ICCV
#445

ICCV
#445

ICCV 2011 Submission #445. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Means in spaces of tree-like shapes

Anonymous ICCV submission

Paper ID 445

Abstract

The mean is often the most important statistic of a dataset as it provides a single point that summarizes the entire set. While the mean is readily defined and computed in Euclidean spaces, no commonly accepted solutions are currently available in more complicated spaces, such as spaces of tree-structured data. In this paper we study the notion of means, both generally in Gromov's $CAT(0)$-spaces (metric spaces of non-positive curvature), but also specifically in the space of tree-like shapes. We prove local existence and uniqueness of means in such spaces and discuss three different algorithms for computing means.

We make an experimental evaluation of the three algorithms through experiments on three different sets of data with tree-like structure: a synthetic dataset, a leaf morphology dataset from images, and a set of human airway subtrees from medical CT scans. This experimental study provides great insight into the behavior of the different methods and how they relate to each other. More importantly, it also provides mathematically well-founded, tractable and robust "average trees". This statistic is of utmost importance due to the ever-presence of tree-like structures in human anatomy, e.g.*, airways and vascularization systems.*

1. Notions of means

Centroids, weighted averages, midpoints of a pair of points, and other variations on the sample mean are the basic building blocks of statistical computations. While they are simple to compute when the underlying sample space is Euclidean, they may become much more complex in non-linear sample spaces. A classical definition of centroids in Euclidean space, dating back to Appolonios of Perga, has a direct extension to general metric spaces [10, 11]: a mean of the finite collection $(x_i)_i$ of points in a metric space (X, d) is a minimizer of the function

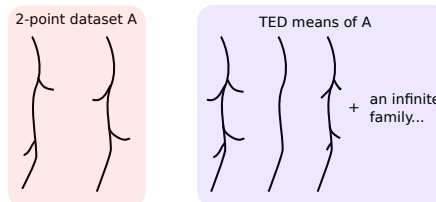$$\Phi(x) = \sum_{i=1}^{n} d(x, x_i)^2. \tag{1}$$

Figure 1. The infinite family of trees on the right are TED means for the set of two trees on the left.

A local minimizer of Φ is called a *Karcher mean* while a global is called a *Fréchet mean*. But when does such a minimizer exist? When is it unique? Although the above definition does not require existence of geodesics, this is often needed in order to compute a minimizer. This reveals important problems already in the simplest of situations. If geodesics exist in X, a solution to the above problem for a set of two points a and b is the point c on the geodesic segment from a to b such that $d(a, c) = d(b, c)$. But what if there is more than one geodesic segment between a and b? The midpoint of each geodesic segment will minimize eq. 1.

A key example where this problem occurs is the (Tree) Edit Distance (denoted TED in the sequel) used in spaces of attributed graphs and trees, *e.g.*, shock graphs [5, 12, 16]. This metric is problematic as *even locally*, geodesics (edit paths) are not unique, and this prevents the existence of well-defined means even in a local context. For the pair of trees on the left in fig. 1 there is an infinite family of geodesics (and hence means) generated by varying the order and amount by which the side branches are shrunk and grown while deforming one tree into the other. A common approach for choosing a typical representative using TED is to choose the simplest possible mean, in this case the one shown in the middle. When iteratively computing means, however, one risks ending up with mean trees that are significantly simpler than the trees in the dataset. This explains the reduced complexity of the TED means found by Trinh and Kimia [16]. Similarly, in the graph embedding work of Bunke and collaborators [5, 15], severe restrictions on

ICCV
#445

ICCV
#445

ICCV 2011 Submission #445. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

the solution space and geodesics (edit paths) have to be imposed in order to have well defined constructions as simple as midpoints; restrictions that are kept implicit.

Even when geodesics are locally unique, means need not be unique. Consider data uniformly distributed along the equator of a sphere: both poles will minimize eq. 1. In shape analysis, this problem often occurs as most shape spaces are non-linear. In Riemannian manifolds, such as the sphere example, existence of at least locally minimizing geodesics is guaranteed. Tools from differential calculus are available for the optimization of eq. 1 and Karcher [11] provides conditions for local existence and uniqueness of the mean. Hence, in most statistical contexts one is content to find a (reasonably large) dataset radius within which geodesics and local minimizers of eq. 1 exist and are unique; for the unit sphere this radius is $\pi/2$.

The Appolonios problem (eq. 1) in Euclidean space can also be solved by geometric constructions using only geodesics and weighted midpoints, which carry over to more general metric spaces. But do different methods provide the same solutions? Even locally? Already in the Riemannian framework, non-linearity introduces difficulties in other constructions. A simple example is Principal Component Analysis, which in the Euclidean case can equally well be defined via maximization of projected variances on subspaces *or* minimization of reconstruction errors. But *these approaches lead to different solutions* in the Riemannian setting, *e.g.*, Fletcher's PGA [6] and Huckemann's GPCA [9] are not equivalent. With this in mind, one should tread carefully when generalizing from Euclidean methods.

Tree-like shapes (and more generally graph-like shapes), among others, present a great challenge, as they are not naturally modeled as elements of smooth manifolds. Any tree-shape can be obtained as a limit of a large number of tree-shapes with very different tree topologies, creating natural self-intersections in tree-space. This may prevent the use of smooth optimization methods. However, this complexity does not prevent the use of treelike shapes. In computer vision they appear in skeletons and shock-graphs for 2D shape recognition and classification [3, 12, 16], and they are often encountered in medical imaging, as airways and blood vasculatures have natural tree-like shapes [14, 17].

Feragen et al. [4] introduced a construction of tree-like shape spaces with a metric called *Quotient Euclidean Distance* (QED), which gives existence and local uniqueness of geodesics. This follows from the fact that they are locally *CAT(0)-spaces* or *spaces of non-positive curvature*, a concept introduced by Gromov [8] and discussed in the monograph [2]. Billera, Holmes and Vogtmann [1] have proposed a $CAT(0)$-space structure for phylogenetic trees, but these trees are abstract objects not encoding 2D or 3D shape, with much more restricted variations.

Spaces of tree-like shapes, in spite of their complexity,

offer a very good framework for computing means through geometric solutions to the Appolonios problem. In this paper we explore three such constructions: the centroid, Birkhoff shortening and weighted midpoints. The methods are tested on leaf vasculature shapes and airway tree shapes.

The rest of the article is organized as follows. In sec. 2 we discuss the space of tree-like shapes along with a metric that gives locally unique geodesics. We review the basics of metric geometry in spaces of non-positive curvature in sec. 3. This leads us to the theoretical novelty of the paper as we prove that unique means exist in such spaces in sec. 3.1. We then discuss various ways of defining and computing means (sec. 4) followed by an experimental comparison of these means (sec. 5). Finally, the paper is concluded with a brief summary and a discussion of open problems.

2. The space of tree-like shapes

We are interested in spaces of tree-like shapes as defined by Feragen et al. [4]. Tree-like shapes are represented as rooted, ordered, binary trees $T = (V, E, r, <)$ with edge attributes $f \colon E \to \mathbb{R}^n$. The attributes take values in $\mathbb{R}^n$, and describe the shape of the particular edge, *e.g.*, via landmark points. The branch order can, for instance, come from a planar ordering of branches. To compare the tree-shapes within one single shape-space, all tree-shapes are parametrized using the same combinatorial tree T, which is *sufficiently large* to describe all the tree-shapes in our dataset. Trees with fewer edges are represented by collapsing redundant branches, as in fig. 2. Higher-order bifurcations are represented in a similar fashion, also using collapsed branches. This gives a representation space $X = \prod_{e \in E} \mathbb{R}^n$, where all tree-like shapes are represented at least once. A space of ordered tree-like shapes $\bar{X}$ is defined as a quotient space of X by identifying different order-preserving representations of the same tree-shape. This corresponds to folding the Euclidean representation space and gluing it along different representations of the same trees. From the Euclidean metric on X, Feragen et al. induce the *quotient metric* on $\bar{X}$, which in this case is called the *QED metric*. The quotient metric is a standard mathematical construction [2], which here creates a piecewise Euclidean metric. If the Euclidean metric on X is replaced with an l_1 product metric, the TED metric is retrieved as quotient metric on $\bar{X}$.

Planar trees come with a given branch order (left to right) on the set of children of each branch, and hence a branch order on the entire tree is easily obtained. Trees that reside in 3D space are not ordered in the same way. In order to optimally compute the distance between two trees, it is necessary to consider all possible orders on the two trees. At a first glance, this seems to give problems with computational complexity. However, when studying shapes that are close together this is not necessary. We can induce an order on each tree by fixing an order on one (sufficiently large) tree,

ICCV
#445

ICCV 2011 Submission #445. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.
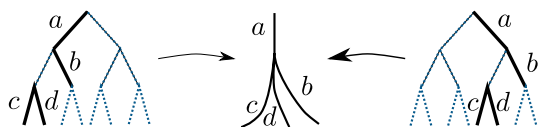
ICCV
#445

Figure 2. Higher-order vertices can be represented by the binary tree by collapsing internal branches, as the dotted blue lines.

aligning the other trees with the chosen one. The alignment can be done by finding the order that minimizes the distance between the trees. This way, we effectively reduce our set of unordered trees to a set of ordered trees. This becomes a great computational relief when we start computing average $3D$ trees, since we make a large number of distance computations as part of the iterative midpoint procedures.

3. Curvature and means in metric spaces

A very favorable property of the space of tree-like shapes is that its local geometry facilitates statistical computations. To be more precise, at "generic points" (that is, at any randomly chosen point) the space is locally $CAT(0)$. This concept is at the heart of our analysis, and is novel in the context of computer vision. Thus, we shall dedicate a few lines to explaining what $CAT(0)$ spaces are and why they are nice geometric objects.

One way of studying the geometry of general metric spaces is to compare them to spaces whose geometry we understand well, referred to as *model spaces*. The model spaces are spheres (positively curved), the Euclidean plane (flat, no curvature) and hyperbolic spaces (negatively curved). Since metric spaces can be rather pathological, a typical approach to defining curvature is to bound the curvature of the space at a given point from above or below. In this article we study spaces of non-positive curvature, *i.e.*, their curvature is bounded from above by 0. Due to the curvature bounded by 0, we study the metric spaces by comparing triangles in the metric space with triangles in the Euclidean plane, as in Gromov's metric geometry [2,8]. Mathematically, this is expressed by the $CAT(0)$ *condition*:

Definition 1 ($CAT(0)$ **spaces, non-positive curvature)**
Let X be a geodesic metric space, that is, a space in which all points can be joined by a geodesic. The length of a path is defined by the metric d in X. A geodesic triangle abc in X consists of three points a, b and c in X, along with geodesic paths joining the points: $[ab]$, $[bc]$ and $[ac]$, see fig. 3. There exists a triangle $\bar{a}\bar{b}\bar{c}$ in the Euclidean plane with vertices $\bar{a}$, $\bar{b}$ and $\bar{c}$ and with edges $[\bar{a}\bar{b}]$, $[\bar{b}\bar{c}]$ and $[\bar{a}\bar{c}]$, whose lengths are the same as the lengths of $[ab]$, $[bc]$ and $[ac]$. This is a comparison triangle for abc, see fig. 3.

For any point x sitting on the segment $[bc]$, there is a corresponding point $\bar{x}$ on the segment $[\bar{b}\bar{c}]$ in the compari-
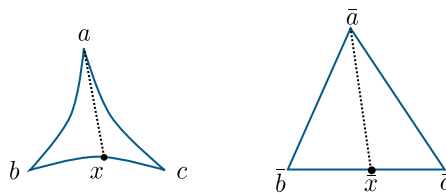

Figure 3. Left: A geodesic triangle, right: the corresponding comparison triangle in the plane $\mathbb{R}^2$.
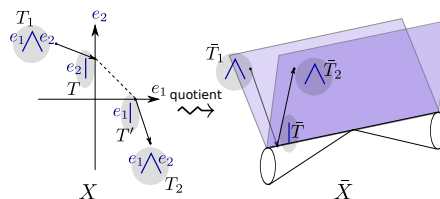

Figure 4. The simplest nontrivial example of a tree-shape space $\bar{X}$ (right) with representation space X (left): Trees with two edges e_1 and e_2, with scalar edge attributes. Here we see the path from $\bar{T}_1$, with two edges, through $\bar{T}$, with one edge, to $\bar{T}_2$, also with two edges. The one-branch tree $\bar{T}$ has two representatives in X on the e_1 and e_2 axes, respectively, which are glued together in the shape space (right). Note how the path from $\bar{T}_1$ to $\bar{T}_2$ in $\bar{X}$ is made from concatenated Euclidean lines in X, with a "teleportation" (dotted line) gluing the representatives T and T' of $\bar{T}$ together.

son triangle, such that $\|\bar{x} - \bar{b}\| = d(x, b)$. If we have

$$d(x, a) \leq \|x - a\| \qquad (2)$$

for every such x, then the geodesic triangle abc satisfies the $CAT(0)$ condition.

The metric space X is a $CAT(0)$ space if any geodesic triangle abc in X satisfies the $CAT(0)$ condition given in eq. 2. Geometrically, this means that triangles in X are thinner than triangles in $\mathbb{R}^2$. Spaces which are locally $CAT(0)$ are **non-positively curved.**

A few examples of $CAT(0)$ spaces are:

1) Euclidean space is a $CAT(0)$ space.
2) Any non-positively curved manifold is locally $CAT(0)$.
3) The union of two intersecting planes is a $CAT(0)$ space [2, Chapter II Theorem 11.3].
4) The space of tree-like shapes, see fig. 4, is locally $CAT(0)$ at generic points [4, Theorem 2].

3.1. Means in $CAT(0)$ spaces

As shown in [4] and discussed above, for generic points in the space of tree-like shapes there exists a radius $r_x > 0$ such that the ball $B(x, r_x)$ is a $CAT(0)$-space. For a point x in the shape space in fig. 4, whose distance to the projected

ICCV
#445

ICCV
#445

ICCV 2011 Submission #445. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

origin is d, $r_x = d\tan(\pi/8)$; namely the radius within which all points are joined by a unique geodesic. Hence, if means exist and are unique in $CAT(0)$ spaces, they exist and are unique for sufficiently dense sets of tree-like shapes.

Theorem 2 *Means exist and are unique in $CAT(0)$-spaces.*

Proof: Let (X, d) be a $CAT(0)$-space. It follows from the $CAT(0)$-inequality that given any fixed $y \in X$, the function $d_y \colon X \to \mathbb{R}$ given by $d_y(x) = d(x, y)$ is convex. Then the function d_y^2 is also convex since the function $g \colon a \mapsto a^2$ is monotone, increasing and convex. Moreover, $((1 - s)d_y(x) + sd_y(x'))^2 < (1 - s)d_y^2(x) + sd_y^2(x')$ for any $s \in]0, 1[$, and hence, d_y^2 is actually strictly convex. Then $\Phi = \sum_{i=1}^{s} d_{x_i}^2$ is strictly convex, and the mean is a minimizer of the strictly convex function Φ; hence it exists and is unique. $\square$

As a direct consequence, *means do exist and are even unique for datasets of sufficiently small diameter in tree-space with the QED metric.* This is in stark contrast to the TED metric where means are practically never unique.

4. Computing means

For finite point sets in the Euclidean space, the easiest way to compute the mean is the closed form solution $\sum x_i/N$. This solution does not carry over to non-linear spaces, but other techniques do.

One way to optimize eq. 1 is to use gradient descent. However, in the case of tree-structured data, the shape space is not even a smooth manifold. Hence, we cannot perform a regular gradient descent, but would have to develop optimization methods for non-smooth spaces. Moreover, we do not have an analytic expression for the gradient even in the smooth parts of the space. In order to numerically approximate derivatives, the distance function would need to be evaluated a large number of times. This makes optimization schemes unattractive for computing means in tree-space, where distances are expensive to compute.

Since the two most obvious methods for computing a mean shape are not applicable in tree-space, we need to look for alternative ways of computing means, which do not require evaluating too many distances. In this paper we study three iterative algorithms, which are all based on halving geodesics, namely centroids, Birkhoff shortening and weighted midpoints.

4.1. The centroid

The *centroid* $c(A)$ of a dataset $A = \{x_1, \ldots, x_N\}$ in a geodesic metric space (X, d), is defined as follows by Billera, Holmes and Vogtmann [1]:

If $N = 2$, then $c(A)$ is the midpoint of the geodesic connecting the two points x_1 and x_2. Assume that we have
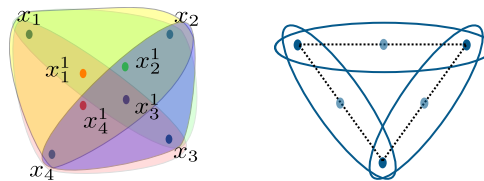

Figure 5. This illustration is best viewed in color. The centroid process is defined recursively with respect to the number of elements in the dataset. On the 4 point dataset $A = \{x_1, x_2, x_3, x_4\}$, each iteration consists finding the centroids of four 3-point subsets of A. Left: The iterative process illustrated for a 4 point dataset, and the centroids are denoted $x_1^1, x_2^1, x_3^1, x_4^1$. Right: For datasets with 3 points, the centroid process coincides with Birkhoff shortening.

a working definition of centroid for datasets with at most $N - 1$ points. Then define a set of subsets $A_1, \ldots, A_N$ by setting $A_i = A \backslash \{x_i\}$. Define a new N-element set $c^1(A) = \{c(A_1), \ldots, c(A_N)\}$, that is replacing A by centroids for each $(N - 1)$-element subset of A. This is illustrated in fig. 5. Define a sequence of sets $c^k(A)$ for $k \in \mathbb{N}$ by setting $c^k(A) = c^1(c^{k-1}(A))$; if the sequence $c^k(A)$ converges to a single point $c \in X$, then c is the *centroid* of A.

It is easy to see that in Euclidean space, the centroid is just the regular mean $\sum_i x_i/N$. Billera, Holmes and Vogtmann [1, Theorem 4.1] prove that in $CAT(0)$-spaces, the centroid construction converges to a unique point. It is, however, unknown whether this point is generally the mean, as defined in eq. 1, or a different point.

4.1.1 Algorithmic properties

The centroid is nice in theory since it defines a well-posed problem: centroids exist and are unique in $CAT(0)$ spaces. However, its algorithmic properties are not attractive. The computational complexity of computing $c(A)$ for a dataset A with N elements, is of the order N times the computational complexity of computing $c(A')$ for a dataset A' with $N - 1$ elements, *i.e.*, $\mathcal{O}(N!)$. In addition, each step involves an iterative convergence procedure, whose complexity is unknown. Combined with an expensive metric, the centroid is essentially intractable for datasets, which are sufficiently large to be interesting. This motivates us to investigate simpler algorithms that compute means in the Euclidean case.

4.2. Birkhoff shortening

Another method for computing means is given by Birkhoff curve shortening, which is used in metric and differential geometry to generate closed geodesic curves. Given a closed curve $\gamma \colon S^1 \to X$ from the unit circle into a (locally) geodesic metric space X, sample the curve by picking N points $z_1, \ldots, z_N$ on S^1 and consider their images $x_i = \gamma(z_i)$, setting $A = \{x_i | i = 1, \ldots, N\}$. Start an iterative process by replacing each point x_i by the midpoint

ICCV
#445

ICCV
#445

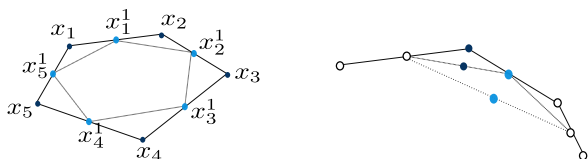ICCV 2011 Submission #445. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Figure 6. Left: Birkhoff curve shortening: Replace a sampled curve by new samples generated at the midpoint of each segment. The process converges to a closed geodesic. Right: By collecting midpoints from every second iteration (in colors), we get two Cauchy sequences, which must converge since the space is complete. It follows that the procedure converges to a closed curve.

x_i^1 of the geodesic connecting x_i to x_{i+1}, where we define $x_{N+1} = x_1$. Together with the geodesic segments we now have a new closed curve; see fig. 6 for an illustration. This process can be continued until convergence, which is ensured by Theorem 3 below. Note that when A contains three elements, the centroid procedure coincides with the Birkhoff shortening, see also fig. 5.

In Euclidean space it is easy to see that this process converges to the mean of the point set A. Trinh and Kimia [16, Conjecture 1] conjecture that the same holds for a wide range of spaces; we are less optimistic and conjecture that the claim holds locally at generic points for tree-space with the QED metric.

Theorem 3 *Let (X, d) be a complete, geodesic metric space and suppose given an ordered set of points $A = \{x_1, \ldots, x_N\}$ in (X, d). The Birkhoff shortening process converges to a closed geodesic, and if the space (X, d) does not have closed geodesics, then the Birkhoff shortening process converges to a single point.*

A simpler form of this theorem is stated without proof by Trinh and Kimia [16, Proposition 1]; we give the proof in the case of a geodesic metric space, as the proof also sheds some light on the difficulties and potential dangers.

Proof of Theorem 3: We first note that each step in the Birkhoff shortening process *will* make the loop in question shorter, assuming that the current curve is not a closed geodesic. To see this, note that

$$
\begin{aligned}
\sum d(x_i, x_{i+1}) \quad &= \sum \left(d(x_i, x_i^1) + d(x_i^1, x_{i+1}) \right) \\
&= \sum \left(d(x_i^1, x_{i+1}) + d(x_{i+1}, x_{i+1}^1) \right) \\
&\geq \sum d(x_i^1, x_{i+1}^1),
\end{aligned}
\tag{3}
$$

where the first equality comes from the fact that the points x_i^1 are midpoints of geodesic segments, the second equality is a rearrangement of terms, and the last inequality comes from the triangle inequality; see fig. 6. This shows that Birkhoff shortening will *not* make the curve longer. If the original loop is not a geodesic curve, then at one of the points x_i, the loop is not a local geodesic. The curve connecting the midpoints before and after x_i is not a geodesic,

and hence, replacing with a geodesic *will* create a strictly shorter loop.

Next, we need to show that the process converges. The lengths of the consecutive curves form a decreasing sequence of non-negative real numbers, which must converge towards some length l. Moreover, the odd/even midpoint sequences shown in fig. 6 are Cauchy (follows from the $CAT(0)$ criterion) and must converge, so the sequence of loops converges to a new loop.

Finally, since we have already shown that for non-geodesic loops, Birkhoff shortening *will* make the curve shorter, the limit loop must be a closed geodesic. But then, if the space does not have closed geodesics, Birkhoff shortening will converge towards *one* point. □

Corollary 4 *Since they have unique geodesics [2, Proposition II.1.4], $CAT(0)$ spaces cannot have closed geodesics. Hence, the Birkhoff shortening procedure converges to a point in $CAT(0)$ spaces.* **In particular, it converges for sets of tree-like shapes with sufficiently small diameter.**

We have shown that for any given initial order on a sufficiently bounded set of tree-like shapes, the Birkhoff shortening procedure converges to a point, but we do not know whether any two orders give the same point. For general $CAT(0)$ spaces, we doubt that this is true. Consider the Birkhoff shortening process shown in fig. 6 in a space whose curvature varies strongly (although staying non-positive). The shortening process would be highly asymmetric and it is unlikely that different set orders would give the same mean. This also gives potential problems with one of the algorithms used for TED-means by Trinh and Kimia, where the set order is permuted in every iteration in order to speed up convergence. However, in tree-space, the local structure is nice almost everywhere, being either flat or an intersection of flat regions. Hence, we believe that the Birkhoff shortening mean may well be independent of its initial order for sets of tree-like shapes in the QED metric.

The computational complexity of each step in the Birkhoff shortening algorithm is $\mathcal{O}(N)$ times the complexity of finding geodesic midpoints. This makes the Birkhoff shortening procedure suitable for computing means when distances are expensive to compute. The iterative nature of the procedure does, however, make it vulnerable to accumulation of numerical noise and approximation errors.

4.3. Weighted midpoints

One of the simplest algorithms for computing the mean of a set in $\mathbb{R}^n$ is based on the following simple observation:

If we denote by $m(A)$ the mean of the finite subset $A = \{x_1, \ldots, x_N\}$, then

$$
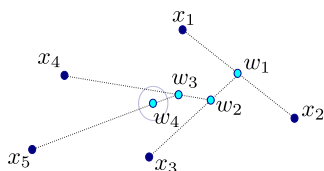m(A) = \frac{x_i + (N-1)m(A \setminus \{x_i\})}{N}.
\tag{5}
$$

Figure 7. The weighted midpoints construction.

Figure 8. Seven synthetic planar trees.

To see that eq. 5 holds for $A \subset \mathbb{R}^n$, just rewrite the equation analytically in the case $i = 1$:

$$\frac{x_1 + \ldots + x_N}{N} = \frac{x_1 + (N-1) \cdot \frac{x_2 + \ldots + x_N}{N-1}}{N}. \quad (6)$$

This indicates a recursive procedure for finding the mean of A in more general spaces: The mean w_1 of $\{x_1, x_2\}$ is the midpoint of the geodesic connecting x_1 to x_2. The mean w_2 of $\{x_1, x_2, x_3\}$ is the point on the geodesic from $w_1 = m(x_1, x_2)$ to x_3 that sits $1/3$ along the way, etc. This is a finite procedure, whose result we call the *weighted midpoints* mean. See also fig. 7.

As the Birkhoff shortening mean, the weighted midpoints procedure also depends on an initial set order, and it is not clear whether different orders give the same results. Again, we believe that the nice local structure of tree-space may be enough to secure independence of initial order.

5. Experiments

We now experimentally compare the different approaches on datasets of tree-structured shapes in the space of tree-like shapes [4] endowed with the QED metric. Specifically, we test the algorithms on three different datasets of varying difficulty and size.

The QED geodesics and midpoints were computed using Algorithm 1 from the article by Feragen et al. [4] on depth 3 trees, leaving space for 1 or 2 structural changes. This is an approximative algorithm, where the geodesics might pass through trees of depth higher than 3 in order to make the structural changes. Whenever the midpoints are of depth > 3, they are cut off at depth 3 in order to initiate the next iteration. This introduces some numerical errors, which may accumulate in long iterative procedures.

5.1. Synthetic data

The first test set consists of synthetic planar trees which are designed to test the system's ability to cope with pairs of bifurcations that are close to forming trifurcations. The whole dataset is shown in fig. 8. Additional figures and movies illustrating the iterative processes are found in the supplementary material.

Small dataset. First, all three different algorithms were ran on a smaller set consisting of the 4 synthetic planar trees shown in the top row of fig. 8. Since the weighted midpoints

and Birkhoff shortening algorithms potentially depend on the order of the dataset, they were ran several times with different randomly selected initial orders. In fig. 9(a) we see the results of the weighted midpoints algorithm for all possible orders on the dataset plotted together. In fig. 9(b), we see the points in the Birkhoff shortening 5th iteration on the same dataset. In fig. 9(c) we see the result of the centroid algorithm. We clearly see that the three algorithms give qualitatively very similar results, also with the different initial orders used in the weighted midpoints and Birkhoff shortening algorithms.

In order to experimentally check whether the found centroid c actually minimizes the function Φ from eq. 1, we selected 100 tree-shapes by adding random normal distributed noise to the centroid edges, and evaluated Φ at each point. The smallest value of Φ was found at the centroid, indicating that the found centroid actually is a mean.

Full dataset. The weighted means and Birkhoff shortening algorithms were also tested on the whole dataset in fig. 8, again using different orders. The results are found in fig. 10 and fig. 11. The centroid was left out of this experiment, since already here, the complexity is too demanding.

Although the Birkhoff shortening and weighted midpoints algorithms depend on the order of the dataset, we clearly see that the attained results are robust with respect to varying initial order. More importantly, we see that the results of the faster methods are practically identical to the centroid, which has good theoretical properties, but is expensive to compute. This is very comforting as it indicates that we can compute usable means with the Birkhoff shortening and weighted midpoints algorithms. In the remainder of this paper, we will only use these algorithms as the centroid is computationally too demanding.

5.2. Leaf morphology

Our first example of natural tree-like structures comes from biology. In botany, tree-like structures are found as vasculatures in leaves, and are studied in order to understand leaf morphology [7]. These structures form excellent proof-of-concept examples, as the tree-structures are necessarily planar, and hence the branches are ordered nicely from left to right.

We extract vascular structures for a set of 10 ivy leaves, see fig. 12, giving the planar trees shown in fig. 13. Using weighted midpoints for six randomly chosen dataset or-
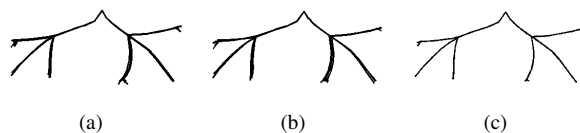
(a)   (b)   (c)

Figure 9. Comparison of algorithms on a set of four synthetic planar trees. (a) Results of weighted midpoints on the first four trees from fig. 8. The algorithm was run for all different initial orders and the results are plotted on top of each other. (b) Result of Birkhoff shortening after 5 iterations, starting from the shown set order on the first four trees from fig. 8. The 5th iteration points are plotted on top of each other. (c) Result of the centroid algorithm on the first four trees from fig. 8.

Figure 10. Weighted midpoints results for nine eight different initial orders on the dataset shown in fig. 8.

Figure 11. Birkhoff shortening results for three different orders on the dataset shown in fig. 8.
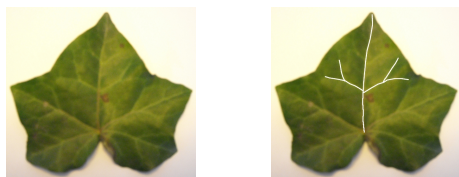

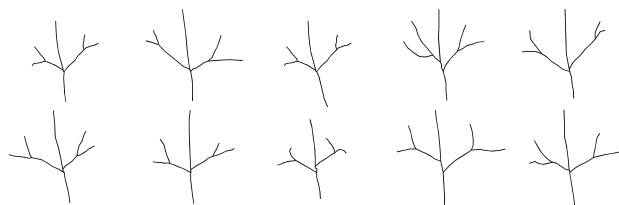Figure 12. An ivy leaf, with a subtree of its vascular tree.

Figure 13. Vascular structures from 10 ivy leaves.

ders we obtain the mean trees showed in fig. 14. Again, the weighted midpoints mean trees initiated with different initial orders look nearly identical. Similarly, in fig. 15, we see 12th iteration Birkhoff shortening mean trees for a random initial order, which are also very similar to the weighted midpoints results. This is a clear indication that the two

Figure 14. weighted midpoints results for the leaves shown in fig. 13 for five random dataset orders.
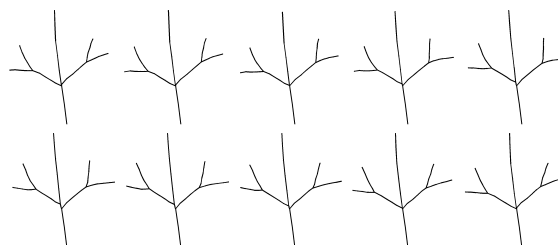

Figure 15. The Birkhoff shortening 12th iteration results for the leaves shown in fig. 13 for a random dataset order.

algorithms compute the same mean.

5.3. Airway tree-shape modeling

The study of tree-like shapes is strongly motivated by their presence in human anatomy, where they appear as delivery systems for fluids and air. Birkhoff shortening and weighted midpoints means are computed for a dataset consisting of 10 $3D$ airway trees extracted from CT scans [13]. The trees represent the centerlines of the first four generations of the human airway tree as shown in fig. 16. The trees are aligned at the endpoint of the root branch (trachea). As with the planar trees, the two algorithms provide very similar means. Both algorithms, and Birkhoff shortening in particular, seem robust even for $3D$ trees.

6. Discussion and conclusion

We have studied the concept of means in non-Euclidean spaces; generally in $CAT(0)$ spaces and specifically in spaces of tree-like shapes. We have shown that means exist in $CAT(0)$ spaces; a result which tells us that means exists locally in spaces of tree-like shapes. The generality of the result allows it to be transferred to other settings. In particular, our results should transfer to the space of attributed graphs defined by Jain and Obermayer [10].

Usually, means can be found in non-Euclidean spaces using standard optimization techniques such as gradient descent. The computational complexity of the QED metric, however, makes this approach infeasible for trees. We consider three different algorithms for computing means in Euclidean space: the centroid, Birkhoff shortening and weighted midpoints. As these rely on on dividing geodesics they are readily generalized to non-Euclidean spaces.

From a theoretical point of view, the centroid is the best

ICCV
#445

ICCV
#445

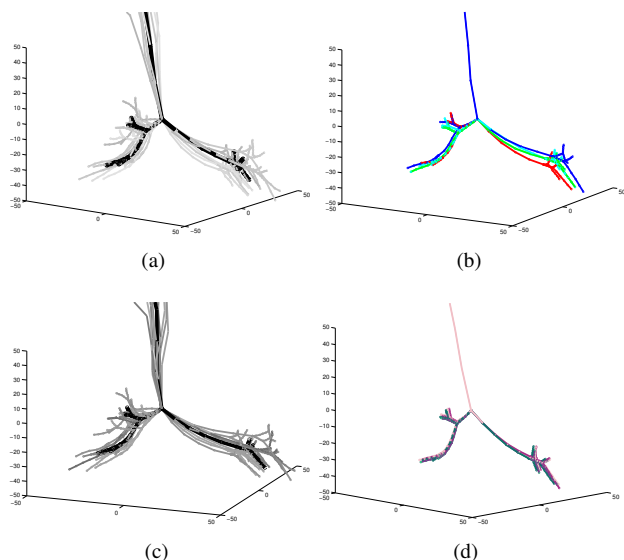ICCV 2011 Submission #445. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Figure 16. (a) The ten airway trees in the dataset are shown in gray; in black is one of the weighted midpoints means. (b) Five weighted midpoints means computed from different initial orders. (c) The ten airway trees in the dataset are shown in gray; in black is the Birkhoff shortening mean. (d) Five Birkhoff shortening means computed from different initial orders.

definition of mean shape. It exists, is unique and is, by definition, invariant of the order of the dataset. Moreover, numerical experiments indicate that it coincides with the mean. The algorithm, however, has complexity $\mathcal{O}(N!)$, which makes it too expensive to be of practical use. On the other hand, the Birkhoff shortening and weighted midpoints means are not quite as nice; while they converge to a single point, this point may depend on the order of the data. This is not ideal. However, experimental results indicate that, up to numerical and approximative noise, different orders actually give the same means for both methods. Even better: these means seem to coincide with the well-defined centroid. On this basis, we conjecture that for tree-structured data, all algorithms compute the same mean tree. As to which method works the best, the simplest seems to win. Birkhoff shortening is an iterative procedure, which makes it slow and vulnerable to accumulating errors. The weighted midpoints mean, on the other hand, comes out as a robust and efficient way of computing mean trees.

We have, thus, provided a practical algorithm for computing mathematically well-defined "average trees"; something that has not been presented elsewhere in the literature. This can potentially serve as a much-needed tool in medical image analysis, where airways and vascularization systems serve as reference structures in the human body. Shape statistics on these structures may provide new insight into how diseases such as COPD (smokers lung) affects the ge-

ometry of human anatomy.

References

[1] L. J. Billera, S. P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.*, 27(4):733–767, 2001. 2, 4

[2] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer-Verlag, 1999. 2, 3, 5

[3] F. Demirci, A. Shokoufandeh, and S. Dickinson. Skeletal shape abstraction from examples. *PAMI*, 31(5):944–952, 2009. 2

[4] A. Feragen, F. Lauze, P. Lo, M. de Bruijne, and M. Nielsen. Geometries in spaces of treelike shapes. In *ACCV*, pages 671–684, 2010. 2, 3, 6

[5] M. Ferrer, E. Valveny, F. Serratosa, K. Riesen, and H. Bunke. Generalized median graph computation by means of graph embedding in vector spaces. *Pattern Recognition*, 43(4):1642 – 1655, 2010. 1

[6] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *TMI*, 23:995–1005, 2004. 2

[7] A. S. Foster and H. J. Arnott. Morphology and dichotomous vasculature of the leaf of kingdonia uniflora. *American Journal of Botany*, 47(8):pp. 684–698, 1960. 6

[8] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. 1987. 2, 3

[9] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica*, 20(1):1–58, 2010. 2

[10] B. Jain and K. Obermayer. Large sample statistics in the domain of graphs. In *SSPR*, volume 6218 of *LNCS*, pages 690–697. 2010. 1, 7

[11] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977. 1, 2

[12] P. Klein, S. Tirthapura, D. Sharvit, and B. Kimia. A tree-edit-distance algorithm for comparing simple, closed shapes. In *SODA*, pages 696–704, 2000. 1, 2

[13] P. Lo, J. Sporring, H. Ashraf, J. J. Pedersen, and M. de Bruijne. Vessel-guided airway tree segmentation: A voxel classification approach. *Medical Image Analysis*, 14(4):527–538, 2010. 7

[14] J. H. Metzen, T. Kröger, A. Schenk, S. Zidowitz, H.-O. Peitgen, and X. Jiang. Matching of anatomical tree structures for registration of medical images. *Im. Vis. Comp.*, 27:923–933, 2009. 2

[15] K. Riesen and H. Bunke. Graph classification based on vector space embedding. *Int. Journal of Pattern Recognition & Artificial Intelligence*, 23(6):1053 – 1081, 2009. 1

[16] N. Trinh and B. Kimia. Learning prototypical shapes for object categories. In *CVPR Workshops*, pages 1–8, 2010. 1, 2, 5

[17] J. Tschirren, G. McLennan, K. Palágyi, E. A. Hoffman, and M. Sonka. Matching and anatomical labeling of human airway tree. *TMI*, 24(12):1540–1547, 2005. 2