Applied Soft Computing Journal 86 (2020) 105942

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

A gradient boosting decision tree based GPS signal reception classification algorithm

Rui Sun [a,b,c,*], Guanyu Wang [a], Wenyu Zhang [a], Li-Ta Hsu [d], Washington Y. Ochieng [e]

[a] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] Shandong Provincial Key Laboratory of Optical Astronomy and Solar-Terrestrial Environment, Shandong University, Weihai 264209, China
[c] State Key Laboratory of Geo-Information Engineering, Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China
[d] Interdisciplinary Division of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong
[e] Center for Transport Studies, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 15 May 2019
Received in revised form 14 October 2019
Accepted 15 November 2019
Available online 26 November 2019

Keywords:
GPS
GBDT
Urban canyon
Multipath
NLOS

ABSTRACT

In urban areas, GPS signals are often reflected or blocked by buildings, which causes multipath effects and non-line-of-sight (NLOS) reception respectively consequently degrading GPS positioning performance. While improved receiver design can reduce the effect of multipath to some extent, it cannot deal with NLOS. Modelling methods based on measurements have shown promise to reduce the effect of NLOS signal reception. However, this depends on their ability to accurately and reliably classify line-of-sight (LOS), multipath and NLOS signals. The traditional method is based on one feature using signal strength as measured by the carrier to noise ratio, C/N_0. However, this feature is ineffective in capturing the characteristics of multipath and NLOS in all environments. In this paper, to improve the accuracy of signal reception classification, we are using the three features of C/N_0, pseudorange residuals and satellite elevation angle with a gradient boosting decision tree (GBDT) based classification algorithm. Experiments are carried out to compare the proposed algorithm with classifiers based on decision tree, distance weighted k-nearest neighbour (KNN) and the adaptive network-based fuzzy inference system (ANFIS). Test results from static receivers in urban environments, show that the GBDT based algorithm achieves a classification accuracy of 100%, 82% and 86% for LOS, multipath and NLOS signals, respectively. This is superior to the other three algorithms with the corresponding results of 100%, 82% and 84% for the Distance-Weighted KNN, 99%, 70% and 65% for the ANFIS and 98%, 35% and 95% for the traditional decision tree. With the NLOS detection and exclusion, the proposed GBDT with multi-feature based method can provide a positioning accuracy improvement of 34.1% compared to the traditional C/N_0 based method.

1. Introduction

Global navigation satellite systems (GNSS), especially the global positioning system (GPS), are widely used in positioning, navigation and timing (PNT), increasingly playing an important role in every aspect of our lives. The development of smart bus systems, emerging unmanned aerial vehicles (UAV) and autonomous vehicles all require accurate and reliable GPS positioning solutions. GPS signals, however, are easily reflected or blocked by buildings, with the consequences of either reduced accuracy or no positioning solutions in some urban environments such as canyons. There are three GPS signal reception types: ① line-of-sight (LOS): direct signal between the satellite and the user receiver; ② multipath: reflected and direct line-of-sight (LOS) signals received together; and ③ non-line-of-sight (NLOS): where the user can only receive a reflected signal. Positioning errors caused by NLOS and multipath cannot be removed by differential techniques [1,2] with the latter potentially resulting in positioning error of around 100 m [3]. Hence, many methods have been proposed to mitigate the effects of multipath and NLOS, including antenna design, signal processing and measurement-based modelling.

Some expensive high-grade antennas could be used to mitigate multipath and NLOS effects [4–6]. Choke-ring antennas can effectively reduce the multipath effect at low elevation. However, they are not only expensive but also bulky and hence, inappropriate for some location-based applications. Dual-polarization antennas could be used for NLOS detection. However, they are mainly used for geodetic applications such as snow depth detection, due to the high cost and bulk of the antennas. Dierendonck et al. showed how narrowing the spacing between early and late receiver code correlators helps to mitigate multipath and NLOS

* Corresponding author at: College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China.
E-mail address: rui.sun@nuaa.edu.cn (R. Sun).

https://doi.org/10.1016/j.asoc.2019.105942
1568-4946/

effects [7]. Subsequently, multipath estimating delay lock loops (MEDLL) [8], vision correlators and strobe correlators have been developed [9–13]. These methods, referred to as signal processing based multipath mitigation, assume that both direct and reflected signals are able to reach the receiver, thus a signal processing technique can be applied to filter the reflected signal. Unfortunately, these methods do not mitigate the effect of NLOS signal reception as it comprises of the reflected signal only.

Measurement-based modelling refers to using satellite observations and related information to mitigate NLOS reception and multipath effects. GPS measurements can be augmented by complementary additional sensors (such as inertial measurement units) to improve performance in urban canyons. However, this is highly dependent on the performance of the inertial measurements and their mechanizations [14–21]. Integrating GNSS with vision sensors can be effective in reducing multipath error and resulting in the improvement of positioning accuracy. However, the performance is affected by weather and environment features [22]. In addition, the expense associated with such sensors precludes their use in location-based services.

Another relatively low-cost approach to mitigate multipath errors is to use a 3D city model as an information source. The model can be used to classify the type of signal reception and thus exclude NLOS signals during positioning [23,24]. However, this is very likely to decrease the number of available satellites, leading both to a weaker geometry and lower redundancy required for the reduction in random errors and integrity monitoring. A more reasonable approach is to use the NLOS signal for positioning with the shadow matching method. Shadow matching utilizes 3D city models to predict satellite visibility and then compares this prediction with the measured satellite visibility to determine the position [25–27]. In addition, a research team at the University of Tokyo used a 3D city model to simulate signal propagation path and computed simulated pseudorange measurements, before weighting each candidate based on the similarity between the simulated and actual pseudorange measurements. The estimated position was then obtained by the weighted average of the candidate positions [28–30]. The performance of this method depends on the correct and reliable classification of LOS, multipath and NLOS signals.

The traditional method for signal classification is to define a threshold value of C/N_0 with higher C/N_0 classified as LOS, while those with lower C/N_0 classified as NLOS. Although Yozevitch et al. have shown that under conditions of no interference, C/N_0 can serve as an excellent classifier [31,32], in practice LOS signals can be detected with low C/N_0 for various reasons (e.g., the antenna location, momentary blocks, etc.) and NLOS signals can be detected with relatively high C/N_0. Hence, additional signal features are required for reception type classification. Therefore, in addition to C/N_0, satellite elevation can be used as a feature for the classification. In general, the higher the elevation, the more likely it is for a signal to be LOS, but as the elevation decreases, the satellite is more likely to be blocked by buildings and other man-made obstacles, with the corresponding signal being NLOS. Deng proposed a satellite selection algorithm based on satellite elevation angle and geometric dilution precision (GDOP). The influence of elevation angle on positioning accuracy is analysed [33]. Wang et al. also use C/N_0 and satellite elevation angle to evaluate the possibility of LOS [34]. In addition, larger pseudorange residuals could point to a higher probability of NLOS or multipath, and therefore, this can also be used to classify signal reception [35]. The other features that could be used for classification are pseudorange change rate, horizontal dilution of precision (HDOP) and vertical dilution of precision (VDOP) [32,36].

Machine learning has the advantages of high speed and accuracy in dealing with various types of features and has been used in recent years to improve the accuracy of GNSS positioning and signal reception classification. Phan et al. used elevation and azimuth angles as the key features of support vector machines (SVM) to mitigate the multipath effect [37]. Yozevitch et al. used C/N_0, elevation and other observations as features of decision trees to classify LOS and NLOS [32]. Monsak et al. proposed a method using machine learning to detect NLOS signals in a collaborative vehicle environment. The results are compared with several different machine learning algorithms [38]. Hsu et al. applied SVM to distinguish received GPS signal types, using four features, including C/N_0, temporal difference of C/N_0, pseudorange residual and pseudorange rate, extracted from raw measurements [36]. Quan et al. proposed a convolutional neural network (CNN) based multipath detection method with the sparse auto-encoder (SAE) for feature extraction [39]. Guermaha et al. proposed a GNSS signal classifier system based on the satellite elevation and the difference of C/N0 value, provided by right-hand circular polarized (RHCP) and left-hand circular polarized (LHCP) antennas, with the decision tree [40]. Sun et al. used nine variables, including the C/N_0, temporal difference of C/N_0, HDOP, VDOP, satellite elevation angle and azimuth angle, pseudorange residual, consistency between delta pseudorange and pseudorange rate and number of visible satellites, derived from the raw GPS measurements together with an algorithm based on an ANFIS to classify LOS, multipath and NLOS measurements from GPS [41]. These studies have shown that machine learning is a potentially effective method for GPS signal reception classification. To date, this potential is still to be exploited, because it is difficult to obtain a trade-off between high classification accuracy and low computational cost with an increasing number of input features.

Therefore, in this paper we are aiming to use a number of input features with various machine learning algorithms to improve classification accuracy with a high computational efficiency. The main known features from GPS raw measurements include C/N_0, HDOP, VDOP, satellite elevation angle, azimuth angle, pseudorange residual, pseudorange rate and number of visible satellites. This paper proposes a signal reception classifier based on the gradient boosting decision tree (GBDT) algorithm using 3 variables as features: C/N_0, pseudorange residuals, and satellite elevation angle. The reason for choosing these 3 features among the known features are that C/N_0, pseudorange residuals, and satellite elevation angle are highly related to the types of signal reception, and their ability to distinguish the signal reception type have been proved by previous studies [31–40]. For the other features, such as HDOP, VDOP and GDOP, represent the geometric distribution of visible satellites in a single epoch without a direct relationship to signal reception type. The pseudorange rate only represents an instantaneous change of the signal and therefore, has a weak relationship to signal reception type. The contributions of this paper are summarized as follows.

- A robust GBDT based GPS signal reception classification algorithm is proposed by using C/N_0, pseudorange residuals and satellite elevation angle as the input features to improve the performance of GPS signal reception classification.
- A sensitivity analysis process is designed and used to select the most appropriate iteration number of the algorithm to ensure the classification accuracy, while avoiding over fitting of the GBDT algorithm
- The raw GPS measurements are collected in various urban canyon environments to evaluate the performance of the designed algorithm

2. Algorithm design

The framework of the algorithm (Section 2.1) consists of the three functions of data labelling (Section 2.2), feature selection (Section 2.3) and the GBDT based signal classification algorithm (Section 2.4).

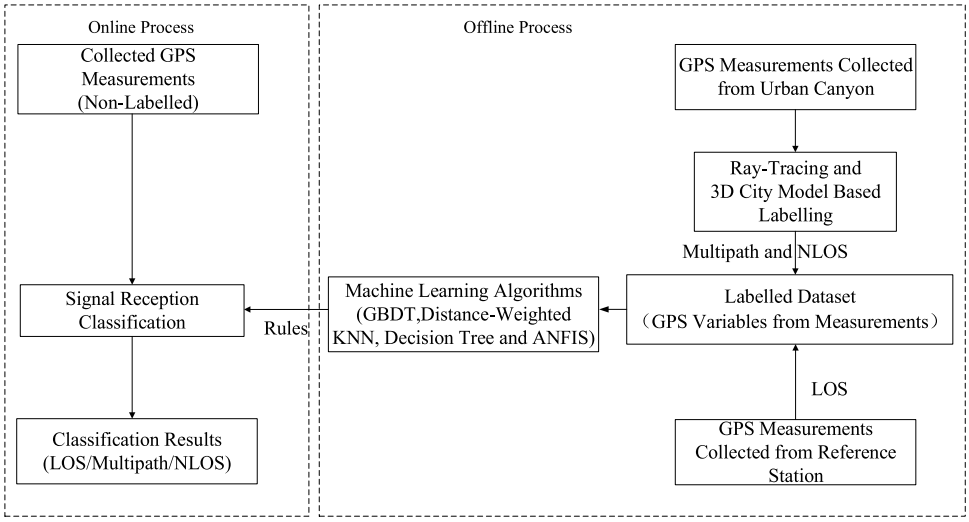

Fig. 1. Classification algorithm functions.

2.1. Algorithm framework

The process of the machine learning based LOS/Multipath/NLOS classification algorithm, including the offline and online parts, is presented in Fig. 1. For the offline process, the dataset containing a large number of LOS, NLOS and Multipath signals is used for training. The LOS signals are labelled and obtained from a GPS reference station located in an open area. With the aid of a 3D city model, GPS measurements collected from urban canyons are classified and labelled as either multipath or NLOS by using a ray-tracing algorithm. The five candidate machine learning algorithms are used separately to train the labelled training dataset for the extraction of classification rules. The extracted rules are then used in the classification of newly collected unlabelled GPS measurements in the online process.

2.2. Data labelling based on the 3D city model and ray-tracing

A 3D city model is an important information source for the algorithm. In this paper, the horizontal coordinates of the 3D city model were from the topographic map from the Land Department of the HK government with a resolution of a 20 cm. The height of the 3D city model was determined from Google map plus the height of the equipment.

With the building corner coordinates from the 3D city model, the ray-tracking method was applied to distinguish between the multipath and NLOS measurements in the datasets collected from urban canyons. The ray-tracing method uses known satellite, reflector and receiver geometry to trace the direct and reflected paths [42]. Satellite positions are obtained from the broadcast ephemeris. The position of the reflectors are derived from a 3D building model. The principle of ray-tracing is shown in Fig. 2. Suppose that A, B, C and D are four vertices of a building, and the corresponding position vectors are $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ respectively. Since Fig. 2 is a side view of the reflector, only $\boldsymbol{c}$ and $\boldsymbol{d}$ are marked in the picture. The normal vector $\boldsymbol{n}$ of the surface can be obtained by taking the cross product of two non-parallel vectors formed by the vertices of the surface. For example, using vertices A, B and C, $\boldsymbol{n}$ can be calculated by:

$$\boldsymbol{n} = (\boldsymbol{b} - \boldsymbol{a}) \times (\boldsymbol{c} - \boldsymbol{a}) \tag{1}$$

The steps of ray tracing are:

(1) Find the position $\boldsymbol{r}$ closest to the receiver antenna position $\boldsymbol{p}$ on the reflecting surface and calculate the position vector difference $\boldsymbol{r} - \boldsymbol{p}$ between the two points.

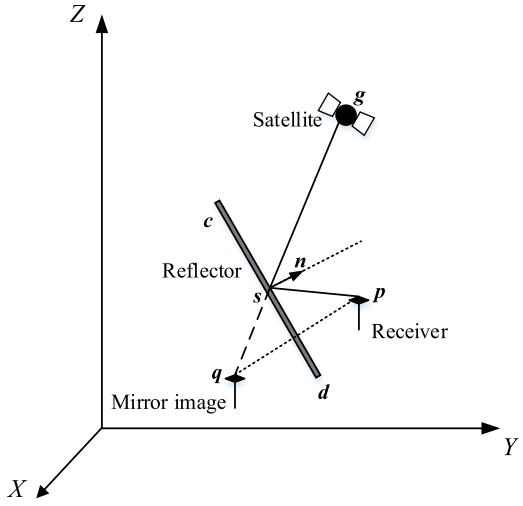

Fig. 2. Ray-tracing technique.

First calculate an intermediate vector $\boldsymbol{t_0}$ before calculating the position $\boldsymbol{r}$, as shown in the following equation:

$$\boldsymbol{t_0} = ((\boldsymbol{c} - \boldsymbol{p}) \cdot \boldsymbol{n}) / (\boldsymbol{n} \cdot \boldsymbol{n}) \tag{2}$$

$\boldsymbol{c}$ can be any point on the plane, here it is taken as one of the vertices on the plane. After finding the intermediate vector $\boldsymbol{t_0}$, calculate the position $\boldsymbol{r}$:

$$\boldsymbol{r} = \boldsymbol{p} + \boldsymbol{t_0 n} \tag{3}$$

(2) Find the mirror image position vector $\boldsymbol{q}$ of the receiver antenna position $\boldsymbol{p}$ relative to the reflection surface:

$$\boldsymbol{q} = \boldsymbol{p} + 2(\boldsymbol{r} - \boldsymbol{p}) \tag{4}$$

(3) Connect the satellite position $\boldsymbol{g}$ to the mirror image $\boldsymbol{q}$ of the antenna and find the intersection point $\boldsymbol{s}$ of the line segment and plane consisting of the reflection surface.

$$\boldsymbol{s} = \boldsymbol{g} + \boldsymbol{t}\,(\boldsymbol{q} - \boldsymbol{g}) \tag{5}$$

where $\boldsymbol{t}$ is an intermediate quantity expressed as:

$$\boldsymbol{t} = ((\boldsymbol{c} - \boldsymbol{g}) \cdot \boldsymbol{n}) / ((\boldsymbol{q} - \boldsymbol{g}) \cdot \boldsymbol{n}) \tag{6}$$

(4) Determine if the intersection is in the reflection surface. If the intersection is outside of the surface, prepare two line

segments to connect the point of reflection and the satellite, and the point of reflection and the receiver. If both line segments are not blocked by some other structure, they are considered as a reflected path. If the intersection is outside the surface, a reflected path does not exist. If the receiver only receives a reflected signal from a satellite, the measurements are labelled as NLOS. If the receiver receives both direct and reflected signals from a satellite, the measurements are labelled as multipath.

2.3. Feature selection

Currently, the signal features can be easily obtained from modern GNSS receivers. The features selected in this paper are: C/N_0, pseudorange residuals and satellite elevation angle.

(1) C/N_0: the signal strength is measured in terms of the C/N_0, which is the ratio of carrier power to noise power per unit of bandwidth in decibel-hertz (dB-Hz). Usually, the C/N_0 of an NLOS signal is smaller than that of a LOS signal. Therefore, it is the most commonly used feature. However, since both low and high C/N_0 values are possible for NLOS in urban canyons due to the different reflection surface materials, classification simply based on C/N_0 can be unreliable and therefore, necessitates the consideration of additional features.

(2) Pseudorange residuals, η: the pseudorange ρ is computed as the time ΔT from the signal being transmitted from the satellite to the detection of the signal in the receiver multiplied by the speed of light c plus the clock synchronization error t multiplied by the speed of light c, which can be expressed in (7):

$$\rho = \Delta T \times c + t \times c \tag{7}$$

The satellite position can be resolved from the broadcast ephemeris, so the positioning solution can be calculated by solving the pseudorange equations using least square estimation in (8):

$$r = \left(G^T G\right)^{-1} G^T \rho \tag{8}$$

where r is the receiver state, including the three-dimensional position and the receiver clock offset. G denotes the design matrix consisting of the unit LOS vector $(u_N^{(i)}, u_E^{(i)}, u_D^{(i)})$ between the satellite and receiver in (9):

$$G = \begin{bmatrix} u_N^{(1)} & u_E^{(1)} & u_D^{(1)} & -1 \\ u_N^{(2)} & u_E^{(2)} & u_D^{(2)} & -1 \\ \vdots & \vdots & \vdots & \vdots \\ u_N^{(i)} & u_E^{(i)} & u_D^{(i)} & -1 \end{bmatrix} \tag{9}$$

Once the positioning solution is calculated, the distance between the receiver and satellite can be obtained. The difference between this distance and the pseudorange is called the pseudorange residual, expressed as η, which can be calculated in (10):

$$\eta = \rho - G \cdot r \tag{10}$$

Pseudorange residuals are important for LOS/Multipath/NLOS signal reception classification [43]. In theory, the absolute value of pseudorange residuals and the probability of NLOS are positively related. This phenomenon becomes more obvious when only a small portion of the signals are NLOS [44]. Hsu et al. showed that the pseudorange residuals could be used as an indicator to classify signal reception type if the number of measurements is sufficient [35].

(3) Satellite elevation, θ: there is a significant positive correlation between satellite elevation and probability of LOS. In general, signals from satellites with higher elevation angles are less likely to be blocked or reflected by buildings. As the elevation decreases, however, there is a more likelihood of signal blockage by buildings and other obstacles. Elevation angle can, therefore, be used as a feature in signal reception classification. The satellite elevation angle θ can be calculated as (11):

$$\theta^{(i)} = -\arcsin(u_D^{(i)}) \tag{11}$$

2.4. Signal classification algorithm based on GBDT

GBDT is a supervised learning algorithm [45], also known as gradient boost regression tree (GBRT) and multiple additive regression tree (MART). It combines regression trees using a gradient boosting technique and has been widely applied in various disciplines, such as credit risk assessment [46], transport crash prediction [47] and fault prognosis in electronic circuits [48]. It replaces the difficult function minimization problem by using least-squares function minimization, followed by only a single parameter optimization based on the original criterion. Therefore, this advantage could potentially facilitate the achievement of high accuracy GPS signal reception classification [45].

In the designed GBDT based algorithm, each sample in the training set is represented as $\boldsymbol{x}_i = (C/N_{0_i}, \eta_i, \theta_i)$, where $i = 1, 2, 3, \ldots, N$ indicating the sequence number of the sample, and N is the number of samples. The labelled training dataset can be expressed as $T = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), (\boldsymbol{x}_3, y_3), \ldots, (\boldsymbol{x}_N, y_N)\}$, where $y_i \in \{-1, 0, 1\}$, is the label of each sample, -1, 0, 1 represent the NLOS, Multipath and LOS signals, respectively. GBDT minimizes the expected value of loss function $L(y_i, f(\boldsymbol{x}_i))$ by iteratively creating a weak learner $h_t(\boldsymbol{x}_i; \boldsymbol{a})$ that points in the steepest-descent direction, i.e., the negative gradient direction. The weak learner $h_t(\boldsymbol{x}_i; \boldsymbol{a})$ is a classification tree, the parameters $\boldsymbol{a}$ are the splitting variables, split locations and the terminal node means of the individual trees. The square loss function (12) is used in this paper:

$$L(y_i, f(\boldsymbol{x}_i)) = \frac{1}{2}(y_i - f(\boldsymbol{x}_i))^2 \tag{12}$$

The input to GBDT is the labelled training dataset T, with M as the number of iterations. The GBDT based GPS signal reception classification algorithm flow is as follows:

1. Initialize a weak learner $f_0(\boldsymbol{x})$ for the training data:

$$f_0(\boldsymbol{x}) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, \gamma) \tag{13}$$

$f_0(\boldsymbol{x})$ is a regression tree consisting of only one root node. Since L is selected to be the square loss function, $f_0(x)$ becomes:

$$f_0(\boldsymbol{x}) = \bar{y} \tag{14}$$

2. For m = 1 to M:

2.1 Compute the negative gradient

$$\tilde{y}_i = -\left[\frac{\partial L(y_i, f(\boldsymbol{x}_i))}{\partial f(\boldsymbol{x}_i)}\right]_{f(\boldsymbol{x}) = f_{m-1}(\boldsymbol{x})} \tag{15}$$

2.2 Replace the label y_i of the training dataset with $\tilde{y}_i$ to obtain a new dataset $T_m = \left\{(\boldsymbol{x}_1, \tilde{y}_1), (\boldsymbol{x}_2, \tilde{y}_2), (\boldsymbol{x}_3, \tilde{y}_3), \ldots, (\boldsymbol{x}_N, \tilde{y}_N)\right\}$, and create a new regression tree $h_m(\boldsymbol{x}_i; a_m)$ by training the new dataset T_m:

$$\boldsymbol{a}_t = \underset{\boldsymbol{a}}{\operatorname{argmin}} \sum_{i=1}^{N} (\tilde{y}_i - h_m(\boldsymbol{x}_i; \boldsymbol{a}))^2 \tag{16}$$

2.3 Update the strong learner:

$$f_m(\boldsymbol{x}) = f_{m-1}(\boldsymbol{x}) + \rho h_m(\boldsymbol{x}; \boldsymbol{a}_m) \tag{17}$$

where ρ is the learning rate, usually chosen to be a value between 0∼1 to prevent the overfitting.

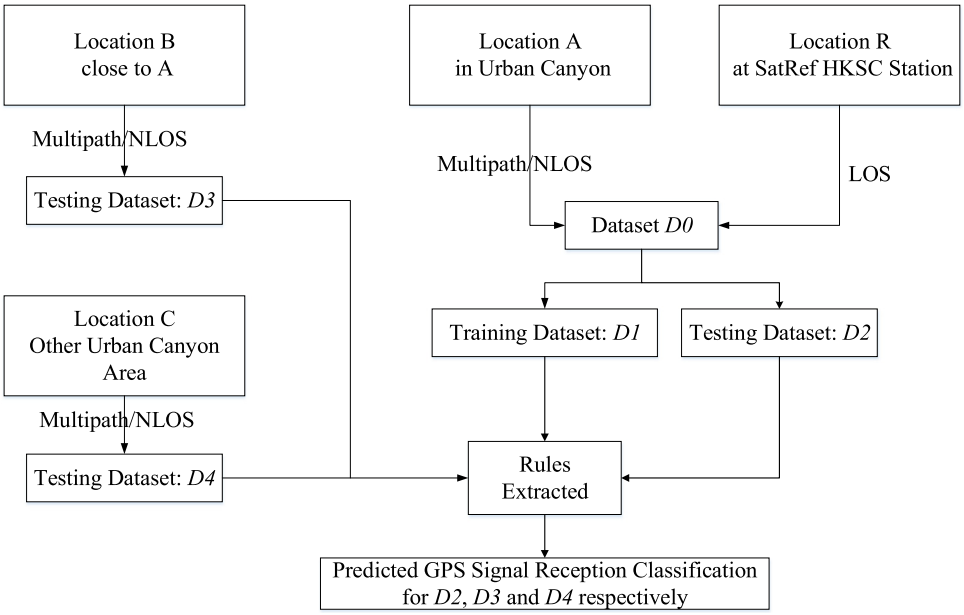

Fig. 3. Data collection and signal reception classification process.

3. After the iteration is terminated, output $f_M(x)$ as the final classifier:

$$f_M(\boldsymbol{x}) = f_0(\boldsymbol{x}) + \sum_{m=1}^{M} \rho h_m(\boldsymbol{x}; \boldsymbol{a}_m) \tag{18}$$

4. $f_M(\boldsymbol{x})$ is used to predict the signal reception type of the newly collected unlabelled sample $\boldsymbol{x} = (C/N_0, \eta, \theta)$ from the testing dataset. The predicted values need to be rounded to the closest value of 1, 0 or −1.

3. Field test and analysis of results

The proposed GPS signal reception classification algorithm based on the GBDT is compared with current classification algorithms, including decision tree, distance weighted k-nearest neighbour (KNN) and the adaptive network-based fuzzy inference system (ANFIS). Decision tree uses a tree structure learning as a predictive model to go from observations about an item to conclusions about the item's target value. Distance weighted KNN is an improved of KNN by heavier weighing of the close neighbours, according to their distances to the query. ANFIS integrates neural network (NN) training with fuzzy inference system (FIS), taking linguistic rules from human experts, and adapting itself by using the input–output data to obtain better training performance [32,41,49]. The Experiment is presented in Section 3.1 and the resulting data analysed in Section 3.2.

3.1. Experimental process

Five datasets were collected from four different locations (Fig. 3). Static GPS data were captured for a period of 24 h at an interval of 30 s at the SatRef HKSC reference station (Location R), Hong Kong. The data collected from the Location R was labelled as LOS. In addition, static GPS data were captured over 24 h at the same interval in a built-up area in Hung Hom (Location A) using a commercial GPS receiver, u-blox NEO-M8T, as shown in Fig. 4. The dataset from Location A contained predominantly multipath and NLOS measurements. Additional static GPS data were collected from different urban canyon environments (Location B and Location C). Since Locations B and C

Fig. 4. GPS data collection environment in a built-up area.

were both in the urban canyon, it was assumed that only NLOS and multipath signals were present. The labelling of NLOS and multipath was also obtained by using a 3D city model and ray-tracing. In one case, however, it was difficult to obtain labelled data due to the limitations of the 3D city model used with the ray-tracing-technique.

Dataset *D0* was created by combining the data collected from Location A and Location R. *D0* was subsequently used to generate two datasets: training set *D1* and testing set *D2*. In order to prevent bias in the training results due to uneven sample distribution, an equal number of LOS, NLOS, and Multipath samples were randomly selected from *D0* to form the training dataset *D1*, which contained 24,000 samples. Thus, in *D*1, the labelling distribution for each of LOS, NLOS, and multipath was one third. Only 24,000 of the 96,992 samples were selected for training, to reduce the computational load and prevent overfitting. There were 18 164 multipath samples in *D0* and 8000 were used in the training dataset. Therefore, 8000 samples of each type were randomly selected from the remainder of *D0* (i.e. excluding *D1*) to form the testing dataset *D2* in order to evenly distribute the

Table 1
Summary of the datasets.

Dataset	*D0*	*D1*	*D2*	*D3*	*D4*
Total samples	96 992	24 000	24 000	11 615	25 039
LOS (labelled as 1)	25 987	8 000	8 000	0	0
Multipath (labelled as 0)	18 164	8 000	8 000	3 114	8 831
NLOS (labelled as −1)	52 841	8 000	8 000	8 501	16 208

three types of samples in the dataset. Thus, the labelling distribution of *D2* for each of LOS, NLOS, and multipath was also one third, but with the labels removed. Each measurement in *D1* and *D2* contained the associated three features of: C/N_0, pseudorange residual and satellite elevation angle. Although some of the features could be correlated over time, in this case, the time dependency of the data was not considered when applying GBDT algorithm.

We used training dataset *D1* with each of the four machine learning algorithms to determine classification rules. The classification rules extracted by the machine learning algorithms were then used to classify the samples in *D2*. These were then compared to the classification results with the reference label for each sample (i.e. the previously removed labels), to calculate the accuracy of the algorithm. To verify the validity of the extracted rules, two more testing datasets collected from other locations were used to feed the rules. Testing dataset *D3* was collected from Location B, close to location A in the urban canyon, while the other testing dataset, *D4*, was from Location C, about three blocks away from location A in the urban canyon. A summary of the datasets is presented in Table 1.

3.2. Results and analysis

The GBDT algorithm was used to train dataset *D1* to determine classification rules, subsequently used to classify the testing datasets to determine classification accuracy. The results were then compared to those from the three other machine learning algorithms. The candidate machine learning algorithms could be used with either single or multiple signal features. To determine the benefit of the additional features, a comparison was made between classification using C/N_0 only and the multiples features of C/N_0, pseudorange residuals and satellite elevation angle.

The confusion matrix of the LOS, multipath and NLOS (1, 0 and −1) classification results for the different algorithms using single feature-based classification (i.e. C/N_0), by using the testing dataset *D2* are compared in Table 2. The accuracy in the table represents the ratio (in percentage) of the number of samples correctly classified to the total number of samples in the dataset. The accuracy of each category refers to the ratio (in percentage) of the number of samples correctly classified to the total number of samples known as being in this category. For example, the NLOS detection accuracy is calculated as the ratio (in percentage) of the number of samples correctly classified as NLOS to the number of total known NLOS samples. For single feature (C/N_0) based classification, the classification accuracies of the four algorithms for multipath and NLOS, are consistently below 80% (ranging from 29% to 79.6%). The classification accuracy of LOS signals is higher ranging from 95.1% to 99.9%.

As discussed earlier, additional features could be used to improve accuracy the traditional single feature, C/N_0 based classification. The scatter diagram in Fig. 5 shows the relationships between the input features (i.e. C/N_0, satellite elevation angle and pseudorange residuals) and their corresponding labelled signal reception types in the dataset *D0*. The green points denoting the LOS signals are concentrated in high C/N_0 areas, while the NLOS and multipath samples are mainly distributed in medium and low

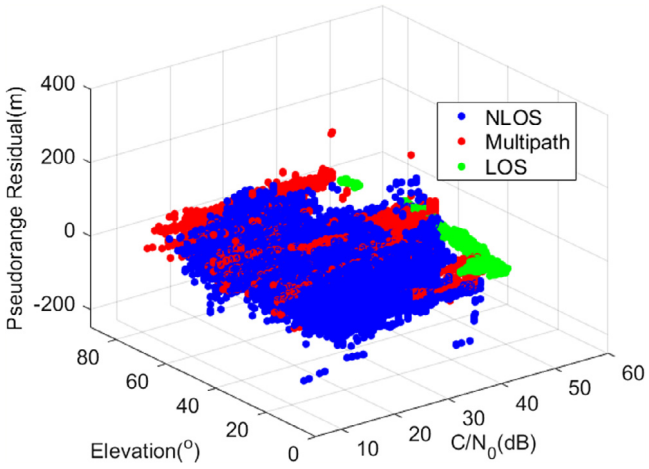

Fig. 5. Relationships between the features (i.e. C/N_0, satellite elevation angles and pseudorange residuals) and the corresponding labelled signal reception types in *D0*.

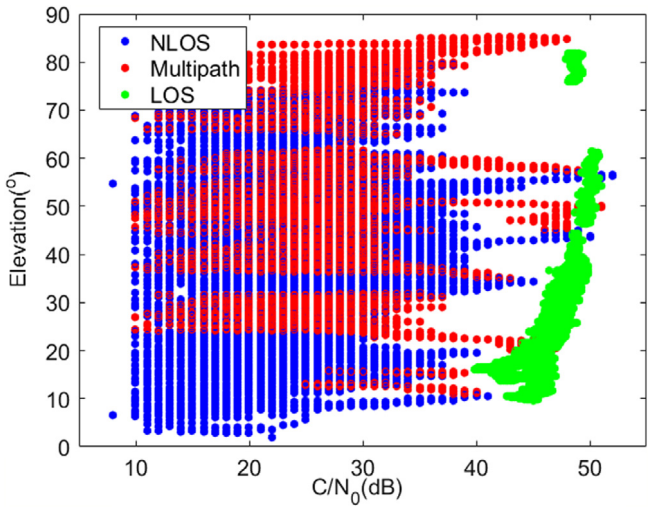

Fig. 6. Relationship between the features (i.e. C/N_0 and satellite elevation angles) and the corresponding labelled signal reception types in *D0*.

C/N_0 areas. The large overlaps makes it difficult to distinguish NLOS from multipath signals by using C/N_0 only.

Fig. 6 shows the relationship between the features (i.e. C/N_0 and satellite elevation angles) and the corresponding labelled signal reception types in dataset *D0*. It can be seen that lower elevation angles are dominated by NLOS signals, with most of their C/N_0 values less than 45dB-Hz. Multipath accounts for the majority of the samples with elevation angles higher than 25 degrees.

Fig. 7 shows the relationship between features (i.e. C/N_0 and pseudorange residuals) and the corresponding labelled signal reception types in dataset *D0*. The pseudorange residuals of LOS signals range from −2 m to 2 m with a mean value close to zero. The pseudorange residuals for multipath signals largely range from −100 m to 100 m, while for NLOS signals the residuals are always over 100 m.

From the analysis of the features and their corresponding labelled signal reception types, consideration of the additional features (satellite elevation angle and pseudorange residuals) has the potential to improve the accuracy of the GPS signal reception classification.

Table 2
Confusion matrix of LOS (Noted as 1), Multipath (0) and NLOS (−1) classification results using different algorithms based on a single feature (C/N_0) for Testing Dataset *D2*.

Algorithms		GBDT			Decision tree		
Label ID		−1	0	1	−1	0	1
No of samples	−1	6365	1635	0	3729	4018	253
	0	4197	3803	0	1836	4611	1553
	1	0	391	7609	0	182	7818
Accuracy (%)			74.1			67.3	
Category accuracy (%)		79.7	47.5	95.1	46.6	57.6	97.7
Algorithms		Distance-weighted KNN			ANFIS		
Label ID		−1	0	1	−1	0	1
No. of samples	−1	15	7735	250	6054	1691	255
	0	0	6170	1830	3934	2329	1737
	1	0	22	7978	0	365	7635
Accuracy (%)			59			66.7	
Category accuracy (%)		69.4	29	99.9	75.7	29.1	95.4

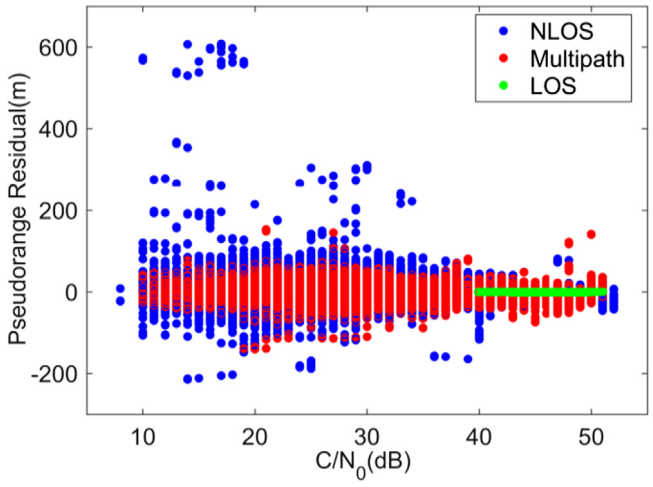

Fig. 7. Relationships between features (i.e. C/N_0 and pseudorange residuals) and the corresponding labelled signal reception types in dataset *D0*.

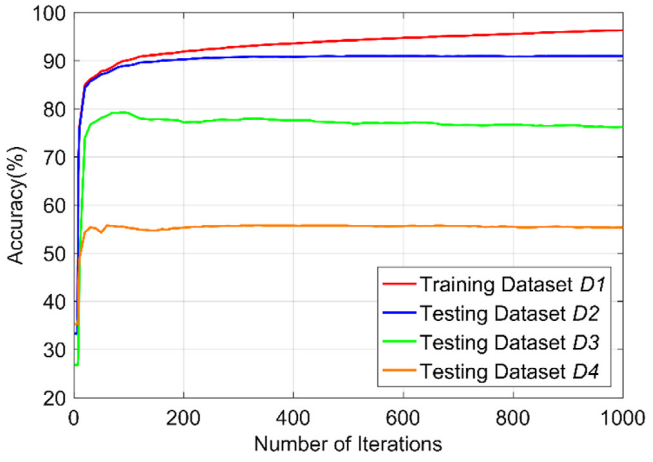

Fig. 8. Relationship between the number of iterations and the GBDT classification accuracy.

The selection of an appropriate number of iterations, i.e. the number of regression trees, is very important for the GBDT algorithm. If the number of iterations is too small, the trained classification rules cannot fully determine the relationships between input features and signal reception types. On the other hand, if there are too many iterations, the results will be prone to overfitting and the computational load is increased. Therefore, in this paper, before comparing GBDT with other algorithms, a sensitivity analysis of the relationship between the number of iterations and classification accuracy of each dataset is carried out. The classification rules determined from different numbers of training iterations with training dataset *D1* are used to classify the signal reception types of the dataset *D1* for internal validation (self-consistency check) and testing sets *D2*, *D3*, and *D4* for external validation (testing). The results are presented in Fig. 8.

Overall the classification accuracy increases with the number of iterations. The classification accuracy of *D2* has been increased greatly as it is from the same data source as the dataset *D1*. This is the reason why the classification rules based on the large number of iterations obtained from *D1* are applicable to *D2*. For the testing datasets *D3* and *D4* from different locations compared to *D1*, the classification accuracy increases initially with the increasing iteration numbers but then starts to decline when the iteration number reaches a certain value (i.e. 100) due to overfitting of the algorithm. Although the more the number of iterations the better the fitting of the extracted rules for the training dataset, it affects the adaptively of classification accuracy for the additional datasets. Therefore, in order to keep a high classification accuracy without overfitting, the number of iterations is set to 100 in this paper. In addition, the classification accuracies of GBDT for datasets *D3* and *D4* are always lower than that of *D1* and *D2*, reflecting the sensitivity of classification rules to different locations.

A comparison of the classification results for different algorithms using C/N_0, pseudorange residuals and satellite elevation angle based on the testing dataset *D2*, is presented in Table 3. The classification accuracy and training time of these algorithms are shown in Figs. 9 and 10, respectively. By comparing the results from Tables 2 and 3, it is obvious that the classification accuracies of multiple feature-based algorithms are higher than that of single feature (C/N_0) based algorithms. In addition, Table 3 shows that the overall classification accuracy of the GBDT algorithm with the three features 89%, better than that with decision tree and ANFIS, and marginally higher than that with distance-weighted KNN. Although the NLOS classification accuracy of the GBDT algorithm is lower than that of traditional decision tree, the latter has a higher number of missed detections, with multipath signals identified as NLOS. In the iterative process, GBDT focuses on the samples with large training residuals (i.e. samples that are difficult to classify). The final learner is the fusion of multiple regression trees created in each iteration with overfitting mitigated to some extent by setting the weight of each regression tree based on the learning rate. For the decision tree, the training dataset can

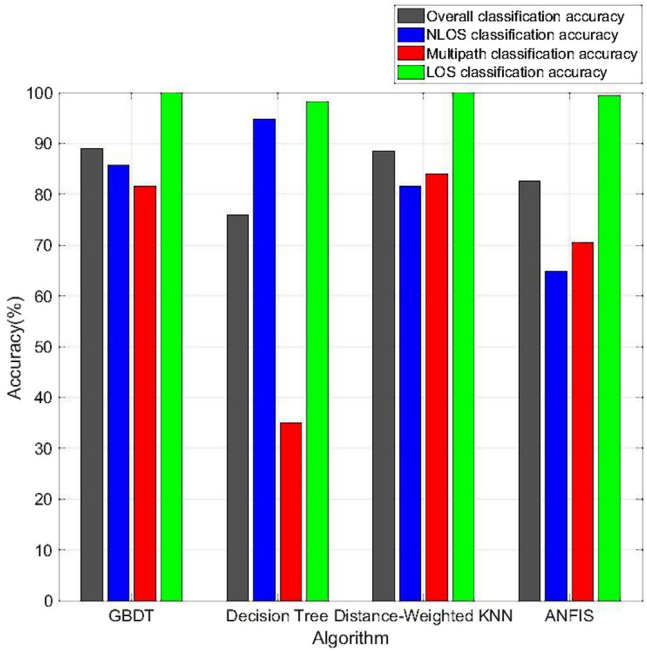

Fig. 9. Classification accuracy for different algorithms using multiple features (C/N_0, Pseudorange Residuals and Satellite Elevation Angle) of the testing dataset *D2*.

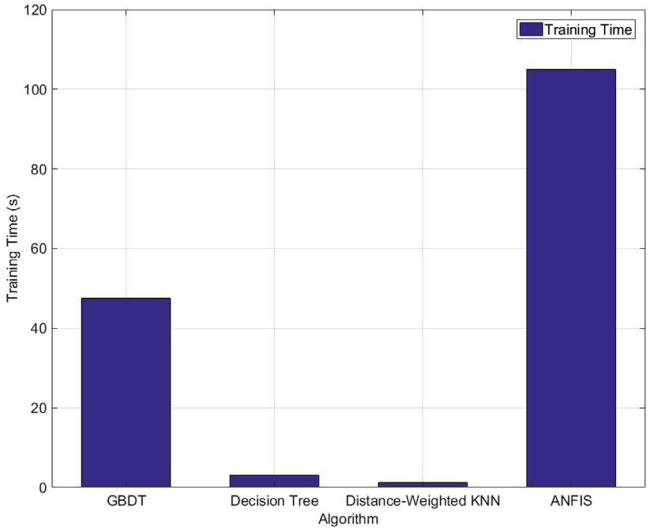

Fig. 10. Training time for different algorithms using multiple features (C/N_0, Pseudorange Residuals and Satellite Elevation Angle) of the testing dataset *D2*.

be better fitted by increasing the number of leaf nodes. However, as the complexity of the tree increases, overfitting is more likely to occur and leads to more misclassification. Therefore, the GBDT composed of multiple regression trees is better than a single complex decision tree. Although the classification accuracy of weighted KNN is close to GBDT and with a lower training time consumption, the classification rules from the KNN-algorithm are less adaptable to other datasets. The classification accuracy of ANFIS is over 80% but with a high misclassification rate between the multipath and NLOS measurements. The ANFIS based algorithm also has the longest training time among all of the candidate algorithms.

Dataset *D3* and dataset *D4* collected from location B and location C in different urban canyon areas are further used to verify the validity of the extracted classification rules. The classification accuracies for the testing datasets *D3* and *D4* are compared in Tables 4 and 5.

The classification accuracy performance based on the testing dataset *D*3, from a similar urban environment to the training data, is illustrated in Table 4. The GBDT based algorithm has an overall classification accuracy of 77%, which is higher than that of distance-weighted KNN (i.e. 68%) and ANFIS (i.e. 71.5%). Although the decision tree-based algorithm has a relative higher overall classification accuracy (i.e. 86.1%), the classification accuracy for the multipath is low (i.e. 54%). More multipath signals are misclassified as NLOS signals by using the decision tree.

The performance based on the testing dataset *D*4, from a different urban environment to the training data, is illustrated in Table 5. The overall classification accuracy for all the candidate algorithms, including the GBDT, are around 55%–60%, reflecting the data sensitivity of the machine learning algorithms.

We have further evaluated the static positioning results based on the elimination of the NLOS signals detected from the proposed GBDT with multi-feature-based method and single C/N_0 based method. Here, the threshold of single C/N_0 is also determined by GBDT, which has a higher classification results compared to the other algorithms in Table 2. The comparisons of the static positioning accuracy Root Mean Square Error (RMSE) of the dataset *D*2, *D*3 and D4 from Location A, B and C, respectively, are shown in Table 6 and the comparisons of positioning results for the three locations are shown in Figs. 11, 12 and 13, respectively.

For the Location A, where the urban training dataset is collected, the 2D and 3D RMSE of the positioning accuracy for the proposed method are 31.23 m and 59.82 m, 30.1% and 34.1% higher than that of the single feature C/N_0 based NLOS elimination method. As shown in Fig. 11, although positioning results for both algorithms failed to cover ground truth due to the low measurement source, compared with the single feature C/N_0 based NLOS elimination method, the positioning results of the proposed method are more concentrated and closer to the ground truth. For the Location B, which is not far away from location A, the 2D and 3D RMSE of the positioning results from the proposed method are 44.06 m and 64.77 m, 19.4% and 19.8% higher than that of the single feature C/N_0 based NLOS elimination method. The positioning results of the proposed method are closer to the ground truth compared to the single C/N_0 based method as depicted in Fig. 12. For the Location C, from a different urban canyon environment, the positioning accuracy improvement from the GBDT with multi-feature based method is limited, with an improvement of 2.2% for the 2D and an deterioration of 5.9% for the 3D positioning results, compared to the single C/N_0 based method due to the similar low performance of the signal reception classification (i.e. about 55% for multi-feature and 57% for single feature). It is indicated from the analysis results that for the Location A and Location B, the proposed method can effectively classify the signal reception types, and therefore improve the positioning results after removal of the NLOS signal detected. While for the Location C, the positioning accuracy could not be improved due to the poor classification performance from the proposed algorithm. It is validated that the higher accuracy of the signal reception classification could result in more improvement of the final GPS positioning results.

To be noted that the positioning based on the NLOS exclusion could result in a 20% to 35% improvement based on proposed algorithm for the similar urban environments. It is potential to further improve the positioning results with the correction of the detected NLOS signals used in the positioning process. The labelling error caused by the inaccuracy borders of the 3D map is also a reason to affect the final positioning accuracy. Therefore, the positioning accuracy could be further improved with the improved high-definition 3D map.

Table 3
Confusion matrix of LOS (1), Multipath (0) and NLOS (−1) classification results using different algorithms based on multiple features (C/N_0, Pseudorange Residuals and Satellite Elevation Angle) for Testing Dataset *D2*.

Algorithms		GBDT			Decision tree		
Label ID		−1	0	1	−1	0	1
No of samples	−1	6858	1134	8	7583	395	22
	0	1322	6522	156	5006	2795	199
	1	0	14	7986	133	5	7862
Accuracy (%)			89			76	
Training Time (s)			47.6			3.1	
Category accuracy (%)		85.7	81.5	99.8	94.8	34.9	98.3
Algorithms		Distance-weighted KNN			ANFIS		
Label ID		−1	0	1	−1	0	1
No. of samples	−1	6521	1472	7	6256	1716	28
	0	1266	6724	10	2237	5634	129
	1	0	1	7999	0	47	7953
Accuracy (%)			88.5			82.7	
Training time (s)			1.2			105	
Category accuracy (%)		81.5	84.1	100	64.9	70.4	99.4

Table 4
Confusion matrix of LOS (1), Multipath (0) and NLOS (−1) classification results using different algorithms based on multiple features (C/N_0, Pseudorange Residuals and Satellite Elevation Angle) for Testing Dataset *D3*.

Algorithms		GBDT			Decision tree		
Label ID		−1	0	1	−1	0	1
No of samples	−1	6239	2262	0	8320	168	13
	0	299	2726	89	1321	1681	112
	1	0	0	0	0	0	0
Accuracy (%)			77.2			86.1	
Category accuracy (%)		73.4	87.5		97.9	54	
Algorithms		Distance-weighted KNN			ANFIS		
Label ID		−1	0	1	−1	0	1
No. of samples	−1	6217	2284	0	7047	1436	18
	0	1024	1687	403	868	1252	994
	1	0	0	0	0	0	0
Accuracy (%)			68			71.5	
Category accuracy (%)		73.1	54.2		82.9	40.2	

Table 5
Confusion matrix of LOS (1), Multipath (0) and NLOS (−1) classification results using different algorithms based on multiple features (C/N_0, Pseudorange Residuals and Satellite Elevation Angle) for Testing Dataset *D4*.

Algorithms		GBDT			Decision tree		
Label ID		−1	0	1	−1	0	1
No of samples	−1	8830	7378	0	14 485	1723	22
	0	3694	5025	112	7 782	851	198
	1	0	0	0	0	0	0
Accuracy (%)			55.3			61.3	
Category accuracy (%)		54.6	56.6		89.4	9.6	
Algorithms		Distance-weighted KNN			ANFIS		
Label ID		−1	0	1	−1	0	1
No. of samples	−1	9217	6991	0	9385	6663	160
	0	2970	5775	86	3141	5374	316
	1	0	0	0	0	0	0
Accuracy (%)			60			59	
Category accuracy (%)		56.9	65.4		57.9	60.9	

In summary, the classification rules extracted from the GBDT algorithm are applicable to environments with largely similar spatial and material characteristics (i.e. testing dataset D2 and D3) but with low adaptively to datasets with different characteristics (i.e. testing dataset D4). Further work is exploring further the issue of adaptively and the development of real-time on-line training algorithms.

4. Conclusions and future work

This paper has presented a GBDT based algorithm, using C/N_0, pseudorange residual and elevation angle as the features, to classify GPS signal reception as LOS, multipath and NLOS, meanwhile, the static positioning solutions are also calculated with the detected NLOS eliminations. The signal reception classification results of the testing dataset *D2* (from Location A), from the

Table 6
Comparisons of RMSE for Location A, B and C.

RMSE (m)		E	N	U	3D	2D
Location A	C/N_0 Based NLOS Elimination	40.92	17.9	79.01	90.76	44.67
	GBDT with Multi-Feature-Based NLOS Elimination	26.19	17.02	51.02	59.82	31.23
	Improvement (%)	36.0	4.9	35.4	34.1	30.1
Location B	C/N_0 Based NLOS Elimination	20.13	45.41	63.72	80.80	49.67
	GBDT with Multi-Feature-Based NLOS Elimination	18.35	35.61	50.89	64.77	40.06
	Improvement (%)	8.8	21.6	20.1	19.8	19.4
Location C	C/N_0 Based NLOS Elimination	25.4	29.5	127.67	133.37	38.59
	GBDT with Multi-Feature-Based NLOS Elimination	25.07	32.27	123.83	130.39	40.86
	Improvement (%)	1.3	−9.4	3.0	2.2	−5.9

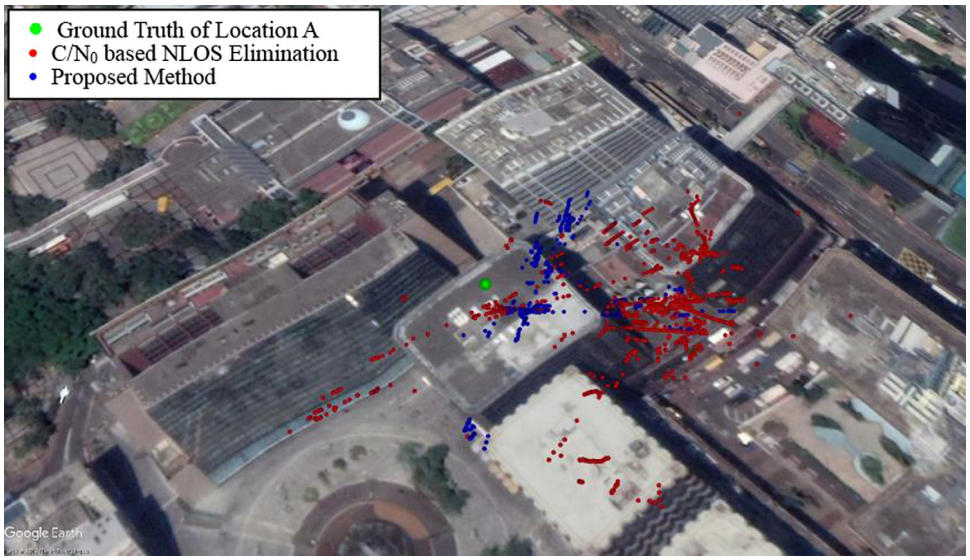

Fig. 11. Positioning results based on the elimination of the NLOS signals detected from the proposed GBDT with multi-feature-based method and single C/N_0 based method in Location A.

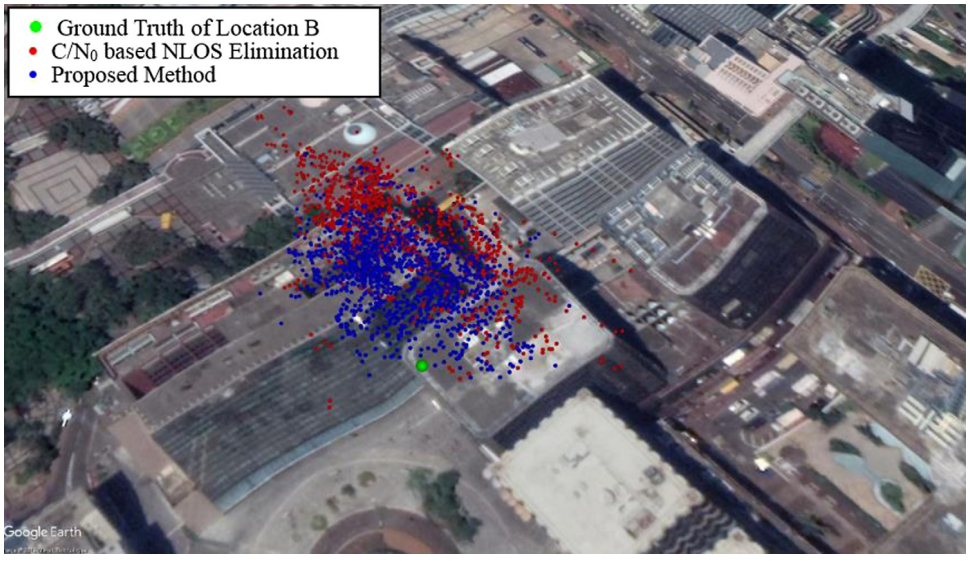

Fig. 12. Positioning results based on the elimination of the NLOS signals detected from the proposed GBDT with multi-feature-based method and single C/N_0 based method in Location B.

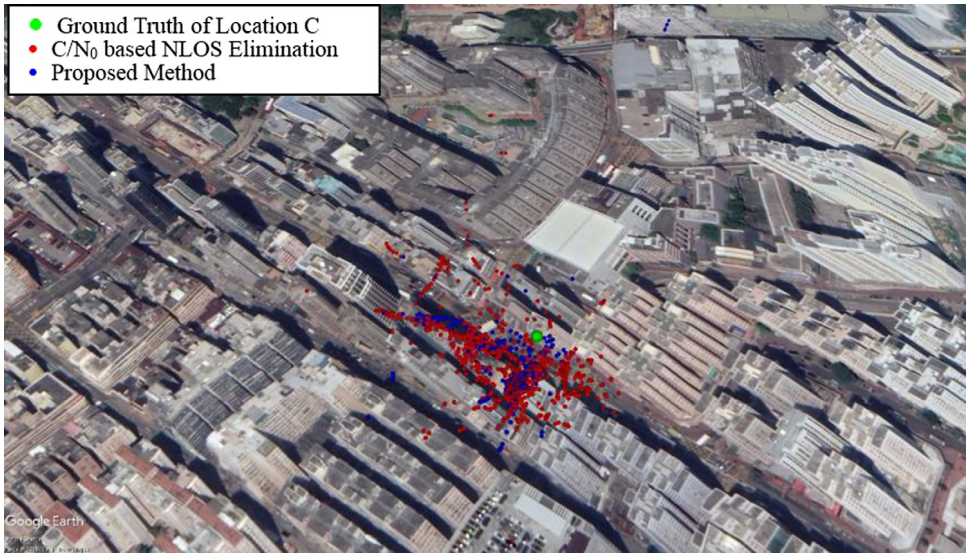

Fig. 13. Positioning results based on the elimination of the NLOS signals detected from the proposed GBDT with multi-feature-based method and single C/N_0 based method in Location C.

same environment as the urban training dataset, have shown that the overall accuracy of the multi-feature-based classification algorithm (i.e. 89% for the static data) is much higher than that of the single-feature-based C/N_0 classification algorithm (i.e.74.1%). Furthermore, for the correct classification of signals in the categories of NLOS and multipath, the local accuracies are 85.7% and 81.5% respectively, better than the decision tree, distance-weighted KNN and ANFIS algorithms. For the testing dataset *D3* (from Location B), from the similar environment as the training dataset, the algorithm achieves an overall classification accuracy of 77.2%. In particular, the detection accuracies are 73.4% and 87.5% for the NLOS and multipath respectively, better than the decision tree, distance-weighted KNN and ANFIS algorithms. The computation time for the GBDT is higher than the decision tree and distance-weighted KNN, however, this could be resolved by higher computing processing power. Overall, considering computation time and classification accuracy, it is shown that GBDT is the best of the algorithms investigated, for GPS signal reception classification. It should be noted that, for some datasets with different characteristics from the training dataset, such as *D4* (from Location B), the classification performance is degraded due to the inapplicability of the rules extracted from the training environment.

Based on the GBDT with multi-feature-based classification results, the application for the static positioning results are further analysed with the detected NLOS elimination. For the Location A, a positioning accuracy improvement of 34.1% (3D RMSE) has been achieved compared to the single C/N_0 based method. For the Location B, the proposed method could also provide an improvement of the positioning accuracy with 19.8% (3D RMSE), lower than that for the Location A. While for the Location C, the proposed method is unable to improve the positioning accuracy due to the different spatial and material characteristics. Therefore, environmental sensitivity is a key issue in the application of classification algorithms. This could be addressed by developing spatio-temporally dynamic algorithms, consideration of more signal related features and training data from a number of different locations in future work.

For static positioning, the proposed method can be used to detect the NLOS and multipath signals, which could then be used in data pre-processing. The experimental results show that removing NLOS based on the proposed method can improve the positioning accuracy to some extent. However, simple elimination cannot meet the requirements of high-precision positioning. The proposed method is a complement to the existing positioning methods based on 3D city model such as shadow matching. In future studies, we will combine the proposed method with 3D map to achieve better static positioning accuracy and therefore, could be used for the civil engineering applications such as building maintenance, and structural integrity/deformation monitoring in the urban canyons. For dynamic positioning, research is ongoing based on a framework of grid of reference points from which data will be captured for training. Users will then automatically obtain the classification rules of nearby reference points for accurate satellite signal reception type classification, to improve positioning. In addition, we will develop online data training for use with the GBDT algorithm for real-time applications such as ground vehicles, pedestrians and unmanned aerial vehicles (UAVs).

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2019.105942.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 41704022, No. 41974033), Natural Science Foundation of Jiangsu Province, China (Grant No. BK20170780), China Postdoctoral Science Foundation funded Project (Grant No. 2017M623360) and Foundation of Graduate Innovation Center in NUAA, China (Grant No. KFJJ20180719) and Specialized Research Fund for Shandong Provincial Key Laboratory, China (Grant No. KLWH201813).

References

[1] P. Misra, P. Enge, GPS Measurements and error sources, in: Global Positioning System: Signals, Measurements and Performance, second ed., Ganga-Jamuna Press, Lincoln, MA, USA, 2011, pp. 174–177.

[2] P.W. Ward, J.W. Betz, C.J. Hegarty, Interference, multipath, and scintillation, in: Understanding GPS: Principles and Applications, second ed., Artech House, Norwood, MA, USA, 2006, pp. 279–292.

[3] G. MacGougan, G. Lachapelle, R. Klukas, K. Siu, L. Garin, J. Shewfelt, G. Cox, Performance analysis of a stand-alone high-sensitivity receiver, GPS Solut. 6 (3) (2002) 179–195, http://dx.doi.org/10.1007/s10291-002-0029-z.
[4] J.M. Tranquilla, J.P. Carr, H.M. Al-Rizzo, Analysis of a choke ring ground-plane for multipath control in global positioning system (GPS) applications, IEEE Trans. Antennas Propag. 42 (7) (1994) 905–911, http://dx.doi.org/10.1109/8.299591.
[5] Z. Jiang, P.D. Groves, NLOS GPS Signal detection using a dual-polarisation antenna, GPS Solut. 18 (1) (2014) 15–26, http://dx.doi.org/10.1007/s10291-012-0305-5.
[6] H. Liang, T. Walter, P. Enge, and, G.X. Gao, GNSS Multipath and jamming mitigation using high-mask-angle antennas and multiple constellations, IEEE Trans. Intell. Transp. Syst. 16 (2) (2015) 741–750, http://dx.doi.org/10.1109/tits.2014.2342200.
[7] V.A. Dierendonck, P. Fenton, T. Ford, Theory and performance of narrow correlator spacing in a GPS receiver, Navigation 39 (3) (1992) 265–283, http://dx.doi.org/10.1002/j.2161-4296.1992.tb02276.x.
[8] R.D.J.V. Nee, J. Siereveld, P.C. Fenton, B.R. Townsend, The multipath estimating delay lock loop: approaching theoretical accuracy limits, in: IEEE/ION Position, Location and Navigation Symposium, Las Vegas, NV, USA, 1994, pp. 246-251, http://dx.doi.org/10.1109/ISSSTA.1992.665623.
[9] L. Garin, J.M. Rousseau, Enhanced strobe correlator multipath rejection for code & carrier, in: ION GPS, Kansas, MO, USA, 1997, pp. 559-568.
[10] B.R. Townsend, P.C. Fenton, A practical approach to the reduction of pseudorange multipath errors in a Ll GPS receiver, in: International Technical Meeting of the Satellite Division of the Institute of Navigation, Salt Lake City, UT, USA, 1994, pp. 143-148.
[11] M.S. Braasch, Performance comparison of multipath mitigating receiver architectures, in: IEEE Aerospace Conference, Big Sky, MT, USA, 2001, pp. 3/1309-3/1315.
[12] B.R. Townsend, P.C. Fenton, K.J.V. Dierendonck, R.D.J.V. Nee, Performance evaluation of the multipath estimating delay lock loop, Navigation 42 (3) (1995) 502–514, http://dx.doi.org/10.1002/j.2161-4296.1995.tb01903.x.
[13] P.C. Fenton, J. Jones, The theory and performance of NovAtel Inc.'s vision correlator, in: ION GNSS, Long Beach, CA, USA, 2005, pp. 2178-2186.
[14] F. Wang, J. Chen, S. Gao, K. Tang, X. Meng, Development and sea trial of real-time offshore pipeline installation monitoring system, Ocean Eng. 146 (2017) 468–476, http://dx.doi.org/10.1016/j.oceaneng.2017.09.016.
[15] M.G. Petovello, C. O'Driscoll, G. Lachapelle, Weak signal carrier tracking of weak using coherent integration with an ultra-tight GNSS/IMU receiver, in: presented at the European Navigation Conference, Toulouse, Lauragais, France, Apr. 23-25, 2008.
[16] M.G. Petovello, G. Lachapelle, Comparison of vector-based software receiver implementations with application to ultra-tight GPS/INS integration, in: ION GNSS, Fort Worth, Texas, 2006, pp. 2977-2989.
[17] A. Soloviev, C. Toth, D. Grejner-Brzezinska, Performance of deeply integrated GPS/ins in dense forestry areas, J. Appl. Geodesy 6 (1) (2012) 3–13, http://dx.doi.org/10.1515/jag-2011-0005.
[18] A. Soloviev, 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. Top. Sign. Proces. 3 (5) (2009) 786–797, http://dx.doi.org/10.1109/JSTSP.2009.2027796.
[19] V. Milanes, J.E. Naranjo, C. Gonzalez, J.A. Ruiz, T. Pedro, Autonomous vehicle based in cooperative GPS and inertial systems, Robotica 26 (5) (2008) 627–633, http://dx.doi.org/10.1017/S0263574708004232.
[20] S. Zhao, Y. Chen, J.A. Farrell, High-precision vehicle navigation in urban environments using an mem's IMU and single-frequency GPS receiver, IEEE Trans. Intell. Transp. Syst. 17 (10) (2016) 2854–2867, http://dx.doi.org/10.1109/TITS.2016.2529000.
[21] K.W. Chiang, Y.W. Huang, An intelligent navigator for seamless INS/GPS integrated land vehicle navigation applications, Appl. Soft Comput. 8 (1) (2008) 722–733, http://dx.doi.org/10.1016/j.asoc.2007.05.010.
[22] J.I. Meguro, T. Murata, J.I. Takiguchi, Y. Amano, T. Hashizume, GPS Multipath mitigation for urban area using omnidirectional infrared camera, IEEE Trans. Intell. Transp. Syst. 10 (1) (2009) http://dx.doi.org/10.1109/tits.2008.2011688.
[23] S. Peyraud, D. Bétaille, S. Renault, M. Ortiz, F. Mougel, D. Meizel, F. Peyret, About non-line-of-sight satellite detection and exclusion in a 3D map-aided localization algorithm, Sensors 13 (1) (2013) 829–847, http://dx.doi.org/10.3390/s130100829.
[24] F. Peyret, D. Betaille, M. Ortiz, S. Miquel, L. Fontenay, How to improve GNSS positioning quality of service for demanding ITS in city environments by using 3D digital maps, in: presented at the 19th ITS World Congress, Vienna, Austria, Oct. 22-26, 2012.
[25] P.D. Groves, Shadow matching: a new GNSS positioning technique for urban canyons, J. Navig. 63 (3) (2011) 417–430, http://dx.doi.org/10.1017/S0373463311000087.
[26] L. Wang, P. Groves, M. Ziebart, Urban positioning on a smartphone: real-time shadow matching using GNSS and 3D city models, in: ION GNSS, Nashville, Tennessee, USA, 2013, pp. 1606-1619.
[27] L. Wang, P.D. Groves, M.K. Ziebart, GNSS Shadow matching: improving urban positioning accuracy using a 3D city model with optimized visibility scoring scheme, Navigation 60 (3) (2013) 195–207, http://dx.doi.org/10.1002/navi.38.
[28] S. Miura, L.T. Hsu, F. Chen, S. Kamijo, GPS Error correction with pseudorange evaluation using three-dimensional maps, IEEE Trans. Intell. Transp. Syst. 16 (6) (2015) 3104–3115, http://dx.doi.org/10.1109/TITS.2015.2432122.
[29] L.T. Hsu, Y. Gu, S. Kamijo, 3D Building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation, GPS Solut. 20 (3) (2016) 413–428, http://dx.doi.org/10.1007/s10291-015-0451-7.
[30] Y. Gu, L.T. Hsu, 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. 65 (6) (2016) 4274–4287, http://dx.doi.org/10.1109/TVT.2015.2497001.
[31] R. Yozevitch, B.B. Moshe, H. Levy, Breaking the 1meter accuracy bound in commercial GNSS devices, in: 27th Convention of the Electrical and Electronics Engineers in Israel (IEEEI), Eilat, Israel, 2012, pp. 1-5.
[32] R. Yozevitch, B.B. Moshe, A. Weissman, A robust GNSS LOS/NLOS signal classifier, Navigation 63 (4pp) (2016) 429–442.
[33] G. Deng, GPS Satellite selection algorithm based on satellite elevation angle and GDOP, Digit. Commun. 37 (2) (2010) 47–50, (in Chinese).
[34] L. Wang, P.D. Groves, M. Ziebart, Smartphone shadow matching for better cross-street GNSS positioning in urban environments, J. Navig. 68 (3) (2015) 411–433, http://dx.doi.org/10.1017/S0373463314000836.
[35] L.T. Hsu, H. Tokura, N. Kubo, Y. Gu, S. Kamijo, Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons, Sensors 17 (6) (2017) 1909–1917, http://dx.doi.org/10.1109/JSEN.2017.2654359.
[36] L.T. Hsu, GNSS multipath detection using a machine learning approach, in: International conference on intelligent transportation systems, Yokohama, Japan, 2017, pp. 1-6. http://dx.doi.org/10.1109/ITSC.2017.8317700.
[37] Q. Phan, S. Tan, I.V. Mcloughlin, L.D. Vu, A unified framework for GPS code and carrier-phase multipath mitigation using support vector regression, Adv. Artif. Neural Syst. (2013) 1–14, http://dx.doi.org/10.1155/2013/240564.
[38] M. Socharoentum, H.A. Karimi, Y. Deng, A machine learning approach to detect non-line-of-sight GNSS signals in Nav2Nav, in: presented at the 23rd ITS World Congress, Melbourne, Australia, 10–14 October, 2016.
[39] Y. Quan, L. Lau, G.W. Roberts, X. Meng, C. Zhang, Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning, Remote Sens. 10 (12) (2018) 2052, http://dx.doi.org/10.3390/rs10122052.
[40] B. Guermah, H.E. Ghazi, T. Sadiki, H. Guermah, A robust GNSS LOS/multipath signal classifier based on the fusion of information and machine learning for intelligent transportation systems, in: 2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), Marrakech, Morocco, 2018, pp. 94-100. http://dx.doi.org/10.1109/ITMC.2018.8691272.
[41] R. Sun, L.T. Hsu, D. Xue, G. Zhang, W.Y. Ochieng, GPS Signal reception classification using adaptive neuro-Fuzzy inference system, J. Navig. (2018) 1–17, http://dx.doi.org/10.1017/S0373463318000899.
[42] L. Lau, P.A. Cross, Development and testing of a new ray-tracing approach to GNSS carrier-phase multipath modelling, J. Geod. 81 (11) (2007) 713–732, http://dx.doi.org/10.1007/s00190-007-0139-z.
[43] P.D. Groves, Z. Jiang, L. Wang, M.K. Ziebart, Intelligent urban positioning using multi-constellation GNSS with 3D mapping and NLOS signal detection, in: International Technical Meeting of the Satellite Division of the Institute of Navigation, Nashville, TN, USA, Sep. 2012, pp. 458-472.
[44] P.D. Groves, Z. Jiang, Height aiding, C/N_0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas, J. Navig. 66 (5) (2013) 653–669, http://dx.doi.org/10.1017/s0373463313000350.
[45] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Statist. 29 (5) (2011) 1189–1232, http://dx.doi.org/10.2307/2699986.
[46] L. Zhou, GBDT-SVM Credit risk assessment model and empirical analysis of peer-to-peer borrowers under consideration of audit information, Open J. Bus. Manag. 6 (2018) 362–372, http://dx.doi.org/10.4236/ojbm.2018.62026.
[47] H. Park, A. Haghani, S. Samuel, A.K. Michael, Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion, Accid. Anal. Prev. 112 (2018) 39–49, http://dx.doi.org/10.1016/j.aap.2017.11.025.
[48] L. Wang, D. Zhou, H. Zhang, W. Zhang, J. Chen, Application of relative entropy and gradient boosting decision tree to fault prognosis in electronic circuits, Symmetry 10 (10) (2018) 495, http://dx.doi.org/10.3390/sym1010049.
[49] J.R. Jang, ANFIS: adaptive-network-based fuzzy inference system, Syst. Man Cybern. 23 (3) (1993) 665–685, http://dx.doi.org/10.1109/21.256541.