Improving GPS Code Phase Positioning Accuracy in Urban Environments Using Machine Learning

Rui Sun, *Member, IEEE*, Guanyu Wang , Qi Cheng, Linxia Fu,
Kai-Wei Chiang, Li-Ta Hsu, *Member, IEEE*, and Washington Yotto Ochieng

***Abstract*—The accuracy of location information, mainly provided by the global positioning system (GPS) sensor, is critical for Internet-of-Things applications in smart cities. However, built environments attenuate GPS signals by reflecting or blocking them resulting in some cases multipath and non-line-of-sight (NLOS) reception. These effects cause range errors that degrade GPS positioning accuracy. Enhancements in the design of antennae and receivers deliver a level of reduction of multipath. However, NLOS signal reception and residual effects of multipath are still to be mitigated sufficiently for improvements in range errors and positioning accuracy. Recent machine learning-based methods have shown promise in improving pseudorange-based position solutions by considering multiple variables from raw GPS measurements. However, positioning accuracy is limited by low accuracy signal reception classification. Unlike the existing methods, which use machine learning to directly predict the signal reception classification, we use a gradient boosting decision tree (GBDT)-based method to predict the pseudorange errors by considering the signal strength, satellite elevation angle and pseudorange residuals. With the predicted pseudorange errors, two variations of the algorithm are proposed to improve positioning accuracy. The first corrects pseudorange errors and the other either corrects or excludes the signals determined to contain the effects of multipath and NLOS signals. The results for a challenging urban environment characterized by high-rise buildings on one side, show that the 3-D positioning accuracy of the pseudorange error correction-based positioning measured in terms of the root mean square error is 23.3 m, an improvement of more than 70% over the conventional methods.**

***Index Terms*—Global positioning system (GPS), gradient boosting decision tree (GBDT), multipath, non-line-of-sight (NLOS), urban positioning.**

This work was sup-ported in part by the sponsorship of the National Natural Science Foundation of China under Grant 41974033 and Grant 41704022; and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170780.*(Corresponding author: Rui Sun.)*

Rui Sun is with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, and also with the State Key Laboratory of Geo-Information Engineering, Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China (e-mail: rui.sun@nuaa.edu.cn).

Guanyu Wang, Qi Cheng, and Linxia Fu are with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: guanyu.wang@nuaa.edu.cn; qicheng@nuaa.edu.cn; flx0125@nuaa.edu.cn).

Kai-Wei Chiang is with the Department of Geomatics Engineering, National Cheng-Kung University, Tainan 701, Taiwan (e-mail: kwchi-ang@mail.ncku.edu.tw).

Li-Ta Hsu is with the Interdisciplinary Division of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, Hong Kong, China (e-mail: lt.hsu@polyu.edu.hk).

Washington Yotto Ochieng is with the Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: w.ochieng@imperial.ac.uk).

I. Introduction

GLOBAL navigation satellite systems (GNSSs) such as global positioning system (GPS), are used widely for positioning, navigation and timing (PNT). GPS is a key technology that provides positioning information and therefore, critical for Internet-of-Things (IoT) applications in urban environments. However, in such environments satellite signals are prone to reflection, diffraction and blockage by proximate obstacles resulting in multipath and non-line-of-sight (NLOS) reception. Multipath occurs when both the direct (referred to as line-of-sight or LOS) and reflected signals are present in the receiver. NLOS on the other hand occurs when only the reflected signal is present at the receiver. The effects of Multipath and NLOS can result in pseudorange errors tens of metres in magnitude, inhibiting the use of GPS for some location-based services in urban areas. To date these effects are largely mitigated through design (receiver and antenna), signal processing and modeling in the measurement domain.

Antenna design methods, such as the choke-ring antenna and the dual-polarisation one, are effective in mitigating multipath effects at low elevation angles [1], [2]. However, they are bulky and expensive inhibiting their widespread use. Van Dierendonck [3] proposed the signal processing-based method employing the correlation technique by reducing the space between early and late receiver code correlators [3]. This was followed by approaches including the multipath estimating delay lock loop (MEDLL), and vision and strobe correlators [4]–[9]. These signal processing-based methods are able to mitigate the multipath effect of middle/long range multipath signal. However, they are invalid for short range multipath signals. Besides, there is no single method for the mitigation of both NLOS and multipath effects.

The methods employed in the measurement domain employ measurements and derivatives, and satellite and signal information to ameliorate the effects of NLOS reception and multipath. Utilizing dual-frequency observations, code-minus-carrier measurements (CMCs) can be obtained to estimate code multipath error and evaluate multipath conditions when carrier-phase and code-phase measurements

The following publication R. Sun et al., "Improving GPS Code Phase Positioning Accuracy in Urban Environments Using Machine Learning," in IEEE Internet of Things Journal, vol. 8, no. 8, pp. 7065-7078, 15 April15, 2021 is available at https://dx.doi.org/10.1109/JIOT.2020.3037074.

are available [10], [11]. The use GPS measurements together with data from additional sensors, such as inertial measurement units (IMUs) and vision sensors, has been shown to enhance the accuracy of positioning in urban canyons [12]–[17]. Although GNSS/IMU integration is effective with high grade IMU sensors, their high cost precludes their widespread use. The integration with visual sensors is also affected by high cost and weather conditions. Apart from sensor integration, recent research has also looked at the use of spatial information (e.g., 3-D city models) to mitigate the effects of multipath and NLOS. Groves developed the shadow matching technique, in which the 3-D city model is used to assist the detection of NLOS reception and then to improve positioning accuracy [18]. Shadow matching is effective in reducing the cross-street positioning error by 77.5%, i.e., from 14.8 m with the conventional solution to just 3.3 m, but is not able to improve positioning accuracy in along street direction [19]. Furthermore, an optimised user position can be calculated based on the weighted average of the estimated candidate positions obtained from a comparison of the simulated and measured pseudoranges based on the 3-D city model [20]–[22]. However, the accuracy of positioning aided by 3-D city model depends on the quality and integrity of model [23]. Although these methods can improve positioning accuracy by reducing the errors caused from NLOS or multipath to a certain extent, as shown there are open issues still to be resolved.

The quality of a signal received depends to a large extent on the physical environment proximate to the antenna. The impact or effect of such an environment on signal reception can occur in one of three ways: 1) LOS; 2) NLOS; and 3) multipath, with the last two traditionally known to have a discernible effect on the carrier to noise ratio (C/N_0) of the signal compared to that of LOS. Hence, the approach taken is to specify a threshold for C/N_0. (i.e., a higher C/N_0 for LOS signals and lower for NLOS signals). Although it has been demonstrated that C/N_0 is an effective classifier in the absence of interference, in practice there are circumstances of detection of LOS signals at low C/N_0 and NLOS signals at high C/N_0. Therefore, single variable (e.g., C/N_0) methods tend to be suboptimal especially in complex physical environments [24]. This calls for the consideration of other variables.

In addition to C/N_0, other variables such as satellite elevation angle, pseudorange residuals and the other derivatives could also be used to classify pseudorange measurements in terms of quality. Deng [25] used satellite elevation angle and geometric dilution of precision (GDOP) to select the satellites to use. Wang *et al.* [26] identified LOS signals using C/N_0 and satellite elevation. Hsu *et al.* [27] showed that the pseudorange residual could be used as an indicator to exclude multipath and NLOS signals if the number of measurements is sufficient. Horizontal and vertical dilutions of precision (HDOP and VDOP, respectively), and pseudorange change rate [24], [28] are the other possible variables.

Traditionally, linear and nonlinear regression are employed to determine the model parameters for observation equations linking the pseudorange error to the multiple variables [29], [30]. Emerging recently is the application of machine learning for modelling and mitigating pseduorange errors. Research has shown that it has a great potential to improve GNSS positioning accuracy by considering various complementary variables that influence measurement error. Phan *et al.* use two variables (elevation and azimuth angles) with a support vector machine (SVM) machine learning model to mitigate the effect of multipath in static applications [31]. By using the labeling results from a camera and compass, Yozevitch *et al.* [24] proposed a decision tree-based approach to classify the LOS and NLOS with the C/N_0, elevation angle and pseudorange variables as the algorithm inputs. Hsu [28] used C/N_0, rate of change of C/N_0, pseudorange residuals and pseudorange consistency with an SVM to classify the received signal types. Labeling of the type of signal reception is through the use of a 3-D city model together with ray-tracing. Socharoentum *et al.* used four machine learning methods (logistic regression, SVM, Naïve Bayes, and decision tree) to detect NLOS signals. They employed a number of inputs, including satellite visibility, position dilution of precision (PDOP), pseudorange corrections and other related variables [32]. The accuracy of the predicted signal reception types is compared in the simulation. Quan *et al.* proposed a convolutional neural network (CNN) method to detect multipath, applying the sparse autoencoder (SAE) for feature extraction. An improvement in positioning accuracy ranging from 18% to 30% was achieved by down-weighting the measurement identified as affected by multipath [33]. Guermaha *et al.* [34] proposed a decision tree-based GNSS signal classifier using satellite elevation and the difference in C/N_0 from a right-hand circular polarized (RHCP) and left-hand circular polarized (LHCP) antennae. Sun *et al.* [35] used nine variables together with principle component analysis (PCA) and artificial neuro fuzzy inference system (ANFIS) to classify LOS, multipath, and NLOS signals. The nine variables are HDOP, PDOP, azimuth angle, satellite elevation angle, number of visible satellites, C/N_0, changes in C/N_0 over time, pseudorange residual, and consistency between delta pseudorange and pseudorange rate [35]. The subsequent research builds on this in an attempt to reduce the number of variables, ultimately applying just the three most representative variables, satellite elevation angle, C/N_0 and pseudorange residuals together with a gradient boosting decision tree (GBDT) model to predict the received signal's classification type before computing the resulting position solutions by excluding the NLOS signals. Although an improvement in positioning accuracy is observed in places, the effect of exclusion on geometry, overall, results in only a marginal improvement in accuracy [36].

It is notable that all of the current methods using machine learning are for the classification of the received signals [24], which is a significant factor in the accuracy of positioning using the signals [31]–[36]. However, the classification accuracy is affected by the errors introduced from the other information sources used in the offline labeling phase of the machine learning algorithms (e.g., 3-D city model, camera or compass etc.) [24], [28], [35]. For example, the accuracy of labeling with a camera or compass often depends on the cost (and thus quality) of the hardware. Using the 3-D city model in the offline labeling phase also has some weaknesses, including:

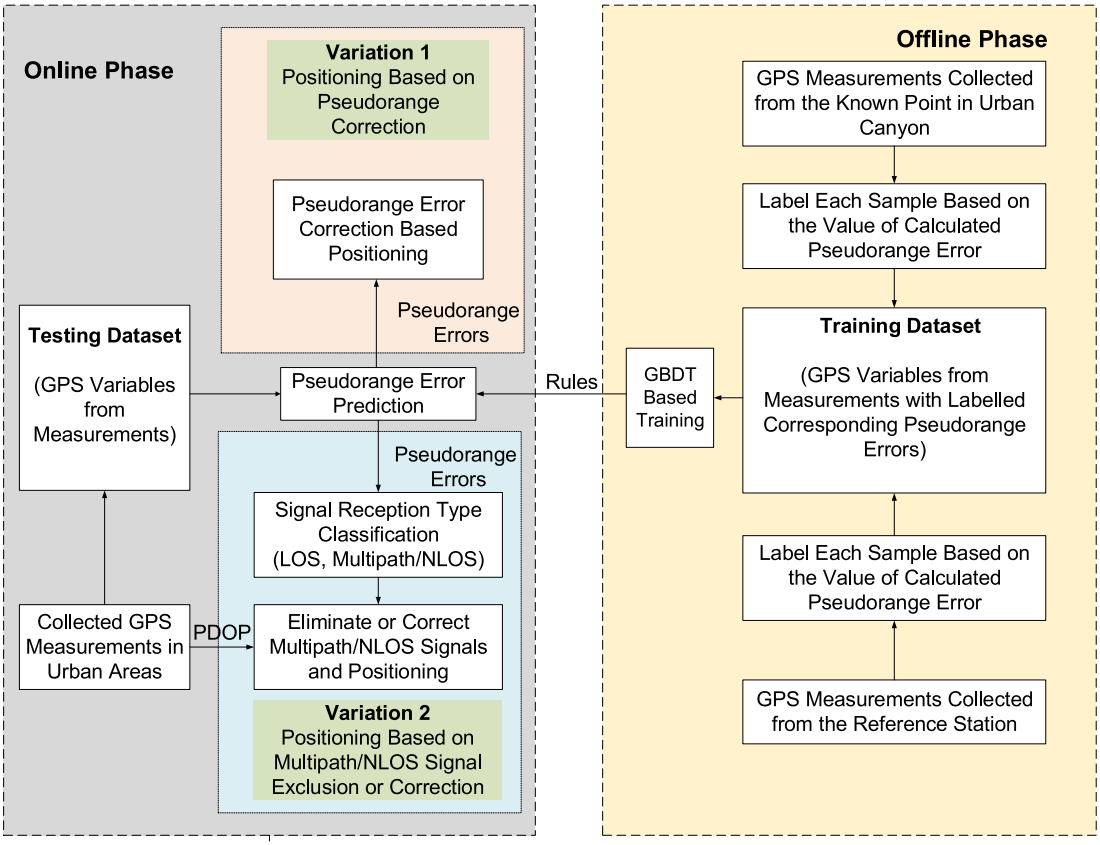

Fig. 1. Framework of the GBDT-based GPS code phase positioning algorithm.

1) the need for significant computational resources for storage, updating and offline processing and 2) the difficulty of generating accurate building borders for some special building shapes. The error introduced by using 3-D city models, cameras, compasses or other sensors in the labeling process for signal reception classification inevitably results in final positioning errors. The key issue for positioning accuracy however, is whether the pseudorange is correctly measured. If we could design a robust algorithm to obtain the pseudorange error measured from each observed satellite, it would potentially be possible to achieve a high accuracy positioning solution based on error corrections.

Building on the above, we have developed a novel GPS pseudorange algorithm to improve positioning accuracy in urban environments. It employs satellite elevation angle, signal strength and pseudorange residuals as the algorithm's inputs. Unlike the traditional methods, we use GBDT to predict the pseudorange errors. Based on the predicted pseudorange errors, two variations of the algorithm for the positioning are proposed. The first variation is based on the pseudorange error correction while the second is based on the multipath/NLOS exclusion or correction. The proposed positioning method could avoid the errors and costs arising from additional hardware or geospatial information during the labeling phase, and therefore, addresses the limitations of current signal reception classification-based positioning approaches using machine learning. The contributions of our research are summarized as follows.

1) Development of a new GBDT-based pseudorange error prediction model using satellite elevation angle, C/N_0 and pseudorange residuals as the input variables. This model is in turn used to improve positioning accuracy by applying the predicted corrections to pseudorange measurements before computing the position solutions.
2) Development of a variation to the model above, by using the predicted pseudorange errors and a threshold to detect and exclude (where the geometric configuration allows) multipath/NLOS affected signals, before computing the position solution with a higher accuracy.

Based on static field tests in urban areas we demonstrate that the proposed approaches are superior to the conventional code phase-based positioning results. Using the scenario of challenging urban areas with high rise buildings on one side as an example, for the first model (application of the predicted pseudorange error as a correction), the 3-D positioning accuracy (RMSE) is enhanced to 23.3 m from about 80 m, an improvement of more than 70%. For the second variation (detection and exclusion), the 3-D positioning accuracy (RMSE) is improved to 60.8 m from about 80 m, an improvement of about 25%.

II. Algorithm Framework

A. Algorithm Framework

The framework of the proposed GBDT-based GPS code phase positioning algorithm is presented in Fig. 1. The

proposed GBDT-based machine learning algorithm includes a training phase conducted offline and a testing phase conducted online.

In the offline phase, the data used are GPS raw pseudoranges collected from a known point in an urban canyon and a reference (known) station. Some of the data from the known location in the urban canyon contain multipath and NLOS effects resulting in relatively large pseudorange errors. These errors are computed from the difference between the raw pseudoranges and the corresponding geometric ranges from the known station coordinates and satellite ephemeris. The raw data from the reference station are not affected by multipath and NLOS, resulting in relatively low pseudorange errors, as determined from the difference between the raw pseudoranges and the corresponding geometric ranges.

Every set of variables at each epoch (satellite elevation angle, C/N_0 and pseudorange residual) is then mapped to, or labeled with, the corresponding pseudorange error. The GBDT algorithm is then used to fit the calculated pseudorange error by means of an offline data set training process, thereby obtaining the rules, that is, the relationship between the input variables (elevation angle, C/N_0, pseudorange residuals) and the corresponding labeled pseudorange errors. The main parts of the offline training process, including variable selection, details on the labeling process and GBDT-based training process are discussed further in the subsequent sections.

In the online phase, new GPS variables from raw measurements in urban canyons, including elevation angle, C/N_0 and pseudorange, are used together with the rules extracted from the offline phase to predict the pseudorange errors. Based on the predicted pseudorange errors, two variations of positioning algorithm are proposed: 1) positioning solutions based on the application of the predicted pseudorange errors as corrections to the new raw pseudoranges and 2) the predicted pseudorange errors are compared with a threshold value to obtain the signal reception type classification, and a geometry allowing position solution based on exclusion or correction of the multipath/NLOS signals in the positioning solutions.

B. Determination of Input Variables

The received GPS signal contains a variety of variables that can be used to determine the pseudorange error. Although, using more variables should improve the accuracy of the training, this is at the cost of computational efficiency. Based on computational cost and training accuracy from previous research by the authors [35], [36], we adopt the three representative variables, satellite elevation θ, C/N_0 and pseudorange residuals η for pseudorange error prediction.

1) *Carrier to Noise Ratio, C/N_0:* The ratio of carrier power to noise power per unit of bandwidth in decibel-hertz (dB-Hz). Under the same noise power, the C/N_0 of a signal with large pseudorange errors (i.e., multipath/NLOS) is lower than that with small pseudorange errors (i.e., LOS). C/N_0 is the most commonly used indicator of pseudorage errors.
2) *Pseudorange Residuals, η:* The inconsistency between the pseudorange measurements and the solution. It is determined from the following equation:

$$\eta = G\Delta x - b \tag{1}$$

where G is the design matrix. b is the difference between observed and computed pseudoranges. Δx is the corrections applied to the approximate values of unknown position and time to determine the final position and time. The solution for Δx for the unweighted measurements in the least-squares process is expressed

$$\Delta x = (G^T G)^{-1} G^T b. \tag{2}$$

Research has shown that given sufficient measurements, the variable pseudorange residual could be used to detect the effects of multipath/NLOS [36]. The pseudorange residual is therefore, a feasible variable for the pseudorange error prediction.

3) *Satellite Elevation, θ:* The higher the satellite elevation angle, the lower the probability of a satellite signal being reflected or blocked by the physical environment proximate to the receiver antenna, resulting in low multipath and NLOS effects

$$\theta^{(i)} = -\arcsin\left(u_D^{(i)}\right). \tag{3}$$

Weighting the measurements based on the elevation angle to reduce the multipath effect is widely used in positioning computation [37]. Therefore, satellite elevation angle is also used for the pseudorange error prediction.

C. Labeling Process

The labeling of the pseudorange error is critical in the offline training process. The ranging errors result from the fact that the contaminated signal (i.e., multipath or NLOS) travels an additional route due to being reflected in the surrounding environment. Typically, in urban canyons, the error can be a few tens of metres. However, larger errors can result if a signal is reflected by a remote tall building. Once the ground truth is known, the pseudorange errors of the received signals can be calculated and labeled with their values. The observed pseudorange ρ between the receiver and a satellite can be expressed in (4) and (5) as follows:

$$\rho = R + c\left(\delta t^r - \delta t^{sv}\right) + I + T + \varepsilon \tag{4}$$

$$R = \sqrt{(x^{sv} - x^r)^2 + (y^{sv} - y^r)^2 + (z^{sv} - z^r)^2} \tag{5}$$

where R is the geometric range between the observed satellite and the receiver; (x^{sv}, y^{sv}, z^{sv}) and (x^r, y^r, z^r) are the coordinates of the satellite and receiver in an earth centred earth fixed (ECEF) system; c is the velocity of light in vacuo; δt^{sv} is the satellite clock offset; δt^r is the receiver clock offset; I is the ionospheric delay; T is the tropospheric delay; ε the errors due to the effects of multipath/NLOS, receiver noise, and antenna delay. The error caused of by the effects of multipath/NLOS dominate ε, since the other errors are relatively small and negligible.

Considering the error sources of the observed pseudorange ρ, and the actual (known) receiver position, the corrected pseudorange ρ^c, with the related error model applied, could be

expressed as

$$\rho^c = R + c\left(\Delta\delta t^r - \Delta\delta t^{sv}\right) + \Delta I + \Delta T + \varepsilon \tag{6}$$

where the geometric range R can be calculated based on the known position of the receiver and the position of the observed satellite from the broadcast ephemeris; c is the velocity of light in a vacuum; $\Delta\delta t^r$ is the residual of the receiver clock offset after applying the calculated receiver clock error from the pseudorange positioning equations with the known ground truth; $\Delta\delta t^{sv}$ is the residual of the satellite clock offset after applying the satellite clock offset obtained from the broadcast ephemeris; $\Delta I + \Delta T$ are the residuals (i.e., noncorrected parts) of the ionospheric and tropospheric delay errors after the corrections from the Klobuchar and Saastamoinen models. ε is the error mainly sourced from multipath/NLOS. With the corrected pseudorange ρ^c from (6), the pseudorange error $\Delta\rho$ can be derived by

$$\Delta\rho = \rho^c - R = c\left(\Delta\delta t^r - \Delta\delta t^{sv}\right) + \Delta I + \Delta T + \varepsilon. \tag{7}$$

The calculated results of the pseudorange error $\Delta\rho$ could then be labeled as the corresponding value from (7). Through the above we can obtain the corresponding pseudorange error $\Delta\rho$ for every set of variables from the GPS measurement, containing satellite elevation θ, carrier to noise ratio C/N_0 and pseudorange residuals η in the offline labeling phase.

D. GBDT-Based Training Process

The performance of machine learning is vital for the prediction of pseudorange error and hence the final positioning accuracy. The decision tree, also referred to as classification or regression tree is a popular machine learning method [38]. Nevertheless, the single decision tree can be unstable affected adversely by small uncertainties and perturbations in the training data sets [39]. One way of improving performance is to use a combination or ensemble of techniques through the accumulation individual learning results. Here, one ensemble decision tree algorithm, GBDT, able to minimize the decision tree training error based on the gradient boosting regression technique, is used for the pseudorange error prediction [40]. The problem can be defined as: Given a training sample $\{\boldsymbol{x}_i, \Delta\rho_i\}_1^N$ of known $(\boldsymbol{x}, \Delta\rho)$ values, the objective is to determine a function that maps $\boldsymbol{x}$ to $\Delta\rho$, such that over the joint distribution of all $(\boldsymbol{x}, \Delta\rho)$ values, the expected value of some specified loss function $L(\Delta\rho_i, f(\boldsymbol{x}_i))$ is minimised. In particular, $\boldsymbol{x}_i = (C/N_{0i}, \eta_i, \theta_i)$, $i = 1, 2, 3, \ldots, N$. i is the sequence number of the sample and N is the total number of the samples. $\Delta\rho_i$ is the corresponding labeled pseudorange error of $\boldsymbol{x}_i$. The GBDT-based pseudorange error prediction algorithm process is as follows.

1) Initiate predictions with a simple decision tree $f_0(\boldsymbol{x})$

$$f_0(\boldsymbol{x}) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L\left(\Delta\rho_i, \gamma\right) \tag{8}$$

where $f_0(\boldsymbol{x})$ is a regression decision tree containing only one root node and γ is a constant value which is the output of $f_0(\boldsymbol{x})$. In order to ensure that the loss function $L(\Delta\rho_i, f(\boldsymbol{x}_i))$ decreases in each iteration, the weak learner $h_m(\boldsymbol{x}_i; \boldsymbol{a}_m) m = 1, \ldots, M$, is created in the direction of steepest descent (i.e., a negative gradient direction). m is the sequence number of iterations. $h_m(\boldsymbol{x}_i; \boldsymbol{a}_m)$ is a decision tree with the parameter $\boldsymbol{a}$, which determines the splitting variable, split locations and terminal node of the individual tree.

2) For $m = 1, \ldots, M$:
   a) compute the negative gradient in the following equation:

$$\tilde{y}_i = -\left[\frac{\partial L\left(\Delta\rho_i, f(\boldsymbol{x_i})\right)}{\partial f(\boldsymbol{x}_i)}\right]_{f(\boldsymbol{x})=f_{m-1}(\boldsymbol{x})} \tag{9}$$

   where the loss function $L(\Delta\rho_i, f(\boldsymbol{x}_i))$ is the square loss function $(1/2)(\Delta\rho_i - f(\boldsymbol{x}_i))^2$ used in the iteration;
   b) create a new data set based on replacing $\Delta\rho_i$ in the training data set by $\tilde{y}_i$. The new data set is expressed as

$$T_m = \{(\boldsymbol{x}_1, \tilde{y}_1), (\boldsymbol{x}_2, \tilde{y}_2), \ldots, (\boldsymbol{x}_i, \tilde{y}_i), \ldots, (\boldsymbol{x}_N, \tilde{y}_N)\} \tag{10}$$

   a new weak predictor $h_m(\boldsymbol{x}_i; a_m)$ is created in the following equation by training of the new data set T_m to minimize the loss function

$$\boldsymbol{a}_m = \underset{\boldsymbol{a}}{\operatorname{argmin}} \sum_{i=1}^{N} (\tilde{y}_i - h_m(\boldsymbol{x}_i; \boldsymbol{a}))^2 \tag{11}$$

   c) update the original predictor with the new predictor multiplied by learning rate β to form a stronger predictor $f_m(\boldsymbol{x})$ in the following equation:

$$f_m(\boldsymbol{x}) = f_{m-1}(\boldsymbol{x}) + \beta h_m(\boldsymbol{x}; \boldsymbol{a}_m) \tag{12}$$

   where β is usually chosen to be a value between $0 \sim 1$ in order to prevent over-fitting. $h_m(\boldsymbol{x}; \boldsymbol{a}_m)$ is the weak predictor and $f_{m-1}(\boldsymbol{x})$ is the strong predictor from the previous iteration.
3) Output $f_M(\boldsymbol{x})$ as the final predictor after the iteration termination

$$f_M(\boldsymbol{x}) = f_0(\boldsymbol{x}) + \sum_{m=1}^{M} \beta h_m(\boldsymbol{x}; \boldsymbol{a}_m). \tag{13}$$

4) Once the final predictor $f_M(\boldsymbol{x})$ (i.e., the rules of the GBDT method) is obtained, the corresponding pseudorange errors of the newly collected variables from GPS measurement can be predicted. The input $\boldsymbol{x} = (C/N_0, \eta, \theta)$ is used together with the rules to predict the pseudorange errors for each observed satellite.

E. Positioning Calculation With Two Variations

The testing data set is used together with the pseudorange errors predicted by GBDT to compute the position with the two proposed variations for accuracy improvement.

1) *Positioning Based on Pseudorange Error Correction:* The newly collected pseudorange measurements can

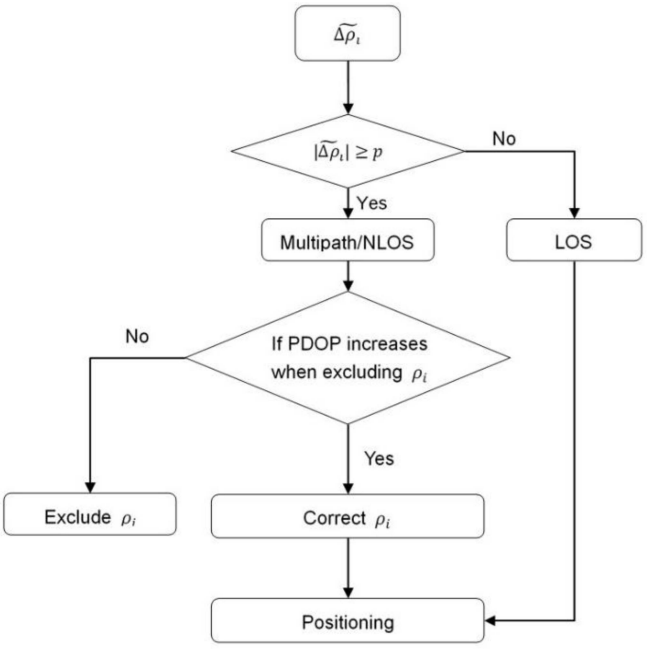

Fig. 2. Positioning based on multipath/NLOS signal exclusion.

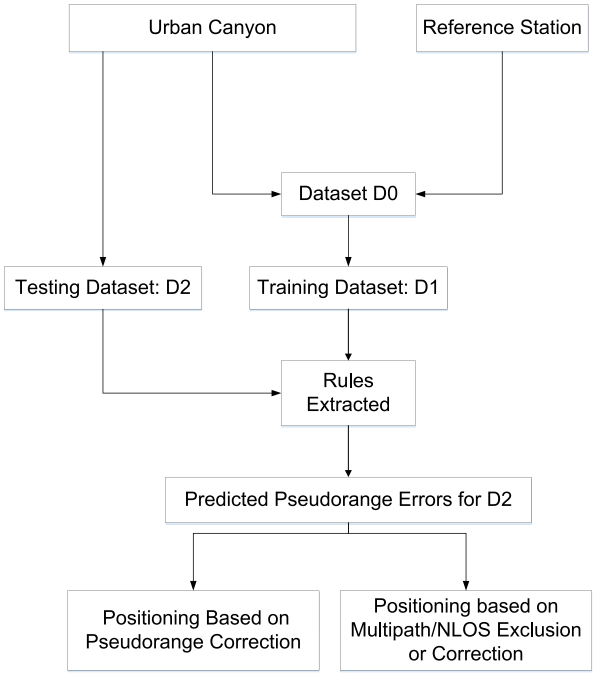

Fig. 3. Experiment procedure.

be corrected by subtracting the predicted pseudorange errors in the following equation:

$$\begin{cases} \rho_1^c = \rho_1 - \widetilde{\Delta\rho}_1 \\ \rho_2^c = \rho_2 - \widetilde{\Delta\rho}_2 \\ \quad\vdots \\ \rho_i^c = \rho_i - \widetilde{\Delta\rho}_i \end{cases} \tag{14}$$

where ρ_i^c is the corrected pseudorange of the ith signals, and $\widetilde{\Delta\rho}_i$ is the predicted pseudorange error of the ith signals. With the corrected pseudorange measurements, a least-squares algorithm is used for position computation.

2) *Positioning Based on Multipath/NLOS Signal Exclusion or Correction:* The process for the positioning based on multipath/NLOS signal exclusion or correction is illustrated in Fig. 2. The predicted pseudorange errors are used to classify multipath/ NLOS signals by comparing their absolute values with a proposed threshold p, see Fig. 2. From the results of field trials, NLOS signals can have an unbounded positive pseudorange error, while the multipath has a bounded positive or negative psudorange error depending on the signal processing method in the receiver design. The LOS measurement could also have small pseudorange errors due to the residual error terms, including the clock offset $\Delta\delta t^r$, satellite clock offset $\Delta\delta t^{sv}$, ionospheric delay ΔI and tropospheric delay ΔT, as introduced in (7). The threshold p is therefore determined empirically from field trials.

The candidate pseudorange with the predicted absolute pseudorange errors $|\widetilde{\Delta\rho_i}|$ within the threshold p is considered as a LOS signal and can be used for the final positioning. The candidate pseudorange with $|\widetilde{\Delta\rho_i}|$ greater than the threshold p is classified as the multipath/NLOS. Nevertheless, we do not remove all of the multipath/NLOS signals since satellite exclusions could degrade the satellite geometry and therefore, affect the final positioning accuracy. The value of the reference PDOP is calculated for all visible satellites in every epoch. At the same time, the candidate PDOPs are calculated for each time excluding one pseudorange. The following conditional strategies are then carried out: if the removal of a pseudorange causes the PDOP to increase, we will correct the pseudorange error following the process in (14) and then use the corrected pseudorange for the final positioning. Alternatively, if the removal of the pseudorange does not cause the PDOP to increase, we will exclude this pseudorange measurement in the final position computation.

III. Field Test and Analysis of Results

The field tests were carried out to validate the proposed algorithms. The two experiments followed the same procedure, with the data being collected on separate occasions in different urban environments, as shown in Fig. 3. The data set D0 consists of data collected from an urban canyon and a reference station. The urban canyon data sets mainly consist of NLOS and multipath signals, and, therefore, generally have large pseudorange errors. The reference station data sets mainly contain LOS signals and have relatively small pseudorange errors. In order to prevent a biased fitting, samples were extracted from the two parts of the data set (i.e., urban canyon and reference station) to form the training data set D1. The rules extracted from the training process were then used to predict pseudorange errors for the testing data set D2. The predicted pseudorange errors are used with the two variations of the proposed algorithm, that is: 1) correction-based positioning and 2) multipath/NLOS signals exclusion or correction-based positioning.

TABLE I
SUMMARY OF THE DATA SETS IN TEST CASE 1

	Training Dataset (D1)			Testing Dataset (D2)	
Collected Location	Urban Canyon (P_0) (two sides of buildings with narrow road)		Reference Station	Urban Canyon (P_0)	
Pseudorange Error	Large	Small	Small	large	small
Sample Size	24000	12000	12000	46759	50338
Total	48000			97097	

A. Test Case With Narrow Road and Buildings on Two Sides

The test case 1 was carried out on a typical urban canyon area with a narrow road and buildings on two sides. Static data at 1 Hz were collected at a point, P0, on a narrow road with buildings on two sides in the campus of Cheng Kung University using a commercial GPS receiver, NovAtel Propak 7, for about seven hours on September 17, 2018. In addition, on the same day, seven hours of static data at an interval of 30 s were collected at the reference station CKSV, located in Cheng Kung University using a geodetic receiver, Trimble NETR9.

The same number of samples were randomly selected from the large pseudorange errors (i.e., 24 000 samples from the urban canyon) and small pseudorange errors (i.e., 12 000 samples from urban canyon and 12 000 from reference station) to form the training data set D1. Another seven hours of data were also collected at 1 Hz at the same urban point, P0, from which 97 097 samples were selected to form the testing data set D2, see Table I.

During the training process, the GBDT was used to fit the pseudorange errors of data set D1. The rules extracted from the training data set D1 were used to predict the pseudorange errors of D2. Internal validation of the training process was based on the determination of the goodness of fit of the rules generated. While external validation, also known as testing, was undertaken using a separate data set, D2, in order to validate the correctness of the fitting results for the predicted pseudorange errors. The RMSE of the fitting results was used to evaluate the accuracy of the internal and external validation. The smaller the RMSE, the closer the fitting result is to the reference pseudorange errors.

A sensitivity analysis was performed to reveal the variation of the training accuracy of GBDT with the number of iterations for different parameters, see Fig. 4. The parameters used to evaluate the performance included the leaf number (LN), learning rate and number of iterations. LN means the number of leaf nodes for each regression tree in GBDT. Each tree grows until the number of leaf nodes reaches this value during the iteration. From the analysis results, the RMSE decreases as the number of iterations increases, and the RMSE decreases very slowly if the number of iterations reaches 1000. In particular, when LN = 20 and learning rate = 0.1 both external and internal validation have the smallest RMSE. We therefore adopt the values of 1000, 20 and 0.1, respectively, for the number of iterations, LN and learning rate during the GBDT-based training.

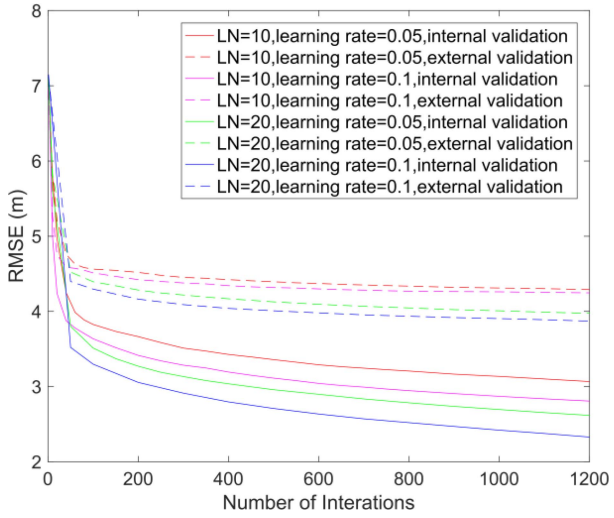

Fig. 4. Variation of the fitting results of GBDT with the number of iterations for different parameters used in test case 1.

TABLE II
COMPARISON OF GBDT AND TRADITIONAL FITTING METHODS IN THE TEST CASE 1

RMSE (m)	Internal Validation (validation)	External Validation (testing)
GBDT	2.19	4.10
Traditional Linear Fitting	5.16	5.55
Quadratic Polynomial Fitting	5.90	6.11
Cubic Polynomial Fitting	5.90	6.11
Quartic Polynomial Fitting	5.89	6.09

The comparisons of the GBDT and the traditional linear and nonlinear regressions, including quadratic polynomial fitting, cubic polynomial fitting and quartic polynomial fitting for the test case 1 are shown in Fig. 5 and Table II. For ease of viewing, only the fitting results of 200 samples are displayed. It is clear that the residual curve of GBDT is closer to zero and more stable than other methods. The RMSEs of the GBDT fitting results for internal and external validation are 2.19 and 4.10 m, respectively, which are smaller than the other traditional fitting methods.

Based on the pseudorange errors predicted by GBDT, the positioning results calculated with the two variations (i.e., correction-based positioning and multipath/NLOS signals exclusion or correction-based positioning) are compared with two conventional positioning methods, i.e., positioning with standard outlier detection and exclusion (conventional positioning method one [41]) and positioning with C/N_0 and elevation angle-based multipath/NLOS exclusion (conventional positioning method two). The classification thresholds for conventional positioning method two are that if the C/N_0 is greater than 30 dB and the elevation angle is higher than 15 degrees, the signal is considered as LOS, with the remaining signals being multipath/NLOS.

Table III presents the results of the second approach which detects and excludes or corrects the multipath/NLOS signals. Here, accuracy of classification is expressed as the percentage ratio of correctly classified signals to the total within the data set. A similar approach is used for each category in which

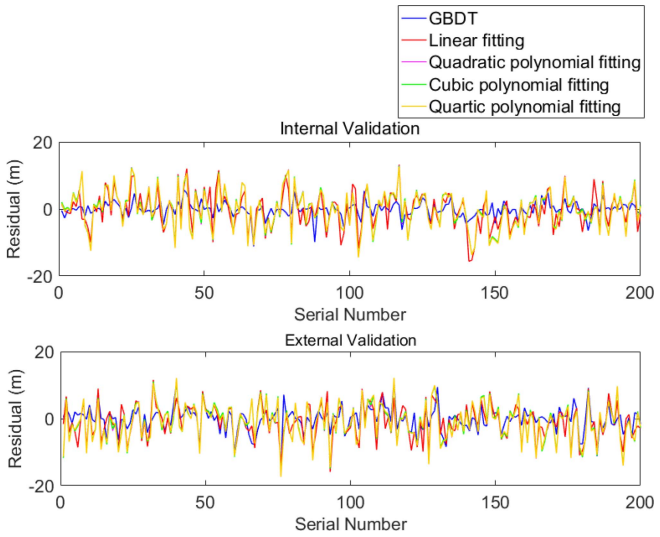

Fig. 5. Comparison of the fitting results in test case 1.

TABLE III
SIGNAL RECEPTION TYPE CLASSIFICATION ACCURACY IN TEST CASE 1

Reception Type	LOS	Multipath/NLOS
Classification Accuracy of Each Category	80.29%	71.49%
Overall Classification Accuracy	75.90%	

accuracy is measured by the ratio of the quantity of signals classified correctly to the quantity known to be in that category.

The classification threshold p was set to 5 in this test case, which means that the samples with absolute pseudorange errors lower than 5 m were considered to be LOS, while those with absolute pseudorange errors of 5 m or higher, were considered to be subject to significant multipath and NLOS effects. An overall classification accuracy of 76% was achieved by the proposed GBDT-based algorithm.

The positioning results of the two conventional methods, pseudorange correction and multipath/NLOS signal exclusion or correction, are depicted in Fig. 6. It can be seen that the results for the two variations of the algorithm proposed are closer to the ground truth than the conventional results, with the pseudorange correction performing better than the multipath/NLOS exclusion or correction. Table IV compares the positioning results of the candidate algorithms. It is shown that the pseudorange corrected based on positioning delivers 3-D positioning accuracies (RMSE) of 45.14m, an improvement of 25.95% on conventional positioning method one and 42.73% on conventional positioning method two. The improvement with the multipath/NLOS signal exclusion or correction method is, however, limited because of the poor classification accuracy of multipath/NLOS signals (i.e., 76%) and a weaker geometric configuration resulted from the reduction of the number of visible satellites.

The improvement in positioning using the proposed algorithm is also depicted in Fig. 7. It is shown that with the conventional positioning results only about 7% to 10% of the epochs are accurate to within 10 m. The proportion increases further to above 53% when the pseudorange correction is applied. In addition, with the application of the method proposed, the proportion of epochs with a positioning accuracy in the range of 10–20 m is much higher than that of conventional positioning, and fewer epochs exhibit positioning errors above 20 m. These results demonstrate the improvements in positioning accuracy using the two variations of the proposed method, in particular, the pseudorange correction method.

In order to investigate further the performance of the algorithm proposed, the positioning accuracy compared with the conventional method one (i.e., the better performing of the two conventional methods) is analysed according to each epoch in Table V. This shows that 81% and 79% of the epochs improved in the 2-D and 3-D positioning results based on the pseudorange correction applied. Although around 20% of the epochs deteriorated due to the GBDT prediction errors, the pseudorange correction is still effective for most of the epochs. The multipath/NLOS exclusion or correction improves the positioning accuracy of about 36% of the epochs, while the positioning accuracy for 58% of the epochs is neither improved nor worsened when the exclusion or correction is applied. One possible reason is that some of the multipath/NLOS are not detected by the algorithm and therefore, lead to no improvement in the positioning accuracy for these epochs. It is also possible that some epochs may have very few multipath/NLOS signals, making our algorithm less effective on these epochs. Only 6% of the epochs got worse with exclusion or correction applied in the form of the exclusion of the signals resulting from the incorrect signal reception classification (e.g., exclusion of LOS but using multipath/NLOS).

B. Test Case With Wide Road and High Rise Buildings on One Side

The test case 2 was undertaken in a typical urban environment with a wide road but with high rise buildings on one side. About 20 min of data were collected from each of two points (P1 and P2) to form part of the training data, with a frequency of 5 Hz using a NovAtel OEM6 geodetic receiver on September 20, 2018, see Fig. 8. In addition, on the same day, four hours of static data with an interval of five seconds were collected at the SatRef HKSC station in Hong Kong using a LEICA GR50 geodetic receiver. The training data set D1 was formed by combining the samples selected from the data collected in the two urban points and the reference station. The testing data set D2 was formed from the 5 Hz samples selected from the other 30 min of data collected in P1. Table VI provides a summary of the data sets used in test case 2.

The results of a sensitivity analysis of the GBDT-based algorithm in test case 2 is shown in Fig. 9. It is shown that the rate of reduction in the RMSE becomes very slow when the iteration number reaches 400. The minimum RMSE is achieved for the external and internal validation when LN = 20 and the learning rate = 0.1. Therefore, we adopt the values of 400, 20, and 0.1, respectively, for the number of iterations, LN and learning rate during the GBDT-based training in this test case.

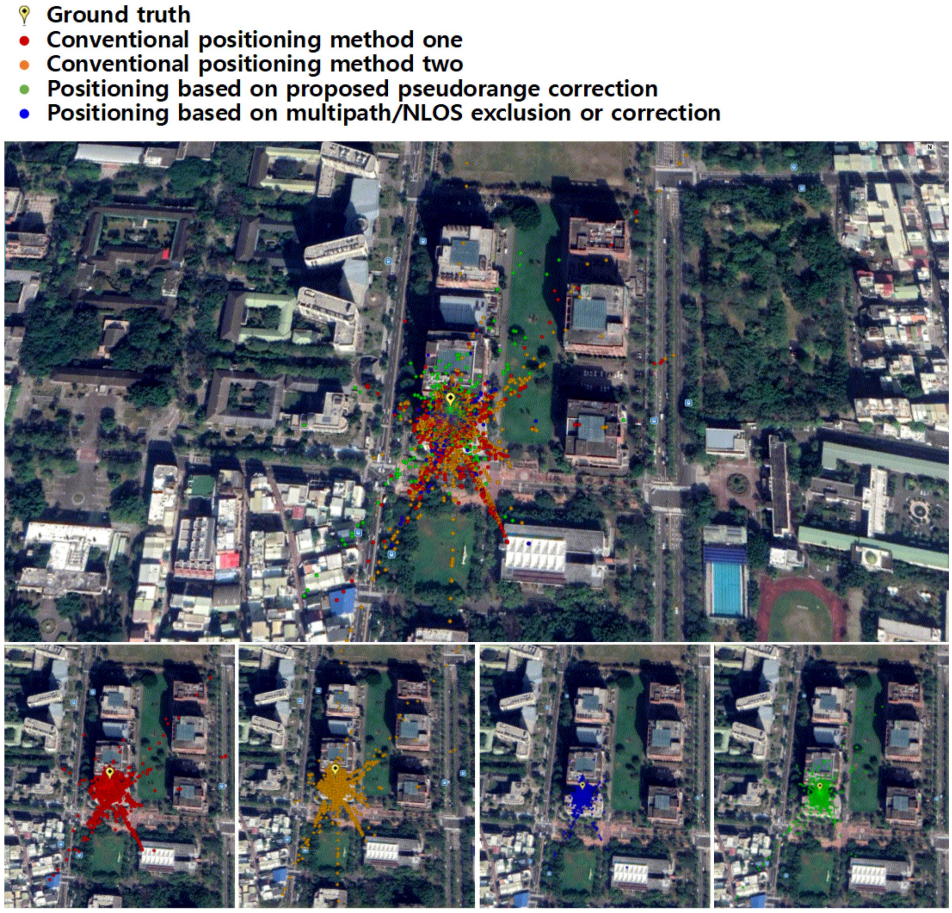

Fig. 6. Positioning results in test case 1.

TABLE IV
POSITIONING ACCURACY COMPARISON IN TEST CASE 1

RMSE (m)	E	N	U	3D	2D
Conventional Positioning Method One	21.38	27.78	49.88	60.96	35.06
Conventional Positioning Method Two	21.93	37.05	65.54	78.82	43.06
Positioning Based on Pseudorange Correction	15.66	17.38	38.60	45.14	23.40
Positioning Based on Multipath/NLOS Signal Exclusion or Correction	20.82	26.77	48.11	58.85	33.91

The comparisons of the GBDT and the traditional linear and nonlinear regressions for the test case 2 are shown in Fig. 10 and Table VII. An example of the first 200 fitting results is displayed in Fig. 10. It is clear that the residual curve for GBDT is closer to zero and more stable than other methods. The RMSEs for GBDT-based fitting results for internal and external validation are 5.96 m and 12.75 m, respectively, which are significantly smaller than the other traditional fitting methods.

The signal reception classification results used for the positioning in variation 2 is shown in Table VIII. In particular, the classification threshold p was set as 50 in this test case,

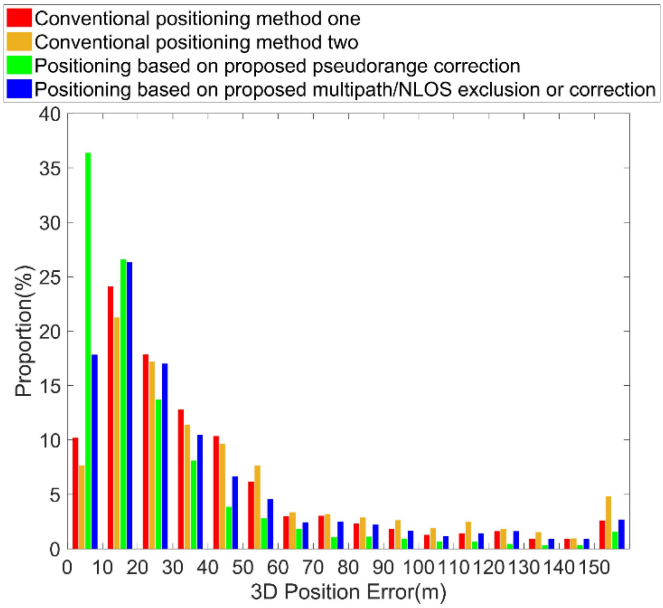

Fig. 7. Positioning accuracy histogram of test case 1.

with absolute pseudorange errors lower than 50 m considered as LOS and above 50 m as multipath/NLOS. Since the multipath/NLOS effect (i.e., especially the NLOS) is more severe than that of test case 1, a larger threshold is used

TABLE V
ALGORITHM PERFORMANCE EVALUATION WITH PROPORTION OF EPOCHS IN TEST CASE 1

Proportion of Epochs (%)	3D			2D		
	Better	Equal	Worse	Better	Equal	Worse
Positioning Based on Pseudorange Corrections	81.34	0.00	18.66	79.46	0.00%	20.54
Positioning Based on Multipath/NLOS Signal Exclusion or Correction	36.49	57.93	5.58	36.05	57.92	6.03

Fig. 8. Locations of P1 and P2.

to ensure a sufficient number of satellites for the positioning after the multipath/NLOS exclusion. Table VIII shows that the proposed GBDT-based algorithm delivers an overall classification accuracy of 91%, with 97% and 73% for the LOS and multipath/NLOS classification accuracy, respectively.

The results for the conventional positioning methods, pseudorange correction and multipath/NLOS signal exclusion or correction, are depicted in Fig. 11 and Table IX. It can be seen that positioning with pseudorange correction provides a 3-D RMSE of 23.27 m, while it is more than 80 m for the two conventional positioning methods, indicating an improvement of about 71%. Furthermore, the horizontal and vertical positioning accuracies are improved significantly, by about 76% and 59%, respectively. The multipath/NLOS signal exclusion or correction could also provide some degree of improvement in the positioning results, with about 30% and 4% for the horizontal and height improvements, respectively.

The positioning accuracy improvement for the proposed algorithm is also depicted in Fig. 12. It is shown that with the pseudorange corrections applied, the epochs with a positioning accuracy within 30 m have increased significantly, while the accuracy for the main epochs when using conventional positioning ranged from 60 to 90 m. Although a less notable improvement, the multipath/NLOS exclusion-based positioning increased the accuracy of the positioning of most epochs to around 30 to 60 m, better than the conventional positioning results.

The positioning accuracy by epoch is further analysed by comparing with conventional method one (which showed

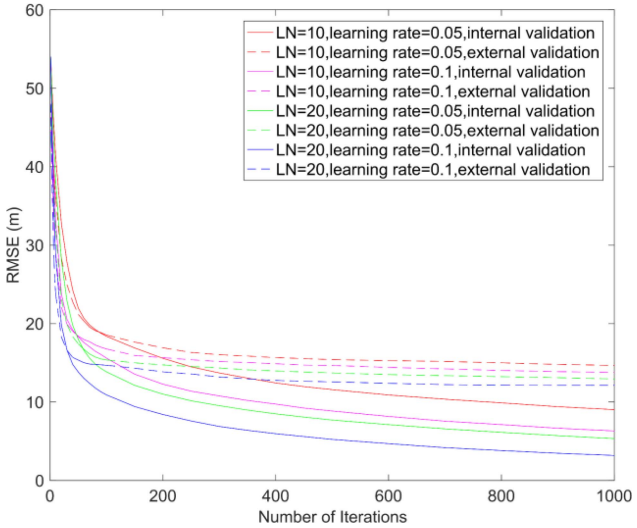

Fig. 9. Variation of the training accuracy of GBDT with the number of iterations for the different parameters used in test case 2.

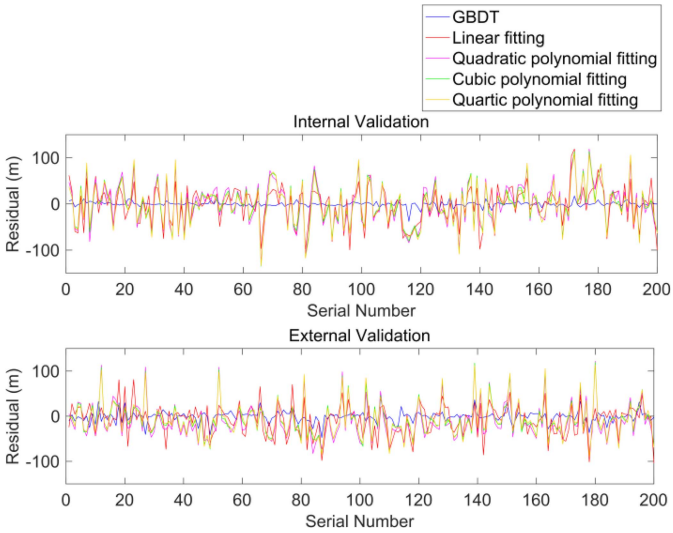

Fig. 10. Comparison of the fitting results in test case 2.

a similar performance to the conventional method two) in Table X. It is shown that the 2-D and 3-D positioning results improved in 98% and 97% of the epochs, respectively, when the pseudorange correction was applied, while only around 3% of the epochs got worse, due to the GBDT prediction errors. The multipath/NLOS exclusion or correction method, meanwhile, improved the positioning accuracy of about 81% (3-D) and 91% (2-D) of the epochs, while the positioning accuracy for 9% (3-D and 2-D) of the epochs did not change when the exclusion or correction was applied. The fact that there were fewer epochs that stayed the same in test case 2 (i.e., 9%) compared to test case 1 (i.e., 58%) was because the quality of the observed satellite measurements were much poorer in test case 2 than in test case 1, which means that there were more severe multipath/NLOS measurements in the test case 2. It is also validated that tall buildings (i.e., in test case 2) will result in more severe multipath/NLOS, i.e., larger pseudorange errors, than short buildings (i.e., in test case 1). The worse

TABLE VI
SUMMARY OF THE TRAINING DATA SET D1 IN TEST CASE 2

	Training Dataset (D1)					Testing Dataset (D2)	
Collected Location	Urban Canyon				Reference Station	Urban Canyon	
	P1		P2				
Pseudorange Error	Large	Small	Large	Small	Small	Small	large
Sample Size	2400	1600	2400	1600	1600	14971	4686
Total	9600					19657	

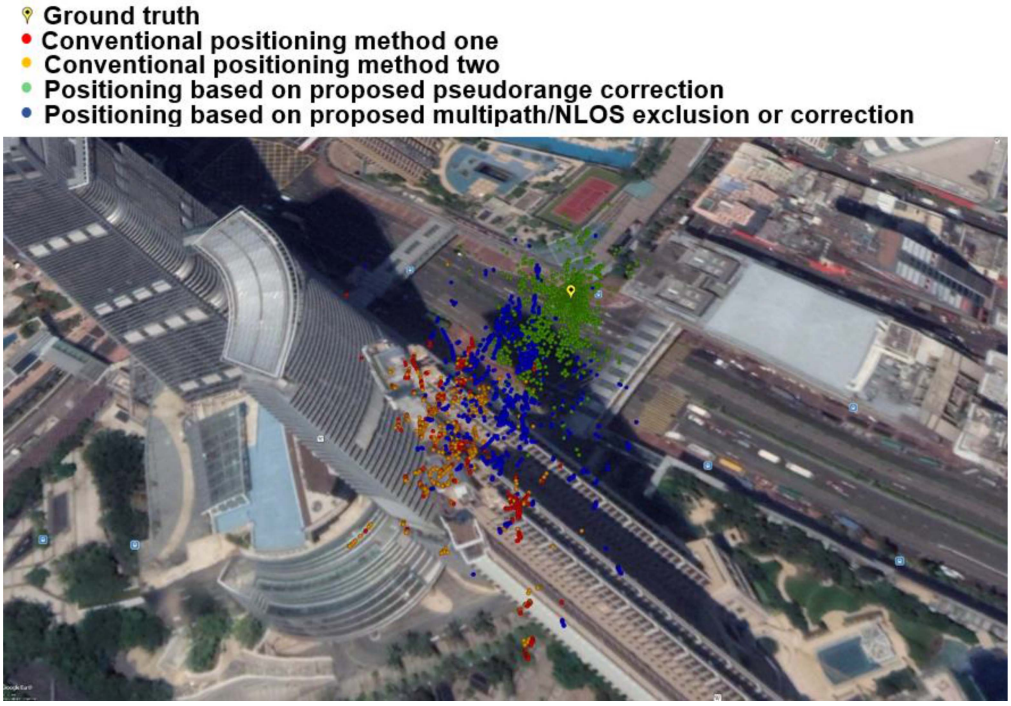

Fig. 11. Positioning results in the test case 2.

TABLE VII
COMPARISON OF GBDT AND TRADITIONAL FITTING METHODS IN THE TEST CASE 2

RMSE (m)	Internal Validation (validation)	External Validation (testing)
GBDT	5.96	12.75
Traditional Linear Fitting	38.29	35.47
Quadratic Polynomial Fitting	44.69	41.50
Cubic Polynomial Fitting	42.77	40.62
Quartic Polynomial Fitting	42.38	40.45

TABLE VIII
SIGNAL RECEPTION TYPE CLASSIFICATION ACCURACY IN TEST CASE 2

Reception Type	LOS	Multipath/NLOS
Classification Accuracy of Each Category	96.91%	73.15%
Overall Classification Accuracy	91.22%	

epochs for the 2-D positioning is only 0.4% but with 10% for 3-D positioning due to the height component which is more affected by the change in pseudorange error and PDOP in challenging urban areas.

TABLE IX
POSITIONING ACCURACY COMPARISON IN TEST CASE 1

RMSE (m)	**E**	**N**	**U**	**3D**	**2D**
Conventional Positioning Method One	35.70	56.27	40.51	81.26	66.64
Conventional Positioning Method Two	35.80	56.06	46.24	80.01	66.52
Positioning Based on Pseudorange Correction	8.77	13.69	16.65	23.27	16.26
Positioning Based on Multipath/NLOS Signal Exclusion or Correction	23.94	40.06	38.94	60.78	46.67

In summary, by analysing the positioning results in two test cases from the proposed algorithm, we found that variation 1 (i.e., pseudorange correction) is superior to variation 2 (i.e., multipath/NLOS exclusion or correction), although both are superior to the conventional methods. This is because the correction-based method could consider all of the pesudorange information, which was especially important in conditions where there were insufficiently healthy satellite observations.

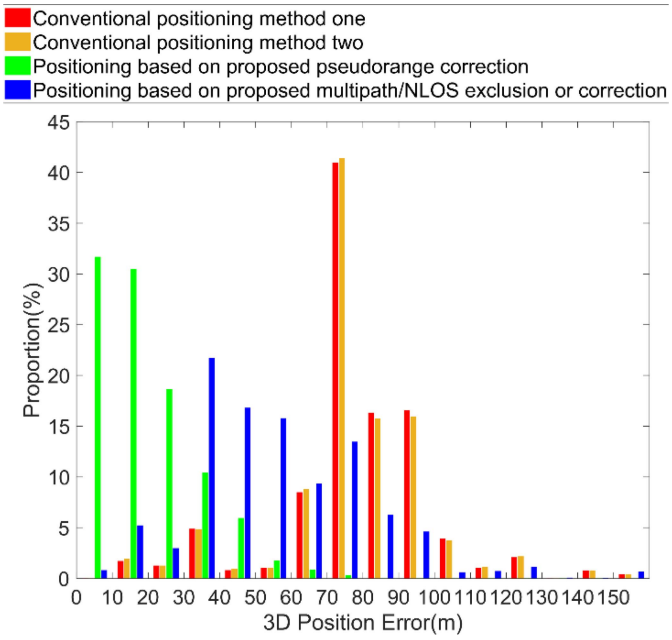

Fig. 12. Positioning accuracy histogram of test case 2.

TABLE X
ALGORITHM PERFORMANCE EVALUATION WITH PROPORTION OF EPOCHS IN TEST CASE 2

Proportion of Epochs (%)	3D			2D		
	Better	Equal	Worse	Better	Equal	Worse
Positioning Based on Pseudorange Corrections	96.50	0.00	3.50	97.80	0.00	2.20
Positioning Based on Multipath/NLOS Signal Exclusion or Correction	81.13	8.77	10.10	90.53	9.11	0.36

In addition, when correcting the multipath errors, this method can correct the residual errors caused by inaccurate ionospheric and tropospheric correction models. The fact that the improvement was more marked in test case 2 than test case 1 indicated that our algorithm could have better performance in environments exhibiting severe pseudorange errors, especially, therefore, challenging urban areas.

IV. CONCLUSION

We have proposed a GBDT-based machine learning algorithm to improve positioning accuracy in built environments, compared to conventional positioning methods. Based on the GBDT-predicted pseudorange errors from the inputs of elevation angle, C/N_0 and pseudorange residual, two variations of the algorithm (positioning with pseudorange correction and multipath/NLOS signal exclusion or correction) are proposed and analysed in two typical urban test scenarios. The results demonstrate that multipath/NLOS signal exclusion, while improving the results, is sensitive to both the proximity of a receiver to signal reflectors/blockers and geometrical configuration. On the other hand, the pseudorange correction approach does not suffer from these limitations and delivers better performance compared to the conventional least-squares all-satellites-received solution. The results, for challenging urban areas with high rise buildings on one side, show an improvement in the 3-D positioning accuracy (RMSE) from 81.3 m, with conventional positioning approaches, to 60.8 m when our multipath/NLOS variation is applied; an improvement of more than 25%. Even more strikingly the corresponding results for the pseudorange error correction-based positioning (compared to the conventional) is from 81.3 m to 23.3 m; an improvement of more than 70%.

The proposed methods, especially the pseudorange correction one, do not require additional sensors and can avoid the errors from the use of additional hardware or geospatial information during the labeling phase, thus improving positioning accuracy with a small amount of calculation and lower cost. Our ongoing research is developing a framework of evenly distributed grid of reference points in urban areas, with the data collected from these reference points used for offline training. With this approach, users will automatically obtain the rules online via communication links to nearby reference points for accurate pseudorange error correction to improve positioning accuracy. This GBDT based online data training mechanism, i.e., with frequently updated rules, will also be developed for real-time and post-processing applications, including robust ground vehicle and pedestrian location-based services [42].

REFERENCES

[1] J. M. Tranquilla, J. P. Carr, and H. M. Al-Rizzo, "Analysis of a choke ring groundplane for multipath control in global positioning system (GPS) applications," *IEEE Trans. Antennas Propag.*, vol. 42, no. 7, pp. 905–911, Jul. 1994.

[2] Z. Jiang and P. D. Groves, "NLOS GPS signal detection using a dual-polarisation antenna," *GPS Solutions*, vol. 18, no. 1, pp. 15–26, Jan. 2014.

[3] A. J. V. Dierendonck, P. Fenton, and T. Ford, "Theory and performance of narrow correlator spacing in a GPS receiver," *Navigation*, vol. 39, no. 3, pp. 265–283, Sep. 1992.

[4] R. D. J. V. Nee, J. Siereveld, P. C. Fenton, and B. R. Townsend, "The multipath estimating delay lock loop: Approaching theoretical accuracy limits," in *Proc. IEEE/ION Position Location Navig. Symp.*, Las Vegas, NV, USA, 1994, pp. 246–251.

[5] B. R. Townsend and P. C. Fenton, "A practical approach to the reduction of pseudorange multipath errors in a Ll GPS receiver," in *Proc. Int. Techn. Meeting Satellite Division Inst. Navig.*, Salt Lake City, UT, USA, 1994, pp. 143–148.

[6] B. R. Townsend, P. C. Fenton, K. J. V. Dierendonck, and R. D. J. V. Nee, "Performance evaluation of the multipath estimating delay lock loop," *Navigation*, vol. 42, no. 3, pp. 502–514, Sep. 1995.

[7] L. Garin and J. M. Rousseau, "Enhanced strobe correlator multipath rejection for code & carrier," in *Proc. ION GPS*, Kansas, MO, USA, 1997, pp. 559–568.

[8] M. S. Braasch, "Performance comparison of multipath mitigating receiver architectures," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, 2001, pp. 1309–1315.

[9] P. C. Fenton and J. Jones, "The theory and performance of NovAtel Inc.'s vision correlator," in *Proc. ION GNSS*, Long Beach, CA, USA, 2005, pp. 2178–2186.

[10] N. Blanco-Delgado and M. U. de Haag, "Multipath analysis using code-minus-carrier for dynamic testing of GNSS receivers," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, 2011, pp. 25–30.

[11] I. Smolyakov, M. Rezaee, and R. Langley, "Resilient multipath prediction and detection architecture for low-cost navigation in challenging urban areas," *Navigation*, vol. 67, no. 2, pp. 397–409, Jun. 2020.

[12] M. G. Petovello, C. O'Driscoll, and G. Lachapelle, "Weak signal carrier tracking of weak using coherent integration with an ultra-tight GNSS/IMU receiver," presented at the Eur. Navig. Conf., Toulouse, France, Apr. 2008, pp. 283–292.
[13] M. G. Petovello and G. Lachapelle, "Comparison of vector-based software receiver implementations with application to ultra-tight GPS/INS integration," in *Proc. ION GNSS*, Fort Worth, TX, USA, 2006, pp. 2977–2989.
[14] V. Milanes, J. E. Naranjo, C. Gonzalez, J. A. Ruiz, and T. Pedro, "Autonomous vehicle based in cooperative GPS and inertial systems," *Robotica*, vol. 26, no. 5, pp. 627–633, Sep. 2008.
[15] J.-I. Meguro, T. Murata, J.-I. Takiguchi, Y. Amano, and T. Hashizume, "GPS multipath mitigation for urban area using omnidirectional infrared camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 22–30, Mar. 2009.
[16] A. Soloviev, C. Toth, and D. Grejner-Brzezinska, "Performance of deeply integrated GPS/INS in dense forestry areas," *J. Appl. Geodesy*, vol. 6, no. 1, pp. 3–13, Mar. 2012.
[17] A. Soloviev and F. V. Graas, "Use of deeply integrated GPS/INS architecture and laser scanners for the identification of multipath reflections in urban environments," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 786–797, Oct. 2009.
[18] P. D. Groves, "Shadow matching: A new GNSS positioning technique for urban canyons," *J. Navig.*, vol. 63, no. 3, pp. 417–430, Jul. 2011.
[19] L. Wang, P. Groves, and M. Ziebart, "Urban positioning on a smartphone: Real-time shadow matching using GNSS and 3D city models," in *Proc. ION GNSS*, Nashville, TN, USA, 2013, pp. 1606–1619.
[20] S. Miura, L. T. Hsu, F. Chen, and S. Kamijo, "GPS error correction with pseudorange evaluation using three-dimensional maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3104–3115, Dec. 2015.
[21] Y. Gu, L.-T. Hsu, and S. Kamijo, "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.*, vol. 65, no. 6, pp. 4274–4287, Jun. 2016.
[22] L. T. Hsu, Y. Gu, and S. Kamijo, "3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation," *GPS Solutions*, vol. 20, no. 3, pp. 413–428, Mar. 2016.
[23] P. D. Groves and M. Adjrad, "Performance assessment of 3D-mapping-aided GNSS part 1: Algorithms, user equipment, and review," *Navigation*, vol. 66, no. 2, pp. 341–362, Mar. 2019.
[24] R. Yozevitch, B. B. Moshe, and A. Weissman, "A robust GNSS LOS/NLOS signal classifier," *Navigation*, vol. 63, no. 4, pp. 429–442, Dec. 2016.
[25] G. Deng, "GPS satellite selection algorithm based on satellite elevation angle and GDOP (in chinese)," *Digit. Commun.*, vol. 37, no. 2, pp. 47–50, Jun. 2010.
[26] L. Wang, P. D. Groves, and M. Ziebart, "Smartphone shadow matching for better cross-street GNSS positioning in urban environments," *J. Navig.*, vol. 68, no. 3, pp. 411–433, May 2015.
[27] L. T. Hsu, H. Tokura, N. Kubo, Y. Gu, and S. Kamijo, "Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons," *Sensors*, vol. 17, no. 6, pp. 1909–1917, Jan. 2017.
[28] L. T. Hsu, "GNSS multipath detection using a machine learning approach," in *Proc. Int. Conf. Intell. Transp. Syst.*, Yokohama, Japan, Oct. 2017, pp. 1–6.
[29] N. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed. New York, NY, USA: Wiley, 1981, pp. 7–13.
[30] S. Chatterjee and A. S. Hadi, "Influential observations, high leverage points, and outliers in linear regression," *Stat. Sci.*, vol. 1, no. 3, pp. 379–416, 1986.
[31] Q. Phan, S. Tan, I. V. Mcloughlin, and L. D. Vu, "A unified framework for GPS code and carrier-phase multipath mitigation using support vector regression," in *Proc. Adv. Artif. Neural Syst.*, Mar. 2013, pp. 1–14.
[32] M. Socharoentum, H. A. Karimi, and Y. Deng, "A machine learning approach to detect non-line-of-sight GNSS signals in Nav2Nav," presented at the 23rd ITS World Congr., Oct. 2016.
[33] Y. Quan, L. Lau, G. W. Roberts, X. Meng, and C. Zhang, "Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning," *Remote Sens.*, vol. 10, no. 12, pp. 1–19, Dec. 2018.
[34] B. Guermah, H. E. Ghazi, T. Sadiki, and H. Guermah, "A robust GNSS LOS/multipath signal classifier based on the fusion of information and machine learning for intelligent transportation systems," in *Proc. IEEE ICTMOD*, Marrakech, Morocco, Nov. 2018, pp. 94–100.
[35] R. Sun, L. T. Hsu, D. Xue, G. Zhang, and W. Y. Ochieng, "GPS signal reception classification using adaptive neuro-fuzzy inference system," *J. Navig.*, vol. 72, no. 3, pp. 685–701, May 2019.
[36] R. Sun, G. Wang, W. Zhang, L. T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," *Appl. Soft Comput.*, vol. 86, pp. 1–12, Jan. 2020.
[37] H. J. Euler and C. C. Goad, "On optimal filtering of GPS dual frequency observations without using orbit information," *J. Geodesy*, vol. 65, no. 2, pp. 130–143, Jun. 1991.
[38] J. H. Cho and P. Kurup, "Decision tree approach for classification and dimensionality reduction of electronic nose data," *Sensors Actuators B Chem.*, vol. 160, no. 1, pp. 542–548, Dec. 2011.
[39] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transp. Res. C Emerg. Technol.*, vol. 58, pp. 308–324, Mar. 2015.
[40] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, Nov. 2000.
[41] L. Biagi and S. Caldera, "An efficient leave one block out approach to identify outliers," *J. Appl. Geodesy*, vol. 7, no. 1, pp. 11–19, Mar. 2013.
[42] R. Sun, Q. Cheng, F. Xie, W. Zhang, T. Lin, and W. Y. Ochieng, "Combining machine learning and dynamic time wrapping for vehicle driving event detection using smartphones," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 5, 2019, doi: 10.1109/TITS.2019.2955760.

Rui Sun (Member, IEEE) received the M.Sc. degree in satellite positioning technology from the University of Nottingham, Nottingham, U.K., in 2007, and the Ph.D. degree in intelligent transport systems from the Imperial College Engineering Geomatics Group, Imperial College London, London, U.K., in 2015.

She is currently an Associate Professor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research interests include seamless positioning and navigation in challenging environments.

Guanyu Wang received the B.S. degree in transportation from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2017, where he is currently pursuing the postgraduation degree in transportation engineering.

His research interest includes the mitigation of multipath effect and improvement of GNSS positioning accuracy using 3-D city maps.

Qi Cheng received the B.S. degree in transportation engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2017, where he is currently pursuing the postgraduation degree in transportation engineering.

His research interest is GNSS NLOS detection algorithms for the intelligent transport system applications.

Linxia Fu received the B.S. degree in electrical engineering from Zhejiang University of Technology, Hangzhou, China, in 2019. She is currently pursuing the postgraduation degree with Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Her research interests include multipath/NLOS detection and GNSS positioning accuracy improvement in urban environments.

Kai-Wei Chiang received the B.Sc. and M.Sc. degrees in geomatics engineering from National Cheng Kung University, Tainan, Taiwan, in 1995 and 1997, respectively, and the Ph.D. degree from the Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada, in 2005.

He is a Professor with the Department of Geomatics, National Cheng Kung University. His areas of expertise include multisensor systems, mobile mapping systems, real-time static and kinematic positioning, and digital photogrammetry and their applications in mapping and geospatial information systems.

Li-Ta Hsu (Member, IEEE) received the B.S. and Ph.D. degrees in aeronautics and astronautics from National Cheng Kung University, Tainan, Taiwan, in 2007 and 2013, respectively.

He was a visiting Ph.D. student with the Faculty of Engineering, University College London, London, U.K., in 2012. He is currently a Japan Society for the Promotion of Sciences Postdoctoral Fellow with the Institute of Industrial Science, University of Tokyo, Tokyo, Japan. His research interests include all aspects of global navigation satellite system (GNSS) navigation, positioning, and signal processing, including wide-area differential GNSS systems and improving GNSS performance under challenging reception conditions and indoor positioning techniques.

Washington Yotto Ochieng received the Ph.D. degree in space geodesy from the University of Nottingham, U.K., in 1993.

He is the Head of the Centre for Transport Studies, the Co-Director of the Institute for Security Science and Technology, and the Chair in Positioning and Navigation Systems, Imperial College London, London, U.K. His research interests are in infrastructure resilience and security, user-centric intermodal mobility and the design of positioning and navigation systems for land, sea, air and space applications.

Dr. Ochieng is a Fellow of the Royal Academy of Engineering.