Earth Observation and Geomatics Engineering 4(1) (2020) 56–65

EARTH OBSERVATION AND GEOMATICS ENGINEERING

website: https://eoge.ut.ac.ir

Inferring geometric similarities of trajectories by an abstract trajectory descriptor

Amin Hosseinpoor Milaghardan[1], Rahim Ali Abbaspour[1*], Christophe Claramunt[2], [3]Alireza Chehreghan

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Iran
[2] Naval Academy Research Institute, Lanveoc-Poulmic, France
[3] Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran

Article history:
Received: 8 January 2020, Received in revised form: 28 April 2020, Accepted: 4 May 2020

ABSTRACT

Investigating the geometric similarity of trajectory data to extract movement patterns in urban environments is an emerging area of research that has attracted several efforts over the past few years. This paper uses a convex-hull algorithm whose objective is first to identify curvatures and turning points in a given trajectory and secondly to provide a computable solution to identify the similarities of trajectories. Moreover, the present paper tries to detect additional capabilities that will support the exploration of regular patterns efficiently. This approach is supported by a series of geometrical definitions and algorithms that reduce the complexity of primary trajectories significantly and identify a trajectory geometrical decomposition modeled by an abstract trajectory descriptor (ATD). The main novelty of this paper is to find out the similarity between the row trajectory's geometry and the results of the ATD method using the known geometric measures as distance, orientation, complexity, and shape. Based on this decomposition principle, trajectory similarities can be studied using physical, geometrical, or both descriptors as considered in the ATD method. The proposed method has been evaluated using Geolife benchmark trajectory database. the results show that the proposed algorithm not only successfully identify curvatures and turning points at different scales, but also proved to provide relevant trajectory similarities with efficient computation times as the overall similarity difference value equals 0.002 between the two row trajectories. The resulted trajectories using the ATD method applying less than 5% of the primary points. In addition, the computation time of about 93% is reduced using the detected critical geometric points using the ATD method.

KEYWORDS

Trajectories
Geometric similarity
Curvature
Turning point
convex-hull

1. Introduction

Nowadays, trajectory data in urban environments has been widely used; taking into account the rapid development of positioning devices, mobile technologies, and user-oriented applications. Several available benchmark trajectory databases use the development of novel applications and cross-comparisons between different research works. Amongst the previous researches, the analysis of the similarities between trajectories is more considerable. This fact might allow us to derive regular patterns and then to provide a better view and potentially to understand of the human behaviors that arise in the large urban environment, this undoubtedly being an asset for many urban planning and transportation behavioral studies. Specifically, it might be useful to examine intra-urban travel patterns and derive some possible scenarios for planning further transportation network developments (Yuan, Qian et al., 2014, Yue, Lan et al., 2014). Additionally, such movement patterns could be beneficial for extracting space-time accessibilities, cost, and travel time between different places in the city (LIN, LV et al. 2014). A movement pattern in the city can be regarded as composed of two complementary geometric and semantic dimensions. While the geometric dimension should characterize a given trajectory according to its origin, destination, and its intrinsic geometrical properties (Widhalm, Yang et al. 2015). While the semantic patterns are crucial when examining the behavioral components of a

* Corresponding author
E-mail addresses: amin_hoseinpoor@ut.ac.ir (A. Hosseinpoor); abaspour@ut.ac.ir (R.Ali Abbaspour); christophe.claramunt@ecole-navale.fr (Ch. Claramunt); Chehreghan@sut.ac.ir (A. Chehreghan)

DOI: 10.22059/eoge.2020.299971.1080

given movement in the city (Lipan & Groza 2010).

When exploring the similarity between different trajectories, choosing the right parameters to represent an important preliminary step. In related work, Buchin et al. (Buchin, et al. 2008) selected several geometric and semantic criteria as well as some context-aware criteria. However, the temporal component has been unconsidered so far, and a given trajectory is considered a linear sequence of points (Yuan & Raubal 2014). Using a primary geometrical approach, two trajectories can be cross-compared using a series of Euclidean, Hausdorff, or Fréchet distances (Demšar, Buchin et al. 2015). However, other geometrical properties as exhibited by curvature and turning points are unconsidered. Another category of approaches considers some additional geometrical or second order parameters such as direction, speed, and acceleration (Soleymani, Cachat et al. 2014, Demšar, Buchin et al. 2015). However, these parameters can be operated to characterize some movement patterns (Dodge, Weibel et al. 2011). Finally, contextual properties like the environmental conditions (e.g., temperature and pressure) can be considered as external parameters worth being related to some trajectory and movement patterns (Buchin, Dodge et al. 2012).

Several methods and computational approaches have been used and implemented to explore the similarity between trajectories in urban environments. For instance, schematic representations of the origin and destination of some trajectories provide some effective means when studying origin destination behaviors in the city (Lu, Wang et al. 2015) and their distribution in space and time at various levels of granularity from intra-city to city-to-city transportation patterns (Jiang, Zheng et al. 2015, Tang, Liu et al. 2015). When considering large sets of trajectories, a generalization should derive the main geometric characteristics of some trajectories. This provides more opportunities to study the correlation between different trajectory categories (Cao, Mamoulis et al. 2005). The temporal dimension can equally be considered as the essential criteria to analyze, particularly when a given trajectory is decomposed according to some given points of interest (Giannotti & Pedreschi 2008). Overall, trajectory parameters can be visualized and different evaluated according to different human profile behaviors (Dodge, Weibel et al. 2009). In another study, Soleymani, Cachat et al. (2014) employed the maximum, minimum, and mean values of the speed, rotation angle, acceleration when comparing two given trajectories. Another group of research has applied some clustering techniques to identify similar trajectories. For example, Morzy (2007) applied a swarm optimization and dynamic time warming (DTW) algorithm to cluster a series of trajectories. In related studies, a Fuzzy c-means method (FCM) (Parent, Spaccapietra et al. 2013) and a support vector machine (SVM) (Lee, Han et al. 2011) were applied for clustering trajectories. As a result, there are variety analysis methods depending on the application goals as the physical, geometrical, or both of them could be utilized in the mentioned works in the analysis.

The approach presented in this paper is unique in several aspects. First, since most of the above contributions consider all surveyed points of trajectories into the analysis while most of these points are unnecessary and just increase the complexity of methods. Hence, the proposed method at first introduces a data-filtering algorithm whose objective is to keep the most relevant geometrical trajectory points to improve computation times, this being a key issue when considering huge data sets. Next, the peculiarity of our algorithmic approach is to consider a different point of view and assumption, giving specific importance and role of the convex-hull geometrical structure as well as a series of additional trajectory parameters such as the shape, the turning points, the direction, and the short-line median Hausdorff distances. In our previous work, we introduced a geometrical and semantic framework to identify the main physical and geometrical critical points of a given trajectory (LIN, LV et al. 2014, Yue, Lan et al. 2014). The distribution of such critical points has been explored using an entropy-based approach (Widhalm, Yang et al. 2015). In the present paper, trajectories similarities are further studied applying the primitive distribution of the critical points of the respective convex-hull s of these trajectories. In addition, this paper seeks to evaluate the efficiency of the ATD method in preserving the geometric specifications of trajectories using geometric descriptors of distance, orientation, complexity, and shape. Therefore, the semantic or topological relations between the trajectories are unconsidered in this paper.

The remaining of the paper is organized as follows. The next section introduces the main principles of our proposed approach. Section 3 presents the experimental implementation and discusses the results are presented. Finally, Section 4 concludes the paper and provides a few directions for further work.

2. Proposed Methodology

Figure 1 shows the framework of the proposed method, which consists of three main steps. The first step involves pre-processing the trajectory data. At the pre-processing stage, the points with low spatial accuracy are withdrawn from the trajectory sequence (e.g., trajectory points with the weak GPS signal, receiver technical problems). In many cases, map matching processes can eliminate these positioning errors. However, in the remaining cases, and when the error value is beyond an acceptable threshold, other methods should be adopted to eliminate these points or match them to the trajectory point sequence appropriately (Zheng 2015). To process such error-prone remaining points, a Kalman filter has been applied. Application of this Kalman filter represents a recursive process that provides a linear fit based on the positions of the input trajectory points; an average standard deviation (σ) between two consecutive points is first calculated, then trajectory points with a distance higher than (twice as much as σ) the trajectory fitting line are eliminated from trajectory points. Next the turning and curvature critical points of the two considered trajectories i and j are detected using the Convex-hull structure described by (Hosseinpoor Milaghardan, Ali Abbaspour et al. 2018). Finally, the spatial similarity degree is calculated using the considered geometric parameters including distance, orientation, complexity, and shape.

Many of the trajectories are made up of a relatively significant number of points. These challenges the assessment of the spatial similarity of each of the trajectories, as well as computation time, are likely to be heavy. This leads us to apply an Abstract Trajectory Descriptor (ATD) approach in which each trajectory is decomposed into a sequence of significant points derived from curvature and turning points. For example, in Figure 2a, consider the trajectory T_i, which consists of 416 points. The significant points of this trajectory were extracted (Figure 2b) and finally, the T_i' graph was derived and provided the representative generalization of this trajectory with 47 significant points to further consider when evaluating spatial similarities with other trajectories (Figure 2c).

Let us consider the spatial characteristics and graphs extracted for each of the trajectories. The similarities are derived from four geometric criteria including the shape, direction, turning point, and Hausdorff distances as described in the following sections. The following is a detailed description of first the trajectory ATD extraction and then similarity evaluations according to the different trajectory geometrical parameters identified so far.

2.1. Trajectory ATD Extraction

To identify the significant points of the trajectory according to the principles introduced in the previous section, a convex-hull geometric structure and a search for turning points are computationally implemented.

Figure 3 illustrates the position of the curvature and turning points for an example of trajectory. One of the critical aspects and the role of these significant trajectory points is to entail the spatial distances and temporal intervals between them. Indeed, this can be applied for further analysis of the typical characteristics embedded in a given trajectory as well as providing a few insights to analyze the underlying behaviors to identify the most minimal possible number of these points a convex hull algorithm is applied to extract the minimum number of critical points that can be used to show all geometric properties of the trajectory including shape, complexity, direction and distance. The role of these parameters is to embed comprehensive information on the underlying spatial and temporal properties that can be used for further analysis and trajectory pattern detection.

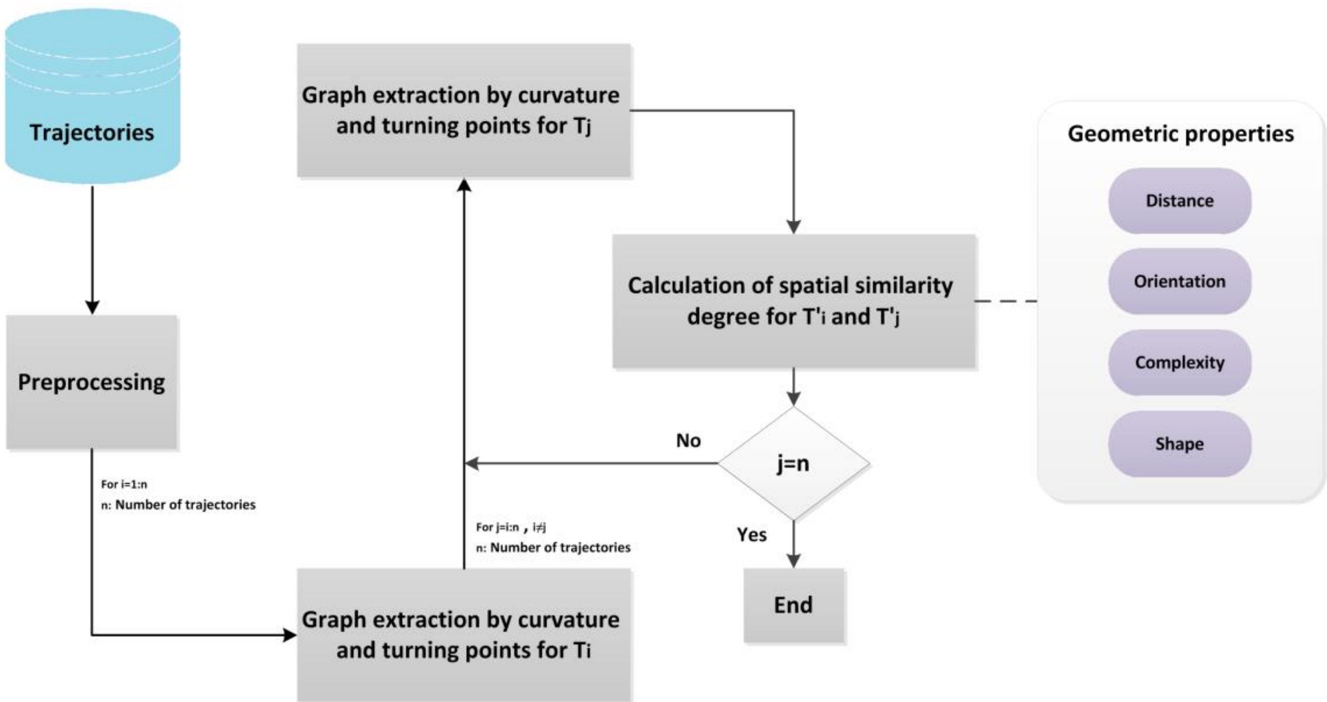

Figure 1. Methodological flowchart.

The convex-hull structure is one of the most commonly used structures in geometric computations based on the concept of convexity, which may be used directly or as a tool for constructing other structures. Figure 4 illustrates the computational process applied to extract curvatures and turning points. In this figure, n_i is the number of nodes in trajectory Ti, and j is the numerator for vertices of a convex polygon.

Considering the framework presented in Figure 4, the computational process begins from the first point of the trajectory while the largest convex polygons are formed along the trajectory as the convex-hull geometric structures.

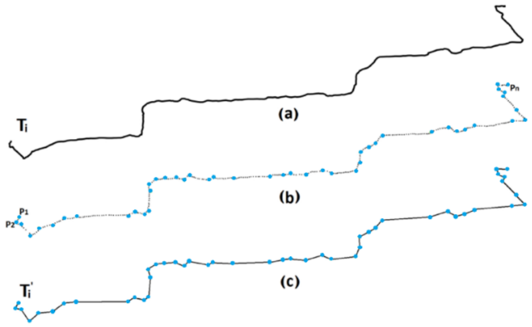

Figure 2. A Sample trajectory with the extracted significant points.

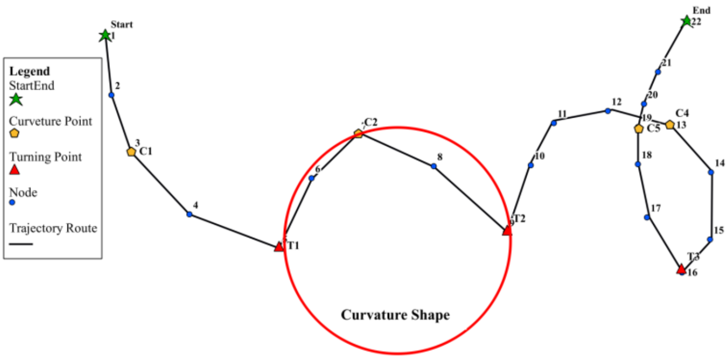

Figure 3. A trajectory along with significant curvature and turning points.

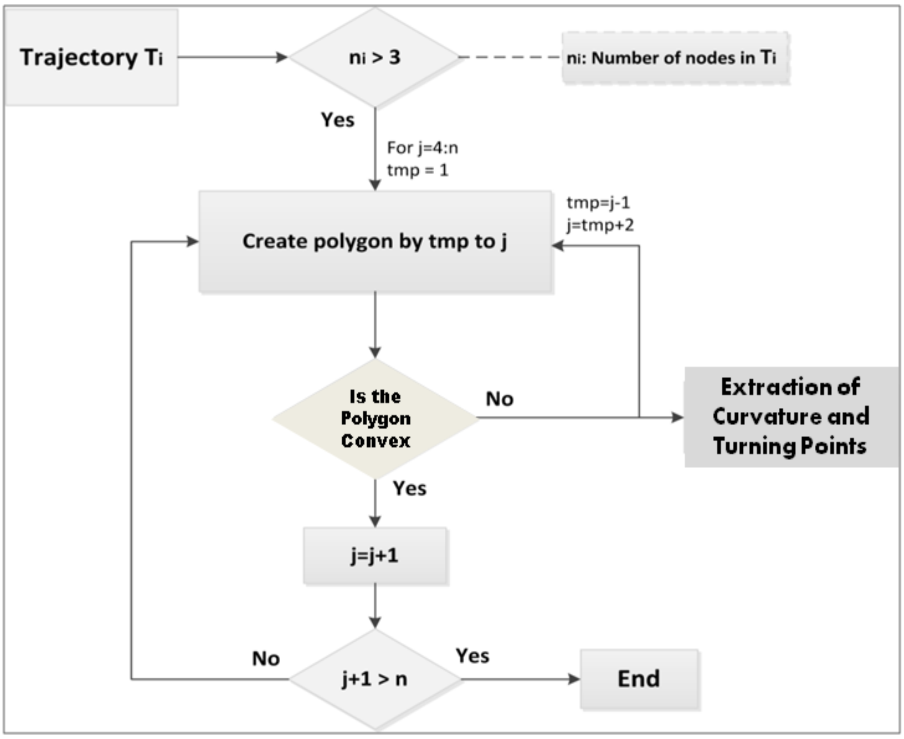

Figure 4. The framework applied to identify significant points.

Broadly, the order of the points to form this structure does not matter when defining the convex-hull structure. Nonetheless, in the approach suggested, the ordering of the points is considered when forming the convex-hull structure imposed the trajectory points successively considered. Figure 5 illustrates the application of the convex-hull structure to an example of a trajectory. Considering the structure of the trajectory data, which can be defined as a time series of trajectory positions, the implementation algorithm of the proposed method is incremental.

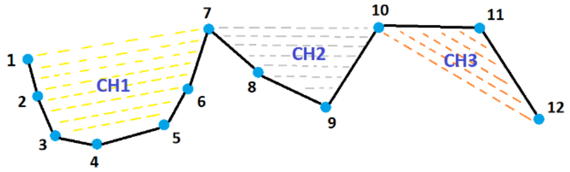

Figure 5. A trajectory with convex-hull structures.

After forming the convex-hull structure, the significant points of the curvature and turning are extracted. Turning points are indeed the points of the beginning and the ending points of a given curvature, and at least one point from the trajectory points set must be between these two turning points. In Figure 6, turning points are identified from a trajectory example. In particular, Figure 6 exhibits two points of sequential turning change, while only one curvature can lie between these two points. Moreover, the two points and the curvature points between them are illustrated by the convex-hull. The generation of the convex-hull s in a trajectory is processed in sequence, and in fact, the end point of convex-hull i and the convex-hull are considered the starting points for the convex-hull (i+1).

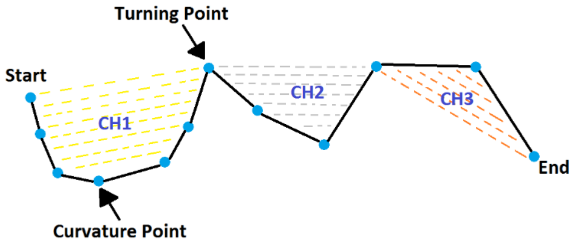

Figure 6. Turning points in a trajectory

More formally, Turning and curvature points are of a given convex-hull are defined as follows.

Definition 2 – Turning point: A point T(ti) materializes a turning point for each i=1,2,….,n from a given trajectory n if and only if:

1- Eqs. (1) and (2) are true for two values of j=i+1 and s<i-1.

$$\left|\mathrm{dist}\left[\overline{T(t_i)T(t_{i-1})}, T(t_j)\right] > 0\right| \quad (1)$$

For $s<i \rightarrow \Delta t_{i,s}$ is minimum and $\mathrm{dist}\left[\overline{T(t_i)T(t_{i-1})}, T(t_s)\right] \neq 0$ (2)

2- The signs of the values extracted from Eq. (3) are the same and opposite to the sign of Eq. (4) for the two values j and s.

$$\begin{vmatrix} x_{T(t_i)} & y_{T(t_i)} & 1 \\ x_{T(t_{i-1})} & y_{T(t_{i-1})} & 1 \\ x_{T(t_s)} & y_{T(t_s)} & 1 \end{vmatrix} \neq 0; \begin{vmatrix} x_{T(t_{s+1})} & y_{T(t_{s+1})} & 1 \\ x_{T(t_s)} & y_{T(t_s)} & 1 \\ x_{T(t_{s-1})} & y_{T(t_s)} & 1 \end{vmatrix} \neq 0 \quad (3)$$

$$\begin{vmatrix} x_{T(t_i)} & y_{T(t_i)} & 1 \\ x_{T(t_{i-1})} & y_{T(t_{i-1})} & 1 \\ x_{T(t_j)} & y_{T(t_j)} & 1 \end{vmatrix} \neq 0 \quad (4)$$

where $x_{T(t_i)}$ and $y_{T(t_i)}$ are the coordinates of the point $T(t_i)$ and the same for the points s and j. Moreover, $\mathrm{dist}\left[\overline{T(t_i)T(t_{i-1})}, T(t_j)\right]$ gives the distance between the connecting line of $T(t_i)$ and $T(t_{i-1})$ to the point $T(t_j)$. Additionally, $\Delta t_{i,s}$ is the time difference between the points i and s. Next, a curvature point is detected according to definition 3 as follows.

Definition 3 - curvature point: For each point, T(ti) and T(tj) respectively denote the start and the end points of a convex-hull (Hosseinpoor Milaghardan, Ali Abbaspour et al. 2018), while T(tr) denotes the curvature point if and only if the Eq. (5) is true;

$$\left|\mathrm{dist}\left[\overline{T(t_i)T(t_j)}, T(t_r)\right] \text{ is Max}\right| \quad (5)$$

After identifying the significant points of the curvature and turning points, the trajectory ATD is compiled. All the geometric parameters listed above are detected and stored in a series of Abstract Trajectory Descriptor (ATD) representations that also take into account the starting and ending points as well as the critical points identified by the geometrical parameters. The nodes of this ATD contain the significant points of the reference trajectory, distances and temporal intervals between successive points are also stored as contextual information.

2.2. Spatial similarity

Based on Eq. (1) and the trajectory ATD extracted in Section 2.1, for each trajectory, a degree of spatial similarity can be derived. Let us initially describe the geometric properties used which are distance, direction, turning point, and shape (Chehreghan & Ali Abbaspour 2017).

2.2.1. Short-line median Hausdorff

Measuring the distance between two trajectories is one of the most primary geometric criteria to check when evaluating the similarity between two trajectories. In many studies, the Euclidean, Hausdorff, DTW, and Fréchet distances have been used to evaluate some degrees of similarity. The

Hausdorff distance (Buchin, Buchin et al. 2008) and the Short-line Median Hausdorff introduced by Tong, Liang et al. (2014) have been used largely in previous trajectory analysis studies. Unlike the other Hausdorff distances such as the median Hausdorff, these types of distances are appropriate against longitudinal anomalies (Tong, Liang et al. 2014, Chehreghan & Ali Abbaspour 2018). Considering the anomalies and turning points in the trajectory data, applying an optimal criterion is mandatory. Eq. (6) gives the Short-line Median Hausdorff distance (Tong, Liang et al. 2014) based on the length between two given trajectories T_i and T_j:

$$C_1(T_i,T_j)=\begin{cases} m(T_i,T_j)\ ;\ \text{if}\ \ \text{length}(T_i)<\text{length}(T_j) \\ m(T_j,T_i)\ ;\ \text{if}\ \ \text{length}(T_i)\geq\text{length}(T_j)\end{cases} \quad (6)$$

where the $\text{length}(T_i)$ and $\text{length}(T_j)$ denote the lengths of the two trajectories T_i and T_j, respectively, and where $m(T_i,T_j)$ and $m(T_j,T_i)$ are derived by Eqs. (7) and (8):

$$m(T_i,T_j)=\text{median}_{P_a\in T_i}\left\{\min_{P_b\in T_j}\left\|P_a-L_b\right\|\right\} \quad (7)$$

$$m(T_j,T_i)=\text{median}_{P_b\in T_j}\left\{\min_{P_a\in T_i}\left\|P_b-L_a\right\|\right\} \quad (8)$$

where,

- L_a and L_b are two optional parts of the trajectories T_i and T_j.
- P_a is a point from the trajectory T_j.
- $\|P_a-L_b\|$ is the vertical distance between one point on the $T_i(P_a)$ and one of the parts of T_j (L_b).
- $\|P_b-L_a\|$ is the vertical distance between one point on the $T_j(P_b)$ and one of the parts of T_i (L_a).

2.2.2. Direction

Another geometric criterion that can be used to check the geometric similarity of two given trajectories is direction. For a given trajectory, the direction can be defined as the angle between the virtual line of the starting point to the endpoint of the trajectory and the horizontal axis. Accordingly, for two trajectories having directions α and β, the direction difference is equal to | α-β | as shown in Eq. (9) (Tong, Liang et al. 2014). In this equation | α-β | is between 0 to π. If the angle difference is about 0 radians, the two trajectories are nearly parallel, if the value of the angle difference is close to π, trajectories are parallel but consider opposite directions; if the angle approximates $\frac{\pi}{2}$, then these two trajectories are perpendicular (Chehreghan & Ali Abbaspour 2017).

$$C_2(T_i,T_j)=\alpha-\beta=\cos^{-1}\left(\frac{\vec{V}_{T_i}.\vec{V}_{T_j}}{\left\|\vec{V}_{T_i}\right\|.\left\|\vec{V}_{T_j}\right\|}\right) \quad (9)$$

where

- $\vec{V}_{T_i}$ is the vector made up of the first point to the last point of the first trajectory.
- $\vec{V}_{T_j}$ is the vector formed from the first point to the last point of the second trajectory.
- the operator ‖ . ‖ gives the Euclidean distance between the first point and the last point of the considered trajectory.
- $C_2(T_i,T_j)$ gives the direction difference between two given trajectories T_i and T_j.

2.2.3. Torsion Complexity

A notion of torsion complexity is another criterion that can be employed to calculate the degree of geometric similarity between two trajectories. Torsion Complexity can be derived by taking into account the weighted average of the distance between the points of the trajectory from the virtual line drawn between the start and the end points (Anderson, Ames et al. 2014). Eq. (10) evaluates Torsion point differences when comparing two trajectories T_i and T_j.

$$C_3(T_i,T_j)=\left|\text{Com}_{T_i}-\text{Com}_{T_j}\right|;\ \ \text{Com}_T=\sum_{k=1}^{n-1}\left(\left(\frac{h_k+h_{k+1}}{2}\right).\left(\frac{d_k}{D}\right)\right) \quad (10)$$

where

- h_k denotes the vertical distance between the point kth and the imagery line between the start and the end points of the trajectory.
- d_k is the length of the kth edge denoting two successive turning points.
- D denotes the length of the imagery line.
- n is the number of the points of the trajectory T.
- $C_3(T_i,T_j)$ returns a measure of torsion complexity difference between two considered trajectories T_i and T_j.

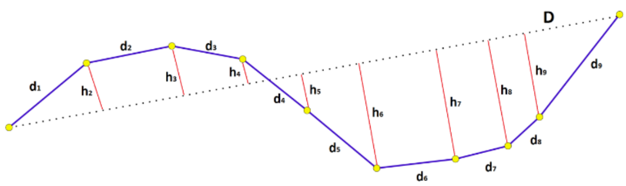

Figure 7. Complexity derivation according to trajectory vertical distances (Chehreghan and Ali Abbaspour 2017).

2.2.4. Shape

One of the most widely-used functions associated with the shape is the cumulative angle function, also known as the Turning function (Veltkamp 2001). The Turning function derives for each point of a given trajectory the value of the angle of the edge connected to the next point relative to the horizontal axis. Having derived these functions for each trajectory, the value of the area enclosed between the two turning functions is considered as the difference between the two trajectories. Eqs. (11) and (12) formally define the shape difference between two trajectories PL_1 and PL_2 (Zhang 2009), where $\theta(s)$ is the turning function of a considered trajectory (Chehreghan & Ali Abbaspour 2017).

$$\left\|\theta_{T_i}(s)-\theta_{T_j}(s)\right\| = \int_0^1 f(\theta_{T_i},\theta_{T_j})ds \tag{11}$$

$$f(\theta_{T_i},\theta_{T_j}) = \begin{cases} \left|\theta_{T_i}-\theta_{T_j}\right| & ;\ \text{if } \left|\theta_{T_i}-\theta_{T_j}\right| \le \pi \\ 2\pi-\left|\theta_{T_i}-\theta_{T_j}\right| & ;\ \text{if } \left|\theta_{T_i}-\theta_{T_j}\right| > \pi \end{cases} \tag{12}$$

while Eq. (11) derives the difference of turning function for two trajectories, the turning function for the mentioned trajectories is derived from Eq. (12). Finally, the degree of similarity $SSD(T_i,T_j)$ of two given trajectories Ti and Tj are derived using the Eq. (13):

$$SSD(T_i,T_j) = \frac{\sum_{k=1}^{m} C_k(P_i,P_j)}{m}; \tag{13}$$

$$P_i = \{P_{i,1},P_{i,2},\ldots,P_{i,n_i}\}\ ,P_j = \{P_{j,1},P_{j,2},\ldots,P_{j,n_j}\}$$

where

- m is the number of geometric properties considered.
- C_k gives the difference between the calculated values for the two trajectories from the viewpoint of the kth property
- P_i is a significant point extracted for the T_i trajectory.
- P_j is a significant point extracted for the T_j trajectory.
- n_i is the number of points of the T_i trajectory.
- n_j is the number of points of the T_j trajectory.
- $SSD(T_i,T_j)$ gives the degree of spatial similarity between the two T_i and T_j trajectories.

Having calculated the degree of spatial similarity, a matrix (Figure 8) is derived as the similarity matrix, and where $SSD(T_i,T_j) = SSD(T_j,T_i)$. (Figure 8)

3. Implementation and results

This section reports the implementation results of the proposed method. Figure 9 illustrates the trajectory dataset used in this study, which includes a part of the Geolife project (Zheng, Zhang et al. 2009) that recorded human trajectories in the city of Beijing from 2007 to 2012. We selected a sample of 326 trajectories recorded by various devices equipped with a GPS such as taxis, personal cars, bicycles and even walking displacements. After pre-processing and withdraw of outlier data, this gives 83412 intermediate trajectory points, and a total distance traveled of 672195 meters. The shortest trajectory is 8.54 meters whereas the longest is 14408.2 meters and the mean length of the trajectory is 2417.97 meters. Similarly, the mean sampling distance is 10.21 meters and the mean sampling time is 5.11 seconds.

3.1. Identifying Trajectory Significant Points

In the first phase of the study, a convex-hull geometric structure is implemented for the 326 trajectories. This gave a total number of 7498 convex-hull structures. The first peculiarity that appeared was a relative difference in convex-hull structures for trajectories with similar origins and destinations.

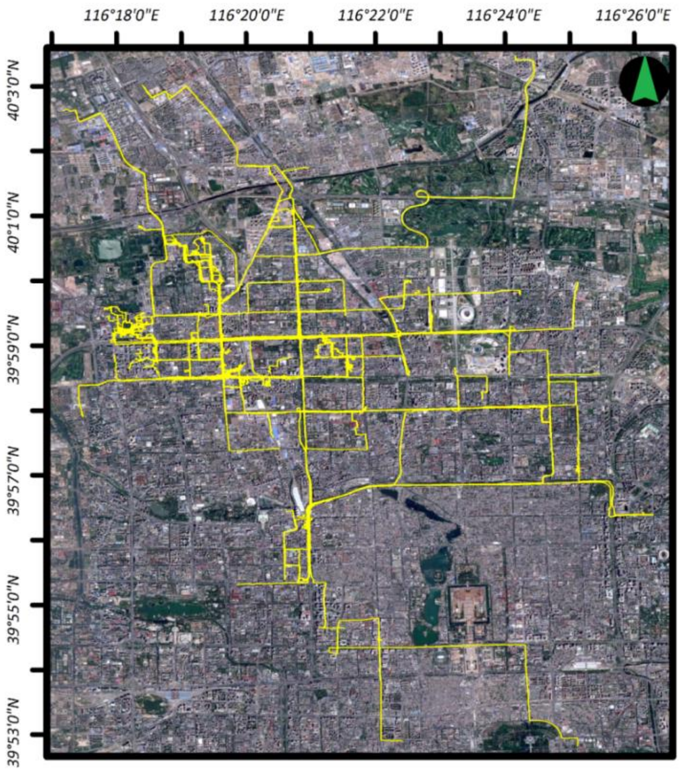

Figure 9. Trajectory sample from the Geolife dataset.

For example, trajectories with ids 76 and 83 are geometrically similar but contain 107 and 148 convex-hull structures, respectively (Figure 10). Different sampling times and an insufficient prior cleaning process are likely to explain these differences. Clearly, the critical points identified from this convex-hull s should give additional insights while analyzing trajectory differences and similarities. Typically, the number of convex-hull s is extremely variable, especially when the chosen thresholds at the pre-processing phase are relatively high values.

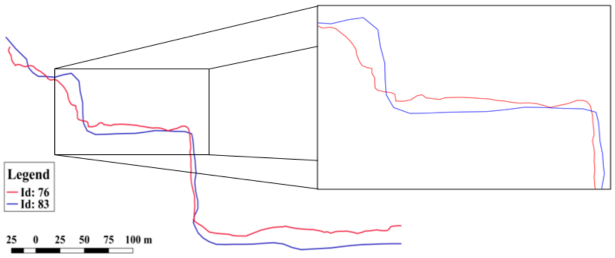

Figure 10. Two similar trajectories of 76 and 83 with different convex-hull s.

The curvature points are derived using definition 2. Subsequently, for the removal of outlier structures, a threshold was defined to evaluate the distance between the curvature point and the convex-hull line (i.e., connecting line as start and end points of a convex-hull), taking into account the length of the trajectory so that we eliminated the structures in which the mentioned distance was more than 0.02D (D is the distance between the origin and destination of the trajectory) as well as the structures with less than four points. In order to achieve the best balance of the need to keep the main semantics of a given trajectory while reducing its complexity, the most appropriate threshold value should be identified. Indeed, this is context dependent. In the context of the dataset used in the experimental validation, it appears after several iterations that a threshold value of 0.02 is appropriate to delete noisy convex. Overall, 2317 structures were removed from the initial set of structures. Table 1 tabulates the number of eliminated structures and the significant points obtained for the trajectories categorized with different lengths. The results presented in Table 1 also show the existence of the greatest outlier structures for trajectories with high length.

What follows is the derivation of trajectory similarities by identifying the significant points and extracting the structure of each trajectory.

3.2. Geometric Criteria Derivation

The derived trajectories generated by the pre-processing introduced in the previous section are compared according to the selected geometric parameters (i.e., shape, turning point, direction, and distances). Without loss of generality, three trajectories (i.e., ids 63, 82, and 88) were chosen from the studied data to examine the efficiency of the proposed method (Figure 11). Practically, in order to evaluate the proposed approach, the results obtained by comparing the similarity between all the main trajectories and the extracted trajectories are compared.

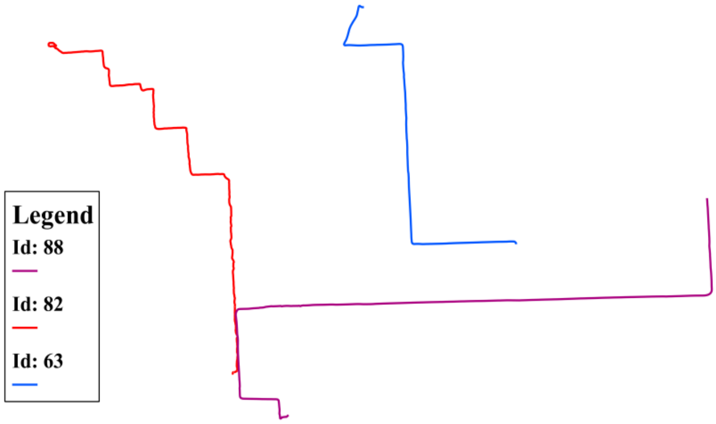

Figure 11. Trajectories sample used for geometric similarity analysis.

Trajectories ids 63, 82 and 88 have 403, 534, and 711 points, respectively. By applying the convex-hull filtering approach, 16, 27 and 17 significant points are extracted for each trajectory, respectively. This overall generates around 5% of the number of initial points of the trajectories, which considerably reduces the number of points and then significantly decreases computation times. Next, the geometric similarity of each pair of trajectories is examined (Table 2). As shown in this table, the ATD of these trajectories is similar to the main trajectories to a great degree. Similarity differences between trajectories 63 82, 63-88, and 82-88 are extracted and presented in Table 2. Moreover, by comparing differences regarding distance, direction, shape and turning point parameters, one can observe a relatively high efficiency of the convex-based approach in extracting the significant points. For example, the trajectories 88-63 and the related ATD exhibit more similarity in shape with respect to other pairs as can be seen in Table 2.

After applying the whole similarity approach to the cross-comparison of all the trajectories from our sample dataset, the overall similarity difference value equals 0.002, which highlighted a high efficiency in extracting the significant points. Moreover, this filtering approach significantly decreases computation times: with a Personal Computer, RAM of 8 GB, and a CPU = Intel Core (TM) i5-2310, @ 2.9 GHz, computing time reduced from 15270 seconds to 1140 seconds. Moreover, MATLAB software is used to implement the proposed method. Therefore, the proposed method caused to approximately 93% reduction in computing time, which is considerable. So, the overall result of the proposed method could be mentioned in two points. At first, the geometry of the trajectories is preserved to be utilized in the related analysis with less than 5% of the primary points of the trajectory since the similarity between two trajectories does not change significantly. Secondly, due to the lower number of processed points of the trajectory, the computing time is reduced by approximately 93%. In addition, considering no similar method to detect the critical geometric points of the trajectory, the comparison with any other methods is not suitable.

Table 1. Extraction of trajectory significant points.

Turning points	Curvature points	The variance of the distance from convex hull	Eliminated convex hulls	Initial convex hulls	Trajectories	Length (m)
449	447	3.41	63	510	56	0 - 1000
1012	1010	7.33	235	1245	82	1000 - 4000
2536	2534	14.57	1376	3910	127	4000 - 8000
1196	1194	17.20	639	1833	62	8000 - 15000

Table 2. Trajectory geometric similarities

Similarity	Turning point	Shape	Direction (Radian)	Distance (km)	Trajectories	
0.510	10.7	329.4	0.44	1.72	82-63	Main trajectoires
0.310	32.8	714.5	2.27	0.87	88-63	
0.360	49.3	642.7	2.01	0.61	88-82	
0.506	10.4	323.1	0.44	1.72	82-63	Related ATD
0.308	33.1	715.6	2.27	0.87	88-63	
0.361	49.1	643.5	2.01	0.61	88-82	

4. Conclusion

Over the past few years, many large urban trajectory databases have been made available thanks to the rapid development of mobile and positioning devices. It opens many avenues of research and applications oriented to the understanding of how humans behave in large urban systems. Amongst many research issues still to address, the search for trajectory similarities in urban environments is still an important direction worth exploring. The research presented in this paper introduced a geometrical approach based on convex-hull s that help to identify a series of the intrinsic trajectory properties, that is, turning points, shapes, direction, and distances. Our work is grounded on our previous work and the development of a geometric framework oriented to trajectory modeling.

In the present paper, we further explore additional capabilities that will efficiently support the exploration of regular patterns. This approach is supported by a series of geometrical definitions and algorithms that first significantly reduce the complexity of incoming trajectories and secondly, identify a trajectory geometrical decomposition modeled by an ATD. The experimental evaluation applied to the Geolife benchmark trajectory database reveals a series of significant patterns considering the geometric functions introduced. The results are presented using the geometric similarity difference between two trajectories before and after applying the ATD method, as there is a mean of 0.002 similarity difference in results considering the mean of 5% of the row points used for the ATD method. Therefore, the quantitative results of this paper show that the ATD method is extremely efficient in reducing the trajectory points in a lower number of critical points. So, the used method could be considered in many domains of trajectory analysis to reduce complexity, especially in big data applications. This paper reveals the efficiency of the ATD method, which could be used for big data applications such as online route finding, extracting the patterns of movement, and major selected roads for business or management goals. Further work will be oriented to the integration of additional contextual properties, especially when considering additional urban dimensions that might provide more insights when studying the behavior and the reasons behind such behaviors. We also plan to apply the whole framework to different urban trajectory databases as well as to different moving objects databases such as the ones available in many biological studies.

References

Anderson, D. L., D. P. Ames & P. Yang (2014). "Quantitative Methods for Comparing Different Polyline Stream Network Models." Journal of Geographic Information System 6(02): 88-98.

Buchin, K., M. Buchin, J. Gudmundsson, M. Löffler & J. Luo (2008). Detecting commuting patterns by clustering subtrajectories. International Symposium on Algorithms and Computation, Springer.

Buchin, M., S. Dodge & B. Speckmann (2012). Context-aware similarity of trajectories. International Conference on Geographic Information Science, Springer.

Cao, H., N. Mamoulis & D. W. Cheung (2005). Mining frequent spatio-temporal sequential patterns. Data Mining, Fifth IEEE International Conference on, IEEE.

Chehreghan, A. & R. Ali Abbaspour (2017). "An assessment of spatial similarity degree between polylines on multi-scale, multi-source maps." Geocarto International 32(5): 471-487.

Chehreghan, A. & R. Ali Abbaspour (2017). "A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm." Cartography and Geographic Information Science: 1-15.

Chehreghan, A. & R. Ali Abbaspour (2018). "An assessment of the efficiency of spatial distances in linear object matching on multi-scale, multi-source maps." International Journal of Image and Data Fusion 9(2): 95-114.

Demšar, U., K. Buchin, F. Cagnacci, K. Safi, B. Speckmann, N. Van de Weghe, D. Weiskopf & R. Weibel (2015). "Analysis and visualisation of movement: an interdisciplinary review." Movement ecology 3(1): 5.

Dodge, S., R. Weibel & E. Forootan (2009). "Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects." Computers, Environment and Urban Systems 33(6): 419-434.

Dodge, S., R. Weibel & P. Laube (2011). "Trajectory similarity analysis in movement parameter space." Plymouth, UK: Proceedings of GISRUK: 27-29.

Giannotti, F. & D. Pedreschi (2008). Mobility, data mining and privacy: Geographic knowledge discovery, Springer Science & Business Media.

Hosseinpoor Milaghardan, A., R. Ali Abbaspour & C. Claramunt (2018). "A Geometric Framework for Detection of Critical Points in a Trajectory Using Convex-hulls." ISPRS International Journal of Geo-Information 7(1): 14.

Jiang, X., C. Zheng, Y. Tian & R. Liang (2015). "Large-scale taxi o/d visual analytics for understanding metropolitan human movement patterns." Journal of Visualization 18(2): 185-200.

Lee, J.-G., J. Han, X. Li & H. Cheng (2011). "Mining discriminative patterns for classifying trajectories on road networks." IEEE Transactions on Knowledge and Data Engineering 23(5): 713-726.

LIN, H., J. LV, C. YANG, M. DENG, K. WANG & X. WANG (2014). "GPS Trajectory Mining: a Survey⋆." Journal of Computational Information Systems 10(16): 6947-6956.

Lipan, F. & A. Groza (2010). Mining traffic patterns from public transportation GPS data. Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on, IEEE.

Lu, M., Z. Wang, J. Liang & X. Yuan (2015). OD-Wheel: Visual design to explore OD patterns of a central region. Visualization Symposium (PacificVis), 2015 IEEE Pacific, IEEE.

Morzy, M. (2007). "Mining frequent trajectories of moving objects for location prediction." Machine Learning and Data Mining in Pattern Recognition: 667-680.

Parent, C., S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo & N. Pelekis (2013). "Semantic trajectories modeling and analysis." ACM Computing Surveys (CSUR) 45(4): 42.

Soleymani, A., J. Cachat, K. Robinson, S. Dodge, A. Kalueff & R. Weibel (2014). "Integrating cross-scale analysis in the spatial and temporal domains for classification of behavioral movement." Journal of Spatial Information Science 2014(8): 1-25.

Tang, J., F. Liu, Y. Wang & H. Wang (2015). "Uncovering urban human mobility from large scale taxi GPS data." Physica A: Statistical Mechanics and its Applications 438: 140-153.

Tong, X., D. Liang & Y. Jin (2014). "A linear road object matching method for conflation based on optimization and logistic regression." International Journal of Geographical Information Science 28(4): 824-846.

Veltkamp, R. C. (2001). Shape matching: similarity measures and algorithms. Shape Modeling and Applications, SMI 2001 International Conference on IEEE.

Widhalm, P., Y. Yang, M. Ulm, S. Athavale & M. C. González (2015). "Discovering urban activity patterns in cell phone data." Transportation 42(4): 597-623.

Yuan, H., Y. Qian, R. Yang & M. Ren (2014). "Human mobility discovering and movement intention detection with GPS trajectories." Decision Support Systems 63: 39-51.

Yuan, Y. & M. Raubal (2014). "Measuring similarity of mobile phone user trajectories–a Spatio-temporal Edit Distance method." International Journal of Geographical Information Science 28(3): 496-520.

Yue, Y., T. Lan, A. G. Yeh & Q.-Q. Li (2014). "Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies." Travel Behaviour and Society 1(2): 69-78.

Zhang, M. (2009). "Methods and implementations of road-network matching." Unpublished PhD Dissertation, Technical University of Munich.

Zheng, Y. (2015). "Trajectory data mining: an overview." ACM Transactions on Intelligent Systems and Technology (TIST) 6(3): 29.

Zheng, Y., L. Zhang, X. Xie & W.-Y. Ma (2009). Mining interesting locations and travel sequences from GPS trajectories. Proceedings of the 18th international conference on World wide web, ACM.