

A Map Matching Algorithm to Eliminate Miscalculation Based on Low-Sample-Rate Data

WANG Dong, WANG Zhiwei, Li Xiaohong, XIAO Zhu

College of Information Science and Engineering

Hunan University

Changsha, China

lixhong@263.net

Abstract—Map-matching is the technology of aligning a sequence of user's GPS positions with the road network on a digital map. However, there exists miscalculation in map-matching for low-sampling-rate GPS trajectories. To address this problem, this paper proposes an algorithm termed MIV-Matching for low-sample-rate GPS trajectories. To improve the accuracy in map-matching, MIV-Matching considers the process of existing algorithms for low-sample-rate GPS trajectories and then discusses and corrects the process of miscalculation. This paper also evaluates the algorithm on real life data set. Experiment results show that the MIV-Matching algorithm outperforms the related method (ST-Matching and IVMM algorithm).

Keywords- GPS Trajectories; Road Network; Low-Sample-Rate Data; Spatial AnalysisIntroduction

I. INTRODUCTION

Nowadays, with the increasing popularity of GPS-enabled device, users have facilitated to track moving objects such as people, vehicles. However, there are many positioning and sampling errors caused by GPS sensor^[1], and these errors impact the precision of GPS application. Therefore, in order to improve the accuracy of GPS data, many researchers use the digital road map to match the GPS errors data and the process is called Map-Matching^[2]. The general purpose of a Map-Matching is to identify the true road segment on which a moving object is travelling. The Map-Matching is a key procedure in many applications based on location information, such as vehicle navigation^[3], fleet management^[4], and intelligent transport system (ITS)^[5], and so on^[6, 7]. Since most of the civilian GPS devices and GPS module in smart phone are cheap and low accuracy, we need a robust map-matching algorithm in the services based on location information to improve the accuracy of GPS data.

The majority of existing map matching algorithms is based on high-sample-rate datasets (e.g. one sample per 10 seconds). However, the acquisition and storage of high-sample-rate datasets need to consume more storage space, GPS module operation and computation time. Therefore, users usually reduce the sampling frequency in collection sample points handheld by GPS devices, which produce many low-sample-rate GPS datasets (The sampling interval is more than 2 minutes).

With lower sampling frequency, many details between two sample points can easily be lost. For example, when the sampling interval is 3 minutes, the distance between two sample points may travel 2 kilometers even if a vehicle's

speed is 40km/h, and it will through several roads in the urban. In this case, the map matching algorithm based on high-sample-rate datasets cannot match the low-sample-rate datasets.

Currently, researchers have proposed some map matching algorithm based on low-sample-rate datasets, but there are any miscalculation. Therefore, we need to establish a more accurate, robust and stable map matching algorithm to solve the problem.

The remainder of this paper is organized as follows. Section II reviews the related work. Section III gives any definitions and shows the miscalculation. Section IV discusses the reason of miscalculation in ST-Matching and IVMM. On this basis, we proposed an algorithm called MIV-Matching. The experiment results are presented in Section V. Section VI concludes the paper and gives directions for our future work.

II. RELATED WORK

Recently, domestic and foreign researchers have highlighted a variety of map matching algorithm to calculate the relationship between GPS data and digital map. According to the different characteristics of the algorithms, we can classify it by various criteria.

A. Map-matching algorithm based on location information

According to the GPS trajectories and road network information, exiting methods can be categorized into three groups: geometric^[8], topological^[9, 10] and probabilistic^[11] methods.

Map matching based on geometric method usually consider the distance relations between individual GPS point and road network, and match the point to road by measuring the distance weight^[12, 13, 14]. But the matching result is imprecise because the geometric method does not consider the linkages between GPS points. And with reduced GPS accuracy, the result errors of map matching will be increased.

Map matching based on topological method usually takes into account the topological relationship between roads, GPS points, etc.^[8, 15, 16]. The algorithms get together the successive, related, restricted points, and combine geometric analysis to match it to the road.

Honey^[15] first introduces the map matching based on probabilistic^[11, 17] method. The algorithms present an oval or rectangular error range for each GPS point, and set an error probability value according to a distance between GPS point and a position in error range.

B. Global Algorithms and Local Algorithms

The local algorithms follow a greedy strategy of sequentially extending the solution from an already matched portion. Such algorithms try to find a local optimal point or edge that is similarity with the distance and direction between two GPS points^[8]. It utilizes the *Dijkstra* algorithm on local free space graph to construct a shortest graph. At high-sample-rate datasets environment, the arithmetic speed of local algorithms is faster than global algorithms. However, when the sample rate is decreased, local algorithms will be "arc-skipping"^[8] and that decreases the result accuracy.

Moreover, the global algorithms try to find a sequential road which is as closer as possible to the sampling track among all available roads in the road network. In order to measure the similarity between matched trajectory and sampling trajectory, almost all the global algorithms uses the "Fr chet distance" or "Weak Fr chet distance" [18]. This work is extended in^[19] where the "average Fr chet distance" is presented.

C. Sampling Rate

According to the sampling density of GPS points in map matching, the existing map matching algorithms can be divided into methods based on high-sample-rate datasets and low-sample-rate datasets.

ST-Matching was the first to be applied to the map matching based on low-sample-rate datasets. It incorporates both the geometric and topologic structures and the speed constrains. Jing extends the ST-Matching to IVMM^[2] which joins the interaction between GPS points to reduce the matching error. Authors in [20] suggested WI-Matching which contained the direction module to solve road roundabout and uncertain problem in matching results.

However ST-Matching and IVMM have the same miscalculation in spatio-temporal analysis and loses any mobile behavior of vehicle. On this basis, this paper introduces an optimization algorithm based on low-sample-rate datasets: MIV-Matching. It not only considers the influence between GPS points but also deals with the spatial analysis. Specially, we focus on the distance relations between two GPS points and matching road in spatial analysis process.

III. MIS CALCULATION OF ST-MATCHING AND IVMM

A. Definitions

Definition 1. GPS point: The GPS point p is a position that is determined by using the GPS device in a certain time interval. GPS point contains latitude and longitude coordinates, the GPS time and other data. We use P to represent the point sets.

Definition 2. GPS Trajectory: A GPS trajectory T is a point sequence $T: p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$ which linked by the GPS points according to a time interval.

Definition 3. Road Network: Road network is a directed graph $G(V, E)$, where include edge number $e.id$, travel speed of edge $e.v$, length of edge $e.l$, starting point $e.start$ and ending point $e.end$.

Definition 4. Road segment: It defines a maximum error of GPS as the radius of circle, and GPS point p as center, the road in part of the circle call road segment. It can be defined as $E(p) = \{ \forall e \in E \mid dis(e, p) < r \}$.

Definition 5. Candidate point: A candidate point c can be determined as a mapping point that the GPS point map in its road segment $E(p)$. We use C to define sets of candidate point.

Definition 6. Path: Given two candidate points c_s and c_e , a path $Path_c$ is a set of connected road segments that starts at c_s and ends at c_e . And $Path_c$ contains a shortest path defined as $Path_{min}$.

Definition 6. Optimal path: The optimal path is a candidate point sequence which the calculated value is maximal in all paths. It can be specified as $Path: c_1^{a_1} \rightarrow c_2^{a_2} \rightarrow \dots \rightarrow c_n^{a_n} (1 \leq a_i \leq s_i)$.

B. Miscalculation

According to the experiments in ST-Matching and IVMM, we discovered that the candidate points are too close in optimal path and it is resulting that the optimal path cannot reflect the true path of vehicles. As shown in Fig.1, the left image is the graph of GPS trajectory whereas the right image is optimal path. In the left one, point 3 and 4 are the GPS points of same position that stopped an intersection to wait for road passage, and point 5 is the point that was moving to next road. In the right one, point 3 and 5 are very adjacent, but point 3 and 4 are far away from each other. In above condition, we cannot distinguish the moving direction and roads of vehicle between point 4 to 6.



Fig.1 GPS trajectory and the optimal path by miscalculation

IV. ANALYSIS OF MIS CALCULATION AND OPTIMIZATION ALGORITHM: MIV-MATCHING

A. The Reason of Miscalculation

ST-Matching and IVMM presents the spatial analysis as:

$$V(c_{i-1}^t \rightarrow c_i^s) = \frac{d_{i-1 \rightarrow i}}{w_{(i-1,t) \rightarrow (i,s)}}. \text{ Where } d_{i-1 \rightarrow i} \text{ is the Euclidean}$$

distance between two successive GPS points, $w_{(i-1,t) \rightarrow (i,s)}$ is the length of $Path_{min}$ between c_{i-1}^t and c_i^s . According to the formula we can present the reason of miscalculation as follows:

a. GPS point p_{i-1} and p_i have road segment $E(p_{i-1})$ and $E(p_i)$, and the collection $E(p_{i-1}) \cap E(p_i) = \{e \in E(p_{i-1}), e \in E(p_i)\}$ is not null.

b. p_{i-1} and p_i map a candidate point set C_{i-1}' and C_i' in $E(p_{i-1}) \cap E(p_i)$ respectively, and the collection $\{c_{i-1}^{a_{i-1}} \in C_{i-1}', c_i^{a_i} \in C_i' \mid path_dis(c_{i-1}^{a_{i-1}}, c_i^{a_i}) < \epsilon\}$ is not null.

Where $\text{path_dis}(c_{i-1}^{a_{i-1}}, c_i^{a_i})$ is the length of Path_{\min} between $c_{i-1}^{a_{i-1}}$ and $c_i^{a_i}$, ε is a value to judge whether two candidate points is closed.

c. The value of $w_{(i-1,t) \rightarrow (i,s)}$ tends to 0, but $d_{(i-1 \rightarrow i)}$ is a nonzero value and causes that $V(c_{i-1}^t \rightarrow c_i^s)$ becomes an infinite value.

d. The $V(c_{i-1}^t \rightarrow c_i^s)$ will impact the matching process in ST-Matching and IVMM.

B. MIV-Matching

We discovered that the essential condition of miscalculation is $w_{(i-1,t) \rightarrow (i,s)} \rightarrow 0$. But in fact, $V(c_{i-1}^t \rightarrow c_i^s)$ should not be large with $w_{(i-1,t) \rightarrow (i,s)} \rightarrow 0$, it should be large with the similar distance interval between $w_{(i-1,t) \rightarrow (i,s)}$ and $d_{(i-1 \rightarrow i)}$. In the conclusion, if

$x = \frac{d_{i-1 \rightarrow i}}{w_{(i-1,t) \rightarrow (i,s)}}$ and the $V(c_{i-1}^t \rightarrow c_i^s)$ then $F(x)$ can be

determined, and the spatial analysis should follow as: 1) If $x=1$, then $F(x)$ is maximum. 2) If $x \in (0,1)$, $F(x)$ is increasing function. 3) If $x \in (1,+\infty)$, $F(x)$ is decreasing function. On this way, we precise $F(x)$ as $F(x) = \frac{1}{(1-x)^2}$, and

$V(c_{i-1}^t \rightarrow c_i^s)$ as:

$$V(c_{i-1}^t \rightarrow c_i^s) = \frac{1}{(1 - \frac{d_{i-1 \rightarrow i}}{w_{(i-1,t) \rightarrow (i,s)}})^2}, 2 \leq i \leq n, 1 \leq t \leq s_{i-1}, 1 \leq s \leq s_i$$

The $d_{(i-1 \rightarrow i)}$ is a nonzero value because the vehicle is locomotive and the GPS data has any errors. On this basis, we check if the formula with $w_{(i-1,t) \rightarrow (i,s)} \rightarrow 0$ and $w_{(i-1,t) \rightarrow (i,s)}$ is nonzero value.

When $w_{(i-1,t) \rightarrow (i,s)} \rightarrow 0$, then $x \rightarrow \infty$, and $F(x) = \frac{1}{(1-x)^2}$

$\rightarrow 0$. It means the spatial analysis value $V(c_{i-1}^t \rightarrow c_i^s) \rightarrow 0$ when position of successive candidate points is same or adjacent.

When $w_{(i-1,t) \rightarrow (i,s)}$ is a nonzero value, then x and $F(x) = \frac{1}{(1-x)^2}$ also is a nonzero value. And when

$w_{(i-1,t) \rightarrow (i,s)} \rightarrow d_{(i-1 \rightarrow i)}$, then $x \rightarrow 1$, and $F(x) \rightarrow \infty$. It means the spatial analysis value is closer infinity when matching result is closer to the true path of vehicle.

Based on the new formula of spatial analysis, and combination of the process of ST-Matching and IVMM, we present a novel MIV-Matching algorithm. Fig.2 is the framework of MIV-Matching and the process can be described to the following steps:

step 1. The algorithm obtains the candidate point set C_i with each GPS point p_i by candidate preparation.

step 2. The algorithm calculates the spatio-temporal value

following as:

- It obtains the spatial analysis probability of each candidate point c_i^s in C_i by optimized spatial analysis.
- Combine with the temporal analysis, the algorithm obtains the value of spatio-temporal analysis.
- The algorithm sets *ST-function* of each candidate point $c_1^{a_1}$ in C_1 to the observation probability of itself, and the *ST-function* between c_{i-1}^t and c_i^s ($2 \leq i \leq n$) is the value of spatio-temporal analysis between them. At the end, the algorithm constitutes the candidate graph $G_T'(C_T', E_T')$ by *ST-function*.

step 3. According to the candidate graph $G_T'(C_T', E_T')$, the algorithm obtains the two different optimal paths by two types of processes.

- The first is ST-Matching process. The algorithm uses $G_T'(C_T', E_T')$ to find a path_c by matching result process in ST-Matching, and the defined path_c is the optimal path.
- The algorithm finds a path_c by mutual influence modeling and interactive voting process in IVMM, and the defined path_c is the optimal path.

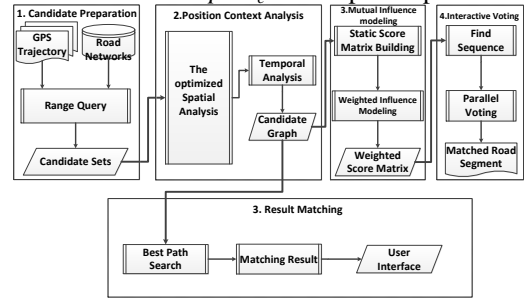


Fig.2 MIV-Matching fragment

V. EXPERIMENTS

A. Experimental settings

1) Dataset description

In our experiments, all GPS trajectories are real data collected from the GPS devices we carried in bus of Changsha. The datasets contains 68 trajectories, and the trajectories cover about 680 kilometers over more than 30 hours.

We establish the road networks from 354 roads in Changsha that contain all the error ranges of each GPS point in 68 trajectories.

2) Parameter Selection

The sample rate: MIV-Matching focuses on the low-sample-rate datasets, thus the sampling interval ranges from 30 seconds to 10.5 minutes in our experiments.

Parameters for MIV-Matching: In order to compare ST-Matching with IVMM, we set $k=5$ as the maximum number of candidates for each GPS point and the range of error is 200 meters. We use a normal distribution (temporal analysis) with $\mu=5$ meters and $\sigma=10$ meters.

The $f(dist(p_i, p_j))$: In order to compare ST-Matching with IVMM, in our experiments, we assign 7 kilometers to β in $f(dist(p_i, p_j))$. We can set β to other value, even set different $f(dist(p_i, p_j))$.

Decision value: We set ε as 20 meters in collection $\{c_{i-1}^{a_{i-1}} \in C_{i-1}, c_i^{a_i} \in C_i | path_dis(c_{i-1}^{a_{i-1}}, c_i^{a_i}) < \varepsilon\}$. It means that we judge two successive candidate points that are closed when the length of $Path_{min}$ between them is less than 20 meters.

B. Evaluation Approaches

Moreover, we design a method of assessment as follows:

a. We record the vehicle mobile distance of each time interval in real datasets, they are determined as dis_i .

b. The experiments calculate the length of $Path_{min}$ between each of the two successive candidate points $c_{i-1}^{a_{i-1}} \rightarrow c_i^{a_i}$ in optimal path $c_1^{a_1} \rightarrow c_2^{a_2} \rightarrow \dots \rightarrow c_n^{a_n}$, and specify it as dis_i' .

c. In each time interval, we calculate the length error between real moving and matching path, and define it as $\square dis_i = |dis_i - dis_i'|$. Then, we compute the variance and standard deviation of $\square dis_i$.

d. We calculate the average distance error $\overline{\square dis}$ distribution by different number of closed candidate point between two successive GPS points.

C. Experimental Result

1) Result of Single Match

Fig.3, Fig.4 and Fig.5 plot the GPS trajectory, result of ST-Matching and result of IVMM. The result means that MIV-Matching is the best to solve the problem where the candidate point in optimal path is too close. And MIV-Matching can more accurately reflect the true path of vehicle.

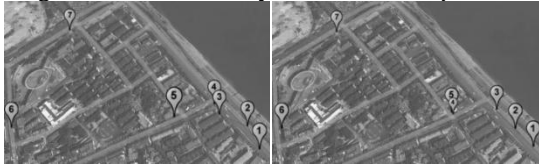


Fig.3 GPS Trajectory



Fig.4 Result of ST-Matching



Fig.5 Result of ST precess in MIV-Matching

2) Δdis Distribution

All GPS points are randomly collected in error range of vehicle by GPS devices, and the candidate point is the vertical point of GPS point in road. But the real position of vehicle is not always the vertical point. So there are more or less error between dis_i' and dis_i , it means Δdis has error.

Fig.6 represents the Δdis distribution of ST process of MIV-Matching and ST-Matching, and Fig.7 shows the $\square dis$ distribution of IVMM process of MIV-Matching and IVMM.

The experimental data contain 20 GPS points in each trajectory. Because of the miscalculation, the error of ST-Matching and IVMM are larger than MIV-Matching. As shown in Fig.6, ST-Matching appeared 3 times error between 114-157 meters and MIV-Matching downsizes it to 19-57 meters. It also happened in Fig.7.

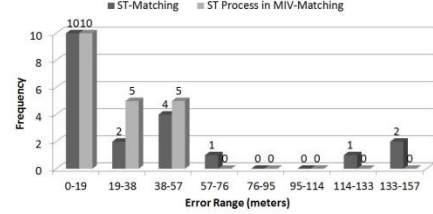


Fig.6 The $\square dis$ distribution of ST-Matching and MIV-Matching

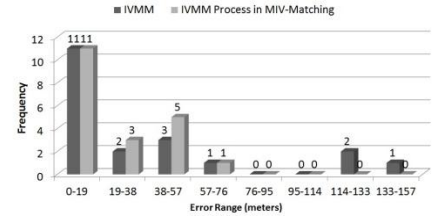


Fig.7 The $\square dis$ distribution of IVMM and MIV-Matching

Fig.8 and Fig.9 present the variance and standard deviation of Δdis in 3 algorithms. It shows that the result of MIV-Matching is smaller than ST-Matching and IVMM. It means the error of MIV-Matching is smaller.

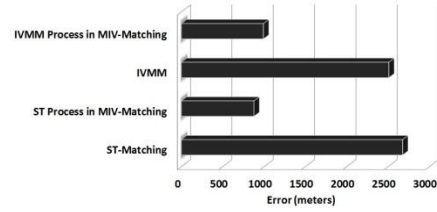


Fig.8 The variance of $\square dis$

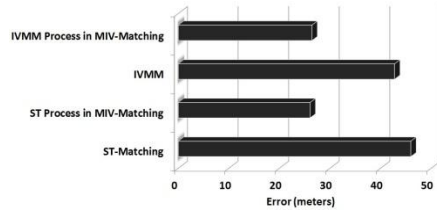


Fig.9 The standard deviation of $\square dis$

3) $\square dis$ distribution with different number of closed candidate point between two successive GPS point

Fig.10 represents the comparison of $\square dis$ distribution between MIV-Matching, ST-Matching and IVMM. When the number of closed candidate point is smaller than 7, the miscalculation is not happened, because the distance between two successive candidate point is larger and the common road segments of two successive GPS points are simplex.

When the common road segments of two successive GPS points are complex, the miscalculation is easy to happen because it meets the essential condition of miscalculation. That is why the $\square dis$ of ST-Matching and

IVMM are larger and MIV-Matching remains unchanged when the number of closed candidate point is larger than 7.

The experiments imply that MIV-Matching eliminated the miscalculation in ST-Matching and IVMM.

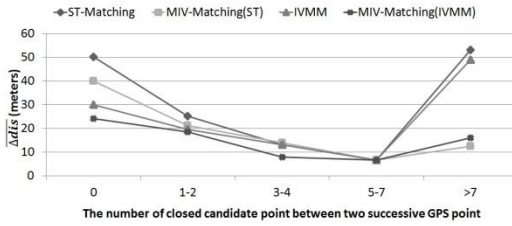


Fig.10 Distribution of \overline{dis}

VI. CONCLUSIONS

In this paper, we introduce the miscalculation of ST-Matching and IVMM. A novel algorithm termed MIV-Matching is proposed and analyzed. The algorithm established the candidate graph by spatio-temporal analysis, and focus on implementation of miscalculation in the graph. In order to obtain the optimal path, the algorithm does the matching from ST and IVMM process by candidate graph. In the experiments, we compared MIV-Matching to ST-Matching and IVMM by real GPS datasets. As a result, the MIV-Matching algorithm outperforms the ST-Matching and IVMM.

In our future work, we plan to focus on accuracy of matching in real time based on MIV-Matching algorithm. In the real time environment, it exists any instability from low-sample-rate, and needs more precision algorithm to solve the problems.

- [1] Pfoser, Dieter, and Christian S. Jensen. "Capturing the uncertainty of moving-object representations." *Advances in Spatial Databases*. Springer Berlin Heidelberg, 1999.
- [2] Yuan, Jing, et al. "An interactive-voting based map matching algorithm." *Mobile Data Management (MDM), 2010 Eleventh International Conference on*. IEEE, 2010.
- [3] Joshi R R. A new approach to map matching for in-vehicle navigation systems: the rotational variation metric/CI//Intelligent Transportation Systems, 2001. *Proceedings. 2001 IEEE*. IEEE, 2001: 33-38.
- [4] Feijoo, C., J. Ramos, and F. Perez. "A system for fleet management using differential GPS and VHF data transmission mobile networks." *Vehicle Navigation and Information Systems Conference, 1993., Proceedings of the IEEE-IEE*. IEEE, 1993.
- [5] Ouddus, Mohammed A., Robert B. Noland, and Washington Y. Ochieng. "A high accuracy fuzzy logic based map matching algorithm for road transport." *Journal of Intelligent Transportation Systems* 10.3 (2006): 103-115.
- [6] Rauschert, Ingmar, et al. "Designing a human-centered, multimodal GIS interface to support emergency management." *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*. ACM, 2002.
- [7] Ouddus, Mohammed A., Washington Y. Ochieng, and Robert B. Noland. "Current map-matching algorithms for transport applications: State-of-the art and future research directions." *Transportation Research Part C: Emerging Technologies* 15.5 (2007): 312-328.
- [8] Greenfeld, Joshua S. "Matching GPS observations to locations on a digital map." National Research Council (US). Transportation Research Board. Meeting (81st: 2002: Washington, DC). Preprint CD-ROM. 2002.
- [9] Chen, W., et al. "Integrated vehicle navigation system for urban applications." *Proceedings of the 7th International Conference on Global Navigation Satellite Systems (GNSS)*, European Space Agency, Graz, Austria. 2003.
- [10] Yin, Huabei, and Ouri Wolfson. "A weight-based map matching method in moving objects databases." *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*. IEEE, 2004.
- [11] Bierlaire, Michel, Jingmin Chen, and Jeffrey Newman. "A probabilistic map matching method for smartphone GPS data." *Transportation Research Part C: Emerging Technologies* 26 (2013): 78-98.
- [12] Bernstein, David, and Alain Kornhauser. "An introduction to map matching for personal navigation assistants." (1998).
- [13] Phuval, Bishnu P. "Method and use of aggregated dead reckoning sensor and GPS data for map matching." *Proceedings of the 15th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2002)*. 2001.
- [14] White, Christopher E., David Bernstein, and Alain L. Kornhauser. "Some map matching algorithms for personal navigation assistants." *Transportation Research Part C: Emerging Technologies* 8.1 (2000): 91-108.
- [15] Honey, Stanley K., et al. "Vehicle navigational system and method." U.S. Patent No. 4,796,191. 3 Jan. 1989.
- [16] He, Zhao-cheng, et al. "On-line map-matching framework for floating car data with low sampling rate in urban road networks." *IET Intelligent Transport Systems* 7.4 (2013): 404-414.
- [17] Newbv, Robert Matthew, et al. "Vehicle location and navigation system." U.S. Patent Application 11/908,285.
- [18] Alt, Helmut, et al. "Matching planar maps." *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2003.
- [19] Brakatsoulas, Sotiris, et al. "On map-matching vehicle tracking data." *Proceedings of the 31st international conference on Very large data bases. VLDB Endowment*, 2005.
- [20] Liu, Siyuan, et al. "Calibrating large scale vehicle trajectory data." *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on*. IEEE, 2012.