**ORIGINAL ARTICLE**

# An efficient GNSS NLOS signal identification and processing method using random forest and factor analysis with visual labels

Linyang Li[1,2,3] · Zhenbang Xu[3] · Zhen Jia[3] · Luguang Lai[3] · Yang Shen[3]

## Abstract

The massive number of global navigation satellite system (GNSS) users and frequent positioning demands in cities, as well as the complexity of urban scenarios, pose many challenges for the accuracy and reliability of precise positioning. Since urban environments tend to suffer from GNSS non-line-of-sight (NLOS) signal conditions, leading to large ranging errors, NLOS signal identification and processing are of great importance. Usually, a visual camera can reflect real occlusion, and machine learning is efficient and accurate in processing multiple types of features. Therefore, an algorithm is proposed that combines the advantages of both methods. First, NLOS labels are generated using a combination of an inertial navigation system (INS) and a fisheye camera, and a total of nine features, namely, the elevation angle as well as the signal-to-noise ratios (SNRs), SNR fluctuation magnitudes, pseudorange consistencies, and pseudorange multipath errors at two frequencies, are extracted. Then, to improve efficiency and avoid overfitting, the nine original features are aggregated into three common factors via factor analysis, and these three factors can be well interpreted. Finally, a NLOS signal identification model based on the random forest (RF) algorithm is designed. In addition, to improve the precise point positioning (PPP) performance, a weighting scheme based on the elevation angle and SNR is optimized in accordance with the probability of NLOS occurrence. In an experiment, the RF model is trained using on-board dynamic multi-GNSS dual-frequency data collected by a low-cost UBLOX F9P receiver in Wuhan, and then validation is performed using data collected in Wuhan and Zhengzhou. The experimental results show that compared with the gradient boosted decision tree (GBDT), support vector machine (SVM), naive Bayes (NB), and convolutional neural network (CNN) algorithms, the RF model shows superior performance. While achieving 87.5% and 72.5% accuracy on the local and remote test datasets, respectively, the RF model costs only 12.2 ms for LOS/NLOS classification per epoch. Moreover, through factor analysis, the computational efficiency is improved by 29.5% for all five algorithms. Additionally, the accuracy and stability of uncombined PPP are improved using the proposed weighting strategy.

**Keywords** GNSS · NLOS signal identification · Visual labels · Random forest · Factor analysis

✉ Zhenbang Xu
   xzb2070@163.com

   Linyang Li
   linyangli@whu.edu.cn

   Zhen Jia
   13523567677@163.com

   Luguang Lai
   llg16690994518@163.com

   Yang Shen
   sy665432@163.com

1  School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

2  State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China

3  Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450001, China

## Introduction

Global navigation satellite systems (GNSSs) are widely used for their ability to provide users with global, all-weather, and high-precision navigation, positioning, and timing services (Hein 2020). However, in complex urban environments characterized by urban canyons, overpasses, and shade trees, GNSS signals are inevitably blocked, interfered with, and reflected, resulting in non-line-of-sight (NLOS) and multipath errors (Hsu et al. 2017a). Usually, the influence of NLOS signal conditions is larger than that of multipath conditions, even leading to hundreds of metres of ranging error. Therefore, efficient recognition and processing of NLOS signals are essential to achieve highly accurate and reliable positioning (Hsu 2018; Li et al. 2022).

Usually, three approaches are utilized in NLOS signal processing. The first is antenna design, in which choke antennas, dual-polarized antennas (Jiang and Groves 2014; Won and Pany 2017), and rotating GNSS antennas (Suzuki et al. 2020) are often exploited. The second approach concerns signal processing at the receiver and mainly focuses on optimizing the receiver tracking loop, e.g., using a multipath estimating delay-locked loop or a vector tracking loop (Jiang et al. 2021). However, the aforementioned two approaches can only partially mitigate NLOS errors while significantly increasing the cost, weight, and size of the hardware equipment. Additionally, these two approaches are mainly used in the geodetic field to achieve high accuracy in open environments and are unsuitable for frequent dynamic positioning in urban environments.

The last approach is observation modelling, in which GNSS observations are combined with other information sources. This approach can be further classified into two directions. The first requires a three-dimensional (3D) building model of the city or a fisheye/panoramic camera. By obtaining a 3D model of a vehicle's surrounding environment in advance, Kumar and Petovello (2017) utilized ray tracing, and Groves (2011) and Zhang et al. (2019) utilized shadow matching to detect NLOS signal conditions. However, these methods will be ineffective in areas where 3D building models are lacking, and the accuracy of NLOS signal identification relies on the accuracy and timeliness of the 3D model. Instead of using an existing 3D model, Meguro et al. (2009) used a panoramic infrared camera, and Marais et al. (2014) used a fisheye camera to distinguish buildings from the sky. Nevertheless, on the one hand, the attitude of the vehicle is ignored in these methods, which will make them invalid when the vehicle is going uphill or downhill. On the other hand, these methods are not accessible to a massive number of users, especially low-cost users, as they require external equipment or a 3D model.

Avoiding the need for external equipment, the second direction is to apply machine learning, which is efficient and accurate in processing multiple types of features (Siemuri et al. 2021). By considering the signal-to-noise ratio (SNR), pseudorange residuals, and elevation angle, Zhu et al. (2021) proposed an unsupervised-learning-based NLOS signal detection method that directly mined data features and performed unsupervised classification by means of cluster analysis, avoiding complicated data preprocessing. However, the identification accuracy was affected by anomalies, resulting in a low reliability. Using the SNR, SNR fluctuation magnitude (SFM), pseudorange residuals, and pseudorange fluctuation rate, an accuracy of 75% for signal classification has been achieved using a support vector machine (SVM) model (Hsu 2017b). In addition, using the SNR, pseudorange residuals, and elevation angle as input features, classification accuracies of 77.3% and 55.3% have been achieved on local and remote test datasets, respectively, based on a gradient boosted decision tree (GBDT) model (Sun et al. 2020). However, Hsu (2017b) and Sun et al. (2020) used ray tracing and a 3D model to tag NLOS signals. Using a decision tree, an accuracy of greater than 85% has also been realized, but the method required manual feature selection and tree design, resulting in a high computational burden (Yozevitch et al. 2016). By inputting time series features into a multivariate long short-term memory fully convolutional network, the signal classification accuracy and positioning accuracy have been improved for a static scene (Lyu and Gao 2020). Liu et al. (2019) designed a NLOS and multipath detecting network with five convolutional layers and two fully connected layers, and the classification accuracy was improved by 45% compared to two traditional SVMs. However, a higher computational load was also incurred, making the model unsuitable for real-time applications. Liu et al. (2023) constructed a NLOS signal detection model based on a convolutional neural network (CNN) using six double-difference features extracted from smartphone observations, and the detection accuracy reached more than 95% under the high-rise building static environment. Using the standard deviation (STD) of the pseudorange, SNR, elevation angle, and azimuth angle, Li et al. (2023) compared machine-learning-based classification algorithms to detect and exclude NLOS signals. In summary, the method of LOS/NLOS labels can be further refined by considering the attitude of the camera, and the selection of GNSS features for multiple epochs or single epochs is still not rich enough. Since the reliability of unsupervised classification methods that directly mine GNSS features is limited, the computational efficiency of supervised classification methods based on decision trees is low without dimension reduction of the GNSS features. In addition, a LOS/NLOS classification model trained on static data cannot be applied to complex dynamic scenes, and classification models trained on dynamic data have not

yet been applied to urban environments far away from the source of their training data; hence, the generality of such models has not been verified.

Additionally, although it is simple to directly remove the identified NLOS signals, the spatial geometry of positioning will be severely influenced by such signal removal in heavily occluded urban scenes. An alternative method is to utilize the NLOS signals through an improved stochastic model, which can be further classified into the model based on environment modelling and the model based on machine learning. The first type of model is represented by 3D-mapping-aiding (Xin et al. 2022), the difference between the observed and nominal carrier-to-noise power density ratios (C/N$_0$) (Sun and Wang 2022), and the combination of geographic cut-off elevation, azimuth, and C/N$_0$ (Zhang et al. 2022). Another model is based on the predicted probability of NLOS occurrence (Li et al. 2023, 2024), in which the differences are the machine learning method and original weighting strategy. Accordingly, to ensure the positioning performance, the processing strategy applied to the identified NLOS signals should be optimized.

To address the above problems, a combination of an inertial navigation system (INS) and a fisheye camera is used in this paper to generate LOS/NLOS labels. Nine undifferenced features are selected and reduced to three common factors via factor analysis. Using the RF algorithm, a NLOS signal identification model is trained. Additionally, a weight matrix updating strategy for a precise point positioning (PPP) Kalman filter is proposed to suppress NLOS signals. In the following sections, the proposed method for NLOS signal identification and processing is described in

detail. Then, the experimental data and processing strategies employed in LOS/NLOS classification and PPP are introduced. Thereafter, the performance and computational efficiency of NLOS signal identification are investigated, and the proposed weighting strategy using uncombined PPP is validated. Finally, the conclusions are summarized.

## Methods

The LOS/NLOS label generation method is introduced first, followed by the NLOS signal identification approach using an RF model and factor analysis. Finally, the weight matrix of the Kalman filter in the PPP model is determined.

### LOS/NLOS label generation using an INS and a fisheye camera

Figure 1 shows the LOS/NLOS tagging process with the assistance of an INS and a fisheye camera. First, the sky map is acquired using a wide-field-of-view fisheye camera facing the sky, and the position and attitude are obtained using a combination of GNSS and INS data. Next, to obtain the surrounding occlusion information, a sky segmentation method, which consists of coarse sky segmentation, missegmentation removal, and segmentation optimization, is used to separate the sky and the occluded regions. Then, using the multi-GNSS real-time ephemeris provided by International GNSS Service (IGS) and the pseudorange-based algorithm (Sunirana et al. 2013), the satellite coordinates are computed, and the receiver position and attitude are obtained

**Fig. 1** INS- and fisheye-camera-assisted LOS/NLOS tagging. The red, yellow, blue, green, and cyan dots represent GLONASS, GPS, QZSS, BDS, and Galileo satellites, respectively
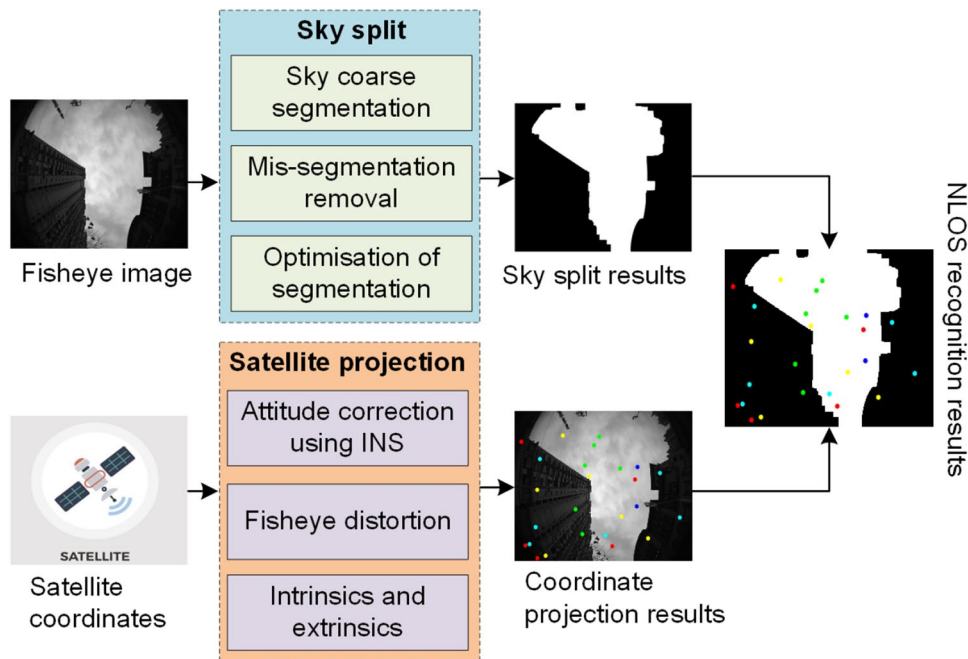
using the tight integration of PPP and INS, followed by the projected satellite coordinate in the image based on attitude correction and the distortion of the fisheye camera. Finally, based on the sky segmentation and satellite projection results, the percentage of nonsky areas within the range of the projected satellite coordinates is calculated to determine whether NLOS signal conditions are present.

## GNSS NLOS signal identification using an RF model with factor analysis

The nine selected representative undifferenced features are introduced first, and these features are then aggregated into three factors using factor analysis. Finally, a NLOS signal identification model is trained using the RF algorithm.

### Nine extracted undifferenced features from GNSS observables

When using machine learning algorithms for LOS/NLOS classification, the selection of reasonable and effective GNSS features is crucial. Since the use of any single feature is not ideal for LOS/NLOS classification and tends to lead to misclassification, multiple GNSS features are considered as input to improve the classification accuracy.

The effectiveness of the selected GNSS features is illustrated using on-board dynamic multi-GNSS dual-frequency observations collected by a low-cost UBLOX F9P receiver in a shaded environment. The B2I frequency of the BDS C06 satellite is selected, with an observation span ranging from 530,750 to 531,000 s in GPS week 2257.
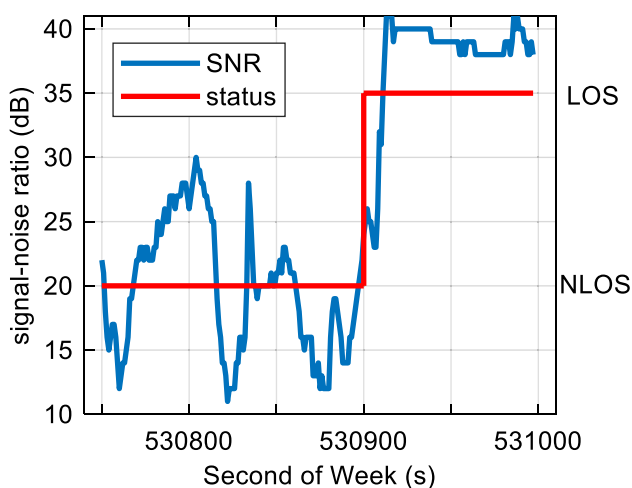
The first and second features are the SNRs at the first and second frequencies. Figure 2 shows the relationship between the SNR time series and the LOS/NLOS status. The average SNRs corresponding to LOS and NLOS signals are approximately 40 dB and 20 dB, respectively, indicating that the average SNR of NLOS signals is much smaller than that of LOS signals. NLOS signal conditions produce reflection or diffraction, increasing the propagation loss and attenuating the received signal strength. However, it is not reasonable to directly compare the SNR with a threshold and simply classify signals above that threshold as LOS signals and those below the threshold as NLOS signals. Low SNRs for LOS signals and high SNRs for NLOS signals are also observed, and therefore, other features must be further introduced to assist in classification.

The third and fourth features are the SFMs at the first and second frequencies. The SFM is calculated using the STD of the SNR within a certain window, as follows:

$$\text{SFM}(k) = \sqrt{\frac{1}{M} \sum_{m=k-M}^{k} \left( \text{SNR}_{k-m} - \langle \text{SNR} \rangle \right)^2} \tag{1}$$

where $M$ is the window size, $k$ is the number of the current epoch, and $\langle \cdot \rangle$ denotes the averaging operation.

Figure 3 shows the relationship between the SFM time series and the LOS/NLOS status. The window size is set to 30 epochs. The average SFMs for LOS and NLOS signals are approximately 1 and 5 dB, respectively. Due to the rapid changes in the reflection points and propagation paths of NLOS signals, the stability of the SNR is poor. Hence, the SFM can effectively distinguish LOS from NLOS signals.
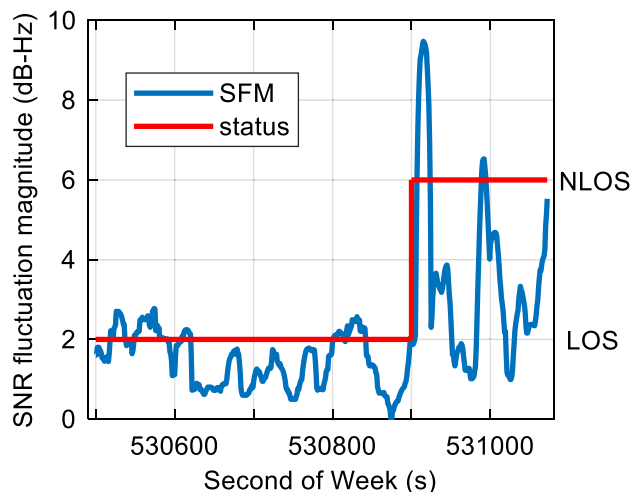


**Fig. 2** The relationship between the SNR time series at the B2I frequency of the C06 satellite and the LOS/NLOS status. The blue and red lines represent the SNR and the LOS/NLOS status, respectively



**Fig. 3** The relationship between the SFM time series at the B2I frequency of the C06 satellite and the LOS/NLOS status. The blue and red lines represent the SFM and the LOS/NLOS status, respectively

The fifth and sixth features are the pseudorange consistencies at the first and second frequencies. The pseudorange consistency is expressed as

$$PD_k = P_k - P_{k-1} - \frac{D_k + D_{k-1}}{2} \cdot \lambda \cdot \Delta t \tag{2}$$

where $P$, $D$, and $PD$ are the pseudorange, Doppler shift, and pseudorange consistency, respectively. $\lambda$ is the wavelength of the carrier phase, and $\Delta t$ is the sampling interval.

Figure 4 shows the relationship between the pseudorange consistency time series and the LOS/NLOS status. The pseudorange consistency of LOS signals is much smoother than that of NLOS signals. The pseudorange and Doppler shift, which are obtained from the code and carrier phase tracking loops, respectively, can be considered independent of each other. The effect of NLOS signal conditions on the pseudorange is greater than that on the Doppler shift; hence, the pseudorange consistency will be poor. The magnitude of the pseudorange consistency is proportional to the probability of NLOS signal conditions.

The seventh and eighth features are the multipath errors at the first and second frequencies. The multipath error at frequency $i$ can be estimated using the carrier phases at frequencies $i$ and $j$ as follows:

$$MP_i = P_i - \frac{f_i^2 + f_j^2}{f_i^2 - f_j^2} \lambda_i \varphi_i + \frac{2f_j^2}{f_i^2 - f_j^2} \lambda_j \varphi_j \tag{3}$$

where $\varphi$ is the carrier phase with a frequency of $f$.

Figure 5 shows the relationship between the multipath time series and the LOS/NLOS status. The multipath error
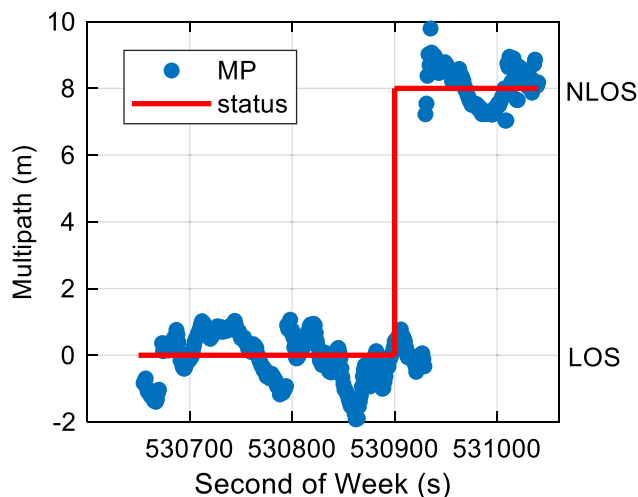


**Fig. 5** The relationship between the multipath error time series at the B2I frequency of the C06 satellite and the LOS/NLOS status. The blue and red lines represent the multipath error and the LOS/NLOS status, respectively

corresponding to LOS signals is close to zero, varying from $-2$ to 2 m, while the multipath error corresponding to NLOS signals is large, ranging from 6 to 10 m.

The last feature is the elevation angle. Usually, the probability of NLOS signal conditions is highly correlated with the elevation angle (Hsu 2018). A satellite signal with a higher elevation angle is less likely to be blocked by buildings and less likely to be reflected, and thus, it is more likely to directly reach the receiver.

### Three aggregated factors with clear meanings using factor analysis

If all nine selected features were to be used as input variables for machine learning, a large computational burden would be incurred. In addition, overfitting would be more likely to occur. Therefore, to improve the LOS/NLOS classification efficiency, factor analysis (Nitin 2006) is further used to aggregate the nine features into a few independent common factors, and these principal component variables are then used as the input for machine learning.

First, to test the suitability of factor analysis, the Kaiser−Meyer−Olkin (KMO) statistic (de la Fuente-Fernández 2011) and the Bartlett spherical test (Leech et al. 2013) are used, with threshold criteria of greater than 0.5 and less than 0.05, respectively. After standardization of the aforementioned nine features, a KMO value of 0.85 and a Bartlett spherical Sig value of 0.00 are obtained using SPSS software, indicating information entanglement between these features, meaning that it is appropriate to extract the effective principal components using factor analysis.
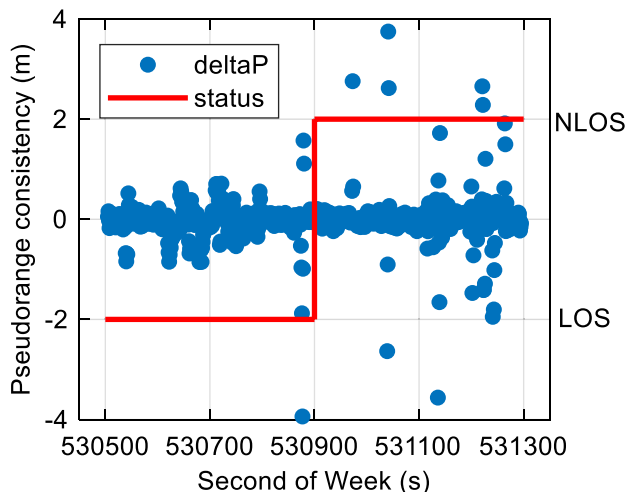


**Fig. 4** The relationship between the pseudorange consistency time series at the B2I frequency of the C06 satellite and the LOS/NLOS status. The blue and red lines represent pseudorange consistency and the LOS/NLOS status, respectively

**Table 1** Explanation of the total variance by the three selected factors

| Factor | Sum of the squares of the loads | | |
|---|---|---|---|
| | Eigenroot | Explanation of variance (%) | Cumulative value (%) |
| F1 | 4.35 | 43.49 | 43.49 |
| F2 | 3.45 | 34.52 | 78.01 |
| F3 | 1.05 | 10.47 | 88.48 |

Second, dimension reduction is performed based on factor analysis, and the eigenroots are calculated to analyse the contribution of each factor in explaining the variance of the variables. Factors with eigenroots larger than 1 are selected, and Table 1 lists how much of the total variance is explained by each of the three selected factors. As seen from the table, the contribution rates of the three most significant factors are 43.49%, 34.52%, and 10.47%, with a cumulative total contribution of 88.48%. Therefore, these three factors can sufficiently represent the aforementioned nine features.

Based on the maximum variance method, Table 2 lists the factor loading coefficients corresponding to the three selected factors. Each variable has a large loading on only one common factor and small loadings on the other factors, highlighting the association between each common factor and the variables that have a large loading on that factor.

Based on the information provided by the rotated factor loading array, Fig. 6 shows the correspondence between the three factors and nine features. SNR1, SNR2, SFM1, and SFM2 are mainly loaded on the first factor, and hence, the first factor is called the signal reception strength factor. PD1, PD2, MP1, and MP2 have large loadings on the second factor, and hence, the second factor is called the observation consistency factor. Since only the elevation angle has a large

**Table 2** Factor loading coefficients corresponding to the three factors

| Variable | Factor loading coefficients | | | Commonality |
|---|---|---|---|---|
| | F1 | F2 | F3 | |
| SNR at first frequency (SNR1) | 0.887 | 0.189 | 0.209 | 0.913 |
| SNR at second frequency (SNR2) | 0.841 | 0.193 | 0.052 | 0.901 |
| SFM at first frequency (SFM1) | 0.913 | 0.196 | 0.168 | 0.913 |
| SFM at second frequency (SFM2) | 0.899 | −0.217 | 0.177 | 0.916 |
| Pseudorange consistency at first frequency (PD1) | 0.116 | 0.922 | 0.050 | 0.937 |
| Pseudorange consistency at second frequency (PD2) | 0.094 | 0.887 | 0.214 | 0.923 |
| Multipath error at first frequency (MP1) | 0.149 | 0.891 | 0.182 | 0.908 |
| Multipath error at first frequency (MP2) | −0.130 | 0.901 | 0.211 | 0.898 |
| Elevation angle | 0.192 | 0.204 | 0.841 | 0.836 |



**Fig. 6** Correspondence between the three factors and nine features
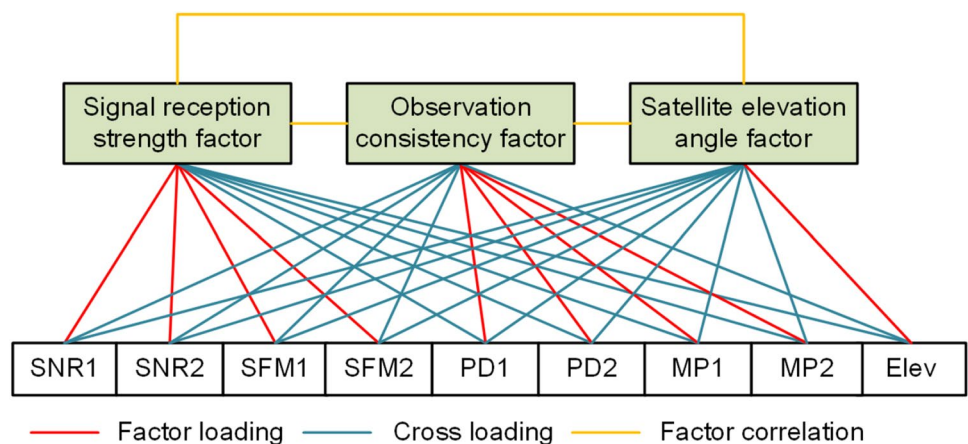
loading on the third factor, the third factor is called the satellite elevation angle factor.

Finally, by transforming the rotated factor loading array, the scores are obtained to express the relationships between the nine features and each of the three factors as

$$F_1 = 0.242\text{SNR}_1 + 0.211\text{SNR}_2 + 0.297\text{SFM}_1 + 0.288\text{SFM}_2 + 0.042\text{PD}_1$$
$$+0.033\text{PD}_2 + 0.074\text{MP}_1 - 0.069\text{MP}_2 + 0.083\text{elevation}$$
$$F_2 = 0.022\text{SNR}_1 + 0.027\text{SNR}_2 + 0.023\text{SFM}_1 - 0.072\text{SFM}_2 + 0.263\text{PD}_1$$
$$+0.241\text{PD}_2 + 0.244\text{MP}_1 + 0.252\text{MP}_2 + 0.043\text{elevation}$$
$$F_3 = 0.051\text{SNR}_1 + 0.034\text{SNR}_2 + 0.041\text{SFM}_1 + 0.044\text{SFM}_2 + 0.034\text{PD}_1$$
$$+0.057\text{PD}_2 + 0.046\text{MP}_1 + 0.053\text{MP}_2 + 0.473\text{elevation}$$

$$(4)$$

By aggregating the nine features into three factors, dimension reduction is achieved, and these three factors have clear meanings. With fewer inputs during machine learning, the computational load can be reduced, and the method becomes more efficient. In addition, the risk of overfitting is reduced, and the generalization ability is improved.

## NLOS signal identification using an RF model

Based on the concept of bagging, the RF algorithm is used to train multiple tree-like LOS/NLOS classifiers in parallel; then, by averaging the probabilities predicted by all trees using soft voting, the probability of LOS/NLOS signal

occurrence can be obtained. Twofold randomness is utilized to further enhance the classification performance of the RF model (Breiman 2001). The first source of randomness is self-sampling, in which the LOS/NLOS training set is randomly sampled for decision tree training; the second is a random subspace, in which random sampling of the features is performed.

Figure 7 shows the flowchart of the RF-based NLOS signal identification process, which consists of three steps. The first step is bootstrap resampling. Each sample in the training set is denoted by $F_i = (F_{1i}, F_{2i}, F_{3i})$, $i \in [1, N]$, where $N$ is the number of samples. Using the bootstrap method, $l$ subsets of samples, denoted by $\{T_1, T_2, \cdots T_l\}$, are randomly selected from the training sample dataset $\mathbf{T} = \{(\mathbf{F}_1, y_1), (\mathbf{F}_2, y_2), \cdots, (\mathbf{F}_N, y_N)\}$, $(y_i = 0, 1)$ for training decision trees, where each subset has the same number of samples. Here, $y_i = 0$ and 1 represent LOS and NLOS signals, respectively.

The second step is the construction of the base classifiers, for which the classification and regression tree (CART) algorithm is used to construct the corresponding decision trees. To form a subset of splitting factors for a decision node, the random subspace method is utilized to randomly draw $d$ subfactors from the three factors with equal probability, and the Gini coefficient is calculated as

**Fig. 7** Flowchart of RF-based LOS/NLOS signal classification. The blue and orange ellipses represent the randomly selected candidate factors and the optimal splitting factors, respectively
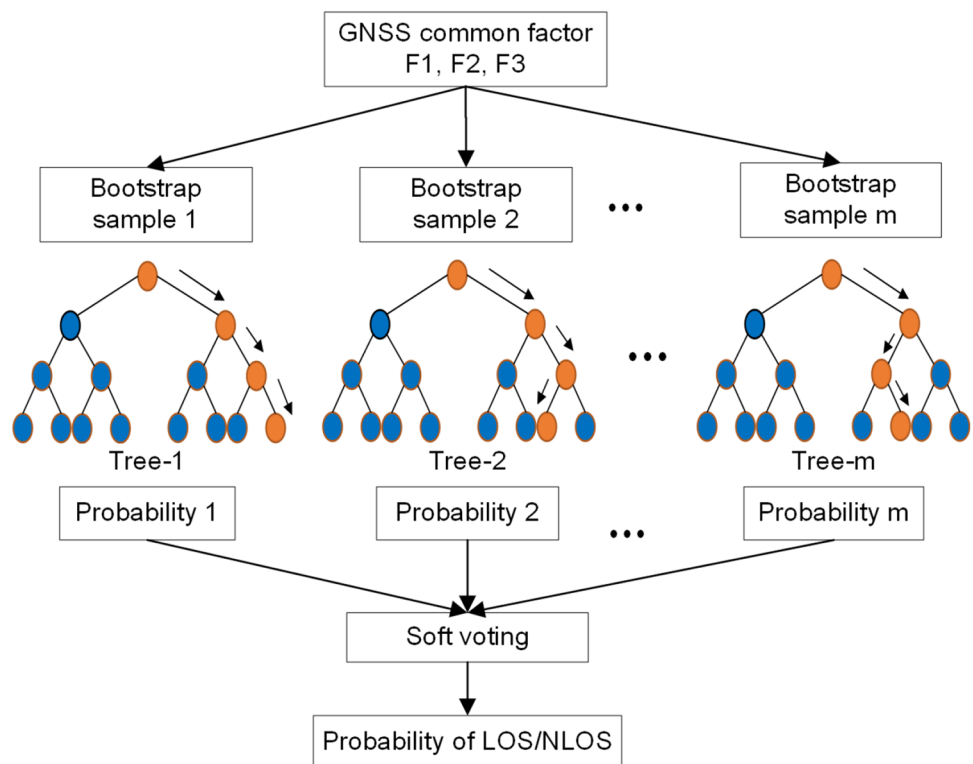
$$\text{Gini}\left(\mathbf{F}_d\right) = 1 - \left(p_0^2 + p_1^2\right) \tag{5}$$

where $\mathbf{F}_d$ is the currently selected subfactor and $p_0$ and $p_1$ denote the probabilities that a sample belongs to the LOS and NLOS signal conditions, respectively.

By selecting the factor with the minimum Gini coefficient, an optimal splitting factor and splitting value are selected to split the tree at the current decision node, and the tree is built recursively in this way until each feature factor becomes a splitting node. Moreover, the above random process is repeated $m$ times, and the $m$ decision trees thus built form a random forest.

The last step is soft voting, in which each decision tree makes a soft prediction and calculates the probability of NLOS signal conditions based on the input test dataset. Then, by averaging the predicted probabilities $y_n(\mathbf{F})$ obtained from all the decision trees, the NLOS probability $Y(\mathbf{F})$ is obtained as follows:

$$Y(\mathbf{F}) = \frac{\sum_{n=1}^{m} y_n(\mathbf{F})}{m} \tag{6}$$

where $Y(\mathbf{F})$ represents the degree of GNSS signal contamination and can be used to optimize.

## Weighting strategy for the Kalman filter in the uncombined PPP model

After a NLOS observation is identified, simply directly removing the corresponding satellite would destroy the satellite spatial geometry. In the case of a heavily occluded scene, for which NLOS observations account for most of the total observations, the removal of all NLOS satellites may even result in a shortage of available satellites and positioning failure. Using the predicted NLOS probability from machine learning, Li et al. (2023) refined the stochastic model with SNR and elevation, while Li et al. (2024) utilized elevation. In the contribution, minor improvements are made compared with the above two researches.

As adopted by Tay and Marais (2013) and Adjrad and Groves (2017), a combination of the SNR and elevation is utilized simultaneously to compute the weight of the observation. To weaken the influence of NLOS observations, the weights of the detected NLOS signals are decreased using the NLOS probability obtained from (6), as follows:

$$\overline{P}_{\text{NLOS}} = P \cdot (1 - Y(\mathbf{F}))$$

$$P = k \cdot \left(a + b \exp\left(-\frac{\theta}{\theta_0}\right)\right)^2 \cdot \left(c \times 10^{-\text{SNR}/10}\right) \tag{7}$$

where $P$ and $\overline{P}_{\text{NLOS}}$ denote the original weight and the optimized weight, respectively, of the NLOS signal, $k$ is the scaling factor between the code pseudorange and carrier phase. In the above expression, the parameter values adopted by Adjrad and Groves (2017) are used, namely, $a = 0.13\,\text{m}$, $b = 0.56\,\text{m}$, $c = 1.1 \times 10^4\,\text{m}^2\,\text{s}^{-1}$, and $\theta_0 = 0.1745\,\text{rad}$. By decreasing their weights in this way, the influence of NLOS observations on the estimated parameters can be reduced. The NLOS probability $Y(\mathbf{F})$ varies from 0 to 1, if the NLOS probability equals to 0, a LOS satellite is observed with normal weight $P$. And if $Y(\mathbf{F}) = 1$, this satellite will be removed. Additionally, the original weight $P$ is similar to Li et al. (2023), except for the coefficients $a$ and $b$ adopted by the well-known GAMIT software.

Figure 8 shows the framework of NLOS signal identification and PPP optimization based on visual labels, the RF algorithm, and factor analysis. First, on-board GNSS dynamic observation data collected in complex urban environments are used to generate LOS/NLOS visual tags with the assistance of an INS and a fisheye camera. Next, nine features are extracted by interpolating the satellite positions using real-time precise orbit products, and the extracted features are aggregated into three common factors via factor analysis. Then, the three factors and corresponding visual labels are fed into an RF classifier for training to extract the rules for LOS/NLOS classification. Finally, based on dynamic observations from a low-cost receiver, LOS/NLOS signal classification is performed using the trained RF classifier, and PPP validation is carried out using the optimized weighting method for NLOS observations.

## Experimental data and processing strategy

This section first describes the experimental platform and data and then specifies the parameter settings and performance evaluation metrics used for the machine learning algorithms as well as the processing strategy for uncombined PPP.

### Experimental platform and data

Figure 9 shows the on-board platform utilized in the experiments. It consisted of a low-cost UBLOX F9P receiver and a helical antenna that can receive dual-frequency observations from GPS, BDS, Galileo, GLONASS, and QZSS; a FLIR fisheye camera (BFS-PGE-16S2) facing the sky; and a tactical INS (StarNeto XW-GI7660). In addition, a time synchronization board was utilized to unify the timestamps of the different sensors to GPS Time. A nearby reference station was set up in an open environment, and tightly coupled GNSS/INS integrated smoothing results from the postprocessing software Inertial Explorer 8.90 were used as a high-accuracy reference trajectory and attitude for the

**Fig. 8** Framework of the proposed NLOS signal identification and processing method using an RF model and factor analysis with visual labels. The framework is separated into an offline model training stage and a service application stage
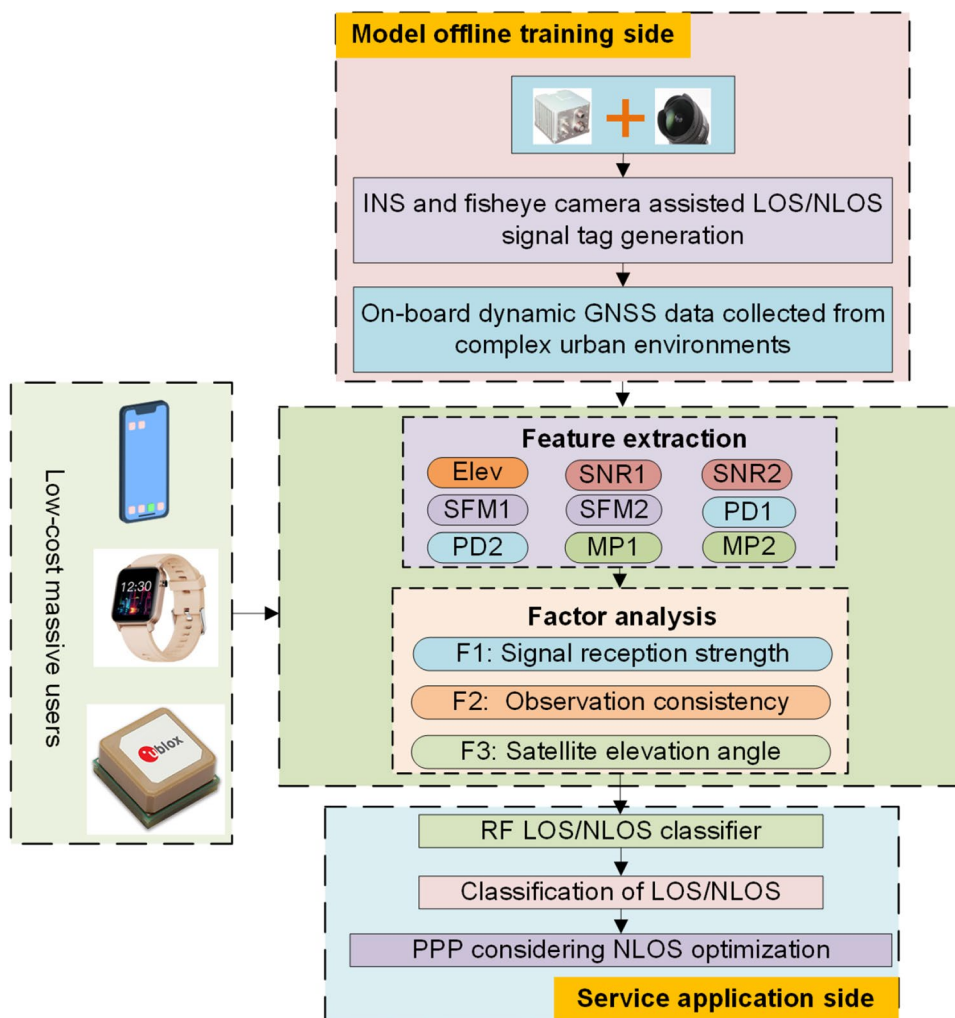




**Fig. 9** On-board experimental hardware platform

vehicle. Additionally, the NLOS signal identification results from the fisheye images were used as the labels fed to the classifier, and NLOS and LOS signals were treated as positive and negative examples, respectively.

Figure 10 shows the experimental trajectories along which data were collected in Wuhan and Zhengzhou, China, using the above hardware platform. The Wuhan and Zhengzhou datasets are collected on January 6th, 2022 and April 15th, and there is a significant difference in the urban environments due to seasonal variations. 2023Dataset Note that the distance between Wuhan and Zhengzhou is more than 500 kms. Figure 11 further gives the PDOP and the number of available satellites during the experiments. As seen from this figure, the GNSS observation conditions were good throughout most of the experimental duration in each scenario, but the numbers of satellites observed in the urban environments varied dramatically. In the case
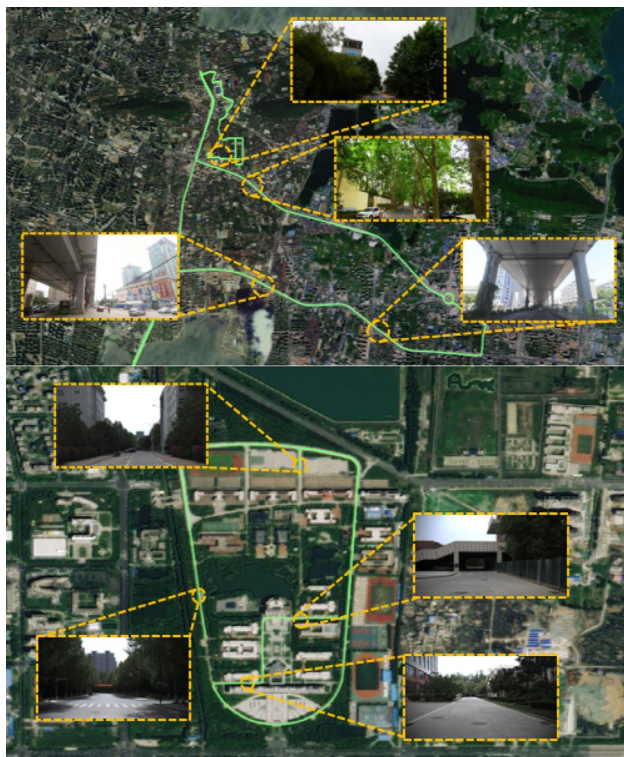
**Fig. 10** Vehicle trajectories in two experiments. The top and bottom panels represent the trajectories in Wuhan and Zhengzhou, respectively

of severe GNSS signal occlusion, the number of available satellites decreased significantly, and the PDOP increased substantially.

Figure 12 gives the velocities of the two datasets. It can be seen that the vehicle speed is less than 1.5 m/s during some sessions in both datasets, therefore the two datasets can simulate the pedestrian motion.

Figure 13 further divides the experimental environments into nine typical scenes. In contrast to scene 9, which represents an open environment on an overpass, scene 1 and scene 2 are typical unilaterally occluded and urban canyon scenes with dense clusters of high-rise buildings near the road. Furthermore, scene 3 and scene 4 are bilaterally occluded scenes with dense clusters of high-rise buildings or tall trees on both sides of the road. Scene 5 and scene 6 are bridge-obstructed environments under overpasses, with scene 5 being especially obscured. Scene 7 and scene 8 are dense tree-shaded environments; between the two, scene 8 is more occluded.

Table 3 specifies the division of the experimental training and test datasets. Since environmental sensitivity is a key issue in NLOS classification, the LOS/NLOS classification rules obtained from one training environment cannot be fully applicable for a different test environment. Therefore,
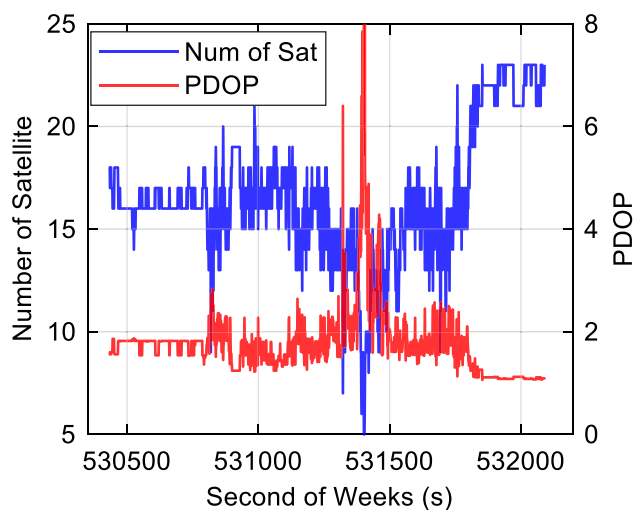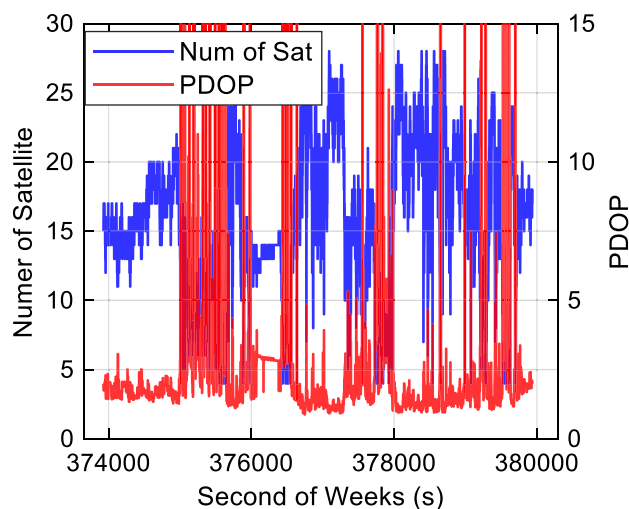


**Fig. 11** The PDOP and the number of available satellites in each experiment. The top and bottom panels present the results for Wuhan and Zhengzhou, respectively

to evaluate the transferability of the obtained NLOS signal identification model to different environmental scenarios, the data from Wuhan were divided into a training set and a local test dataset, and the data collected in Zhengzhou were used as a remote test dataset. The training dataset and both test datasets each contained scenes of all types corresponding to scenes (1) to (9) in Fig. 13. A total of 51,061 sample data points were used for the experimental analysis, including 25,958 positive and 25,103 negative cases, meaning that the ratio between the numbers of positive and negative cases was approximately 1:1. A dedicated laptop with a 2.50 GHz Intel Core (TM) i5-10500H processor and 8.0 GB of memory was used in the experiments.
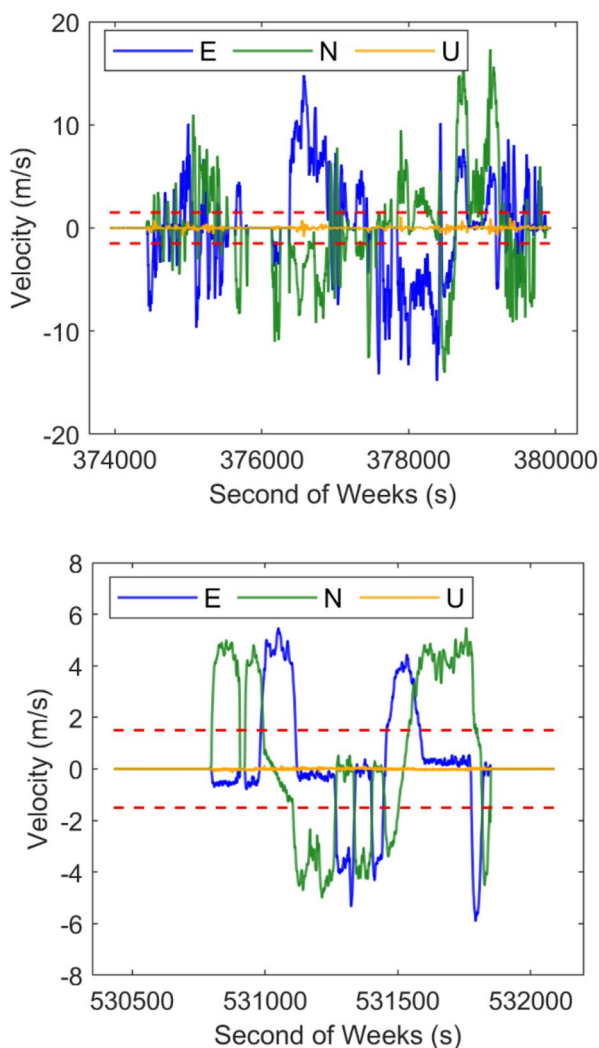
**Fig. 12** The velocities of the two datasets. The top and bottom panels represent the vehicle velocities in Wuhan and Zhengzhou, respectively. The red dashed line indicates that the vehicle speed is 1.5 m/s

### Processing strategy

Three machine learning algorithms, namely, the GBDT, SVM, naive Bayes (NB), and CNN algorithms, were selected for comparison. Cross-validation was used to determine the hyperparameters of the RF and GBDT models, and the number of classifiers and the maximum depth of the decision trees were set to 100 and 30, respectively. For the SVM model, soft interval classification with a penalty factor of 0.5 was used, and to achieve nonlinear classification, a Gaussian kernel function was used to map the input features to a high-dimensional space. In addition, the Gaussian NB classification method was used for NB model training, and as Liu et al. (2023) adopted, the CNN was constructed with two convolutional layers with the

hyperbolic tangent function (tanh) as activation function, one max pooling layer, one flatten layer, and one dense layer.

Validation was performed on the above two test datasets based on the trained classification models. The performance evaluation metrics are expressed as

$$
\begin{aligned}
\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\
F_{1\_}\text{score} &= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\end{aligned}
\tag{8}
$$

where TP and FN refer to cases in which a NLOS sample is determined to be a NLOS or LOS sample, respectively, and TN and FP refer to cases in which a LOS sample is classified as LOS or NLOS, respectively. $F_{1\_}$score is the harmonic mean of precision and recall.

Table 4 gives the detailed processing strategy for uncombined PPP. In all experiments, dual-frequency uncombined observations were used, and if any feature was missing for a certain satellite, e.g., in the case of single-frequency observables, that satellite was removed.

## Experimental results and analysis

This section first compares the performance of the five algorithms for NLOS signal identification, then analyses their efficiencies with and without factor analysis, and finally presents a validation of uncombined PPP.

### Performance of NLOS identification using local test dataset and remote test dataset

Table 5 gives the NLOS signal identification performance of the five machine learning algorithms on the local and remote test datasets. The LOS/NLOS classification performance of the ensemble learning models (RF and GBDT) and machine learning model CNN is basically the same, with the differences in all four metrics being less than 2%. Moreover, for all four metrics, values of more than 85% and 71% are achieved on the local and remote test datasets, respectively. In addition, using a DJI M300 drone and a P1 camera, a decimetre-accuracy 3D model corresponding to the remote dataset was obtained eight months before the experiments, and this 3D model was utilized in place of the fisheye camera and INS to generate visual labels. Compared with the NLOS signal identification results obtained using the fisheye camera and INS, the accuracy achieved using the 3D model and ray tracing is slightly lower. The cause of this is

**Fig. 13** Nine representative urban scene characteristics and observation environments selected in the experiments
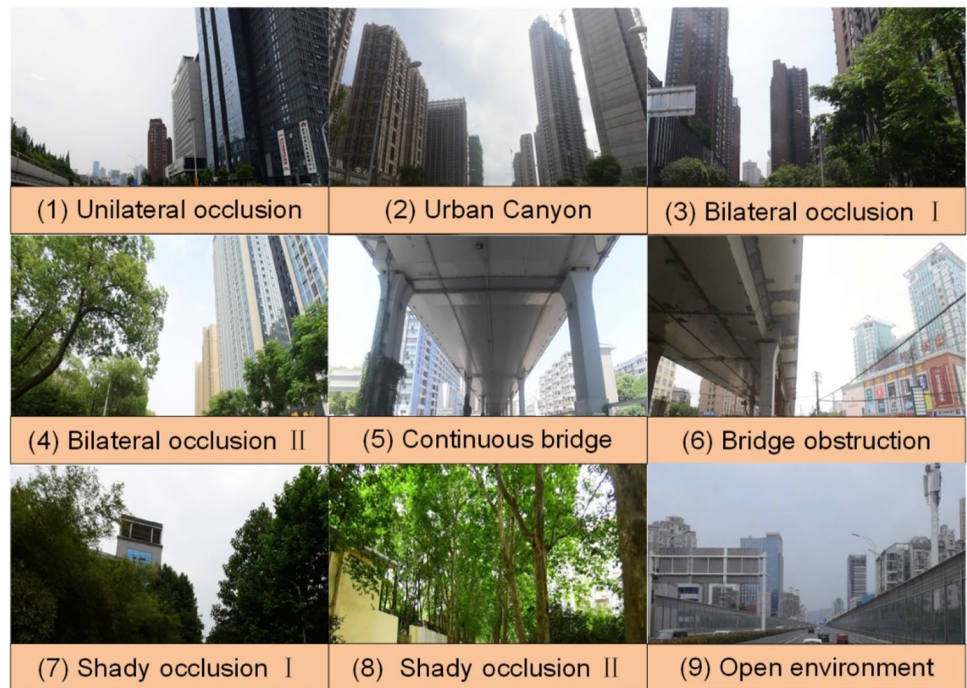


**Table 3** Description of the training dataset and two test datasets

| Dataset | Collection location | Data span (min) | Included scenes | NLOS/LOS |
|---|---|---|---|---|
| Training dataset | Wuhan, China | 60 | (1)~(9) | 16,452/15,408 |
| Local test dataset | Wuhan, China | 30 | (1)~(9) | 6302/5087 |
| Remote test dataset | Zhengzhou, China | 20 | (1)~(9) | 3204/4608 |

**Table 4** Uncombined PPP processing strategy

| Type | Processing strategy |
|---|---|
| Sampling interval | 1 s |
| Frequencies | GPS: L1/L2; Galileo: E1/E5b; BDS: B1I/B2I; GLONASS: G1/G2; QZSS: L1/L2 |
| Elevation mask | 7° |
| Satellite ephemeris and clock offsets | Post precise products from the Centre for Orbit Determination in Europe (CODE) |
| Tropospheric delay | A priori tropospheric delay corrected using the Saastamoinen model, global pressure and temperature 3 (GPT3), and the vienna mapping functions 3 (VMF3) |
| Ionospheric delay | Estimated per satellite |
| Ambiguity | Modelled as a float constant without cycle slips |

**Table 5** NLOS signal identification performance using five machine learning algorithms on the local and remote datasets

| Algorithm | Accuracy | | Precision | | Recall | | F1_score | |
|---|---|---|---|---|---|---|---|---|
| | Local set (%) | Remote set (%) | Local set (%) | Remote set (%) | Local set (%) | Remote set (%) | Local set (%) | Remote set (%) |
| RF | 87.5 | 72.5 | 86.4 | 71.4 | 86.8 | 73.2 | 86.6 | 72.3 |
| CNN | 88.7 | 71.9 | 88.6 | 73.2 | 85.2 | 73.0 | 86.9 | 73.1 |
| GBDT | 86.9 | 71.7 | 85.3 | 71.0 | 87.2 | 73.3 | 86.2 | 72.1 |
| SVM | 74.8 | 65.8 | 72.6 | 68.4 | 68.1 | 64.1 | 70.3 | 66.2 |
| NB | 71.3 | 63.3 | 71.1 | 66.5 | 67.2 | 67.6 | 69.1 | 67.0 |

surmised to be the limited accuracy and timeliness of the 3D model due to the different season at the time of acquisition (affecting, e.g., the vegetation states of trees) as well as the possible presence of new buildings and occlusions since the time of acquisition. And hence this is the first time that both a fisheye camera and an INS have been used in the generation of NLOS labels, leading to more convincing results.

For the SVM and NB classifiers, for which none of the four metrics exceeds 75% on the two test datasets, the NLOS signal identification performance is relatively poor compared to that of the GBDT, RF, and CNN models. The NB model exhibits the lowest recall, successfully identifying the fewest NLOS signals, whereas the SVM model has the worst accuracy, implying the highest probability of misclassifying NLOS signals.

Hence, taking all four metrics into consideration, the two integrated learning methods (RF and GBDT) and CNN outperform the single learners (SVM and NB) in classification. This is because the GBDT and RF models use boosting and bagging, respectively, as integrated learning methods, and CNN applies the multi-convolutional layers, pooling layer, local connection, and weight sharing, resulting in better LOS/NLOS classification accuracy, precision, and recall as well as better F1_score values.

## Computational efficiency of NLOS signal identification

The high sampling rate and continuous long-term nature of GNSS observations lead to considerable computational pressure, and hence, a fast algorithm with low computational effort is urgently needed. Therefore, in addition to the LOS/NLOS classification performance, the efficiency of NLOS signal identification is of particular importance.

Table 6 reports the time consumptions for 1800 epochs and a single epoch when using the different algorithms. Between the two traditional machine learning algorithms using a single learner, the NB algorithm has the lower computational complexity and the faster speed for NLOS signal identification, consuming only 11.1 ms per epoch. In contrast, with high computational complexity and low

**Table 6** Time consumptions for LOS/NLOS classification using the five algorithms

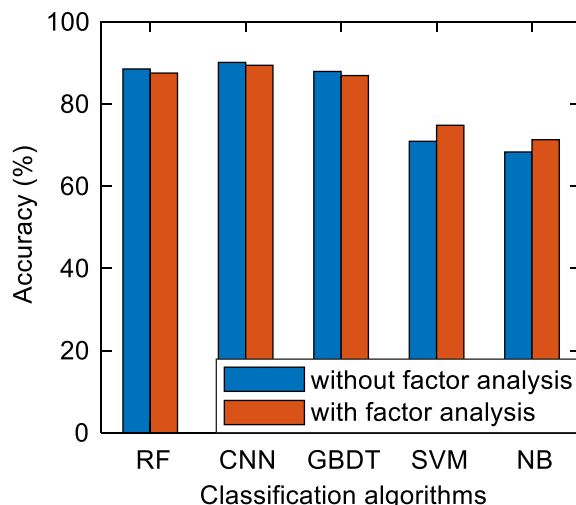| Type | RF | CNN | GBDT | SVM | NB |
|---|---|---|---|---|---|
| Processing time for 1800 epochs (s) | 21.9 | 57.4 | 38.0 | 30.1 | 19.9 |
| Average time for a single epoch (ms) | 12.2 | 31.9 | 21.1 | 16.7 | 11.1 |



**Fig. 14** Effect of factor analysis on the classification accuracy of the five signal classifiers

classification efficiency, the SVM model has a time consumption of 16.7 ms per epoch.

The RF algorithm uses the strategy of bagging integration, in which the training subset is first determined via random sampling, several base learners are selected to train models in parallel, and a voting method is finally used to integrate the classification results, greatly improving the efficiency of LOS/NLOS classification. In contrast, the GBDT algorithm uses the strategy of boosting integration, which relies on multiple serial trees, each tree depending on the residuals of the previous tree. And the depth and width of the multi-layer neural networks bring in large computational complexity and burden, the CNN, which belongs to the deep
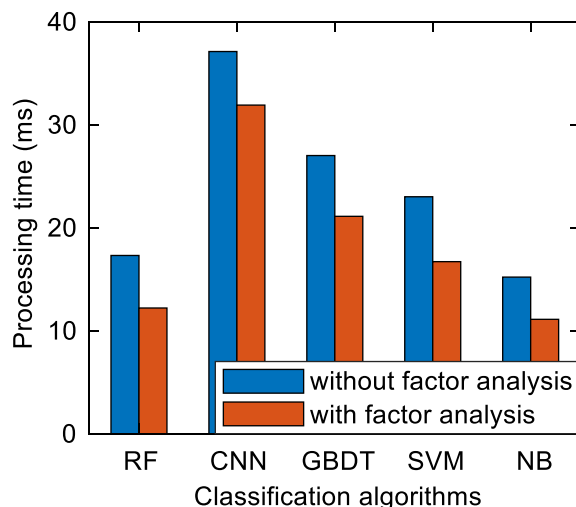


**Fig. 15** Effect of factor analysis on the classification efficiency of the five signal classifiers

learning approach, has a higher computational complexity and is better at handling high-dimensional data, i.e., the images and speeches. Consequently, the GBDT and CNN models require more time for training and classification and hence is less efficient. Therefore, considering all four metrics and the LOS/NLOS classification efficiency, the RF algorithm shows the best performance.

To further evaluate the impact of factor analysis on the performance of the LOS/NLOS signal classifiers, the local test dataset was utilized. Figures 14 and 15 present the classification accuracy and efficiency of the five signal classifiers with and without factor analysis, respectively. The statistical results show a variation of no more than 4.0% between the cases with and without factor analysis; therefore, the difference in the classification accuracy for each classifier is not significant. However, the classification efficiency increases by 29.5% for all five classifiers. After factor analysis, only three common factor variables are used to represent almost all of the variance of the original nine GNSS features, thereby greatly



Fig. 16 The experimental data in the local test dataset. The top and bottom panels represent the trajectory and the PPP accuracy, respectively

reducing the computational burden with little loss of feature information.

## Uncombined PPP validation using local test dataset and remote test dataset

Figure 16 presents the relationship between the PPP accuracy and the trajectory for the local test dataset, with different colours representing different error ranges of the uncombined PPP results. In some sections of the trajectory with severe occlusions, the positioning deviations were large, and some short interruptions even occurred, making it difficult to provide highly accurate and continuous positioning services. Therefore, it is necessary to enhance the PPP performance by making use of the identified NLOS observations.

Figure 17 shows the error series of the original PPP and the optimized uncombined PPP obtained using the proposed weighting strategy, and four local features are further enlarged in the horizontal component. It can be seen that the accuracy and stability of the uncombined PPP results are improved after the weight matrix of the NLOS observations is updated. For example, in the local test dataset, the number of available satellites decreased at 379,090 s, and the PDOP abruptly increased. In the subsequent time range of 379,100–379200 s, the convergence speed of the proposed weighting method was faster, and its positioning accuracies in the horizontal and vertical components were significantly better. In the remote test dataset, the number of NLOS satellites increased at 531,305 s, and positioning reconvergence was necessary when using the traditional uncombined PPP method. However, the weights of NLOS observations were reduced based on the probability of NLOS signal occurrence, the positioning performance in the east, north, and vertical components improved considerably. For the two kinds of results, their positioning accuracies in the horizontal component are close to each other, especially in the open environment, while these of vertical component are obviously different.

Table 7 further gives the statistics of the positioning accuracy on the two test datasets. Compared with the original uncombined PPP method, the proposed weighting method improves the positioning accuracy and stability in both the horizontal and vertical components. For the local test dataset, the positioning accuracies in the horizontal and vertical components improve by 13.24% and 18.24%, respectively, and the corresponding improvements on the remote test dataset are 6.10% and 3.46%. The magnitude of accuracy improvement on the local test dataset is higher than that on the remote test dataset. On the one hand, the migration of the environment leads to an increased probability of missing or misclassifying NLOS signals using the RF method, which is consistent with Table 5. On the other hand, the remote test dataset was mainly collected on a university campus,
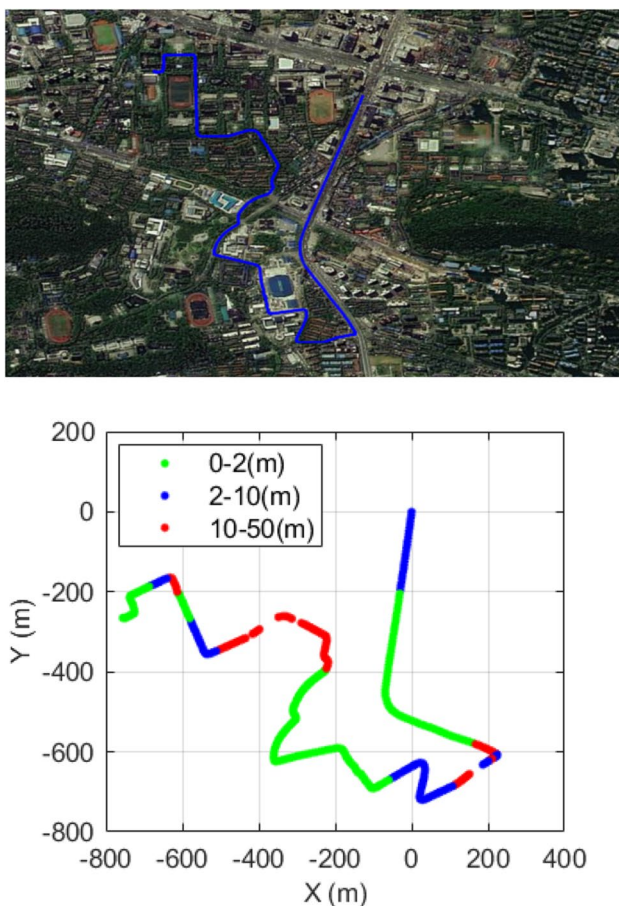
**Fig. 17** The error series of the
original PPP results and the
optimized PPP results obtained
with the proposed weighting
strategy. The top and bottom
panels represent the results
on the local and remote test
datasets, respectively. The
orange and green dots represent
the original uncombined PPP
results and the optimized
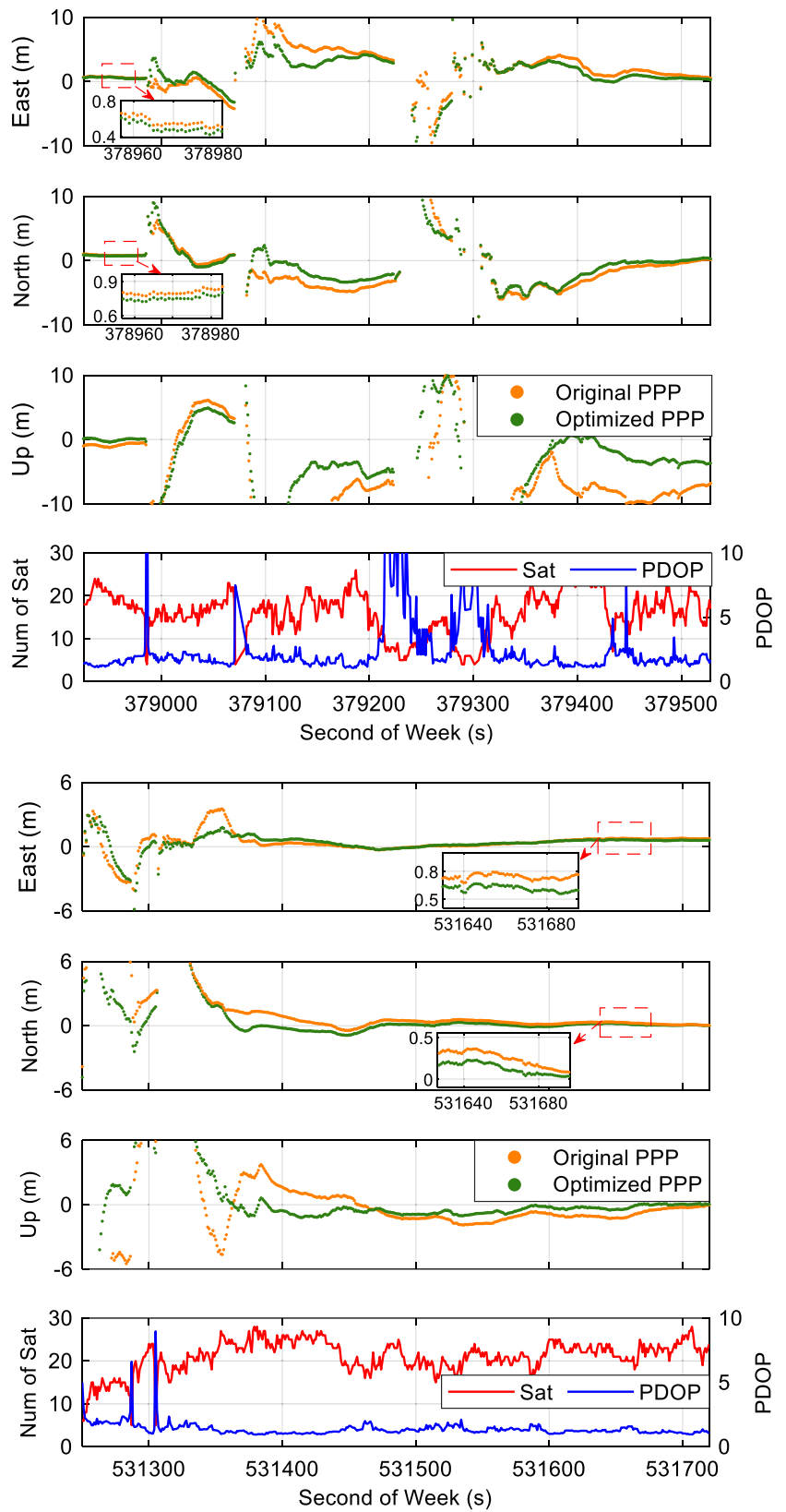PPP results obtained with the
weighting scheme, respectively

**Table 7** Statistics of the positioning performance of uncombined PPP with and without optimization of weighting

| Dataset | Component | Solution type | RMS (m) | Improvement |
|---|---|---|---|---|
| Local test dataset | Horizontal | Original PPP | 10.27 | – |
| | | PPP with optimization | 8.91 | 13.24% |
| | Vertical | Original PPP | 12.61 | – |
| | | PPP with optimization | 10.31 | 18.24% |
| Remote test dataset | Horizontal | Original PPP | 2.95 | – |
| | | PPP with optimization | 2.77 | 6.10% |
| | Vertical | Original PPP | 4.63 | – |
| | | PPP with optimization | 4.47 | 3.46% |

which is not as complex as the observation environment of the local test dataset, and hence, the accuracy of the original uncombined PPP method on the remote test dataset was already higher than that on the local test dataset; hence, the improvement in positioning accuracy is less obvious.

## Conclusions and remarks

Many challenges arise in GNSS-based navigation and positioning under complex urban environments, among which NLOS signal conditions are a nonnegligible issue. In this work, first, an INS and a fisheye camera are used together to generate accurate NLOS labels. Next, a total of nine features are selected, and to reduce their dimensionality and avoid overfitting in machine learning, they are reduced to three common factors via factor analysis. Then, an efficient RF-based NLOS signal identification model is designed. Finally, to improve the PPP performance, the weights of the detected NLOS observations are decreased using the corresponding NLOS probability.

Compared with the GBDT, SVM, NB, and CNN algorithms, when the classification accuracy, precision, recall, F1-score, and classification efficiency are considered, the RF algorithm better balances accuracy and efficiency in NLOS/LOS classification. Specifically, the proposed RF model has a time cost of only 12.2 ms per epoch for LOS/NLOS classification while achieving 87.5% and 72.5% accuracy on local and remote test datasets, respectively. In addition, with factor analysis, the computational efficiency is improved by approximately 30.0%.

When the proposed weighting method is used, both the positioning accuracy and stability of uncombined PPP are improved, with improvements in the horizontal and vertical components of 13.24% and 18.24%, respectively, on the local test dataset and 6.10% and 3.46%, respectively, on the remote test dataset.

The efficient machine learning model proposed in this paper can be used on a massive scale for low-cost users faced with real-time dynamic scenarios in complex urban environments. However, only dual-frequency observations are used in this paper, although the factor analysis is expected to be more efficient when using multifrequency observations. And the visual labels can also be generated at night using an infrared fisheye camera. Moreover, in addition to PPP validation, other models should be used for further validation, e.g., PPP/INS and PPP-RTK, which will be a focus of our future research.

**Data availability** The data analysed during the current study are available from the corresponding author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no conflicts of interest.

**Ethical approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** The authors confirms that the work described has not been published before; that it is not under consideration for publication elsewhere; that its publication has been approved by all coauthors, if any; and that its publication has been approved by the responsible authorities at the institution where the work was carried out.

# References

Adjrad M, Groves PD (2017) Enhancing least squares GNSS positioning with 3D mapping without accurate prior knowledge. Navig J Inst Navig 64(1):75–91. https://doi.org/10.1002/navi.178

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

De la Fuente-Fernández S (2011) Factorial analysis. Dissertation, University Autónoma Madrid

Groves PD (2011) Shadow matching: a new GNSS positioning technique for urban canyons. J Navig 64(3):417–430. https://doi.org/10.1017/S0373463311000087

Hein GW (2020) Status, perspectives and trends of satellite navigation. Satell Navig 1:22. https://doi.org/10.1186/s43020-020-00023-x

Hsu LT (2018) Analysis and modeling GPS NLOS effect in highly urbanized area. GPS Solut 22(1):7. https://doi.org/10.1007/s10291-017-0667-9

Hsu LT, Tokura H, Kubo N, Gu Y, Kamijo S (2017) Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons. IEEE Sens J 17(6):1909–1917. https://doi.org/10.1109/JSEN.2017.2654359

Hsu LT (2017b) GNSS multipath detection using a machine learning approach. In: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC). IEEE, Yokohama, Japan, pp 1–6. https://doi.org/10.1109/ITSC.2017.8317700

Jiang Z, Groves P (2014) NLOS GPS signal detection using a dual-polarisation antenna. GPS Solut 18(1):15–26. https://doi.org/10.1007/s10291-012-0305-5

Jiang C, Xu B, Hsu LT (2021) Probabilistic approach to detect and correct GNSS NLOS signals using an augmented state vector in the extended Kalman filter. GPS Solut 25(2):72. https://doi.org/10.1007/s10291-021-01101-6

Kumar R, Petovello M (2017) 3D building model-assisted snapshot positioning algorithm. GPS Solut 21(4):1923–1935. https://doi.org/10.1007/s10291-017-0661-2

Leech N, Barrett K, Morgan GA (2013) SPSS for intermediate statistics: use and interpretation. Routledge, New York. https://doi.org/10.4324/9781410616739

Li X, Huang J, Li X, Shen Z, Han J, Li L, Wang B (2022) Review of PPP-RTK: achievements, challenges, and opportunities. Satell Navig 3:28. https://doi.org/10.1186/s43020-022-00089-9

Li L, Elhajj M, Feng Y, Ochieng WY (2023) Machine learning based GNSS signal classification and weighting scheme design in the built environment: a comparative experiment. Satell Navig 4:12. https://doi.org/10.1186/s43020-023-00101-w

Li X, Xu Q, Li X, Xin H, Yuan Y, Shen Z, Zhou Y (2024) Improving PPP-RTK-based vehicle navigation in urban environments via multilayer perceptron-based NLOS signal detection. GPS Solut 28:29. https://doi.org/10.1007/s10291-023-01567-6

Liu Q, Huang Z, Wang J (2019) Indoor non-line-of-sight and multipath detection using deep learning approach. GPS Solut 23(3):75. https://doi.org/10.1007/s10291-019-0869-4

Liu Q, Gao C, Shang R, Peng Z, Zhang R, Gan L, Gao W (2023) NLOS signal detection and correction for smartphone using convolutional neural network and variational mode decomposition in urban environment. GPS Solut 27(1):31. https://doi.org/10.1007/s10291-022-01369-2

Lyu Z, Gao Y (2020) A new method for non-line-of-sight GNSS signal detection for positioning accuracy improvement in urban environments. In: ION GNSS 2020, Institute of Navigation, Virtual, pp 2972–2988. https://doi.org/10.33012/2020.17662

Marais J, Meurie C, Attia D, Ruichek Y, Flancquart A (2014) Toward accurate localization in guided transport: combining GNSS data

and imaging information. Transp Res Part C Emerg Technol 43(2):188–197. https://doi.org/10.1016/j.trc.2013.11.008

Meguro J, Murata T, Takiguchi J, Amano Y (2009) GPS multipath mitigation for urban area using omnidirectional infrared camera. IEEE Trans Intell Transp Syst 10(1):22–30. https://doi.org/10.1109/TITS.2008.2011688

Nitin K (2006) Dimensionality reduction using factor analysis. Dissertation, Griffith university

Siemuri A, Kuusniemi H, Elmusrati MS, Valisuo P, Shamsuzzoha A (2021) Machine learning utilization in GNSS-use cases, challenges and future applications. In: 2021 International conference on localization and GNSS (ICL-GNSS), Tampere, Finland, IEEE, pp 1–6. https://doi.org/10.1109/ICL-GNSS51451.20219452295

Sun Y, Wang J (2022) Mitigation of multipath and NLOS with stochastic modeling for ground-based indoor positioning. GPS Solut 26(2):47. https://doi.org/10.1007/s10291-022-01230-6

Sun R, Wang G, Zhang W, Hsu LT, Ochieng WY (2020) A gradient boosting decision tree based GPS signal reception classification algorithm. Appl Soft Comput 86:105942. https://doi.org/10.1016/j.asoc.2019.105942

Sunirana J, Zornoza JM, Hernández-Pajares M (2013) GNSS data processing. In volume I: fundamentals and algorithms, ESA Communications, Paris, pp 98

Suzuki T, Matsuo K, Amano Y (2020) Rotating GNSS antennas: simultaneous LOS and NLOS multipath mitigation. GPS Solut 24(3):86. https://doi.org/10.1007/s10291-020-01006-w

Tay S, Marais J (2013) Weighting models for GPS Pseudorange observations for land transportation in urban canyons. In: 6th European workshop on GNSS signals and signal processing, Munich, Germany

Won JH, Pany T (2017) Signals processing. In: Teunissen PJG, Montenbruck O (eds) Springer handbook of global navigation satellite systems. Springer, New York, pp 401–442. https://doi.org/10.1007/978-3-319-42928-1

Xin S, Geng J, Zhang G, Ng HF, Guo J, Hsu LT (2022) 3D-mapping-aided PPP-RTK aiming at deep urban canyons. J Geod 96(10):78. https://doi.org/10.1007/s00190-022-01666-1

Yozevitch R, Ben MB, Weissman A (2016) A robust GNSS LOS/NLOS signal classifier. Navig J Inst Navig 63(4):429–442. https://doi.org/10.1002/navi.166

Zhang G, Wen W, Hsu LT (2019) Rectification of GNSS-based collaborative positioning using 3D building models in urban areas. GPS Solut 23(3):83. https://doi.org/10.1007/s10291-019-0872-9

Zhang Z, Li Y, He X, Chen W, Li B (2022) A composite stochastic model considering the terrain topography for real-time GNSS monitoring in canyon environments. J Geod 96(10):79. https://doi.org/10.1007/s00190-022-01660-7

Zhu B, Yang C, Liu Y (2021) Analysis and comparison of three unsupervised learning clustering methods for GNSS multipath signals. Acta Geod et Cartogr Sin 50(12):1762–1771. https://doi.org/10.11947/j.AGCS.2021.20210233

**Linyang Li** is an associate professor at the Information Engineering University, China. He obtained his doctoral degree from the same university in 2019. His research focuses on high-accuracy positioning.



**Luguang Lai** obtained his M.S. degree from the Information Engineering University, China, in 2023, and he is now a Ph.D. candidate. His research focuses on multisensor fusion.



**Zhenban Xu** is a Ph.D. candidate at the Information Engineering University, China. He obtained his B.S. degree from the same university in 2023. His research interests include geodetic data processing and satellite gravimetry applications.



**Yang Shen** is a Ph.D. candidate at the Information Engineering University, China. He obtained his B.S. degree from the same university in 2022. His research is focused on high-accuracy navigation and positioning in polar regions.



**Zhen jia** is an engineer at the Information Engineering University, China. His research focuses on GNSS data processing.