There and Back Again: Using Fréchet-Distance Diagrams to Find Trajectory Turning Points

Lukas Beckmann, Benedikt Budig,*
Thomas C. van Dijk
Chair for Computer Science I
Universität Würzburg
{benedikt.budig,
thomas.van.dijk}@uni-wuerzburg.de

Johannes Schamel
Institute for Geography and Geology
Universität Würzburg
johannes.schamel@uni-wuerzburg.de

ABSTRACT

There is much interest in the automatic analysis of GPS trajectories, motivated amongst others by animal tracking, traffic modelling and tourism studies. Core questions involve for example detecting stopping points and flocking behaviour, and developing routing models. In this paper we consider turning points. We develop a formal definition of "turning point" based on analysing the structure of free-space diagrams for Fréchet distance. We give an efficient algorithm for computing an optimal set of turning points under this criterion.

Our method is evaluated in the context of a current study on tourist preferences at Berchtesgaden National Park, Germany. We evaluate the suitability of our definition and the efficiency of our algorithm on a large set of real-world GPS trajectories collected by outdoor recreationists. A ground truth of turning points was established by hand for the complete data set. Experiments show that the runtime and quality of our method are suitable for practical applications.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Geometrical problems and computations*; H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS*

General Terms

Algorithms, Experimentation, Theory

Keywords

GPS Trajectories, Fréchet Distance, Segmentation, Turning Points, Computational Geometry

*Corresponding author.

SIGSPATIAL '15, November 03 - 06, 2015, Bellevue, WA, USA

DOI: http://dx.doi.org/10.1145/2820783.2820878.

1. INTRODUCTION

GPS tracking is an important tool for tourism studies. In the field of outdoor recreation, it is used to investigate intra-destination movement patterns. Lew and McKercher [14], for instance, distinguish between point-to-point patterns (meaning that tourists visit an attraction and follow the same way back) and circular patterns (that do not feature such turning points). For this classification, it is important to determine turning points in GPS trajectories.

The identification of turning points was also an essential step in an ongoing study on the behavior of visitors to Berchtesgaden National Park [16]. In this study more than five hundred GPS trajectories were recorded and segmented at turning points for further analysis. In this paper, we introduce a novel method for the detection of turning points and use the data collected in the study for experimental evaluation.

Related work

The algorithmic analysis and segmentation of trajectories has been intensively studied in the past years. Segmentation algorithms have been based, for example, on finding *hotspots* of a certain shape [12], or based on so-called stable criteria [1], hereditary properties [4], monotone criteria [8] or nonmonotone criteria [3]. Van Kreveld and Luo [17] give algorithms to efficiently calculate the similarity of trajectories and subtrajectories, taking temporal aspects into account (in addition to geometric properties). The specific problem of identifying turning points has been attacked by Liu and Schneider [15], who segment trajectories using an angle criterion. A similar technique has been proposed by Chen et al. [9]. However, their method is completely local, which makes it susceptible to scale issues such as the sampling rate and the presence of noise. Our algorithm provides robustness in these settings by making a more global analysis.

The concept of free-space diagrams for of Fréchet distance is due to Alt and Godau [2]. Several groups of authors intersecting on Buchin & Buchin have applied this concept to the analysis of GPS trajectories in various ways. They propose a measure for partial similarity of curves [7]. The free-space diagrams are used to find subtrajectories that are near each other and to match vertices. They introduce *constrained free-space diagrams* [5] as a tool to perform various other analyses, proposing restrictions on the free space and what paths are considered admissible. A sweepline algorithm [6] is used to cluster similar parts in one or many trajectories.

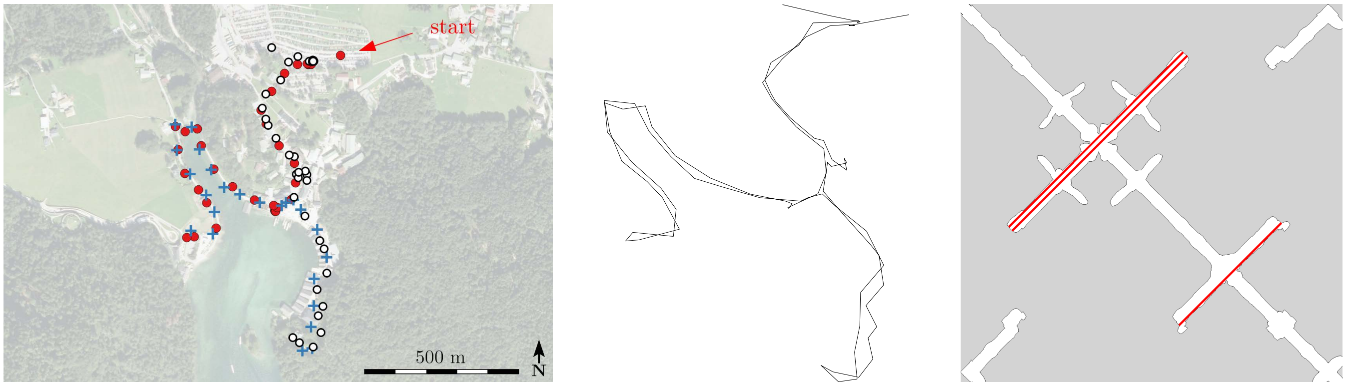

Figure 1: Three views of a trajectory from our data set. Left: GPS points as segmented into three parts by the ground truth, overlaid on satellite imagery.[1] Middle: The entire trajectory as a single polyline. Right: The resulting free-space diagram. The tourist starts at a parking lot in the north and goes west of the lake (●). The tourist turns around, doubles back and then goes east of the lake (+). A second turning point occurs and the tourist returns to the parking lot (○).

2. HOLES IN FRÉCHET DIAGRAMS

In this paper, we interchangeably interpret GPS trajectories as polylines and vice versa. We now consider the free-space diagrams introduced by Alt and Godau [2] for computing the Fréchet distance between two (directed) polylines. In their paper, they search for an admissible path through this particular free-space diagram and prove that such a path exists if and only if the two input polylines have Fréchet distance at most ε. See the original paper for details. When just looking for the existence of such a path, the exact parameterisation of parameter space is irrelevant. In this paper, we parameterise by the length of the line segments.

These free-space diagrams (henceforth: *Fréchet diagrams*) contain considerable information about the relative position and shape of the polylines. Recall that a point f in a Fréchet diagram for polylines P and Q corresponds to a point on P and a point on Q, and that f is free if and only if these two points have distance at most ε. We will find that interesting information can be found even for *one* polyline T by comparing T to its reverse T_r. Since T equals T_r except in reverse, a point in the Fréchet diagram corresponds to *two* points on T. See Fig. 1 for an example.

A visually prominent feature of such diagrams is the strip of free space connecting the upper left and the lower right corner along the diagonal: see Fig. 2. In particular, for points *on* the diagonal from the upper left to the lower right corner, the point on T and on T_r necessarily coincide. Call this the *main diagonal* of the diagram. It also follows that ε is a lower bound on the width of this strip in parameter space: Any point on T is free when compared to any other point that lies within distance ε *along* T: by the triangle inequality, the Euclidean distance between these two points is at most ε. Then a line segment of slope 1, placed symmetrically about the main diagonal, must be completely free if it has length at most $\sqrt{2}\varepsilon$. We call this the *basic length* since it will serve as unit of scale later. Note that the Fréchet diagram for T and T_r has a mirror symmetry along the main diagonal.

Fig. 2a shows a trajectory that passes close to itself, but in reverse. The trajectory contains two subtrajectories that are parallel and next to each other. This corresponds to a hole in the Fréchet diagram as indicated by the red line. The line has positive slope, but does not intersect the main diagonal. This means that the corresponding intervals on T and T_r do not overlap. Parallel subtrajectories that are traversed in the same direction lead to a similar hole, but with negative slope. Fig. 2b shows the same situation, except that one of the two subtrajectories is substantially longer than the other. This is expressed by the different slope of the corresponding red line segment. Fig. 2c shows a trajectory that contains a hotspot and a corresponding large disc in free space.

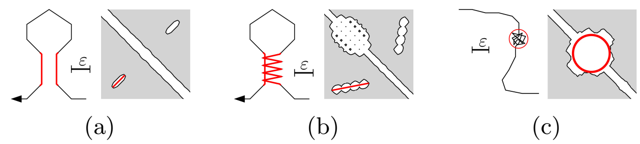

Figure 2: Different shapes of holes in Fréchet diagrams and their corresponding subtrajectories.

Now we look at Fig. 1 again. The Fréchet diagram on the right includes several maximally-long slope-1 line segments that cross the main diagonal. By our earlier observations, we see that these correspond to long nearly-parallel subtrajectories that are traversed in reverse direction. We call such line segments *turning intervals* and predict a turning point at their intersection point with the main diagonal. Note that this is accurate in Fig. 1.

3. CLASSIFYING TURNING POINTS

Based on the discussion above, we are interested in long, slope-1 line segments in the free space that cross the main diagonal. We call a line segment *maximal* if it cannot be moved within the free space. Given two polylines of n and m vertices, the set of maximal line segments of given slope in the Fréchet diagram can be enumerated in $\mathcal{O}(nm)$ time (algorithm omitted for space). This set includes many irrelevant turning points: not all maximal line segments correspond to actual turning intervals. First we remove all line segments that do not intersect the main diagonal. Short line segments

[1]Imagery: DigitalGlobe, GeoBasis-DE/BKG, GeoContent, Geomimage Austria, Salzburg AG / Wenger Oehn.

might correspond to small bends in the trajectory that do not actually represent turning points from a semantic point of view. We use a discrimination threshold τ that only passes long line segments, where τ is a factor on the basic length (see Section 2): a line segment of length 2ℓ will be selected if and only if $\ell \geq \tau \cdot \sqrt{2} \cdot \varepsilon$. (Note that τ is a dimensionless constant.) Appropriate values for τ in real-world data are discussed in the next section.

4. CASE STUDY

In this section we apply our algorithm to a set of real-world data and show that it is suitable for practical applications. As part of a research project in Berchtesgaden National Park, 637 GPS trajectories tracking outdoor activities have been collected by Schamel [16]. The trajectories in this data set have a high sampling rate (1 Hz) and an average length of 9.7 km. They were manually segmented by a geography student temporarily employed for this purpose, splitting at turning points and when the mode of travel changes (the latter being determined from external information). This manual process took approximately 50 hours.

In this section we compare the results of our algorithm to this manually-generated *ground truth*. Since our algorithm knows nothing about travel-modes, we focus on the subset of trajectories that were only cut at turning points. We have run our experiments on 284 such trajectories, with an average of 1.51 ground-truth turning points (median 1). Note that we do not expect to replicate the ground truth exactly, since the concept of turning point is vague to begin with, and the student was instructed to omit "very short" segments (without formal criterion).

We have used only every 150$^{\text{th}}$ point of each trajectory. This results in an effective sampling rate of once per two and a half minutes, which is reasonable for the activities in this case study (mostly hiking). This reduced sampling rate represents lower quality data that might be more easily obtainable in practice (e.g. the energy-saving mode on consumer GPS devices). Even on this relatively crude data, our algorithm performs well.

We have implemented our algorithm in Java. No effort was made to optimise this implementation for runtime; significantly improved constant factors are surely possible. On a representative trajectory of 100 points we measured a runtime of 0.24 s on commodity hardware; $n \approx 600$ took 2.7 s.

4.1 Statistical Accuracy

We have run our algorithm on all 284 trajectories and used $\varepsilon = 0.003°$ as value for the Fréchet distance calculations. This corresponds to about 333 m in this area and was picked by hand based on visual inspection of some of the trajectories. Given a discrimination threshold τ, we compare the position of the turning points according to the ground truth and the detected turning *intervals*. If a detected interval contains at least one ground-truth turning point, it is considered a true positive; otherwise, a false positive. The ground-truth turning points that were not contained in any detected interval are considered false negatives. Based on these classes, we evaluated our algorithm using ROC analysis [10] and calculate the *area under the curve* (AUC). In this experiment we find an AUC of 0.914, which is considered excellent [13]. This shows that the maximal line segments in Fréchet diagrams are a successful data reduction step for finding turning intervals in GPS trajectories and that the length of the line segment is a meaningful feature for classifying the hits.

Table 1: Classification performance depending on discrimination threshold τ.

τ	Recall	Precision	F_1 Score	Distance (median)
1.13	0.94	0.73	0.82	50.2 m
1.55	0.78	0.85	0.82	55.0 m
4.12	0.50	0.99	0.66	63.2 m

The ROC analysis aggregates over the discrimination threshold τ. In practice, a particular value for τ must be used. Inspection of the ROC curve suggests 4.12, 1.55 and 1.13 as promising values for τ. Table 1 shows the statistics. Note that $\tau = 1.13$ and $\tau = 1.55$ lead to the same F_1 score, where the lower threshold leads to an increased recall at the cost of some precision. Choosing $\tau = 4.12$ leads to a lower F_1 score, but ensures a very high precision (99 %). This might be desirable depending on the application.

4.2 Geographic Accuracy

Now we consider the (geographic) distance between our predicted turning points and their counterparts in the ground truth. Table 1 gives the median distance, which lies between 50.2 m and 63.2 m depending on τ. This can be considered very precise, since the median distance between subsequent points in the trajectories is 91.2 m. Specifically for $\tau = 1.55$, more than a third of the ground-truth turning points have a predicted turning point within 25 m.

When considering Table 1, it is noteworthy that the distance value actually *improves* for lower thresholds, that is, more loose classification. Recall that our algorithm initially detects *intervals* that it predicts contain a turning point, and that the exact locations of these turning points is only estimated afterwards. This means that long intervals are fairly certain to include a turning point (improved precision), but allow many possible geographic locations of the actual turning point (increased geometric error). Shorter intervals, while being less reliable indicators, provide tighter constraints on where the turning point is located geographically. For this reason, admitting shorter intervals decreases statistical precision, but improves the geographic precision.

4.3 Influence of Scale

Recall that ε can be interpreted as a Fréchet distance that must be admissible within a detected turning interval. This shows that ε influences the *scale* of the turning intervals our algorithm locates. Consider Fig. 3. With ε_1, the resulting Fréchet diagram contains only one hole that allows a slope-1 line segment of significant length. The resulting turning interval captures turning point A well and ignores point B altogether. Depending on the context, the scale, and the size of turning areas one is interested in, this might well be the desired result. If we pick ε_2 instead, we increase the scale of allowed turning areas. In the Fréchet diagram for ε_2, we see an additional distinct hole that will contain long maximal line segments. The turning interval described by this hole corresponds to turning point B. This experiment shows that our algorithm is capable of locating turning intervals at different scales, intuitively controllable by setting ε: it is a natural parameter for which a reasonable value can be picked

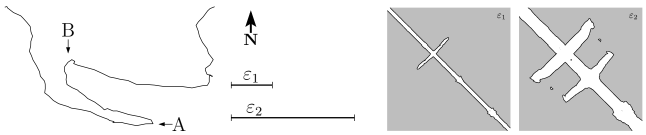

Figure 3: A crop of a trajectory from our data set. The value of ε in the Fréchet-distance calculation influences the scale at which turning points are detected.

based on the input data and the intended scale of turning points. Thus, it can also be applied to trajectories that contain turning behaviour on considerably larger or smaller scales than the hiking trajectories considered in this paper.

5. CONCLUSION

We have introduced a new characterisation of turning points in trajectories, based on free-space diagrams for Fréchet distance. In particular, we interpret the trajectory as a polyline and compare this polyline with its reverse. If intervals of these two polylines are Fréchet-similar, then those intervals represent a part of the trajectory that is similar in both directions. This directly relates to turning points and we have shown how to ‘read’ this from the holes of the free-space diagram. We chose to find long slope-1 line segments within the free space. While this may seem overconstrained and somewhat arbitrary, it has a clear interpretation and works well in our case study. We provide a classifier to actually detect the turning points from this set. An advantage of our method over previous work is that it takes the entire trajectory into account and has a natural scale parameter (ε).

In this paper we have focused specifically on our characterisation of turning points—to the exclusion of possible additional features such as local turning angles [15] and kernel density approaches. Additional information may come from a point-of-interest data set, or map matching when a road network is available. A practical system will want to combine multiple features in order to arrive at better classification performance (for example through classifier committees); here we have demonstrated that this particular feature already works well on its own.

Further work may want to investigate if other semantically meaningful properties of trajectories can be derived from the Fréchet diagram. A clear extension of the current work is a more general analysis of parallel subtrajectories: What can we detect that is relevant for the various sciences dealing with GPS trajectories, such as tourism studies and biology? The inclusion of turning points could improve segmentation of trajectories in tourism, which by now is for the most part based on detecting stop locations [11, 18]. In our case study, we have detected slope-1 line segments. It may be fruitful to generalise this to other shapes within the free space. (Compare how Buchin et al. [6] detect single-file movement in sets of trajectories.)

Acknowledgements

The research in this paper was partially supported by the German Research Foundation (DFG grant GZ JO 388/9-1). The free-space diagrams in this paper were generated using the *freespace* Ipelet by Günter Rote.

6. REFERENCES

[1] S. P. A. Alewijnse, K. Buchin, M. Buchin, A. Kölzsch, H. Kruckenberg, and M. A. Westenberg. A Framework for Trajectory Segmentation by Stable Criteria. In *Proc. ACMGIS’14*, pages 351–360, 2014.

[2] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *IJCGA*, 5(1–2):75–91, 1995.

[3] B. Aronov, A. Driemel, M. van Kreveld, M. Löffler, and F. Staals. Segmentation of Trajectories on Non-Monotone Criteria. In *Proc. SODA’13*, pages 1897–1911, 2013.

[4] D. Bokal, S. Cabello, and D. Eppstein. Finding All Maximal Subsequences with Hereditary Properties. In *SoCG’15*, pages 240–254, 2015.

[5] K. Buchin, M. Buchin, and J. Gudmundsson. Constrained free space diagrams: a tool for trajectory analysis. *IJGIS*, 24(7):1101–1125, 2010.

[6] K. Buchin, M. Buchin, J. Gudmundsson, M. Löffler, and J. Luo. Detecting Commuting Patterns by Clustering Subtrajectories. *IJCGA*, 21(03):253–282, 2011.

[7] K. Buchin, M. Buchin, and Y. Wang. Exact Algorithms for Partial Curve Matching via the Fréchet Distance. In *Proc. SODA’09*, pages 645–654, 2009.

[8] M. Buchin, A. Driemel, M. van Kreveld, and V. Sacristán. An Algorithmic Framework for Segmenting Trajectories Based on Spatio-temporal Criteria. In *Proc. ACMGIS’10*, pages 202–211.

[9] Y. Chen, K. Jiang, Y. Zheng, C. Li, and N. Yu. Trajectory Simplification Method for Location-based Social Networking Services. In *Proc. LBSN’09*, pages 33–40, 2009.

[10] T. Fawcett. An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[11] A. Y. Grinberger, N. Shoval, and B. McKercher. Typologies of tourists’ time-space consumption: a new approach using GPS data and GIS tools. *Tourism Geographies*, 16(1):105–123, 2014.

[12] J. Gudmundsson, M. van Kreveld, and F. Staals. Algorithms for Hotspot Computation on Trajectory Data. In *Proc. ACMGIS’13*, pages 134–143, 2013.

[13] D. W. Hosmer Jr and S. Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, 2004.

[14] A. Lew and B. McKercher. Modeling Tourist Movements: A Local Destination Analysis. *Annals of Tourism Research*, 33(2):403–423, 2006.

[15] H. Liu and M. Schneider. Similarity Measurement of Moving Object Trajectories. In *Proc. IWGS’12*, pages 19–22, 2012.

[16] J. Schamel. Ableitung von Präferenzen aus GPS-Trajektorien bei landschaftsbezogenen Erholungsaktivitäten. In *AGIT – Journal für Angewandte Geoinformatik*, pages 56–64, 2015.

[17] M. van Kreveld and J. Luo. The Definition and Computation of Trajectory and Subtrajectory Similarity. In *Proc. ACMGIS’07*, pages 44:1–44:4, 2007.

[18] H. Xiao-Ting and W. Bi-Hu. Intra-attraction Tourist Spatial-Temporal Behaviour Patterns. *Tourism Geographies*, 14(4):625–645, 2012.