

A dynamic two-dimensional (D2D) weight-based map-matching algorithm



M.N. Sharath^a, Nagendra R. Velaga^a, Mohammed A. Quddus^{b,*}

^a Transportation Systems Engineering, Department of Civil Engineering, Indian Institute of Technology (IIT) Bombay, Mumbai 400076, India

^b Transport Studies Group, School of Architecture, Building and Civil Engineering, Loughborough University, Leicestershire LE11 3TU, UK

ARTICLE INFO

Keywords:

Lane-level localization
Topological map-matching
Dynamic coefficients
2-D map-matching
Adaptive regression

ABSTRACT

Existing map-Matching (MM) algorithms primarily localize positioning fixes along the centerline of a road and have largely ignored road width as an input. Consequently, vehicle lane-level localization, which is essential for stringent Intelligent Transport System (ITS) applications, seems difficult to accomplish, especially with the positioning data from low-cost GPS sensors. This paper aims to address this limitation by developing a new dynamic two-dimensional (D2D) weight-based MM algorithm incorporating *dynamic weight coefficients* and *road width*. To enable vehicle lane-level localization, a road segment is virtually expressed as a matrix of homogeneous grids with reference to a road centerline. These grids are then used to map-match positioning fixes as opposed to matching on a road centerline as carried out in traditional MM algorithms. In this developed algorithm, vehicle location identification on a road segment is based on the total weight score which is a function of four different weights: (i) proximity, (ii) kinematic, (iii) turn-intent prediction, and (iv) connectivity. Different parameters representing network complexity and positioning quality are used to assign the relative importance to different weight scores by employing an *adaptive regression method*. To demonstrate the *transferability* of the developed algorithm, it was tested by using 5,830 GPS positioning points collected in Nottingham, UK and 7,414 GPS positioning points collected in Mumbai and Pune, India. The developed algorithm, using stand-alone GPS position fixes, identifies the correct links 96.1% (for the Nottingham data) and 98.4% (for the Mumbai-Pune data) of the time. In terms of the correct lane identification, the algorithm was found to provide the accurate matching for 84% (Nottingham) and 79% (Mumbai-Pune) of the fixes obtained by stand-alone GPS. Using the same methodology adopted in this study, the accuracy of the lane identification could further be enhanced if the localization data from additional sensors (e.g. gyroscope) are utilized. ITS industry and vehicle manufacturers can implement this D2D map-matching algorithm for liability critical and in-vehicle information systems and services such as advanced driver assistant systems (ADAS).

1. Introduction

GPS is the backbone of many location-based Intelligent Transport Systems (ITS) such as navigation and route guidance, public transport operations and road pricing (e.g. Velaga and Pangbourne, 2014; Mukheja et al., 2017). Applications such as road network generation from crowdsourced trajectories can resort to offline map-matching (MM) algorithms (e.g., Knapen et al., 2018; Yang et al., 2018). However, more stringent vehicle-based ITS applications such as collision avoidance systems, lane-level navigation assistance,

* Corresponding author.

E-mail address: m.a.quddus@lboro.ac.uk (M.A. Quddus).

<https://doi.org/10.1016/j.trc.2018.12.009>

Received 7 July 2018; Received in revised form 17 December 2018; Accepted 17 December 2018

0968-090X/ © 2018 Loughborough University. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

lane departure warning, emergency response and enhanced driver awareness systems require lane-level positioning and mapping in real time (Toledo-Moreo et al., 2010; Velaga and Sathiasel, 2011). Unfortunately, position coordinates obtained from a GPS receiver suffer from different types of failures resulting in positioning error of between 5 and 20 m (2 σ) (e.g., Kaplan and Hegarty, 2006; Thanh-Son et al., 2006; Velaga et al., 2013; Aly et al., 2016). Digital road maps have also a deficiency in the forms of missing segments, translated segments, rotated segments and misrepresentation of intersections resulting in an inaccuracy in the range of 5–20 m between the ground truth and the corresponding map point (Toledo-Moreo et al., 2010). A MM algorithm integrates these erroneous positions and map data to derive position solutions that are more accurate and reliable. MM algorithms are categorized as: geometric, topological, probabilistic and advanced (Quddus, 2006; Quddus et al., 2007). For a detailed review of MM algorithms, readers are directed to Quddus et al. (2007) and Hashemi and Karimi (2014).

In order to further enhance the performance of a positioning system with respect to accuracy and reliability, researchers developed integrated positioning systems consisting of multiple sensors such as dead-reckoning, RADAR, LiDAR and cameras (Tao et al., 2013; Li et al., 2014; Alkhorshid et al., 2015; Kamijo et al., 2015; Cui et al., 2016; Kogan et al., 2016; Rabe et al., 2016a, 2016b; Kim et al., 2017). However, in the context of developing economies (such as India), a stand-alone GPS receiver shall be employed as the only ubiquitous sensor. This paper, therefore, aims to develop a MM algorithm capable of achieving high performance by using only a stand-alone GPS receiver.

Typically, in MM algorithms, roads are represented by centerlines (1-dimensional) ignoring the road width, which is one of the major limitations of existing algorithms. A traditional weight-based MM algorithm assigns different weighting scores to different elements of the road network (segments/nodes). The probable position is decided based on the total weighting score determined after considering the relative importance of the different weighting scores. Literature review suggests that these weighting scores depend on the operational environment (i.e., road network complexity) and indicators for the quality of positioning fixes (e.g., speed and dilution of precision) (Velaga et al., 2009, 2012b; Hashemi and Karimi, 2016). Road network complexity is *spatially* dynamic while the positioning quality indicator of GPS is *temporally* dynamic. These indicators are henceforth termed as dynamic indicators. The present study develops a two-dimensional (2D) weight-based MM algorithm by considering the road width that strives to localize the vehicle at lane-level. In doing so, it incorporates dynamic weight coefficients to address different operational environments. Therefore, the algorithm is termed as a dynamic two-dimensional (D2D) map-matching algorithm. The key features of the developed D2D MM algorithm are as follows:

1. The road segment is not represented by its centerline, rather as a matrix of homogeneous grids. These virtual grids are generated using road centerline data and road width data, which are represented by their centroids (see Fig. 3). Lane-level positioning is principally possible because of the consideration of road width;
2. The data from stand-alone GPS is used without any augmentation;
3. Four different weighting scores are considered: proximity, kinematic, connectivity, and turn-intent prediction;
4. Virtual grids (not road segments) are awarded different weighting scores. The relative importance of the weighting scores is computed based on prevailing road network complexity and the quality of positioning information. Therefore, the coefficients are both *spatially* and *temporally* dynamic;
5. Adaptive regression method is used to express the weight coefficients as a function of dynamic indicators.

2. Literature review

MM algorithms perform the task of localizing a vehicle on the road network using digital map data and information from a GPS receiver. MM algorithms support a range of ITS applications such as the estimation of parameters of a route choice model (Oyama and Hato, 2018), travel-time estimation (Rahmani et al., 2017; Tang et al., 2018), taxi trip anomaly detection (Qin et al., 2019) and gridlock network analysis (Oyama and Hato, 2017). Different approaches for map-matching have their specific benefits and drawbacks. For instance, topological weight-based MM algorithms stand out because of its ease of implementation and high performance (both in terms of speed and accuracy) (Quddus, 2006; Velaga et al., 2009; Hashemi and Karimi, 2016). When topological information such as the shape of each link, turn restrictions, speed limits, historical matching information are incorporated in deriving a position, the performance of the MM algorithm is enhanced (Bernstein and Kornhauser, 1996; Bouju et al., 2002; Srinivasan et al., 2003). In this paper, a review of weight-based topological map-matching algorithms is provided.

Greenfeld (2002) pioneered the concept of weight-based MM algorithm. Three weighting scores were defined: (i) heading of the vehicle, (ii) bearing of road link and (iii) proximity of raw position information relative to road links. The consideration of equal weighting coefficients is one of the main drawbacks of the Greenfeld's (2002) algorithm. However, the differential importance of different factors influencing a MM process was realized soon after the introduction of the Greenfeld's algorithm. For instance, Quddus et al. (2003) developed a weight-based MM algorithm fusing GPS with Inertial Navigation System (INS) using Extended Kalman Filter (EKF). Three weighting functions were used with heading weight being given the highest importance followed by the weight for relative position and the proximity. They identified the dependence of the weight coefficients on the performance of MM algorithms and empirically determined them. Unequal coefficients for different weighting scores were used to enhance the performance of the MM algorithm (Quddus et al., 2003).

Li et al. (2013) developed a weight-based topological MM algorithm capable of localizing the vehicle and providing its integrity. The MM algorithm developed by Velaga et al. (2012a) was adopted by Li et al. with suitable modifications to incorporate 3-dimensional Digital Elevation Model (DEM) and additional topological features. Because a DEM has road-link elevation information, level separated road networks (e.g., fly-overs) can be efficiently map-matched. Their work claimed to identify 97.7% of the links

correctly. This high level of accuracy may be attributed to the application of DEM that is not readily available for a road network (particularly in developing countries).

The advantage of incorporating the shortest path algorithm is not easily realized when using high-frequency GPS data. But, the performance of most of the MM algorithms declines under low-frequency GNSS data. The connectivity information (which is a topological information) cannot be used for MM under this scenario (Quddus and Washington, 2015). A weight-based MM algorithm that works well even at low-frequency GPS sampling rate was developed by Quddus and Washington (2015) to address this issue. A *shortest path* algorithm was used in determining the shortest path between two successive GPS position fixes. However, most of the ITS applications demand high-frequency vehicle localization.

Yeh et al. (2016) resorted to using the MM algorithm developed by Velaga et al. (2009) with suitable modifications to incorporate DEM. Urban road network typically consists of a multi-level road network, which is a difficult scenario for map-matching. Using the DEM and an inclinometer, they infer whether the vehicle is using the ramp to change the elevation. This served as additional topological information to resolve a position in complex situations.

A more rational study was performed by Velaga et al. (2009) and reported a range of optimal weighting score coefficients for different operational environments with a premise that the road network complexity is indicated by operational environments (e.g., rural, suburban or urban). Four weight functions were used, and a novel optimization technique was adopted for estimation of optimal weighting score coefficients. The study confirms the dependency of weight coefficients with road network complexity. Further, the complexity of the operational environment is spatially dynamic, which motivates us to consider the dynamic weight coefficients in this research.

2.1. Research gap and motivation

Discrete characterization of the operational environments (i.e., rural, suburban and urban) may not be sufficient to comprehensively express the road network complexity, as it is rather continuous. This situation demands dynamic coefficient of weighting scores to represent a variety of road network complexity. The literature review clearly indicates the dependency of coefficients of the weights on road network complexity (i.e., topological parameters) (Velaga et al., 2009). However, Hashemi and Karimi (2016) have presented a weight-based MM algorithm with dynamic coefficients, wherein, no topological parameter was assumed to influence the coefficients. Only horizontal dilution of precision (HDOP), speed and distance traveled by the vehicle in the successive time epochs were considered in formulating the dynamic weighting coefficients (Hashemi and Karimi, 2016). Inspired by these two studies (Velaga et al., 2009; Hashemi and Karimi, 2016), road network complexity and positioning quality indicators are assumed to affect the weight coefficients.

As mentioned earlier, most of the existing MM algorithms ignored the road width parameter. Therefore, vehicle localization is always restricted along the centerline of the road; this is a major limitation with existing MM algorithms. A lane-level vehicle localization is therefore principally not possible. The works that claim to have achieved lane-level positioning adopt a combination of expensive sensors/technology such as LiDAR, RADAR, camera, inertial navigation sensors, connected vehicles technology, satellite/area augmentation, etc. so as to aid the localization process (example: Du et al., 2004; Thanh-Son et al., 2006; Du and Barth, 2008; Peyret et al., 2008; Toledo-Moreo et al., 2009, 2010; He and Head, 2010; Vu et al., 2012; Szottka, 2013; Tao et al., 2013; Aly et al., 2015, 2016; Rabe et al., 2016b; 2016a; Li et al., 2017). Affording these sensors/systems for localization in developing economies may not be possible in the near future. Therefore, data from only a stand-alone GPS receiver is used in developing a lane-level map-matching algorithm.

To address the above drawbacks, a weight-based topological MM algorithm is developed; it considers the road to be a two-dimensional element without ignoring the road width. Coefficients of the weighting scores are made dynamic to account for non-discrete road network complexity. No hardware augmentation is used and only the information from a stand-alone GPS receiver is used in MM at lane-level.

2.2. Research contributions

Since a large body of literature exists on map-matching algorithms, it is important to highlight the intended contributions of this research. This is briefly summarized below

Lane-level localization using standalone GPS receiver and opensource map: The developed MM algorithm makes use of ubiquitous standalone GPS receiver and road network data from OpenStreetMap. Despite these, the MM algorithm is capable of achieving very high performance in terms of correct link identification, correct lane identification and radial accuracy. Such a low-cost system configuration is desirable in the developing economies like India.

Dynamic weight coefficients: The magnitude of the weight coefficients has a significant effect on the performance of the MM algorithm. Literature clearly indicates the dependency of coefficients on the road network topology and quality of positioning information. Although Hashemi and Karimi (2016) have used dynamic coefficients, their study does not take into account any topological parameters. Five topological parameters and three indicators of positioning quality are used in this study so as to determine the magnitude of different weight coefficients.

A novel method to obtain optimal weight coefficients: A novel method to obtain the coefficients of different weights that maximizes the performance of the MM algorithm is provided in this study.

Two new weight functions: Kinematic weight and turn-intent prediction weight scores are introduced in this study. Further, a data driven approach for formulation of the different weight functions is being adopted. This ensures that different grids acquire appropriate weight scores.

Table 1

Description of GPS datasets used in the present study.

Dataset	Carrier-phase GPS data	Stand-alone GPS data	Data used for		Sampling frequency
			Calibration	Evaluation	
Dataset 1 (Nottingham)	8,245	8,245	2,415 (Dataset 1a)	5,830 (Dataset 1b)	1 Hz
Dataset 2 (Mumbai-Pune)	—	7,414	—	7,414	1 Hz
Dataset 3	Data used only for empirical analysis				1 Hz

2.3. Target applications

Many ITS applications that require lane-level localization are safety-of-life (SoL) critical and hence are strictly to be fail-proof. Therefore, MM algorithms capable of providing lane-level localization often make use of complementary and redundant sensors to ensure safety. However, the MM algorithm to be developed in this paper largely relies on open-source map and standalone GNSS receiver. Hence, owing to situations such as GNSS outages, map errors, the system may not be fail-proof. Therefore, the algorithm to be developed is intended to support the following liability critical ITS applications (see Table 5 for details):

1. Electronic toll collection
2. Lane specific traffic information
3. Lane allocation
4. Lane level navigation
5. Prediction of driver's intent.

3. Methodology

3.1. Data

Table 1 summarizes the dataset used in this study. Raw GPS data and corresponding precise carrier-phase GPS data for 8,245 epochs sampled at 1 Hz originally collected by Velaga et al. (2009) from Nottingham, UK were employed as Dataset 1. Along with positioning fixes, the data also has positioning quality indicators such as instantaneous velocity, horizontal and vertical dilution of precisions (DOPs). Dataset 1 also contains corresponding positioning fixes obtained from a high accuracy carrier-phase GPS receiver. The accuracy of the carrier phase GPS receiver was found to be better than 5 cm over 97.5% of the time (Aponte et al., 2009; Velaga, 2010). Therefore, positioning fixes from this carrier-phase GPS receiver is employed as ground truth to which the estimated positional accuracy of the MM algorithm is compared.

Dataset 2 was collected as a part of this research on 16th December 2016 along a route from Mumbai to Pune, India. The standalone GPS receiver used for collecting Dataset 1 and Dataset 2 is the same. In the absence of facility to measure the true position, a dashboard camera was used along with a stand-alone GPS receiver to record the video and GPS positioning data simultaneously. Information regarding the number of lanes, the lane occupied or road segments travelled by the experimental vehicle was deciphered manually from the video and used for the performance evaluation. This data sampled at 1 Hz has 7,414 data points.

The GPS vehicle trajectory data (1 Hz) collected by Velaga (2010) in Mumbai is used only for an empirical analysis. This dataset, which is referred to as Dataset 3, is neither used for calibration nor evaluation of the developed MM algorithm.

Road network data available from the OpenStreetMap is used in the study. Road width data (e.g. number of lanes) is not directly available in the OpenStreetMap. According to Lopes and Nogueira (2011), percentage error in horizontal distance measurement using the Google Earth is between 0.02% and 0.44%. Other studies have also used the Google Earth for measuring horizontal distances (e.g. Potere et al., 2009; Benker et al., 2011). In this study as well, the Google Earth is used to measure road width of the chosen roads.

3.2. Description of dynamic indicators

The premise of the study is that the coefficients of the weighting scores are governed by the prevailing road network complexity and positioning data quality at each epoch. The coefficients of the weighting scores are therefore assumed to depend on the following explanatory variables:

X1: Number of grids within a radius of 220 m considering the raw position as the centre. X1 indicates a macro-level road network complexity within the surrounding operational environment.

X2: Number of intersections within a radius of 50 m with the raw position as the centre. Map-matching at intersections is more complicated.

X3: Number of grids with a radius of 50 m. The road network complexity within the immediate surrounding is represented by X3. The difficult scenarios such as parallel road segments, dense road segments and intersections are effectively represented by this indicator.

X4: Number of segments joining at the intersection that is closest to the current position. The complexity in MM increases with the increased number of segments joining to an intersection.

X5: Distance to the nearest intersection in meters. The closer the vehicle to an intersection, the greater the difficulty in carrying out map-matching process.

X6: Horizontal Dilution of Precision (HDOP) from a GPS receiver, which is an indicator of 2-D horizontal position accuracy.

X7: Position Dilution of Precision (PDOP) from a GPS receiver, which is an indicator of 3-D position accuracy.

X8: Speed of the ego-vehicle in m/s.

Two radii were used to describe some of the dynamic indicators. The larger radius was assumed to indicate the macro-level road network complexity, whereas, the smaller radius was assumed to indicate the road network complexity of immediate surrounding. A sensitivity analysis was conducted using Dataset 1a wherein, different regression models were built using Multivariate Adaptive Regression Splines (MARS) for different combinations of radii to establish the relationship between dynamic indicators and weight coefficients. MARS analysis permits us to build an adaptive regression model (some details of MARS are provided in Section 3.3.2). Sensitivity analysis revealed that the radii of 220 m and 50 m are the combination that resulted in the best model. It can be noted that the radius of 220 m is comparable to the radius of 200 m specified by Velaga et al. (2012a) wherein the 200 m radius is used to categorize the road network complexity (operational environment) as rural, suburban or urban.

The positioning quality depends upon several spatial and temporal dynamic factors including ionosphere (Bhuyan and Borah, 2007), presence of trees and buildings (due to satellite masking) (Ohno et al., 2003; DeCesare et al., 2005), satellite-receiver geometry, etc. Exact quantification and controlling of some of these factors are rather difficult. However, the dilution of precision (DOP) parameters are used in this study. DOP parameters indicate the quality of the positioning solution derived at a particular time epoch considering the geometry of the satellite constellation and the receiver (Quddus, 2006). HDOP and PDOP are considered in this study as these indicate the quality of positioning fixes. Further, the velocity of the ego-vehicle (both heading and speed) is generally considered unreliable at low displacement rates of the vehicle (Quddus, 2006; Quddus and Washington, 2015). Therefore, the instantaneous speed of the ego-vehicle is considered as another indicator for quality of positioning fixes.

An empirical analysis is conducted using Dataset 1a to find the optimal grid size. More specifically, the performance parameters such as correct link identification, correct lane identification, data processing rate and radial error are determined for different grid sizes. The summary result of this empirical analysis is presented in Table 2.

Since the road network used in this analysis has varied complexities, the dynamic indicators employed in the algorithm address the heterogeneity in road network configurations. Therefore, the optimal grid size for different road configurations is not determined. With the main focus on data processing rate and radial accuracy, a virtual rectangular grid of length 3 m (measured along the centerline) and width 1.875 m is found to be the best performing grid. A longer length of the grid may marginally enhance the percentage correct link identification, but the repercussion is an increased radial error. The width of the grid is considered as a half the lane width because any further resolution would increase the number of grids and hence the computation time.

3.3. Calibration

Calibration is a pre-processing phase that involves establishing the relationship between dynamic indicators and the weight coefficients. A total of 2,415 data points from Dataset 1a are used for calibration while Dataset 1b and Dataset 2 are used for

Table 2
Summary results of the sensitivity analysis.

Width (m)	Length (m)	Correct lane detection (%)	Correct link detection (%)	^a Data processing rate (Hz)	Radial error (m)
0.9375	1	82.93	95.80	8.81	2.11
1.875	1	85.96	95.26	9.93	2.29
2.8125	1	82.54	95.30	13.09	2.52
3.75	1	82.98	95.51	19.43	3.43
0.9375	2	83.22	95.13	10.06	2.31
1.875	2	86.09	95.17	18.56	2.43
2.8125	2	82.83	95.22	23.57	2.56
3.75	2	83.19	95.26	34.06	3.89
0.9375	3	83.42	95.72	14.05	2.30
1.875	3	85.88	95.93	25.26	2.63
2.8125	3	83.00	95.05	33.10	2.85
3.75	3	83.23	95.71	46.41	4.01
0.9375	4	83.05	96.89	18.31	2.71
1.875	4	85.96	95.51	32.60	2.86
2.8125	4	83.00	96.55	42.94	3.09
3.75	4	83.27	95.97	55.42	4.20
0.9375	5	83.00	96.43	21.75	2.77
1.875	5	86.17	96.59	37.46	3.01
2.8125	5	83.41	96.17	50.31	3.27
3.75	5	82.52	96.01	65.11	4.31
0.9375	6	82.58	96.59	26.92	2.97
1.875	6	86.63	96.05	43.33	3.23
2.8125	6	83.28	95.97	61.60	3.50
3.75	6	83.14	96.22	77.72	4.54

^a Using Intel i7 3.4 GHz processor, 8 GB RAM.

evaluation. Validation is the process of evaluating the developed MM algorithm using the calibrated MARS model.

The process of determining the set of optimal weight coefficients is explained in detail in the following sub-sections, followed by a method for constructing the MARS model. The procedural difference between the calibration and the evaluation phases is depicted in Fig. 4.

3.3.1. Procedure for determination of optimal weight coefficient combination

The steps involved in the determination of the optimal coefficient combination are enumerated here:

Step 1: Random generation of 1,000 coefficient combinations such that

$$\sum_{i=1}^{1000} C_{P,i} + C_{K,i} + C_{C,i} + C_{T,i} = 1 \quad \text{and} \quad (1)$$

$$0 \leq C_{P,i}, C_{K,i}, C_{C,i} \text{ and } C_{T,i} \leq 1$$

where,

$C_{P,i}$ is the magnitude of the coefficient of the proximity weighting score in combination i

$C_{K,i}$ is the magnitude of the coefficient of the kinematic weighting score in combination i

$C_{C,i}$ is the magnitude of the coefficient of the connectivity weighting score in combination i

$C_{T,i}$ is the magnitude of the coefficient of the turn-intent prediction weighting score in combination i

Step 2: Determination of the Total Weighting Score (TWS) for each of the coefficient combinations as follows:

$$TWS_{ij} = C_{P,i} \times W_{P,j} + C_{K,i} \times W_{K,j} + C_{C,i} \times W_{C,j} + C_{T,i} \times W_{T,j} \quad (2)$$

where, TWS_{ij} is the total weighting score for grid j obtained using the i^{th} coefficient combination

$W_{P,j}$ is the proximity weighting score awarded to grid j

$W_{K,j}$ is the kinematic weighting score awarded to grid j

$W_{C,j}$ is the connectivity weighting score awarded to grid j

$W_{T,j}$ is the turn-intent prediction weighting score awarded to grid j .

Step 3: Determination of potential grids: a grid that has the highest TWS for a particular coefficient combination is a candidate grid. This process results in 1,000 candidate grids.

Step 4: Determination of candidate coefficient combination: the radial error (Euclidean distance) between candidate grids and the true grid is computed. True grid is known because the true position is known during the calibration phase. All the grids among the candidate grids that result in the least radial error are shortlisted. The corresponding coefficient combinations are potential coefficient combinations.

Step 5: If there are multiple potential coefficient combinations, the combination that is the closest in Euclidean space to the set of combinations identified in the previous time epoch is taken as the optimal combination of the coefficients in the present time epoch.

Step 6: Repetition of Step 1 to Step 5 for the remaining epochs.

Sample size selection: The number of random coefficient combinations used in Step 1 can influence the MARS model. Therefore, the coefficient estimation error for different sample sizes is determined to identify the sample size. The coefficient estimation error, E is defined as follows: $E = \frac{1}{m} \sum_{i=1}^m \text{abs}(C_{P,i}^{1000} - C_{P,i}^N) + \text{abs}(C_{K,i}^{1000} - C_{K,i}^N) + \text{abs}(C_{C,i}^{1000} - C_{C,i}^N) + \text{abs}(C_{T,i}^{1000} - C_{T,i}^N)$, where,

$C_{P,i}^{1000}$ is the proximity weight coefficient determined using 1000 random samples for i^{th} epoch

$C_{K,i}^{1000}$ is the kinematic weight coefficient determined using 1000 random samples for i^{th} epoch

$C_{C,i}^{1000}$ is the connectivity weight coefficient determined using 1000 random samples for i^{th} epoch

$C_{T,i}^{1000}$ is the turn-intent prediction weight coefficient determined using 1000 random samples for i^{th} epoch

$C_{P,i}^N$ is the proximity weight coefficient determined using N random samples for i^{th} epoch

$C_{K,i}^N$ is the kinematic weight coefficient determined using N random samples for i^{th} epoch

$C_{C,i}^N$ is the connectivity weight coefficient determined using N random samples for i^{th} epoch

$C_{T,i}^N$ is the turn-intent prediction weight coefficient determined using N random samples for i^{th} epoch

m is total number of epochs.

Fig. 1 depicts the variation of the coefficient estimation error with the sample size. The coefficient estimation error became negligible for all the sample sizes larger than 600. The calibration being an offline task, computational efficiency is not a concern. Therefore, a sample size of 1000 is chosen in the study.

3.3.2. Brief description of MARS model

For each time epoch, a total of eight variables were derived. Five variables were employed to indicate the road network complexity; whereas, the other three variables were utilized to indicate the quality of position information in the calibration phase. Simultaneously, the optimal coefficient combination at every time epoch is also determined. The task is to express each coefficient as a function of dynamic indicators. Regression analysis can be resorted to in these situations. In the present study, an adaptive regression technique is used. This allows us not to assume any functional relationship between independent ($X1, X2, \dots, X8$) and dependent variables (C_P, C_K, C_C, C_T).

Friedman (1991a) proposed a flexible non-parametric regression modeling method, which is popular and known as Multivariate Adaptive Regression Splines. The MARS method is very popular because:

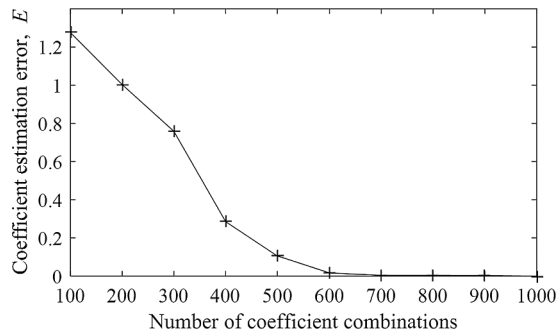


Fig. 1. Variation of coefficient estimation error with sample size.

- i. No assumption regarding the functional relationship between the variables is required. Underlying complex relationships between the variables can still be established.
- ii. The relative importance of different independent variables can be known. The method is equipped to identify the important variables and therefore the irrelevant variables are not considered while building a model. This ability is useful when several independent variables are taken into account. The issue of multicollinearity is inherently addressed.
- iii. Results in a simple, comprehensible and parsimonious model.

The domain is segmented, and a piecewise linear regression model is developed in MARS. Non-linearity is well handled by incorporating different regression slopes into different segments. Any degree of interaction between variables is allowed. Generalized Cross Validation (GCV) measure is considered as a measure of lack of fit which is given by:

$$GCV(\lambda) = \frac{\sum_{i=1}^N [y_i - \hat{f}(x_i)]^2}{\left[1 - \frac{M(\lambda)}{N}\right]^2} \quad (3)$$

where, $M(\lambda)$ is the effective number of parameters in the model, N is the sample size, $y_i = \hat{f}(x_i)$ is the model that is being estimated, y_i indicates the response variables and x_i is a set of corresponding explanatory variables.

There are primarily two stages in the model building: (i) forward model building stage: a large number of basis functions are used resulting in an over-fit model, (ii) backward pruning stage: all the basis functions that are making the insignificant contribution in reducing the GCV are removed from the model resulting in a much simpler, yet effective model. Interested readers are referred to Friedman (1991a, 1991b) for further information.

In the present research, an implementation of MARS by Jekabsons (2016) is used. A piecewise linear model is used for simplicity. A multivariate regression model is developed such that:

$$y = \hat{f}(X_1, X_2, \dots, X_8) + \varepsilon \quad (4)$$

$$C_p = \sum_{\forall i} K_{p,i} \times BF_i \quad (5)$$

$$C_K = \sum_{\forall i} K_{K,i} \times BF_i \quad (6)$$

$$C_C = \sum_{\forall i} K_{C,i} \times BF_i \quad (7)$$

$$C_T = \sum_{\forall i} K_{T,i} \times BF_i \quad (8)$$

where,

$K_{p,i}$ is the coefficient of the i^{th} basis function BF_i for coefficient of proximity weight

$K_{K,i}$ is the coefficient of the i^{th} basis function BF_i for coefficient of kinematic weight

$K_{C,i}$ is the coefficient of the i^{th} basis function BF_i for coefficient of connectivity weight

$K_{T,i}$ is the coefficient of the i^{th} basis function BF_i for coefficient of turn-intent prediction weight

ε is the fitting error.

3.4. Map-matching algorithm

Map-matching process is divided into two modules: (i) Initial MM (IMM) and (ii) Subsequent MM (SMM). Fig. 2 depicts the map-matching algorithm developed in this study. Radial Error (RE) is the radial distance between the estimated position and the position

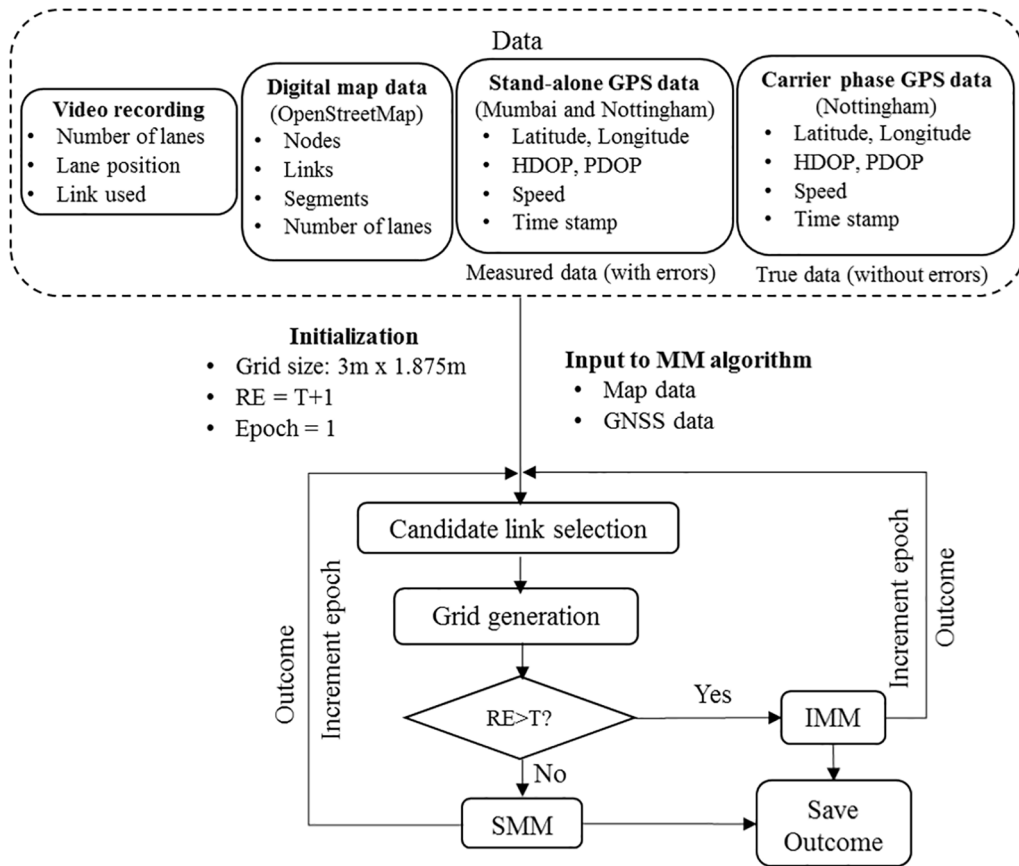


Fig. 2. Flowchart of MM algorithm.

fix used at any time epoch. Because of the measurement and process errors, RE may grow larger. Since the previously identified position is also an input in estimating the future position, the effect tends to be accumulative. To prevent the error accumulation, IMM is re-initiated if RE exceeds a threshold, T . Here, T is taken as 5 m for calibration as a larger value may not be useful in establishing model capable of position estimation at lane-level. T is taken as 40 m in validation phase, which is borrowed from Velaga et al. (2009). This conservative value for T is selected to consider the cascaded errors caused by the errors associated with positioning and digital map data.

3.4.1. Input

Road network data including the road width and GPS positioning information are the inputs to both IMM and SMM.

3.4.2. Candidate link selection

Candidate links are those that are within/crossing a circle of radius 220 m with GPS position as the center. The radius of 220 m is considered to not expel out the true link during initial stages of map-matching.

3.4.3. Grid generation

Knowing the length and width of the road segments, they are represented as array of standard grids (3 m × 1.875 m) by their centroids. A vehicle (say a car in this case) can occupy only one grid at a time. Fig. 3 depicts the road segments represented by grids.

3.4.4. Outcome of map-matching algorithm

There are three possible outcomes of either IMM or SMM process:

1. Estimated position of the vehicle (i.e., centroid of the identified grid)
2. Lateral position of the vehicle with respect to the centerline of the road
3. Radial Error (RE).

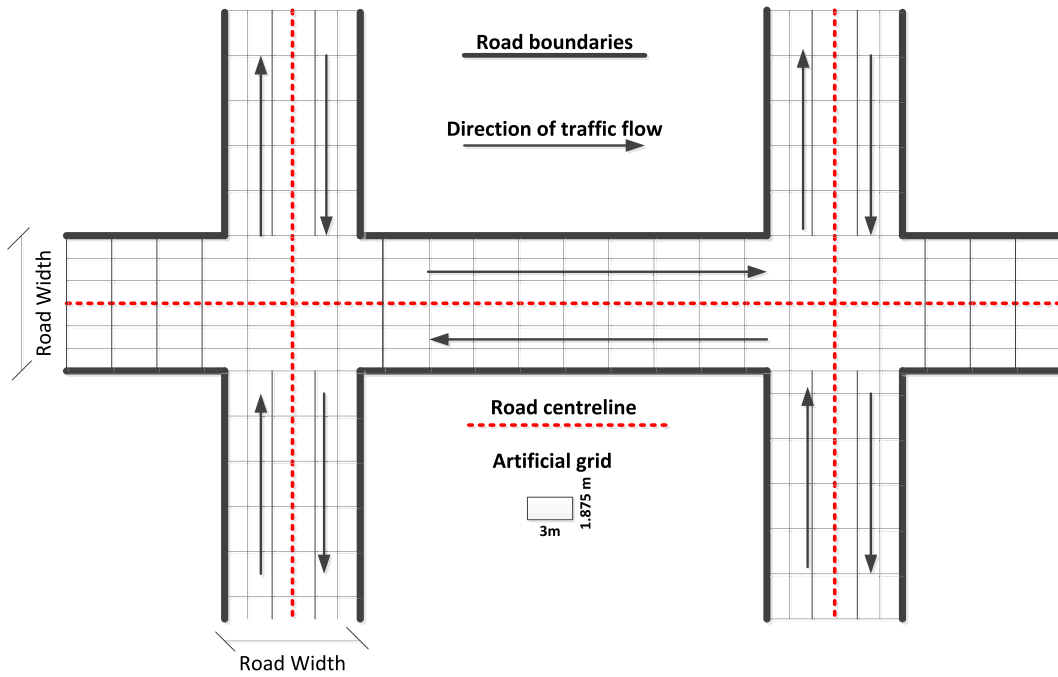


Fig. 3. A 2-D representation of fictitious road network with virtual grids.

3.5. Initial map-matching

IMM is initiated either when the historical position information is not available or when the radial error exceeds a threshold. The radial error, as mentioned earlier, may be resulted from the presence of trees, buildings, satellite-receiver geometry, quality of road maps, ionosphere, etc. The cascaded error sometimes can be very large. The threshold is, therefore, appropriately chosen to consider all these error sources (see Section 3.4).

Since the true position at a time epoch is known in the calibration phase, the grid that is the closest to the true position is considered as the grid that the vehicle occupies. Whereas during the validation phase, the grid that is the closest to the raw GPS position is considered as the grid that hosts the vehicle. Enumeration of the IMM process is as follows:

1. Consideration of the centroid of the grid that is the closest to the true/raw GPS position as the estimated position
2. Computation of lateral position of the vehicle with respect to the road centerline
3. Computation of radial error between the true/GPS position and the estimated position.

3.6. Subsequent map-matching

SMM strives to identify the probable grid the vehicle would occupy in the subsequent time epochs. Fig. 4 highlights the procedural distinctions in SMM process between the calibration and validation processes.

Different steps in SMM are enumerated in the description as follows:

1. Weighting score calculation: W_p , W_k , W_c , W_t are proximity, kinematic, connectivity, and turn-intent prediction weighting scores respectively, which are computed using Eqs. (10), (16), (17), and (21)
2. Determination of dynamic indicators: $X1$, $X2$... $X8$ are calculated, which indicate the prevailing road network complexity and position quality
3. Either Calibration or Validation/Evaluation:
 - (a) Calibration: MARS model is built to express the coefficients of weighting scores as a function of dynamic indicators. This calibration is made using the 2,415 data points of the first dataset, which is a pre-processing phase.
 - (b) Validation: For evaluating the performance of the MM algorithm, 5,830 data points from the Dataset 1b and 7,414 data points from the Dataset 2 are used. This phase involves back calculation of the dynamic weight coefficients using the calibrated MARS models.
4. Computation of the total weighting score (TWS): TWS for every grid is determined using the following equation

$$TWS = C_p \times W_p + C_k \times W_k + C_c \times W_c + C_t \times W_t \quad (9)$$

where, C_p , C_k , C_c , C_t are the coefficient of proximity, kinematic, connectivity, and turn-intent prediction weighting scores.

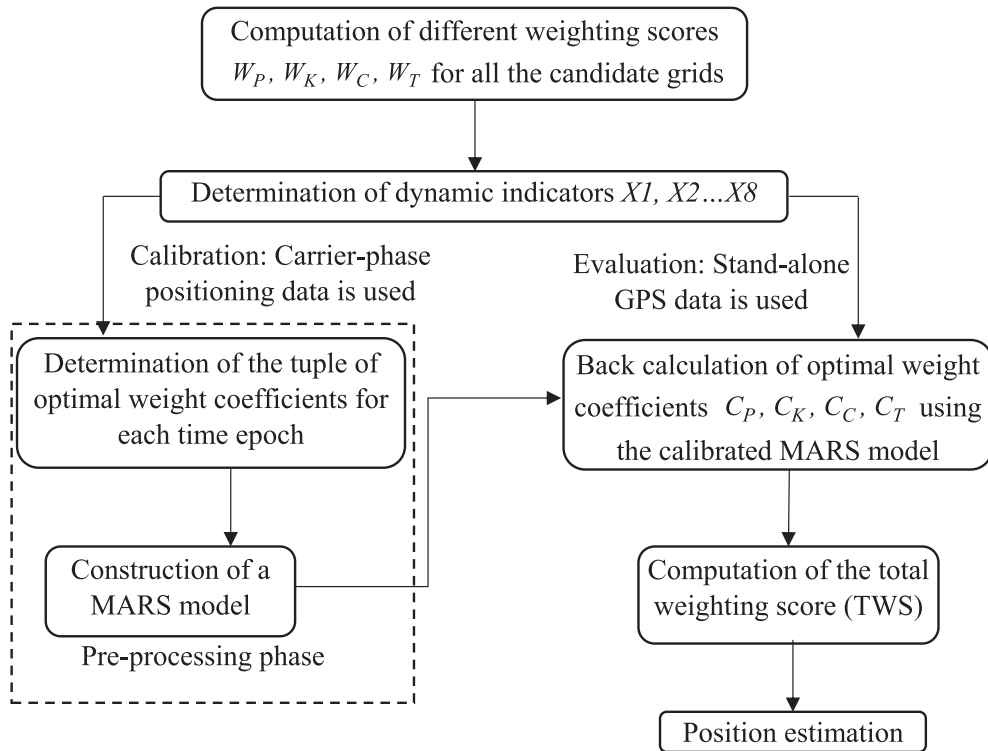


Fig. 4. Subsequent map-matching module.

5. Position estimation:

- Consideration of the grid having the highest TWS as the grid to host the vehicle
- Computation of lateral position of the vehicle with respect to the road centerline
- Computation of radial error between true/GPS position and estimated position.

A condition is imposed in both IMM and SMM to not allow the vehicle to travel in restricted direction on one-way road segments. This condition is imposed supposing no driver will break the rule.

3.7. Description of different weighting scores

All the candidate grids are awarded four different weighting scores. Each of the weight functions is described in detail in the following sub-sections.

3.7.1. Proximity weighting score

The proximity of a raw positioning point to a grid is a factor that indicates the grid that a vehicle occupies. From randomly selected 1,000 carrier-phase GPS data points from Dataset 1a, the distribution of Euclidean distance (i.e., a radial error) between a true position and a GPS raw position was determined to be lognormal with $\mu = 1.35$ and $\sigma = 0.668$; where μ and σ are parameters of the lognormal distribution. Proximity weight function is intended to award higher weighting score to the grids that are closer to the raw position information than to those that are away. The area under the lognormal distribution (i.e., cumulative distribution function) is more for the grids that are away from the raw position. Therefore, higher weighting score is provided to the grids that are closer to the raw position using the cumulative distribution function of the lognormal distribution as follows:

$$W_{P,i} = \left(1 - \frac{1}{\sigma\sqrt{2\pi}} \int_0^{d_i} \frac{1}{x} \times \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) dx \right) \times 100 \quad (10)$$

where, d_i is the radial distance between raw GPS position and centroid of the grid i . $W_{P,i}$ indicates the proximity weighting score the grid i gets.

3.7.2. Kinematic weighting score

Kalman filters are popularly used for position (state) estimation (Krakiwsky et al., 1988; Sasiadek et al., 2000; Najjar and Bonnifait, 2005; Gomez-Gil et al., 2013). Interested readers are referred to Faragher (2012), Kalman (1960) and Ribeiro (2004) for

further details. In this research, a linear Kalman filter is used to estimate the position of the vehicle at a future time epoch ($t + 1$) knowing the position of the vehicle and velocity at time t . The vehicle is assumed to travel in a two-dimensional Cartesian plane in which x and y represent the vehicle's position on that plane and the vehicle is assumed to not accelerate within one time epoch. This assumption is justified as the time epoch is small (i.e., one second). A linear Kalman filter is described to estimate the vehicle's position using equations of kinematics. The following equations describe the predictor and corrector equations used for future position estimation (Faragher, 2012).

$$S_{t+1}^- = AS_t^+ + BC_{t+1} \quad (11)$$

$$P_{t+1}^- = AP_t^+ A^T + N_P \quad (12)$$

$$K_{t+1} = P_{t+1}^- H^T (HP_{t+1}^- H^T + N_M)^{-1} \quad (13)$$

$$S_{t+1}^+ = S_{t+1}^- + K_{t+1} (M_{t+1} - HS_{t+1}^-) \quad (14)$$

$$P_{t+1}^+ = (I - K_{t+1}H)P_{t+1}^- \quad (15)$$

where,

$S = \begin{pmatrix} x \\ y \end{pmatrix}$ is State Vector

$C = v \times \begin{pmatrix} \sin \theta \\ \cos \theta \end{pmatrix}$ is Control Vector

and,

$M = \begin{pmatrix} x_{GPS} \\ y_{GPS} \end{pmatrix}$ is the Measurement Vector

M is raw GPS position (in the Cartesian plane) provided by GPS receiver.

θ is vehicle heading.

v is instantaneous speed. θ and v are not computed using current and previous positions. Rather, the apparent shift in the radio signals (Doppler effect) is analyzed to decipher the velocity. v and θ are measured but not computed (Kaplan and Hegarty, 2006; Qaddus, 2006).

K is Kalman gain.

t indicates the time epoch.

Superscript – indicates a priori estimate or predictor equation.

Superscript + indicates a posteriori estimate or corrector equation.

A , B and H are transformation matrices given by

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} B = \begin{pmatrix} \Delta t & 0 \\ 0 & \Delta t \end{pmatrix} H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

N_P is process noise matrix

N_M is measurement noise matrix

P is Variance-covariance matrix.

Knowing the position and instantaneous velocity at time $t - 1$, Kalman filter estimates the position the vehicle would acquire at time t . A grid closer to this Kalman estimated position should be awarded a higher kinematic weighting score while the grids that are farther are to be awarded lesser scores. Further, an analysis of Dataset 1a suggested that the radial error between the Kalman estimated position and true position is lognormally distributed with $\mu = 0.946$ and $\sigma = 0.558$. The kinematic weighting score can be expressed in a way similar to that of the proximity weighting score as follows:

$$W_{K,i} = \left(1 - \frac{1}{\sigma\sqrt{2\pi}} \int_0^{d_i} \frac{1}{x} \times \exp\left(\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right) dx \right) \times 100 \quad (16)$$

where, $\mu = 0.946$ and $\sigma = 0.558$.

$W_{K,i}$ indicates the kinematic weight the grid i gets.

d_i is the radial distance between Kalman filter estimated position and centroid of the grid i .

3.7.3. Connectivity weighting score

Dijkstra's algorithm is used to determine the cost of reaching different nodes from a position. The length of the segment is considered as the cost of traveling on a segment. Segment cost (S_C) is then computed as the smaller of the two associated nodal costs. The distance traveled (D) by the vehicle in one time epoch is computed as the Euclidean distance between two successive time epochs. Any segment having the segment cost equal to D will have a high probability of hosting the vehicle. Therefore, that segment (all the grids in that segment) is awarded high connectivity score. The weighting score is assumed to linearly reduce on either side of D according to Eq. (17). It is to be noted that the turn restrictions are inherently considered in this approach. Connectivity weighting

score is defined as follows:

$$W_{C,i} = \begin{cases} \frac{100}{D} S_{C,i} & \text{if } 0 \leq S_{C,i} \leq D \\ 100 \left[1 - \frac{(S_{C,i} - D)}{D} \right] & \text{if } D < S_{C,i} \leq 2D \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

where, $W_{C,i}$ is the connectivity weighting score to be awarded to all the grids of the segment i ,

$S_{C,i}$ is the cost of reaching segment i and

D is the Euclidean distance between current and previous position.

3.7.4. Turn-intent prediction weighting score

The speed with which a vehicle is traversing an intersection and the lateral position of the vehicle with respect to road centerline gives us a clue about whether the driver is willing to make a turn at an intersection. For example, if a vehicle is not slowing down below a certain threshold, we can assume that the vehicle does not intend to make any turn at that intersection. On the contrary, if a vehicle is traveling at low speed along the most extreme lane, the vehicle might be willing to make a turn (there could be several reasons for slowing including traffic interaction, speed breakers, etc.). Deviation angle is the change in the direction of travel of the vehicle between two successive time epochs. Analysis of about 6 h of GPS data from Dataset 3 and Dataset 1a suggests that a deviation angle has a normal distribution with zero mean and a standard deviation of $7^\circ/\text{s}$. 95th percentile deviation angle for all speeds less than 8 m/s was found to be $15^\circ/\text{s}$. 95th percentile deviation angle at speeds greater than 15 m/s was found to be $\approx 0^\circ/\text{s}$. This reflects the fact that the vehicles cease to make larger deviations at higher speeds. Between speeds of 8 m/s and 15 m/s, the 95th percentile deviation angle almost linearly reduced. Therefore, a function is constructed that attempts to predict the deviation angle a vehicle would make based on the speed and its lateral position with respect to the road centerline as follows:

$$A = \begin{cases} 15 & \text{if } v < 8 \\ 15 - \frac{15}{7}(v - 8) & \text{if } 8 \leq v \leq 15 \\ 0 & \text{if } v > 15 \end{cases} \quad 0 \leq A \leq 15 \quad (18)$$

where, A gives the maximum possible deviation angle at speed v (in m/s)

$$B = \frac{|LP|}{LP_{\max}} \quad 0 \leq B \leq 1 \quad (19)$$

B is a coefficient between 0 and 1. It approaches zero when the vehicle is traveling along the centerline of the road and approaches to unity when the vehicle is traveling along the road edge.

LP is the lateral position of the vehicle with respect to centerline; negative if on the left side of the centerline and vice versa.

LP_{\max} is the maximum lateral position possible, which is equal to half the road width.

$$\Delta = A \times B \times \frac{LP}{|LP|} \quad (20)$$

Δ gives the deviation angle (in degrees/second) a vehicle is expected to make in subsequent time epoch based on its current speed and lateral position.

The grids when occupied causes a deviation angle of Δ shall be awarded the highest weighting score as follows:

$$W_{T,i} = 100 \times \exp\left(\frac{-(x_i - \Delta)^2}{2\sigma^2}\right) \quad (21)$$

where, x_i is the deviation angle the vehicle has to make to occupy grid i . σ is taken as $7^\circ/\text{s}$.

4. Results and discussions

Table 3 provides the parameters of the multivariate MARS model. The weight coefficients can be computed as given in Eqs. (4)–(8). These equations relate the weight coefficient to the prevailing road network complexity and position quality. Therefore, the weight coefficients are dynamic. It is noticeable that the variables X_2 , X_5 , and $PDOP$ are not used in the MARS model. These are not significant variables or could be thought as proxy for other variables. Using these dynamic equations, the performance evaluation of the developed model is carried out and presented in Fig. 5, Tables 4 and 5.

4.1. Calibration results

see Table 3.

4.2. Validation results

Fig. 5 compares the performance of some existing map matching algorithms with the performance of the developed algorithm in

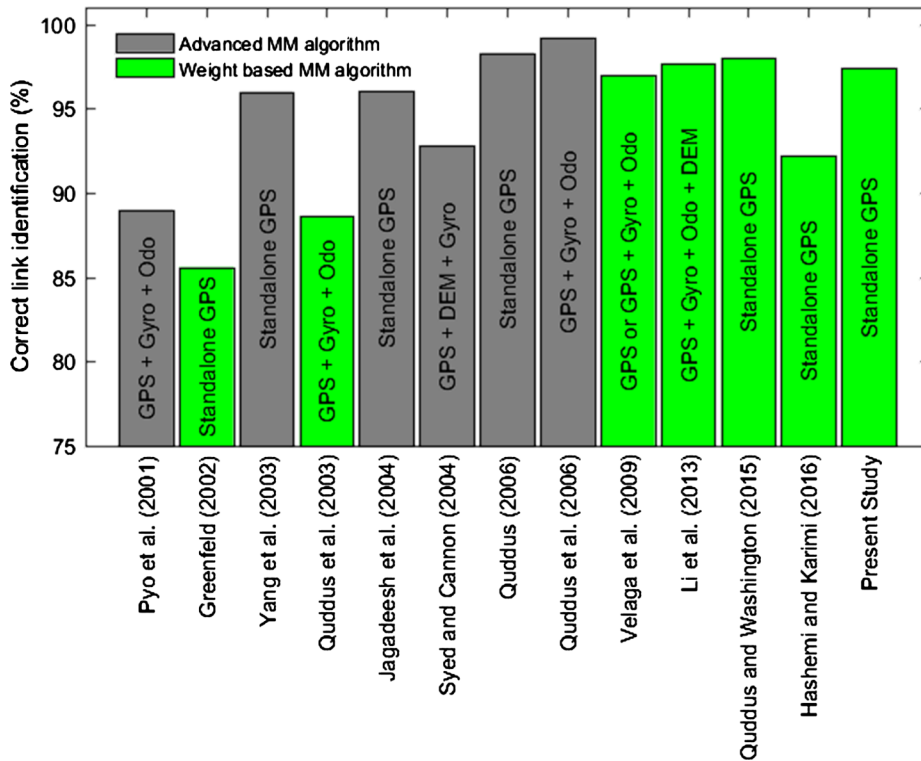


Fig. 5. Performance comparison of different map-matching algorithms. Gyro: Gyroscope. Odo: Odometer. DEM: Digital Elevation Model. (See above-mentioned references for further information.)

Table 3
Calibrated MARS model.

GCV	0.0065	0.0121	0.0053	0.0030	
Basis functions	K_P	K_K	K_C	K_T	Sum
$BF_0 = 1$	3.460E-01	3.480E-01	2.760E-01	2.995E-02	1.00
$BF_1 = \max(0, 283 - X3)$	-2.594E-03	9.040E-04	2.182E-03	-4.924E-04	0.00
$BF_2 = \max(0, X4 - 3)$	-7.233E-03	9.418E-02	-8.558E-02	-1.374E-03	0.00
$BF_3 = \max(0, 3 - X4)$	6.614E-03	4.750E-03	3.413E-03	-1.478E-02	0.00
$BF_4 = BF_1 \times \max(0, X1 - 6095)$	1.083E-05	-1.457E-05	7.461E-06	-3.717E-06	0.00
$BF_5 = BF_1 \times \max(0, 6095 - X1)$	1.089E-05	-1.264E-06	-9.859E-06	2.304E-07	0.00
$BF_6 = BF_5 \times \max(0, Speed - 19.54)$	6.266E-07	-1.765E-07	-4.574E-07	7.329E-09	0.00
$BF_7 = BF_5 \times \max(0, 19.54 - Speed)$	-5.948E-07	9.826E-08	5.032E-07	-6.627E-09	0.00
$BF_8 = \max(0, Speed - 27.13)$	-8.951E-02	1.364E-01	-2.030E-02	-2.663E-02	0.00
$BF_9 = \max(0, 27.13 - Speed)$	-1.113E-03	-2.194E-03	3.032E-03	2.753E-04	0.00
$BF_{10} = \max(0, X1 - 1242)$	-1.458E-05	1.282E-05	-8.405E-06	1.017E-05	0.00
$BF_{11} = \max(0, 1242 - X1)$	9.707E-05	1.201E-04	-2.002E-04	-1.705E-05	0.00
$BF_{12} = BF_1 \times \max(0, 1.18 \cdot Speed)$	2.461E-03	7.727E-04	-2.724E-03	-5.099E-04	0.00
$BF_{13} = \max(0, HDOP - 0.9)$	5.398E-02	-2.247E-02	-3.113E-02	-3.854E-04	0.00
$BF_{14} = \max(0, 0.9 - HDOP)$	2.791E-01	-1.250E-01	-1.124E-01	-4.174E-02	0.00
$BF_{15} = \max(0, 283 - X3) \times \max(0, Speed - 1.18) \times \max(0, X1 - 5625)$	-1.294E-07	8.828E-07	-8.457E-07	9.231E-08	0.00
$BF_{16} = \max(0, 283 - X3) \times \max(0, Speed - 1.18) \times \max(0, 5625 - X1)$	-6.394E-07	8.180E-08	5.642E-07	-6.618E-09	0.00
$BF_{17} = BF_9 \times \max(0, X3 - 245)$	-9.976E-07	5.016E-06	3.962E-07	-4.415E-06	0.00
$BF_{18} = BF_9 \times \max(0, 245 - X3)$	1.081E-04	-1.454E-04	3.763E-05	-3.597E-07	0.00

this study. In terms of the correct link identification metric, the performance of the D2D map matching algorithm that used data from stand-alone GPS is in-line with that of some existing algorithms which employed data from integrated positioning systems.

Two datasets (i.e., Dataset 1b and 2) defined earlier were used to evaluate the performance of the developed MM algorithm. The magnitude of the percentage of the correct link identification is the most popular way of describing the performance of a MM algorithm. The two-dimensional representation of the road segments permits us to carry out lane-level localization. Therefore, it is possible to express the performance of the algorithms in terms of other measurable quantities such as the percentage of the correct grid identification and the correct lane identification, which is presented in Table 4.

Table 4
Performance of the developed D2D MM algorithm.

Performance measures	Dataset 1b	Dataset 2
Sample size	5,830	7,414
Location	Nottingham, UK	Mumbai-Pune, India
Correct link identification	96.14%	98.41%
Mean horizontal accuracy	2.82 m	n/a
Horizontal accuracy (2σ)	6.61 m	n/a
Correct grid identification	71.21%	n/a
Correct lane identification	84%	79%

As the D2D map matching algorithm is capable of lane level-localization and mapping, Table 5 provides the comparison between different lane-based map-matching algorithms. The performance of the algorithm depends on many factors such as the use of additional sensors (e.g., cameras, RADARs) and lane-based road network. For example, Rabe et al. (2016a, 2016b, 2017) have reported a very high performance of a lane-level localization algorithm. However, the test track used in their study is very small. Multiple trajectories along this track cumulatively form 36 km length or 80 min duration. The test track does not represent a wide variety of operational environments. In addition, 32% of the time, vehicle travelled on a single-lane road, which certainly increases the correct lane identification. Another such example is the study by Toledo-Moreo et al. (2010), where the performance is evaluated against a very small dataset. The largest scenario considered by Toledo-Moreo et al. (2010) has the test duration of only 617 s, which may not encompass different road network complexities.

It can be observed from Table 5 that the use of high quality localization aids such as cameras, LiDAR, enhanced maps, RADARs, satellite/land-based augmentation, V2X communication, IMUs results in enhanced accuracy. However, the cost of such a system skyrockets. Furthermore, some of the localization aids such as 3D map, terrain map, lane-based map may not be readily available for all the geographical regions, especially in developing countries. These facts make many of the high performing algorithms mere academic studies and not practically viable for mass deployment in the content of a developing nation. However, the present study aims to address the situation in developing economies where standalone GPS is ubiquitous and the only option. Therefore, a standalone GPS receiver and map data obtained from OpenStreetMap have been utilized in the study.

Despite using only a standalone GPS receiver and the road network data from the OpenStreetMap, the correct lane identification by the developed D2D map matching algorithm is comparable to that of by Zheng and Hansen (2017) and Aly et al. (2016) who used additional low-cost sensors. This suggests that the integration of the low-cost localization aids has an immense potential to enhance the performance of the developed algorithm.

The algorithm can process about 15 data points per second (2.5 GHz processor with 4 GB RAM) that makes it suitable for real-time applications. From Fig. 5 and Table 5, it can be clearly observed that the developed MM algorithm outperforms (or performs at par) advanced MM algorithms based on stand-alone GNSS. The performance of the developed MM algorithm is also comparable with that of the algorithms that do not rely on a stand-alone GPS receiver. This indicates the capability of the developed MM algorithm and potential to further improve the performance with the fused sensors.

4.2.1. Understanding the source/reason of errors

Understanding the nature of the error sources and the reasons for mismatching can further assist in modifying the algorithm so as to improve the performance. Therefore, raw position, map-matched positions and the true positions were superimposed on the road network and a visual examination was conducted for Dataset 1b.

Error can creep into the MM algorithm due to the following reasons:

1. Error in road map
2. Error in GNSS position
3. Error in map matching process

The present study uses OpenStreetMap and hence it is not possible to get a quantitative knowledge regarding the map errors. However, missing links (Fig. 6), lateral displacement of links (Fig. 7) and improper representation of the intersections (Figs. 8 and 10) are evident at places.

All these three error sources could simultaneously act, but it is not possible to identify such errors (due to absence of error free roadmap). Therefore, here it is assumed at most two of the three error sources could contribute to the mismatch. Table 6 presents the breakup of these error combinations.

From Table 6, the following can be observed:

1. Correct lane identification process is more sensitive to map error (e.g. lateral translation, missing link, rotated link, etc.) as compared to correct link identification.
2. Percentage incorrect lane identification and incorrect link identification in the case of only GNSS error differ considerably. Thus, the correct lane identification is more sensitive to GNSS errors as compared to the correct link identification.
3. Percentage of errors due to map-matching process alone is less in the case of correct lane identification. This reflects the advantage

Table 5

Comparison of different lane-level localization algorithms in relation to the developed D2D map-matching algorithm.

Authors	Localization aids	Performance	Target application
Dao et al. (2007)	V2V communication, GPS	Correct lane identification = 95%	<ul style="list-style-type: none"> • Lane departure warning • Lane level navigation to reduce lane changes • Lane specific traffic information
Toledo-Moreo et al. (2010)	EGNOS, gyroscope, odometer, enhanced map	Radial accuracy Scenario 1: (0.289 ± 0.287) m Scenario 2: (0.691 ± 0.709) m Scenario 3: (0.296 ± 0.324) m	<ul style="list-style-type: none"> • Safety applications • Electronic fee collection • Lane allocation
Vu et al. (2012)	Camera, DGPS, gyroscope, accelerometer	Lateral accuracy is typically less than 1 m	<ul style="list-style-type: none"> • Vehicle control • Vehicle navigation • Safety systems • Traffic management
Schindler (2013)	High precision map, GPS, IMU, camera, LiDAR	Lateral accuracy = (0.12 ± 0.096) m	<ul style="list-style-type: none"> • Advanced Driving Assistance Systems (ADAS)
Szotzka (2013)	High frequency GPS, cameras, odometer, steering angle sensor	Only visual comparison. No quantitative comparison	<ul style="list-style-type: none"> • Vehicle navigation • ADAS
Tao et al. (2013)	GPS, speedometer, gyroscope, camera	Lateral accuracy = (0.26 ± 0.34) m	<ul style="list-style-type: none"> • Lane keeping • Parking • Obstacle avoidance
Gruyer et al. (2014)	2 cameras, accurate digital map, GPS, speedometer, gyroscope, accelerometer	Lateral accuracy < 0.1 m	<ul style="list-style-type: none"> • Obstacle avoidance
Gu et al. (2014)	3D map, gyroscopes, speedometer, GPS	Correct lane identification = 50% as reported by Kamijo et al. (2015)	<ul style="list-style-type: none"> • Autonomous driving • ADAS
Aly et al. (2015)	GPS, gyroscope, accelerometer, magnetometer, road map with known anchor locations	Correct lane identification 80%	<ul style="list-style-type: none"> • ADAS • Autonomous cars • Lane level traffic estimation • Electronic toll collection • Prediction of driver's intent
Kamijo et al. (2015)	GNSS, Speedometer, gyroscope, Camera, enhanced map	Correct lane identification = 90%	Not mentioned
Ahmed et al. (2016)	Gyroscope, accelerometer, terrain map	There is only one lane change in the experiment and is detected	<ul style="list-style-type: none"> • Lane keeping and active steering systems • Lane allocation • Automated toll collection • ADAS • Autonomous navigation • Prediction of driver's intent
Aly et al. (2016)	GPS, gyroscope, accelerometer, magnetometer, road map with known anchor locations	Correct lane identification = 84%	<ul style="list-style-type: none"> • Vehicle cooperation • Vehicle control • Autonomous navigation
Gu et al. (2016)	GNSS, 3D building model, gyroscope, speedometer	Radial accuracy (1.42 ± 0.95) m	
Kogan et al. (2016)	Camera, 3D map	Lateral accuracy = 0.18 m	
Kuhnt et al. (2016)	LiDAR, Odometer, gyroscope, enhanced map,	Radial accuracy (0.525 ± 0.279) m	<ul style="list-style-type: none"> • Autonomous driving • Collision avoidance • ADAS
Nur et al. (2016)	Camera, GPS, OpenStreetMap	Radial accuracy = (1.636 ± 0.851) m for dataset 1 and (2.522 ± 1.566) m for dataset 2	<ul style="list-style-type: none"> • Automated driving • Lane change advise • Autonomous navigation
Rabe et al. (2016a)	Camera, RADAR, GPS, enhanced map, odometer, gyroscope	Correct lane identification* = 99.68%	<ul style="list-style-type: none"> • Lane change advise • Autonomous navigation
Rabe et al. (2016b)	Camera, RADAR, GPS, enhanced map, odometer, gyroscope	Correct lane identification* (mean = 99.25%, 95th percentile = 99.5%)	<ul style="list-style-type: none"> • Lane change advise • Autonomous navigation
Kim et al. (2017)	Cameras, GPS, enhanced map, speedometer, gyroscope	Lateral accuracy (0.072 ± 0.067) m	<ul style="list-style-type: none"> • Autonomous navigation
Rabe et al. (2017)	GPS, camera, RADAR, odometer, gyroscope, enhanced map	Correct lane identification* (mean = 99.56%, 95th percentile = 99.75%,)	<ul style="list-style-type: none"> • Lane change advise • Autonomous navigation
Zheng and Hansen (2017)	Steering angle sensor, speedometer,	Lane change detection accuracy (left = 80.36% and right = 83.22%)	<ul style="list-style-type: none"> • Autonomous driving • Driver's intent prediction

(continued on next page)

Table 5 (continued)

Authors	Localization aids	Performance		Target application
Present study	Standalone GPS, map data from OpenStreetMap	Dataset 1b Correct lane identification = 84% Radial accuracy (2.82 ± 3.3) m	Dataset 2 Correct lane identification = 79% —	<ul style="list-style-type: none">• Lane departure warning• Electronic toll collection• Lane specific traffic information• Lane allocation• Lane level navigation to reduce lane changes• Prediction of driver's intent

Note:
* The values are interpreted from graphs/figures.

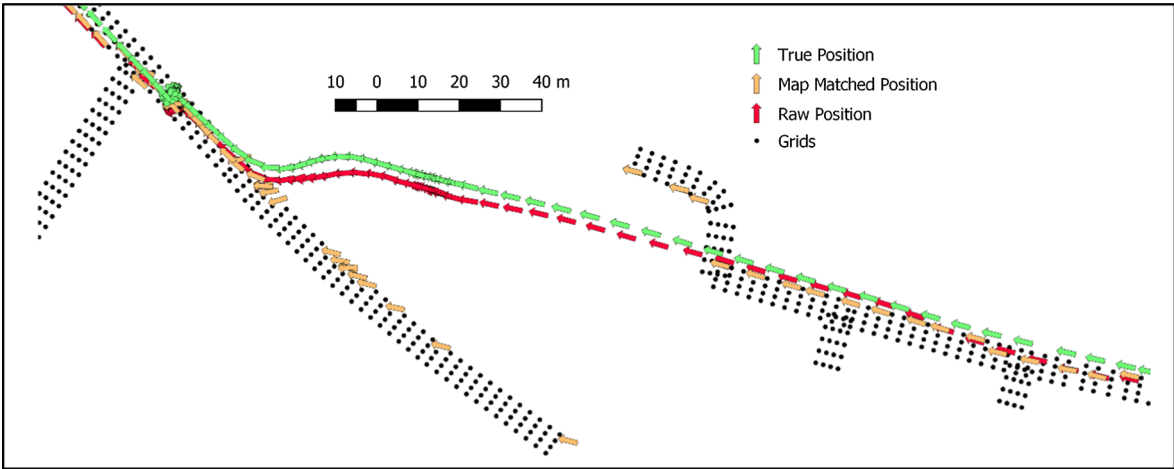


Fig. 6. Missing link resulting in erroneous map-match.

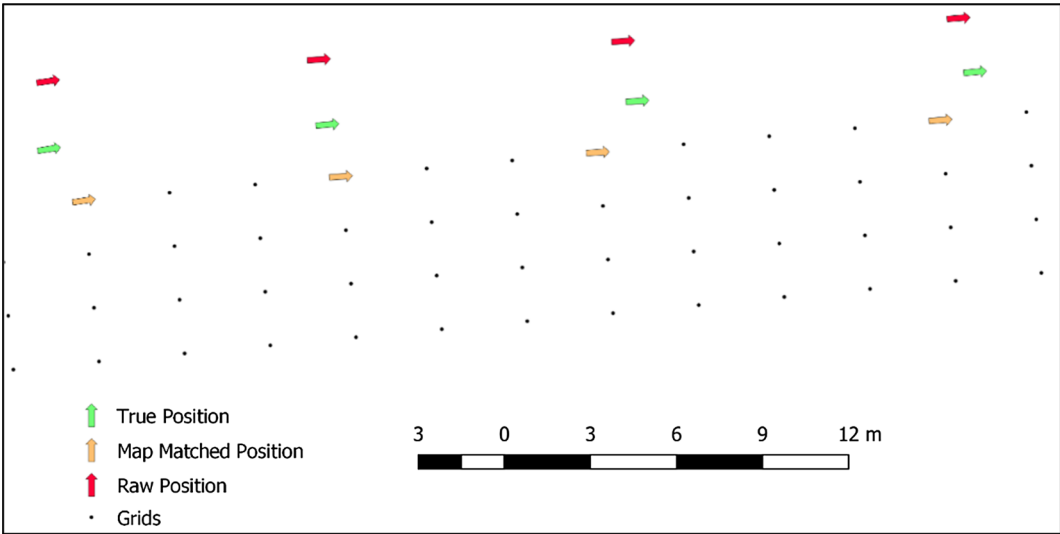


Fig. 7. Simultaneous map error and GNSS error.

of 2-dimensional representation of the road segments.
4. In general, both the map error and GNSS error contribute significantly to the incorrect lane identification and link identification process appears to be less affected.

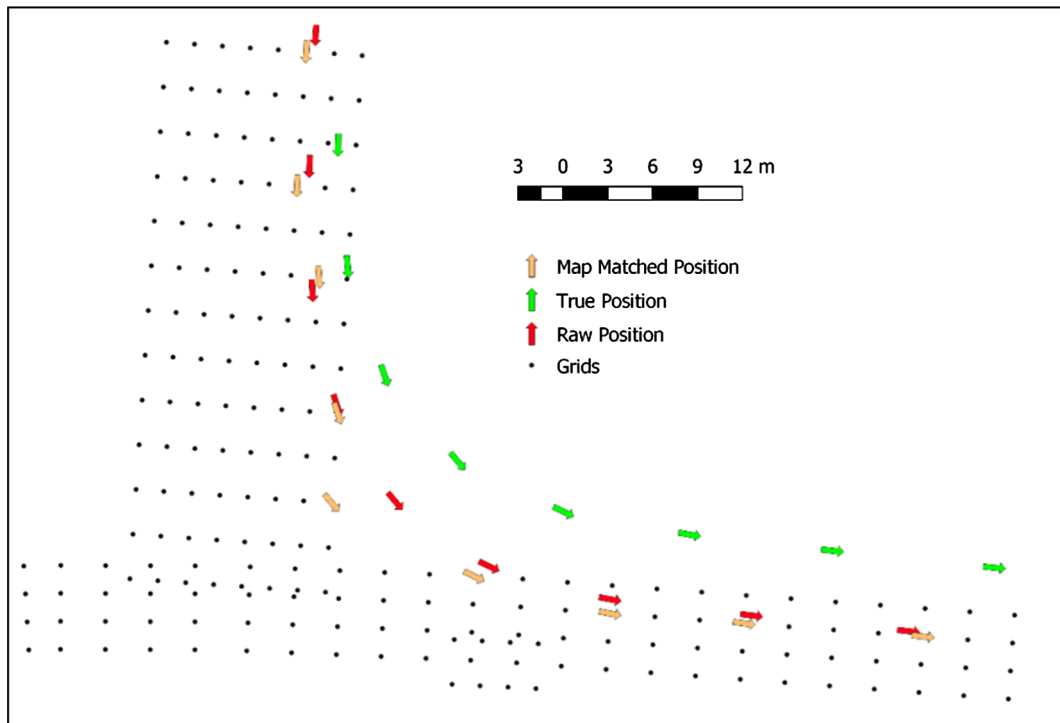


Fig. 8. Geometric error of map.

Table 6

Causes and instances for map-matching errors.

	Mismatch due to map error	Mismatch due to GNSS error	Mismatch due to map matching process
Causes and instances for the incorrect link identification (226 instances)			
Mismatch due to map error	31 (13.7%)	9 (4%)	18 (8%)
Mismatch due to GNSS error		20 (8.8%)	54 (23.9%)
Mismatch due to map matching process			94 (41.6%)
Causes and instances for the incorrect lane identification (931 instances)			
Mismatch due to map error	67 (7.2%)	137 (14.7%)	169 (18.1%)
Mismatch due to GNSS error		240 (25.8)	97 (10.4%)
Mismatch due to map matching process			221 (23.7%)

4.2.2. Critical scenarios for map-matching

Some of the typical scenarios that can result in erroneous map-matching are presented in this section.

1. Map-matching at intersection

Fig. 9c depicts a vehicle trajectory of a part of Dataset 1b. Fig. 9a provides the satellite view of a roundabout obtained from the Google Map while Fig. 9b represents the corresponding grid (2-dimensional) representation generated using the data obtained from the OpenStreetMap.

The error in correct lane identification at the roundabout can be observed from Fig. 9b. Such an error prevails whenever the vehicle makes a sharp turn until the MM algorithm recovers from the error. This error may be attributed to the assumption that a vehicle does not accelerate or change the direction within one time epoch (i.e. 1 s). A sampling rate greater than 1 Hz may alleviate this problem.

Another major source of error relates to the inaccurate representation of a road element in OpenStreetMap. For instance, Fig. 10a is a snapshot of a road element obtained from the Google map and this clearly shows that there is a roundabout. The same road element is being represented by a four-legged intersection in OpenStreetMap.

Some of the major issues with opensource maps are the oversimplified representation of a road network and the measurement of quality and accuracy. Complex road geometry with curved segments is simplified by linear segments. In this case, it can be clearly observed from Fig. 10 that the roundabout representation in OpenStreetMap is erroneous. Under such situations, the map-matching algorithm fails. It can be observed from Fig. 10b that the algorithm recovers from the error quickly indicating its resilience.

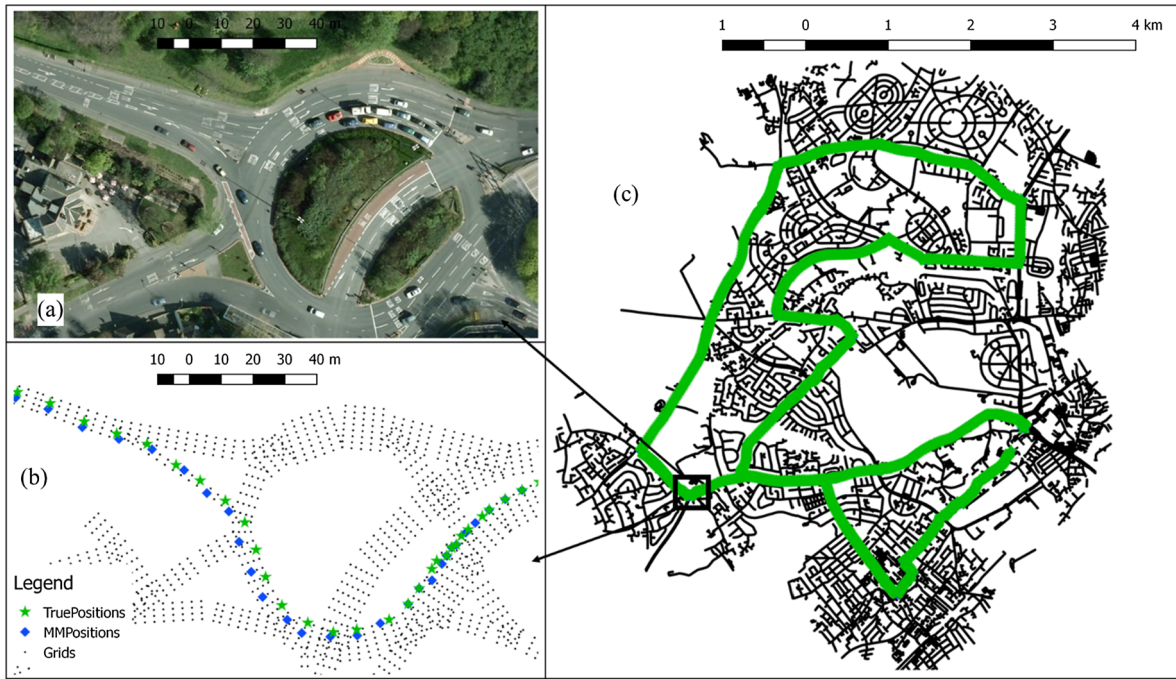


Fig. 9. Vehicle trajectory along the roundabout, grid representation and OpenStreetMap.

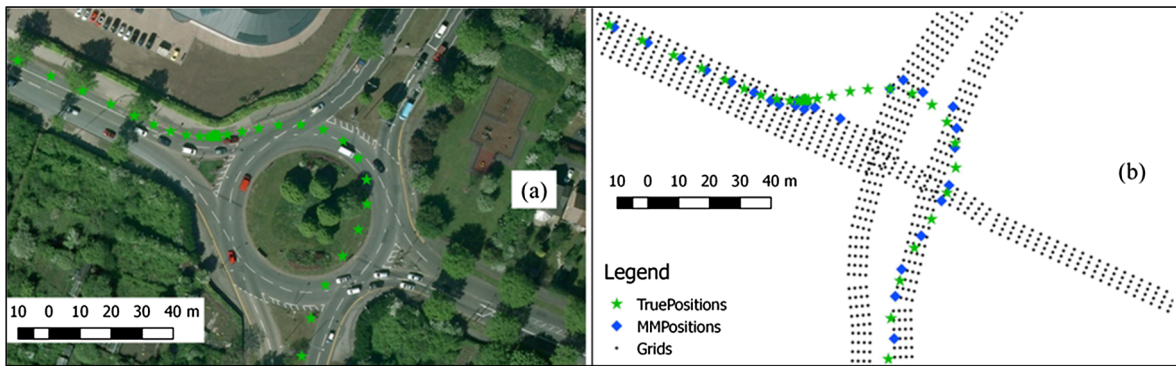


Fig. 10. An example of erroneous intersection representation.

2. Map-matching of slow-moving vehicle

Whenever the vehicle stops or moves with very low speeds, the GNSS information becomes erroneous, which is depicted as clustered red arrows in Fig. 11. Map-matching process gets hampered under this situation.

3. Map-matching on parallel road links

Fig. 12 provides a typical scenario where two roads run parallel to each other. Despite the raw position being erroneous, the developed map matching algorithm correctly identifies the correct lane and link.

5. Reasoning for improved localization accuracy

This section briefly explains the reasons why the developed D2D map-matching algorithm can provide lane-level localization even though it only uses data from standalone GPS and OpenStreetMap.

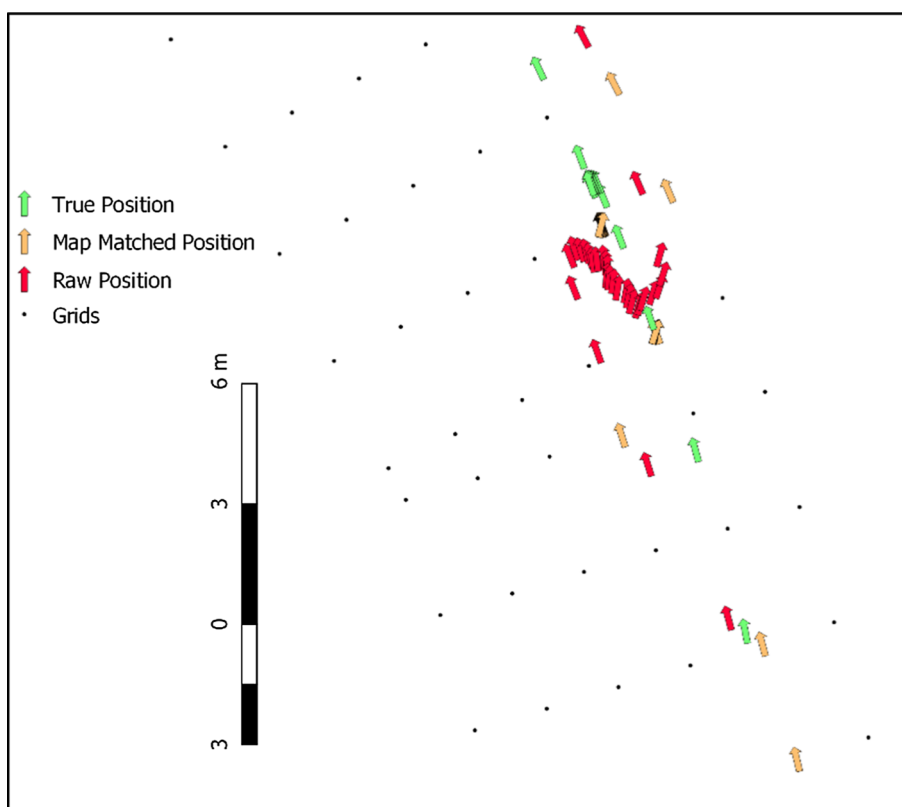


Fig. 11. GPS error during slow movement of the vehicle.

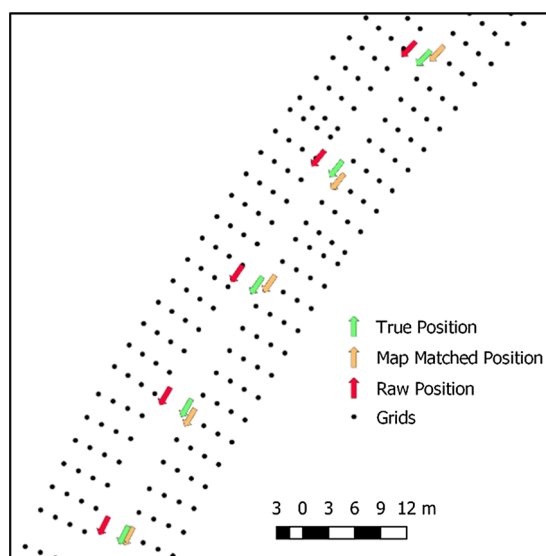


Fig. 12. GPS error in parallel road links.

5.1. Ability to overcome lateral translation error

The ability of the developed D2D MM algorithm to overcome the lateral translation error of a road segment is demonstrated via a hypothetical example. As the objective is to highlight how the lateral translation error (i.e. a type of map error) is addressed, we only consider translational error and assume that all other error sources are non-prevalent (which is of course far from reality). Therefore, we assume that the positioning sensor is error free.

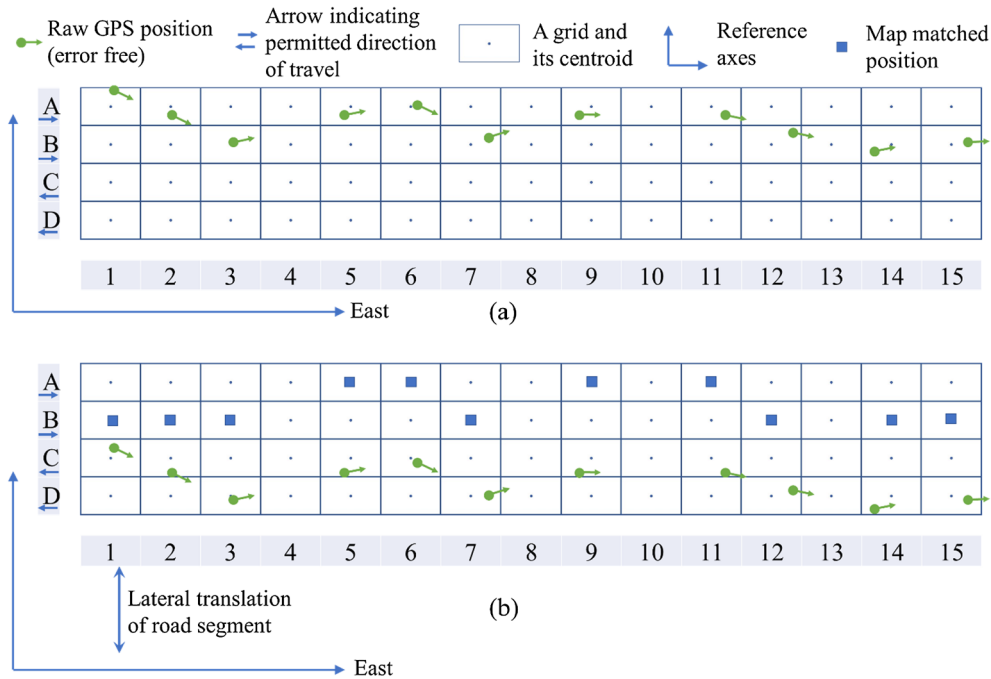


Fig. 13. Two-way two-lane road segment (a) without translation error and (b) with lateral translation error.

Fig. 13 shows a two-way two-lane road segment represented by grids and their centroids. To demonstrate the inherent ability of the developed algorithm, we consider that the road segment in Fig. 13a is free from any lateral translation error whereas the road segment in Fig. 13b suffers an error. The grids along lanes A and B host vehicles heading East while the grids along lanes C and D host the vehicles travelling in the opposite direction.

The subject vehicle is assumed to be heading East and hence the algorithm assigns the vehicle's positions only along lanes A and B. Green stars indicate the true positions (i.e. error free GPS positions) of this vehicle at different time epochs. It can be observed that the vehicle positions with respect to the reference axes is the same in both Fig. 13a and b demonstrating that Fig. 7b contains translational map error.

Few observations can be made with reference to Fig. 13b. Using only the proximity information, the initial map matched position for the first epoch would be grid B1, which is the closest valid grid that can host the vehicle. Then the subsequent map-matching module tries to map-match the subsequent positioning points. Vehicle kinematics is also considered in this phase. The turn-intent prediction weight score and the connectivity weight score are not considered as they are not relevant in this hypothetical example.

The kinematic weight score considers the kinematics of the vehicle. Knowing the present position, instantaneous speed and heading, the position of the vehicle at the future epoch is estimated. Considering the vehicle's present position (grid B1) and the instantaneous heading (i.e. the direction of the line that connects the first and the second green stars) at the first epoch, the vehicle is predicted to be in grid C2 at the second time epoch and hence, grid C2 acquires the highest kinematic weight score for the second epoch. However, C2 being an invalid grid, grid B2 that acquires the second highest kinematic weight score is identified by the algorithm as the vehicle's position. Similarly, at the third epoch, both the kinematic and proximity weight scores identify the same grid.

For the fourth epoch, A5 receives higher kinematic weight score compared to the proximity weight score obtained by grid B5. The total weight score acquired by grid A5 (considering the dynamic weight coefficients) would be greater than that of all other grids. Thus, grid A5 would be chosen as the vehicle's position. On the same lines, all the grid IDs identified by the MM algorithm for subsequent time epochs are provided in Table 7.

Table 7 provides true grids and the grids identified by different weight scores at different time epochs. It illustrates how the lateral translation error of the road segment is overcome within the three time-epochs. It can be observed that the lane-level localization

Table 7

Grids identified by map matching algorithm based on different weight scores.

Epoch	1	2	3	4	5	6	7	8	9	10	11
True grid	A1	A2	B3	A5	A6	B7	A9	A11	B12	B14	B15
Grid identified by kinematic weight score	n/a	B2	B3	A5	A6	B7	A9	A11	B12	B14	B15
Grid identified by proximity weight score	B1	B2	B3	B5	B6	B7	B9	B11	B12	B14	B15
Grid with the highest total weight score	B1	B2	B3	A5	A6	B7	A9	A11	B12	B14	B15

accuracy would drastically reduce if only proximity information is used (however, the correct link identification accuracy may not be affected). Therefore, the above demonstration clearly depicts the ability of the developed D2D map-matching method to overcome the lateral displacement error of a road. It can be noted that the longitudinal displacement error of a road map may not affect the lane prediction accuracy.

Inclusion of the kinematic weight score is the main reason for the algorithm's capability to overcome any potential lateral translation error. Because of the translation error, the distance between true grids on which the vehicle is traveling and raw position increases, and as a result the true grids naturally are assigned lower proximity weight scores. On the contrary, the kinematic weight score function awards a higher score to the true grids. Because of this disparity between magnitudes of proximity and kinematic weight scores whenever a road segment suffers lateral translation error, kinematic weight score assumes higher importance and contributes the largest towards the total weight score. This can be observed in the hypothetical example presented in Table 7, wherein, the corresponding elements in third and the fifth rows are the same.

5.2. Data-driven weighting functions

Most of the other weight-based map-matching algorithms arbitrarily define the weight functions. For example, Hashemi and Karimi (2016) have defined the Proximity weight function as $W_{p,i} = \frac{p_i}{\sum_j p_j}$; where, p_i is the proximity or the shortest distance between a GPS point and link i .

There is no logical reasoning behind this linear functional formulation of the weight score but the formulation is rather arbitrary. Such weight functions do not guarantee that different road segments would receive appropriate weight scores in order to achieve the optimal performance.

However, in the present study, the formulation of the weight score is data driven. For example, the proximity weight score function is obtained after studying the distribution of the radial errors between the true and raw GPS positions. The data driven non-linear proximity weight function ensures that the grids are awarded appropriate proximity weight score.

5.3. Dynamic weight coefficients

Literature clearly states that the magnitude of coefficients of the weight scores plays a significant role in determining the performance of the map matching algorithm (Quddus et al., 2003; Velaga et al., 2009; Quddus and Washington, 2015; Hashemi and Karimi, 2016). Literature has also given clear directions regarding the dependence of these coefficients on road network complexity (Velaga et al., 2009) and quality of the positioning information (Binjammaz et al., 2016; Hashemi and Karimi, 2016). The road network complexity and the positioning quality are spatially and temporally dynamic. Therefore, the coefficients should be governed by these dynamic indicators.

The usage of dynamic weight coefficients ensures that the grids receive appropriate total weight score considering the prevailing road network complexity and the quality of the positioning information. The fact that the magnitude of the coefficients adapts according to these dynamic indicators gives us a strong direction towards developing a map-matching algorithm that is transferrable.

5.4. Optimal weight coefficients

Optimality of the weight coefficient ensures the best performance of the map-matching algorithm (Velaga et al., 2009; Quddus and Washington, 2015). As mentioned earlier, the magnitude of the weight coefficients should be governed by the road network complexity and the quality of the positioning information. However, the magnitude of each coefficient should also ensure optimal performance of the map-matching algorithm.

We have presented a novel methodology, which ensures dynamic weight coefficients obtained at every time epoch results in an optimal estimate of the vehicle's position.

In summary, despite the capability of lane-level localisation, the system is not fool-proof as a standalone GNSS receiver is used instead of complementary and expensive sensors such as radars. However, the algorithm can still be applied to many liability critical and in-vehicle information systems.

Although the link connectivity information used in the developed MM algorithm can prevent some mismatches in grade separated roads, for accurate map-matching, the algorithm needs to utilize the elevation information from road network as well as from GNSS. The state space presented in Section 3.7 is defined primarily for achieving the lane-level localization. Furthermore, elevation information of the road segments is not available within OpenStreetMap data used in this study. Hence the elevation information from the GPS is not utilised. Therefore, the algorithm in the presented form may not efficiently distinguish between multi-level roads although the connectivity information between the segments prevents mismatches.

6. Conclusions

A new dynamic 2-dimensional (D2D) weight-based topological MM algorithm was developed to achieve lane-level localization. The novel features of the developed algorithm are: (i) the use of positioning fixes from stand-alone GPS only, (ii) data-driven dynamic weight coefficients and (iii) its ability to overcome lateral translational errors and (iv) its transferability. The condition was warranted considering the situation in the developing economies, where augmentation may not be feasible. In the developed MM algorithm, the

road network is represented as an array of virtual grids in contrast to road centerlines. This is a novel idea allowing us to localize a vehicle at a lane-level with only one positioning sensor - GPS. Three new weighting scores were introduced to aid map-matching: (i) kinematic weight was based on the vehicle kinematics and prediction of its position in future time epochs, (ii) connectivity weight considered link connectivity and traffic movement restrictions simultaneously, and finally (iii) the intent to make a turn at an intersection was captured by the turn-intent prediction weighting score.

Eight dynamic indicators were used; five to express the road network complexity and three to indicate the quality of positioning fixes from a GPS receiver. An approach to determine the optimal coefficients of different weighting scores in minimizing the radial error between the estimated and true position was developed. These set of optimal coefficients were then expressed as a function of the dynamic indicators using an adaptive regression technique. The issue of transferability of the developed D2D map-matching algorithm was addressed by calibrating and testing the algorithm by employing data from heterogeneous road networks.

Dynamic weight coefficients used in this study provided a means to effectively consider the quality of the position information and prevailing complexity of the road network. The dynamic coefficients, therefore, facilitated the MM algorithm to resolve the ambiguity that arose in complex road network more effectively resulted in a high performing MM algorithm.

The algorithm was tested on two real-world datasets obtained from stand-alone GPS receivers. The algorithm identifies 96.1% and 98.4% links correctly for the GPS datasets collected in Nottingham, UK and Mumbai-Pune, India respectively. This performance is comparable with 98.6% as reported by Quddus and Washington (2015). Despite using *OpenStreetMap* as a base map, the present study achieves a horizontal accuracy of 6.61 m (2σ) in comparison with 7.35 m (2σ) as achieved by Quddus and Washington (2015). Furthermore, the mean horizontal accuracy of 2.82 m achieved in the present study is primarily attributed to the lane-level representation of the road network. In terms of the correct lane prediction, the algorithm correctly predicted 84% and 79% of the time for both the datasets. The developed MM algorithm is capable of processing about 25 data points per second (with Intel i7 3.4 GHz processor, 8 GB RAM), which renders it suitable for real-time ITS applications demanding lane-level localization.

This new D2D map-matching algorithm is expected to make a profound impact on ITS industry and vehicle manufacturers as this algorithm can be employed in real-time to support several liability critical and in-vehicle information systems such as lane specific traffic diversion, lane allocation, lane-based electronic toll collection, lane-based navigation and the prediction of driver's intent.

Acknowledgment

This research was funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC) (Grant reference: EP/F018894/1). The author takes full responsibility for the content of the paper and any errors or omissions. Research data for this paper are available on request from Mohammed Quddus (m.a.quddus@lboro.ac.uk).

References

- Ahmed, H., Tahir, M., Ali, K., 2016. Terrain based GPS independent lane-level vehicle localization using particle filter and dead reckoning. In: 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall). IEEE, pp. 1–5. <https://doi.org/10.1109/VTCFall.2016.7881249>.
- Alkhorshid, Y., Aryafar, K., Wanielik, G., Shokoufandeh, A., 2015. Camera-based lane marking detection for ADAS and autonomous driving, pp. 514–519. https://doi.org/10.1007/978-3-319-20801-5_57.
- Aly, H., Basalamah, A., Youssef, M., 2016. Robust and ubiquitous smartphone-based lane detection. *Pervasive Mob. Comput.* 26, 35–56. <https://doi.org/10.1016/j.pmcj.2015.10.019>.
- Aly, H., Basalamah, A., Youssef, M., 2015. LaneQuest: An accurate and energy-efficient lane detection system. In: 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, St. Louis, MO, USA, pp. 163–171. <https://doi.org/10.1109/PERCOM.2015.7146523>.
- Aponte, J., Meng, X., Moore, T., Hill, C., 2009. Assessing network RTK wireless delivery. *GPS World* 14–19.
- Benker, S.C., Langford, R.P., Pavlis, T.L., 2011. Positional accuracy of the Google Earth terrain model derived from stratigraphic unconformities in the Big Bend region, Texas, USA. *Geocarto Int.* 26, 291–303. <https://doi.org/10.1080/10106049.2011.568125>.
- Binjammaz, T.A., Al-Bayatti, A., Al-Hargan, A.H., 2016. Context-aware GPS integrity monitoring for intelligent transport systems. *J. Traffic Transp. Eng. (English Ed.)* 3, 1–15. <https://doi.org/10.1016/j.jtte.2015.09.002>.
- Bernstein, D., Kornhauser, A., 1996. An Introduction to Map Matching for Personal Navigation Assistants.
- Bhuyan, P.K., Borah, R.R., 2007. TEC derived from GPS network in India and comparison with the IRI. *Adv. Sp. Res.* 39, 830–840. <https://doi.org/10.1016/j.asr.2006.12.042>.
- Bouju, A., Stockus, A., Bertrand, R., Boursier, P., 2002. Location-based spatial data management in navigation systems. In: Intelligent Vehicle Symposium, 2002. IEEE, pp. 172–177. <https://doi.org/10.1109/IVS.2002.1187947>.
- Cui, D., Xue, J., Zheng, N., 2016. Real-time global localization of robotic cars in lane level via lane marking detection and shape registration. *IEEE Trans. Intell. Transp. Syst.* 17, 1039–1050. <https://doi.org/10.1109/TITS.2015.2492019>.
- Dao, T., Leung, K.Y.K., Clark, C.M., Huissoon, J.P., 2007. Markov-based lane positioning using intervehicle communication. *IEEE Trans. Intell. Transp. Syst.* 8, 641–650. <https://doi.org/10.1109/TITS.2007.908574>.
- DeCesare, N.J., Squires, J.R., Kolbe, J.A., 2005. Effect of forest canopy on GPS-based movement data. *Wildl. Soc. Bull.* 33, 935–941. [https://doi.org/10.2193/0091-7648\(2005\)33\[935:EOFCOG\]2.0.CO;2](https://doi.org/10.2193/0091-7648(2005)33[935:EOFCOG]2.0.CO;2).
- Du, J., Barth, M.J., 2008. Next-generation automated vehicle location systems: positioning at the lane level. *IEEE Trans. Intell. Transp. Syst.* 9, 48–57. <https://doi.org/10.1109/TITS.2007.908141>.
- Du, J., Masters, J., Barth, M., 2004. Lane-level positioning for in-vehicle navigation and automated vehicle location (AVL) systems. In: Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749). IEEE, pp. 35–40. <https://doi.org/10.1109/ITSC.2004.1398868>.
- Faragher, R., 2012. Understanding the basis of the Kalman filter via a simple and intuitive derivation [lecture notes]. *IEEE Signal Process. Mag.* 29, 128–132. <https://doi.org/10.1109/MSP.2012.2203621>.
- Friedman, J.H., 1991a. Multivariate adaptive regression splines. *Ann. Stat.* 19, 1–67. <https://doi.org/10.1214/aos/1176347963>.
- Friedman, J.H., 1991b. Estimating Functions of Mixed Ordinal and Categorical Variables Using Adaptive Splines.
- Gomez-Gil, J., Ruiz-Gonzalez, R., Alonso-Garcia, S., Gomez-Gil, F., 2013. A Kalman filter implementation for precision improvement in low-cost GPS positioning of tractors. *Sensors* 13, 15307–15323. <https://doi.org/10.3390/s131115307>.
- Greenfield, J.S., 2002. Matching GPS observations to locations on a digital map. In: Transportation Research Board 81st Annual Meeting.
- Gruyer, D., Belaroussi, R., Revilloud, M., 2014. Map-aided localization with lateral perception. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, pp.

- 674–680. <https://doi.org/10.1109/IVS.2014.6856528>.
- Gu, Y., Hsu, L.-T., Kamijo, S., 2016. GNSS/onboard inertial sensor integration with the aid of 3-D building map for lane-level vehicle self-localization in urban canyon. *IEEE Trans. Veh. Technol.* 65, 4274–4287. <https://doi.org/10.1109/TVT.2015.2497001>.
- Gu, Y., Wada, Y., Hsu, L., Kamijo, S., 2014. Vehicle self-localization in urban canyon using 3D map based GPS positioning and vehicle sensors. In: 2014 International Conference on Connected Vehicles and Expo (ICCVE). IEEE, pp. 792–798. <https://doi.org/10.1109/ICCVE.2014.7297660>.
- Hashemi, M., Karimi, H.A., 2016. A weight-based map-matching algorithm for vehicle navigation in complex urban networks. *J. Intell. Transp. Syst.* 20, 573–590. <https://doi.org/10.1080/15472450.2016.1166058>.
- Hashemi, M., Karimi, H.A., 2014. A critical review of real-time map-matching algorithms: Current issues and future directions. *Comput. Environ. Urban Syst.* 48, 153–165. <https://doi.org/10.1016/j.compenvurbysys.2014.07.009>.
- He, Q., Head, K.L., 2010. Pseudo-Lane-Level, Low-Cost GPS Positioning with Vehicle-to-Infrastructure Communication and Driving Event Detection. In: 13th International IEEE Conference on Intelligent Transportation Systems. IEEE, Funchal, Madeira Island, Portugal, pp. 1669–1676.
- Jagadeesh, G.R., Srikanthan, T., Zhang, X.D., 2004. A map matching method for GPS based real-time vehicle location. *J. Navig.* 57, 429–440. <https://doi.org/10.1017/S0373463304002905>.
- Jekabsons, G., 2016. ARESLab: adaptive regression splines toolbox.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35. <https://doi.org/10.1115/1.3662552>.
- Kamijo, S., Gu, Y., Hsu, L.-T., 2015. GNSS/INS/on-board camera integration for vehicle self-localization in urban canyon. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, pp. 2533–2538. <https://doi.org/10.1109/ITSC.2015.407>.
- Kaplan, E.D., Hegarty, C.J., 2006. *Understanding GPS: Principles and Applications*, second ed. Artech House Inc, Norwood.
- Kim, D., Kim, B., Chung, T., Yi, K., 2017. Lane-level localization using an AVM camera for an automated driving vehicle in urban environments. *IEEE/ASME Trans. Mechatron.* 22, 280–290. <https://doi.org/10.1109/TMECH.2016.2533635>.
- Knapen, L., Bellemans, T., Janssens, D., Wets, G., 2018. Likelihood-based offline map matching of GPS recordings using global trace information. *Transp. Res. Part C Emerg. Technol.* 93, 13–35. <https://doi.org/10.1016/j.trc.2018.05.014>.
- Kogan, V., Shimshoni, I., Levi, D., 2016. Lane-level positioning with sparse visual cues. In: 2016 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 889–895 <https://doi.org/10.1109/IVS.2016.7535493>.
- Krakiwsky, E.J., Harris, C.B., Wong, R.V.C., 1988. A Kalman filter for integrating dead reckoning, map matching and GPS positioning. In: IEEE PLANS '88., Position Location and Navigation Symposium, Record. "Navigation into the 21st Century". IEEE, pp. 39–46 <https://doi.org/10.1109/PLANS.1988.195464>.
- Kuhnt, F., Orf, S., Klemm, S., Zollner, J.M., 2016. Lane-precise localization of intelligent vehicles using the surrounding object constellation. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 526–533 <https://doi.org/10.1109/ITSC.2016.7795605>.
- Li, F., Bonnfait, P., Ibanez-Guzman, J., Zinoune, C., 2017. Lane-level map-matching with integrity on high-definition maps. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 1176–1181 <https://doi.org/10.1109/IVS.2017.7995872>.
- Li, L., Quddus, M.A., Zhao, L., 2013. High accuracy tightly-coupled integrity monitoring algorithm for map-matching. *Transp. Res. Part C Emerg. Technol.* 36, 13–26. <https://doi.org/10.1016/j.trc.2013.07.009>.
- Li, Q., Chen, L., Li, M., Shaw, S.-L., Nuchter, A., 2014. A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios. *IEEE Trans. Veh. Technol.* 63, 540–555. <https://doi.org/10.1109/TVT.2013.2281199>.
- Lopes, E.E., Nogueira, R.E., 2011. Proposta Metodológica para Validação de Imagens de Alta Resolução do Google Earth para a Produção de Mapas, in: XV Simpósio Brasileiro de Sensoriamento Remoto - SBSR. Curitiba, Brazil, p. 2308.
- Mukheja, P., Kiran, K.M., Velaga, N.R., Sharmila, R.B., 2017. Smartphone-based crowdsourcing for position estimation of public transport vehicles. *IET Intell. Transp. Syst.* 11, 588–595. <https://doi.org/10.1049/iet-its.2016.0247>.
- Najjar, M.E.E., Bonnfait, P., 2005. A road-matching method for precise vehicle localization using belief theory and Kalman filtering. *Auton. Robots* 19, 173–191. <https://doi.org/10.1007/s10514-005-0609-1>.
- Nur, D., Dharmiza, S., Salleh, A., Seignez, E., 2016. Monocular visual odometry with road probability distribution factor for lane-level vehicle localization. In: 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, pp. 1–6 <https://doi.org/10.1109/ICARCV.2016.7838777>.
- Ohno, K., Tsubouchi, T., Shigematsu, B., Maeyama, S., Yuta, S., 2003. Outdoor navigation of a mobile robot between buildings based on DGPS and odometry data fusion. In: 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422). IEEE, pp. 1978–1984 <https://doi.org/10.1109/ROBOT.2003.1241884>.
- Oyama, Y., Hato, E., 2017. A discounted recursive logit model for dynamic gridlock network analysis. *Transp. Res. Part C Emerg. Technol.* 85, 509–527. <https://doi.org/10.1016/j.trc.2017.10.001>.
- Oyama, Y., Hato, E., 2018. Link-based measurement model to estimate route choice parameters in urban pedestrian networks. *Transp. Res. Part C Emerg. Technol.* 93, 62–78. <https://doi.org/10.1016/j.trc.2018.05.013>.
- Peyret, F., Bétaille, D., Laneurit, J., Toledo-Moreo, R., 2008. Lane-level positioning for cooperative systems using EGNOS and enhanced digital maps. *Proc. ENC GNSS*.
- Potere, D., Schneider, A., Angel, S., Civco, D., 2009. Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *Int. J. Remote Sens.* 30, 6531–6558. <https://doi.org/10.1080/01431160903121134>.
- Pyo, J.S., Shin, D.-H., Sung, T.-K., 2001. Development of a map matching method using the multiple hypothesis technique. In: ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585). IEEE, pp. 23–27 <https://doi.org/10.1109/ITSC.2001.948623>.
- Qin, G., Huang, Z., Xiang, Y., Sun, J., 2019. ProbDetect: A choice probability-based taxi trip anomaly detection model considering traffic variability. *Transp. Res. Part C Emerg. Technol.* 98, 221–238. <https://doi.org/10.1016/j.trc.2018.11.016>.
- Quddus, M.A., 2006. *High Integrity Map Matching Algorithms for Advanced Transport Telematics Applications*. Imperial College London, London.
- Quddus, M.A., Noland, R.B., Ochieng, W.Y., 2006. A high accuracy fuzzy logic based map matching algorithm for road transport. *J. Intell. Transp. Syst. Technol. Planning, Oper.* 10, 103–115. <https://doi.org/10.1080/15472450600793560>.
- Quddus, M.A., Ochieng, W.Y., Noland, R.B., 2007. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transp. Res. Part C Emerg. Technol.* 15, 312–328. <https://doi.org/10.1016/j.trc.2007.05.002>.
- Quddus, M.A., Ochieng, W.Y., Zhao, L., Noland, R.B., 2003. A general map matching algorithm for transport telematics applications. *GPS Solut.* 7, 157–167. <https://doi.org/10.1007/s10291-003-0069-z>.
- Quddus, M.A., Washington, S., 2015. Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. *Transp. Res. Part C Emerg. Technol.* 55, 328–339. <https://doi.org/10.1016/j.trc.2015.02.017>.
- Rabe, J., Hubner, M., Necker, M., Stiller, C., 2017. Ego-lane estimation for downtown lane-level navigation. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 1152–1157 <https://doi.org/10.1109/IVS.2017.7995868>.
- Rabe, J., Meinke, M., Necker, M., Stiller, C., 2016a. Lane-level map-matching based on optimization. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1155–1160 <https://doi.org/10.1109/ITSC.2016.7795702>.
- Rabe, J., Necker, M., Stiller, C., 2016b. Ego-lane estimation for lane-level navigation in urban scenarios. In: 2016 IEEE Intelligent Vehicles Symposium (IV). IEEE, Gothenburg, Sweden, pp. 896–901 <https://doi.org/10.1109/IVS.2016.7535494>.
- Rahmani, M., Koutsopoulos, H.N., Jenelius, E., 2017. Travel time estimation from sparse floating car data with consistent path inference: A fixed point approach. *Transp. Res. Part C Emerg. Technol.* 85, 628–643. <https://doi.org/10.1016/j.trc.2017.10.012>.
- Ribeiro, M.I., 2004. Kalman and Extended Kalman Filters: Concept, Derivation and Properties, Institute for Systems and Robotics.
- Sasiadek, J.Z., Wang, Q., Zeremba, M.B., 2000. Fuzzy adaptive Kalman filtering for INS/GPS data fusion. In: Proceedings of the 2000 IEEE International Symposium on Intelligent Control. Held Jointly with the 8th IEEE Mediterranean Conference on Control and Automation (Cat. No.00CH37147). IEEE, pp. 181–186 <https://doi.org/10.1109/ISIC.2000.882920>.
- Schindler, A., 2013. Vehicle self-localization with high-precision digital maps. In: 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 141–146 <https://doi.org/10.1109/IVS.2013.6629461>.

- Srinivasan, D., Cheu, R.L., Tan, C.W., 2003. Development of an improved ERP system using GPS and AI techniques. In: *Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems*. IEEE, pp. 554–559 <https://doi.org/10.1109/ITSC.2003.1252014>.
- Syed, S., Cannon, M.E., 2004. Three dimensional fuzzy logic based-map matching algorithm for location based service applications in urban canyons. In: *Proceedings ION GNSS 17th*, pp. 1–12.
- Szotkka, I., 2013. Particle filtering for lane-level map-matching at road bifurcations. In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, pp. 154–159 <https://doi.org/10.1109/ITSC.2013.6728226>.
- Tang, K., Chen, S., Liu, Z., Khattak, A.J., 2018. A tensor-based Bayesian probabilistic model for citywide personalized travel time estimation. *Transp. Res. Part C Emerg. Technol.* 90, 260–280. <https://doi.org/10.1016/j.trc.2018.03.004>.
- Tao, Z., Bonnifait, P., Fremont, V., Ibanez-Guzman, J., 2013. Lane marking aided vehicle localization. In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, pp. 1509–1515 <https://doi.org/10.1109/ITSC.2013.6728444>.
- Thanh-Son, D., Kit Leung, K.Y., Clark, C.M., Huissoon, J.P., 2006. Co-operative lane-level positioning using Markov localization. In: *2006 IEEE Intelligent Transportation Systems Conference*. IEEE, pp. 1006–1011 <https://doi.org/10.1109/ITSC.2006.1707353>.
- Toledo-Moreo, R., Betaille, D., Peyret, F., 2010. Lane-level integrity provision for navigation and map matching with GNSS, dead reckoning, and enhanced maps. *IEEE Trans. Intell. Transp. Syst.* 11, 100–112. <https://doi.org/10.1109/TITS.2009.2031625>.
- Toledo-Moreo, R., Betaille, D., Peyret, F., Laneurit, J., 2009. Fusing GNSS, dead-reckoning, and enhanced maps for road vehicle lane-level navigation. *IEEE J. Sel. Top. Signal Process.* 3, 798–809. <https://doi.org/10.1109/JSTSP.2009.2027803>.
- Velaga, N.R., 2010. Development of a Weight-based Topological Map-Matching Algorithm and an Integrity Method for Location-based ITS Services. Loughborough University, Loughborough.
- Velaga, N.R., Nelson, J.D., Edwards, P., Corsar, D., Sripada, S., Sharma, N., Beecroft, M., 2013. Development of a map-matching algorithm for rural passenger information systems through mobile phones and crowd sourcing. *J. Comput. Civ. Eng.* 27, 732–742. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000238](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000238).
- Velaga, N.R., Pangbourne, K., 2014. Achieving genuinely dynamic road user charging: issues with a GNSS-based approach. *J. Transp. Geogr.* 34, 243–253. <https://doi.org/10.1016/j.jtrangeo.2013.09.013>.
- Velaga, N.R., Quddus, M.A., Bristow, A.L., 2012a. Improving the performance of a topological map-matching algorithm through error detection and correction. *J. Intell. Transp. Syst.* 16, 147–158. <https://doi.org/10.1080/15472450.2012.691852>.
- Velaga, N.R., Quddus, M.A., Bristow, A.L., 2009. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems. *Transp. Res. Part C Emerg. Technol.* 17, 672–683. <https://doi.org/10.1016/j.trc.2009.05.008>.
- Velaga, N.R., Quddus, M.A., Bristow, A.L., Zheng, Y., 2012b. Map-aided integrity monitoring of a land vehicle navigation system. *IEEE Trans. Intell. Transp. Syst.* 13, 848–858. <https://doi.org/10.1109/TITS.2012.2187196>.
- Velaga, N.R., Sathiasel, A., 2011. The role of location based technologies in intelligent transportation systems. *Asian J. Inf. Technol.* 10, 227–233. <https://doi.org/10.3923/ajit.2011.227.233>.
- Vu, A., Ramanandan, A., Chen, A., Farrell, J.A., Barth, M., 2012. Real-time computer vision/dgps-aided inertial navigation system for lane-level vehicle navigation. *IEEE Trans. Intell. Transp. Syst.* 13, 899–913. <https://doi.org/10.1109/TITS.2012.2187641>.
- Yang, D., Cai, B., Yuan, Y., 2003. An improved map-matching algorithm used in vehicle navigation system. In: *Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems*. IEEE, pp. 1246–1250 <https://doi.org/10.1109/ITSC.2003.1252683>.
- Yang, X., Tang, L., Niu, L., Zhang, X., Li, Q., 2018. Generating lane-based intersection maps from crowdsourcing big trace data. *Transp. Res. Part C Emerg. Technol.* 89, 168–187. <https://doi.org/10.1016/j.trc.2018.02.007>.
- Yeh, A.G., Zhong, T., Yue, Y., 2016. Angle difference method for vehicle navigation in multilevel road networks with a three-dimensional transport GIS database. *IEEE Trans. Intell. Transp. Syst.* 1, 1–13. <https://doi.org/10.1109/TITS.2016.2567451>.
- Zheng, Y., Hansen, J.H.L., 2017. Lane-change detection from steering signal using spectral segmentation and learning-based classification. *IEEE Trans. Intell. Veh.* 2, 14–24. <https://doi.org/10.1109/TIV.2017.2708600>.