A novel compression approach for truck GPS trajectory data

Sijing Liu¹,² | Gang Chen¹,² | Long Wei¹ | Guoqi Li¹,³

¹ School of Transportation and Logistics, Southwest Jiaotong University, Sichuan, People’s Republic of China
² National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Sichuan, People’s Republic of China
³ National United Engineering Laboratory of Integrated and Intelligent Transportation, Sichuan, People’s Republic of China

Abstract

Nowadays, the application of location-aware devices, such as global positioning system (GPS)-enabled mobile phones and personal digital assistants (PDAs) is increasing. Based on this, a new demand for efficiently storing trajectory data has arisen. Trajectory compression reduces the storage space and cost, the cost and time of data transmission by retaining critical trajectory points and effectively eliminating redundant data. Herein, a novel trajectory compression method based on stay points (TCSP) is proposed. By using the road network data and the stay points of GPS trajectory data to compress the GPS trajectory data, this method can greatly improve the ratio and accuracy of compression. Experimental results based on real data sets show that the compression rate of this method can stably reach 8.61%, and the compression effect can be restored to the original trajectory route. Compared with the critical point algorithm and the Douglas–Peucker (DP) algorithm, the TCSP method can reach a better compression effect and compression ratio.

INTRODUCTION

In recent years, with the rapid development and extensive use of location-aware devices, a large amount of global positioning system (GPS) trajectory data is also recorded. To improve data processing speed and decision efficiency, many enterprises and organizations conduct data mining on these GPS trajectory data to facilitate data commercialization. Generally, these bulky GPS trajectories often contain a lot of redundant data [1]. Therefore, the pre-processing of trajectory data becomes the key step of trajectory compression.

These data usually includes a very complete track chain, which has a broad application prospect in behaviour analysis [2–5], urban planning [6], and regional industry situation analysis [7–9]. However, the interval time of the GPS trajectory data is too short, these huge data take up a lot of storage space and bandwidth resources, posing a big challenge in data processing and mining. Therefore, trajectory compression is necessary for mining these trajectory data. Trajectory compression technology refers to eliminating redundant data in trajectory data by detection means or methods and represents a piece of complete trajectory information with as few trajectory points as possible, thereby achieving the purpose of saving storage space and facilitating data mining [8, 10, 11].

Trajectory compression has been studied for decades. Several trajectory compression algorithms or methods have been proposed, and most of them can be classified into lossy compression. The main idea is to delete some trajectory points while maintaining an acceptable degree of error [12]. We consider the problem of lossy compression when maintaining stay points in GPS trajectory data. To ensure its usability in practical applications, there are some requirements need to be meet: (a) The compression ratio should be improved compared to other similar algorithms or methods; (b) The compression process should be fast; (c) The compressed trajectory should be able to be restored to the original trajectory with high accuracy.

Herein, we present a novel trajectory compression method based on stay points (TCSP) that combines the critical point algorithm [13] and the stay point identification method [14].
Firstly, according to the needs of the GPS trajectory compression process, this study designs the calculation method of the speed threshold in the stay point identification method. Secondly, based on the critical point algorithm [13], it proposes a GPS trajectory compression method combining road network data. Finally, the feasibility of the TCSP method is carefully verified through comparative experiments. The results show that the TCSP method can effectively compress the GPS trajectory data.

The contributions are as follows:

1. A novel trajectory compression method called TCSP is proposed, which uses the critical points of the road network and the stay points of trajectory to obtain the key information of GPS data while compressing the trajectory.
2. We design a method of determining the speed threshold in the stay point identification algorithm. The calculation method is proved to be superior to the K-means algorithm and the different density and join-based clustering (DJCluster) algorithm.
3. The evaluation results show that the superiority of our approach over previous the critical point algorithm and the Douglas–Peucker (DP) algorithm.

The rest of this article is organized as follows: Section 2 reviews and summarizes the related works, and discusses the advantages and disadvantages of them. Section 3 introduces the main rules for data pre-processing before trajectory compression. Section 4 describes the main steps of stop point identification and the calculation method of speed threshold. Section 5 describes the main steps and the main idea of the TCSP method we present. Section 6 discusses our preliminary evaluation results. Conclusions and future work are discussed in Section 7.

2 RELATED WORKS

Currently, there are two commonly used categories of trajectory compression strategies proposed, aiming to reduce the size of a trajectory while not compromising much precision in its new data representation [15]. One is lossless compression algorithms, which performs a more accurate reconstruction of the original data without causing any distortions [12, 16]. The other is lossy compression algorithms, which is to lose some secondary information to improve the data compression rate while maintaining an acceptable degree of deviation [17, 18]. Moreover, lossy compression can be further divided into online compression technique and batched compression technique [19]. Online compression technique can compress a trajectory instantly as the object travels, and make judgment of retention or deletion for each arriving trajectory point in time. Meanwhile, batched compression technique reduces the size of the GPS trajectory only after the trajectory has been fully generated. Thus, batched compression technique can always get the globally optimal compressed results at the cost of more time than online compression.

One of the most classic trajectory compression algorithms is the DP algorithm [19]. It uses the line connecting the start point and the endpoint of the trajectory as the approximate trajectory, and then calculates the vertical Euclidean distance between each point of the trajectory data and the approximate trajectory, with vertical Euclidean distance threshold better than the distance threshold. The trajectory points are used as segmentation points to segment the current trajectory, and then the above process is repeated until all vertical Euclidean distances are less than the distance threshold. The last segmentation points are the compressed trajectory points. Based on summarizing the Top-Down, the Bottom-Up, the Sliding Window and the Opening Window (OPW) algorithms, Meratnia et al. proposed a synchronous Euclidean distance (SED) to replace perpendicular distance, and the Top Down Time Ratio (TD-TR) algorithm to improve the DP algorithm and the Opening Window Time Ratio (OPW-TR) to improve the OPW algorithm by adding the time variable [20]. Potamias et al. put forward the spatiotemporal trace (ST-Trace) algorithm in 2006 to estimate the safe zone of successor point by location, velocity, and direction [21]. Schmid et al. proposed that trajectories stored in the form of trajectory points can replace semantic information of road networks in 2009 [22]. The critical point algorithm in [13] uses the speed, direction and other attributes in GPS trajectory data to eliminate trajectory points that contain little information and compresses the GPS trajectory data. This method is also the concept of “line simplification” in the geographic information system (GIS) field. Song et al. proposed paralleled road-network-based trajectory compression (PRESS) framework to compress trajectory under road network constraints, which separates the spatial representation of a trajectory from the temporal representation and uses a hybrid spatial compression (HSC) algorithm and error bounded temporal compression (BTC) algorithm to compress the spatial and temporal information of trajectories [23].

These algorithms or methods successfully solve the compressed GPS trajectory data. However, among the above algorithms, the DP algorithm and the OPW algorithm mainly focus on the overall contour geometric features of the trajectory, ignoring the way of movement and internal features on the trajectory [23]. Besides, the error rate of the TD-TR algorithm is high [24]; while the processing time of the ST-Trace algorithm is considerable [24, 25]. Most importantly, the purpose of trajectory compression is not only to save the storage space, but also to facilitate the extraction of the motion characteristics of the subject in the trajectory, such as speed, staying point, and direction. Then the extracted characteristics can be used for further research such as vehicle logistics characteristics, and route optimization. However, most of the aforementioned algorithms or methods cannot effectively retain the key information on the road network [24, 25].

Thus, to fill the gap, the TSCP method is proposed. we use both the time dimension and the space dimension of the GPS data to improve the stay point identification in [14]; the TCSP
TABLE 1  Data structure of the GPS track data set

| Data type   | Remark                                      |
|-------------|---------------------------------------------|
| Id          | GPS trajectory record id                    |
| Car number  | License plate number                        |
| Company     | Vehicle registered company                  |
| Kind        | Vehicle type                                |
| Colour      | Vehicle colour                              |
| GPS time    | GPS track point recording time              |
| Longitude   | Number                                      |
| Latitude    | Number                                      |
| Direction   | Vehicle driving direction: degree           |
| Speed       | Instantaneous velocity: kilometres per hour (km/h) |
| Speed limit | Limit velocity: kilometres per hour (km/h)  |

3 | **PRE-PROCESSING OF TRAJECTORY DATA**

For GPS trajectory data, data pre-processing is usually the premise and basis for further research. Its main purpose is to solve the problems of missing data, data errors and data redundancy in the original GPS trajectory data. The process of data pre-processing usually analyses the overall characteristics of the data, and then adopts corresponding processing rules or methods for processing according to the characteristics of the data.

3.1 | **Page GPS track data characteristics**

The source of GPS trajectory data is usually the data centre transfers the location records generated by the terminal equipment to the local in the form of logs. In this article, the data comes from the Sichuan provincial department of transportation highway monitoring and settlement centre. The track record structure of the GPS track data set is shown in Table 1.

As shown in Table 1, in addition to the space and time attributes, the GPS trajectory data record table also contains the instantaneous speed of the GPS trajectory point, the speed limit of the road at the time, and the information such as the plate number, registered company, vehicle type, and colour. In the research of this article, its time, space property and speed properties are relatively important features. Therefore, in this data pre-processing, the time, space, and velocity attributes of GPS trajectory data are mainly processed.

3.2 | **Page GPS track data cleaning**

Currently, due to errors of GPS-related equipment itself and errors that occur during data transmission and reception of vehicle-mounted GPS equipment, GPS data often exhibit problems such as data duplication, data loss, and GPS drift. However, the use of GPS data containing duplicates, deletions and drifts affects the results and efficiency of subsequent trajectory compression, so it is necessary to pre-process the GPS trajectory data.

The authors of this article believe that before any GPS data is used in actual research, it is necessary to ensure that the basic attributes of the data are complete, the time attribute information of the data is unique. Based on the above principles, combined with the actual research needs and data sets of this study, the data pre-processing rules are as follows:

1. Delete GPS track data record with incomplete data attribute information [26];
2. If the GPS track data have the same time attribute information, only retain the last record with remaining records deleted;
3. If the GPS track data have a jump, first determine the jump point by the map matching method, and then remove it. Then the transition points are connected into a trajectory. If the transition points can form a trajectory separately, it is regarded as an independent trajectory data [26].
4. If the speed attributes of the GPS track data are all less than a predetermined threshold, we can suggest that these track points in track always stay in an area so that they do not have research value and need to be deleted.

4 | **STAY POINT IDENTIFICATION ALGORITHM**

The core of stay point identification is to determine whether the stay point is valid. There are three kinds of stay points: the first is mainly parking on the road due to waiting for traffic lights. Generally, the parking time of this mode is short and meaningless; the second is mainly due to refueling, half-way breaks, traffic congestion and so forth. This kind of parking time is generally longer, which has a greater impact on the research in this article. The third is parking, loading or unloading, and so forth. Through these three types of stay points, information such as the starting point of the truck line can be identified. Considering the characteristics of GPS data, we can judge whether it is a stay point according to the speed and time attributes of the GPS track points. Herein, a stay point identification algorithm based on speed and time is used [14]. The algorithm including two steps: judging the suspicious stay point based on the speed of the trajectory point, and identifying the stay point based on the stay time of the suspicious stay point. In [14], the speed threshold is determined by experience or previous researches, which cannot scientifically set the speed threshold for the research.
object. To solve the above problem, this study innovates the method of determining the speed threshold in this algorithm.

### 4.1 Suspicious stay point identification

The identification of the suspicious stay point depends on the speed threshold. A speed upper limit \( V_{set} \) can be set to determine whether to stay. First, according to \( V_{set} \), all GPS points are divided into two categories: suspicious stay points and driving points. Since there should not be only one point which speed \( V < V_{set} \) while parking, there should be at least two points, so a stay point candidate area can be formed. The process is shown in Figure 1.

In [14], the speed threshold is determined by previous experience. In [27], the speed threshold is replaced by the distance threshold. However, the distance threshold of different vehicles in different parking positions is different, or even too large. Considering that the speed of different vehicles in different parking positions is very close, this study proposes to use the speed threshold for stay point identification. Under actual conditions, the speed threshold should be determined according to all the suspicious stay points in a piece of GPS trajectory data, that is, the average speed of the trajectory point approaching 0 km/h over a period of time. Considering the general acceleration and deceleration of the vehicle, the time is set to 5 s. Thus, the method for setting the speed limit \( V_{set} \) is as follows: assume that in a piece of trajectory data, the number of GPS trajectory data is \( N \), and the number of the speed of the GPS trajectory points is \( n \); \( G \) is a GPS trajectory point with velocity of 0 km/h in the trajectory data, and the recording time at this point is \( T \), the speed is \( V \); \( G_{i} \): \( \{ G_{i} \mid T - 5s < t < T + 5s \} \) is used to represent the trajectory points taken within 5 s of the forward or backward trajectory point \( G \), the number is \( m \); The speed set of each point in \( G_{i} \) is: \( V_{G_{i}} \): \( \{ V_{G_{i}} \mid G \subseteq G_{i} \} \); then \( V_{set} \) can be expressed by the following mathematical formula:

\[
V_{set} = \frac{\sum_{i=1}^{m} V_{G_{i}} + \sum_{j=1}^{m} V_{G_{j}}}{m + 1}
\]  

### 4.2 Stay point identification

Through a series of suspicious stay points obtained earlier, combining the time threshold and selecting the last point in the set of stay points as the final stay point, the specific process is as follows:

1. Read the first suspicious stay point \( P_{1} \) in the trajectory and put it into the stay point set \( P \).
2. Identify whether there is still a suspected stay point \( P_{i} \) \((i = 1, 2, 3, 4, 5 ...)\), and if there is, place the point in the stay point set \( P \), and repeat step (2), otherwise proceed to step (3).
3. Calculate the stay time of the stay point set \( P \). If \( t_{stay} \) is shorter than the time threshold \( t \), the set \( P \) is cleared. If it is longer than the time threshold \( t \), the stay record is retained and the last stay point in the stay point set \( P \) is used as the final GPS trajectory stay point for this stay point set \( P \).

Through the stay point identification algorithm, a series of stay points of GPS trajectory data can be obtained, which is convenient for finding the driving law of the vehicle, such as the stay point and stay time, laying a good foundation for subsequent trajectory compression.

### 5 THE TCSP METHOD

The purpose of the trajectory is not only to compress the storage space of the trajectory data, but also to facilitate the extraction of distance, space and time features in the trajectory, to optimize the vehicle path in conjunction with the road network or to study and analyse the characteristics of urban freight. Therefore, it is very important to combine the road network with trajectory compression and determine the key points of the trajectory. Therefore, we propose the TCSP method in this study. The key intersection points of road network data and the stay points identification algorithm are used in the TCSP method to improve the compression ratio and to retain the key information in the compressed trajectory. On the one hand, using the intersection points of road network data can clearly
express the trajectory of the vehicle with the least trajectory points. On the other hand, the use of stay point identification may retain the key stay point information in the trajectory. And it contributes to researchers such as routing optimization, and logistics characteristics, helping achieve vehicle high efficiency, low fuel consumption and low carbon. Additionally, the threshold \( t \) in the stay point recognition algorithm can be flexibly set according to vehicle characteristics or research needs.

This method first parses the road network data and simplifies the road network data to obtain all the critical intersection information in the road network data; then uses the key intersection point data of the road network to improve the critical point selection method of the critical point method and further improve the trajectory compression rate, compression effect and reduce the storage space of the GPS trajectory data; finally, the dwell point identification is used to retain the dwell point information in the trajectory, and the trajectory is sorted according to the “id” attribute of the trajectory, which facilitates the feature extraction and deep mining of the GPS trajectory data in the future.

5.1 Related definitions

GPS trajectory: the GPS trajectory \( T \) of a vehicle is a continuous sequence composed of trajectory points arranged in the order of recording time. \( G = \{G_{id}, id, <P_1, ..., P_i, ..., P_j>\} \), where \( P \) represents the trajectory point, includes latitude and longitude, record time information, instantaneous speed, and other basic information; \( G_{id} \) means track number; id means vehicle number.

Critical points: refers to the trajectory points with significant changes in direction or speed in the GPS trajectory. These GPS trajectory points usually contain critical trajectory points with more positioning information.

Road network segment: a section in the road network is a directional (including bidirectional) straight line or polyline, including: (a) The starting point; (b) The termination point; (c) The section number road ID; (d) The section width \( w \); (e) The section direction \( d \); (f) The section name.

5.2 Critical point selection rules

According to the critical point method in [13], firstly, critical points need to be selected from the GPS trajectory data, the selection of critical points refers to the trajectory points whose direction or speed changes significantly in the GPS track. The GPS track data is compressed by excluding track points other than the critical points.

The TCSP method proposed herein is to improve the method of selecting critical intersection points by combining road network data and further improve the compression rate of GPS trajectory without affecting the path recognition of the vehicle.

Due to complex phenomena such as the interleaving of various roads in the road network, this study explains the selection of critical points based on whether the roads in the road network are staggered with each other.

1. If the road on which the vehicle is driving intersects with other roads, record the point in the vehicle’s driving trajectory that is the same or similar to the intersection of the road network, and use this as a critical point.
2. If the road on which the vehicle is traveling does not intersect with other roads, record the trajectory points in and out of the road in the vehicle’s trajectory and use it as the critical point in this road.

5.3 Method description

The trajectory compression method proposed mainly includes two parts, which are the trajectory compression and stay point superposition. The main purpose of trajectory compression is to compress the pre-processed trajectory data based on the filtered critical points of the road network. The main purpose of the stay point overlay is to retain the key stay information in the trajectory data, and further compress the distance between the stay point and the compressed trajectory point to form a compressed trajectory that retains the key stay information of the trajectory data.

The screening of the critical intersection points of the road network is implemented by the Python programming language. The specific code of this part is as follows:

**Algorithm 1: Identifying Critical Intersections Points in Road Network**

| Input | The location where the critical intersections of the road network are screened |
|-------|--------------------------------------------------------------------------------|
| Output| Storage unit \( r_1, r_2, r_3, ... \) representation of the filtered results |
| 1     | `import osmnx as ox` |
| 2     | `place = 'place' //input address//'` |
| 3     | `road = ox.graph_from_place(place,simplify=False)` |
| 4     | `fig=ax = ox.plot_graph(road,node_color='b',node_zorder=2)` |
| 5     | `road_simplify = ox.graph_from_place(place,simplify=False,simplify=True)` |
| 6     | `fig,a = ox.plot_graph(road_simplify,node_color='b',node_zorder=2)` |
| 7     | `nc = [b if ox.is_endpoint(road, node) else 'r' for node in road.nodes()]` |
| 8     | `fig,a = ox.plot_graph(road,node_color=nc,node_zorder=2)` |

As shown in the Algorithm 1, we first specify the filtering area (line 2); then we visualize the road network data in the filtering area (lines 3–4), “simplify = False” means that no filtering simplification is currently performed, and the road network data in the area is simplified and visualized (lines 5–6); and we visualize the key points and non-key points in the road network data at the same time, and distinguish them by colour,
where the key points are blue and non-the key points are red (lines 7–8).

The specific process of the trajectory compression part is shown in Algorithm 2, where distance (TD, JCD_point) indicates the distance between the trajectory point and the intersection of the road network. The purpose is to determine whether the trajectory point is a candidate key based on the distance between the two points, and finally return the set of candidate key points to the intersection of the road network among the candidate key points; the ‘mindistance TD’ represents the trajectory point closest to the intersection of the road network among the candidate key points; the parameter ‘10m’ is set as average turning radius of large vehicles on main roads, ramps and intersections in [28].

**ALGORITHM 2: Trajectory Compression**

**Input:** track_data: Track data  
JCD: critical intersections points in the road network

**Output:** compressed_track

1 Function compression ()
2 for each point JCD_point in JCD
3   do for each data TD in track_data
4     do if distance (TD, JCD_point) < 10m then
5       select mindistance TD save as MINTD
6     endif
7   endfor
8 save MINTD as compressed_track  //Save all filtered MINTD as compressed_track files//
9 Endfor

As shown in Algorithm 2, it requires the input of trajectory data and road network intersection point data; the distance between the intersection point and the trajectory point are calculated by lines 4–5. When the distance is less than 10 m, the trajectory point can be seen as a trajectory critical point and be stored into a set, and then the point with the smallest distance is selected as the final key point and retained; the 8th line is to save all processed trajectory critical points as compressed trajectory data of the trajectory. After that, this data is processed through the “Stop Overlay” section to form the final compressed trajectory.

The process of the stay point superposition part is shown in Algorithm 3, where distance (SP, CT) represents the calculation of the distance between the stay point SP and the trajectory point processed by the trajectory compression part. And here, the distance (SP, CT) is used to determine whether the stay point overlaps the trajectory point. INSERTION-SORT (added_data) indicates that the trajectory data superimposed by the stay point is sorted by the “id” attribute using the insertion sort method, and the returned result is the final compressed trajectory.

**ALGORITHM 3: Stay Point Overlay**

**Input:** compressed_track  
stay_points

**Output:** final_track_data

1 Function Stay_point_add ()
2 for each track CT in compressed_track
3   do for each point SP in stay_point
4     do if distance(SP,CT) < 10m then
5       delete CT
6     endif
7   endfor
8 save compressed_track as added_data
9 Endfor
10 INSERTION-SORT(added_data)//Sort by data using the ordinal attribute of the data//
11 output: final_track_data

This code requires the input of stay point data and compressed trajectory data processed by the “trajectory compression” step. Lines 3–8 are to calculate the distance between the compressed trajectory point data and the dwell point data. When the distance is less than 10 m, the dwell point data is retained and the trajectory point data is deleted, eventually forming uncompressed trajectory data; The 12th line is sorted according to the “id” attribute of the compressed track data formed after the processing of 3–8 lines by the insertion sort method, and finally the final compressed track “final_track_data” is output.

6 | EXPERIMENTS AND ANALYSIS

6.1 | Data preparation

The data comes from the Sichuan provincial department of transportation highway monitoring and settlement centre. The sample is selected from the GPS trajectory data of heavy trucks in March 2019. It contains a total of 679,849 GPS trajectory data. Each data records the truck number, license plate number, recording time, latitude and longitude coordinates, direction, instantaneous speed, and speed limit. This study first screens and pre-processes the trajectory data in this data set. After processing, a total of 413,453 GPS trajectory data are selected, and the data amount is about 40.2 MB. It is used for the recognition and compression of the stay points, the compression effect and compression ratio of the experiment are analysed.

6.2 | Analysis

Because the TCSP method mainly involves the stay point identification and the trajectory compression method based on critical point of the road network method, the analysis of the experimental results mainly includes the inspection of the identifica-
6.2.1 Stay point identification

The identification of stay points is mainly divided into two steps: identification of suspicious stay points and identification of stay points. In the experiment, the trajectory data in the GPS trajectory of the truck is calculated and analyzed. The speed threshold is calculated according to formula (1), and the result is 5.6 km/h. The time threshold is determined based on the type of vehicle and its usual activities. The research object of this article is the trajectory data of heavy goods vehicles, and the activities of heavy goods vehicles during their stay usually are refueling, breaks, traffic congestion, or loading and unloading of goods. The minimum stay time of these activities is the time threshold determined in this article, and the result is 600 s. Figure 2 shows the identification results of the trajectory.

The evaluation indicators for stop point identification mainly include accuracy rate (P), recall rate (R), and comprehensive evaluation index (F1-Measure) [27]. Among them, the accuracy rate is the ratio of the exact number in the results to the total number of results; the recall rate is the ratio of the number of related results identified to the total number of results in the data, which measures the recall rate; F1-Measure comprehensively considers the results of Precision and Recall. The larger the three indicators are, the better the experimental results are. Herein, the accuracy rate refers to the ratio of the identified true stay points to the total number of identified stay points; the recall rate refers to the ratio of the identified true stay points to the total stay points. The specific calculation formulas for the three evaluation indicators are as follows. Among them, P1 and P2 refer to the number of stay points that are correct and incorrect, respectively, and R represents the number of actual stay points.

\[
\text{Precision} = \frac{P_1}{P_1 + P_2} \quad (2)
\]
\[
\text{Recall} = \frac{P_1}{R} \quad (3)
\]
\[
F_1 - \text{Measure} = \frac{2P_1}{P_1R + P_1P_2 + P_2} \quad (4)
\]

Table 2 shows the comparison results of the stay point identification results in this study with K-means and DJ-Cluster algorithms:

As shown in Table 2, compared with the K-Means method and the DJ-Cluster method, the recall and precision of the method in this study have improved to some extent. The reasons may be:

1. In the data pre-processing part, the noise in the data set is basically eliminated effectively.
2. The stay point identification method used here is based on the characteristics of trucks.

6.2.2 GPS trajectory compression

The road network data used in this experiment comes from the latest Chinese road data downloaded from the OpenStreetMap. The road network data is composed of dots and lines. The required intersection of road network data is filtered using the osmnx library in Python. Since the research object of this experiment is Sichuan’s heavy trucks, the map data screening area is Sichuan road data. The acquisition method is obtained by Algorithm 1, and since the required data is the critical point data of the road network of Sichuan Province, the ‘place’ in Algorithm 1 is entered into ‘Sichuan, China’.

Figure 3 indicates that using this method to filter the road network data can well filter the key points of each road in the road network, and it is convenient to use this data to compress the trajectory data later.

The method designed in this article is a GPS trajectory data compression based on the critical point of the road network.
method. Therefore, the experimental analysis mainly compares the experimental results of the critical point method and MMTC to analyse the compression rate and compression effect.

A total of 63 truck GPS trajectory data files are compressed using the critical points algorithm, the DP algorithm and the method proposed herein, and the compression ratios of the three methods are calculated respectively. The compression rate (CR) calculation formula is as follows. Among them, $C_0$ represents the number of track points in the track data before the track compression, and $C_1$ represents the number of track points in the track data after the track compression. The results are shown in Figure 4.

$$CR = \frac{C_1}{C_0}$$  \hspace{1cm} (5)

Figure 4 illustrates the compression ratio of the TCSP method, the DP algorithm, and the critical point algorithm when compressing the same GPS trajectory data. It is obvious that the TCSP method can achieve a higher and more stable compression ratio than the critical point algorithm and the DP algorithm. Through calculation, the average compression ratio of the TCSP method is 8.61%, which is 4.04% lower than the average compression rate of the critical point algorithm of 12.65%, and 9.28% lower than the DP algorithm. However, it can be seen from the figure that the compression ratio of some trajectory files is too much higher than the average compression rate. Through analysis, we find that the compression ratio of 24th, 30th, 39th, 53th, and 55th trajectory files are excessively higher than the average compression rate, and the reasons are as follows: (a) the number of trajectory points contained in the trajectory file is small; (b) the trajectory points contained in the trajectory file are mostly key trajectory points such as
FIGURE 5 The compression effect of the TCSP method. (a) Sample diagram of raw trajectory data, (b) Sample diagram of compressed trajectory data. (c) Sample diagram of compressed trajectory data connected into a line by character id

intersections; (c) the interval between points is more than one road segment.

An example of the compression effect of the TCSP method is shown in Figure 5. Among them, diagram a is the raw trajectory data; diagram b is the compressed trajectory data; diagram c is the trajectory line after the compressed trajectory data is connected into a line in order.

As shown in Figure 5, the TCSP method can completely represent the original trajectory data route without affecting the path recognition of the trajectory influences. And it can be directly used for data mining of trajectory data stay points.

7  |  CONCLUSIONS AND RECOMMENDATIONS

This study presents a new technique called TCSP for compressing GPS trajectory data. First, the stay points in the trajectory are determined by the stay points identification. Second, the critical intersection points extracted from the road network data are combined with the critical method to perform the trajectory compression. Finally, superimposed on the compressed trajectory as well as replacing the trajectory points at the intersection in the trajectory data, the trajectory compression is completed.

Experiments based on real data sets verify the feasibility and accuracy of the method. This method can ensure the integrity of the vehicle trajectory and express the dwell point information of the trajectory according to the spatiotemporal properties of the trajectory data itself. Its accuracy and compression ratio are improved to a certain extent compared with the previously proposed method.

The method will be further improved from the following aspects: (a) further consider the directional properties of the trajectory data, and improve the trajectory compression rate. (b) in-depth research on the threshold determination standard based on socioeconomic characteristics of different stay points. (c) combine with feature extraction of transportation distance, objects.

ACKNOWLEDGEMENTS

This research is funded by the National Natural Science Foundation of China (NSFC) under project codes NSFC 71603219, 41501123 and Research project of the China Railway Eryuan Engineering Group CO. LTD under project codes KYY2019093(19-21).
REFERENCES

1. Birnbaum, J., et al.: Similarity-Based Compression of GPS Trajectory Data. In: 2013 Fourth International Conference on Computing for Geospatial Research and Application, San Jose, 22–24 July 2013
2. Li, Q., et al.: Mining user similarity based on location history. In: GIS ’08: Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, (ACM), Irvine, 5–7 November 2008
3. Zheng, Y., et al.: Understanding Mobility Based on GPS Data. In: 10th International Conference on Ubiquitous Computing, Ubicomp 2008, (ACM), Seoul, 21–24 September 2008
4. Giannotti, F., et al.: Unveiling the complexity of human mobility by querying and mining massive trajectory data. VLDB J. 20(5), 695–719 (2011)
5. Grengs, J., et al.: Using GPS data to understand driving behavior. J. Urban Technol. 15(2), 33–53 (2008)
6. Veloso, M., et al.: Sensing urban mobility with taxi flow. In: 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN 2011, Chicago, 1 November 2011
7. Doytsher, Y., et al.: Storing routes in socio-spatial networks and supporting social-based route recommendation. In 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN 2011, Chicago, 1 November 2011
8. Zheng, Y., et al.: Mining interesting locations and travel sequences from GPS trajectories. In: WWW’09 - Proceedings of the 18th International World Wide Web Conference, Madrid, 20–24 April 2009
9. Zheng, Y., et al.: Recommending friends and locations based on individual location history. ACM Trans. Web 5(1), 1–44 (2011)
10. Jiang, J., Xiaoling, W.: Review on trajectory data compression. J. East China Normal Univ. (Nat. Sci.) (05), 61–76 (2015)
11. Zhang, Z.: Deriving trip information from GPS trajectories. Dissertation, East China Normal University (2010)
12. Jing, W.: Trajectory stops identification algorithm based on correlation coefficient, Dissertation, Central China Normal University (2016)
13. GB 50647-2011: Code for planning of intersections on urban roads. Tongji University (2011)

How to cite this article: Liu S, Chen G, Wei L, Li G. A novel compression approach for truck GPS trajectory data. IET Intell Transp Syst. 2021;15:74–83. https://doi.org/10.1049/itr2.12005