# Research Article
# Route Reconstruction from Floating Car Data with Low Sampling Rate Based on Feature Matching

Jun Li, Liang-Hui Xie and Xin-Jun Lai
Research Center of Intelligent Transportation System, Sun Yat-Sen University, China

**Abstract:** Floating Car Technology is widely used to collect traffic information. To reappear the actual trips of drivers, a bi-level probability method is proposed to reconstruct routes from floating car data, address two issues: the first one is incorrect map matching caused by GPS accuracy and complexity of road network; and the second one is the link missing duo the low sampling rate of floating car. Using confidence region, GPS points are divided into three types: zero-feature points with no feature matching, single-feature points that have a unique matched link or node and multiple-features points that have multiple features. The matching probability for GPS points to the possible feature according to the distance between GPS point and the link, which is assumed to be normal distribution. The missing links between two single-feature points are reconstructed by the shortest path algorithm with consideration of the probability of multiple matched features. A case study of Guangzhou floating car data shows that the proposed method can produce reasonable routes on complicated urban road network.

**Keywords:** Feature matching, floating car, map matching, route choice

## INTRODUCTION

The use of Floating Car Technology (FCT) to collect traffic information is becoming increasingly popular among traffic managers and researchers. Floating Car Data (FCD) which records real-time position, direction and speed information can help to produce trajectories of the vehicles. Unfortunately, it is not precise due to the measurement error caused by the limited GPS accuracy and the link missing caused by the sampling rate. The complexity of urban road network and the occlusion of tall buildings and trees severely reduce the GPS positioning accuracy. On the other hand, for floating car, there is a large numbers of data need to be stored, so it is impossible to get high sampling rate under the existing system. These make it difficult to obtain the routes of vehicles in urban road network.

In this study, we propose a method of route reconstruction from Floating Car Data with low sampling rate. Based on an objective as route, the GPS points don't need to be matched to points or line segments immediately. We define three types of points by confidence region: zero-feature points with no feature matching, single-feature points that have a unique matched link or node and multiple-features points that have multiple features. It helps to reconstruct possible paths for a set of GPS trip data. Then we get the most probable path of the candidates by bi-level probabilistic model. The experiment of

Guangzhou Floating car verifies the effectiveness of the proposed method.

## LITERATURE REVIEW

Route reconstruction is a process of map matching. Traditional map matching algorithms is to identify the correct link among the candidate links and determine the vehicle location on that link (Greenfeld, 2002). And the applying of map matching algorithm is mostly for vehicle navigation with high sampling rate (Christopher *et al.*, 2000; Ochieng *et al.*, 2003; Yang *et al.*, 2003; Nagendra *et al.*, 2009), so the accuracy rate is relatively high. But for floating car data with low sampling, it becomes more difficult.

When point-to-point matching method is firstly used to the map matching, it only considered the geographic information. Map the GPS point to the closest node or shape point in the network, the closer to the GPS trick the more easily be matched, without considering other information, so bring great error (Kim, 1996). Point-to-line matching, mapping the GPS point to the closest arc in the network (minimum distance from the point to the curve), is not stable (Taylor *et al.*, 2001). For the curve to curve matching algorithm, GPS points and candidate node are jointed into piece-wise linear curves respectively. Through distance calculation the closed curve is selected, then matching GPS points to this curve (White *et al.*, 2000). But this method is so sensitive to exceptional value to
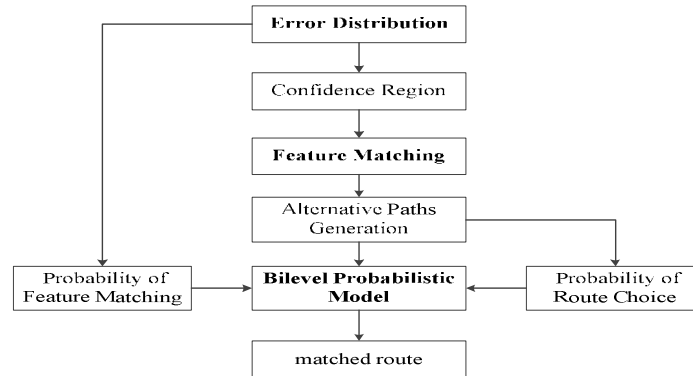
Fig. 1: Flowchart of route reconstruction algorithm

generate outrageous results sometimes (Quddus *et al.*, 2006).

Weight-based algorithms primarily consider the similarity in vehicle heading and bearing of the link, proximity of a point to a link (White *et al.*, 2000). Enhanced weight-based algorithms are including the performance of weights for turn-restriction at intersection; link connectively, roadway classification and road infrastructure information (Srinivasan *et al.*, 2003; Blazquez and Vonderohe, 2005). Weight setting increases the use of network information and GPS data, as well as the complexity, especially for the intersection, every weigh value will cause great impact on the results.

Chawath promotes a segment-based matching method taking into account every different confidence values. It outlines a 90% confidence region but the error distribution is not further used. Confidence values are assigned for different sampling points, giving priority to match the high-confidence segments and then match low-confidence segments using previously matched results. The method does good performance when the sampling rate is high, otherwise the accuracy will decrease significantly (Chawathe, 2007).

The global methods aim to match the entire trajectory with the road network. It aims to find a minimum weight path based on edit distance or Fréchet distance (Alt *et al.*, 2003; Yin and Wolfson, 2004; Brakatsoulas *et al.*, 2005). Moreover, Yin brings the spatial-temporal information in the trajectories, it generates candidate sequences and takes the highest score sequence matching results as the final matching results (Yin *et al.*, 2009). Wei *et al.* (2011) propose an algorithm framework, which incorporates curve matching and probabilistic analysis modules. But in the processing interval of the above two methods, the initial position of the vehicle can't ensure a good matching, which will lead to the matching error of follow-up point.

The generated results of above methods are matched point or line segment, or matched paths in the processing interval of the global methods, even if the sections or intervals matching with high accuracy, when connected into a full paths, the accuracy rate will decline in geometric level, not to mention that the matching accuracy of above methods in intersection and elevated roads are not high enough.

## THE PROPOSED ALGORITHM

We propose a framework of route reconstruction algorithm. The entire block diagram is as follows: (Fig. 1) According to the error distribution of distance between GPS point and the matched road section, we can get the probability for GPS points matching to the candidate matched feature. We can also determine the confidence region which judges whether a feature is candidate. The shortest path algorithm is used to connect the candidates and the route choice model is used to calculate the probability of each candidate route to be chosen. A bi-level probabilistic model is used to get the most probable route of GPS points which is the final matched route.

In the following sections, the blocks will be elaborated in details.

**Error distribution:** There are two major sources of errors between the location of the positioning point and the actual position of vehicle. The first is GPS inherent error that results from the inevitable factors in GPS method. GPS inherent error is subject to normal distribution. The second source of errors is tree cover, urban canyons and other problems. The error arises from these effects can also be described by normal distribution. Therefore, GPS error can be modeled using random variables with normal distribution.

Denote the distance between GPS point and the actual position of vehicle as $x$, the probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2\right] \quad (1)$$

where,
$\mu$ = The expectation and
$\sigma$ = The standard deviation of the normal distribution.

The heading of vehicle is used to judge whether the distance is positive or negative. The vehicles run by the
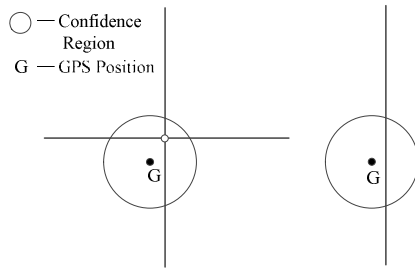
Fig. 2: Two types of single-feature point

Table 1: Graphic symbols and results of multiple features matching

| Graphic symbols | Results of matching |
| --- | --- |
|  | Line segment*2 |
|  | Node*2 |
|  | Node*1<br>Line segment*1 |
| …… | The superposition of above results |

right in China, so we define that the distance is positive when the GPS point is at the right of line segment.

**Confidence region:** This study uses a circular confidence region around a position fix based on GPS error model. Map features within the confidence region are taken as the candidate features. In order to cover the correct feature as possible, we choose the 99.7% confidence region and the radius of confidence region is $3\sigma$.

**Feature matching:** Every feature in the digital map in confidence region has a chance to be matched. These features maybe point, line, arc and so on. Because the method in this study is not eager to match GPS point to the actual position of vehicle, these features can be fuzzy matching.

According to the different kinds of features in confidence region, it can be divided into following tree types:

- **Zero-feature matching:** If the confidence region of a GPS point does not contain any map feature, it would be passed and not be calculated in the following algorithm. These GPS points are called zero-feature points.
- **Single-feature matching:** After confirming the confidence region, when there is only one line segment in confidence region, it is inferred that the actual position of vehicle is locating on this line segment.

When there are several candidate line segments in confidence region, which have a same node (no matter the node is origin or destination for them), this method only extract this node while ignore the actual vehicle position. Because no matter which way the driver chooses, the vehicle must pass through the crossing point.

In the two situations as Fig. 2 shows, the floating car must pass through the intersection in Fig. 2a and must pass through the road in Fig. 2b. GPS points are matched to the unique node or line segment without any other possibility. These GPS points are defined as single-feature points.

In addition to line segment matching, the complex intersection matching is simplified into a crossing point

and makes it possible for the GPS points which near the intersection become single-feature point. Vehicle trajectories must pass through the single-feature points.

- **Multiple-features matching:** Besides the two situations above, sometimes there are several line segments in confidence region, which do not have the same node. But we can simplify them based on the former two situations, just as Table 1 shows.

We can calculate the correct matching probability of each GPS feature according to the probability density function of GPS error.

**Route reconstruction:** After matching each GPS point onto digital map, these features form a sequence. Different GPS points are connected by shortest paths and then make up alternative paths, as the Fig. 3.

If the impedance of one candidate is far greater than the minimum impedance of all candidates between adjoining GPS points, this candidate is unreasonable (Leurent, 1997). The threshold value is named elongation ratio which needs to be estimated. The unreasonable paths should be excluded, just as the dash lines in Fig. 3.

Route choice model of drivers is used in probability calculation of path matching. Here we use the multinomial logit route choice model (MNL) under relative impedance. The probability of path *m* chosen is:

$$P_m = \frac{\exp(-\theta L_m / L_{\min})}{\sum\limits_{n=1}^{N} \exp(-\theta L_n / L_{\min})} \tag{2}$$

where,
$\theta$ = The parameter of model
$L_m$ = The distance impedance of route m
$L_{\min}$ = The minimum distance impedance of route $N$

Because the elevated roads have higher service level, drivers are more inclined to it. When calculate the distance impedance of elevated road, it should multiply
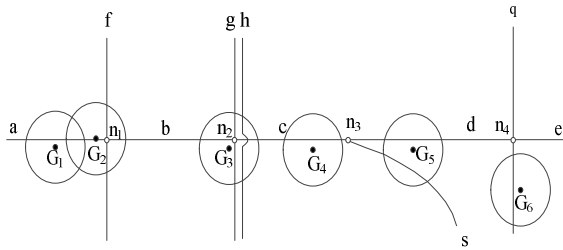
Fig. 3: Process of route reconstruction

by conversion coefficient which is a parameter need to be estimated.

The path between adjoining single-feature points has been confirmed since they had been matched to certain features. The possibility of a group of consecutive multiple-feature points matching to a certain path $k$ is decided by bi-level probabilistic model. It composes of two aspects: One is the GPS error distribution probability of corresponding matching feature defined as $p_{tk}$. The other is the choice probability of path $k$ defined as $P_k$. The probability of correct travel track on path $k$ is:

$$P_k^{'} = P_k \sum_{t=1}^{T} p_{tk} \tag{3}$$

where,
$T$ = The number of GPS points

The highest probability path is supposed to the matched path. All the certain paths and highest probability paths compose a full trip route which is the result of route reconstruction.

## EXPERIMENT

There are more than 17,000 taxies as floating car with data upload interval of 20-120s in Guangzhou. Over 30,000,000 data records on July 6th, 2011 are used in experiment. With these data, we accomplish error statistics, type statistics of points and route reconstruction in a complex road network.

**Error statistics:** Extract 30,000 taxi GPS data among the OD pairs and distinguish trip track of each GPS data group with handwork matching (we think it is very accurate). Then we calculate the distances between GPS points and matched line segments.

We get normal distribution curve by fitting them with normal distribution as Eq. (1). The expectation $\mu = $ 5.5846 m and the standard deviation $\sigma = $ 14.876 m.

According to the definition of positive-negative distance, because of running by the right, the expectation of result is positive and this is consistent with the practice.
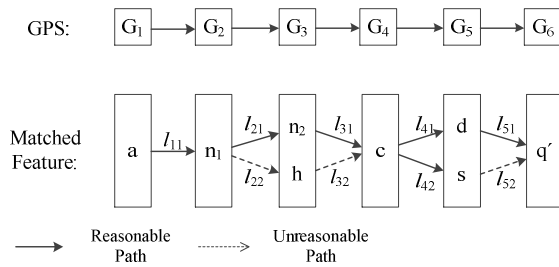
Table 2: The result of point type statistics

| Feature type | Point type | Number | Percentage |
|---|---|---|---|
| No Feature | Zero-feature | 134 | 1.36% |
| Single line segment | Single-feature | 4715 | 47.68% |
| Single intersection | Single-feature | 2904 | 29.37% |
| Two features | Multiple-features | 468 | 4.73% |
| Three features | Multiple-features | 1250 | 12.65% |
| Four features | Multiple-features | 376 | 3.80% |
| Five features | Multiple-features | 3 | 0.03% |
| Six features | Multiple-features | 38 | 0.38% |
| Total | | 9888 | 100% |

We are meant to choose $3\sigma$ as the radius of confidence region and the confidence region is [-39, 50]. For convenience, we set confidence region as [-50, 50]. As a result, candidate features can be searched on a circular whose center is GPS point and whose radius is 50 meters.

**Type statistics of points:** The ratio of single-feature points has a strong impact on the efficiency of route reconstruction, so type statistics is quite important. We search for candidate features within confidence region and process 9888 GPS data records under the definition of feature matching which has discussed before. The statistical result is as Table 2 shows.

It is indicated that GPS points with unique line segment feature are 47.68%, with unique intersection feature are 29.37%. What's more, GPS points with two candidate features are 4.73%, with three candidate features are 12.65% and with over three candidate features are only 0.41%. Because of the fuzzy matching of intersection feature, the percentage of single-feature point increases from 47.68% to 77.08%, all this lay a solid foundation for the efficiency of route reconstruction and route choice.

**The example:** Route reconstruction from Guangzhou Railway Station to Tiyuxi whose trip involves complex intersections and elevated roads is completed as an example. The value of elongation ratio is taken as 1.6 and the value of parameter $\theta$ is 10 according to the previous study of Guangzhou road network.

As Fig. 4 shows, elevated roads and the roads below are the matching candidate features of GPS points all the time. The introduction of the reasonable
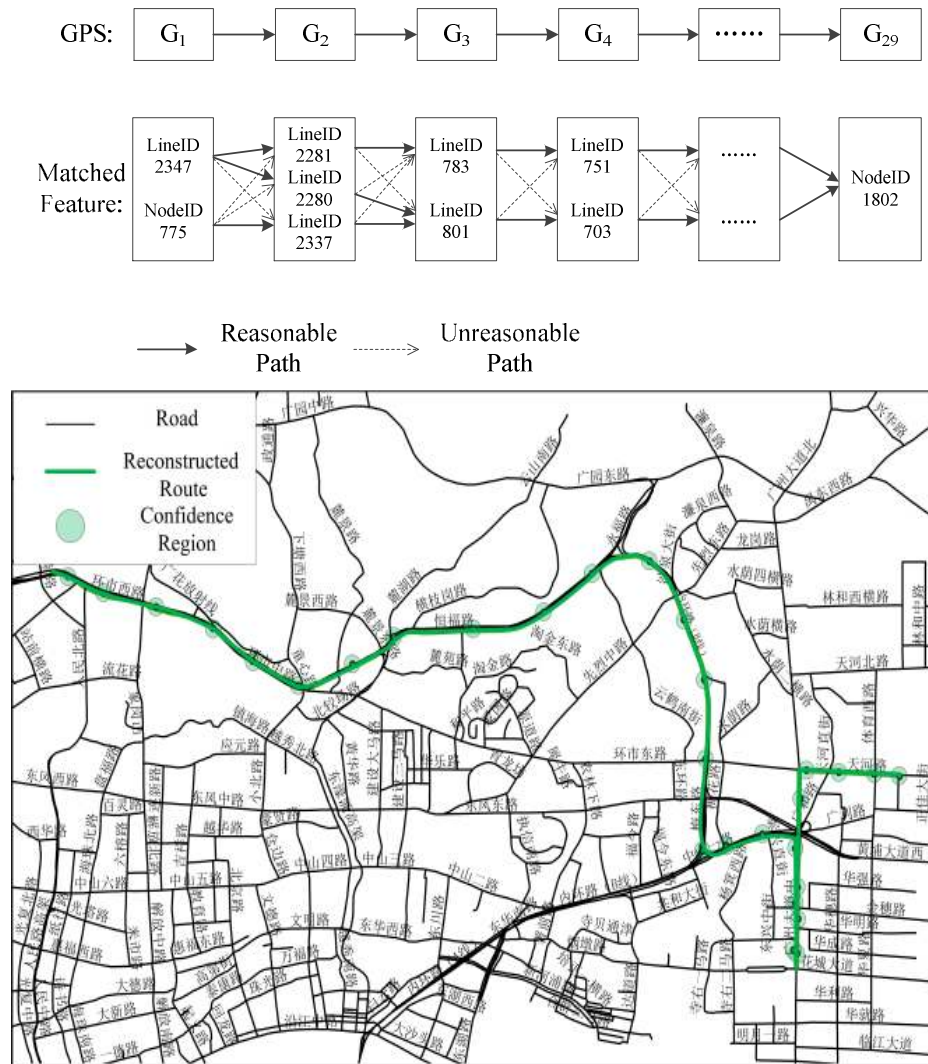
Fig. 4: Route reconstruction of Guangzhou railway station-Tiyuxi

path has avoided incessant transitions between elevated roads and the roads below. Their route distance and error distribution probability are also similar. Because the distance impedance of elevated roads should multiply by conversion coefficient, it has obvious superiority and becomes the matching route of this trip finally. The result is coincided with practice, as we know that when the travel track is coincident with both elevated roads and the roads below, drivers prefer run along the elevated roads than the roads below.

## CONCLUSION

We propose a method of route reconstruction from Floating Car Data in urban road network. The framework is made up of error statistics, feature matching and route reconstruction. The highest probability route by bi-level probabilistic model is supposed to the route which GPS points matched to.

The experiment shows that after fuzzy matching in intersection, the ratio of single-feature points can reach up to 70%. What's more, using elongation ratio can exclude unreasonable paths. The example initially indicates that the method is quite useful for route matching in urban common road network. More analysis using more complicated routes would be researched to illustrate the robustness of the proposed algorithm.

Moreover, considering the heading information, the matching line segment will be more practicable if its direction is closer to the GPS point heading. The heading information would be used when matching GPS point to the feature in the further research.

## REFERENCES

Alt, H., A. Efrat, G. Rote and C. Wnkb, 2003. Matching planar maps. IEEE T. J. Algorith., 49(2): 262-283.

Blazquez, C.A. and A.P. Vonderohe, 2005. Simple map-matching algorithm applied to intelligent winter maintenance vehicle data. Transport. Res. Rec., 1935: 68-76.

Brakatsoulas, S., D. Pfoser, R. Salas and C. Wenk, 2005. On map-matching vehicle tracking data. Proceeding of the 31st International Conference on Very Large Data Bases, Trondheim.

Chawathe, S.S., 2007. Segment-Based Map Matching. Proceeding of IEEE Symposium on Intelligent Vehicles, Istanbul, pp: 1190-1197.

Christopher, S.A., X. Li, R.M. Welch, J.S. Reid, P.V. Hobbs, T.F. Eck and B. Holben, 2000. Estimation of surface and Top-Of-Atmosphere (TOA) shortwave irradiance in biomass-burning regions during SCAR-B. J. Appl. Meteorol., 39: 1742-1753.

Greenfeld, J.S., 2002. Matching GPS observations to locations on a digital map. Proceeding of 81th Annual Meeting of the Transportation Research Board, Washington, D.C.

Kim, J.S., 1996. Node based map matching algorithm for car navigation system. Proceeding of the 29th ISATA Symposium, Florence, pp: 121-126.

Leurent, F., 1997. Curbing the computational difficulty of logit equilibrium assignment model. Transport. Res. B, 31: 315-326.

Nagendra, R.V., M.A. Quddus and A.L. Bristow, 2009. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems. Transport. Res. Part C Emerg. Technol., 17(6): 672-683.

Ochieng, W., M. Quddus and R. Noland, 2003. Map-matching in complex urban road networks. Brazilian J. Cartog., 55(2): 1-18.

Quddus, M.A., R.B. Noland and W.Y. Ochieng, 2006. A high accuracy fuzzy logic based map matching algorithm for road transport. J. Intell. Transport. Syst., 10(3): 103-115.

Srinivasan, D., R.L. Chue and C.W. Tan, 2003. Development of an improved ERP system using GPS and AI technique. Proceeding of IEEE Conference on Intelligent Transportation Systems, 1: 554-559.

Taylor, G., G. Blewitt, D. Steup, S. Corbett and A. Car, 2001. Road reduction filtering for GPS-GIS navigation. Proceeding of 3rd AGILE Conference on Geographic Information Science, Helsinki, pp: 114-120.

Wei, W., J. Jing, B. Ran and X. Guo, 2011. Large-scale freeway network traffic monitoring: A map-matching algorithm based on low-logging frequency GPS probe data. J. Intell. Transport. Syst., 15(2): 63-74.

White, C.E., D. Bernstein and A.L. Kornhauser, 2000. Some map matching algorithms for personal navigation assistants. Transport. Res. Part C Emerg. Technol., 8(1): 91-108.

Yang, D., C. Baigen and Y. Yuan, 2003. An improved map-matching algorithm used in vehicle navigation system. Proceeding of IEEE Intelligent Transportation Systems, 2: 1246-1250.

Yin, H. and O.A. Wolfson, 2004. Weight-based map matching method in moving objects databases. Proceedings of the 16th International Conference on Scientific and Statistical Database Management, Dept. of Comput. Sci., UIC, Chicago, IL, USA, pp: 437-438.

Yin, L., C. Zhang, Y. Zheng, X. Xie and W. Wang, 2009. Map-matching for low-sampling-rate GPS trajectories. Proceeding of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, pp: 352-361.