Identifying Common Portions between Two Trajectories

Swaminathan Sankararaman* Pankaj K. Agarwal* Thomas Mølhave* Arnold P. Boedihardjo†
*Dept. of Computer Science, Duke University. email: {swami,pankaj,thomasm}@cs.duke.edu
†U.S. Army Corps of Engineers. email: arnold.p.boedihardjo@erdc.dren.mil

1 Introduction

Trajectories are functions from a time domain—an interval on the real line—to $\mathbb{R}^d$ with $d > 1$, and observed as sequences of points sampled from them. A fundamental problem in analyzing this data is that of identifying common patterns between pairs or among groups of trajectories observed as sequences of sampled points.

Let $P = \langle p_1, \ldots, p_m \rangle$ and $Q = \langle q_1, \ldots, q_n \rangle$ be two sequences of points in $\mathbb{R}^d$, sampled from two trajectories γ_1 and γ_2 defined over the time interval $[0, 1]$. For simplicity, we assume that P and Q are points sampled from the images of the trajectories and we ignore the temporal component.[1] Since, in practice, the underlying continuous trajectories γ_1 and γ_2 are not known but we observe only the sampled points P and Q, we will work in the discrete setting where we are only concerned with these sample points. In this abstract, we will refer to the discrete sample points P, Q as the input trajectories.

We wish to compute correspondences between points belonging to similar portions of these trajectories while distinguishing these portions from the dissimilar ones. The following issues with trajectory sampling must be taken into account when identifying similarity: (i) significantly different sampling rates, (ii) presence of noise/outliers which must be distinguished from dissimilarities, and (iii) presence of significant unobserved portions on the trajectories with no sample points.

Background. A common choice for measuring trajectory similarity is the Fréchet distance [1] defined as follows. A *reparameterization* is a continuous non-decreasing surjection $\alpha : [0, 1] \to [0, 1]$, such that $\alpha(0) = 0$ and $\alpha(1) = 1$. The Fréchet distance $\mathrm{Fr}(\gamma_1, \gamma_2)$ is given by:

$$\mathrm{Fr}(\gamma_1, \gamma_2) = \inf_{\alpha,\beta} \max_{t \in [0,1]} \|\gamma_1(\alpha(t)) - \gamma_2(\beta(t))\|,$$

where $\|\cdot\|$ is the underlying norm (typically the Euclidean norm), and α and β are reparameterizations of $[0, 1]$. Since we only observe a finite set of sample points, we may define a discrete version of the Fréchet distance where the reparameterizations are discrete functions restricted to the sampled points P and Q.

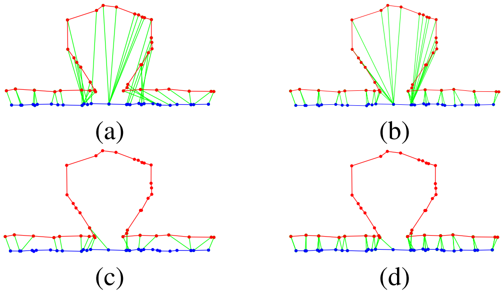


Figure 1. Comparison of measures: (a) Fréchet distance, (b) average Fréchet distance, (c) sequence alignment based method, (d) our model. Green edges indicate correspondences.

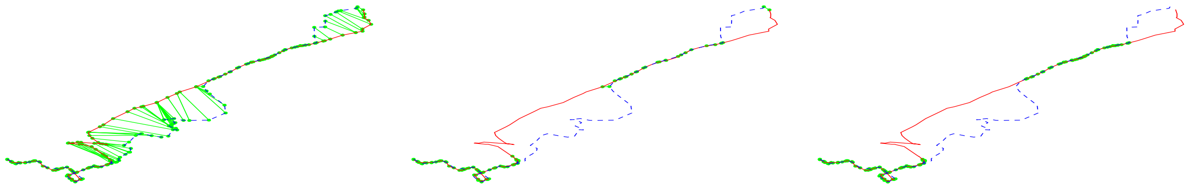
A set of correspondences yielding the optimal Fréchet distance is not necessarily a good indicator of similarity due to the large number of such correspondences; see Fig. 1(a). The *average Fréchet distance* which minimizes the average distance of the correspondences rather than the maximum distance provides a better set of correspondences. However, if there are significant dissimilar portions, possibly due to actual deviations rather than outliers, the results are not meaningful due to the requirement of correspondences for all points; see Fig. 1(b).

In computational biology, the technique of pairwise sequence alignment [2, cf. Chapter 2] is designed to distinguish similar and dissimilar portions between biological sequences. Given two sequences A and B, their alignment is expressed by writing them in two rows such that similar characters are placed in the same column. Characters in one sequence with no similar character in the other sequence are aligned with a blank character. A maximal contiguous sequence of blank characters is termed a *gap*. The goal is to optimize a scoring function which assigns a score for aligning two characters (incentive or penalty depending on their similarity) and a penalty for gaps.

We may extend the sequence-alignment model to the

---

[1] Strictly speaking, a trajectory is the graph of the underlying function, and what we have are the curves traced by the two trajectories, but we will not distinguish between the two.



(a) (b) (c)

Figure 2. Real data: (a) average Fréchet distance, (b) sequence alignment based approach, (c) our model. Green edges indicate correspondences.

alignment of trajectories with the choice of an appropriate scoring function. However, as Fig. 1(c) shows, non-uniform sampling rates cause similar portions to be designated as gaps since correspondences are one-to-one.

2 Model

As noted above, the average Fréchet distance yields good correspondences for similar portions even with different sampling rates while sequence alignment identifies dissimilarities accurately. We capture the advantages of the these two methods under a unifying notion of assignments.

Definition 2.1. An ***assignment*** for P and Q is a pair of functions $\alpha : P \to Q \cup \{\perp\}$ and $\beta : Q \to P \cup \{\perp\}$ for the points of P and Q respectively. If $\alpha(p_i) = \perp$ (or $\beta(q_j) = \perp$), then p_i (or q_j) is called a *gap point*. A maximal contiguous sequence of gap points in P or Q is called a *gap*. An assignment is ***monotone*** if it satisfies the following conditions: (i) if $\alpha(p_i) = q_j$ implies that for all $i' > i, \alpha(p_{i'}) \in \{\perp\} \cup \{q_{j+1}, \ldots, q_n\}$, (ii) $\beta(q_j) = p_i$ implies that for all $j' > j$, $\beta(q_{j'}) \in \{\perp\} \cup \{p_{i+1}, \ldots, p_m\}$.

Intuitively, if a point $p_i \in P$ lies on a similar portion of the two trajectories then $\alpha(p_i)$ defines the point on Q to which p_i corresponds, and p_i is a gap point otherwise. A similar interpretation holds for $\beta(\cdot)$. Let $\Gamma(\alpha, \beta)$ denote the set of gaps in P and Q for the assignment α, β. We define the ***score*** of α, β, denoted by $\sigma(P, Q; \alpha, \beta)$, as

$$\sigma(P, Q; \alpha, \beta) = \sum_{\substack{p_i \in P \\ \alpha(i) \neq \perp}} \frac{1}{c + \|p_i - \alpha(p_i)\|^2} + \sum_{\substack{q_j \in Q \\ \beta(j) \neq \perp}} \frac{1}{c + \|q_j - \beta(q_j)\|^2} + \sum_{g \in \Gamma(\alpha, \beta)} \Big(a + \Delta \cdot |g|\Big),$$

where a, Δ and c are carefully chosen parameters, $\|\cdot\|$ is the L_2-norm and $|g|$ is the length of a gap g. For a pair of points $p_i \in P$ and $q_j \in Q$, the difference in values $1/(c + \|p_i - q_j\|^2$ versus Δ dictates the choice of whether to assign $\alpha(p_i) = q_j$ or $\beta(q_j) = p_i$ versus assigning one or both as gap points. Thus, Δ is chosen based on a distance threshold for similarity. The parameter a is used to avoid extremely short gaps (of length less than l for some $l > 0$) which may be due to outliers rather than actual deviations and is set to $-l\Delta$.

A monotone assignment α, β which maximizes $\sigma(P, Q; \alpha, \beta)$ may be found by a dynamic programming algorithm in time $O(mn)$. This is essentially a more complicated version of the algorithm for sequence alignment. Fig. 1(d) shows the results which perform similarly to the average Fréchet correspondences in the similar portions while distinguishing the dissimilar portions as accurately as the sequence alignment based approach.

3 Discussion

Our framework is not limited to the scoring function described. For example, the sequence-alignment based approach, average Fréchet distance or other measures such as adaptations of edit-distance are easily incorporated into our model. Further, we may extend the dynamic programming algorithm to compute locally similar portions instead of global trajectory similarity in a manner similar to local alignment of sequences.

We have conducted experiments comparing the average Fréchet distance, sequence alignment and our model on a dataset of 145 trajectories of school buses in Athens, Greece [3]. Fig. 2 shows the comparison for a pair of trajectories from this set. As is clearly seen, sequence alignment "finds" the dissimilar portions accurately but the on close examination, we note that there are gaps even in the similar portions. This is rectified by our model which performs similarly to average Fréchet distance in the similar portions while avoiding dissimilar portions accurately.

References

[1] H. Alt and M. Godau. Computing the Fréchet Distance between Two Polygonal Curves. *Int. J. Comput. Geometry Appl.*, 5:75–91, 1995.

[2] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[3] E. Frentzos, K. Gratsias, N. Pelekis, and Y. Theodoridis. Nearest neighbor search on moving object trajectories. In *SSTD 2005*, pages 328–345. 2005.