This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2826000, IEEE Sensors Journal

GOI: A Novel Design for Vehicle Positioning and Trajectory Prediction under Urban Environments

Zhu Xiao, *Member, IEEE*, Pingting Li, Vincent Havyarimana, Hassana Maigary Georges, Dong Wang, *Member, IEEE*, and Keqin Li, *Fellow, IEEE*

***Abstract*—In this paper, we propose a new paradigm of GPS and OBD Integration (GOI) based on GPS receiver and On Board Diagnostics (OBD) reader, which offers a feasible way for large-scale trajectory collection especially suitable for private cars. With GOI, we adopt GPS receiver to obtain vehicle location and design a low-cost OBD reader to retrieve the driving information such as vehicle velocity and steering direction from the in-vehicle motion sensors through the OBD interface. In order to deal with the inherent errors of the motion sensors and the GPS outage issue, we propose a vehicle positioning approach by employing Supporting Vector Machine for Regression (SVR) to achieve accurate and reliable vehicle position and trajectory prediction based on GOI. The low-dimensional non-linear GOI trajectory data is transferred into high-dimensional linear problems by using kernel function so that it reduces the computational complexity and overcomes the problem of dimension disaster. Furthermore, we design a Local Shrinking Particle Swarm Optimization (LS-PSO) algorithm to cope with the parameter selection for SVR based GOI approach. Experiments from real urban environment demonstrate the effectiveness of our approach, which outperforms the existing methods in terms of prediction accuracy under various GPS outages and road conditions.**

***Index Terms*—GPS outage, local shrinking particle swarm optimization, on board diagnostics, SVR, trajectory collection, vehicle positioning.**

I. Introduction

NOWADAYS, the explosive increasing of automobiles has brought unprecedented pressure to a series of problems in modern cities such as municipal transportation, traffic management and environmental protection, etc. In the meantime, large number of vehicles driving in the road networks also generate huge volume of trajectory data, which create new opportunities for solving traffic congestion and improving transport services. Hence, high-quality vehicle trajectory data has important social and application value since it not only offers an effective way to resolve city problems but also is of special significance to understand people's travel activities [1], [2]. In this context, obtaining large-scale vehicle trajectory will facilitate the applications and development in various areas such as Internet of Vehicle (IoV), Intelligent Transportation Systems (ITS) and Smart City [3]–[6].

In particular, it is noteworthy that the private cars, i.e., a class of small motor vehicles usually registered by individual and for personal use, compose the vast majority of city automobiles. For instance, the number of automobiles had reached 194 million by the end of 2016 in China, more than 75 percent of which are private cars [7]. Therefore, there is a severe need to exploit effective method to implement accurate large-scale trajectory acquisition especially for private cars by making use of land vehicle localization technologies.

A. Related Works

In the process of trajectory collection, Global Positioning System (GPS) is the most widely-used method for vehicle positioning since it can provide accurate and reliable vehicle localization and navigation performance in outdoor environment. However, the accuracy degradation of GPS is inevitable due to the blockage of satellite signals and multipath effect in dense urban environments. To resolve this, multi-source data fusion method [8], [9], such as GPS combined Inertial Navigation System (INS), has become one of the most promising technologies since INS is able to amend GPS error under challenging urban environment [10], [11].

In order to deal with the heterogeneous information fusion in GPS/INS integration, Kalman Filter (KF) is the classical method for vehicle positioning and trajectory estimation [12], [13]. However, it is unable to solve non-linear/non-Gaussian problems, which leads to inaccurate modeling and tedious calculation. Particle Filter (PF) is considered as benchmark of filtering method for predicting vehicle position [14]. The drawback of PF method requires a large number of particles and hence leads the algorithm computationally expensive [15], [16]. In recent years, machine learning methods, such as Artificial Neural Network (ANN), Back Propagation Neural Networks (BPNN), Random Forest Regression (RFR), Gauss Process Regression (GPR) [17]–[20], have drawn wide attention due to their abilities to solve the problem of non-linearity and uncertainty. Although these methods perform well in GPS/INS based vehicle positioning, they are unable to meet the needs for real-time and high-precision for vehicular

Manuscript received XX XX, 2018.

This work was supported in part by the National Natural Science Foundation of China (Grants Nos. 61301148, 61502162 and 61702175), the Fundamental Research Funds for the Central Universities of China, Fund of State Key Laboratory of Geoinformation Engineering (No. SKLGIE2016-M-4-2), Hunan Natural Science Foundation of China (No. 2018JJ2059), and Open Fund of State Key Laboratory of of Integrated Services Networks (No. ISN17-14).

(*Corresponding authors: Dong Wang and Keqin Li.*)

The authors are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, China. Zhu Xiao is also with State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China. e-mails: {zhxiao, lipingting, wangd}@hnu.edu.cn, hamjorg@yahoo.fr.

Vincent Havyarimana is also with Department of Applied Sciences, Ecole Normale Suprieure, 6983, Bujumbura, Burundi. e-mail: havincent12@gmail.com.

Keqin Li is also with Department of Computer Science, State University of New York, New Paltz, New York 12561, USA. e-mail: lik@newpaltz.edu.

position prediction due to the inherent sensor errors and bad adaptability in complex GPS-denied urban environment.

In addition to the data fusion method in vehicle positioning, external Inertial Measurement Unit (IMU) is the most widely-used equipment for implementing INS in GPS/INS integration based vehicle positioning and trajectory collection, for instance, smartphones embedded with GPS receiver and IMU sensors can be adopted for vehicle trajectories collection [21], [22]. Nevertheless, this may not be suitable for long-time and large-scale trajectory collection. The main reason is that vehicle trajectory data is incomplete and discontinuous because the smart phone navigation would not be in use throughout the driving. Moreover, most drivers don't need the smartphone navigation since they are familiar with the route for their frequent-visit places in their city driving. The on-board navigation offers an alternative solution for vehicle tracking and could be used for trajectory collection. However, even though vehicle trace can be displayed in console screen, the datasets from on-board navigation are not in an easy-to-use format hence restrict the usefulness of raw trajectory data without intensive processing. Besides, only the high-grade vehicles are deployed with advanced functions such as on-board navigation. Therefore, it is essential to design a usable and friendly method that can be employed for large-scale trajectory acquisition.

B. Contributions

In this work, unlike the widely-used GPS/INS method, we propose a novel GPS and On Board Diagnostics (OBD) integration approach (GOI) to implement vehicle positioning and trajectory collection. Within the proposed GOI, we adopt off-the-shelf GPS receiver to obtain the position of vehicle and design an OBD reader to read the motion information such as vehicle velocity and steering direction from the in-vehicle motion sensors through the OBD interface hence no external INS devices are needed. To the best of our knowledge, the design of so-called GOI provides a feasible way for large-scale trajectory acquisition which is especially suitable for private cars. More importantly, it offers a new paradigm to achieve heterogeneous information fusion for vehicle positioning and trajectory prediction. Notice that OBD reader suffers from the inherent noise of the motion sensors which results in location error accumulated over time. Hence, the performance of vehicle positioning based on GOI trajectory data degrades in case of error fluctuation especially with GPS outage which is likely to take place in urban environments. In order to enhance the accuracy of vehicle position and correct the error in the collected trajectory, we propose a novel GOI based vehicle positioning approach by employing Supporting Vector Machine for Regression (SVR) for solving the vehicle position prediction under GPS outage conditions. Compared to other methods in existing GPS/INS application, the proposed SVR model transforms low-dimensional non-linear problems in GOI based trajectory data into high-dimensional linear space by using kernel function so that reduces the computational complexity and overcomes the problem of dimension disaster. Moreover, it can avoid over-fitting by choosing a specific hyperplane among the feature spaces. Therefore, the proposed SVR based GOI approach can achieve reliable and accurate vehicle position during trajectory collection in complex urban environments.

Besides, it should be recognized that deciding appropriate regression parameters is a rather cumbersome task. To solve this, we design a Local Shrinking Particle Swarm Optimization Algorithm (LS-PSO) to cope with parameter optimization during the process of SVR modoling based on GOI. The proposed LS-PSO significantly improves both the global optimization ability and the convergence speed and in the meantime preserves the simple structure with the traditional PSO. To validate the performance of our method, we implement road experiments in real urban area. The experimental results demonstrate that our method outperforms the existing methods including Back Propagation Neural Network (BPNN) [17], Gauss Process Regression (GPR) [19], Partial Least Squares Regression (PLSR) [23] and the Ensemble Kalman Filter with Delayed Particle Smoother (EKF-DPS) [24], in terms of both vehicle positioning accuracy and root mean square error (RMSE).

The remainder of this paper is organized as follows. We describe our design for GOI device in Section II. In Section III, we present the proposed approach for vehicle position and trajectory prediction based on GOI. The experiment results based on real-world road tests, together with the performance analysis, are given in Section IV. Finally, we conclude the paper in Section V.

II. Design of GOI Device

In this section, we present a new design of GOI which is dedicated to implementing vehicle positioning and trajectory collection. As shown in Figure 1, the device consists of two components, i.e. the GPS module and OBD reader. We deploy the u-Blox GPS receiver in the GPS module so as to obtain vehicle positions including longitude and latitude. We design an OBD reader, which is connected with the vehicle's OBD interface via ISO 14230 (KWP2000), to retrieve the motion information from the vehicle motion sensors such as the velocity, acceleration and steering direction. Notice that due to the increasing demand of driving safety, vehicles are mandatory to install Anti-Lock Braking System (ABS) and Electronic Stability Program (ESP), hence the OBD reader is able to read the driving status through the in-built vehicle motion sensors. Furthermore, a lightweight communication unit with a SIM card (see the back of GPS module in Figure 1), which supports 3G and LTE [25], is embedded with the GPS module so that the vehicle position and driving status can be sent back to the data center.

By installing the GPS and OBD device as shown in Figure 1, we obtain various types of trajectory information including location, vehicle velocity, steering direction, fuel consumption and mileage, etc. The details of the trajectory information are given in Table 1. For example, ObjectID (The ID number of the vehicle), StartTime represents the time when the vehicle starts the engine, StartLon and StartLat give the longitude and latitude when the vehicle starts the engine. StopTime

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2826000, IEEE Sensors Journal

Fig. 1: GOI device

TABLE I: GOI based trajectory information

Field definition	Note
ObjectID	ID number of the vehicle
StartTime	Time of starting the vehicle
StartLon	Longitude of starting the vehicle
StartLat	Latitude of starting the vehicle
StartPos	Position of starting the vehicle
StopTime	Time of turning off the vehicle's engine
StopLon	Longitude of turning off the vehicle's engine
StopLat	Latitude of turning off the vehicle's engine
StopPos	Position of turning off the vehicle's engine
Lon	Longitude of the current position
Lat	Latitude of the current position
Startmileage	Mileage of staring the vehicle
OilNum	Current fuel consumption of the vehicle
Speed	Instantaneous speed of the vehicle
Direct	Current direct of the vehicle's steering
GPSTime	Time of positioning
TravelMileage	Current mileage of the vehicle

represents the moment that vehicle shuts down the engine. StopLon and StopLat denote its longitude and latitude once the vehicle shuts down the engine.

Different with the existing works wherein the external IMU sensors [26] or additional on-vehicle sensors [27] are installed on the vehicle, we design a low-cost OBD reader in the proposed GOI method, which is connected with the vehicles OBD interface, to directly retrieve the motion information from the vehicle-in motion sensors instead of using external IMU devices. In a word, the proposed GOI method is totally affordable and easily to be accepted for the private car users.

During the process of trajectory collection via GOI device, we discover that the data sampling rate of vehicle trajectory is unstable. In addition, the trajectory information may be missing due to the GPS signal interruption caused by the complex urban environment or the problem of OBD reading. To resolve these problems and enhance the reliability of the collected trajectory, we propose a data fusion approach for vehicle positioning and trajectory prediction based on GOI method, the details of which are presented in next section.

III. Vehicle Position Prediction Approach based on GOI

Figure 2 shows vehicle positioning approach based on GOI. In the absence of GPS outages, we input GOI trajectory data to implement training stage of SVR and create a prediction model for vehicle position and trajectory prediction. We design a Local Shrinking Particle Swarm Optimization Algorithm (LS-PSO) to cope with parameter selection problem. When GPS is available, the trajectory data is used to train the SVR model and then update the knowledge that GPS information can remove the accumulation error of OBD readings. After the updating stage, the trained SVR model integrated with vehicle status data to predicts next-time vehicle position when the GPS is unavailable.

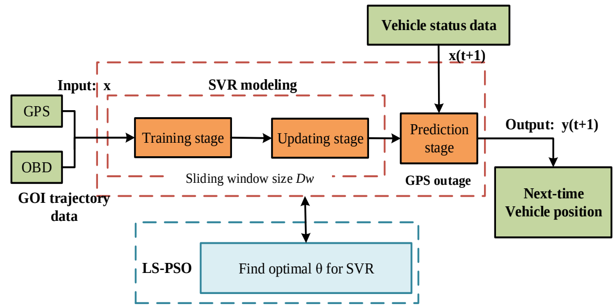

Fig. 2: Vehicle position prediction based on GOI

A. SVR Model for Vehicle Position Prediction

To cope with the inherent non-linearity in the collected GOI trajectory data, we propose a Support Vector Regression (SVR) based vehicle positioning method, with which the low-dimensional nonlinear GPS/OBD data can be mapped into linearly separable high-dimensional space for regression estimation by using kernel function. We assume that the training set $\mathbf{x}$ has n samples, which can be denoted by $\mathbf{x} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_i, y_i), ...(\mathbf{x}_n, y_n)\}$. $\mathbf{x}_i = \{x_i^1, x_i^2, ..., x_i^d\} \in R$ is the input GOI data including velocity, azimuth, acceleration and history latitude/longitude information obtained from GOI device. $y_i \in R$ denotes the spatial displacement between the i-th predicted vehicle position and its $(i-1)$-th position, which can be used to calculate the next-time predicted vehicle position. Our proposed SVR based vehicle positioning approach is a supervised learning method [28], and its goal is to find a function $f(\mathbf{x})$ for which all the training points have a maximum deviation. The linear regression function can be described by:

$$f(\mathbf{x}) = \mathbf{W}^T \cdot \phi(\mathbf{x}) + b, \quad (1)$$

where $\mathbf{W}$ denotes the weight vector and b denotes the intercept. $\phi(\mathbf{x})$ is used to map training set $\mathbf{x}$ to the feature space, then the main task of SVR is to minimize the following loss function :

$$\underset{(\mathbf{W},\xi_i,\xi_i^*)}{\text{Min}} \quad \frac{1}{2} \parallel \mathbf{W}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*), \quad (2)$$

$$s.t. \begin{cases} y_i - \mathbf{W}^T\phi(\mathbf{x}_i) - b \le \varepsilon + \xi_i^*, i = 1, 2, ..., n, \\ -y_i + \mathbf{W}^T\phi(\mathbf{x}_i) + b \le \varepsilon + \xi_i, i = 1, 2, ..., n, \\ \xi_i, \xi_i^* \ge 0, i = 1, 2, ..., n. \end{cases} \quad (3)$$

In (2), $||\mathbf{W}||$ denotes the Euclid-norm of the vector $\mathbf{W}$. C denotes the penalty factor, which is associated with penalty for excess deviation during the training. A larger C means more weights are assigned to training errors and less weights to model complexity. In (3), the insensitive ε plays a decisive role, since it determines the allowed deviation of function $f(\mathbf{x})$ during the training stage. The selection of ε can achieve a compromise between prediction error and generalization ability. It controls the width of the insensitivity region of the regression function to the sample data and affects the number of support vectors. ξ_i and ξ_i^* are slack variables that correspond to the size of this excess deviation for positive and negative deviations, respectively.

Let α_i, α_i^* denote the Lagrange multipliers. Based on the Lagrangian multiplier method, we eliminate $\mathbf{W}$ and b and obtain the duality problem of the optimization problem, which are expressed in (4) and (5).

$$\begin{aligned} \underset{(\alpha_i,\alpha_i^*)}{\text{Max}} \quad & \sum_{i=1}^{n} y_i(\alpha_i-\alpha_i^*) - \varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*) \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(\mathbf{x}_i)^T(\alpha_i-\alpha_i^*)\phi(\mathbf{x}_j)(\alpha_j-\alpha_j^*), \end{aligned} \tag{4}$$

$$s.t. \sum_{i=1}^{n}(\alpha_i-\alpha_i^*) = 0, 0 \le \alpha_i, \alpha_i^* \le C, i = 1, 2, ..., n. \tag{5}$$

Define the kernel function $K(\mathbf{x}_i,\mathbf{x}_j)$, which can be given by:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j), i, j = 1, 2, 3, ..., n. \tag{6}$$

By fulfilling the KKT (Karush-Kuhn-Tucker) conditions, we solve the value of α_i, α_i^* and obtain $\mathbf{W} = \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\phi(\mathbf{x}_i)$ according to (4) and (5). Therefore, we obtain the regression function which can be written as:

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\phi(\mathbf{x}_i)^T\cdot\phi(\mathbf{x}) + b \\ &= \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)K(\mathbf{x},\mathbf{x}_i) + b. \end{aligned} \tag{7}$$

The kernel function $K(\mathbf{x}_i,\mathbf{x}_j)$ can project linearly indivisible data in low-dimensional space, i.e. the GPS and OBD data in our study, into high-dimensional space and hence make data linearly separable. Furthermore, it maps the training samples to an appropriate feature space and reduces the complexity of the SVR model. To address the non-linear vehicle positioning problems, we choose the radial basis function as kernel function, due to its advantages of simple model selection, low computational complexity, high computational efficiency and easy implementation. The kernel function can be expressed by:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp\frac{(-|\mathbf{x}_i-\mathbf{x}_j|)}{2\gamma^2}, i, j = 1, 2, 3, ...n, \tag{8}$$

where γ is the width coefficient of kernel function. The smaller γ will lead the promotion and generalization ability of the model become worse, on the other hand, a larger γ will reduce the learning accuracy and increase the training error.

According to the above analysis, the parameters, i.e. penalty factor C , the kernel coefficient γ and the insensitive factor ε, are crucial variables when modeling vehicle position prediction by employing SVR. In existing studies, these parameters are usually set based on experience or determined by using large scale search. Recently, some parameter optimization methods have been proposed, for instance, Particle Swarm Optimization (PSO) and Bayesian optimization. PSO is one of the most effective evolutionary optimization methods [29], [30], which can achieve fast convergence while is prone to fall into local optimum in the post-search. Bayesian optimization is widely applied to tune the parameters of machine learning algorithms [31], [32]. Although this method has a good global convergence, the convergence rate is too slow and the time to find the optimal solution is particularly long. Therefore, the computational accuracy and the search ability for the parameter optimization are not to be guaranteed.

By taking optimal solution and searching efficiency into consideration, we adopt PSO as the basic optimization method for solving the parameter selection in the vehicle positioning and trajectory prediction. Furthermore, in order to avoid the local optimum of PSO, we wish to conduct global search first so that the search will cover a certain area, then the local detailed search will be implemented to achieve a better solution. To this end, we devise a Local Shrinking Particle Swarm Optimization algorithm (LS-PSO) to balance the role of global search and local search and then obtain the optimal parameters of the SVR based vehicle position prediction.

B. Optimization of $\theta = [C, \gamma, \varepsilon]$ based on LS-PSO

In the standard PSO algorithm, the potential solution to optimization problem is a particle in the search space. Each particle has a corresponding position and velocity, which can be denoted as $T_s = (T_{s1}, T_{s2}, ..., T_{sd}; s = 1, 2, 3, ..., S)$ and $V_s = (v_{s1}, v_{s2}, ..., v_{sd}; s = 1, 2, 3, ..., S)$, respectively. S is the number of particles and $d(d = 1, 2, 3, ..., n)$ is the dimension of the solution. The speed update formula is the most important part of the process of finding the optimal solution, which can be expressed by:

$$\begin{aligned} v_{sd}^{k+1} = \omega_{sd}^k \times v_{sd}^k &+ c_1 \times rand(\cdot) \times (p_{sd}^k - T_{sd}^k) \\ &+ c_2 \times rand(\cdot) \times (pg_d - T_{sd}^k), \end{aligned} \tag{9}$$

where k denotes the current generation, c_1 and c_2 are learning factors and ω_{sd}^k is the inertia weight. $T_s = (T_{s1}, T_{s2}, ..., T_{sd})$ represents the current position of the particle s. $p_s = (p_{s1}, p_{s2}, ..., p_{sd})$ is the best position found by particle s, whereas $pg = (pg_1, pg_2, ..., pg_d)$ is the best position which has been found by the whole population up to now. $\omega_{sd}^k \times v_{sd}^k$ is capable of exploring new areas and expanding the search range, hence the PSO algorithm has excellent performance for global search. The second term in (9) describes the cognitive part, which indicates that the particle has a strong local search ability by studying the best information of their own history. The third term in (9) is the social part, which reflects the collaboration and information shared between the particles.

The balance between exploration and development is determined by ω_{sd}^k in the traditional PSO. In general, the value of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2826000, IEEE Sensors Journal

ω_{sd}^{k} is constant and only affects its own speed to search for a global weight rather than the local search, thus it is easy to fall into the local optimum. To solve this, we devise a local shrink operation for PSO, i.e. LS-PSO method, by which we give ω_{sd}^{k} a dynamic adjustment with the increase of iterations. We use the following formula to adjust the linear dynamic:

$$\omega_{sd}^{k} = \omega_{\max}^{\mathrm{k}} - (k/H) \cdot (\omega_{\max}^{\mathrm{k}} - \omega_{\min}^{\mathrm{k}}), \tag{10}$$

where k represents the number of iterations and H represents the maximum number of iterations. When ω_{sd}^{k} is initialized, we take a larger value to speed up the global search, and then reduce the value of ω_{sd}^{k} to obtain more accurate results. In this study, each particle represents a set of SVR parameters, i.e., $\theta = [C, \gamma, \varepsilon]$, LS-PSO searches for optima by updating generations. Each particle has a fitness value decided by decision function. Based on (11) and (12), LS-PSO updates the particle's velocity and position during evolution.

$$\begin{aligned} v_{sd}^{k+1} =& \omega_{sd}^{k} \times (v_{sd}^{k} + c_1 \times rand(\cdot) \times (p_{sd}^{k} - T_{sd}^{k})) \\ &+ c_2 \times rand(\cdot) \times (pg_d - T_{sd}^{k}), \end{aligned} \tag{11}$$

$$T_{sd}^{k+1} = T_{sd}^{k} + v_{sd}^{k+1}. \tag{12}$$

During the process of particle velocity update, our dynamic weight also affects the particle's own speed and cognitive part so as to make the algorithm achieve the balance between global search ability and local search capability.

C. Proposed Algorithm for Vehicle Positioning and Trajectory Prediction

In the proposed vehicle position prediction approach, the current vehicle position depends on the historical trajectory and vehicle status data. Since the vehicle trajectory is more closely related to its recent historical data, we design a sliding window for collecting trajectory data during the actual process of vehicle position prediction. Let D_W denote the size of sliding window. The window slides on the input trajectory data, as new data arrives, the old data out of each observation window is discarded. The sampling rate of GOI device is set to 1s. When the D_W is 10, the processing time of training and prediction is approximately 1.5s. Therefore, the proposed method is able to provide a real-time vehicle position prediction. For a given observation time, the data in the window can be expressed as:

$$D_W = \{[x_{t-w+1}, ..., x_{t-1}, x_t]^T, [y_{t-w+1}, ..., y_{t-1}, y_t]^T\}. \tag{13}$$

In order to guarantee the accuracy and reliability of the vehicle position prediction, we obtain the input and output model by training the valid GPS data set. In particular, when GPS is available, the trajectory data is used to train the SVR model and then learn the knowledge of calibrating the output of OBD readings. When the GPS signal is unavailable, we utilize the trained model and vehicle motion data that is obtained from OBD to predict $(t+1)$-th output y_{t+1}.

The parameters of the SVR are optimized by LS-PSO. According to the criterion of mean square error (MSE), we define the fitness function as follow:

$$Fit(s) = \frac{1}{n} \sum_{d=1}^{n} (y_s - \hat{y}_d)^2, \tag{14}$$

where $Fit(s)$ denotes the fitness value of s-th particle, y_s denotes the output value of the n samples of the s-th particle, $\hat{y}_d$ is expected output value of n samples. The details of proposed SVR method with LS-PSO are presented in Algorithm 1.

Algorithm 1 Vehicle position prediction via GOI data fusion

Input: GOI trajectory data x.
Output: The predicted position of the next time y_{t+1}.

1: Initialization: Determine the training number of particles S and the maximum number of iterations H, set the initial population of particle pairs $(c_1; c_2)$ with random position T_s and velocity v_s; Set the initial values $\theta = [C, \gamma, \varepsilon]$,the max inertia weigh ω_{max}^{k} and the min inertia weigh ω_{min}^{k}.
2: Training the window based SVR model with the initialized values;
3: Calculate the fitness $Fit(s)$ of the new particle T_s according to (14), where $s = 1; ::: S$;
4: Update the velocity v_{sd}, position T_{sd} and ω_{max}^{k} of particle s , according to (11), (12) and (10) respectively.
5: **if** $Fit(T_s)$ is better than $Fit(p_s)$ **then**
6:     Set T_s to be the p_s
7: **end if**
8: **if** $Fit(T_s)$ is better than $Fit(pg)$ **then**
9:     Set T_s to be the pg
10: **end if**
11: **if** $k = H$, or the optimal solution is not changed in a certain number of iterations **then**
12:     Output the optimal solution i.e. the updated $\theta = [C, \gamma, \varepsilon]$
13: **else**
14:     Let k = k + 1 and return to Step 2
15: **end if**
16: **if** $k < H$ **then**
17:     Let k = k + 1 and return to Step 2
18: **end if**
19: Utilize the updated $\theta = [C, \gamma, \varepsilon]$ and D_W to build window based SVR model;
20: Process data in sliding window according to (13);
21: Predict vehicle position y_{t+1} according to (7).

IV. EXPERIMENTS AND EVALUATION RESULTS

A. Experimental Setup

The performance of the proposed GOI based vehicle positioning is examined with respect to real-world trajectory data, which is obtained from an urban area in Changsha, Capital city of Hunan Province, China. The test trajectory is depicted in Figure 3, which contains typical road conditions such as straight roads, curved road, the upslope and downslope roads, acceleration and deceleration roads. This is a typical

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2826000, IEEE Sensors Journal

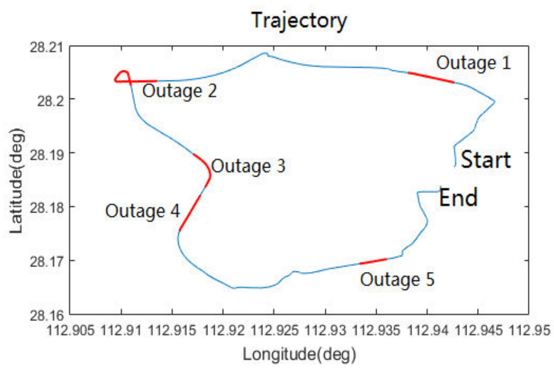

Fig. 3: Field test trajectory.

urban scenario with multipath effect and and GPS outage environment. To simulate driving conditions under GPS signal interruption, we consider five GPS outages in various road conditions including straight section (driving with direction and speed), viaduct area (continues turning), curved road (decelerations and accelerations), high speed portion and low speed portion (low average speed and stop/go driving). Then we conduct the comparative studies based on the vehicle position prediction performance of the proposed method and four existing methods including BPNN [17], GPR [19], PLSR [23] and EKF-DPS [24]. In the proposed GOI method, the u-Blox GPS receiver is adopted in the GPS module. To compare the proposed GOI method with the widely-used integration of GPS and INS, we collect the raw trajectory data by using GPS receiver and external IMU device. To specify, The vehicle motion information and vehicle position (longitude/latitude) are obtained from the external IMU sensor MPU 6500 and the u-Blox GPS receiver, respectively. We then implement the vehicle position and trajectory prediction by using four comparative methods.

Fig. 4: Performance during GPS outage 1: the straight road.

B. Experimental Results Analysis

Figure 4 illustrates the trajectories from straight section with 30s GPS outage. It can be seen that the proposed method generates a well-matched trajectory with the reference line (see the blue line in Figure 4). In the straight section, the direction and speed remain relatively stable, hence the position prediction results of methods such as PLSR and GPR are relatively accurate. EKF-DPS and BPNN have a large degree of deviation when comparing with other methods, which is mainly because of their weak ability in reducing the inherent error from the motion sensors.

Figure 5 presents the results from outage 2, in which the test vehicle drives through the viaduct area with turns and slopes. In the real area, this is a section that the GPS signal reception can be affected by signal blockage and multipath effect. In addition, the vehicle passes this area with frequent changes in the direction and speed. We choose a relatively long period of GPS outage (almost 70s) to illustrate the advantages of our method. According to the results from Figure 5, the proposed method outperforms the others in terms of position prediction performance. Large position errors can be seen in the trajectories from PLSR, BPNN and EKF-DPS, the reason is that large GPS outage leads to gradual accumulation of the prediction error and thus decreases the position prediction accuracy, while those methods fail to provide sufficient error correction based on OBD due to their weak generalization ability when dealing with the long-term error accumulation. The proposed method and GPR can generate good position prediction and pose less drift in comparison to others. Moreover, the proposed method achieves better performance than GPR when the vehicle drives with sharp turn especially in the ramp section.

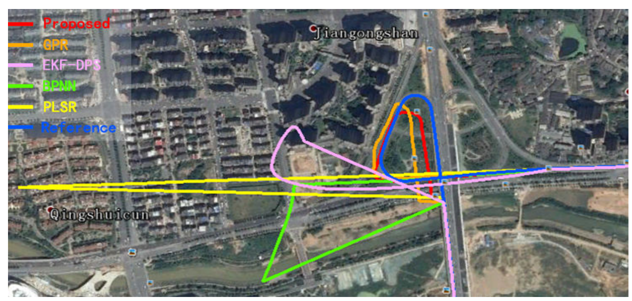

Fig. 5: Performance during GPS outage 2: the viaduct portion.

Figure 6 describes predicted trajectories in the curve portion with 43s GPS outage. In this long turning driving, the vehicle speed and direction are easy to change and the prediction modeling is more difficult. PLSR, BPNN and EKF-DPS create large positioning errors because they are unable to provide error compensation during the GPS outage. Our proposed method shows superior performance compared with others, which not only obtains the best improvement on positioning accuracy but also generates a smooth curve that fits the reference trajectory quite well.

Figure 7 and Figure 8 illustrate the predicted trajectories in high speed portion and low speed portion, where the GPS outages are 43s and 40s, respectively. In Figure 7, the vehicle is with high and stable speed, i.e. 70km/h, and the driving status is steady, hence all methods can obtain well-matched trajectories since the intrinsic error of the vehicle position system is small and the position prediction is accurate. In Figure 8, the vehicle drives at a slow speed and with lots of accelerations and brakes, the comparative methods have large deviations due to the accumulation of errors. In overall, the proposed method can achieve best positioning performance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2826000, IEEE Sensors Journal

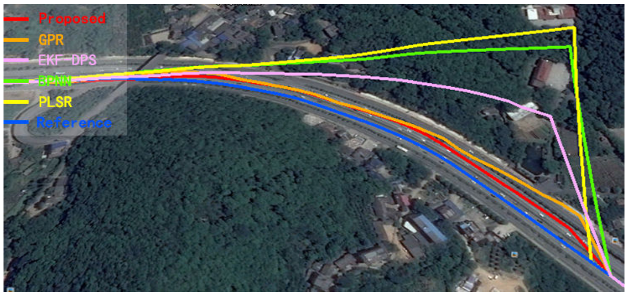

Fig. 6: Performance during GPS outage 3: the curve portion.

Fig. 7: Performance during GPS outage 4: the high speed portion.

and provide nearly perfect trajectories based on the results from Figure 7 and Figure 8.

Fig. 8: Performance during GPS outage 5: the low speed portion.

To further evaluate the position prediction performance, we use Root Mean Square Error (RMSE) to quantify the position prediction performance, which is given in (15):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M} \left[\hat{D}_i - D_i\right]^2}{M}}, \tag{15}$$

where $\hat{D}_i$ is the predicted data, D_i is the reference data. $\hat{D}_i - D_i$ represents the distance between the reference trajectory and the predicted trajectory, and M is the outage duration.

The RMSE of various position prediction methods at different outages are shown in Table 2. The proposed method achieves a smaller RMSE value on position prediction when comparing with GPR, BPNN, EKF-DPS and PLSR. Combining with Figures 4-8, the predicted RMSE of our method is the smallest and the predicted trajectories are closer to the reference trajectory. For instance, in the straight section, the trajectory predicted by the proposed method achieves RMSE

TABLE II: Trajectory prediction errors (RMSE)

Outage	**Proposed**	GPR	EKF-DPS	BPNN	PLSR
Outage 1(37s)	**3.158**	8.537	23.570	32.244	6.238
Outage 2(69s)	**12.633**	28.851	173.254	197.653	264.124
Outage 3(49s)	**8.704**	13.201	46.264	71.592	65.385
Outage 4(43s)	**2.256**	3.074	13.310	5.383	6.373
Outage 5(33s)	**3.496**	14.990	6.214	7.788	5.684

of 3.1585m, while the RMSE of BPNN is reached to 32m. This can also be seen from Figure 4, the trajectory predicted by BPNN has larger deviation. The percentage improvement in the positioning accuracy of the proposed method is 90.2% against BPNN. In the complex viaduct portion and curve portion, in which long interruption with more cumulative errors turn out, the proposed method also obtains better performance than BPNN, GPR, PLSR and EKF-DPS. In particular, the percentage improvement in the positioning accuracy can be found to improve by 48.37% ,92.7% ,93.9% and 95.3% against GPR, EKF-DPS, BPNN and PLSR when the vehicle passes the viaduct portion.

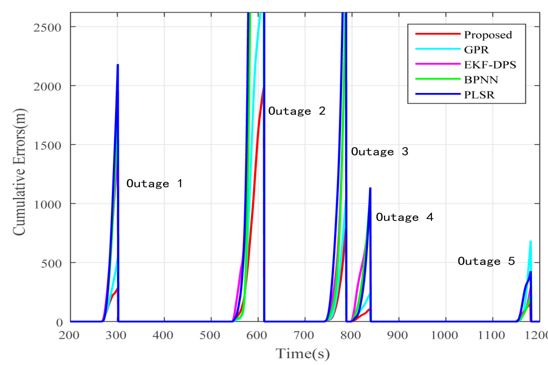

Fig. 9: Cumulative errors.

Figure 9 presents the cumulative errors of various methods under five outages. It is shown that the proposed method outperforms the comparative methods. In particular, in the outage 1, outage 4 and outage 5, the cumulative errors of the proposed method (the red line) are less than 500m, which is obviously lower than the results in other methods. In outage 2 and outage 3, the cumulative errors of the comparative methods are far larger than 2500 m, and the proposed method can keep the cumulative errors below 2000m and 1000m in these two outages, respectively. The results in Figure 9 are also supported from Figures 4-8 and the results of RMSE. In conclusion, during both short and long GPS outages, the experimental results reveal that our proposed method shows great superiority to other adopted prediction methods in all considered GPS outages.

V. CONCLUSIONS

In this paper, we present a new design of GPS/OBD integration approach (GOI) that can be used to implement vehicle positioning and trajectory collection. The so-called GOI method offers a user-friendly solution for large-scale trajectory acquisition which is particularly suitable for private cars. To guarantee the vehicle positioning accuracy during trajectory collection, we propose a SVR based method for modeling vehicle trajectory data collected from GPS and OBD device in order to obtain optimal predicted vehicle position. Moreover, we design a Local Shrinking Particle Swarm Optimization algorithm (LS-PSO) to optimize the parameters of SVR model in terms of both convergence rate and accuracy. Experiments from real urban environments demonstrate the feasibility and effectiveness of the GOI based vehicle positioning and trajectory prediction approach. The results reveal that the proposed approach achieves significant improvement in the positioning accuracy under different GPS outages and road conditions when comparing with existing methods. In addition, the proposed GOI has been successfully applied to real world trajectory collection and we have obtained a large database of private car trajectory data. As presented in our previous work [33], we investigate the travel regularity of private cars based on the long-term trajectory data collected from large-scale private cars via GOI approach.

References

[1] F. Zhang, B. Jin, Z. Wang, H. Liu, J. Hu, and L. Zhang, "On geocasting over urban bus-based networks by mining trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1734–1747, 2016.

[2] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, "Mining periodic behaviors for moving objects," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2010, pp. 1099–1108.

[3] Y. Zheng, "Trajectory data mining: An overview," *Acm Transactions on Intelligent Systems and Technology*, vol. 6, no. 3, pp. 1–41, 2015.

[4] F. Zhenni and Z. Yanmin, "A survey on trajectory data mining: Techniques and applications," *IEEE Access*, vol. 4, pp. 2056–2067, 2017.

[5] R. Sundar, S. Hebbar, and V. Golla, "Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1109–1113, 2014.

[6] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of things using big data analytics," *Computer Networks*, vol. 101, no. C, pp. 63–80, 2016.

[7] C. N. S. Bureau, *China Statistical Yearbook*. Beijing: China Statistical Publishing House, 2011-2016.

[8] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[9] H. M. Georges, X. Zhu, and W. Dong, "Hybrid cooperative vehicle positioning using distributed randomized sigma point belief propagation on Non-Gaussian noise distribution," *IEEE Sensors Journal*, vol. 16, no. 21, pp. 7803–7813, 2016.

[10] X. Li and Q. Xu, "A reliable fusion positioning strategy for land vehicles in gps-denied environments based on low-cost sensors," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3205–3215, 2017.

[11] V. Havyarimana, D. Hanyurwimfura, P. Nsengiyumva, and X. Zhu, "A novel hybrid approach based-SRG model for vehicle position prediction in multi-GPS outage conditions," *Information Fusion*, vol. 41, pp. 1–8, 2018.

[12] C. Barrios and Y. Motai, "Improving estimation of vehicle's trajectory using the latest Global Positioning System with Kalman Filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 12, pp. 3747–3755, 2011.

[13] T. B. Karamat, M. M. Atia, and A. Noureldin, "Performance analysis of code-phase-based relative GPS positioning and its integration with land vehicles motion sensors," *IEEE Sensors Journal*, vol. 14, no. 9, pp. 3084–3100, 2014.

[14] D. Bhatt, P. Aggarwal, V. Devabhaktuni, and P. Bhattacharya, "A novel hybrid fusion algorithm to bridge the period of GPS outages using low-cost INS," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2166–2173, 2014.

[15] P. Aggarwal, Z. Syed, and N. El-Sheimy, "Hybrid Extended Particle Filter (HEPF) for integrated civilian navigation system," in *Position, Location and Navigation Symposium, IEEE/ION*, Monterey, CA, USA, 2008, pp. 984–992.

[16] H. M. Georges, D. Wang, Z. Xiao, and J. Chen, "Hybrid global navigation satellite systems, differential navigation satellite systems and time of arrival cooperative positioning based on iterative finite difference Particle Filter," *IET Communications*, vol. 9, no. 14, pp. 1699–1709, 2015.

[17] N. El-Sheimy, K. W. Chiang, and A. Noureldin, "The utilization of Artificial Neural Networks for multisensor system integration in navigation and positioning instruments," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 5, pp. 1606–1615, 2006.

[18] S. Adusumilli, D. Bhatt, H. Wang, P. Bhattacharya, and V. Devabhaktuni, "A low-cost INS/GPS integration methodology based on Random Forest Regression," *Expert Systems with Applications*, vol. 40, no. 11, pp. 4653–4659, 2013.

[19] Z. Xiao, S. Zhan, Z. Xiang, D. Wang, and W. Chen, "A GPR-PSO incremental regression framework on GPS/INS integration for vehicle localization under urban environment," in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications*, Valencia, Spain, 2016, pp. 1–6.

[20] S. Choi, E. Kim, and S. Oh, "Real-time navigation in crowded dynamic environments using Gaussian process motion control," in *IEEE International Conference on Robotics and Automation*, Hong Kong, China, 2014, pp. 3221–3226.

[21] F. Calabrese, D. Mi, G. D. Lorenzo, J. F. Jr, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example," *Tramsportation Research Part C Emerging Technologies*, vol. 26, no. 1, pp. 301–313, 2013.

[22] R. Hostettler, W. Birk, and M. L. Nordenvaad, "Joint vehicle trajectory and model parameter estimation using road side sensors," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 5075–5086, 2005.

[23] S. Wold, H. Martens, and H. Wold, *The multivariate calibration problem in chemistry solved by the PLS method*. Springer Berlin Heidelberg, 1983.

[24] Z. Xiao, V. Havyarimana, T. Li, and D. Wang, "A nonlinear framework of Delayed Particle Smoothing method for vehicle localization under Non-Gaussian environment," *Sensors*, vol. 16, no. 5, p. 692, 2016.

[25] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall, "LTE: The evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[26] E. Choi and S. Chang, "A consumer tracking estimator for vehicles in gps-free environments," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 450–458, 2018.

[27] C. Melendez-Pastor, R. Ruiz-Gonzalez, and J. Gomez-Gil, "A data fusion system of gnss data and on-vehicle sensors data for improving car positioning precision in urban environments," *Expert Systems with Applications*, vol. 80, pp. 28–38, 2017.

[28] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support Vector Regression Machines," *Advances in Neural Information Processing Systems*, vol. 28, no. 7, pp. 779–784, 1997.

[29] Y. Bazi and F. Melgani, "Semisupervised PSO-SVM regression for biophysical parameter estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1887–1895, 2007.

[30] S. Chen, X. Hong, and C. J. Harris, "Particle Swarm Optimization aided orthogonal forward regression for unified datamodeling," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 4, pp. 477–499, 2010.

[31] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, "Hyperparameter tuning for big data using Bayesian optimisation," in *International Conference on Pattern Recognition*, Cancun, Mexico, 2016, pp. 2574–2579.

[32] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of Machine Learning algorithms," *Computer Science*, vol. 4, pp. 2951–2959, 2012.

[33] D. Wang, Q. Liu, Z. Xiao, J. Chen, H. Yourong, and W. Chen, "Understanding travel nehavior of private cars via trajectory big data analysis in urban environments," in *IEEE DataCom*, Orlando, USA, Nov 2017.