A Reeb Graph Approach to Tractography

Jonathan Sun UC Santa Barbara jsun@cs.ucsb.edu

Matthew Cieslak UC Santa Barbara matthew.cieslak@psych.ucsb.edu

Scott Grafton UC Santa Barbara scott.grafton@psych.ucsb.edu suri@cs.ucsb.edu

Subhash Suri UC Santa Barbara

ABSTRACT

We propose an efficient algorithm for discovering the highlevel topological structure of a collection of 3-dimensional trajectories. Our algorithm computes a sparse graph representing the latent "bundling" and "unbundling" structure of the trajectory data. This graph can serve both as a compact signature of the trajectory data set as well as a tool for efficient comparison among different data sets. Our problem formulation and the algorithms are broadly applicable and general-purpose but we focus on a particular neuroscience application to highlight the key features. In particular, our motivation stems from the emerging area of brain tractography, which aims to construct the connectome of human brain white matter fibers. These fibers can be inferred noninvasively using magnetic resonance imaging (MRI) diffusion scans of the brain interior and modeled abstractly as a set of time-independent geometric trajectories in a threedimensional brain space. Real neuronal fiber pathways exhibit complex but natural bundling structures, which elude existing MRI reconstruction techniques, but are easily captured by our algorithm. We validate our algorithms both theoretically (uniqueness of the graph representation and provably efficient algorithms) and empirically (using both synthetic and real scanned brain data sets).

Categories and Subject Descriptors

[Big Spatial Data, Computational Geometry, Spatial Data Analytics

1. INTRODUCTION

In this paper, we are interested in extracting a sparse but structurally significant representation of a complex set of 3dimensional trajectories. Broadly defined as a point-sequence in some coordinate space, a trajectory is a convenient abstraction in a wide range of applications dealing with mobility patterns, spatial phenomena, geographical networks or geometric connectivity, among others. Trajectory analvsis often centers on discovering geometric and topological structure that is common to subsets of trajectories.

The focus of this paper is the discovery of *bundling structure*, namely, how the groups of trajectories branch and merge. In some settings, the bundling phenomenon is wide-spread but easily visible in the data. Our work looks at a specific situation where the branching and merging is critical but requires significant computational effort to discover because of the (1) large scale of data, (2) complex behavior of trajectories, (3) non-uniformity of bundles, and (4) difficulties caused by noise and missing data. While our techniques and results are general-purpose, we describe them in the context of human brain's white matter connectivity as a specific application.

Tractography: Fiber Pathways of the Brain. Brain connections are built from morphologically complex 3-dimensional fiber bundles that are not randomly organized; in fact, much of the white matter structure is constrained by the brain's anatomic development process. White matter fibers tend to start at a functional region, merge with nearby fibers into larger pathways for efficiency, and split off as they near their terminal functional regions. As such, the structure discovery problem here is a mix of geometry, graph theory, and topology. The use of geometry alone is highly susceptible to noise because of the fiber resolution, which is at sub-micron scale, as well as natural physical disparities (in size and shape) between human brains. On the other hand, general-purpose graph-theoretical methods are both computationally intractable (i.e. subgraph isomorphism) and lose vital information regarding the physical proximity and bundling structure of fibers.

Results. We propose using the topological concept of *Reeb* Graph as a way to represent the bundling structure in a discrete fashion that remains faithful to the original geometric features. Our representation combines the geometry of the fiber pathways with graph-theoretic tools in a complemen-

^{*}Brain research and datasets funded by the General Electric/National Football League Head Health Challenge and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office.

tary way. In particular, we apply techniques from computational geometry to track fiber trajectories, grouping and splitting them as needed to encode the spatial branching and merging structure that underlies the neuronal pathways. We then process the groups into a Reeb graph structure that allows the relevant geometric structure to be recovered while providing a high-level view of the overall topological structure. While the basic idea is simple and elegant, there are a number of computational and conceptual challenges in ensuring algorithmic efficiency and being able to identify important structural variations from random variation and noise.

Our main contributions include the following: (1) an efficient algorithm for partitioning the trajectory data into ε connected bundles; the algorithm runs in worst-case time $O(N \log N)$, where N is the total input size, and (2) an implementation and empirical validation of the algorithm on real brain imaging data.

2. PREVIOUS WORK

The analysis of trajectory data is a well-studied problem with a vast literature, especially due to the proliferation of GPS data in recent years. Existing theoretical approaches for comparing and grouping trajectories are explored in [23, 7, 20, 19, 12, 11]. Real-world applications of trajectory analysis, such as tracking animal herds and vehicle traffic, are discussed in [15, 16, 2, 10, 18, 1]. *Tractography*, the science of using trajectories to model human brain white matter from diffusion magnetic resonance imaging (dMRI) data, is described in [4, 3, 21, 8]. Brain fiber trajectories pose a distinct challenge over GPS-based data because they are purely *spatial* (time-independent) and three-dimensional, but are known to follow well-established anatomical constraints.

More recent literature addresses the major shortcoming of early trajectory clustering techniques in requiring each trajectory to be classified into a group as a whole. Trajectory grouping algorithms that allow segmentation in order to discover common substructure are described in [14, 17, 2, 19]. Methods for summarizing the grouping structure are explord in [17, 5, 6, 22]. Notably, [6] introduces the application of the Reeb graph as a way to succinctly represent trajectory data in a way that encodes the merging and splitting structure. Our approach extends their technique to time-independent trajectories, which is a strictly harder case, and refines the grouping definition so that the resulting Reeb graph also partitions the set of input points, so that the groups can be labeled with domain-specific information such as anatomical features.

3. REEB GRAPH OF TRAJECTORIES

A trajectory T is an ordered sequence of points in some metric space \mathcal{M} . We denote a trajectory T as a sequence of points $p_1p_2p_3\cdots p_m$, where m is the number of points in T. Equivalently, a trajectory can be represented by a sequence of segments $s_1s_2s_3\cdots s_{m-1}$, where $s_i = p_ip_{i+1}$. A subtrajectory T' of a parent trajectory T is a subsequence of segments in T beginning at some index s and ending at t, i.e. $p_sp_{s+1}\cdots p_{t-1}p_t$. Two disjoint subtrajectories of the same parent T are adjacent if they share an endpoint, or, equivalently, if their union is also a subtrajectory of T. (Depending on the application, trajectories can be *time-dependent*, where each point has an associated timestamp, and the points in the trajectory are ordered by increasing time, or *time-independent* where no timestamps are involved. Our focus in this paper is the case of *time-independent*, or purely spatial, trajectories.) Throughout we assume that the ambient space is the 3-dimensional Euclidean space with n trajectories, with a total of N points.

Many real-world trajectories exhibit highly structured *merg*ing and splitting behavior, where a group of trajectories travel close together in one region, then separate and travel farther apart in another region. In order to discover the large-scale topological structure of trajectory data, we represent the bundling structure of trajectories in the form of a *Reeb graph*. Originally defined in Morse theory as a way to describe critical points of a manifold, we use the Reeb graph as a way to represent the merging and splitting behavior (which we call the *branching structure* of a trajectory set.

Intuitively, if a continuous portion of a set of trajectories moves together, i.e. stays within close distance of one another, then these subtrajectories share a common behavior. We formalize this by introducing the notion of subtrajectory bundles. For any segments s_1 and s_2 in \mathcal{M} let $d(s_1, s_2)$ be a distance metric on segments (for the brain connectome data, we use the simple metric of the maximum distance between endpoints). For any $\varepsilon > 0$, we say two trajectories T_i and T_j are ε -connected if every segment in T_i is within ε of some segment in T_j and vice versa. Note that the definition of ε -connected extends to subtrajectories as well.

Given an input set of trajectories $\mathcal{I} = \{T_1, T_2, \cdots, T_n\}$, a bundle B is a set of subtrajectories which contains at most one subtrajectory from each trajectory in the input set. Finally, a bundle B is an ε -bundle if every subtrajectory $T \in B$ is ε -step connected, i.e. between any two subtrajectories in B there exists an ε -connected sequence of subtrajectories connecting them.

Two bundles B and B' are *adjacent* if a pair of their subtrajectories is adjacent, i.e. if there exists some subtrajectory $T \in B$ that is continued as T' in B'. A bundle partition $\mathcal{P} = \{B_1, B_2, \cdots, B_k\}$ for \mathcal{I} is a set of bundles such that every segment in I is assigned to exactly one bundle. A bundle partition can be thought of as a clustering on subintervals of trajectories in \mathcal{I} . For every input set \mathcal{I} there may be multiple bundle partitions that are wholly composed of ε -bundles; for example, the trivial partition where each trajectory is a bundle. However, for a bundle partition to capture *all* the grouping structure we restrict our consideration to the partition composed of *max-width* ε -bundles. Formally, an ε -bundle B of \mathcal{I} is max-width if no other possible ε -bundle of \mathcal{I} intersects B and contains a superset of the trajectories represented in B.

We can show that for any input set \mathcal{I} and any parameter ε , the max-width ε -bundle partition that *minimizes* the total number of bundles is unique *if* all the trajectories satisfy the property that two segments from the same trajectory do not lie in each other's ε -step neighborhoods. This property is exhibited in real-world data sets like brain fibers, where trajectories do not form "loops" or other structures with ambiguous connectivity. We defer the proof to the full paper.

Let \mathcal{I} be a set of trajectories, and let \mathcal{P} be the max-width bundle partition for \mathcal{I} . The Reeb graph \mathcal{R} for \mathcal{P} is an undirected graph that associates for each bundle $B_i \in \mathcal{P}$ an edge e_i in \mathcal{R} , and connects two edges e_i and e_j with a vertex if their corresponding bundles are adjacent. Intuitively, the Reeb graph succinctly captures the branching structure of the trajectories: its vertices are either endpoints of subtrajectory groups or critical points in the data at which merging and splitting behavior occur, and the edges adjacent to each vertex are the subtrajectory groups that are involved in the critical behavior at that region. The algorithm parameter ε controls the granularity of the bundling desired-small values of ε allow only very dense sets of subtrajectories to be considered for grouping, while larger values of ε relax the groups and allow larger, sparser groups to form. We summarize our algorithmic result in the following theorem, deferring the details of the algorithm to the full paper.

THEOREM 3.1. Given an input set of trajectories $\mathcal I$ and a distance parameter ε , the Reeb graph \mathcal{R} for \mathcal{I} can be computed in $O(N \log N)$ time, where N is the total number of points in the input set.

ALGORITHM IMPLEMENTATION AND 4. **EXPERIMENTAL RESULTS**

In order to form a *qualitative* impression of the kind of bundle formations that our algorithm discovers, we first illustrate the behavior of the algorithm on a few synthetically generated but distinctive trajectory data sets. These small examples are designed to highlight the split-merge phenomena among white matter fibers considered significant by neuroscientists in their study of brain imaging data. These examples show how the Reeb graph both simplifies the structure of trajectory sets when they share common behavior but also preserves geometric faithfulness, especially for substructures within the same trajectory.





A basic ex- (b) An example of (c) An example distinct groupings. branching events.

overlapping

Figure 1: Synthetic trajectory samples demonstrating basic, but nontrivial branching structure where simple clustering methods fail to identify all the latent substructure.

We construct three different trajectory data sets with nontrivial topologies (shown in Figure 1, each highlighting a distinct type of bundling structure. In each case, the input set of trajectories can be easily *clustered* using traditional clustering methods, but those groupings fail to capture the



Figure 2: Reeb graph representations of the synthetic data. White vertices identify termination regions, while grey vertices identify branching points. Edges record the common subtrajectories of the groupings.

unique and distinctive spatial and topological pattern that is vital to these sets, and is correctly found by our algorithm. The Reeb graph representation in Figure 2 succinctly identifies the these substructures and, if trajectory information is stored with each edge, provides means for recovering the original geometric structure.

4.1 **Brain Imaging Data**

We collected a series of scans from a single subject from the UCSB Brain Imaging Center. White matter trajectories were generated using the algorithm described in [25]. To test the Reeb graph computation, we reconstructed individual trajectory sets for the arcuate fasciculus and the superior longitudinal fasciculus. The anatomy of these two structures is well-understood within the neuroscience community and is known to display strong geometric structure, with welldefined termination points that branch off from larger and larger bundles. Consequently, the branching and merging structure of the physical tissue of the arcuate fasciculus is thoroughly described in [13, 9] and the superior longitudinal fasciculus is described in [24]. Thus, each Reeb graph's attributes can be verified against neurological ground truth.

The trajectories reside in a normalized three-dimensional brain space that is $182 \times 218 \times 182$ mm in size. Selecting an optimal value of ε depends on the application in question; we use a value of $\varepsilon = 2.25$ for two reasons: it is close enough to the minimum dMRI resolution of 2mm to capture the smallest reasonable groups, but large enough to ensure that macroscopic groupings are identified.





fasciculus [9].

(a) Brain atlas of the arcuate (b) Brain atlas of the superior longitudinal fasciculus [24].

Figure 3: Anatomical ground-truth labeling for two wellknown neurological structures. Important macroscopic termination regions are labeled based on physical dissection.

We use manually labeled diagrams of the arcuate fasciculus and superior longitudinal fasciculus from Figure 3 as our



(a) Reeb graph partition of the trajectories in Figure 3a.

(b) Reeb graph representation of arcuate fasciculus.

Figure 4: Reeb graph labeling and representation of the arcuate fasciculus, a well-known neurological structure, produced by our algorithm. Colored bundles are reflected in the graph edges. White vertices identify termination regions, grey vertices identify branching points.



(a) Reeb graph partition of the trajectories in Figure 3b.

(b) Reeb graph representation of longitudinal fasciculus.

Figure 5: Reeb graph labeling and representation of the superior longitudinal fasciculus, another well-known neuro-logical structure, produced by our algorithm.

reference points and compare them with colored diagrams of the Reeb graph on the trajectory set for our subject in Figures 4 and 5. We find that our algorithm discovers most of the known termination regions and assigns them edges in the Reeb graph, and also discovers unique substructures within the individual that are absent from the brain atlas. Furthermore, our algorithm discovers the large central arcuate bundle and superior longitudinal bundle that unites all of the separate connections in each dataset. In regions where the trajectories are more sparsely distributed or noisy, the algorithm has some difficulty finding distinct groups. Methods for reducing the effect of noisy imaging data or outliers are an ongoing area of work.

4.2 Runtime Analysis

We evaluated the computational efficiency of our algorithm on an AMD dual-core 1.4Mhz processor with 4GB of RAM. Our input size ranged from 50 to 5,000 trajectories representing sets of white matter pathways in the human brain, and our average trajectory length was 86 points per trajectory. The resulting runtime ranged from 2 seconds to 78 minutes.

5. REFERENCES

- M. Ahmed, S. Karagiorgou, D. Pfoser, and C. Wenk. A comparison and evaluation of map construction algorithms using vehicle tracking data. *GeoInformatica*, 19(3):601–632, 2015.
- [2] M. Andersson, J. Gudmundsson, P. Laube, and T. Wolle. Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, 12(4):497–528, 2008.
- [3] P. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44(4):625–632, 2000.
- [4] P. J. Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. NMR in Biomedicine, 8:333–344, 1995.
- [5] K. Buchin, M. Buchin, M. van Kreveld, M. Löffler, R. . Silveira, C. Wenk, and L. Wiratma. Median trajectories. *Algorithmica*, 66(3):595–614, 2013.
- [6] K. Buchin, M. Buchin, M. J. van Kreveld, B. Speckmann, and F. Staals. Trajectory grouping structures. *CoRR*, abs/1303.6127, 2013.
- [7] L. Chen, M. T. Özsu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *Proc. ACM SIGMOD*, pages 491–502, 2005.
- [8] M. Cieslak and S. T. Grafton. Local termination pattern analysis: a tool for comparing white matter morphology. *Brain* imaging and behavior, 2013.
- [9] J. C. Fernández-Miranda, Y. Wang, S. Pathak, L. Stefaneau, T. Verstynen, and F. Yeh. Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain* structure & function, 2014.
- [10] Z. Fu, W. Hu, and T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *IEEE International Conference on Image Processing*, 2005.
- [11] S. Gaffney, A. Robertson, P. Smyth, S. . Camargo, and M. Ghil. Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics*, 29(4):423–440, 2007.
- [12] S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In Proc. ACM KDD, pages 63–72, 1999.
- [13] M. F. Glasser and J. K. Rilling. DTI Tractography of the Human Brain's Language Pathways. *Cerebral Cortex*, 18(11):2471–2482, 2008.
- [14] J. Gudmundsson, M. Kreveld, and B. Speckmann. Efficient detection of patterns in 2d trajectories of moving points. *Geoinformatica*, 11:195–215, 2006.
- [15] J. Gudmundsson and M. van Kreveld. Computing longest duration flocks in trajectory data. In Proc. 14th Annual Symposium on Advances in Geographic Information Systems, pages 35–42, 2006.
- [16] Y. Huang, C. Chen, and P. Dong. Modeling herds and their evolvements from trajectory data. In *Geographic Information Science*, pages 90–105. 2008.
- [17] J. Lee and J. Han. Trajectory clustering: A partition-and-group framework. In Proc. SIGMOD, pages 593–604, 2007.
- [18] X. Li, W. Hu, and W. Hu. A coarse-to-fine strategy for vehicle motion trajectory clustering. In 18th International Conference on Pattern Recognition, pages 591–594, 2006.
- [19] S. Sankararaman, P. K. Agarwal, T. Mølhave, and A. P. Boedihardjo. Computing similarity between a pair of trajectories. *CoRR*, 2013.
- [20] E. Tiakas, A. N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic, and S. Djordjevic-Kajan. Trajectory similarity search in spatial networks. In Prof. 10th Int. Database Engineering and Applications Symposium, pages 185–192, 2006.
- [21] D. C. Van Essen and K. Ugurbil. The future of the human connectome. *Neuroimage*, 62(2):1299–1310, 2012.
- [22] M. J. van Kreveld, M. Löffler, and F. Staals. Central trajectories. CoRR, abs/1501.01822, 2015.
- [23] M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. In Proc. 18th International Conference on Data Engineering, pages 673–684, 2002.
- [24] X. Wang, S. Pathak, L. Stefaneanu, F.-C. Yeh, S. Li, and J. C. Fernandez-Miranda. Subcomponents and connectivity of the superior longitudinal fasciculus in the human brain. *Brain Structure and Function*, 2015.
- [25] F.-C. Yeh, V. Wedeen, and W.-Y. Tseng. Generalized q -sampling imaging. *Medical Imaging, IEEE Transactions on*, 29(9):1626–1635, 2010.