

## A Method of Urban Road Network Extraction Based on Floating Car Trajectory Data

Chunlei Mi<sup>1,2</sup>, Feng Lu<sup>1,2,3,4\*</sup>

<sup>1</sup> State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China- (mcl, luf)@lreis.ac.cn

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou 350003, China

<sup>4</sup> Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

**KEY WORDS:** floating car trajectory data, road network extraction, adaptive radius centroid drift clustering, WaveCluster, GPS data

### ABSTRACT:

With the gradual opening of floating car trajectory data, it is possible to extract road network information from it. Currently, most road network extraction algorithms use unified thresholds to ignore the density difference of trajectory data, and only consider the trajectory shape without considering the direction of the trajectory, which seriously affects the geometric precision and topological accuracy of their results. Therefore, an adaptive radius centroid drift clustering method is proposed in this paper, which can automatically adjust clustering parameters according to the track density and the road width, using trajectory direction to complete the topological connection of roads. The algorithm is verified by the floating car trajectory data of a day in Futian District, Shenzhen. The experimental results are qualitatively and quantitatively analyzed with ones of the other two methods. It indicates that the road network data extracted by this algorithm has a significant improvement in geometric precision and topological accuracy, and which is suitable for big data processing.

### 1. INTRODUCTION

In recent decades, the rapid development of urban construction, whether it is urban expansion or internal functional restructuring, will cause changes in urban road network. The traditional road network renewal mainly through the professional surveying and Mapping Department field survey, remote sensing image interpretation and digitization two ways, not only the cost of human and financial resources investment is high, the professional and technical requirements of practitioners are high, but also the acquisition period is long, the management workload of late renewal is large, it is difficult to meet the current urban development of road network data low-cost, high timeliness requirements. Therefore, there is an urgent need for a low-cost, short update cycle of automatic access to road network data. With the gradual opening of floating car trajectory data, it is possible to extract geometric and topological information of road network from floating car trajectory data and build and update road network automatically in real time.

At present, many scholars have carried out a lot of research in this field. We adopt the terminology introduced in recent surveys by Biagioni and Eriksson (Biagioni and Eriksson, 2012) and Ahmed et al. (Ahmed et al., 2015), who also describe and compare several approaches in great detail. Even so, for better classify the algorithms, we divide them into three categories according to the main principles of them. One is the method of clustering trajectory points and lines (Davies et al., 2006; Edelkamp and Schrödl, 2003; Kuntzsch et al., 2016; Li et al., 2019; Worrall and Nebot, 2007). The road components are obtained by spatial clustering of GPS, and then they are connected according to different rules. Another is the method of graph theory and image processing (Biagioni and Eriksson, 2012; Chen and Cheng, 2008; Jiang et al., 2012; Steiner and Leonhardt,

2011). The trajectory data are transformed into 2D raster images according to geographical location, and the road skeleton is extracted by different thinning methods. These methods can improve the speed of road network extraction, but they are often difficult to deal with at complex intersections. The third method is incremental merging of trajectories (Ahmed and Wenk, 2012; Karagiorgou and Pfoser, 2012; Karagiorgou et al., 2013; Li et al., 2015). that is, finding common segments among multiple trips, mainly based on spatial proximity and similarity of additional features such as heading. Tracks are sequentially integrated into an initially empty map, leading to modifications to the existing network whenever a new track is inserted.

Besides extracting geometric and topological information, some researchers also attempt to extract other spatial features of road networks (Cao and Krumm, 2009; Chen and Krumm, 2010; Deng et al., 2018; Guo et al., 2007; Schroedl et al., 2004; Zhang et al., 2010), such as the width and number of lanes per road, turning restrictions and other traffic-related features.

In summary, although many research achievements have been made in road network extraction, there are still many problems in current extraction methods: 1) Most algorithms extract road network information by setting a unified threshold, but the density of vehicle trajectory data is very different, which makes it difficult to apply; 2) Many algorithms only use vehicle trajectory shape, without considering the direction of the trajectory, affecting road topological accuracy; 3) Most of the road network information extracted by the algorithm contains too much redundant data and is not easy to simplify. Therefore, this paper designs an adaptive radius centroid drift clustering method, which can automatically adjust clustering parameters according to trajectory density and road width and realize road topological connection by using trajectory direction.

\* Corresponding author

## 2. METHODOLOGY

This paper designs an anti-noise and easy clustering method, which can effectively extract the road network using floating car trajectory data (Figure 1). The specific steps are as follows: 1) Preprocessing the original floating car trajectory data, including projection coordinate transformation, calculation the moving direction of two neighboring points and removal of stagnation points. 2) For the processed trajectory point data, the improved centroid drift clustering algorithm is used to extract the network skeleton points (clustering centers). Traditional centroid drift clustering algorithm uses fixed radius to cluster, which is difficult to apply to roads with different widths. This paper improves the traditional algorithm and realizes the adaptive adjustment of clustering radius according to road width. 3) Determine the direction of road network skeleton points. Wavelet clustering algorithm is used to cluster the direction set of the clustering trajectory point set (obtained from step 2). The clustering centers of direction set are calculated and assigned to the corresponding road network skeleton points (clustering centers). 4) Road network skeleton point connection. According to the clustering radius and skeleton point direction, the road network skeleton points are connected recursively to generate road network data.

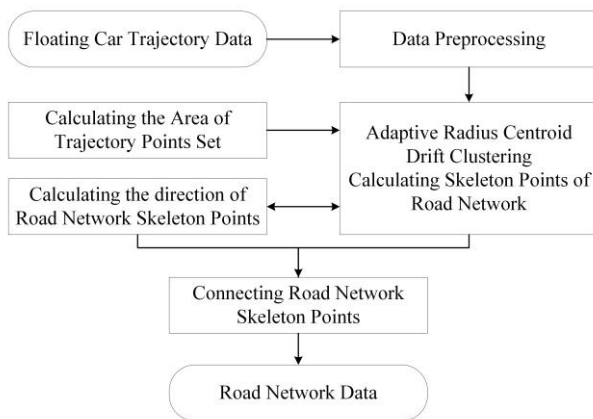


Figure 1. The flow chart of road network extraction

### 2.1 Data Preprocessing

Because the floating car uploads data at intervals, only one point of data is uploaded near the corner. Which results in the dispersed direction of track points at road corners (this is the main direction is not obvious). In order to solve this problem, this paper defines the direction from one data upload point to the next time interval data upload point as the interval moving direction of the point. There are two directions for each floating car trajectory point, instantaneous moving direction  $V_{instantaneous}$  and interval moving direction  $V_{interval}$ . The absolute value of the difference between two directions on the straight road  $\Delta a = |V_{instantaneous} - V_{interval}|$  is analyzed to obtain the maximum value  $\Delta a_{max}$  of  $\Delta a$  which can reflect the characteristics of the straight road. Add a field (*straight\_flag*) to the original data, if  $\Delta a \leq \Delta a_{max}$  then *straight\_flag* = 1, otherwise *straight\_flag* = 0. In this way, the *straight\_flag* = 1 floating car trajectory point set retains only the road extension direction at the intersections. (Figure 2).

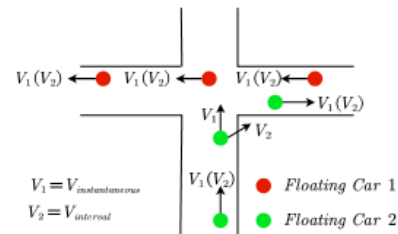


Figure 2. The direction of  $V_{instantaneous}$  and  $V_{interval}$

### (2) Removal of stagnation points

When a floating car is waiting at a traffic light intersection or parking in congestion, it will generate many repetitive (or very small location difference) data points nearby, which are of little significance in road network extraction and will increase the amount of calculation. Therefore, the distance between a trajectory data point and the previous trajectory data point is calculated. If the distance is less than 5 m (about GPS civilian precision), the trajectory data point is deleted. In addition, the original data is geographic coordinate system, in order to facilitate calculation, it is converted into projection coordinate system.

### 2.2 Extracting Skeleton Points of Road Network

In this paper, an adaptive radius trajectory point clustering method is proposed. The clustering radius is adaptively adjusted according to the road width by using the ratio of the area of clustering trajectory point set to the area of clustering circle.

#### (1) Calculating the Area of Trajectory Point Set

In order to improve the calculation efficiency, a method of calculating the area of trajectory point set with only time complexity  $O(n)$  is designed. For a set of trajectory points, the maximum  $x_{max}$  and minimum  $x_{min}$  abscissa coordinates, the maximum  $y_{max}$  and minimum  $y_{min}$  longitudinal coordinates and the number of trajectory points  $n$  are calculated. In this way, a rectangle  $Q$  can be determined by point  $(x_{min}, y_{min})$  and point  $(x_{max}, y_{max})$ , and then the rectangle  $Q$  is divided into grids (Figure 3). The size of each grid is as follows:

$$\begin{cases} k = \lfloor \frac{n^{1/2}}{2} \rfloor + 1 \\ \Delta x = \frac{x_{max} - x_{min}}{k} \\ \Delta y = \frac{y_{max} - y_{min}}{k} \end{cases} \quad (1)$$

This partition ensures an average of 1 to 4 points in each grid. The number  $N_{ij} (1 \leq i, j \leq k)$  of trajectory points in each grid is counted, then the number  $m$  of grids which  $N_{ij} \geq \lfloor \frac{1}{2} \cdot \frac{n}{k^2} \rfloor$  is counted. This to say, if the number of trajectory points in the grid is greater than half of the average number of trajectory points in each grid, the number of trajectory points can be counted. The area of the set of trajectory points is  $S = m \cdot \Delta x \cdot \Delta y$ .

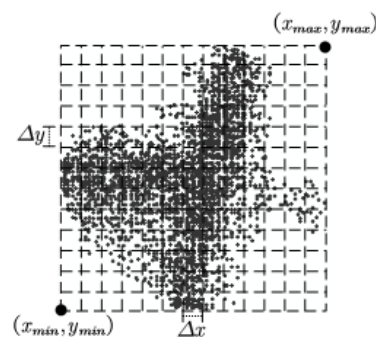


Figure 3. The calculation of trajectory points set's area

## (2) Adaptive Radius Centroid Drift Clustering

The basic principle of trajectory points clustering is to divide the trajectory points set  $I = \{P_1, P_2, P_3, \dots, P_n\}$  into  $l (l \geq 1)$  smaller trajectory points set  $I = \{S_1, S_2, S_3, \dots, S_l\}$  by clustering.  $S_i = \{P_{i1}, P_{i2}, P_{i3}, \dots, P_{ik}\}, S_i \cap S_j = \emptyset, 1 \leq i, j \leq l$ . Then the cluster centers  $C_i (1 \leq i \leq l)$  of each trajectory points set are obtained, and the cluster centers  $G = \{C_1, C_2, C_3, \dots, C_l\}$  are the skeleton points of the road network.

This method can adaptively change the clustering radius according to the road width, and get the skeleton points set  $G = \{C_1, C_2, C_3, \dots, C_l\}$  and the corresponding clustering radius set  $\Phi = \{R_{C_1}, R_{C_2}, R_{C_3}, \dots, R_{C_l}\}$  of the road network. Firstly, given the initial clustering radius  $R_{start}$ , the minimum clustering radius  $R_{min}$ , the step size  $R_{dt}$  of each change of the clustering radius, the minimum number  $N_{min}$  of clustering trajectory points, judging whether the clustering has a stable drift distance threshold  $d_{stable}$  and whether the clustering result accepts the ratio range *Range* between the area  $S_{points}$  of clustering trajectory point set and the area  $S_R$  of clustering circle. If the set  $I$  of trajectory points is not empty, a trajectory point is selected arbitrarily as the initial clustering center  $O_{start} = (x_{start}, y_{start})$  in the set  $I$  of trajectory points, and then the trajectory points set  $ST_0 = \{P_{01}, P_{02}, P_{03}, \dots, P_{0k}\}$  in the neighborhood  $U(O_{start}, R = R_{start})$  with radius  $R = R_{start}$  is counted. calculate the new clustering center of the set of trajectory points. Among them,

$$\begin{cases} x_1 = \frac{1}{k} \sum_{i=1}^k x_{P_{0i}} \\ y_1 = \frac{1}{k} \sum_{i=1}^k y_{P_{0i}} \end{cases} \quad (2)$$

The distance  $d_{drift}$  of between points  $O_{start}$  and  $O_{end}$  is calculated. If  $d_{drift} > d_{stable}$  and the number of elements  $Card(ST_0) \geq N_{min}$  in the trajectory points set  $ST_0$ , the clustering is not stable. Let  $O_{start} = O_{end}$ , the clustering is restarted again (Figure 4a).

If  $d_{drift} \leq d_{stable}$  the number of elements  $Card(ST_0) \geq N_{min}$  in the trajectory points set  $ST_0$ , the clustering is stable. The area  $S_{ST_0}$  of the trajectory points set  $ST_0$  is calculated. If  $\frac{S_{ST_0}}{S_R} \in Range$ , the trajectory points set  $ST_0$  is called Cluster Trajectory Point Set (CTPS  $ST_0$ ), and the cluster center  $O_{end}$  is added to the skeleton points set of road network and the cluster radius  $R$  is added to the cluster radius set  $\Phi$ . The directions of the skeleton point of road network is determined by using the CTPS  $ST_0$ , and the trajectory points in the CTPS  $ST_0$  are deleted from the total trajectory points set. If the trajectory point set is not empty, the centroid drift clustering is restarted; if  $\frac{S_{ST_0}}{S_R} > \max(Range)$ , the cluster radius is enlarged  $R = R + R_{dt}$ . Let  $O_{start} = O_{end}$ , the clustering is restarted (Figure 4b); if  $\frac{S_{ST_0}}{S_R} < \min(Range)$  and  $R > R_{min}$ , the radius of clustering is reduced  $R = R - R_{dt}$ , the clustering is restarted. if  $\frac{S_{ST_0}}{S_R} < \min(Range)$  but  $R \leq R_{min}$ , deleting the trajectory points in the CTPS  $ST_0$  from the total trajectory points set, the clustering is restarted (Figure 4c).

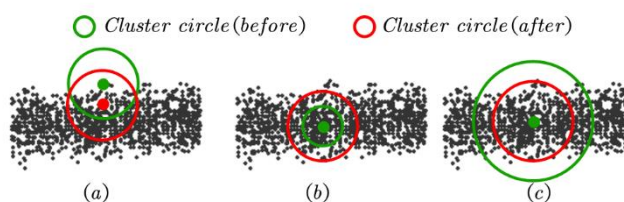


Figure 4. The adaptive radius centroid drift clustering

If the number of elements  $Card(ST_0) \geq N_{min}$  in the trajectory points set  $ST_0$ , the clustering results are usually discrete points. The trajectory points in the  $ST_0$  are deleted from the total trajectory points set  $I$ . If the trajectory points set  $I$  is not empty, the centroid drift clustering is restarted.

## 2.3 Data Preprocessing

After obtaining a CTPS  $S$  and the corresponding road network skeleton point (cluster center), the direction set of the skeleton point should be extracted in order to connect the skeleton points. However, the direction data of CTPS  $S$  is not only large in number but also contains a lot of noise. Therefore, a fast, strong anti-noise and less manual intervention clustering method is needed to obtain the skeleton point direction. WaveCluster is very suitable for the clustering of one-dimensional and huge trajectory point direction data sets, so this paper uses the wavelet clustering algorithm to cluster the trajectory point direction set to determine the direction of the road network skeleton points (Sheikholeslami et al., 1998). The specific steps are as follows:

(1) Generating quantitative feature space.

The trajectory direction of the floating car is within  $[0^\circ, 360^\circ)$ . From  $0^\circ$  the beginning, the whole angle space is divided into 60 even intervals. For a CTPS  $S$ , if the number of *straight\_flag* = 1 elements in points set  $S$  accounts for more than 50% of the total number of elements, the direction set  $A$  of the CTPS  $S$  is generated using the elements' direction which is *straight\_flag* = 1; Otherwise, the direction set  $A$  of  $S$  is generated using all elements' direction. The data points in set  $A$  are placed in the partitioned angle space, and then the number  $C_i (1 \leq i \leq 60)$  of data points in each interval is counted separately. Because the number of elements in each CTPS is not uniform, in order to set a uniform threshold for clustering, the quantitative feature space  $H_i = \frac{C_i}{C} \times 100$  is generated by using the percentage of the number  $C_i$  of data points in each interval to the total number  $C$  of data points (Figure 5a).

(2) Mean filtering. In order to better overcome the impact of noise on clustering, a mean filter is used to reduce the impact of noise on clustering in the experiment. The angle space is actually a cyclically continuous space, so when filtering the boundary position, it fills in in a periodic manner. And a new quantized feature space  $\{F_i\}$  is generated (Figure 5b).

(3) Wavelet transform is applied. For quantized feature space  $\{F_i\}$ , Haar discrete wavelet transform of different scales is applied, and periodic mode is used for wavelet transform due to the characteristics of angle space. After discrete wavelet transform, approximate coefficients form a new feature space  $\{T_i\}$  (Figure 5c).

(4) Clustering. Zeroing (denoising) is less than the given threshold 3.3 (the percentage of the number of data points in the average interval). In  $\{T_i\}$ , the non-zero values are detected, which are connected by the same cluster, and the sequence numbers of cluster categories are given in turn. Since the angular space is cyclically continuous, the  $\{T_i\}$  is considered to be connected at the beginning and end of the detection. (Figure 5d).

(5) Construct a lookup table. Through the unique correspondence between the clustering trajectory point direction data and the quantized feature space  $\{F_i\}$ , the new feature space  $\{T_i\}$ , and the clustering sequence number, the clustering category sequence number of each data point in the clustering trajectory point direction set is determined. The data points of the sequence number of the same clustering category constitute a clustering direction set of the direction set  $A$  of the clustering trajectory points.

(6) Calculate the skeleton point direction. Through step (1) - (5), several clustering direction sets  $A_1, A_2, A_3, \dots, A_n$  of the direction set  $A$  can be obtained, among which  $A_i = \{a_1, a_2, a_3, \dots, a_k\} (1 \leq i \leq n)$ . For any clustering direction set  $A_i$ , the sum vector  $(x_{A_i}, y_{A_i})$  of unit directional vectors  $(\cos a_j, \sin a_j)$  of all angle elements  $a_j (1 \leq j \leq k)$  in  $A_i$  is

calculated. The direction  $a_{A_i}$  of sum vector  $(x_{A_i}, y_{A_i})$  is regarded as the direction of the set  $A_i$ . Find out the direction of all clustering direction sets of the set  $A$  and form a new direction set  $\{a_{A_1}, a_{A_2}, a_{A_3}, \dots, a_{A_n}\}$ . That is the direction set of a skeleton point.

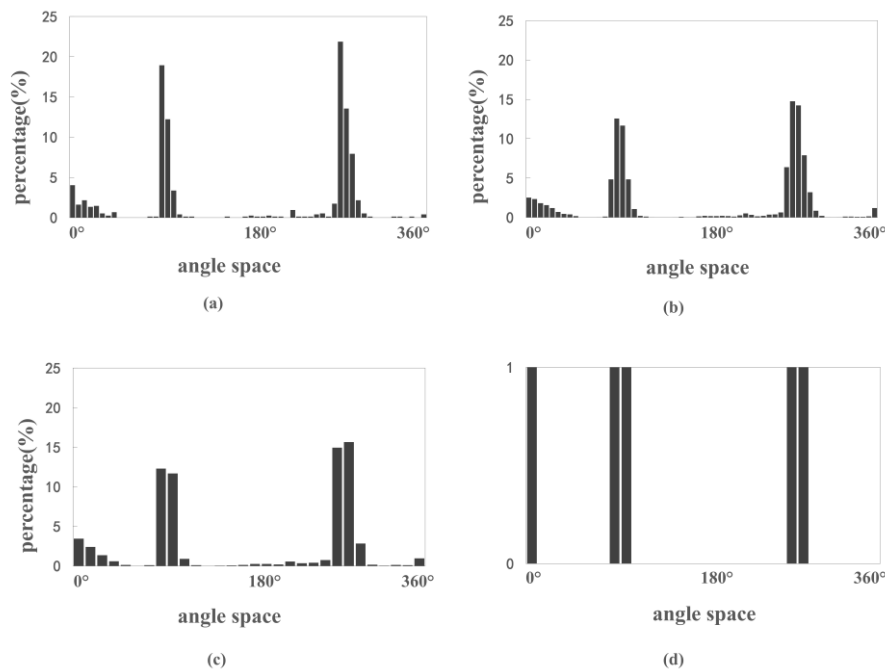


Figure 5. The calculation of skeleton point's direction

#### 2.4 Connecting Road Network Skeleton Points

The number of directions of skeleton points is used to divide sections. If the number of directions of a skeleton point is greater than or equal to 3, the skeleton point may be an intersection in the road network; if the number of directions is 2, it may be a common point in the middle of the road; if the number of directions is 1, it may be an end of a road. These intersections and endpoints can be used to divide the road network. Neighbor skeleton points are searched by clustering radius of skeleton points, and the direction of skeleton points is used to connect the searched neighbor skeleton points, iterating repeatedly until the intersection or the end of the road stops, forming a complete road. Specific steps are as follows:

Calculating the maximum clustering radius  $R_{max}$  in the clustering radius set  $\Phi$ , it is easy to know that the nearest connection point of any road network skeleton point  $C_i$  must be in the circle of the radius  $R_{C_i} + R_{max}$  of it; According to the  $\Delta a$  distribution statistics, the angle error range of searching the next neighboring connection point along a certain direction is within  $\pm \Delta a_{max}$ . Arbitrarily determining a road network skeleton point  $C_i$  as the starting point:

1) If the number  $N_{Dir}$  of directions of the point  $C_i$  is 1. In the road network skeleton points set, along the range of this direction  $\pm \Delta a_{max}$ , finding the nearest neighbor connection point in the circle of radius  $R_{C_i} + R_{max}$ . After the nearest neighbor connection point  $C_j$  is found, deleting this direction of points  $C_i$ , calculating the direction  $a_{C_j \rightarrow C_i}$  of point  $C_j$  to point  $C_i$ , and deleting these directions of point  $C_j$  in the range  $a_{C_j \rightarrow C_i} \pm \Delta a_{max}$ . Then finding the next neighbor connection point of the point  $C_j$ , iterate repeatedly until a path is generated when the next

nearest neighbor connection point is not found, or the number of directions of the next nearest neighbor connection point is greater than or equal to 3 or 1. Removing empty skeleton points whose direction set is empty from road network skeleton points set  $G$  (Figure 6a).

2) If the number  $N_{Dir}$  of directions of the points  $C_i$  is 2, then search along the two directions of the points and other synchronization step 1 to generate a path (Figure 6b).

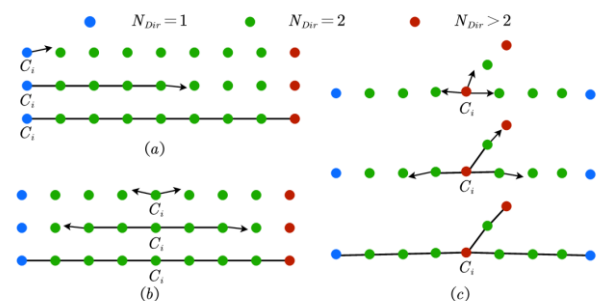


Figure 6. The connection of skeleton points

3) If the number  $N_{Dir}$  of directions of the point  $C_i$  is greater than 2, then starting from point  $C_i$ , we recursively search for  $N$  directions of point  $C_i$  and generate  $M$  paths ( $M < N$ ), because some directions may not have adjacent points. The specific operation is also the same step 1 (Figure 6c).

Repeatedly perform the above operations until there is no data point in the network skeleton points set  $G$  and the network data is output.

### 3. EXPERIMENTAL ANALYSES

#### 3.1 Experimental Data Set and Environment

The floating car trajectory data used in our work are the one-day floating car trajectory of Shenzhen on Thursday, December 1, 2016. Floating vehicle trajectory data include vehicle ID, vehicle instantaneous direction, GPS time, GPS longitude and latitude, etc. Futian District of Shenzhen and its surrounding areas were selected as experimental areas. The average sampling interval of trajectory points was 20s. There were 64,887,486 trajectory points, and 38,387,446 trajectory points after preprocessing. In the P4/4GB/2GB/Win7 environment, experimental verification of the algorithm is carried out based on ArcGIS platform and python programming language.

For purpose of qualitatively evaluating the experimental results, we selected GF-2 satellite remote sensing images for superposition and contrast analysis. Furthermore, in order to verify the experimental results, we use OSM data as the ground truth. Among them, non-motorized lanes are deleted from OSM data, and multi-lanes are not merged, so the recall rate of all algorithms is low.

#### 3.2 Explanation of Algorithmic Parameters

(1) Selection of  $\Delta a_{max}$ . In the experiment, trajectory points of straight lanes, each of which has a length greater than 5 km, were selected from the total floating cars trajectory data. A total of 3,284,534 trajectory points (about 5% of the total data) were analyzed statistically. It was found that 75% of the absolute value  $\Delta a$  fell within  $30^\circ$ . When  $\Delta a > 30^\circ$ , the curve is very smooth and can not reflect the characteristics of the straight lane. This article  $\Delta a_{max} = 30^\circ$  is taken (Figure 7).

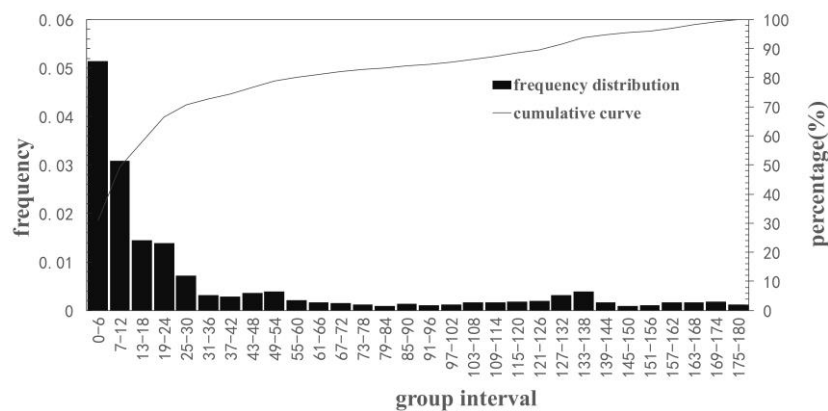


Figure 7. The distribution statistics of  $\Delta a$

(2) Selection of clustering parameters. The clustering parameters used in the experiment are as follows:  $R_{start} = 28m$ ,  $R_{min} = 5m$ ,  $R_{dt} = 0.5m$ ,  $N_{min} = 10$ ,  $d_{stable} = 1m$ ,  $Range = [0.5, 0.6]$ .  $R_{start}, R_{min}$  according to the width of common motor lanes in China;  $R_{dt}, d_{stable}$  according to the error of road digitization;  $Range = [0.5, 0.6]$  ensuring the clustering radius and road width are approximately equal.

#### 3.3 Analysis of experimental results

The skeleton point data (Figure 8a) and the road network data (Figure 8b) of the road network in the experimental area are obtained by using this method. Qualitative evaluation is made by superimposing the extracted road network data with remote sensing image maps and OSM vector maps (Figure 8c and Figure 8d). It is shown that the overall results of road network data cover the roads in the experimental area with high accuracy; Road geometry and topology information can be extracted more accurately in the area of road intersection; And the density difference of track points under different road grades can be handled adaptively. It can extract the road data that is near and control the minimum width of the extraction road by setting the minimum clustering radius.

In order to empirically evaluate the proposed approach, we adopt three indicators, namely precision, recall, and F-value, computed using the following formula, to quantitatively evaluate our method and the methods proposed by Davies et. al (Davies et al., 2006) and Kuntzsch et. al. (Kuntzsch et al., 2016)

$$\left\{ \begin{array}{l} Precision = \frac{\text{correctly roads}}{\text{all extraction roads}} \\ Recall = \frac{\text{correctly roads}}{\text{all roads in OSM}} \\ F - value = 2 * \frac{Precision * Recall}{Precision + Recall} \end{array} \right. \quad (3)$$

Taking OSM road vector data of Shenzhen city as ground truth, the accuracy (P), recall rate (R) and F-value of road length extracted by each method are counted by establishing matching threshold of 5 m, 10 m and 20 m respectively (Table 1). From Table 1, we can see that the extraction results of this method are better than those of the other two methods in accuracy. The road extracted by our method is represented by skeleton points, and there is less data redundancy compared with the other two methods. With the increase of matching threshold, this redundant information may be calculated into recall rate. Therefore, the road network recall rate we proposed is on the low side.

Table 1. The evaluation of experimental results

	Matching Threshold (%)								
	P(5m)	R(5m)	F(5m)	P(10m)	R(10m)	F(10m)	P(20m)	R(20m)	F(20m)
Davies	44.10	34.50	38.71	58.73	37.52	45.79	74.01	43.41	54.72
Kuntzsch	42.60	<b>34.54</b>	38.15	60.48	<b>38.32</b>	46.92	76.71	<b>46.46</b>	57.87
Our	<b>53.25</b>	32.80	<b>40.59</b>	<b>64.69</b>	37.39	<b>47.39</b>	<b>83.25</b>	44.79	<b>58.24</b>

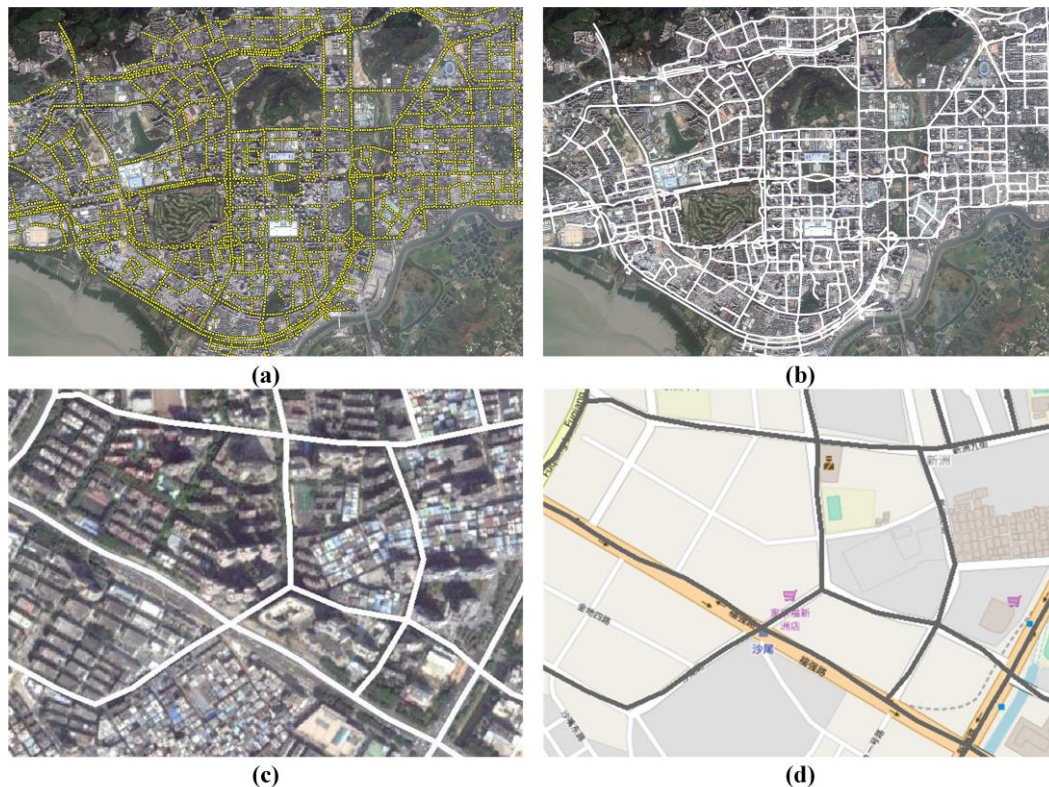


Figure 8. The results of experiment and qualitative evaluation of road data

#### 4. CONCLUSIONS

In order to extract road network data efficiently and accurately from the trajectory data of floating cars, an adaptive radius centroid drift clustering method considering the direction information of floating cars is proposed in this paper. The experimental analysis is carried out based on the data of floating car trajectory points in Shenzhen at a certain time. OSM road network vector data is used as the ground truth, and the experimental results of other extraction methods are compared qualitatively and quantitatively. The results show that the geometric and topological accuracy of the road network data extracted by this method has been significantly improved, and the algorithm of this method is easy to implement, which is very suitable for big data processing.

At the same time, there are still many aspects to be further studied and perfected: 1) This paper only extracts partial geometric and topological information of road network. The extraction of attribute information such as road width, road name, lane and traffic, needs to be further explored; 2) For road extraction with three-dimensional hierarchical structure such as complex intersections and overpasses, it still needs to be further studied; 3) How to fuse the floating car spatio-temporal trajectory data with other related multi-source heterogeneous large data and perceive more urban traffic fluctuations will be the next research focus.

#### ACKNOWLEDGEMENTS

This research is supported by the State Key Research Development Program of China (GrantNo. 2016YFB0502104)

#### REFERENCES

Ahmed, M. and Wenk, C., 2012. Constructing street networks from GPS trajectories, European Symposium on Algorithms. Springer, pp. 60-71.

Ahmed, M., Karagiorgou, S., Pfoser, D. and Wenk, C., 2015. A comparison and evaluation of map construction algorithms using vehicle tracking data. *GeoInformatica*, 19(3): 601-632.

Biagioni, J. and Eriksson, J., 2012. Map inference in the face of noise and disparity, Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, pp. 79-88.

Cao, L. and Krumm, J., 2009. From GPS traces to a routable road map, Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, pp. 3-12.

Chen, C. and Cheng, Y., 2008. Roads digital map generation with multi-track GPS data, 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing. IEEE, pp. 508-511.

Chen, Y. and Krumm, J., 2010. Probabilistic modeling of traffic lanes from GPS traces, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 81-88.

Davies, J.J., Beresford, A.R. and Hopper, A., 2006. Scalable, distributed, real-time map generation. *IEEE Pervasive Computing*, 5(4): 47-54.

Deng, M. et al., 2018. Generating urban road intersection models from low-frequency GPS trajectory data. *International Journal of Geographical Information Science*, 32(12): 2337-2361.

Edelkamp, S. and Schrödl, S., 2003. Route planning and map inference with global positioning traces. *Computer science in*

perspective. Springer, pp. 128-151.

Guo, T., Iwamura, K. and Koga, M., 2007. Towards high accuracy road maps generation from massive GPS traces data, 2007 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 667-670.

Jiang, Y., Li, X., Li, X. and Sun, J., 2012. Geometrical characteristics extraction and accuracy analysis of road network based on vehicle trajectory data. *Journal of Geo-information Science*, 14(2): 165-170.

Karagiorgou, S. and Pfoser, D., 2012. On vehicle tracking data-based road network generation, *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM, pp. 89-98.

Karagiorgou, S., Pfoser, D. and Skoutas, D., 2013. Segmentation-based road network construction, *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, pp. 460-463.

Kuntzsch, C., Sester, M. and Brenner, C., 2016. Generative models for road network reconstruction. *International Journal of Geographical Information Science*, 30(5): 1012-1039.

Li, J. et al., 2019. An Automatic Extraction Method of Coach Operation Information from Historical Trajectory Data. *Journal of Advanced Transportation*, 2019.

Li, J., Qin, Q., Han, J., Tang, L.A. and Lei, K.H., 2015. Mining trajectory data and geotagged data in social media for road map inference. *Transactions in GIS*, 19(1): 1-18.

Schroedl, S., Wagstaff, K., Rogers, S., Langley, P. and Wilson, C., 2004. Mining GPS traces for map refinement. *Data mining and knowledge Discovery*, 9(1): 59-87.

Sheikholeslami, G., Chatterjee, S. and Zhang, A., 1998. Wavecluster: A multi-resolution clustering approach for very large spatial databases, *VLDB*, pp. 428-439.

Steiner, A. and Leonhardt, A., 2011. Map-generation algorithm using low-frequency vehicle position data, *TRB 90th Annual Meeting Compendium of Papers*. Transportation Research Board, pp. 1-17.

Worrall, S. and Nebot, E., 2007. Automated process for generating digitised maps through GPS data compression, *Australasian Conference on Robotics and Automation*. Brisbane: ACRA.

Zhang, L., Thiemann, F. and Sester, M., 2010. Integration of GPS traces with road map, *Proceedings of the Third International Workshop on Computational Transportation Science*. ACM, pp. 17-22.