the baseline method and in the right column for the proposed method. Both methods have a high precision for 1D prediction. In addition, the proposed method produces more 1D fragments than the simulated ones since our current simulation model does not produce hierarchical dimensions where 1D lines can be found in 2D planes. For 2D and 3D predictions, the proposed method works has higher accuracy than the baseline even at high noise level ($\sigma = 11\mu m$). As seen in the top of Figure 1C, with our proposed method we obtain an average accuracy (F1 score) $> 90\%$ at $\sigma \leq 5\mu m$ (medium noise level), $> 85\%$ at $\sigma \leq 10\mu m$ (high noise level) and $\sim 80\%$ for $\sigma > 10\mu m$. The accuracy is much higher compared to the baseline method with a fast decay of 2D accuracy ($< 60\%$ for $\sigma \geq 3\mu m$) and a very low 3D accuracy ($\sim 40\%$). More precisely, both methods give similar results for precision (Figure 1C, middle panel) but not for recall (Figure 1C, bottom panel). The baseline method's recall is very low and rapidly decreasing for 2D and 3D, compared to the proposed method. This can be explained by the fact that the baseline method is not robust to noise. It decomposes a 2D plane into

sequence of 2D arcs and 1D lines, whereas a 3D fragment can be corrupted by false positive 2D fragments. We thus conclude that our intrinsic dimensionality decomposition is accurate and capable of recovering the dimensionality of curve fragments even in the presence of noise.

2.2 Computation of local 3D scale of neuronal traces from multiscale local intrinsic dimensions

Decomposition of local intrinsic dimensions depends on the scale at which we analyze the 3D trace. The notion of *scale* can be abstractly interpreted as a level of detail varying from high, when observing at smaller scales, to low, at larger scales. Computationally, a scale-space is determined through convolution with a Gaussian, building versions of a given curve at all spatial scales. In the context of dimension decomposition, scales are typically measured in micrometers. If one thinks of the evolution of a local fragment on the trace as we look at it through increasing scales, it is typically first described as 3D at the smallest scales (high levels of detail), then becomes 2D or 1D at larger scales (low levels of detail), as details are progressively smoothed out. Examples of decomposition results at discrete scales are shown in the top three rows of Figure 2 where such transformations are made clear. In order to make use of the local dimension decomposition across multiple scales, we propose to define the *local 3D scale* as the highest scale at which the trace still remains in 3D. The trace will then transform to 2D or 1D for scales higher than that local 3D scale. Examples of local 3D scales are shown in the bottom row of Figure 2.

We examined the local 3D scales for cell types with different arbor shapes, using traces from the NeuroMorpho.Org database. The first column in Figure 2 corresponds to the neuronal reconstruction of a spiny ganglia in mouse from (Li *et al.*, 2019). Although this neuron has a relatively simple structure, its local 3D scales express variations in spatial structure across local regions. Indeed, comparing a region of the axonal arbor more distal and more proximal to the soma, the more distal region converts to 2D and 1D faster than the proximal region (Figure 2, first column i & j, respectively). This results in a lower local 3D scale in region i ($\sim 70\mu m$) compared to region j ($\sim 100\mu m$), corresponding to the fact that the neuronal trace in the more proximal segment is more tortuous than that in the more distal portion.

A mouse retinal ganglion cell from (Badea and Nathans, 2011) is shown in the second column with longer axonal and dendritic arbors compared to collumn 1. We observed small local 3D scales along the axonal trajectory within region i ($\sim 30\mu m$) except in region j where high local 3D scales ($\sim 100\mu m$) could be found. This reflect the structure of the axon as it leaves the retina to join the optic nerve. The dendritic arbor presents a local 3D scale on the whole ($\sim 50\mu m$) but with higher variation than the axonal part. Some dendritic branches exhibit complex shapes (region k) whereas others are more direct (region l).

Then, we focused on an interneuron from the optic lobe of the blowfly (Borst and Haag, 1996) with a more complex structure than the retinal ganglion cell (Figure 2, third column). This blowfly interneuron consists of a long axon and a well-organized dendritic arbor with a long main branch and many other smaller subbranches. We observed again a small local 3D scale ($\sim 30\mu m$) along axonal traces as emphasized in region i, and comparatively high local 3D scales ($\sim 80\mu m$) along the main dendritic branch highlighted in region j. The spatial configuration of the dendritic sub-branches are in contrast more diverse than the main branch as shown in region k (high local 3D scales) and region l (small local 3D scales).

Overall, we compute local measurements of the geometric complexity of a neuron trace, and find they are robust across neuron types, shapes and sizes.

2.3 Local 3D scale in single neuron development

To show the relevance and usefulness of our measure in practice, we reanalyze data from (Santos *et al.*, 2018), as hosted in the NeuroMorpho.Org database. In (Santos *et al.*, 2018), down syndrome cell adhesion molecule (DSCAM) is shown to affect the development of dendritic and axonal arbors in Xenopus laevis tadpoles. Downregulation (downreg) of

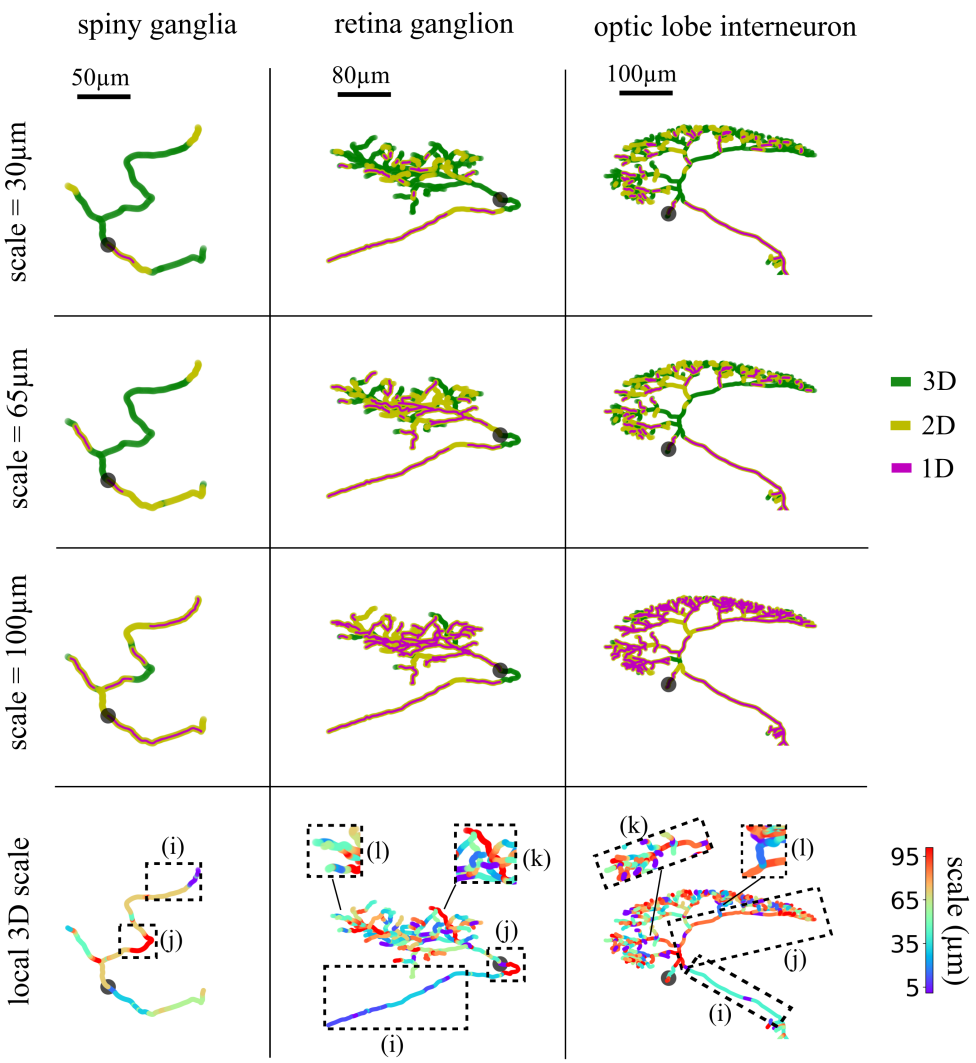

Figure 2: Computation of 3D local scale of different cell types from multiscale dimension decompositions. Top three rows represent the dimension decompositions in different scales, the local 3D scale is then computed from those decompositions (last row). Three cell types at different arbor sizes and animals were selected, the spiny ganglia and the retina ganglion in mouse at the left and middle columns (Li *et al.*, 2019; Badea and Nathans, 2011), the optic lobe interneuron in blowfly at the right column (Borst and Haag, 1996).

DSCAM in tectal neurons enhances the number of branchings and total arbor lengths while in contrast overexpression (overexpr) of DSCAM limits their development. An increase of these features over the course (in days) of tadpole development was also reported upon both downreg and overexpr of DSCAM. Here we reanalyzed this dataset and studied how local 3D scale can reports the formation of dendritic tectal neurons under the overexpr of DSCAM. The local 3D scales between control and overexpr DSCAM in three days are computed and Figure 3A show that the local 3D scale in DSCAM is on average smaller than that in the control. The quantification of mean local 3D scale (Figure 3B, top panel) showed a significant difference between DSCAM and the control at days 2 and 3 while it stays constant across days..

A morphological metric classically used to measure geometrical complexity is tortuosity, the ratio between Euclidean and curvilinear distance. (Santos *et al.*, 2018) show that downreg DSCAM can also increase the tortuosity of the longest

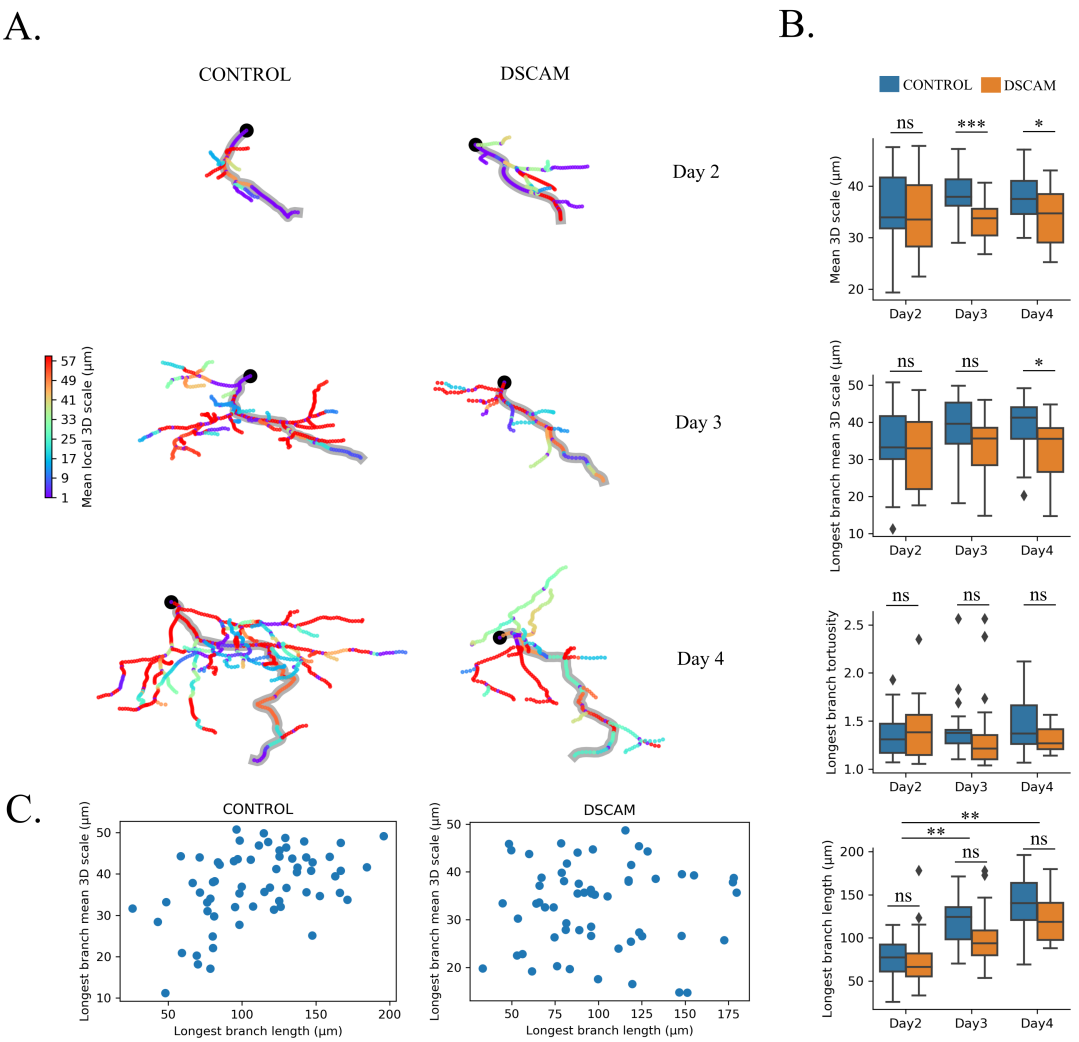

Figure 3: Study of local 3D scale of tectal neurons in developing Xenopus laevis tadpoles with Down syndrome cell adhesion molecule (DSCAM) overexpression (Santos *et al.*, 2018). (A) Examples of local 3D scales between Control and DSCAM in three days. (B) Mean 3D local scale of entire neuronal trace (first), mean 3D local scale of the longest branch (second), tortuosity of the longest branch (third) and longest branch length (forth) were reported between Control and DSCAM in three days. Two-way ANOVA, Student's t-test with Holm-Sidak for multiple comparisons were used, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.005$, **** $p \leq 0.001$. (C) Correlation between mean 3D scale and length of the longest branch in Control and DSCAM.

dendritic branch, but do no report this result in overexpr DSCAM. We thus computed the tortuosity and the length of the longest branch in overexpr DSCAM and observed that there is neither significant difference of the tortuosity between Control and DSCAM nor between days (third of Figure 3B). Contrastingly, the difference in length is significant between days but not between DSCAM-Control (Figure 3B, bottom panel). By computing the mean local 3D scale (Figure 3B, second column), we report a significant difference of DSCAM compared to Control in day 4. This can be seen in Figure 3A where the longest branch is marked as dark gray band. Even the difference among days is not clear in case of mean local 3D scale for the longest branch but we also found an interesting correlation between its length and mean local 3D scale in case of Control but not for DSCAM (Figure 3C). This indicates not only a relationship of local

3D scales to the development of tectal neurons but also a non-invertible effect between downreg and overexpr DSCAM. The study of local 3D scale in case of downreg DSCAM of tectal neurons and retinal ganglion cells was also reported (Figure 3-figure supplements 1 and 2).

From this we conclude that our local 3D scale is a biologically relevant information that can measure subtle differences that were missed using classical tools.

2.4 Local 3D scale in different cell types of larval zebrafish brain

A database including over 2000 neuronal traces across the whole brain of the larval zebrafish has been published in (Kunst *et al.*, 2019). They clustered the neuronal traces into classes using the NBLAST algorithm (Costa *et al.*, 2016) and found that while some are well known such as retina ganglion cells (RGCs), mitral cells or oculomotor neurons, others were discovered for the first time. We reanalyzed this dataset by studying the local 3D scales in some of those cell types. We first looked at the RGCs consisting of two groups that are laterally symmetric. We studied the local 3D scale along each RGC's trajectory starting from retina and arriving in the tectum. Here we focused on a single group due to the symmetry. Local 3D scales of some example RGCs are shown in Figure 4A. These present different spatial patterns along their trajectories notably in the segment just prior to entering the tectum and in the region of the optic nerve following exit from the retina. Figure 4B indicates the local 3D scales of all neurons in the group where the heterogeneity is clear acrossed regions. To quantify this, we divided the RGC trajectories according to segments along the path from the retina to tectum, i.e. within the retina, at the exit from retina, at the midline, just prior to entering the tectum and within the tectum. We observed some significant differences among these regions. For example, the mean local 3D scales within the retina and tectum were lower ($\sim 45\mu m$) than at the exit from the retina and entrance to the tectum ($\sim 60\mu m$) while the midline region ($\sim 50\mu m$) was not significantly different compared to within the retina and tectum.

In contrast to RGCs which in zebrafish project to a single target, we then analyzed the trajectories of mitral cells originating in the olfactory bulb and projecting to multiple destinations such as the pallium of the cerebellum or the habenula (Miyasaka *et al.*, 2014). Similarly to the analysis of RGCs, we focused on a group of mitral cells from one side. We found different shape complexities across regions (Figures 4D and E) with low local 3D scales at the olfactory bulb ($\sim 40\mu m$) and habenular formation ($\sim 37\mu m$) compared to the highest value in the pallium ($\sim 57\mu m$).

Overall, our analysis demonstrates that our method can be applied to large datasets, and is capable of identifying both heterogeneity across single axons of the same type and significant pattern whithin each type.

2.5 Local 3D scale in different regions of larval zebrafish brain

In addition to a collection of over 2000 neuronal traces, the whole zebrafish brain atlas in (Kunst *et al.*, 2019) was manually divided by the author into 36×2 bilateral symmetric regions. A brain-wiring diagram showing the connection strength between those regions was then built and different connection strengths were reported in fore-, mid- and hindbrain. The connection strength in each region was deduced from the sum of neurite lengths of neurons originating from/connecting to each region. We computed the local 3D scales of all neuronal traces in the database (Figure 5A). The mean local 3D scales varied across brain regions (Figure 5B). The mean local 3D scale (Figure 5C) in the midbrain ($\sim 47\mu m$) is on average higher than in the forebrain ($\sim 53\mu m$). This correlates with the observation from the authors that connection strengths in those three parts of zebrafish brain. In particular, regions such as olfactory bulb (OB) and habenula have (Hb) have a small local 3D scale compared to the trigeminal ganglion (TG) or pituitary (Pi) with much higher local 3D scales (Figure 5D). However, the distributions of the local 3D scales within each region have a high variance, which could be due to diverse cell types adopting different local shape structures when starting from or arriving to a brain region.

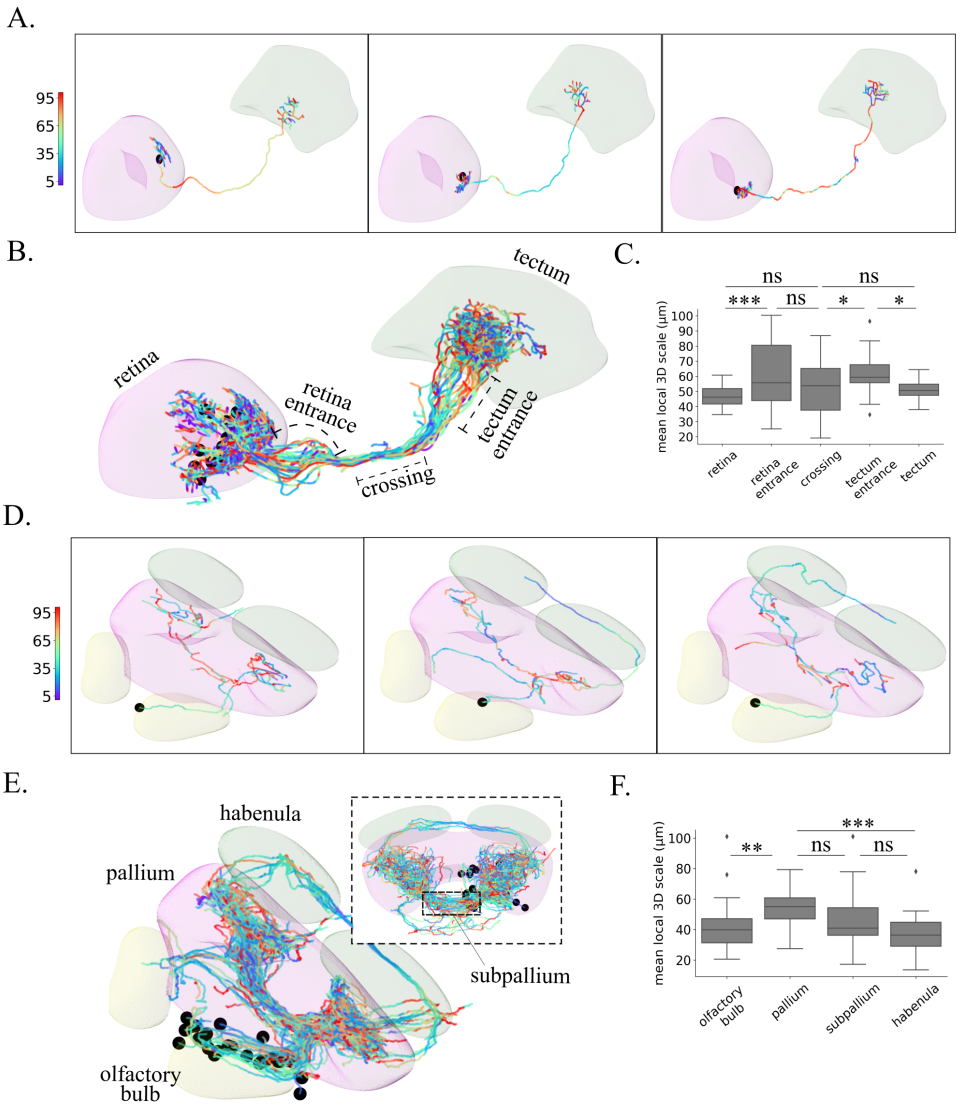

Figure 4: Study of local 3D scales of retina ganglion cells (A,B,C) and mitral cells (D,E,F) in larval zebrafish (Kunst *et al.*, 2019). (A) some examples of local 3D scale of retina ganglion cell starting from right retina and connecting to left tectum. (B) Local 3D scale of all right retina ganglion cells, different regions along neuronal trace were studied, quantification of local 3D scales in those regions were reported in (C). (D) some examples of local 3D scale of mitral cell starting from left olfactory bulb and having multiple outputs in pallium of cerebellum and habenula. (E) Local 3D scale of all left mitral cells, different regions along neuronal trace were studied, quantification of local 3D scales in those regions were reported in (F). One-way ANOVA, Student's t-test with Holm-Sidak for multiple comparison were used, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.005$, **** $p \leq 0.001$.

We then compared the mean local 3D scale with neuron density and the volume normalized connection strength and number of branches in individual regions (Figure 5E). In contrast to the correlation found in three parts of brain, we observed no clear relationship between the mean local 3D scale and those three features at the individual region level. We found that very high connection strength, branch numbers or neuron number ratios do not imply a high mean local 3D scale as in the example of the anterior reticular formation (aRF) where a medium local 3D scale was reported (Figure 5F, middle panel). In contrast, regions with small values of those features can have small/high mean local 3D

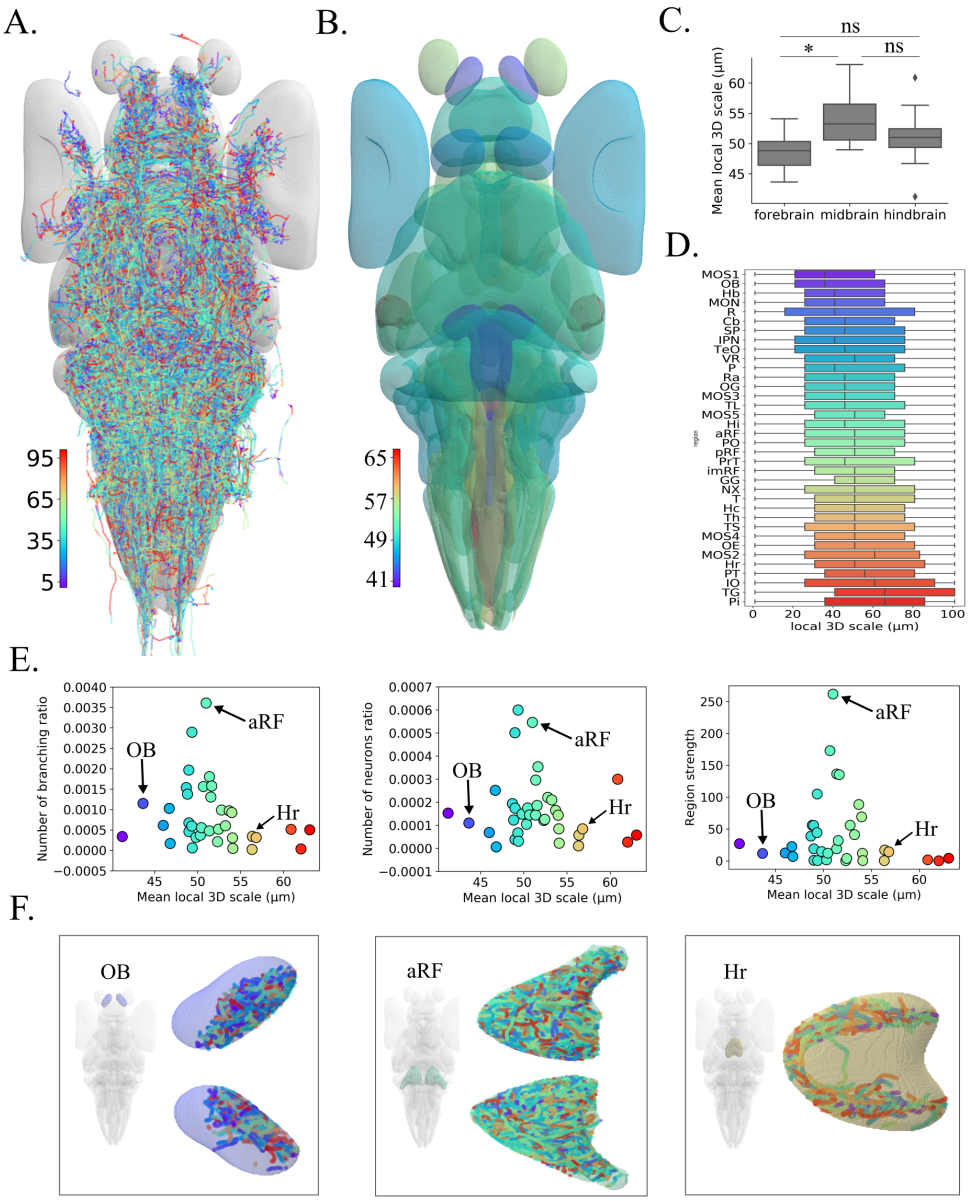

Figure 5: Study of local 3D scales across brain regions of larval zebrafish (Kunst *et al.*, 2019). (A) Local 3D scales of all neuronal traces. (B) Mean local 3D scale by brain regions. (C) Mean local 3D scales in fore- mid- and hindbrain .(D) Local 3D scales distribution by brain regions, the color is mapped to mean 3D local scale. (E) correlation between mean local 3D scale and number of branchings (left), number of neurons (middle), connection strength (right). All features are normalized by region volume, the connection strength is calculated based on sum of neurite lengths for neurons from/to the brain region. (F) Examples of 3D local scales of neuronal traces in some brain regions, the OB (left) with low mean 3D scale and low connection strength, the aFR (middle) with medium mean 3D scale and high connectivity strength, the Hr (right) with high mean 3D scale and low connectivity strength; only the portion of the neurons within the brain regions are shown but the 3d scales were computed with the whole traces. One-way ANOVA, Student's t-test with Holm-Sidak for multiple comparison were used, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.005$, **** $p \leq 0.001$.

scale as in the example of the OB and rostral hypothalamus (Hr) (Figure 5F, left & right panels). There is no clear relationship between those features, showing that the computed 3d local scale provide additional orthogonal information. This shows that the local 3D scale is significant at the whole brain level, adding a level of description to previously used information.

3 Materials and Methods

3.1 Intrinsic dimension decomposition of neuronal branches based on scale-space theory

The *intrinsic dimension* of a curve is the smallest dimension in which it can be expressed without significant loss of information. We modeled a neuronal branch as a 3D parametric curve $\gamma(u) = (x(u), y(u), z(u))$ for $u \in [0, 1]$, which we then decompose into local fragments lying intrinsically on 1D line, 2D plane or 3D. Note that such decomposition needs not be unique. For example, not taking scales into account, two consecutive lines followed by one plane can in theory be decomposed into one 3D fragment, one 1D and one 3D fragments, two consecutive 1D and one 2D fragments, or two consecutive 2D fragments (Figure 1-supplement figure 1). The last two decomposition schemes are both meaningful and their combination represents a hierarchical decomposition of γ where linear fragments are usually parts of a larger planar fragment. We compute the decomposition of a curve intrinsically by looking at the curvature and torsion. Let us define the curvature as κ of γ by $\kappa = \|\gamma' \times \gamma''\|/\|\gamma'\|^3$, which corresponds to the inverse of the radius of the best approximation of the curve by a circle locally, the osculating circle. The torsion τ of γ is $\tau = ((\gamma' \times \gamma'') \cdot \gamma''')/\|\gamma' \times \gamma''\|^2$ and corresponds to the rate of change of the plane that includes the osculating circle. We then determine that if κ is identically equal to zero on a fragment then that fragment is a 1D line, and similarly that τ being identically equal to zero defines a 2D arc curve. Thus we define a linear indicator L of γ:

$$L(u) = \begin{cases} 1, & \text{if } \kappa(u) \leq \varepsilon_\kappa \\ 0, & \text{otherwise,} \end{cases}$$

and a planar indicator H of γ:

$$H(u) = \begin{cases} 1, & \text{if } |\tau(u)| \leq \varepsilon_\tau \\ 0, & \text{otherwise,} \end{cases}$$

where ε_κ and ε_τ are the tolerances for computed numerical errors. We note that only using the H indicator is not sufficient to estimate the 2D plane. For example, a 2D plane fragment composed of a 1D line followed by a 2D arc cannot be entirely identified as 2D since the torsion is not defined as the curvature tends to 0. We therefore consider the planar-linear indicator $T = L \cup H$ instead of H for characterizing the curve according to such hierarchical order.

In practice, the resulted dimension decomposition will be closely tied to a "scale" at which the curve is studied. Scale-spaces of curves have been described in 2D (Mackworth and Mokhtarian, 1988) or 3D (Mokhtarian, 1988) in particular, giving a robust meaning to that intuition. They define γ_s such that $\gamma_0 = \gamma$ and increasing s lead to increasingly simplified curves. The scaled curve γ_s can be calculated by convolving γ with a Gaussian kernel of standard deviation s (Witkin, 1983), or by using mean curvature flow (Gage and Hamilton, 1986). In 2D, it has been shown for example that a close curve under a mean curvature motion scale space will roundup with increasing s and eventually disappear into a point (Grayson, 1987).

However, the selection of a scale of interest representing a meaningful description of a curve is not straightforward. For example, the scale defined as the standard deviation of the Gaussian kernel above can be affected by the kernel length and the curve sampled points. To make the scales become a more relevant physical/biological measurement, we employed the radius of curvature r_κ, the inverse of the curvature κ. Indeed, the level of detail on the curve can be characterized by r_κ where small r_κ represents high level of detail looking at small objects such as small bumps and tortuosity and large r_κ represents low level of detail/large object and features such as plateaus or turns []. To ease the

interpretation and usage of scale spaces, we associate to standard deviation s a scale in micron $\hat{r}_\kappa$, by determining for each point u whose $r_\kappa(u) < \hat{r}_\kappa$ a standard deviation s such that $r_{\kappa_s}(u) = \hat{r}_\kappa$. This therefore creates a mapping function between a scale of interest $\hat{r}_\kappa$ which is physical relevant and a list of standard deviations which are less familiar.

Assuming S a list of standard deviations deduced from the scale $\hat{r}_\kappa$, to estimate the dimension decomposition of γ at those scales, the first step is to compute L_s and T_s, for all $s \in S$. This can be done by calculating the curvatures κ_s and the torsions τ_s. We then exclude the fragments having lengths smaller than a given fragment length threshold ε_ω at every s to eliminate small irrelevant fragments. From L_s and T_s, the linear fragments denoted as $\mathcal{D}_{L_s}$ and planar-linear fragments denoted as $\mathcal{D}_{T_s}$ are deduced across all $s \in S$. An example of this sequence of dimension decompositions is illustrated in Figure 1-figure supplement 2 where $\mathcal{D}_{L_s}$ and $\mathcal{D}_{T_s}$ are superimposed.

The curve is completely 3D at small s, several linear and planar fragments then appear at higher s, progressively fuse to form larger linear and planar fragments and eventually converge to one unique line at a very high value of s.

In the second step, we estimate the most durable combination of fragments from the linear segments $\mathcal{D}_{L_s}$ and linear-planar segments $\mathcal{D}_{T_s}$ calculated in the first step. We compute the number of fragments at every $s \in S$, to measure how long each combination of fragments exists for. We then select the combination of fragments remaining for the longest subinterval of S. Knowing that each fragment in the combination can have different lengths among that subinterval, we thus select the candidate with the longest length. In cases of overlapping fragments, we split the overlap in half. Also of note, we repeat the step twice, first to estimate the best combination of linear-planar segments from $\mathcal{D}_{T_s}$ input and mark the corresponding subinterval, then to estimate the best combination of linear fragments from $\mathcal{D}_{L_s}$ input within that subinterval. As seen in Figure 1-figure supplement 2, the curve is decomposed at every s by consecutive planar/nonplanar fragments, then the linear fragments are identified within each planar fragment. The final result is the hierarchical dimension decomposition of the curve γ at a given scale of interest $\hat{r}_\kappa$.

3.2 Evaluation of dimension decomposition on simulated curve

The simulated curve consists of consecutive fragments of 1D lines, 2D planes and 3D. The simulation of 1D line is done by simply sampling a sequence of points on an arbitrary axis (e.g. x axis), then rotating it to a random orientation in 3D. The simulation of random yet regular fragments in 3D embedded in a 2D plane is more challenging. We tackle this issue by using the active Brownian motion model (Volpe *et al.*, 2014). Active Brownian motion of particles is an extended version of the standard Brownian motion (Uhlenbeck and Ornstein, 1930) by adding two coefficients: the translational speed to control directed motion and the rotational speed to control the orientation of the particles. At each time point, we generate the new coordinates $(x, y, 1)$ by the following formulas.

$$
\begin{aligned}
D_T &= \frac{k_B\, T}{6\,\pi\,\eta\, R} \\
D_R &= \frac{k_B\, T}{8\,\pi\,\eta\, R^3} \\
\frac{d}{dt}\,\varphi(t) &= \Omega + \sqrt{2\,D_R}\,W_\varphi \\
\frac{d}{dt}\,x(t) &= v\,\cos\varphi(t) + \sqrt{2\,D_T}\,W_x \\
\frac{d}{dt}\,y(t) &= v\,\sin\varphi(t) + \sqrt{2\,D_T}\,W_y,
\end{aligned}
$$

where D_T and D_R are the translational and rotational coefficients, k_B the Boltzmann constant, T the temperature, η the fluid viscosity, R particle radius, φ rotation angle, Ω angular velocity and W_φ, W_x, W_y independent white noise. After generating the simulated intrinsic 2D fragment, a random rotation is applied. For simulating a 3D fragment, we extended the Active Brownian motion model in 2D (Volpe *et al.*, 2014) to 3D as follows:

$$
\begin{aligned}
\frac{d}{dt}\,\varphi(t) &= \cos\Omega + \sqrt{2\,D_R}\,W_\varphi \\
\frac{d}{dt}\,\theta(t) &= \sin\Omega + \sqrt{2\,D_R}\,W_\theta \\
\frac{d}{dt}\,x(t) &= v\,\cos\theta(t)\,\sin\varphi(t) + \sqrt{2\,D_T}\,W_x \\
\frac{d}{dt}\,y(t) &= v\,\sin\theta(t)\,\sin\varphi(t) + \sqrt{2\,D_T}\,W_y \\
\frac{d}{dt}\,z(t) &= v\,\cos\varphi(t) + \sqrt{2\,D_T}\,W_z,
\end{aligned}
$$

where (φ, θ) are spherical angles. We simulate the curve γ with a sequence of n_ω consecutive fragments of varying random intrinsic dimensions. The curve γ is then resampled equally with n_γ points and added by white noise $\mathcal{N}(0, \sigma^2)$. We set $n_\omega \leq 5$, $n_\gamma = 1000$ points and vary σ between 1 and 30 μm. The upper value of $\sigma = 30\mu m$ is high enough to corrupt local details of a simulated fragment with about $100\mu m$ of length in our experiment. The metric used to measure the accuracy of the intrinsic dimension decomposition is defined by $(1/n_\omega)\sum_i \max_j \xi(\omega_i, \hat{\omega}_j)$, where $\{\omega_1, \omega_2, \ldots, \omega_{n_\omega}\}$ are the simulated intrinsic fragments, $\{\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_{n_{\hat{\omega}}}\}$ the estimated intrinsic fragments and $\xi(.,.)$ the F_1 score (Rijsbergen, 1979) calculated by:

$$
2 \times \frac{Precision \times Recall}{Precision + Recall},
$$

where

$$
Precision = \frac{|\omega \cup \hat{\omega}|}{|\omega|} \quad \text{and} \quad Recall = \frac{|\omega \cup \hat{\omega}|}{|\hat{\omega}|}.
$$

The metric is between 0.0 and 1.0 and measures an average accuracy across fragments. In order to evaluate the decomposed intrinsic dimensions, we look at the curve γ at various scales $r_\kappa = 1, 2, \ldots, 100\mu m$, then execute the dimension decomposition for every r_κ and take the one corresponding to the largest accuracy. Setting $r_\kappa = 100\mu m$ is large enough to get a smoothed curve whose simulated intrinsic fragments are no longer preserved. In addition, we also removed estimated fragments whose lengths were smaller than ε_ω. Here we set $\varepsilon_\omega = 5\%$ of the curve length since it does not surpass the minimal length of simulated intrinsic fragments. Our method is also compared with a baseline method (Yang *et al.*, 2016; Ma *et al.*, 2017) that first iteratively assigns each point on the curve as linear/nonlinear by a collinearity criterion, then assigns nonlinear points as planar/nonplanar by a coplanarity criterion. Prior to this, the curve was also denoised based on (Czesla *et al.*, 2018).

3.3 Local 3D scale computation

Assuming a scale interval of interest, the local 3D scale of a point on the curve is computed by iteratively checking at each scale the intrinsic dimension of that point and taking the scale at which the local scale is not 3D anymore. In some cases, the dimension transformation of a point is not very smooth since the point at a 2D/1D scale can convert back to 3D. We therefore select the first scale of the longest 2D/1D subinterval as local 3D scale. In case no subintervals are found, it means that the point remains 3D across the whole interval and its local 3D scale is set as the highest scale.

3.4 Local 3D scale analysis of neuronal trace

We consider a neuronal trace as a 3D tree and first resample it with a spacing distance between two points equal to $1\mu m$. The B-spline interpolation (Boor, 1978) with order 2 is used for the resampling. We then extract the neuronal trace

into distinct sub-branches, since the intrinsic dimension decomposition and local 3D scale are computed on individual neuronal branches. To decompose the tree, we first take the longest branch, then repeat this process for all subtrees extracted along the longest branch. Of note, there are different ways to decompose a tree for example cutting the tree based on its branching positions. We examined both methods and found that the "longest branch" method produced more accurate results. Next, the local 3D scale is computed for each extracted sub-branch within a scale interval of interest. The scale is defined as the radius of curvature and can be selected based on lengths of branch or studied region. For example in the case of the Xeonopus Laevis tadpoles dataset (Santos *et al.*, 2018) we select scales from 1 to $60\mu m$ which is sufficient to study neuronal branches with about $180\mu m$ in length (i.e. a $60\mu m$ radius of curvature corresponds to a semicircle with $60\pi \approx 180\mu m$). In the case of the zebrafish brain dataset (Kunst *et al.*, 2019), we select the scales from 1 to $100\mu m$ based on the length of the largest region, which is about $300\mu m$. Moreover, we can exclude the sub-branches whose lengths are smaller than a given threshold. For example when analyzing the zebrafish brain, we set the threshold to $5\mu m$ which is half the length of the smallest brain region. Branches with lengths smaller than $5\mu m$ have a very small contribution to the local 3D scale result. This is also useful to reduce artifact caused by manual tracing of neurons. Finally, the local 3D scales at branch points along the neuronal trace are ignored due to the ambiguity.

3.5 Tool for neuronal traces manipulation

We provide a free Python toolset for seamless handling of 3D neuronal traces data. It supports reading from standard formats (swc, csv, etc), visualizing, resampling, denoising, extracting sub-branches, calculating basic features (branches, lengths, orientations, Strahler order (Horton, 1945), etc), the proposed intrinsic dimension decompositions and the local 3D scales. The toolset is part of the open source Python library GeNePy3D (https://msphan.gitlab.io/genepy3d) giving full access to manipulation and interaction of various kinds of geometrical 3D objects (trees, curves, point cloud, surfaces).

4 Discussion and conclusion

We present a novel method for the multiscale estimation of intrinsic dimensions along an open 3D curve and use it to compute a *local 3D scale*, which we propose as a new measurement to characterize neuronal arbors. We demonstrate that our methods are accurate and robust, and apply them to published trace data to present its relevance and usefulness from single axon to whole brain scales.

Mathematically, scale spaces of curves are trickier and a lot less studied in 3D than in 2D. In 2D for example, it is known that scale spaces based on curvature motion have better properties than Gaussian ones (Cao, 2003) but are not readily available in 3D. While the approach chosen here works in practice, additional studies of the mathematical foundation of 3D scale spaces would lead to simpler and more robust algorithms. Moreover, this work is complementary to other neural arbor analysis techniques; for example it does not concern itself with the topology of the arbor but decomposes it in 3D curves.

Exploration of local geometrical properties of neurites may be of particular importance in the precise study of neuronal traces. For example in (K. Matho et al., in preparation) we use it to locally and globally probe the topology and topography of axons innervating the Medial Nucleus of the Trapezoid Body (MNTB). More generally, we know numerous phenomena governing the final shape of neurites, like attractive or repulsive chemical guidance cues (Lowery and Van Vactor, 2009), mechanical cues (Gangatharan *et al.*, 2018), pruning (Holcomb *et al.*, 2013) or migration. A fine analysis of resulting neuronal traces would provide key information for those studies toward a deeper understanding of brain development.

Importantly, an implementation of the proposed algorithms, along with all codes to reproduce the figures, is openly available and easy to use, making it a potential immediate addition to computational neuroanatomy studies. More generally, it is part of GeNePy3D, a larger quantitative geometry Python package with methods from spatial statistics or

computational geometry as well. These tools hold great promise for providing theoretical and computational metrics needed to transform advanced microscopy into neurobiological understanding.

References

Abdeladim, L., Matho, K. S., Clavreul, S., Mahou, P., Sintes, J.-M., Solinas, X., Arganda-Carreras, I., Turney, S. G., Lichtman, J. W., Chessel, A., Bemelmans, A.-P., Loulier, K., Supatto, W., Livet, J., and Beaurepaire, E. (2019). Multicolor multiscale brain imaging with chromatic multiphoton serial microscopy. *Nature Communications*, **10**(1), 1–14.

Ascoli, G. A. (2006). Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nature Reviews. Neuroscience*, **7**(4), 318–324.

Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). NeuroMorpho.Org: A Central Resource for Neuronal Morphologies. *Journal of Neuroscience*, **27**(35), 9247–9251. Publisher: Society for Neuroscience Section: Toolbox.

Badea, T. C. and Nathans, J. (2011). Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: analysis of wild type and mutant cells using genetically-directed sparse labeling. *Vision Research*, **51**(2), 269–279.

Bates, A. S., Janssens, J., Jefferis, G. S., and Aerts, S. (2019). Neuronal cell types in the fly: single-cell anatomy meets single-cell genomics. *Current Opinion in Neurobiology*, **56**, 125–134.

Bates, A. S., Manton, J. D., Jagannathan, S. R., Costa, M., Schlegel, P., Rohlfing, T., and Jefferis, G. S. (2020). The natverse, a versatile toolbox for combining and analysing neuroanatomical data. *eLife*, **9**, e53350. Publisher: eLife Sciences Publications, Ltd.

Boor, C. d. (1978). *A Practical Guide to Splines*. Applied Mathematical Sciences. Springer-Verlag, New York.

Borst, A. and Haag, J. (1996). The intrinsic electrophysiological characteristics of fly lobula plate tangential cells: I. Passive membrane properties. *Journal of Computational Neuroscience*, **3**(4), 313–336.

Cao, F. (2003). *Geometric Curve Evolution and Image Processing*. Lecture Notes in Mathematics. Springer-Verlag, Berlin Heidelberg.

Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S., and Jefferis, G. S. X. E. (2016). NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. *Neuron*, **91**(2), 293–311.

Cuntz, H., Forstner, F., Borst, A., and Häusser, M. (2010). One Rule to Grow Them All: A General Theory of Neuronal Branching and Its Practical Application. *PLOS Computational Biology*, **6**(8), e1000877. Publisher: Public Library of Science.

Czesla, S., Molle, T., and Schmitt, J. H. M. M. (2018). A posteriori noise estimation in variable data sets - With applications to spectra and light curves. *Astronomy & Astrophysics*, **609**, A39.

Digne, J., Morel, J.-M., Souzani, C.-M., and Lartigue, C. (2011). Scale Space Meshing of Raw Data Point Sets. *Computer Graphics Forum*, **30**(6), 1630–1642. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2011.01848.x.

Gage, M. and Hamilton, R. S. (1986). The heat equation shrinking convex plane curves. *Journal of Differential Geometry*, **23**(1), 69–96.

Gangatharan, G., Schneider-Maunoury, S., and Breau, M. A. (2018). Role of mechanical cues in shaping neuronal morphology and connectivity. *Biology of the Cell*, **110**(6), 125–136. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/boc.201800003.

Gerhard, S., Andrade, I., Fetter, R. D., Cardona, A., and Schneider-Mizell, C. M. (2017). Conserved neural circuit structure across Drosophila larval development revealed by comparative connectomics. *eLife*, **6**, e29089. Publisher: eLife Sciences Publications, Ltd.

Grayson, M. A. (1987). The heat equation shrinks embedded plane curves to round points. *Journal of Differential Geometry*, **26**(2), 285–314.

Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, **500**(7461), 168–174.

Holcomb, P. S., Hoffpauir, B. K., Hoyson, M. C., Jackson, D. R., Deerinck, T. J., Marrs, G. S., Dehoff, M., Wu, J., Ellisman, M. H., and Spirou, G. A. (2013). Synaptic Inputs Compete during Rapid Formation of the Calyx of Held: A New Model System for Neural Development. *Journal of Neuroscience*, **33**(32), 12954–12969. Publisher: Society for Neuroscience Section: Articles.

Horton, R. E. (1945). Erosional Development Of Streams And Their Drainage Basins; Hydrophysical Approach To Quantitative Morphology. *GSA Bulletin*, **56**(3), 275–370. Publisher: GeoScienceWorld.

Kunst, M., Laurell, E., Mokayes, N., Kramer, A., Kubo, F., Fernandes, A. M., Förster, D., Dal Maschio, M., and Baier, H. (2019). A Cellular-Resolution Atlas of the Larval Zebrafish Brain. *Neuron*, **103**(1), 21–38.e5.

Li, R., Zhu, M., Li, J., Bienkowski, M. S., Foster, N. N., Xu, H., Ard, T., Bowman, I., Zhou, C., Veldman, M. B., Yang, X. W., Hintiryan, H., Zhang, J., and Dong, H.-W. (2019). Precise segmentation of densely interweaving neuron clusters using G-Cut. *Nature Communications*, **10**(1), 1549.

Lindeberg, T. (1990). Scale-space for discrete signals. *IEEE Trans. PAMI*, **12**(3), 234–254.

Lowery, L. A. and Van Vactor, D. (2009). The trip of the tip: understanding the growth cone machinery. *Nature reviews. Molecular cell biology*, **10**(5), 332–343.

Ma, X., Zhao, T., Wen, R., Wu, Z., and Wang, Q. (2017). Motion recognition based on concept learning. In *I2MTC*, pages 1–6.

Mackworth, S. and Mokhtarian, F. (1988). The renormalized curvature scale space and the evolution properties of planar curves. In *Proceedings CVPR '88: The Computer Society Conference on Computer Vision and Pattern Recognition*, pages 318–326. ISSN: 1063-6919.

Magliaro, C., Callara, A. L., Vanello, N., and Ahluwalia, A. (2019). Gotta Trace 'em All: A Mini-Review on Tools and Procedures for Segmenting Single Neurons Toward Deciphering the Structural Connectome. *Frontiers in Bioengineering and Biotechnology*, **7**.

Meinertzhagen, I. A. (2018). Of what use is connectomics? A personal perspective on the Drosophila connectome. *Journal of Experimental Biology*, **221**(10). Publisher: The Company of Biologists Ltd Section: COMMENTARY.

Mihaljević, B., Larrañaga, P., Benavides-Piccione, R., Hill, S., DeFelipe, J., and Bielza, C. (2018). Towards a supervised classification of neocortical interneuron morphologies. *BMC Bioinformatics*, **19**(1), 511.

Miyasaka, N., Arganda-Carreras, I., Wakisaka, N., Masuda, M., Sümbül, U., Seung, H. S., and Yoshihara, Y. (2014). Olfactory projectome in the zebrafish forebrain revealed by genetic single-neuron labelling. *Nature Communications*, **5**(1), 3639. Number: 1 Publisher: Nature Publishing Group.

Mokhtarian, F. (1988). Multi-scale description of space curves and three-dimensional objects. In *CVPR*, pages 298–303.

Petilla Interneuron Nomenclature Group, Ascoli, G. A., Alonso-Nanclares, L., Anderson, S. A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsáki, G., Cauli, B., Defelipe, J., Fairén, A., Feldmeyer, D., Fishell, G., Fregnac, Y., Freund, T. F., Gardner, D., Gardner, E. P., Goldberg, J. H., Helmstaedter, M., Hestrin, S., Karube, F., Kisvárday, Z. F., Lambolez, B., Lewis, D. A., Marin, O., Markram, H., Muñoz, A., Packer, A., Petersen, C. C. H., Rockland, K. S., Rossier, J., Rudy, B., Somogyi, P., Staiger, J. F., Tamas, G., Thomson, A. M., Toledo-Rodriguez, M., Wang, Y., West, D. C., and Yuste, R. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews. Neuroscience*, **9**(7), 557–568.

Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworth-Heinemann, USA, 2nd edition.

Santos, R. A., Fuertes, A. J. C., Short, G., Donohue, K. C., Shao, H., Quintanilla, J., Malakzadeh, P., and Cohen-Cory, S. (2018). DSCAM differentially modulates pre- and postsynaptic structural and functional central connectivity during visual system wiring. *Neural Development*, **13**(1), 22.

Sapiro, G. (2006). *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, Cambridge, 1 edition.

Schneider-Mizell, C. M., Gerhard, S., Longair, M., Kazimiers, T., Li, F., Zwart, M. F., Champion, A., Midgley, F. M., Fetter, R. D., Saalfeld, S., and Cardona, A. (2016). Quantitative neuroanatomy for connectomics in Drosophila. *eLife*, **5**, e12059. Publisher: eLife Sciences Publications, Ltd.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review*, **36**(5), 823–841.

Volpe, G., Gigan, S., and Volpe, G. (2014). Simulation of the active Brownian motion of a microswimmer. *American Journal of Physics*, **82**(7), 659–664.

Witkin, A. P. (1983). Scale-space Filtering. IJCAI, pages 1019–1022, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Yang, J., Yuan, J., and Li, Y. (2016). Parsing 3d motion trajectory for gesture recognition. *Journal of Visual Communication and Image Representation*, **38**, 627–640.