



## Article

# Multipath/NLOS Detection Based on K-Means Clustering for GNSS/INS Tightly Coupled System in Urban Areas

Hao Wang <sup>1,2</sup>, Shuguo Pan <sup>1,2,\*</sup>, Wang Gao <sup>1,2</sup>, Yan Xia <sup>3</sup> and Chun Ma <sup>1,2</sup>

<sup>1</sup> School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China; whao@seu.edu.cn (H.W.); gaow@seu.edu.cn (W.G.); machun@seu.edu.cn (C.M.)

<sup>2</sup> Key Laboratory of Micro-Inertial Instrument and Advanced Navigation Technology, Southeast University, Nanjing 210096, China

<sup>3</sup> School of Geomatics Science and Technology, Nanjing Tech University, Nanjing 211800, China; yanx@njtech.edu.cn

\* Correspondence: psg@seu.edu.cn; Tel.: +86-137-7660-4834

**Abstract:** Due to the massive multipath effects and non-line-of-sight (NLOS) signal receptions, the accuracy and reliability of GNSS positioning solution can be severely degraded in a highly urbanized area, which has a negative impact on the performance of GNSS/INS integrated navigation. Therefore, this paper proposes a multipath/NLOS detection method based on the K-means clustering algorithm for vehicle GNSS/INS integrated positioning. It comprehensively considers different feature parameters derived from GNSS raw observations, such as the satellite-elevation angle, carrier-to-noise ratio, pseudorange residual, and pseudorange rate consistency to effectively classify GNSS signals. In view of the influence of different GNSS signals on positioning results, the K-means clustering algorithm is exploited to divide the observation data into two main categories: direct signals and indirect signals (including multipath and NLOS signals). Then, the multipath/NLOS signal is separated from the observation data. Finally, this paper uses the measured vehicle GNSS/INS observation data, including offline dataset and online dataset, to verify the accuracy of signal classification based on double-differenced pseudorange positioning. A series of experiments conducted in typical urban scenarios demonstrate that the proposed method could ameliorate the positioning accuracy significantly compared with the conventional GNSS/INS integrated navigation. After excluding GNSS outliers, the positioning accuracy of the offline dataset is improved by 16% and 85% in the horizontal and vertical directions, respectively, and the positioning accuracy of the online dataset is improved by 21% and 41% in the two directions. This method does not rely on external geographic information data and other sensors, which has better practicability and environmental adaptability.

**Keywords:** GNSS/INS tightly coupled system; multipath/NLOS detection; K-means clustering algorithm; urban areas



**Citation:** Wang, H.; Pan, S.; Gao, W.; Xia, Y.; Ma, C. Multipath/NLOS Detection Based on K-Means Clustering for GNSS/INS Tightly Coupled System in Urban Areas. *Micromachines* **2022**, *13*, 1128. <https://doi.org/10.3390/mi13071128>

Academic Editors: Shuai Chen, Changhui Jiang, Yuwei Chen and Jianxin Jia

Received: 18 June 2022

Accepted: 15 July 2022

Published: 17 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Global Navigation Satellite System (GNSS) can provide all-day and all-weather global Positioning, Navigation, and Timing (PNT) services for global users, and its positioning errors will not accumulate over time [1,2]. However, satellite signals are frequently blocked or even lose lock in complex urban scenarios, which cannot guarantee the effectiveness of positioning [3]. Inertial Navigation System (INS) has the advantages of strong autonomy and strong anti-interference and can obtain short-term, high-precision navigation and positioning results [4]. However, INS errors accumulate over time, and long-term independent solution can result in reduced accuracy or even divergence [5]. GNSS and INS possess highly complementary characteristics, which can effectively overcome the adverse effects of the harsh environment with the two systems integrated [6]. Therefore, GNSS/INS integrated navigation is widely utilized in naturalistic driving, high-precision vehicle navigation, Intelligent Transportation system (ITS), and autonomous driving [7–10].

Based on the integrated method, GNSS/INS integrated navigation can be divided into loosely coupled, tightly coupled, and deeply coupled. Among them, the GNSS/INS tightly coupled system can maintain continuous positioning despite the insufficiently visible satellites and has the advantage of simple integrated structure and easier implementation, which has been extensively recognized by many scholars [11–13].

In the application of vehicle dynamic positioning, a continuous, reliable, and high-precision positioning method is urgently required. Balsa-Barreiro (2013) implemented an innovative methodology based on vehicle speed for the geo-referencing naturalistic driving method, which can overcome the problems related to the lack of positioning data [14]. However, this method requires extra geographic information. In addition, some scholars use GNSS and multiple sensors for positioning, which can integrate the advantages of each sensor, and obtain continuous and reliable positioning information in urban scenes [15–17]. But, the cost of multi-sensor is high, the weight is difficult to determine, and the amount of calculation is large. Therefore, based on the above discussion, this paper chooses GNSS/INS tightly coupled system positioning as the method to obtain the vehicle position in urban scenarios. However, the performance of GNSS/INS tightly coupled positioning is inevitably plagued by various outliers in the complex urban scenario [18]. For the GNSS receiver, the multipath effect and NLOS signal are the commonly adverse factors to restrict the positioning performance of GNSS, which further subsides the GNSS/INS integration system positioning capability [19,20]. Multipath effect occurs when a signal is received through multiple paths between the satellite and the receiver antenna. Additionally, multipath interference can affect the phase detection characteristics of the receiver tracking loop, resulting in tracking and measurement errors [21]. Multipath includes both direct and reflected signals, and the reflected signals can be multiple. NLOS signal reception occurs when the direct path from the satellite to the receiver is blocked and the signal can only be received through the reflected path [22]. Due to the relatively low cost and the ability to provide accurate time reference and absolute coordinate information, GNSS solution is still an important and preferred technical method in the field of high-precision navigation and location services. Unfortunately, the degradation of GNSS positioning accuracy may negatively affect the performance of the entire system [23]. We need to mitigate the error of the standalone GNSS positioning with innovative signal-processing methods to promote the performance of GNSS/INS integration.

Accurate multipath/NLOS detection and subsequent processing are the basis for the signal quality control and positioning strategy optimization of the GNSS/INS integration Navigation system [24]. With the prosperity of artificial intelligence technology, more and more scholars began to employ machine learning or deep-learning methods to identify multipath/NLOS signals. These methods constructed the mapping relationship between multiple feature parameters and GNSS signal categories to reduce the operating cost of traditional methods and enhance the availability of the algorithm, which have achieved outstanding results [25–28]. The above approaches are classifiers based on supervised learning that need to label the training samples in advance. Meanwhile, the accuracy of the labeling is directly associated with the performance of the classifiers. In practical applications, it is challenging and expensive to obtain an accurately labeled dataset that covers multiple scenarios and represents all state types. In order to refrain from the limitations of signal classification methods caused by the above problems, scholars have begun to use unsupervised learning techniques for signal classification research, which can sufficiently excavate the information of the observation data itself [29–32]. Compared with the supervised-learning multipath/NLOS signal detection algorithm, the unsupervised technology has better advantages in availability and environmental applicability. Furthermore, the positioning accuracy after excluding contaminated GNSS satellites is significantly better than the traditional threshold method and the classic RAIM algorithm.

Given that the multipath/NLOS signal restricts the dynamic positioning accuracy and reliability of the GNSS/INS integrated system, apart from GNSS alone positioning, multipath/NLOS detection and mitigation for GNSS/INS integrated systems were developed in

past years [33–35]. However, these technologies are dependent on geographic information data or external sensor equipment, such as 3D building models, cameras, and LiDAR, which have a certain level of deficiencies in terms of availability, cost, and security. In addition, the equivalent weight model is employed to construct robust algorithms that can weaken the influence of the gross error of the observation on the positioning accuracy [36,37]. This method utilizes the robust factor to adjust the filter gain moment or the observation noise for GNSS/INS integrated positioning, which plays a role in suppressing multipath and NLOS errors to a certain extent. However, the robust estimation algorithm has difficulty handling multiple outliers on the same epoch and relies heavily on the correctness of the robust model. Therefore, when the original observations are seriously polluted by the multipath/NLOS signals in harsh environments, the reliability of the algorithm cannot be guaranteed.

This paper aims to further promote the accuracy of GNSS/INS tightly coupled positioning results by using unsupervised techniques to detect multipath/NLOS signals. A clustering algorithm is utilized to label GNSS data in offline system as normal and abnormal observations, the latter mainly caused by multipath/NLOS signals. The clustering criteria obtained by offline dataset training are applied to detect multipath/NLOS signals for online data, which can enhance the performance of GNSS/INS real-time positioning. Additionally, it can provide an innovative perspective for the research on GNSS signal quality control methods of vehicle positioning systems in highly complex urban areas.

The rest of this paper is organized as follows: In Section 2, the relevant mathematical methodology is presented for K-means Clustering and GNSS/INS tightly coupled positioning algorithm. Next, Section 3 implements the data collection and experiment analysis, which validate the accuracy and reliability improvements of the proposed method. Finally, the conclusion and outlook will be given in Section 4.

## 2. Methodology

### 2.1. Feature Extraction

Reasonable feature value is critical to the capability of machine-learning algorithms, and this paper refers to the feature parameters of multipath/NLOS signal detection in supervised learning classifiers. Most of the current machine learning methods for multipath/NLOS detection adopt feature values at the observation data level. We only extract the feature parameters from the RINEX format file output by the GNSS receiver, including pseudorange, carrier phase, carrier-to-noise ratio (signal strength), and Doppler frequency shift, etc., which are all closely related to GNSS signal types. However, it is impossible for any single feature to effectively classify GNSS signals. Hence, a combination of different features is needed to ameliorate the classification accuracy [22,38–40].

(1) Satellite elevation angle: It is a common method to assign weights to each observation value based on the satellite elevation angle to reduce the influence of multipath and NLOS signal reception on the positioning results. Generally speaking, satellite signals from high elevation angles are less likely to be blocked and reflected by buildings, but this is not always the case in reality. Due to the height and distribution of buildings in urban areas, satellite signals at high elevation angles may also be NLOS signals, while signals at low elevation angles may be direct signals. Nonetheless, satellite elevation angle is still an important feature indicator to distinguish NLOS signals.

(2) Carrier-to-noise ratio: The GNSS receiver will output the observations of the tracked satellite signal strength. According to the signal propagation theory, supernumerary propagation and reflection will increase the path loss of the GNSS signal. As an important indicator reflecting the signal quality, the C/N<sub>0</sub> observation value is also a common parameter to alleviate the multipath effect. Similar to the elevation angle, the satellite signal strength or carrier-to-noise ratio also has a certain correspondence with the type of signal. The signal strength of the satellite received by the survey antenna is usually higher in an open environment. However, the magnitude of C/N<sub>0</sub> does not have a clear correspondence

with the type of GNSS signal in a multipath environment, because constructive multipath will increase the received signal, while destructive multipath reduces signal strength.

(3) Pseudorange residual: When there are more observation equations than unknown parameters and the position estimation is accurate enough, the magnitude of the pseudorange residual can reflect the inconsistency between the pseudorange measurements and the geometric distance of the satellite.

In addition, multi-constellation GNSS integrated positioning increases the number of available observation satellites and observation redundancy. Therefore, the pseudorange residual can be used as an indicator to detect the quality of GNSS signals.

(4) Pseudorange rate consistency: The pseudorange observations originate from the receiver code tracking loop, and the Doppler shift of the signal is determined by the receiver frequency tracking loop. Compared with the code tracking loop, the multipath/NLOS signal has less influence on the frequency tracking loop, so the consistency between the pseudorange change rate and the Doppler frequency shift can reflect the interference degree of the reflected signal. Its formula is expressed as:

$$\zeta = |\Delta\rho - \dot{\rho} \cdot \Delta t| \quad (1)$$

where  $\Delta\rho$  and  $\Delta t$  represent the pseudorange variation and time interval between adjacent epochs, respectively. According to the Doppler effect, the pseudorange rate  $\dot{\rho}$  is calculated from the Doppler shift.

$$\dot{\rho} = -\lambda_i \cdot f_{D_i} \quad (2)$$

where  $\lambda_i$  and  $f_{D_i}$  indicate the wavelength of frequency  $i$  and the Doppler shift in Hz, individually.

Since all of the above single features are uncertain and interdigitated with each other for NLOS signals, it is impossible for any single feature to effectively classify GNSS signals. Thus, NLOS signals need to be determined by a combination of different features. In summary, this paper comprehensively selects the above four parameters to form the feature vector of cluster analysis. Then the data are standardized to eliminate the influence of different dimensions on the clustering results, that is, each feature value conforms to the standard normal distribution after data processing.

## 2.2. K-Means Clustering Algorithm and Its Evaluation Indicator

It is considered that GNSS signals in complex environments are generally divided into two main types: direct signals and indirect signals (including multipath and NLOS signals), and each type of signal has a certain internal relationship with the above four feature parameters.

In accordance with this characteristic, the K-means algorithm is used for signal clustering. When the sample is closest to one of the cluster centers, it is classified into this class.

For a given sample set  $S = \{x_1, x_2, \dots, x_m\}$ , where  $x_m = \{e_m, CN0_m, v_m, \dots, con_m\}$  represents the standardized feature vector of satellite elevation angle, carrier-to-noise ratio, pseudorange residual, and pseudorange rate consistency. This paper assigns corresponding weights to the feature parameters based on experience [22], which are set as  $[\sqrt{0.2}, \sqrt{0.3}, \sqrt{0.2}, \sqrt{0.3}]$ , respectively. The K-means algorithm divides them into  $k$  clusters  $\bar{C} = \{C_1, C_2, \dots, C_k\}$  so that the sum of squared Euclidean distances from each data point to its nearest cluster center is minimized, namely:

$$\operatorname{argmin}_C \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (3)$$

where  $\mu_i$  is the mean vector of the cluster  $C_i$ , and can be expressed as:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (4)$$

The basic process of the K-means algorithm is as follows [41]:

- (1) This algorithm randomly selects  $k$  samples as the initial cluster center;
- (2) For the remaining samples, according to the distance of their cluster centers, they are classified into the nearest cluster;
- (3) For each cluster, the mean of all samples is calculated as the new cluster center;
- (4) Repeat steps (2) and (3) until the cluster centers no longer change.

Based on the above calculations, all GNSS signals are classified into different clusters. The K-means algorithm needs to specify the value of  $k$  in advance, which is usually defaulted to 2 or 3. However, due to the complexity of the scenario and the correlation between GNSS signals, 2 and 3 are often not the optimal  $k$  values. Additionally, the difference between the clustering results corresponding to different  $k$  values is not obvious. Therefore, this paper chooses the Davies–Bouldin Indicator (DBI) as the internal evaluation index of the clustering effect [42].

DBI is defined as the average similarity between each cluster  $C_i, i = 1, 2, \dots, k$  and its the most similar one  $C_j$ , where the similarity is expressed by the ratio of the intra-cluster distance to the inter-cluster distance. The minimum value of DBI is 0, and the smaller the value, the better the clustering effect. The specific calculation formula is:

$$I_{\text{DBI}} = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad (5)$$

where  $k$  is the number of clusters;  $R_{ij}$  denotes similarity that can be constructed by a simple choice as follows so that it can keep nonnegative and symmetric:

$$R_{ij} = \frac{d(s_i) + d(s_j)}{d_{ij}} \quad (6)$$

where  $d(s_i)$  and  $d(s_j)$  mean the average distance between each point of cluster data to the centroid of that cluster also known as cluster diameter, individually;  $d_{ij}$  is the distance between cluster centroids  $i$  and  $j$ , which represents the dispersion degree of data point for cluster centroids  $i$  and  $j$ .

### 2.3. GNSS/INS Tightly Coupled Positioning Model

GNSS/INS integrated navigation system adopts Extended Kalman Filter (EKF) for system fusion to realize high-precision navigation and positioning by effectively detecting and rejecting Multipath/NLOS signals in complex urban areas. In vehicle navigation, due to the strong reliability and highly real-time performance of the pseudorange/INS system, this paper employs the integrated positioning solution of GNSS double-differenced pseudorange (DGNSS) and INS observation.

The system state model depends on the INS error model and the description of the inertial sensor system error. The INS error equation based on the psi angle is adopted in this paper [43].

$$\delta \mathbf{r}^n = -\boldsymbol{\omega}_{en}^n \times \delta \mathbf{r}^n + \delta \mathbf{v}^n \quad (7)$$

$$\delta \mathbf{v}^n = -(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \delta \mathbf{v}^n - \boldsymbol{\psi}^n \times \mathbf{f}^n + \delta \mathbf{g}^n + \mathbf{C}_b^n \delta \mathbf{f}_{ib}^b \quad (8)$$

$$\dot{\boldsymbol{\psi}}^n = -(\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \boldsymbol{\psi}^n - \mathbf{C}_b^n \delta \boldsymbol{\omega}_{ib}^b \quad (9)$$

where  $\delta \mathbf{r}^n$ ,  $\delta \mathbf{v}^n$ , and  $\delta \boldsymbol{\psi}^n$  indicate position error, velocity error, and attitude angle error, respectively;  $\mathbf{C}_b^n$  is the rotation matrix from the body frame ( $b$ -frame) to the navigation frame ( $n$ -frame);  $\delta \mathbf{f}_{ib}^b$  and  $\delta \boldsymbol{\omega}_{ib}^b$  are the accelerometer and gyroscope error vector in the  $b$ -frame, separately; In addition, the specific force vector measured by the accelerometer, the rotation velocity of the earth, and the transfer rate are represented by  $\mathbf{f}^n$ ,  $\boldsymbol{\omega}_{ie}^n$ , and  $\boldsymbol{\omega}_{en}^n$ , respectively.

The accelerometer error and gyroscope error are the main factors affecting the accuracy of GNSS/INS tightly coupled system, and the bias errors are modeled by a random walk process. Their specific forms can be expressed as:

$$\begin{cases} \dot{\mathbf{b}}_a = w_{ba} \\ \dot{\mathbf{b}}_g = w_{bg} \end{cases} \quad (10)$$

where  $\mathbf{b}_a$  is bias of the accelerometer;  $\mathbf{b}_g$  is the bias of the gyroscope;  $w_{ba}$  and  $w_{bg}$  express the corresponding random white noise.

The equation of the state of the system is as follows:

$$\dot{\mathbf{X}}_{ins} = \mathbf{F} \cdot \mathbf{X}_{ins} + \mathbf{G} \cdot \mathbf{W} \quad (11)$$

where  $\mathbf{F}$  is the state transition matrix;  $\mathbf{G}$  and  $\mathbf{W}$  represent the dynamic noise matrix and the noise vector;  $\mathbf{X}_{ins}$  is the state parameter,  $\mathbf{X}_{ins} = (\delta\mathbf{r}, \delta\mathbf{v}, \delta\boldsymbol{\psi}, \mathbf{b}_g, \mathbf{b}_a)_{15 \times 1}$ .

Based on the INS error model,  $\mathbf{G}$  and  $\mathbf{W}$  are optimized and presented as follows:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{rr} & \mathbf{F}_{rv} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{F}_{vr} & \mathbf{F}_{vv} & \mathbf{f}^l \times & \mathbf{0} & \mathbf{C}_b^n \\ \mathbf{F}_{\psi r} & \mathbf{F}_{\psi v} & -(\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times & -\mathbf{C}_b^n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_b^n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{C}_b^n & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (12)$$

where  $\mathbf{I}$  indicates the unit matrix;  $\mathbf{F}_{rr}$  and  $\mathbf{F}_{rv}$  represent state coefficients of position;  $\mathbf{F}_{vr}$  and  $\mathbf{F}_{vv}$  represent coefficients of velocity;  $\mathbf{F}_{\psi r}$  and  $\mathbf{F}_{\psi v}$  represent coefficients of attitude. The specific derivation of the above symbols can be found in [33].

The difference between the distance from the satellite to the ground station predicted by the INS and the GNSS double-differenced pseudorange is solved, which is used as the EKF measurement to achieve high-precision positioning for GNSS/INS tightly coupled system. The measurement equation is written in matrix form:

$$\mathbf{Z}_k = \mathbf{H}_k \mathbf{X}_k + [e_\rho, e_{ins}]^T \quad (13)$$

where  $\mathbf{Z}_k$  represents the measurement vector at time epoch  $k$ ,  $\mathbf{Z}_k = \Delta\nabla\rho^* - \Delta\nabla\rho_{ins}$ , and  $\Delta\nabla\rho_{ins}$  indicates the satellite-to-ground distance predicted by INS; “\*” represents different satellite systems including uses GPS and BDS in this paper;  $\mathbf{H}_k$  is the measurement model coefficient matrix;  $e_\rho$  and  $e_{ins}$  are the pseudorange observation noise and INS observation noise, respectively.

The final GNSS/INS tightly coupled positioning results can then be solved based on the following EKF procedures.

Prediction stage:

$$\hat{\mathbf{X}}_{k/k-1} = \boldsymbol{\Phi}_{k,k-1} \hat{\mathbf{X}}_{k-1} \quad (14)$$

$$\mathbf{P}_{k/k-1} = \boldsymbol{\Phi}_{k,k-1} \mathbf{P}_{k-1} \boldsymbol{\Phi}_{k,k-1}^T + \mathbf{Q}_{k-1} \quad (15)$$

Update stage:

$$\mathbf{K}_k = \mathbf{P}_{k/k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k/k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (16)$$

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_{k/k-1} + \mathbf{K}_k (\mathbf{Z}_k - \mathbf{H}_k \hat{\mathbf{X}}_{k/k-1}) \quad (17)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k/k-1} \quad (18)$$

where  $\hat{\mathbf{X}}_k$ ,  $\boldsymbol{\Phi}_k$ , and  $\mathbf{P}_k$  express the state vector estimates, the state transition matrix, and the error covariance matrix at time epoch  $k$ , respectively;  $\mathbf{Q}_{k-1}$  represents the system noise covariance matrix at time epoch  $k-1$ ;  $\mathbf{R}_k$  indicates the measurement noise covariance matrix at time epoch  $k$ ;  $\mathbf{H}_k$  denotes the measurement matrix at time epoch  $k$ ;  $\mathbf{K}_k$  represents the EKF

gain matrix at time epoch  $k$ ; In addition,  $\mathbf{A}_{k,k-1}$  represents matrix/vector propagation from time epoch  $k-1$  to  $k$ .

2.4. Overview of the Proposed Method

The flowchart of the proposed method is shown in Figure 1. Firstly, four essential features are extracted from GNSS raw observation data, namely, satellite elevation angle, carrier-to-noise ratio, pseudorange residual, and pseudorange rate consistency, which are comprehensively used to enhance the classification accuracy. Secondly, satellite signals received by GNSS receivers in complex scenarios are generally divided into two categories: direct and indirect signals, the latter including multipath and NLOS signals, and each type of signal has a certain internal relationship with the above four main features. According to this characteristic, this paper adopts the k-means clustering algorithm to cluster the signals and selects the DBI as the internal evaluation index of the clustering effect. Finally, both the raw INS measurements and the GNSS double-differenced pseudorange observations are tightly integrated with EKF filtering, resulting in reliable and high-precision positioning results.

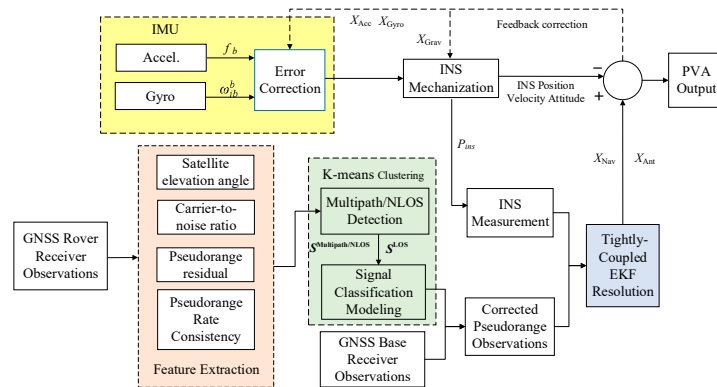


Figure 1. Flowchart of the proposed Multipath/NLOS Detection method based on K-means Clustering for GNSS/INS tightly coupled system.

It should be pointed out that the measured circumstance of the GNSS receiver of the base station is open and superior with no obstacle occluding all around as shown in Figure 2. Additionally, the Trimble GNSS-Ti earth-type chock-ring antenna is installed. This paper purports that there is no NLOS signal and the multipath is well suppressed, which is small relative to the rover and does not affect the subsequent positioning solution. Therefore, multipath/NLOS signal detection for base station GNSS observations is not required.

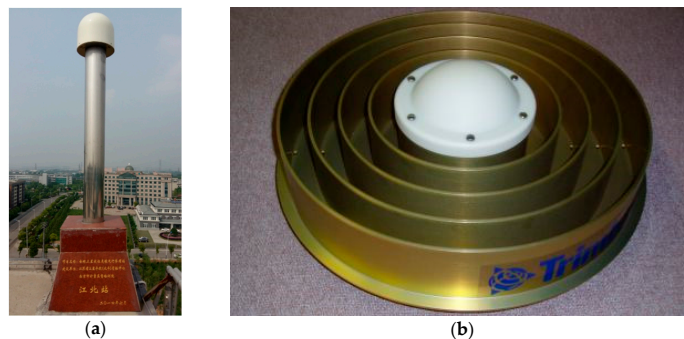


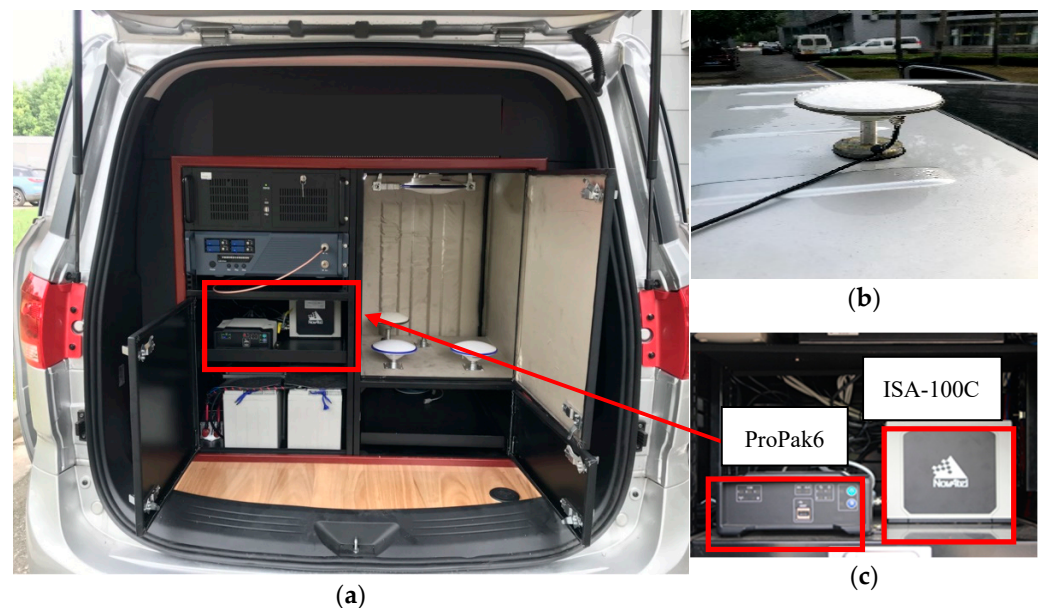
Figure 2. Observation environment of the base station and GNSS receiver: (a) The base station; and (b) Trimble GNSS-Ti earth-type chock-ring antenna.

3. Data Collection and Experiment Analysis

3.1. Data Collection

This paper takes advantage of GPS L1 and BDS B1 frequency observation data, and the sampling rate is 1 Hz. The reference values of the three-dimensional position, velocity, and

attitude of the vehicle are acquired through GNSS RTK/INS tightly coupled positioning, which is used to verify the improvement of positioning accuracy after excluding outliers. The NovAtel ProPak6 receiver of the rover station and the Trimble receiver of the three static base stations provide the GNSS carrier phase observations. Moreover, the solution of the reference value is realized using Inertial Explorer post-processing software developed by NovAtel Company [44]. The experimental platform is shown in Figure 3, and Table 1 is the specific technical parameters of the IMU.



**Figure 3.** The test platform and equipment of the experimental vehicle: (a) The measurement vehicle; (b) satellite signal receiving antenna; (c) integrated system sensors.

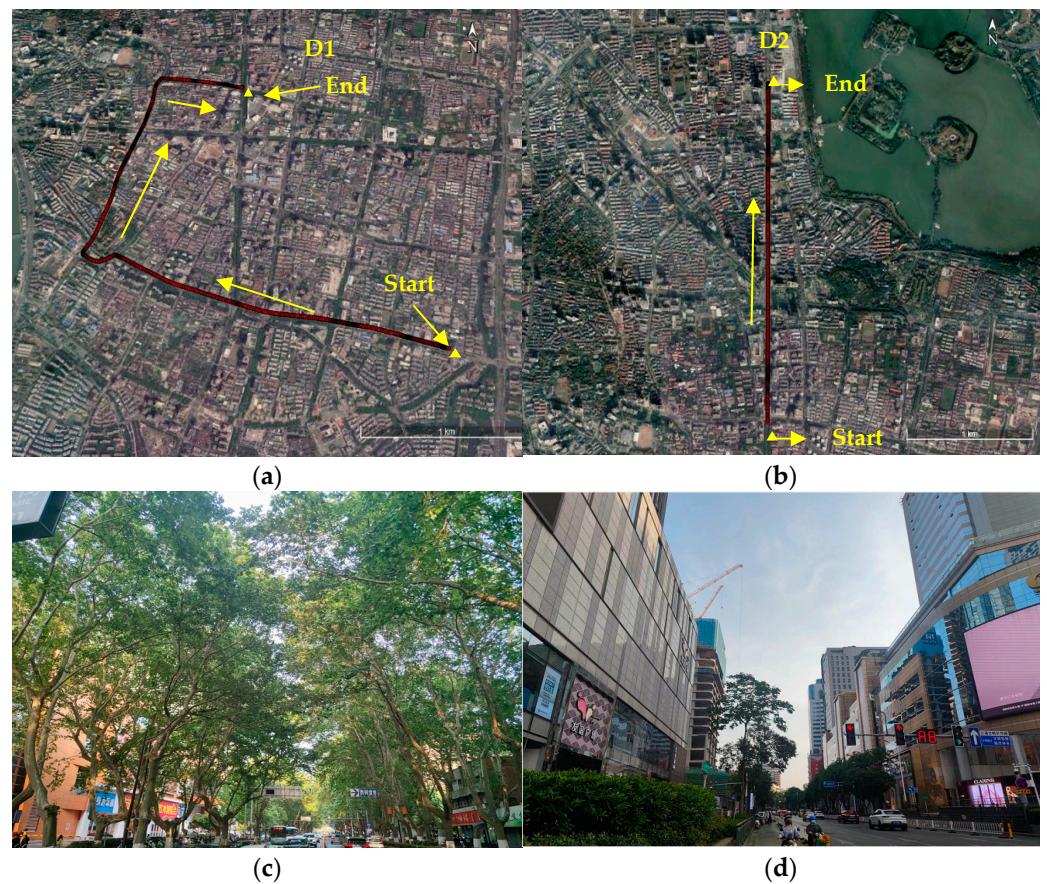
**Table 1.** IMU technical data.

Parameters	Gyroscope	Accelerometer
Bias stability	0.5 deg/h	1250 $\mu$ g
Scale factor	100 ppm	100 ppm
Random walk	0.3deg/ $\sqrt{h}$	100 $\mu$ g / $\sqrt{h}$

The experiment was carried out in the urban area of Nanjing, where the typical scenarios included tree shade and urban canyons, etc. Figure 4 shows the vehicle trajectories corresponding to two different observation periods. The two sets of datasets are marked as D1 and D2 in the order of time. D1 is an offline dataset, and D2 is an online dataset for real-time positioning verification. In addition to the occlusion of satellite signals caused by obstacles such as huge buildings, the strong reflection effect of modern building materials can also cause serious multipath and NLOS signal reception. Furthermore, the lush tree canopy on both sides of urban roads can also lead to complex multipath effects.

The specific information of the dataset is displayed in Table 2. The number of valid epochs refers to the number of epochs for which the position solution is obtained. Owing to the influence of the observation environment, some epochs do not have GNSS data output, or the number of observation satellites is too small to conduct the double system positioning solution. Therefore, these epochs will be considered to be invalid epochs.





**Figure 4.** The route map of the actual measurement of the urban complex scenarios: (a) the route maps of offline dataset D1 for vehicle dynamic testing; (b) the route maps of online dataset D2 for vehicle dynamic testing; (c) the boulevard; (d) the urban canyon.

**Table 2.** Valid epoch and sample size of each dataset.

Dataset	Start Time (UTC)	End Time (UTC)	Valid Epoch	Sample Size
D 1	2017-04-20 05:12:58	2017-04-20 05:31:43	1052	11,454
D 2	2017-04-20 05:31:44	2017-04-20 05:41:28	531	6011

### 3.2. Outliers Detection for the Offline Data Based on K-Means

This section will discuss in detail how to use the K-means algorithm to detect GPS/BDS multipath/NLOS signals on dataset D1, which employs GNSS/INS tightly coupled post-processing algorithm to verify its effectiveness. This lays the foundation for the subsequent real-time application for GNSS/INS integrated navigation.

In order to broaden the feature value range of the sample and upgrade the performance of the machine learning algorithm, the cut-off elevation angle and the signal-to-noise ratio are not set during the positioning process. The number of epochs that do not satisfy the chi-square test accounts for about 10% of the valid epochs, indicating that the observation environment has a certain complexity. The epochs that satisfy the chi-square test still contain a certain number of satellite observations disturbed by multipath and NLOS signals. Therefore, identifying them accurately is the key to enhancing the performance of GNSS positioning and GNSS/INS tightly coupled positioning. Figure 5 illustrates the sky map of GPS and BDS observed by the receiver during vehicle driving.

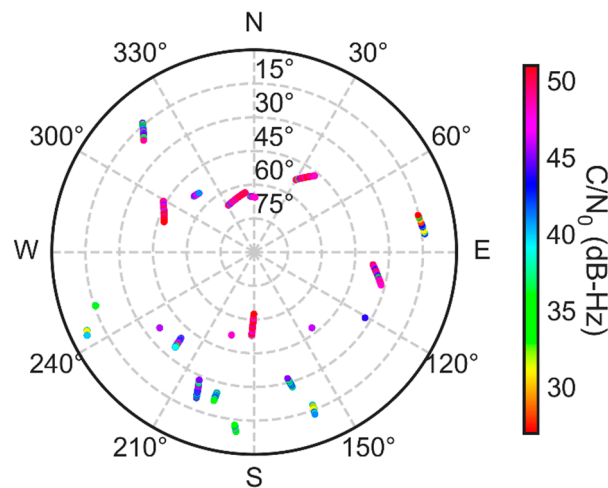


Figure 5. GPS/BDS sky map and observation signal-to-noise ratio.

### 3.2.1. Cluster Analysis

Since only four typical feature parameters are used as the input of the machine learning model, there is no need to use methods such as Principal Components Analysis (PCA) for dimension reduction of the sample data. The feature value is normalized by a z-score to ensure that they are in the same order of magnitude. Here, different weights are assigned to the four feature parameters elaborated above for clustering calculation. Due to the possible connectivity and correlation between different GNSS signal types and the potential non-integrity of feature parameters, we will no longer constrain the value of  $k$  to 3 or 2, but judge according to the actual situation.

Figure 6 demonstrates the corresponding DBI values for different values of  $k$ . It can be revealed that when the value of  $k$  is 3, 4, 5, and 7, the DBI value is small with inconspicuous disparity. Hence, this paper selects the number of clusters from these four groups of values.

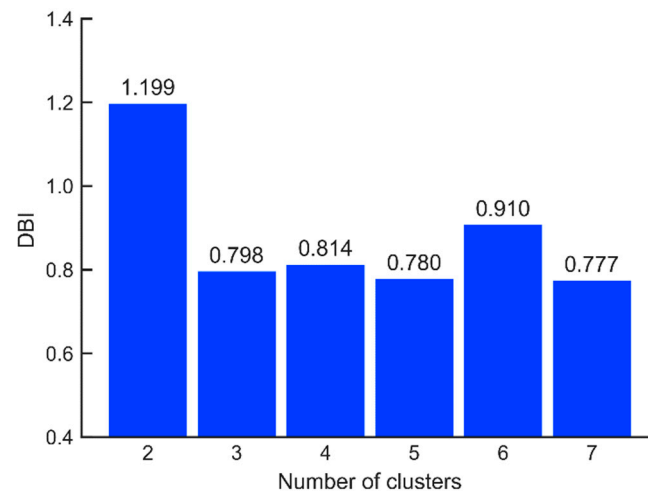
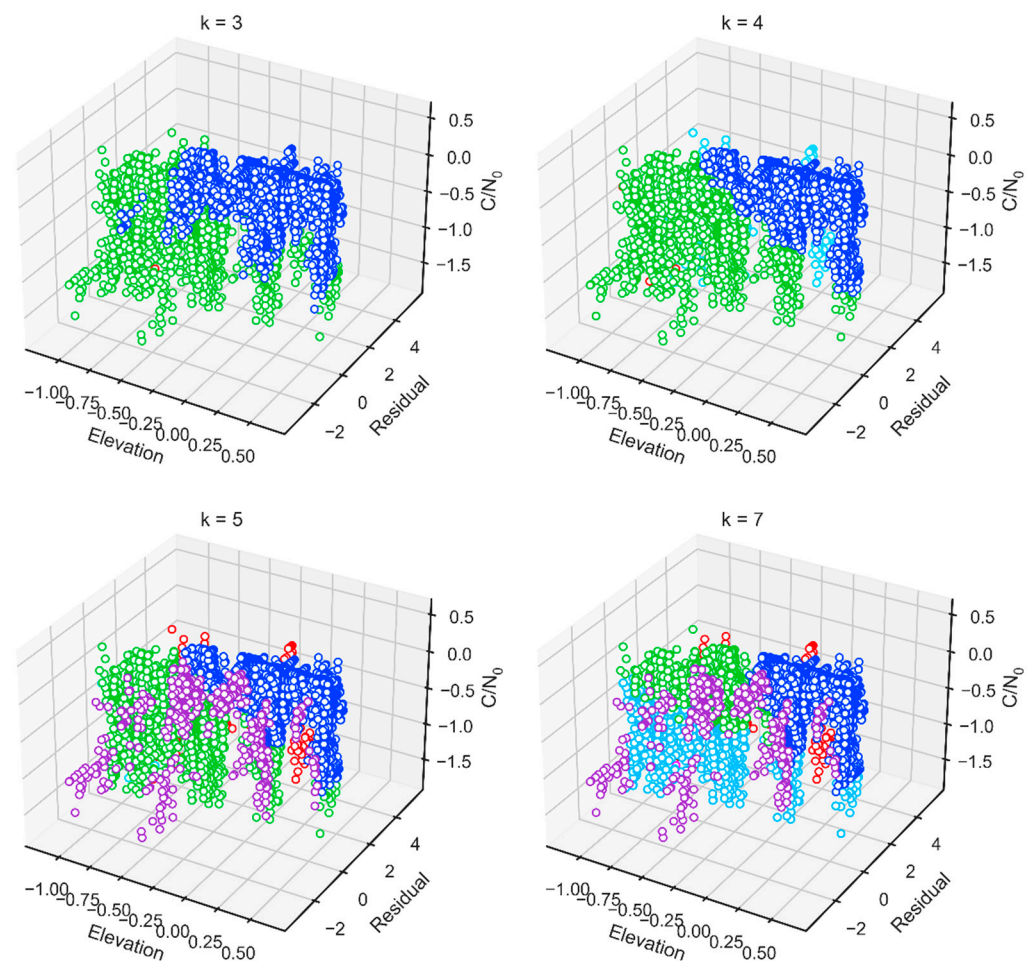


Figure 6. DBI values of clustering results correspond to different  $k$  values.

It is worth noting that due to the incompleteness of feature parameters, the complexity of environmental influences and the fuzzy correlation between signal types, the number of clusters do not necessarily correspond strictly to the type of signal (three types of LOS/Multipath/NLOS); that is, 3 is not necessarily the optimal value of  $k$ . Here only the value of DBI is used as a reference for us to determine the value of  $k$ .  $k = 3$  is great, but if the DBI value of  $k > 3$  is also low, we will also consider it. In this case, we can think of them as clusters with varying degrees of signal interference. Due to the existence of inertial measurement data in tightly combined positioning, we do not need to care too much about

the perfect value of  $k$  as in GNSS alone positioning for the time being. A slightly larger division of GNSS signals into multipath /NLOS will not cause the underdetermination of the tightly combined observation equation.

Figure 7 shows the visualization graphics of the clustering results corresponding to different  $k$  values in the three-dimensional feature space. In general, the larger the value of  $k$ , the smaller the number of observations in the dataset that are not disturbed by multipath and NLOS signals. For example, when  $k$  is 3, we know from observation in Figure 7, and from experience, that the number of line-of-sight (LOS) observations is 7807, and the number of samples of the other two clusters is 3646 and 1. When  $k$  is 7, the number of LOS observations is reduced to 6563.



**Figure 7.** Clustering effect illustration based on different  $k$  values.

Since the larger  $k$  is, the smaller is the number of LOS observations, if only the LOS observations are retained during the positioning process, the risk of  $k = 3$  lies in missed detection, and the false detection rate will increase when  $k = 7$ . Considering that there is no extraordinary explicit boundary between the observation data of different characteristics, this paper chooses a relatively compromised method that the value of  $k$  is set to 5. This way can alleviate the probability that observations are contaminated by multipath and NLOS, which are mixed into the LOS clusters during the clustering process; thus, it can avoid a reduction in the accuracy of GNSS/INS tightly coupled positioning. Furthermore, based on the existence of INS observation data, positioning results can be output, despite an insufficient number of LOS satellites. At this time, the number of samples of LOS observations is 7286.

Figure 8 further expresses the distribution pattern of the clustering results under four different feature parameters when  $k = 5$ . The diagonal represents the probability

distribution of the sample points of different clusters on the variable. Additionally, the off diagonal represents the horizontal distribution of the sample points on the corresponding two-dimensional feature.

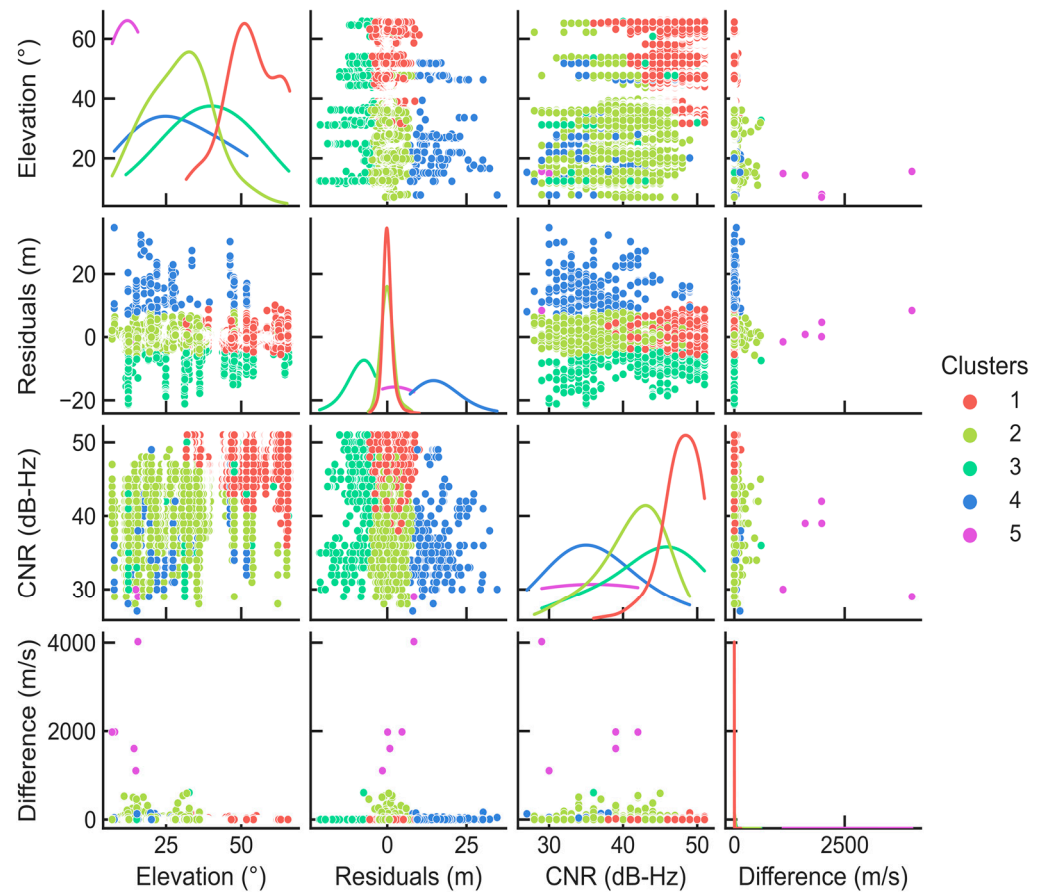
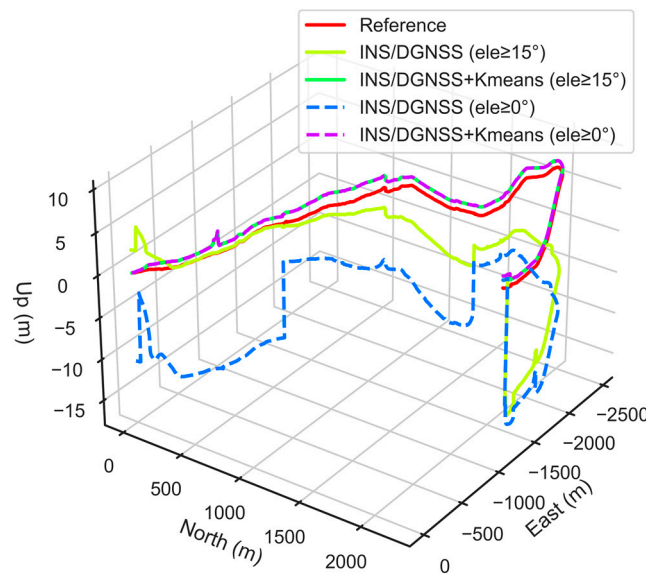


Figure 8. Four feature value distribution pattern diagrams when the  $k$  value is 5.

It is worth noting that here for better visualization, the probability-density curves of each cluster are independently normalized so that their respective acreage under each curve is 1. According to the probability distribution, it can be inferred that cluster 1 in Figure 8 is the LOS observation sample. It can be seen that its main distribution area is concentrated in the space with the maximum satellite elevation angle and carrier-to-noise ratio, the smallest pseudorange residual, and the consistency of the pseudorange rate, in which the elevation angle of the satellites is basically above 25°. This is highly similar to the nature of LOS observations in urban environments. The remaining four clusters can be considered to be different distribution patterns of the sample set composed of observations contaminated by multipath or NLOS signals.

### 3.2.2. Analysis of Positioning Results

In order to validate the correctness of the clustering results, we execute the DGNSS/INS tightly coupled positioning algorithm to conduct comparative experiments. Figure 9 demonstrates the comparison of the 3D position of the vehicle obtained by different positioning methods, where the red solid line is the ground reference value of the vehicle on the ground. This paper takes the starting point of the vehicle as the origin of the site-centric coordinate system. From Figure 9, we can recognize that the tightly coupled positioning result without multipath/NLOS observation processing deviates more from the reference coordinate value. Furthermore, the result without the height cut-off elevation angle has the most serious position deviation. The deviation in the starting position from the reference value causes an adverse influence on the positioning accuracy.



**Figure 9.** 3D position comparison of positioning results of different tightly coupled schemes.

Since the multipath/NLOS observations are not eliminated in complex urban scenarios, it makes the tightly coupled observation equation unreliable. It can also be concluded that installing the satellite cut-off elevation angle can suppress the influence of multipath/NLOS observations to a certain extent. However, the overall positioning performance of the GNSS/INS tightly coupled system is still unsatisfactory, especially in the vertical direction. However, after using the K-means clustering algorithm to identify and exclude multipath/NLOS observations, the positioning results are significantly ameliorated, and the deviation level from the true value becomes small.

Moreover, even if the satellite cut-off elevation angle is set to  $0^\circ$ , the position solution of this method hardly changes. This phenomenon indicates that the cut-off elevation angle of most LOS observations in the scenario identified by the clustering algorithm are above  $15^\circ$ . However, it does not mean that all satellite observations above  $15^\circ$  come from LOS signal reception.

The above experiments also reveal that simply relying on setting the cut-off elevation angle cannot improve the positioning accuracy of GNSS/INS tightly coupled system, and even destroy its satellite geometry, which is proved and expounded in detail by Xia et al. (2020). The LOS signal identification method proposed in this paper sufficiently considers the influence of the four feature parameters on the properties of GNSS signals with higher reliability.

Furthermore, this paper calculates the RMS of the position error for the above positioning method, as shown in Table 3. The experimental results show that the positioning accuracy is 0.63 m in the horizontal direction and as high as 6.50 m in the vertical direction without multipath/NLOS observation processing (the cut-off elevation angle is set to  $15^\circ$ ).

**Table 3.** Accuracy comparison of positioning results for different tightly coupled schemes.

Positioning Scheme	RMS/m			
	E	N	U	2D
INS/DGNSS (ele $\geq 15^\circ$ )	0.33	0.53	6.50	0.63
INS/DGNSS (ele $\geq 0^\circ$ )	0.32	2.75	11.57	2.76
INS/DGNSS + K-means (ele $\geq 15^\circ$ )	0.37	0.38	0.98	0.53
INS/DGNSS + K-means (ele $\geq 0^\circ$ )	0.38	0.38	0.98	0.53

After excluding the multipath/NLOS observations, the GNSS/INS tightly coupled positioning accuracy is improved by 16% and 85% in the horizontal and vertical directions,

respectively. The horizontal accuracy reaches 0.53 m, and the vertical accuracy is specially promoted to within 1 m. It is particularly acknowledged that if using the K-means algorithm to eliminate multipath and NLOS observations, we do not need to set the cut-off elevation angle, because the algorithm has taken into account the factor that the low elevation angle is susceptible to multipath/NLOS signal interference. Additionally, compared to the accuracy of the positioning result in the horizontal direction, the accuracy in the vertical direction tends to be substantially promoted as the multipath/NLOS signals are eliminated. This is because in the harsh urban scenarios, the positioning results in the vertical direction are more severely affected by multipath/NLOS signals than the positioning results in the horizontal direction, which is elaborated upon by Sun et al. (2022).

To more clearly show the contribution of K-means clustering to the positioning results, we implemented GNSS pseudorange double-differenced positioning for the experiments. Figure 10 shows the comparison of positioning trajectories in several typical real urban scenarios (skyscrapers and trees), which demonstrates that the results of GNSS/INS tightly coupled position are significantly better than GNSS. Further, the accuracy of GNSS pseudorange differential positioning has been greatly promoted based on K-means to exclude multipath/NLOS observations. With the addition of INS observation data, not only the accuracy of the positioning consequence is further improved, but also possesses better continuity and availability.

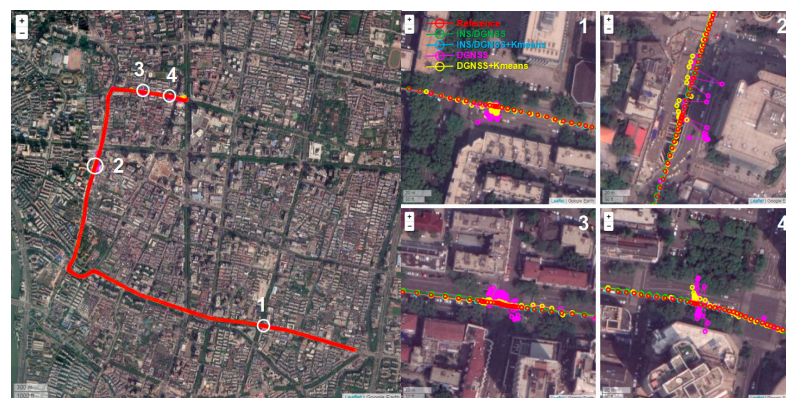


Figure 10. Comparison of positioning results in typical occlusion environments.

Figures 11 and 12 show the comparison of the number of satellites participating in the position calculation and the PDOP value before and after excluding the multipath/NLOS observations, respectively. By comparison, it is known that after excluding satellites, the number of effective satellites is expected to decrease significantly, which causes the geometric distribution of satellites to be more deteriorated. This phenomenon indicates that the experimental environment is relatively harsh, and multipath/NLOS satellites account for a considerable proportion; yet, despite this, the positioning accuracy has been greatly ameliorated, which shows that the multipath/NLOS signal detection algorithm in this paper is effective.

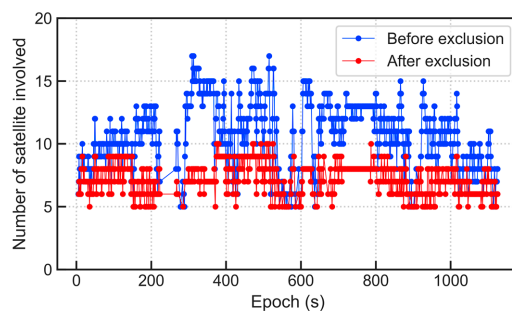
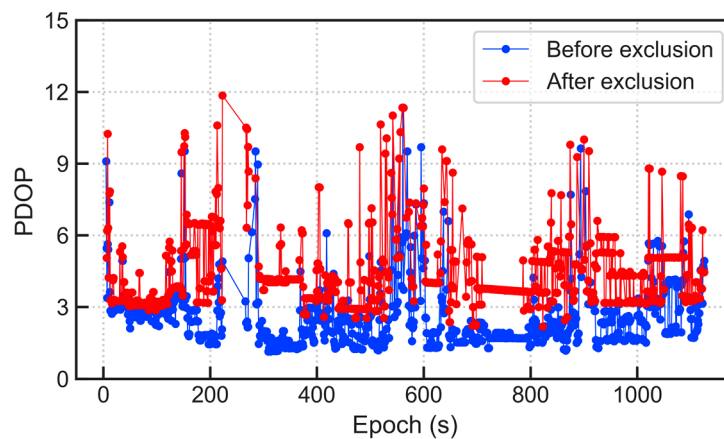


Figure 11. Comparison of the number of satellites before and after excluding multipath/NLOS signals.

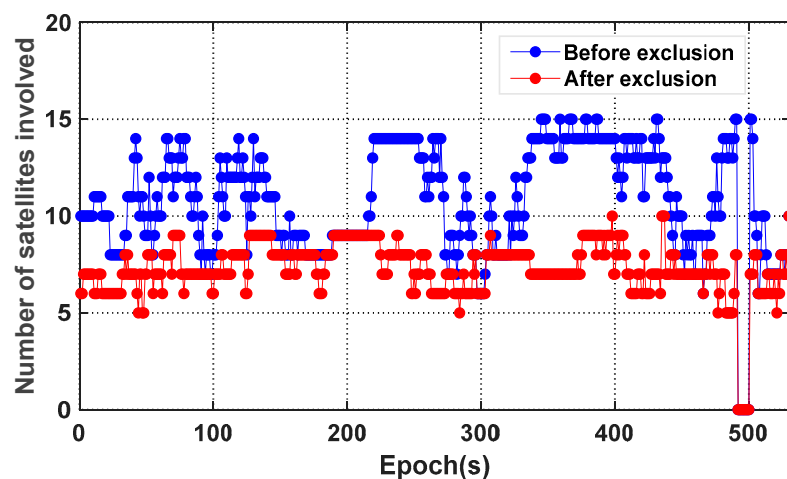


**Figure 12.** Comparison of PDOP values before and after excluding multipath/NLOS signals.

### 3.3. Outliers Detection for Real-Time Positioning

We use offline data for unsupervised learning and training to obtain the signal type identification rule. This rule is finally used for real-time GNSS observation classification and position calculation, and we call the real-time observed data—online data. Due to the good scalability of the K-means algorithm, there is no need for re-supervised learning on offline data (the labels have been obtained through the above process), but the clustering rules obtained by training are directly used for real-time detection of signals. Therefore, the above model is then employed to identify the observation outliers containing multipath/NLOS signal on the online observation dataset D2 for GNSS/INS real-time positioning. Feeding new GNSS/INS observations into the classifier, outliers can be obtained in real time. Similar to the offline system, the accuracy and availability of INS/DGNSS double-differenced pseudorange positioning results are used to evaluate the performance of anomaly detection algorithms after excluding multipath/NLOS signals.

Figure 13 indicates the comparison of the number of satellites participating in the position calculation before and after excluding the multipath/NLOS observations.



**Figure 13.** Comparison of the number of satellites for D2 before and after excluding multipath/NLOS signals.

Figure 14 exhibits the comparison of the 3D position of the vehicle obtained by different positioning methods, where the red solid line is the ground truth value of the vehicle. From Figures 4b, 13 and 14, we can see that the observation conditions of the first part of the epoch of the data set D2 are poor. Multipath/NLOS signals seriously affect the positioning accuracy, which causes the position result to deviate from the reference value to a large extent. Without detecting and eliminating multipath/NLOS on the observational data, the trajectory diagram of the vehicle's 3D position can even produce spurs. The observation

circumstance of the subsequent epochs is gradually expanded, and the results obtained by the two algorithms are close to the reference value. This demonstrates that in urban complex scenarios, detecting and eliminating multipath/NLOS signals plays an important role in improving the performance of GNSS/INS integrated real-time positioning.

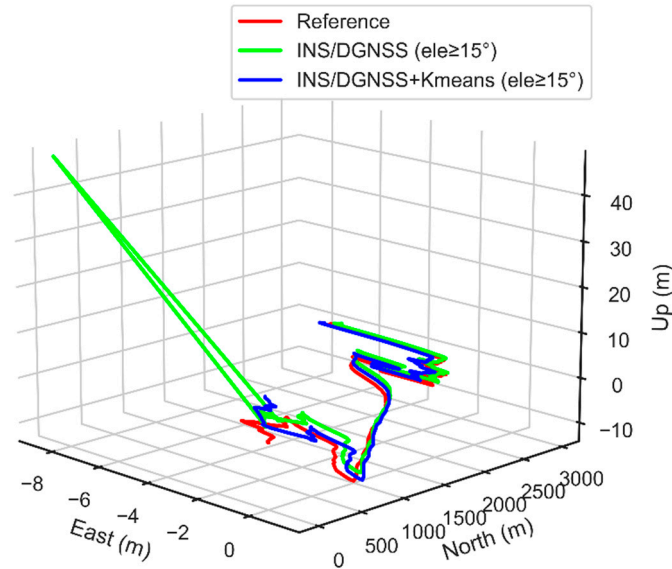


Figure 14. 3D position comparison of positioning results of different tightly coupled schemes.

Further, this paper calculates the RMS value of the position error of the real-time positioning of the online dataset, as shown in Table 4. It can be seen that in the urban harsh scenarios, the K-means clustering algorithm has a better effect on the improvement of the positioning results. The positioning accuracy of the D2 online dataset is improved by 21% and 41% in the horizontal and vertical directions, respectively.

Table 4. Accuracy comparison of positioning results for different tightly coupled schemes.

Positioning Scheme	RMS/m			
	E	N	U	2D
INS/DGNSS (ele ≥ 15°)	0.53	0.54	2.79	0.75
INS/DGNSS + K-means (ele ≥ 15°)	0.53	0.25	1.66	0.59

#### 4. Discussion

Multipath effects and NLOS signals are the main factors restricting the accuracy and reliability of GNSS/INS positioning, especially in challenging environments such as urban canyons, shaded trees, etc. Therefore, given the interference of multipath/NLOS signals, this paper proposes an outliers detection method composed of an offline learning system and an online learning system for GNSS/INS tightly coupled positioning in urban areas. We believe that GNSS signals in complex environments are generally divided into three categories: LOS, multipath, and NLOS signals. Each type of signal has a certain internal relationship with the four feature parameters of satellite elevation angle, carrier-to-noise ratio, pseudorange residual, and pseudorange rate consistency. According to this characteristic, K-means algorithm is used for signal clustering. When the sample is closest to one of the cluster centers, it is classified into this class. In an offline system, the K-means clustering algorithm is employed to detect observation outliers and construct an offline training set with labels, without resorting to 3D building model and external sensor. On this basis, due to the good scalability of the K-means clustering algorithm, the above-mentioned model is then utilized to identify multipath/NLOS signals on the online observation dataset for real-time positioning.



As can be seen from Figures 11 and 12, after excluding GNSS observation outliers, the number of satellites participating in the position calculation decreases, and the GDOP value has generally increased. Yet, despite this, the positioning accuracy has been improved. However, while ensuring the positioning accuracy, the continuity of the dynamic positioning results is also crucial. Although correct outlier detection and exclusion can effectively improve the performance of positioning results, it must be admitted that directly removing abnormal observations will reduce the number of available GNSS satellites and weaken the satellite geometric distribution. This will reduce the positioning performance to a certain extent, especially when the number of available satellites is small. Therefore, it is not advisable to blindly pursue positioning accuracy and lose a large number of original valid epochs, but there will certainly be more space for optimizing signal selection when the available GNSS constellations are enough.

Unfortunately, although this paper has made some meaningful explorations in multipath/NLOS detection and elimination, the above research work still needs to be further improved because of the complexity of multipath/NLOS signals, for example, using more GNSS/INS observation data to establish offline label datasets, so that the training set covers more scenarios and satellite constellations, and improves the generalization ability of the classification model. In addition, based on different anomaly distribution assumptions, a more suitable detection method for observation outliers is pursued under the condition of ensuring positioning accuracy. We have also made preparations for this and will further study it in the future.

## 5. Conclusions

GNSS/INS integrated navigation possesses excellent characteristics so that it plays a significant role in vehicle positioning requirements. However, the performance of GNSS/INS integration suffers from excessive unexpected GNSS outliers such as multipath/NLOS signal in dense urban areas.

This paper put forward an urban vehicle GNSS multipath/NLOS observation detection algorithm based on K-means clustering, which can effectively promote the accuracy of GNSS/INS tightly coupled positioning results. The method is essentially an offline learning system that can be used for post-processing solution of GNSS/INS observation data. Simultaneously, we employ K-means to detect observation outliers and obtain LOS/NLOS classification rules, which can be further broadened to GNSS/INS integrated navigation vehicle position in real-time. The proposed method obtains the signal type label by sufficiently excavating the information of the GNSS observation data itself, without the assistance of external software and hardware. Based on the good scalability of the K-means clustering algorithm, the above model is used to identify the multipath/NLOS of online observation data for real-time positioning. As a result, it can effectively enhance the performance of GNSS/INS tightly coupled system with higher availability and environmental adaptability.

In future work, we will continue to research the influence of the GNSS signal distribution pattern in different scenarios and test data on the positioning performance of GNSS/INS tightly coupled system, and study a more robust outlier boundary determination rule. Additionally, when the number of visible satellites is relatively small, simply excluding the multipath/NLOS signal will deteriorate the satellite geometric distribution, which reduces the positioning accuracy or even fails to execute the positioning solution. Therefore, in future research, we will further consider reasonable multipath/NLOS processing strategies, such as optimizing the stochastic model of the observation equation [45].

**Author Contributions:** H.W. conceived the idea and designed the experiments with S.P. and W.G., H.W. and S.P. wrote the main manuscript. Y.X. reviewed the paper. C.M. collected GNSS/INS dynamic observation by vehicle. All components of this research were carried out under the supervision of H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is partially supported by the National Key Research and Development Program of China (Grant No. 2021YFB3900804) and the Research Fund of Ministry of Education of China and China Mobile (Grant No. MCM2020J01).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We would like to thank everyone who contributed to this research, including data collection, manuscript review and project support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bjanis, J.P.; Jones, M.R.; Quackenbush, N.F. Benefits of operating multiple atomic frequency standards for GNSS satellites. *GPS Solut.* **2021**, *25*, 141.
2. Wang, Z.; Xia, Y.; Li, J.; Wu, Q. A new method of integer parameter estimation in linear models with applications to GNSS high precision positioning. *IEEE Trans. Signal Process.* **2021**, *69*, 4567–4579. [[CrossRef](#)]
3. Chen, Q.; Zhang, Q.; Niu, X. Estimate the pitch and heading mounting angles of the IMU for land vehicular GNSS/INS integrated system. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 6503–6515. [[CrossRef](#)]
4. Kim, L.; Lee, Y.; Lee, H.K. Kalman–Hatch dual-filter integrating global navigation satellite system/inertial navigation system/on-board diagnostics/altimeter for precise positioning in urban canyons. *IET Radar Sonar Navig.* **2022**, *16*, 379–397. [[CrossRef](#)]
5. Li, Z.; Liu, Z.; Zhao, L. Improved robust Kalman filter for state model errors in GNSS-PPP/MEMS-IMU double state integrated navigation. *Adv. Space Res.* **2021**, *67*, 3156–3168. [[CrossRef](#)]
6. Shen, C.; Zhang, Y.; Tang, J.; Cao, H.; Liu, J. Dual-optimization for a MEMS-INS/GPS system during GPS outages based on the cubature Kalman filter and neural networks. *Mech. Syst. Signal Process.* **2019**, *133*, 106222. [[CrossRef](#)]
7. Balsa-Barreiro, J.; Valero-Mora, P.M.; Pareja-Montoro, I.; Sánchez-García, M. Proposal of geographic information systems methodology for quality control procedures of data obtained in naturalistic driving studies. *IET Intell. Transp. Syst.* **2015**, *9*, 673–682. [[CrossRef](#)]
8. Chen, K.; Chang, G.; Chen, C. GINav: A MATLAB-based software for the data processing and analysis of a GNSS/INS integrated navigation system. *GPS Solut.* **2021**, *25*, 108. [[CrossRef](#)]
9. Liao, J.; Li, X.; Wang, X.; Li, S.; Wang, H. Enhancing navigation performance through visual-inertial odometry in GNSS-degraded environment. *GPS Solut.* **2021**, *25*, 50. [[CrossRef](#)]
10. Jagelčák, J.; Gnap, J.; Kuba, O.; Frnda, J.; Kostrzewski, M. Determination of Turning Radius and Lateral Acceleration of Vehicle by GNSS/INS Sensor. *Sensors* **2022**, *22*, 2298. [[CrossRef](#)]
11. Han, H.; Wang, J.; Wang, J.; Moraleda, A.H. Reliable partial ambiguity resolution for single-frequency GPS/BDS and INS integration. *GPS Solut.* **2017**, *21*, 251–264. [[CrossRef](#)]
12. Miller, I.; Campbell, M. Sensitivity analysis of a tightly-coupled GPS/INS system for autonomous navigation. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 1115–1135. [[CrossRef](#)]
13. Jiang, C.; Zhang, S.; Li, H.; Li, Z. Performance evaluation of the filters with adaptive factor and fading factor for GNSS/INS integrated systems. *GPS Solut.* **2021**, *25*, 130. [[CrossRef](#)]
14. Balsa-Barreiro, J.; Valero-Mora, P.M.; Pareja Montoro, I.; Sanchez Garcia, M. Geo-referencing naturalistic driving data using a novel method based on vehicle speed. *IET Intell. Transp. Syst.* **2013**, *7*, 190–197. [[CrossRef](#)]
15. Ando, T.; Kugimiya, W.; Hashimoto, T.; Momiyama, F.; Aoki, K.; Nakano, K. Lateral control in precision docking using RTK-GNSS/INS and LiDAR for localization. *IEEE Trans. Intell. Veh.* **2020**, *6*, 78–87. [[CrossRef](#)]
16. Qian, C.; Zhang, H.; Li, W.; Tang, J.; Liu, H.; Li, B. Cooperative GNSS-RTK Ambiguity Resolution with GNSS, INS, and LiDAR Data for Connected Vehicles. *Remote Sens.* **2020**, *12*, 949. [[CrossRef](#)]
17. Jiang, H.; Li, T.; Song, D.; Shi, C. An Effective Integrity Monitoring Scheme for GNSS/INS/Vision Integration Based on Error State EKF Model. *IEEE Sens. J.* **2022**, *22*, 7063–7073. [[CrossRef](#)]
18. Alqurashi, M.; Wang, J. Performance analysis of fault detection and identification for multiple faults in GNSS and GNSS/INS integration. *J. Appl. Geod.* **2015**, *9*, 35–48. [[CrossRef](#)]
19. Li, T.; Zhang, H.; Gao, Z.; Chen, Q.; Niu, X. High-accuracy positioning in urban environments using single-frequency multi-GNSS RTK/MEMS-IMU integration. *Remote Sens.* **2018**, *10*, 205. [[CrossRef](#)]
20. Zhang, G.; Wen, W.; Xu, B.; Hsu, L.T. Extending shadow matching to tightly-coupled GNSS/INS integration system. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4979–4991. [[CrossRef](#)]
21. Hannah, B.M. Modelling and Simulation of GPS Multipath Propagation. Ph.D. Thesis, Queensland University of Technology, Brisbane, QLD, Australia, 9 March 2001.
22. Groves, P.D.; Jiang, Z.; Rudi, M.; Strode, P. A portfolio approach to NLOS and multipath mitigation in dense urban areas. In Proceedings of the ION GNSS+ 2013, Nashville, TN, USA, 16–19 September 2013; pp. 3231–3247.
23. Drawil, N.M.; Amar, H.M.; Basir, O.A. GPS localization accuracy classification: A context-based approach. *IEEE Trans. Intell. Transp. Syst.* **2012**, *14*, 262–273. [[CrossRef](#)]

24. Wen, W.; Zhang, G.; Hsu, L.T. GNSS outlier mitigation via graduated non-convexity factor graph optimization. *IEEE Trans. Veh. Technol.* **2021**, *71*, 297–310. [[CrossRef](#)]
25. Yozevitch, R.; Moshe, B.B.; Weissman, A. A robust GNSS LOS/NLOS signal classifier. *Navig. J. Inst. Navig.* **2016**, *63*, 427–440. [[CrossRef](#)]
26. Hsu, L.T. GNSS multipath detection using a machine learning approach. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.
27. Quan, Y.; Lau, L.; Roberts, G.W.; Meng, X.; Zhang, C. Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning. *Remote Sens.* **2018**, *10*, 2052. [[CrossRef](#)]
28. Sun, R.; Wang, G.; Zhang, W.; Hsu, L.T.; Ochieng, W.Y. A gradient boosting decision tree based GPS signal reception classification algorithm. *Appl. Soft Comput.* **2020**, *86*, 105942. [[CrossRef](#)]
29. Xia, Y.; Pan, S.; Zhao, P. Satellite NLOS signal Detection method based on unsupervised learning. *J. Southeast Univ.* **2019**, *49*, 565–572.
30. Xia, Y.; Pan, S.; Meng, X.; Gao, W.; Ye, F.; Zhao, Q.; Zhao, X. Anomaly detection for urban vehicle GNSS observation with a hybrid machine learning system. *Remote Sens.* **2020**, *12*, 971. [[CrossRef](#)]
31. Savas, C.; Dervis, F. Multipath detection based on K-means clustering. In Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019), Miami, FL, USA, 16–20 September 2019; pp. 3801–3811.
32. Bin, Z.H.U.; Cheng, Y.A.N.G.; Yan, L.I.U. Analysis and comparison of three unsupervised learning methods for GNSS multipath signals. *Acta Geod. Et Cartogr. Sin.* **2021**, *50*, 1762–1771.
33. Suzuki, T.; Kitamura, M.; Amano, Y.; Hashizume, T. Multipath mitigation using omnidirectional infrared camera for tightly coupled GPS/INS integration in urban environments. In Proceedings of the 24th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2011), Portland, OR, USA, 20–23 September 2011; pp. 2914–2922.
34. Obst, M.; Bauer, S.; Reisdorf, P.; Wanielik, G. Multipath detection with 3D digital maps for robust multi-constellation GNSS/INS vehicle localization in urban areas. In Proceedings of the 2012 IEEE Intelligent Vehicles Symposium, Madrid, Spain, 3–7 June 2012; pp. 184–190.
35. Wen, W.; Bai, X.; Kan, Y.C.; Hsu, L.T. Tightly coupled GNSS/INS integration via factor graph and aided by fish-eye camera. *IEEE Trans. Veh. Technol.* **2019**, *68*, 10651–10662. [[CrossRef](#)]
36. Han, H.; Wang, J. Robust GPS/BDS/INS tightly coupled integration with atmospheric constraints for long-range kinematic positioning. *GPS Solut.* **2017**, *21*, 1285–1299. [[CrossRef](#)]
37. Wang, D.; Dong, Y.; Li, Z.; Li, Q.; Wu, J. Constrained MEMS-based GNSS/INS tightly coupled system with robust Kalman filter for accurate land vehicular navigation. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 5138–5148. [[CrossRef](#)]
38. Strode, P.R.; Groves, P.D. GNSS multipath detection using three-frequency signal-to-noise measurements. *GPS Solut.* **2016**, *20*, 399–412. [[CrossRef](#)]
39. Hsu, L.T.; Tokura, H.; Kubo, N.; Gu, Y.; Kamijo, S. Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons. *IEEE Sens. J.* **2017**, *17*, 1909–1917. [[CrossRef](#)]
40. Jiang, Z.; Groves, P.D.; Ochieng, W.Y.; Feng, S.; Milner, C.D.; Mattos, P.G. Multi-constellation GNSS multipath mitigation using consistency checking. In Proceedings of the 24th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2011), Portland, OR, USA, 20–23 September 2011; pp. 3889–3902.
41. Likas, A.; Vlassis, N.; Verbeek, J.J. The global K-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461. [[CrossRef](#)]
42. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [[CrossRef](#)]
43. Groves, P.D. *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*; Artech House: Norwood, MA, USA, 2013.
44. Kennedy, S.; Hamilton, J.; Martell, H. Architecture and system performance of SPAN-NovAtel's GPS/INS solution. In Proceedings of the IEEE/ION PLANS 2006, San Diego, CA, USA, 25–27 April 2006; pp. 266–274.
45. Sun, R.; Zhang, Z.; Cheng, Q.; Ochieng, W.Y. Pseudorange error prediction for adaptive tightly coupled GNSS/IMU navigation in urban areas. *GPS Solut.* **2022**, *26*, 28. [[CrossRef](#)]