On the Statistical Analysis of Complex Tree-shaped 3D Objects

Guan Wang, Hamid Laga, Anuj Srivastava Fellow, IEEE

Abstract—How can one analyze detailed 3D biological objects, such as neurons and botanical trees, that exhibit complex geometrical and topological variation? In this paper, we develop a novel mathematical framework for representing, comparing, and computing geodesic deformations between the shapes of such tree-like 3D objects. A hierarchical organization of subtrees characterizes these objects – each subtree has the main branch with some side branches attached – and one needs to match these structures across objects for meaningful comparisons. We propose a novel representation that extends the Square-Root Velocity Function (SRVF), initially developed for Euclidean curves, to tree-shaped 3D objects. We then define a new metric that quantifies the bending, stretching, and branch sliding needed to deform one tree-shaped object into the other. Compared to the current metrics, such as the Quotient Euclidean Distance (QED) and the Tree Edit Distance (TED), the proposed representation and metric capture the full elasticity of the branches (i.e., bending and stretching) as well as the topological variations (i.e., branch death/birth and sliding). It completely avoids the shrinkage that results from the edge collapse and node split operations of the QED and TED metrics. We demonstrate the utility of this framework in comparing, matching, and computing geodesics between biological objects such as neurons and botanical trees. The framework is also applied to various shape analysis tasks: (i) symmetry analysis and symmetrization of tree-shaped 3D objects, (ii) computing summary statistics (means and modes of variations) of populations of tree-shaped 3D objects, (iii) fitting parametric probability distributions to such populations, and (iv) finally synthesizing novel tree-shaped 3D objects through random sampling from estimated probability distributions.

Index Terms—Tree-shape space, geodesics, metrics, 3D shape variability, 3D tree synthesis, symmetry analysis, symmetrization, 3D atlas, tree classification, square-root velocity function (SRVF), correspondence, registration, topological variability.

1 INTRODUCTION

Tree-like 3D objects are ubiquitous, especially in delivery systems such as vascular systems, airway trees, blood vessels in the eye, plant roots and shoots, and neuronal structures in the brain. Analyzing and understanding the 3D structural variability of such tree-like biological objects are of great interest from multiple scientific perspectives. Shapes of objects both constrain and enable their functionalities in larger biological systems. Thus, a mathematical characterization of shapes can help provide insights into objects’ functional roles in such biological phenomena as genesis, growth, and disease. For instance, studies of changes in the 3D structure of plant roots could improve their water and nutrient uptake efficiency. Several papers have also shown that neuronal morphology, i.e., their types, geometry, and topology, and their connections are the key to discerning how neurons integrate information and subsequently explaining brain activity and its functions [1], [2], [3]. Also, statistical analysis of neuron morphology is crucial to understanding brain functionality and in characterizing cognitive health. Alterations in neuron morphology are not only due to a normal aging [4] but can also be the consequence of a pathology, e.g., senile dementia [5] and Alzheimer disease [6].

Existing techniques for modeling shape variability are mostly limited to objects with fixed topology i.e., objects that only bend and stretch [7], [8], [9], [10], [11]. Tree-shaped objects, however, are more challenging because they exhibit variability in (1) their geometry, in terms of the shape of the individual branches, e.g., axons and dendrites in neuronal structures, which can bend and stretch, and (2) topology, in terms of the structural relationships between those branches. One consequence of variable geometry and topology is that finding an optimal registration, i.e., putting in correspondence parts (points, curves, and branches) across such objects, becomes a challenging problem. Several recent works have explored this problem. In particular, Feragen et al. [12], [13], [14], [15] proposed a tree-shape space for computing statistics on tree-shaped objects such as airway trees. In this framework, tree-shaped objects are represented as graphs whose nodes correspond to bifurcation points, and their edges encode the shape of the segment between two bifurcation points. Due to its computational complexity, since finding an optimal matching between trees is an NP-complete problem, the framework is limited to simple trees with a few branches. Wang et al. [16], [17] extended the framework to handle complex botanical trees by pre-computing the correspondences. A such, correspondences and geodesics are computed separately, using different optimality criteria, which can be sub-optimal. Both works use metrics such as the Tree Edit Distance (TED) or the Quotient Euclidean Distance (QED), where topological changes are modeled by edge collapse and node split operations.
Consequently, geodesics exhibit significant shrinkage, especially between trees with large topological differences and between trees that significantly bend.

Duncan et al. [18] addressed this problem by representing 2D tree-shaped objects using curves that meet at the bifurcation points. While this representation is efficient, it is limited to simple 2D neuronal trees, i.e., trees composed of one main branch and some lateral branches. In this paper, we generalize the framework of Duncan et al. [18] to complex and arbitrary tree-shaped 3D objects. We propose a novel recursive representation where a tree-shaped object is characterized by a hierarchical organization of subtrees, i.e., some side branches attached to the main branch. The task of comparing and registering two individual branches requires an optimal bending and stretching of one branch to align it onto the other. The task of comparing and registering two subtrees, each composed of the main branch and multiple side branches, is complex. It requires: (i) optimally sliding branches along the main branch to capture the topological differences between the two subtrees, and (ii) bending and stretching the branches to align them. By defining a metric that quantifies these deformations, i.e., bending, stretching, and sliding branches, one can specify geodesic deformations between two subtrees as the shortest path under this metric, after factoring out shape-preserving transformations such as translation, scale, rotation, and re-parameterization. (Recall that parameterization defines correspondence.) The geodesic length between two tree-shaped objects is then a measure of dissimilarity between the two objects. Defining this metric recursively across the entire hierarchy of the tree-shaped objects allows us to analyze complex tree-shaped 3D objects.

In this paper, we develop this mathematical framework for deriving tools that:

- Compute correspondences and geodesic deformations between tree-like 3D objects even when undergoing significant bending, stretching, and complex topological deformations.
- Compute statistical atlases, i.e., means and principal modes of variation, of collections of tree-like 3D objects.
- Characterize the geometric and structural variability within a collection using probability distributions.
- Develop a mechanism for synthesizing 3D tree-like structures randomly using random sampling.

We demonstrate these tools using datasets such as 3D botanical trees and neuronal structures.

The remainder of this paper is organized as follows; After reviewing the related work in Section 2, we introduce a novel tree-shape space (Section 3) and metric on this space (Section 4) to quantify bending, stretching, and topological changes in tree-shaped 3D objects. We then develop algorithms for computing one-to-one correspondences and geodesics, i.e., optimal deformation paths with respect to the chosen metric, between tree-shaped 3D objects (Section 5). Using these building blocks, we develop computational tools for computing statistical summaries of collections of tree-shaped 3D objects (Section 6) for synthesizing novel tree-shaped 3D models (Section 7). Finally, we present results in Section 8 and conclude in Section 9.

2 RELATED WORK

Statistical shape analysis has been studied extensively by the computer vision, computer graphics, and statistics communities. There are two subproblems, which are essential for statistical shape analysis: registration and optimal deformation. Registration is the problem of finding a one-to-one matching of points across objects, i.e., deciding which point on one object matches which point on the other. Optimal deformation, or geodesics, is the problem of finding an optimal continuous sequence of shapes starting from one shape and ending at the other. The optimality is measured with respect to a physically motivated metric.

There is a large body of literature that investigates these problems. Many papers are restricted to 3D objects that can be studied by bending and stretching. The idea is to treat shapes as points in a shape space equipped with a proper metric that measures the amount of bending and stretching that the shapes undergo. Equipping the space with a proper metric allows comparing objects based on their shapes, computing geodesics, and performing statistical analysis, including regressions and shape synthesis. The well-known work from Kendall’s school [19], [20] represents shapes with point sets that are already registered and focused only on deformations. Approaches such as medial surfaces [21], [22] and level sets [23] either presume registration, or solve for it using some independent pre-processing criterion such as MDL [24]. Kilian et al. [25] represent surfaces by discrete triangulated meshes and compute geodesic deformation paths between them while assuming that the meshes are registered. Heeren et al. [26] propose a method for computing geodesic-based deformations of thin shell shapes, with extensions for computing summary statistics in the shell space [27], but with known registration.

Some other papers solve for registration using shape descriptors while ignoring deformation; see [28], [29] for a detailed survey on the topic. These methods assume uniform sampling on the shape surfaces to address the registration. This assumption is a significant restriction, as it limits the registration of corresponding features across surfaces. These methods are not suitable for 3D objects that undergo complex geometric (bending and stretching) and topological deformations and for objects with complex self-similarities and symmetries.

Recent works, on the other hand, tried to address the registration and optimal deformation problems jointly. Central to this problem is the definition of a Riemannian metric (on the space of parameterized objects) preserved by the action of the relevant transformations, typically reparameterizations, translations, rotations, and perhaps scale. In particular, Srivastava et al. [30] introduced a particular elastic metric in conjunction with a representation called the square-root velocity function (SRVF). It has been used for joint registration and geodesic computation between planar shapes and between curves in $\mathbb{R}^d$, $d \geq 3$ [9]. It has been later extended to the analysis of parameterized surfaces [8], [10], [31], [32]. These methods, however, are limited to 3D models with fixed topology. They cannot capture and model topological variabilities such as those present in plant roots, botanical trees, and neuronal structures.

Closer to this paper are techniques based on tree stas-
tics. The seminal work of Billera et al. [33] proposed the notion of continuous tree-space and its associated tools for computing summary statistics. Some variants of this idea were developed for statistical analysis of tree-structured data, e.g., [34], [35]. However, these works only consider the topological structure of trees and ignore the geometric attributes of edges, limiting their usage. As a result, several papers have defined a more general tree-space. Examples include Feragen et al.'s framework [12], [13], [14], [15], which proposed a tree-shape space for computing statistics of airway trees, and its extension to complex botanical trees [16], [17]. Despite their efficiency and accuracy in certain situations, these techniques exhibit three main fundamental limitations. First, they use the Quotient Euclidean Distance (QED), which is not suitable for capturing large elastic deformations, i.e., bending and stretching, of the branches. Second, they represent tree shapes as a father-child branching structure, which leads to significant shrinkage along the geodesics between trees that exhibit large topological differences. Third, branch-wise correspondences need to be manually specified, especially when dealing with complex tree-like structures. In contrast, we propose in this paper a statistical framework that is more suitable for analyzing complex tree-shaped 3D objects such as botanical trees, plant roots, and neuronal structures. It builds upon and extends the recent work of Adam et al. [18], which showed that main-side branching representation is more efficient for capturing topological changes. To the best of our knowledge, this is the first approach that deals with the statistical modeling of such complex tree-shaped 3D objects. In addition to classical applications such the classification of tree-shaped 3D objects based on their shape and the computation of correspondences and geodesics, we demonstrate through experiments that this framework can produce reasonable statistical summaries. It also enables the synthesis of tree-shaped 3D objects, either randomly or in a controlled manner, and thus can be used to populate virtual environments and to generate simulated data for training deep neural networks.

3 Representation

The input to our framework is a collection of tree-like 3D objects such as plant roots, botanical trees, or neuronal structures. We first skeletonize each object and convert it into a set of curves, one for each branch. We then augment the skeletal curves with additional attributes such as the thickness at each skeletal point of the branch it represents. Each branch $\beta$ can then be seen as a continuous curve in $\mathbb{R}^3 \times \mathbb{R}^+$ of the form:

$$\beta : [0, 1] \rightarrow \mathbb{R}^3 \times \mathbb{R}^+,$$

$$\beta(s) \equiv (f(s), r(s)) = (x(s), y(s), z(s), r(s)).$$

(1)

Here, $f(s) = (x(s), y(s), z(s))$ is the 3D coordinates of the skeletal curve at $s$, and $r(s)$ is its thickness. The parameter $s \in [0, 1]$ is expressed as the proportion of arc-length along the skeletal curves of $\beta$.

We organize the branches into layers, with layer zero representing the main branch. With this setup, a tree-like object can be represented recursively by writing $\beta = (\beta^0, \{\beta^i, s^i\}_{i=1}^n)$, where:

- $\beta = \beta^0$ is the main branch, e.g., the trunk in the case of a botanical tree.
- $\{\beta^i\}_{i=1}^n$ is a subtree attached to $\beta^0$ at the bifurcation point $\beta^0(s^i)$ with $s^i \in [0, 1]$.

This is illustrated in Figure 1. If a subtree $\beta^i$ further contains sub subtrees, then that can be represented recursively by repeating this idea at every level.

To make the representation translation and scale invariant, we first translate each tree so that the start point of its main branch is located at the origin, and then scale the entire tree so that the length of its main branch is one. Note that invariance to scale is optional as it may not be required for some applications such as growth analysis. In what follows, we assume that all the objects have been normalized for translation and scale, and thus they are elements of a pre-tree-shape space $C_{\beta}$, which can be recursively defined as follows.

- Let $C_{\beta} = C_{\beta}^0$ be the pre-tree shape space of all trees that only have the main branch, a single curve. We will call such structures a level-0 tree. It can be shown that

$$C_{\beta}^0 = A_f \times \mathbb{R}^+,$$

(2)

where $A_f$ is the pre-shape space of all parameterized curves in $\mathbb{R}^3$. Note that $C_{\beta}^0$ is in fact the space of all curves defined by Eqn. (1).

- Let now $C_{\beta}^k, k \geq 1$ be the pre-tree shape space of all trees that can have up to $k$ levels. This space can be recursively defined as follows;

$$C_{\beta}^k = \prod_{n=0}^{\infty} \left( C_{\beta}^{k-1} \times [0, 1] \right)^n.$$

(3)

This mean that a level-$k$ tree can have $n$ structures attached the main branch, each structure being a level-$(k - 1)$ tree, at specified attachment points.

4 Metric for Tree-Shapes

In order to perform statistical shape analysis and modeling of the geometry and topology of a collection of tree shapes, we need to define a distance metric in the shape space of trees, and a mechanism for computing correspondences and geodesics between elements in this space.
4.1 The elastic metric for the shape of branches

We propose to use an elastic metric that captures the bending and stretching of the branches. Consider a path \( \alpha : [0,1] \rightarrow \mathcal{C}_\beta \), which deforms \( \beta_1 \) onto \( \beta_2 \). In other words, \( \alpha(0) = \beta_1, \alpha(1) = \beta_2 \), and \( \forall t \in [0,1], \alpha(t) \in \mathcal{C}_\beta \). Its length \( L[\alpha] \) is given by:

\[
L[\alpha] = \int_0^1 \langle \langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle \rangle \, dt ,
\]

where \( \dot{\alpha} = \frac{\partial \alpha}{\partial t} \) is the velocity vector, and \( \langle \langle \cdot, \cdot \rangle \rangle \) is a certain metric. (Note that here we dropped the indices of the path \( \alpha \) for clarity.) Instead of using the \( L^2 \) metric as in [14], [16], we use a metric that quantifies the bending and stretching of the branches.

Physically, bending can be quantified by measuring changes in the orientation of the tangent vectors to the skeletal curve \( f \) of \( \beta \) along the deformation path. Stretching can be decomposed into two components; the first one is related to the elongation of the skeletal curve \( f \), which can be quantified by looking at how the magnitude of the tangent vector to \( f \) at every point \( s \) changes along the deformation path. The second component is related to changes in the thickness of the branch and can be measured by looking at how the thickness \( r \) at each point \( s \) varies along the deformation path. Let us write:

\[
f'(s) = \frac{\partial f(s)}{\partial s} = \theta(s)e^\phi(s) ,
\]

where \( \theta(s) \) is the unit tangent vector to \( f \) at \( s \), and \( e^\phi(s) \) is the magnitude of the tangent vector to \( f \) at \( s \). (We also refer to \( e^\phi(s) \) as the speed.) In other words:

\[
e^\phi(s) = \| f'(s) \| = \left\| \frac{\partial f(s)}{\partial s} \right\| , \text{ and } \theta(s) = \frac{f'(s)}{\| f'(s) \|} .
\]

One can then define an elastic metric as the weighted sum of the changes of \( \phi, \theta \), and \( r \) along the path \( \alpha \). In other words:

\[
\langle \langle \dot{\alpha}, \dot{\beta} \rangle \rangle = a \int \langle \dot{\theta}(s), \dot{\theta}(s) \rangle e^\phi(s) \, ds + b \int \dot{\phi}(s)^2 e^\phi(s) \, ds + c \int \dot{r}(s)^2 e^\phi(s) \, ds . (7)
\]

Here, \( \dot{\cdot} \) denotes the derivative of the function \( \cdot \) with respect to \( t \). The first term of Eqn. (7) quantifies bending by measuring changes in the orientation of the tangent vector to the skeletal curve. The second and third terms quantify stretching. The first two terms are equivalent to the elastic metric between curves in \( \mathbb{R}^3 \), which has been introduced by Srivastava et al. [30]. While the full elastic metric of Eqn. (7) is the most appropriate for analyzing the shape of branches, it is computationally very expensive. To reduce the computation time, Srivastava et al. [30] introduced the Square Root Velocity Function (SRVF) of a curve \( f \) as a mapping \( \mathcal{Q} \) of the form:

\[
\mathcal{Q}(f)(s) = \begin{cases} 
\frac{f'(s)}{\sqrt{\| f'(s) \|}} & \text{if } f'(s) \text{ exists and } \| f'(s) \| \neq 0 , \\
0 & \text{otherwise} .
\end{cases} \tag{8}
\]

More importantly, Srivastava et al. [30] showed that the \( L^2 \) metric in the space of SRVFs is equivalent to the elastic metric of Eqn. (7) when setting \( a = 1, b = \frac{1}{2} \), and \( c = 0 \). In our case, we define the Extended SRVF (ESRVF) of a branch \( \beta = (f, r) \) as the pair \( q = (\text{SRVF}(f), r) \). We can easily show that the \( L^2 \) metric in the space of ESRVFs, hereinafter denoted by \( \Omega \), is equivalent to the elastic metric of Eqn. (7) when setting \( a = c = 1, b = \frac{1}{2} \). This is very important since, instead of working with the complex elastic metric, we can represent the shape of each branch using its ESRVF and use the associated \( L^2 \) metric to capture the bending and stretching of branches.

The ESRVF representation inherits all the properties of the SRVF, including the fact that the action of the reparameterization group on the ESRVF is by isometry. In other words, \( \forall \gamma \in \Gamma, \| q_1 - q_2 \| = \| q_1 - \gamma \circ q_2 \| \). (Here \( \Gamma \) is the space of all diffeomorphisms of the domain \([0,1] \) to itself.) With this, the aligning two branches \( \beta_1 \) and \( \beta_2 \) becomes the problem of finding the optimal rotation \( O^* \in SO(3) \) and reparameterization \( \gamma^* \in \Gamma \) such that:

\[
(O^*, \gamma^*) = \arg \min_{O \in SO(3), \gamma \in \Gamma} \| q_1 - O(q_2, \gamma) \| . \tag{9}
\]

We can easily show that the \( L^2 \) metric in the space of SRVFs, hereinafter denoted by \( \Omega \), is equivalent to the elastic metric of Eqn. (7) when setting \( a = c = 1, b = \frac{1}{2} \). This is very important since, instead of working with the complex elastic metric, we can represent the shape of each branch using its ESRVF and use the associated \( L^2 \) metric to capture the bending and stretching of branches.

The ESRVF representation inherits all the properties of the SRVF, including the fact that the action of the reparameterization group on the ESRVF is by isometry. In other words, \( \forall \gamma \in \Gamma, \| q_1 - q_2 \| = \| q_1 - \gamma \circ q_2 \| \). (Here \( \Gamma \) is the space of all diffeomorphisms of the domain \([0,1] \) to itself.) With this, the aligning two branches \( \beta_1 \) and \( \beta_2 \) becomes the problem of finding the optimal rotation \( O^* \in SO(3) \) and reparameterization \( \gamma^* \in \Gamma \) such that:

\[
(O^*, \gamma^*) = \arg \min_{O \in SO(3), \gamma \in \Gamma} \| q_1 - O(q_2, \gamma) \| . \tag{9}
\]

We can easily show that the \( L^2 \) metric in the space of SRVFs, hereinafter denoted by \( \Omega \), is equivalent to the elastic metric of Eqn. (7) when setting \( a = c = 1, b = \frac{1}{2} \). This is very important since, instead of working with the complex elastic metric, we can represent the shape of each branch using its ESRVF and use the associated \( L^2 \) metric to capture the bending and stretching of branches.

The next step is to define a metric in the space of SRVFTs and a mechanism for computing correspondences and geodesics between points in this space. We do this by generalizing the metric proposed by Duncan et al. [18] for the analysis of simple trees, composed of the main branch and a set of branches. The ESRVF representation inherits all the properties of the SRVF, including the fact that the action of the reparameterization group on the ESRVF is by isometry. In other words, \( \forall \gamma \in \Gamma, \| q_1 - q_2 \| = \| q_1 - \gamma \circ q_2 \| \). (Here \( \Gamma \) is the space of all diffeomorphisms of the domain \([0,1] \) to itself.) With this, the aligning two branches \( \beta_1 \) and \( \beta_2 \) becomes the problem of finding the optimal rotation \( O^* \in SO(3) \) and reparameterization \( \gamma^* \in \Gamma \) such that:

\[
(O^*, \gamma^*) = \arg \min_{O \in SO(3), \gamma \in \Gamma} \| q_1 - O(q_2, \gamma) \| . \tag{9}
\]
side branches, to complex 3D trees of arbitrary levels of hierarchy. Let $\beta_1$ and $\beta_2$ be two tree shapes represented with their SRVFTs $q_1$ and $q_2$ in $C_q$. Let us first assume that the two trees have the same number of branches and are in branch-wise correspondence. We define the distance between two tree shapes $\beta_1$ and $\beta_2$ recursively as follows:

$$d_{C_q}(q_1, q_2) = \lambda_m \| q_1^0 - q_2^0 \|^2 + \lambda_s \sum_{i=1}^n d_{C_q}(q_1^i, q_2^i) + \lambda_p \sum_{i=1}^n (s_1^i - s_2^i)^2.$$  \hspace{1cm} (12)

The first term of Eqn. (12) measures the amount of bending and stretching needed to align one main branch to another. The second term, which is computed recursively using Eqn. (12), is the dissimilarity between two subtrees $q_1^i$ and $q_2^i$ attached, respectively, to the bifurcation points $s_1^i$ and $s_2^i$. Note that when $q_1$ and $q_2$ are null trees, then we set $d_c(q_1, q_2) = 0$. The third term is the distance between the locations of the corresponding main branches of the subtrees $q_1^0$ and $q_2^0$. The parameters $\lambda = (\lambda_m, \lambda_s, \lambda_p)$ control the relative cost of deforming the main branch, deforming the subtrees connected to the main branch, and moving the positions of the subtrees connected to the main branch. The latter enables topological changes.

### 4.3 Invariant metric

A proper metric for comparing the shape of tree-like objects should be invariant to shape-preserving transformations. The translation and scale have already been factored out in a pre-processing step. Now, we consider rotations, reparameterization of the branches, and permutations of the orders of the lateral subtrees attached to a branch. Let:

- $O \in SO(3)$ be a global rotation applied to the entire tree,
- $\gamma = (\gamma^0, \{\gamma^i\}_{i=1}^n)$ be the reparameterization of the main branch and its subtrees. Specifically, $\gamma^0 \in \Gamma$ is a diffeomorphism that applies to the main branch of $q$ and $\gamma^i$ the reparameterization, defined recursively, of the $i$-th subtree,
- $\sigma = (\sigma^0, \{\sigma^i\}_{i=1}^n)$ such that $\sigma^0 \in \Gamma$ is the permutation of the orders of the lateral subtrees on their corresponding main branch $q^0_k$ (here, $\sigma$ is the space of such permutations), and $\sigma^i$ defines recursively these permutations for the $i$-th subtree.

Let $(q, O, \gamma, \sigma)$ denote the result after applying these three transformations to $q$. Since these transformations are shape-preserving, $q$ and $(q, O, \gamma, \sigma)$ have the same shape. They thus are equivalent under the action of global rotations, branch reparameterizations, and permutations of the indices of the lateral subtrees of the different branches. Thus, we define the rotation, reparameterization, and index permutation-invariant distance between $q^1$ and $q^2$ as the infimum over all possible rotations, branch reparameterizations, and branch index permutations, i.e.,:

$$d(q_1, q_2) = \inf_{O \in SO(3), \gamma \in \Gamma, \sigma \in S} d_{C_q}(q_1, q_2, O, \gamma, \sigma).$$  \hspace{1cm} (13)

The optimal registration of $q_2$ onto $q_1$ can then be found by solving the following optimization problem:

$$(\hat{O}, \hat{\gamma}, \hat{\sigma}) = \arg\min_{O, \gamma, \sigma} d_{C_q}(q_1, q_2, O, \gamma, \sigma).$$  \hspace{1cm} (15)

The challenge now is how to optimize simultaneously over all these transformation spaces. This will be discussed in Section 5.

### 4.4 Trees with different numbers of side branches

The formulation presented in Sections 4.2 and 4.3 assumes that the trees have the same number of branches and the same topology. Thus a one-to-one branch-wise correspondence exists. However, this is not the case in practice. Two tree models rarely have the same number of bifurcation points and branches. As a result, these trees lie on disjoint subspaces, making it difficult to compare them and compute geodesics. One way to overcome this issue is by adding null branches, i.e., branches of length zero.

First, we define the order of a tree $\beta$, and subsequently of $q$, by taking the maximum of the number of bifurcation points on its branches. Let $n_i, i = 1, 2$ be the order of the trees $q_1$ and $q_2$, and let $n = \max(n_1, n_2)$. Then, we add to each branch $l$ of the $i$-th tree, $n - n_i$ null branches. (Here, $n_i$ is the order of branch $l$.) By doing so, all $k$-level trees become elements of the same pre-tree shape space. This facilitates their comparison, and subsequently putting them in correspondence and computing geodesic paths.

### 5 Correspondences and geodesics

Since the metric in the pre-tree shape space $C_q$ is a weighted sum of $L^2$ distances, the combined space is flat. This fact has significant practical benefits. In fact, instead of working in the original space of 3D objects, which is nonlinear and equipped with a complex metric, one can map the input tree-like objects into the SRVFT space, use the simple metric of Eqn. (13) to compute correspondences, geodesics, and statistics using standard tools from vector calculus, and finally map the results back to the original space of trees for visualization. Since the SRVFT mapping is one-to-one and onto (up to global translation), the inverse mapping exists, is unique, and more importantly, it has a closed analytical form. This invertibility is very important in practice.

Given two tree-shaped objects $\beta_1$ and $\beta_2$ represented with their SRVFTs $q_1$ and $q_2$, we obtain one-to-one correspondences by solve the optimisation problem of Eqn. (15). This is done by alternating the optimisation over $SO(3), \Gamma$, and $S$, i.e.,

- Optimize over $O \in SO(3)$, assuming fixed parameterizations and permutations, which is straightforward to achieve using Procrustes analysis, then

$$d_{C_q}(q_1, q_2, O, \gamma, \sigma) = \lambda_m \| q_1^0 - O(q_2^0, \gamma_0) \|^2 + \lambda_s \sum_{i=1}^n d(q_1^i, q_2^i) + \lambda_p \sum_{i=1}^n (s_1^i - s_2^i)^2.$$  \hspace{1cm} (14)

The optimal registration of $q_2$ onto $q_1$ can then be found by solving the following optimization problem:

$$(\hat{O}, \hat{\gamma}, \hat{\sigma}) = \arg\min_{O, \gamma, \sigma} d_{C_q}(q_1, q_2, O, \gamma, \sigma).$$  \hspace{1cm} (15)

The challenge now is how to optimize simultaneously over all these transformation spaces. This will be discussed in Section 5.
• Optimize over $\gamma \in \Gamma$ and $\sigma \in S$ while assuming a fixed rotation.

In this section, we discuss the process of optimizing over $\gamma \in \Gamma$ and $\sigma \in S$. We first consider the simple case of trees with two levels of hierarchy, i.e., trees that only have one main branch and multiple side (or lateral) branches (Section 5.1). We then show how this method can be generalized to complex tree-shaped 3D objects of arbitrary branching structures (Section 5.2). Algorithms 1 and 2 summarize the overall procedures.

5.1 Correspondence between simple tree shapes

Assume the simple case where the two trees are composed of only one main branch and some lateral branches. In this case $\gamma = (\gamma^0, (\gamma^1_i)_{i=1}^n)$ and $\sigma = (\sigma_0)$. Since the first term of Eqn. (14), which measures the distance between the two main branches, is disentangled from the remaining two terms, the case studied in Duncan et al. [18] results. We first find, using the approach of Srivastava et al. [30], the optimal re-parameterization $\gamma^0$ that elastically aligns the main branch of $q_2$ onto the main branch of $q_1$.

Next, to find which side branch of $q_2$ is matched to which branch of $q_1$ (i.e., using the third term of Eqn. (14)) and also elastically register the corresponding branches (i.e., using the second term of Eqn. (14)), we formulate it as a linear assignment problem. That is, we build a pairwise distance matrix $E$ of size $n_1 \times n_2$ where

$$E_{ij} = \lambda_s \inf_{\gamma, \sigma} ||q^1_i - O(q^2_i, \gamma)||^2 + \lambda_p (s^1_i - s^2_i)^2. \quad (16)$$

We then apply the Kuhn-Munkres algorithm, also known as the Hungarian Algorithm [36], to compute the optimal matching $\sigma_0$. Next, for each matching branches $q^1_i$ and $q^2_{\sigma_0(i)}$, we compute the reparameterization $\gamma_i$ as

$$(O_i, \gamma_i) = \arg\min_{O, \gamma} ||q^1_i - O(q^2_i, \gamma)||^2 \quad (17)$$

by using the approach of Srivastava et al. [30]. Note that, in contrast to Duncan et al. [18], which builds an $(n_1 + n_2) \times (n_1 + n_2)$ cost matrix, we solve the linear assignment problem by building a $\max(n_1, n_2) \times \max(n_1, n_2)$ cost matrix. In this way, we reduce the worst-case time complexity from $O((n_1 + n_2)^3)$ to $O(\max(n_1, n_2)^3)$ and reduce the computation time by a factor of 8. This is important especially when processing tree shapes with many layers recursively.

5.2 Extension to complex tree shapes

To handle complex trees, one can recursively run the procedure described in Section 5.1 over the entire hierarchy of the trees. By doing so, however, the computation time would increase exponentially. We approximate this procedure by considering three levels at a time. That is, we first consider the main branch and the subtrees attached to it. However, for each subtree, we only consider its main branch and the side branches attached to it. To align two trees, we first align their main branches, compute the assignment matrix where each entry $E_{ij}$ is the optimal distance between the subtree $q^1_i$ and $q^2_j$. Since these two subtrees have only two levels, $E_{ij}$ can be computed using the procedure described in Section 5.1. Similar to Section 5.1, we apply the Hungarian algorithm to find the assignment $\sigma_0$ which matches subtrees on $q_1$ onto subtrees on $q_0$. By repeating this procedure, recursively, on $q^1_i$ and $q^2_{\sigma_0(i)}$, we obtain a full matching between the complex trees. Algorithms 1 and 2 summarize the overall procedure.

5.3 Computing geodesics

Let $\tilde{q}_2 = (q_2, \tilde{O}, \tilde{\gamma}, \tilde{\sigma})$. Since the metric is a weighted Euclidean norm of $L^2$ distances and Euclidean distances, the combined space is flat. Thus, the geodesic $\tilde{\alpha}$ between $q_1$ and $q_2$ is just the straight line that connects $q_1$ to $q_2$, i.e.,

$$\tilde{\alpha}(t) = (1 - t)q_1 + tq_2, t \in [0, 1]. \quad (18)$$

Finally, for visualization, we map $\alpha(t), t \in [0, 1]$, back to the space of tree shapes $C_q$ using inverse SRVF mapping, which has a closed analytical form, see [30].

6 Statistics on the tree-shape space

The ability to compute correspondences and geodesics between 3D tree-shaped objects enables a wide range of shape analysis tasks. In this section, we show how these fundamental tools can be used to compute means and modes of variability of a collection of 3D botanical trees. Let $\{\beta_i, i = 1, 2, \ldots, m\}$ be a set of tree-shaped 3D objects, e.g., botanical trees, plant roots, or neuron structures. Let $\{q_i, i = 1, 2, \ldots, m\}$ be their corresponding SRVF representations. To compute the mean tree $\mu$, we first compute the mean SRVF representation $\mu_q$ and then map it back to the space of tree shapes.

Mathematically, the mean $\mu_q$ can be regarded as the point in $C_q$ that is as close as possible to all tree shapes in $\{q_i, i = 1, 2, \ldots, m\}$. The closeness is defined with respect to the metric of Eqn. (14). In other words,

$$\mu_q = \arg\min_{O_{i}, \gamma_{i} \in SO(3), \sigma_{i} \in S} \sum_{i=1}^{m} d_{q}^{2}(q,O(q_{i}, \gamma_{i}, \sigma_{i})).$$

(19)
Algorithm 2 Reparameterization and permutation.

Input:
- \( n_{\text{levels}} = 2, 3 \) or 4.
- \( q^1, q^2 \), the SRVFT representations of two tree-shaped objects to be aligned.
- \( \lambda = (\lambda_m, \lambda_s, \lambda_p) \), the distance weight parameters.
- \( n_1, n_2 \), the number of side branches in \( q^1 \), respectively, \( q^2 \).

Output:
- \( \gamma \) and \( \sigma \).

1: procedure REPARAMPERMUTE \((O, q^1, q^2, \lambda, n_{\text{levels}})\)
2: \( N = \max(n_1, n_2) \).
3: for \( i = 1 : N, j = 1 : N \) do
4: if \( n_{\text{levels}} = 3 \) or 4 then
5: \( (O, \gamma_{ij}, \sigma_{ij}) = \text{AlignTrees}(q^1_i, q^2_j, n_{\text{levels}} - 1) \).
6: else
7: \( \gamma_{ij} = \text{DynamicProgQ}(q^1_i, Oq^2_j) \).
8: \( E_{ij} = \lambda_m \|q^1_i - O(q^2_j)\|^2 + \lambda_p (s^1_i - s^2_j)^2 \).
9: end if
10: end for
11: \( (\sigma_0, E_{\text{sides}}) = \text{LAPJV}(E) \).
12: \( \gamma_0 = \text{DynamicProgQ}(q^1_0, q^2_0) \).
13: for \( i = 1 : N, j = 1 : N \) do
14: if \( n_{\text{levels}} = 3 \) or 4 then
15: \( \gamma_{ij} = \text{AlignTrees}(q^1_i, q^2_{\gamma_0(i)}, n_{\text{levels}}) \).
16: \( \gamma = \gamma \cup \gamma_i \).
17: \( \sigma = \sigma \cup \{\sigma_0\} \).
18: else
19: \( \gamma = \gamma \cup \gamma_i \).
20: \( \sigma = \sigma \cup \{\sigma_0\} \).
21: end if
22: end for
23: Return \( \gamma, \sigma \).
24: end procedure

Solving Eqn. (19) involved finding \( \mu_q \), also known as the Karcher mean, while simultaneously registering every \( q_i \) on \( \mu_q \). This can be efficiently done via a gradient descent approach. That is:

1) Set \( \mu_q = q_1 \).
2) For \( i = 1 : m \):
   - Optimally register \( q_i \) onto \( \mu_q \), using Eqn. (15).
   - Let \( (\bar{O}_i, \tilde{\gamma}_i, \tilde{\sigma}_i) \) be the solution to Eqn. (15).
3) Set \( \mu_q = \frac{1}{m} \sum_{i=1}^{m} \bar{O}_i(q_i, \tilde{\gamma}_i, \tilde{\sigma}_i) \).
4) Repeat steps 2 and 3 until convergence.
5) Return \( \mu_q \) and \((\bar{O}_i, \tilde{\gamma}_i, \tilde{\sigma}_i)_{i=1}^{m} \).

Finally, the mean tree \( \mu \) is obtained by mapping \( \mu_q \) back to the space \( C_B \). This mean shape characterizes the primary morphological properties of shapes in the collection \( \{\beta_i, i = 1, 2, \ldots, m\} \). In what follows, let \( q_i = \bar{O}_i(q_i, \tilde{\gamma}_i, \tilde{\sigma}_i) \) be the SRVFT representation of \( \beta_i \) but optimally registered onto the mean \( \mu_q \). Let \( v_i = q_i - \mu \). The covariance matrix \( \Lambda_i \) define the principal directions of variations while the corresponding eigenvalue \( \lambda_i \) defines the variance along the \( i \)-th direction.

With this setup, each SRVFT sample \( q \) can be modeled as a linear combination of the \( k \) leading eigenvectors:

\[
q = \mu_q + \sum_{i=1}^{k} a_i \sqrt{\lambda_i} \Lambda_i v_i, b_i \in \mathbb{R},
\]

where \( a_i \)'s are standard normal random variables. Note that with this representation, we are fitting a Gaussian distribution to the input data \( \{q_i, i = 1, 2, \ldots, m\} \). In general, however, especially since the SRVFT space is flat, we can fit any arbitrary distribution from the parametric or non-parametric families.

7 Tree-shape synthesis

7.1 Shape synthesis by random sampling

We characterize the shape variability of an input population of tree shapes by fitting to the shape population a multivariate Gaussian of mean \( \mu_q \) and a diagonal covariance matrix \( C \) whose diagonal elements are the eigenvalues \( \lambda_i \). A random 3D tree-shape can then be generated by randomly sampling from this multivariate Gaussian. That is, we first randomly sample \( k \) real numbers \( a_1, \ldots, a_k \sim \mathcal{N}(0, 1) \).

The SRVFT \( q \) of a random tree-shape \( \beta \) is then given by Eqn. (20), and the random tree-shape \( \beta \) is obtained by mapping \( q \) back to the space of 3D tree-shapes. In this paper, we only consider the \( k \)-leading eigenvalues such that \( k \leq m \) and \( \sum_{i=1}^{k} \lambda_i > 0.99 \). To ensure that the synthesized treeshapes are plausible, one can restrict \( a_i \) to be within a certain range, e.g., \([-1, 1]\).

8 Results and discussion

In this section, we demonstrate the performance of the proposed framework in finding correspondences and computing geodesics between pairs of tree-like 3D objects (Section 8.1), symmetry analysis and symmetrization of such shapes (Section 8.3), computing summary statistics such as means and modes of variation (Section 8.4), and finally synthesizing novel tree-shaped 3D objects either randomly or in a controlled manner through regression in the three-shape space (Section 8.5). We use two types of datasets: 3D botanical trees and 3D neuronal structures. In both cases, the models have complex branching structures composed of multiple layers, unlike Duncan et al. [18], which is limited to trees of level one. The models used in this section also exhibit large geometric and topological variability, making it challenging to find one-to-one correspondences between such tree-like 3D objects. The Supplementary Material includes more results as well as videos of all the examples shown in this paper.

8.1 Correspondence and geodesics

Figs. 2 and 3 show examples of geodesic paths between 3D botanical trees and 3D neuronal structures, respectively. For each example, we show the geodesic between the leftmost (source) and the rightmost (target) trees. The botanical trees
(a) Geodesic length $d_s = 31.7$. 

(b) Geodesic length $d_s = 131.4$. 

(c) Geodesic length $d_s = 121.8$. 

Fig. 2: Geodesic deformations between the most left and the most right 3D botanical trees in each row. In this experiment, we use $\lambda_m = \lambda_s = \lambda_p = 1.0$. The Supplementary Material includes more examples.

Fig. 5 compares the results of our approach with the approaches of Feragen et al. [14] and Wang et al. [16]. Conceptually, the approach of Feragen et al. [14], which automatically finds correspondences and geodesics, does not consider the thickness of the branches that compose the tree shapes. It is also computationally costly, especially for complex trees. Wang et al. [16] extended the approach by pre-computing the correspondences in an ad-hoc manner. In both methods, the nodes representing bifurcation points, and edges representing the branch segments between the bifurcation points. Topological changes are modeled using edge collapse and node split operations. As such, the intermediate shapes along a geodesic exhibit significant shrinkage; see the bottom two rows in Fig. 5. The approach proposed here solves this problem by modeling topological changes as sliding branches. It also uses a full elastic metric, which explicitly quantifies the bending and stretching of the branches. Subsequently, the geodesics look more natural compared to the state-of-the-art; see the top row in Fig. 5.

**Computation time.** Our framework is implemented using Matlab and runs entirely on CPU, configured with 2.4GHz Intel Core i5, 8GB RAM. The most time-consuming part is the shape matching process, which takes 60.1 to 4923.1 seconds for botanical trees and 22.4 to 30.2 seconds for neuron trees. The following geodesic process takes relatively less time (0.4 to 1.9 seconds for botanical trees and 0.2 seconds for neuron trees).

### 8.2 Ablation study

**Effect of the weights $\lambda_m$, $\lambda_s$, and $\lambda_p$.** These three parameters control the importance of each of the three terms in Eqn. (14). In the experiments of Fig. 2, these parameters were manually set to $\lambda_m = \lambda_s = \lambda_p = 1.0$ while in Fig. 3, they were set to $\lambda_m = 0.2, \lambda_s = 1.0, \lambda_p = 0.2$. Fig. 6 shows how the correspondences and geodesics vary while varying these parameters. For the first example in Fig. 6 with $\lambda_m = 1.0, \lambda_s = 10^{-4}, \lambda_p = 1.0$, the approach favours creating new virtual subtrees (see the subtree indicated by an arrow) rather than sliding the existing one, which happens in the second example with $\lambda_m = 1.0, \lambda_s = 1.0, \lambda_p = 10^{-4}$. This is reasonable since in the first example, the last term in Eqn. 12, which measures distance between bifurcation
(a) Geodesic length $d_s = 156.7$.

(b) Geodesic length $d_s = 181.8$.

(c) Geodesic length $d_s = 165.2$.

Fig. 3: Geodesic deformations between the most left and the most right 3D neuronal trees in each row. In this experiment, we use $\lambda_m = 0.2, \lambda_s = 1.0, \lambda_p = 0.2$.

8.3 Reflection symmetry analysis and symmetrization

Symmetry is an essential feature of artificial and biological objects and can be helpful in many different applications. Existing symmetry analysis techniques are mainly based on feature detection and matching; see, for example, the survey by Mitra et al. [37]. The proposed framework provides a proper metric and a mechanism for computing geodesics. It can easily analyze the reflection symmetry and symmetrize tree-shaped 3D objects without extracting and computing shape descriptors.

To analyze the level of asymmetry of a given tree-shaped object $\beta$ using our framework, we first obtain its reflection across an arbitrary plane $\Delta$. Let $v \in \mathbb{R}$ be a normal vector to this plane. Then, the reflection of $\beta$ with respect to $\Delta$ is given by

$$ \tilde{\beta} = (I - 2 \frac{vv^T}{v^Tv}) \beta. \quad (21) $$

Here, $I$ is the $3 \times 3$ identity matrix and $v^T$ refers to the transpose of the vector $v$. Next, we use the approach described in this paper to compute a geodesic path, $\alpha$, between $\beta$ and $\tilde{\beta}$. $\alpha$ provides valuable information about the symmetry of $\beta$. First, its length gives a proper measure of the asymmetry of $\alpha$. Second, the halfway point along this geodesic, i.e. $\alpha(0.5)$, is symmetric. Moreover, amongst all symmetric trees, $\alpha(0.5)$ is the nearest to $\beta$. Thus, the process of computing a geodesic between a tree-shaped object and its reflection is equivalent to symmetrizing the object.

Fig. 9 shows several examples of highly complex botanical trees, and Fig. 10 shows one example of neuron structures symmetrized using the proposed approach. Observe that the midpoints of each geodesic are symmetric; thus, the geodesic path provides a natural symmetrization of the given tree-shaped 3D object.

8.4 Summary statistics

Fig. 11 show one example of mean shapes computed from three 3D neuron trees. More examples about mean shapes of 3D neuron trees and mean shape examples for 3D botanical trees are can be found in the Supplementary Material. From these results, we can see that mean shapes capture the overall shape information of the input samples.

Next, we perform a covariance analysis of a collection of 36 models of botanical trees. Fig. 19 shows the first three leading modes variation (one per row). The middle shape in each row (at zero standard deviation) is the mean shape of the input collection. We also perform a similar experiment of 51 neuron structures; see Fig. 21 in Supplementary Material. These two experiments clearly show that the leading modes of variation capture the main geometric and structural variation of the input 3D tree shape collections.

8.5 3D Tree-shape synthesis

Once a statistical model is fit to a collection of 3D tree-shaped objects, one can synthesize new 3D shape instances via random sampling from the statistical model. Fig. 13 show examples of randomly synthesized 3D tree-shaped objects. These tree shapes have been produced without any professional knowledges. We can clearly see they share some shape similarities with the input trees (see Fig. 19 in the Supplementary Material), but are not identical.
Fig. 5: Comparison of the quality of the geodesics obtained using our approach, the approach of Feragen et al. [14], and the approach of Guan et al. [16]. The Supplementary Material includes more comparison examples.

Fig. 6: The influence of weights $\lambda_m$, $\lambda_s$, and $\lambda_p$ of Eqn. 12 on the geodesics computed with our approach.

Fig. 7: Ablation study on neuronal tree geodesics. (a) Geodesic generated without graph matching, (b) geodesic generated with graph matching.

9 CONCLUSIONS

In this paper, we develop a comprehensive framework for quantifying biological 3D shapes that deform both in geometry and topology. These shapes are characterized by variable branching structures, sizes and shapes. This framework provides tools for quantifying shape differences, computing geodesic deformations, and modelling shape variability. A key idea here is to match common substructures across objects to result in more natural deformations and shape summaries. The resulting framework provides a powerful setting for quantifying complex structural variability in tree-like 3D shapes such as neuron structures and botanical tree models.

Although effective as evidenced by the results presented in this paper, there are several improvements that could be considered for future work. First, the geodesic quality depends on the weights of the three terms of the distance metric defined in Eqn. 12. In this work, we manually set those weights, while in practice they depend on specific knowledge lying in plant and medicine domains. Thus, one could consider learning these parameters directly from data labelled with domain knowledge. Second, currently, we have only considered synthesis by random sampling and not taking into account biological factors such as light...
intensity, ageing processes, disease, etc., which affect the 3D shape of tree-like structures such as botanical trees and neurons. One potential direction for future work is to incorporate these factors into the modelling process, e.g., via regression. Finally, the proposed framework can be used to enrich data collections, and thus can benefit data-driven applications such as deep learning based 3D reconstruction [38].

**References**

[1] X. Jiang, S. Shen, C. R. Cadwell, P. Berens, F. Sinz, A. S. Ecker, S. Patel, and A. S. Tolias, “Principles of connectivity among morphologically defined cell types in adult neocortex,” *Science*, vol. 350, no. 6264, p. aac9462, 2015.

[2] N. Spruston, “Pyramidal neurons: dendritic structure and synaptic integration,” *Nature Reviews Neuroscience*, vol. 9, no. 3, pp. 206–221, 2008.

[3] C. Lin, Y. Huang, T. Quan, and Y. Zhang, “Modelling brain-wide neuronal morphology via rooted cayley trees,” *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[4] D. Kabaso, P. J. Coskren, B. I. Henry, P. R. Hof, and S. L. Wearne, “The electrotonic structure of pyramidal neurons contributing to prefrontal cortical circuits in macaque monkeys is significantly altered in aging,” *Cerebral Cortex*, vol. 19, no. 10, pp. 2248–2268, 2009.

[5] V. Chan-Palay and E. Asan, “Alterations in catecholamine neurons of the locus coeruleus in senile dementia of the Alzheimer type and in parkinson’s disease with and without dementia and depression,” *Journal of comparative neurology*, vol. 287, no. 3, pp. 373–392, 1989.

[6] M. J. West, P. D. Coleman, D. G. Flood, and J. C. Troncoso, “Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer’s disease,” *The Lancet*, vol. 344, no. 8925, pp. 769–772, 1994.

[7] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3d faces,” in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[8] S. Kurttek, A. Srivastava, E. Klassen, and H. Laga, “Landmark-guided elastic shape analysis of spherically-parameterized surfaces,” in *Computer graphics forum*, vol. 32, no. 2, pp. 1–204. Wiley Online Library, 2013, pp. 429–438.

[9] H. Laga, S. Kurttek, A. Srivastava, and S. J. Miklavcic, “Landmark-free statistical analysis of the shape of plant leaves,” *Journal of theoretical biology*, vol. 363, pp. 41–52, 2014.

[10] H. Laga, Q. Xie, I. H. Jermyn, A. Srivastava et al., “Numerical inversion of srf maps for elastic shape analysis of genus-zero surfaces,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2451–2464, 2017.

[11] I. H. Jermyn, S. Kurttek, H. Laga, and A. Srivastava, “Elastic shape analysis of three-dimensional objects,” *Synthesis Lectures on Computer Vision*, vol. 12, no. 1, pp. 1–185, 2017.

[12] A. Feragen, F. Lauze, P. Lo, M. de Bruijne, and M. Nielsen, “Geometries on spaces of treelike shapes,” in *Asian Conference on Computer Vision*. Springer, 2010, pp. 160–173.

[13] A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze, “Means in spaces of tree-like shapes,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 736–746.

[14] A. Feragen, P. Lo, M. de Bruijne, M. Nielsen, and F. Lauze, “Towards a theory of statistical tree-shape analysis,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 2008–2021, 2013.

[15] A. Feragen, M. Owen, J. Petersen, M. M. Wille, L. H. Thomsen, A. Dirksen, and M. de Bruijne, “Tree-space statistics and approximations for large-scale analysis of anatomical trees,” in *International Conference on Information Processing in Medical Imaging*. Springer, 2013, pp. 74–85.

[16] G. Wang, H. Laqa, N. Xie, J. Jia, and H. Tabia, “The shape space of 3d botanical tree models,” *ACM Transactions on Graphics (TOG)*, vol. 37, no. 1, p. 7, 2018.

[17] G. Wang, H. Laga, J. Jia, N. Xie, and H. Tabia, “Statistical modeling of the 3d geometry and topology of botanical trees,” in *Computer Graphics Forum*, vol. 37, no. 5. Wiley Online Library, 2018, pp. 185–198.

[18] A. Duncan, E. Klassen, A. Srivastava et al., “Statistical shape analysis of simplified neuronal trees,” *The Annals of Applied Statistics*, vol. 12, no. 3, pp. 1385–1421, 2018.

[19] I. Dryden and K. Mardia, *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[20] B. Allen, B. Curless, B. Curless, and Z. Popović, “The space of human body shapes: reconstruction and parameterization from range scans,” in *ACM transactions on graphics (TOG)*, vol. 22, no. 3. ACM, 2003, pp. 587–594.

[21] S. Bouix, J. Pruissner, D. Collins, and K. Siddiqi, “Hippocampal shape analysis using medial surfaces,” in *Medical Image Computing and Computer-assisted Intervention—MICCAI 2009*. Springer, 2009, pp. 339–348.
Fig. 12: Mean tree (highlighted in the middle) and the first three principal modes of variation (one per row) for 36 botanical trees. The input botanical trees are shown in the supplementary material.

Fig. 13: Randomly synthesized botanical trees. The Supplementary Material includes more examples.

10 Supplementary Material

In this Supplementary Material, we provide additional results that could not fit into the main manuscript due to the page limit.

10.1 Geodesics and symmetry analysis

Fig. 14 shows another geodesic example between two botanical trees. Fig. 15 shows a geodesic comparison between our approach, the approach of Feragen et al. [14], and the approach of Guan et al. [16]. Fig. 16 shows two additional symmetry analysis and symmetrization of 3D botanical and neuron trees.

10.2 Summary statistics

Fig. 17 shows examples of mean shapes of three and five neuronal trees. Fig. 18, on the other hand, shows additional examples of mean shapes of three and five botanical trees.

10.2.1 Inputs, Modes and Random samples

Fig. 19 shows the input 3D botanical tree used to generate the random samples in Fig. 13 of the main manuscript.

Figs. 21 and 22 are, respectively, examples of principal modes of variation and randomly synthesized neuron trees, computed from the collection of 51 neuron trees shown in Fig. 20.

Finally, Fig. 24 shows the leading principal modes of variation computed from the 36 botanical tree models of Fig. 23. Fig. 25, on the other hand, shows 3D botanical trees randomly synthesized by sampling from the probability distribution fitted to the population of Fig. 23. We can clearly see that sampling from the probability distribution fitted to the input population allows the synthesis of rich 3D botanical tree models and neuron tree shapes.
Fig. 14: Geodesic deformations between the most left and the most right 3D botanical trees in each row. In this experiment, we use $\lambda_m = \lambda_s = \lambda_p = 1.0$.

Fig. 15: Comparison of the quality of the geodesics obtained using our approach, the approach of Feragen et al. [14], and the approach of Guan et al. [16].
Fig. 16: Analysis of the symmetry and symmetrization of (a) 3D botanical trees and (b) 3D neuron structures.
Fig. 17: Mean shape of 3D neuron trees.

(a) Mean shape of 3 neuron trees.

(b) Mean shape of 5 neuron trees.

Fig. 18: Mean shape of 3D botanical trees.

(a) Mean shape of 3 botanical trees.

(b) Mean shape of 3 botanical trees.

(c) Mean shape of 5 botanical trees.

(d) Mean shape of 5 botanical trees.

Fig. 18: Mean shape of 3D botanical trees.
Fig. 19: Input 3D tree shapes used to generate the random samples of Fig. 13 in the main manuscript.

Fig. 20: Input 3D neuron trees for Figs. 21 and 22.
Fig. 21: First three principal modes of variation (one per row) for 51 neuronal trees. The mean 3D neuron tree is highlighted in red.

Fig. 22: Randomly synthesized neuronal trees.
Fig. 23: Input botanical trees for Figs. 24 and 25.
-2 std  -1 std  0 std  1 std  2 std

Fig. 24: Principal modes of variation for 36 botanical trees. The mean 3D tree is highlighted in red.
Fig. 25: Randomly synthesized botanical trees