A Resilient Smartphone Positioning Approach by Tightly Integrating the Monocular Camera and GNSS Signals

Han Xu[1], Lei Wang[1,2*], Ruizhi Chen[1,2], Haitao Zhou[1], Tao Li[1], Yi Han

1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
2. Collaborative Innovation Center for Geospatial Technology, Wuhan 430079, China
*   Corresponding Author, lei.wang@whu.edu.cn

Abstract: The Global Navigation Satellite System (GNSS) signals are often blocked or interfered in complex geographical or electromagnetic environments, which may make GNSS receivers unable to provide satisfying navigation and positioning services. There have been many ground-based or space-based GNSS augmentation systems to improve the resilience of GNSS positioning, of which most of them rely on additional infrastructures. In this study, a smartphone-based tightly-coupled positioning method was developed using the images from a build-in monocular camera and GNSS signals. In this method, the feature points with the known coordinates are regarded as 'visual pseudolite' and the distance between the camera and the feature points was calculated according to the photogrammetry approaches and used to estimate the user positioning with GNSS signals. The experimental results showed the feasibility of the tightly-coupled positioning algorithm and reached the positioning accuracy of ±5.56 m (1σ), which is significantly higher than that of GNSS-only and vision-only positioning solutions.

Keywords: visual positioning; GNSS; tightly-coupled positioning; high availability; resilient PNT

1. Introduction

Global Navigation Satellite Systems (GNSS) have been widely used in positioning and navigation. GNSS signals work well in the open sky, but are often blocked or interfered in less GNSS-friendly environments, which degrade or even interrupt the positioning service [1]. On the other hand, GNSS interference, jamming and fraud occur frequently, which makes the GNSS-based positioning more difficult [2,3]. The future alternative positioning, navigation and timing (PNT) framework will expectedly achieve flexible and tough navigation and positioning services by integrating multiple heterogeneous navigation sources [4]. The method of multi-source fusion from the assured PNTs ensures the high availability of navigation and positioning services [5].

Generally, there are three types of augmentation methods to improve satellite-based positioning availability: signal augmentation, matching, and dead reckoning. There are multiple GNSS signal augmentation approaches, such as the pseudolite [6,7], low earth orbit (LEO) navigation signal augmentation [8,9], cellular network, Wi-Fi signal [10], acoustic signal, and Radar, Loran-C, etc. This augmentation employs external infrastructures or actively transmitting signals to obtain more geometrical information for positioning. The matching algorithm relies on certain prior information, such as the magnetic field[11], the strength field of radio frequency (RF) signals, or

image features. Since the matching algorithm does not require geometry observation，it is often loosely coupled with the other positioning techniques, such as GNSS. The dead reckoning technique can be applied for navigation using the accumulated position change, such as the inertial sensors[12,13], visual odometer with cameras and/or LiDAR [14], simultaneous localization and mapping (SLAM)[15], etc. The dead reckoning methods can be either loosely coupled or tightly coupled with GNSS to obtain a smoother trajectory. However, most dead reckoning approaches are not capable of determining the initial states, which may lead to navigation failure when starting navigation in GNSS denied environments.

The visual information is often used for local navigation, such as the SLAM technique and the visual odometers. These approaches determine the user movement by matching the features between two consecutive images. It can also be used for absolute positioning with cooperative targets, such as the QR code[16], the text[17], or other encoded targets [18]. The vision-based positioning is particularly suitable for positioning locally since the positioning accuracy will significantly decrease as the distance increases between the object and the camera. As a result, vision-based positioning techniques are mainly used to solve indoor positioning problems [19]. Because of the anti-jamming characteristics of the visual ranging signal, as well as the low price and portability of the cameras and GNSS receivers, the integration of GNSS and visual positioning systems has become a momentous issue for researchers [12-14].

In this research, a new GNSS/Vision tightly-coupled positioning method was proposed and developed toward improving the positioning availability and accuracy in the GNSS-denied environment. We viewed those feature points with their known coordinates as the 'visual pseudolites' and calculated the distance between the feature points and users via the photogrammetry methods. Then these visual ranges were combined with the GNSS signals to estimate the user positions. The visual pseudolite does not requires additional infrastructure deployment and immune from electromagnetic inference. In the future, a measurable 3D map can be used as the feature database to provide enough 'visual pseudolite', which is particularly beneficial for these GNSS challenging environments, such as the city canyon.

The remainder of this paper is organized as follows: the related work is reviewed in Section 2. Then, Section 3 presents the GNSS/Visual Localization (VL) tightly coupled model while the procedure of extracting the visual ranges is introduced in Section 4. Section 5 provides the results and discussion of the experiments. In the end, Section 6 concludes the manuscript and outlines the potential future research.

2. **Related Work**

Since GNSS-based positioning has been extensively studied, we put out focus on vision-based localization. The literature review here includes the visual localization approaches and the existing GNSS/VL integration approaches.

2.1 Visual Localization

Visual localization is an essential research topic in the field of computer vision, and the specific technology can be roughly divided into three categories. The most common method depends on monocular vision, including positioning by processing single images or multiple images. Fischler et al. [20] proposed the PnP problem (perspective-n-point problem), which is a positioning method according to the projection relationship between n feature points and their corresponding spatial positions. Liu et al[21] established a geometric model to determine the camera position according to the correspondence between 2D and 3D lines or points. The second type of visual localization is the binocular vision-based approach, which demands a large amount of computation and is difficult to match image points. The SIFT features have been widely used in binocular stereo matching due to their robustness to scale, rotation, angle of view, and other changes [22,23]. The third type relies on panoramic vision with complex measurement depth. Yagi et al. [24] firstly applied the

omnidirectional vision system with a hyperbolic mirror to the navigation of mobile robots. Since then, the panoramic vision-based localization method has been widely studied.

2.2 GNSS/vision combined Localization

GNSS provides real-time and all-weather global positioning service and do not accumulate positioning errors with time. However, the GNSS signals are vulnerable to obstructions and interferences, so their performance needs to be improved in the GNSS challenging environment. There have been quite a few attempts to integrate GNSS with the vision-based localization approaches, which provide position increments most likely using inertial sensors and/or different odometers [25]. On the other hand, position and velocity information from GNSS can also be used for global optimization and geo-referencing in visual SLAM computation[26]. In terms of their integration architectures, the GNSS and vision information can be integrated with either loosely or tightly coupled approaches.

Due to their complementary characteristics, the GNSS and vision integration can fully take their advantages. The monocular visual odometer suffers from a rank deficiency with the trajectory scale, but it provides high relative positioning precision. GNSS positioning results in absolute positioning solutions, but its performance is vulnerable to the observation conditions. In the form of loosely coupled methods, one preprocesses two subsystems to deliver their results separately and then fuses them to obtain the integrated solution by the factor graph or filters. Dusha and Mejias [27] proposed a loosely coupled filtering method for monocular cameras and GPS, which is similar to the traditional GPS/INS loosely coupled filtering method. Chen et al. [28] confirmed that the monocular camera could significantly improve the GNSS positioning accuracy when in GNSS-denied environments. The fusion of visual information and GNSS data based on iterative optimization is also verified to be feasible[26]. The limitation of the loosely coupled localization system is that the visual information cannot be utilized to improve the availability of GNSS positioning. When

GNSS signals become invalid, the loosely coupled system can only rely on the visual information, which may degrade the performance in terms of long-time GNSS loss-of-lock.

With the tightly coupled GNSS/vision localization integration architecture, the carrier phases and pseudoranges from GNSS receivers are directly fused with the visual information [29]. The current tightly coupled system resorts to the camera information to identify the none-line-of-sight (NLOS) signals. Paul and Kyle [30] proposed an NLOS effect suppression algorithm based on LOS satellite selection for harsh environments. The images collected by the sky-pointing camera were divided into the open sky and obstructed regions. The satellites falling into the obstructed region would be rejected to participate in the final position calculation. On this basis, the satellite signal and visual information were tightly coupled through Kalman filter to provide the positional solution. Similarly, Suzuki employed the sky-pointing photo matching to eliminate wrong position candidate and thus improve the positioning accuracy[31]. These approaches can improve positioning accuracy in GNSS challenging environments, but cannot improve the GNSS signal availability in GNSS-denied environments. Another tightly coupled system is to tightly fuse GNSS signals with sensors such as the visual odometer. The changing information of relative positions provided by the visual odometer constrains the GNSS trajectory toward improving the positioning accuracy. Schreiber and Konigshof [32] proposed a method to combine the local visual odometer obtained by a vehicle stereo camera system with a low-cost GNSS receiver. In order to solve the positioning problem when the number of satellites was insufficient, the pseudorange measurements were directly fused with the sensor data. Because the GNSS/VO integration is similar to the GNSS/INS integration, it suffers from a similar issue, the performance declining due to a long period of GNSS outage [33]. In this study, a resilient smartphone based positioning approach is proposed by tightly integrating the monocular camera and GNSS signals to ensure the positioning availability issue in GNSS challenging and/or denied

environments.

3. GNSS/VL Tightly-coupled Positioning Model

In order to solve the positioning feasibility issue in the GNSS-denied environment, alternate navigation techniques should be involved to overcome the rank defect issue in positioning estimation. Currently, multiple transmitters are introduced such as the pseudolite stations, networks, or stations with Wi-Fi, 5G and acoustic signals, etc. All these signals are used to measure distances between transmitters and user's devices and enhance the overall positioning performance. In this study, a ranging approach using the monocular camera is proposed. The features with their known coordinates are viewed as the 'visual pseudolites', which transmit light signals. The light signals are captured by the monocular camera. According to the similarity between the object and image spaces, the range between a camera focus and a visual pseudolite is calculated and used to collaborate with the GNSS signals tracked by the GNSS chip and further position the smartphone. The principle of the tightly coupled approach is illustrated in **Error! Reference**
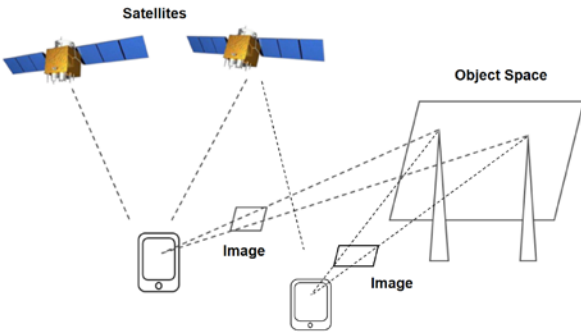

Figure 1 Illustration of the tightly-coupled positioning approach using GNSS receiver and camera

source not found.. The visual pseudolites are natural objects with certain easy-to-identify features, so no extra infrastructure is required. As the visible light is not affected by electromagnetic interference, it is more resilient than the GNSS-only positioning technique. In the future, a measurable 3D real map with enormous known coordinates information will be employed to assistant the identification of the visual pseudolites. Conceivably, the proposed method

can also provide a resilient kinematic positioning solution in the GNSS denied environment.

On the assumption that the range to a visual pseudolite has been successfully determined, then the tight fusion model of the two types of observations can be used in the positioning process. At first, we introduce the model for tightly coupling the GNSS and vision. The procedure of the visual range extraction is discussed in Section 4.

The observation equations for the tightly-coupled positioning can be expressed as follows

$$\begin{cases} P_G = \rho + \delta_{orb} + c(\delta t^S - \delta t_R) + I_i + \delta_{trop} + \varepsilon_P \\ P_C = \rho + \delta_{match} + \varepsilon_C \end{cases} \tag{1}$$

where P_G and P_C are the GNSS pseudorange and the distance measured by camera respectively; ρ presents the geometric distance between a satellite or a visual pseudolite and a receiver; δ_{orb} is the error in GNSS satellite orbit; c is the speed of light in vacuum; δt^S and δt_R are the clock error of satellites and the receiver, respectively; I_i and δ_{trop} are the ionospheric and tropospheric delays on the GNSS signals, respectively; δ_{match} is the position error of a visual pseudolite via the matching process, and ε_P and ε_C are the white noises of the GNSS observations and the visual ranges, respectively. Apparently, for the smartphone's built-in camera, the resolution and lens dissertation may lead to position errors during the feature matching process.

How to handle the error sources is decisive on the final accuracy of the integrated positioning system. In GNSS pseudorange based positioning, the satellite orbit and clock correction can be obtained from the broadcast ephemeris. Currently, the error associated with the broadcast satellite orbit and satellite clock is about ±1-2 meters [34] and the remaining error is neglected in the processing procedure. The ionosphere and troposphere delays can be corrected with the empirical models. The

former can be corrected with the Klobuchar model, BDGIM model or the NeQuick model[35], while the latter can be corrected with the Saastamoinen model or the Hopfield model. The receiver clock bias needs to be estimated along with the user's unknown position [36].

For the vision-derived ranges, the errors come from the visual pseudolite position and the ranges. Although we assume that the true position of a visual pseudolite is always known, the error is still brought by the matching process. However, the error in the matching process cannot be handled directly and can be adapted by adjusting the stochastic model. The distortion of the lens also affects the visual range accuracy, but it can be calibrated ahead. Since the visual range typically varies from a few meters to hundreds of meters, the atmospheric refraction can be ignored in this particular application. The visual range is independent of time, so it is free of time-dependent error sources. However, it is still possible to add time-tags for the visual ranges for kinematic positioning. As a result, the vision-based ranges suffer from fewer error sources and can be used to estimate the user coordinates directly.

Another important issue for the integration is the unification of the spatial and temporal datums. The spatial datum for the visual range is defined by the coordinates of the visual pseudolite. Hence, only when the coordinate system of the visual pseudolite is compatible with the current geodetic datum of a GNSS, e.g. the World Geodetic System 84 (WGS84) for GPS, the Beidou coordinate systems (BDCS) or the International terrestrial reference frame (ITRF). Since the visual range does not have time information, it is not necessary to synchronize the visual information and the GNSS signals, but the visual range information should be updated once the user position changed.

Estimating the user's coordinates from the range information is not a linear problem, so the distance observation equation needs to be linearized first. The geometric distance can be expanded with the Taylor series as follows:

$$\rho = \rho^{(0)} + \frac{\partial \rho}{\partial x}dx + \frac{\partial \rho}{\partial y}dy + \frac{\partial \rho}{\partial z}dz + \varepsilon \qquad (2)$$

$$= \rho^{(0)} + \frac{x^{(0)} - x_i}{\rho^{(0)}}\delta x + \frac{y^{(0)} - y_i}{\rho^{(0)}}\delta y + \frac{z^{(0)} - z_i}{\rho^{(0)}}\delta z + \varepsilon$$

where $\rho^{(0)}$ is the approximate geometric distance from the receiver to satellites or visual pseudolite; ε is the nonlinear error; (x_i, y_i, z_i) are the coordinates of satellites or visual pseudolite; (δx, δy, δz) are the increment of the receiver coordinates. Neglecting the high order nonlinear term ε yields

$$E(\rho - \rho^{(0)}) \approx (\frac{x_i - x^{(0)}}{\rho^{(0)}} \quad \frac{y_i - y^{(0)}}{\rho^{(0)}} \quad \frac{z_i - z^{(0)}}{\rho^{(0)}}) \begin{pmatrix} \delta x \\ \delta y \\ \delta z \end{pmatrix} \qquad (3)$$

The linearized observation model is given as follows:

$$E(y_m) = A \cdot \delta X \qquad (4)$$

$$\delta X = [\delta x, \delta y, \delta z, \delta t]^T \qquad (5)$$

$$A = \begin{pmatrix} \frac{-(x_1 - x^{(0)})}{\rho_1} & \frac{-(y_1 - y^{(0)})}{\rho_1} & \frac{-(z_1 - z^{(0)})}{\rho_1} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-(x_m - x^{(0)})}{\rho_m} & \frac{-(y_m - y^{(0)})}{\rho_m} & \frac{-(z_m - z^{(0)})}{\rho_m} & 1 \\ \frac{-(x_1 - x^{(0)})}{\rho_1} & \frac{-(y_1 - y^{(0)})}{\rho_1} & \frac{-(z_1 - z^{(0)})}{\rho_1} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-(x_n - x^{(0)})}{\rho_n} & \frac{-(y_n - y^{(0)})}{\rho_n} & \frac{-(z_n - z^{(0)})}{\rho_n} & 0 \end{pmatrix} \qquad (6)$$

where A is the design matrix, in which the first m distance observations are the GNSS satellites, while the last n observations are obtained from the n visual pseudolites. δX is the parameter vector to be estimated containing the receiver coordinates and clock bias. y_m is the observations. The least-square solution of this equation system is formed as:

$$\delta \hat{X} = (A^T P A)^{-1} A^T P y_m \qquad (7)$$

where P is the observation weighting matrix determined according to the prior ranging accuracy of camera and GNSS measurements.

The stochastic model is crucial for the tightly coupled GNSS/vision fusion. For satellite data, the popular elevation-based weighting strategy can be

used. The weighting function between the observed variance and the elevation angle can be expressed as[37]:

$$\sigma^2 = a^2 + b^2 / (\sin(\alpha))^2 \qquad (8)$$

where σ^2 is the variance, α is the elevation angle of a satellite, and a and b are the constant error components, which can be determined empirically or derived from the variance component estimation.

The stochastic model for the visual observations is characterized to be inversely proportional to the distance. A feature that is closer to the camera may have better resolution and is identified with the smaller matching error. So it should be more accurate. According to the experience, the distance errors corresponding to different distances are obtained, and the corresponding weight value is calculated by fitting the linear function.

4. Range Extraction From the Images

There is no doubt the key of the GNSS/Vision tightly coupled approach is how to extract the visual ranges. Extracting the precise visual ranges follows a stepwise procedure, which will be introduced in this section.

4.1 Principle of Extracting the Visual Ranging

The principle of the visual ranging is the space resection, which can be illustrated in Figure 2 Figure 2. The 'visual pseudolite' are A, B and C in real-world with known 3D coordinates. The coordinates of the features in the photo plane a, b and c are the observables. S is the focus of the smartphone camera and the focal length can be precisely obtained from a calibration process. Ideally, the position and the attitude of the smartphone can be estimated with 3 or more matched features. However, due to the limitation of the smartphone camera quality, the initial position cannot achieve satisfactory accuracy, particularly for the large scene scenario. The visual range can be generated with the initial position and the known coordinates of the visual pseudolite. The user position can be re-estimated by integrating the visual range and the GNSS signals. Since the precision of the GNSS signals does not affect by the visual range length, the precision of the re-estimated user position can be improved. On the other hand, when the visible GNSS satellite is too few to fix the position, the integration of the visual range and the GNSS signals can provide reliable positioning results.
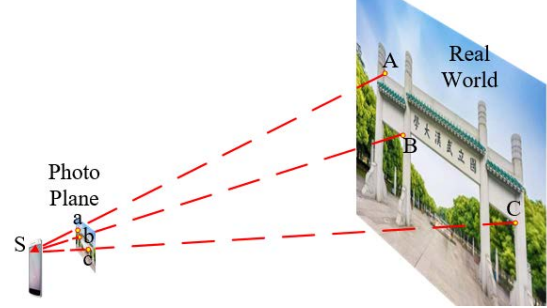

Figure 2 Illustration of the feature-based space resection

In practice, the visual range is affected by many error sources. In order to obtain high precise visual ranges, a procedure with four steps is designed and presented in **Error! Reference source not found.**. After acquired the images, the distortion caused by
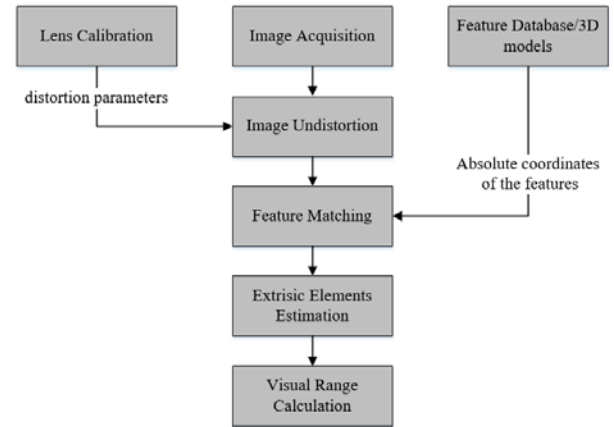

Figure 3 Flowchart of the visual ranging extraction

the lens should be calibrated first. The second step is obtaining the coordinates of the features by matching them with the feature database or 3D models. The absolute coordinates of the features in the database are known, so the matching procedure can connect the coordinates of the feature in the photo plane and the real world. The third step is estimating the initial position using the multiple matched features. The last step is deriving the visual ranges from the initial position so that they can work together with the GNSS signals to improve the positioning availability by

and reliability.

4.2 Camera Lens Calibration

The similarity relationship between the image plane and the real world is based on the light propagation along a straight line, however, this assumption is generally not true for the camera lens due to the lens design or the manufacturing precision. Two typical distortions caused by the camera lens are presented in **Error! Reference source not found.**.
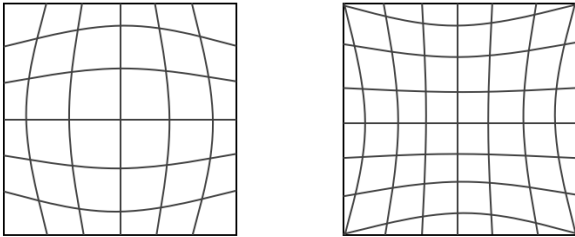

Figure 4 Examples of typical distortion caused by the camera lens

Camera lens distortion includes barrel distortion, pillow distortion, and linear distortion [38], which can be calibrated in advance. The calibration process is quite straightforward. It employs a series of photos with regular shapes. Then the lens distortion can be captured. Then, the distortion can be expressed as a set of distortion parameters, which can be estimated from the photo set. The distortion can be decomposed as radial distortion, tangential distortion, and thin prism distortion [39], The tangential distortion and the thin prism distortion are mainly caused by installation error and imperfection manufacturing[40]. In the calibration, the distortion parameters are expressed as a linear function of the radius and the origin coordinates in the photo [41]. These distortion parameters are estimated to describe the non-linear distortion over the two-dimensional photo plane.

After the distortion parameter estimated, these parameter is used to calibrate the new coming photos. The coordinates of the pixel in the photo can be calibrated one by one and the distortion caused by the lens can be significantly mitigated. For the smartphone, camera calibration is particularly important since its lens is not as good as the professional camera.

4.3 Feature Matching

The second step to extract the visual range is feature matching. The target of feature matching is obtaining the absolute coordinates of the feature in the feature database. In order to automatically matching the features, a feature database with labeled data should be established. The feature matching procedure is similar to look up the dictionary. At first, extracting features from the newly captured photos, then automatically find the best matches in the database. Since the illumination condition and the view angle of the new photo and the database may not the same, so it is important to select the features and the matching algorithms. There are many feature matching approaches, such as the gray intensity-based approach, edge contour-based and corner detection-based approaches [42], while some of them only perform well for a particular type of image. In this paper, the Scale-invariant feature transform (SIFT) [43] feature detection and matching algorithm is adopted due to its outstanding computation stability in the case of illumination, rotation and scale change. Figure 5 gives an example of the matched features in two photos. The two photos have different view angles and the same features in the two photos are connected with solid lines. If there is no feature matched in the database, then the feature extracted from the new images will be discarded. If the feature is matched successfully, the known coordinates in the database will be given to the features in the new photo, so that the connections between the 2D photo plane and the 3D real world are established. In the future, the feature database can be replaced with a measurable 3D map, so that it can be used for large-scale feature matching to support autonomous driving applications.

Figure 5 Example of matched features using SIFT features in two photos with different view angles.

4.4 Visual Range Estimation

With the coordinates of the features known, the extrinsic elements of the camera can be estimated. The visual range estimation is based on the principle of space resection in photogrammetry. In Figure 2, the coordinates of the visual pseudolite A and the camera focus S in the real-world coordinate system are denoted as (X_A, Y_A, Z_A) and (X_S, Y_S, Z_S). The coordinates of the corresponding feature a in the image plane coordinate system are denoted as (x, y, -f). Based on the collinear condition equation, the observation equation of the known point A can be expressed as [44]:

$$x - x_0 = -f \frac{a_1(X - X_s) + b_1(Y - Y_s) + c_1(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \left. \atop \right\}$$
$$y - y_0 = -f \frac{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \quad (9)$$

where a_i, b_i and c_i (i=1,2,3) are nine direction cosines of attitude parameters ψ, ω and κ.

The six exterior orientation elements $(X_S, Y_S, Z_S, \psi, \omega, \kappa)$ can be obtained through the coordinates of more than three non-collinear visual pseudolite. The collinear condition equation can be expanded by the Taylor series and the obtained linearized observation equation can be shown as:

$$x = x_0 + \frac{\partial x}{\partial \varphi}d\varphi + \frac{\partial x}{\partial \omega}d\omega + \frac{\partial x}{\partial \kappa}d\kappa + \frac{\partial x}{\partial Xs}dXs + \frac{\partial x}{\partial Ys}dYs + \frac{\partial x}{\partial Zs}dZs \left. \atop \right\}$$
$$y = y_0 + \frac{\partial y}{\partial \varphi}d\varphi + \frac{\partial y}{\partial \omega}d\omega + \frac{\partial y}{\partial \kappa}d\kappa + \frac{\partial y}{\partial Xs}dXs + \frac{\partial y}{\partial Ys}dYs + \frac{\partial y}{\partial Zs}dZs \quad (10)$$

In order to ensure the accuracy and reliability of parameter estimation, it is usually necessary to measure four or more image control points and corresponding image point coordinates, and use the least square adjustment method to solve the problem. At this time, the coordinates of image points (x, y) are taken as observation values, and the error equation of each image point can be listed as

$$v_x = a_{11}dX_s + a_{12}dY_s + a_{13}dZ_s + a_{14}d\varphi + a_{15}d\omega + a_{16}d\kappa - l_x \left. \atop \right\}$$
$$v_y = a_{21}dX_s + a_{22}dY_s + a_{23}dZ_s + a_{24}d\varphi + a_{25}d\omega + a_{26}d\kappa - l_y \quad (11)$$

where these coefficients can be referred to [44].

Using the least-squares method to solve the exterior orientation elements can be given as

$$\hat{\theta} = (A^T PA)^{-1}(A^T Pl) \quad (12)$$

where A represents the design matrix and P is the weight matrix of the observation value and l contains the observation information.

Based on the exterior orientation elements, the approximate coordinates of the camera can be calculated, which is used to figure out the distance from the camera to the target points with the accurate coordinates of the target points together.

In general, the longer the visual range between the camera and the visual pseudolite, the lower the accuracy of external orientation elements and position solution. So the variance of the distance observation value can be determined by establishing an empirical model according to the distance of observation.

5. Experiments

The performance of the proposed approach was verified with the experiments in the Wuhan University campus. The method is evaluated by images in outdoor scenarios and compared with other methods. Meanwhile, we will discuss the effect of different factors on the performance of the method.

5.1 Experiment Setup

A test field was set up on the Youyi square on the campus. The distance from the observation point to the target building varies from 50 m to 100 m. The GNSS data and the pictures are collected with a RedMi k30 ultra smartphone. The smartphone was put on a tripod for the GNSS data collection and the skyplot of the tracked GNSS satellite is presented in Figure 6. Signals from 10 GPS satellites are tracked and only pseudorange on L1 frequency is used for positioning in this study. In the experiment, only two or three GNSS satellites are used to simulate the GNSS denied environments. Based on GNSS static data, the ground truth of the position is determined by Post-processing kinematic (PPK) technique which can achieve high-precision results. A feature database was established with a set of historical image data. The features were extracted from these images and their true coordinates in the WGS84 coordinate system were obtained by combining GNSS and the total station. The absolute coordinate of the total station site was computed by GNSS relative static positioning. The international GNSS service (IGS)

station WUH2 station is used as the reference station, which is only a few hundred meters far away from the experiment site. The precise coordinate of the WUH2 station is precisely known. Two GNSS points are measured in the test site. One is the total station site and the other one is used as the back sight. A Leick GS60 prism-free total station was used to precisely obtain the coordinates of the visual pseudolite. The visual pseudolites are randomly distributed on the wall or the ground. The features extracted from new photos would be matched by comparing with the established feature database and the corresponding visual ranging information would be calculated.
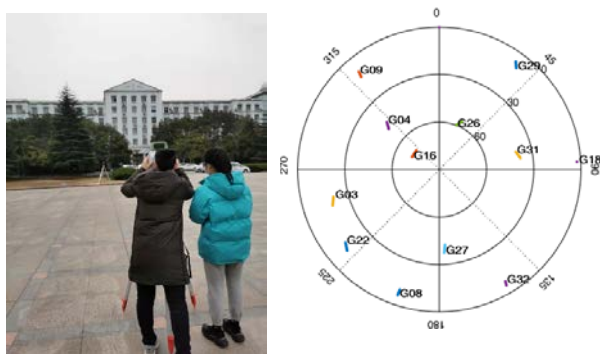

Figure 6 Experiment setup (left) and the skyplot of the tracked GNSS satellites (right)

5.2 GNSS/VL integrated Positioning performance

In order to evaluate the effectiveness of the proposed method, the positioning solutions estimated from GNSS-only positioning, visual-only, and tightly-coupled methods were compared. In the GNSS-only positioning scheme, all tracked GNSS signals by the smartphone were used. In the visual-only positioning scheme, all the matched feature points are involved to solve the user position. In GNSS/vision tightly-coupled positioning scheme, only three visible satellite is selected to simulate the GNSS denied environment. In this study, GNSS signals from satellite G22, G09 and G31 are used.

GNSS and image data were collected on 4 different sites using the same smartphone in this experiment, the positioning results and the matched feature location are presented in Figure 7. The figure presents the position obtained with the visual localization-only, GNSS-only and tightly-coupled solution respectively. The true coordinates were computed with the static relative GNSS data processing. Most features in this experiment are located on the building wall. Each image may only contain a portion of the features. The figure indicates that all the three positioning schemes achieve meter level accuracy, but the tightly-coupled solution is closer to the reference solution than the other two solutions.
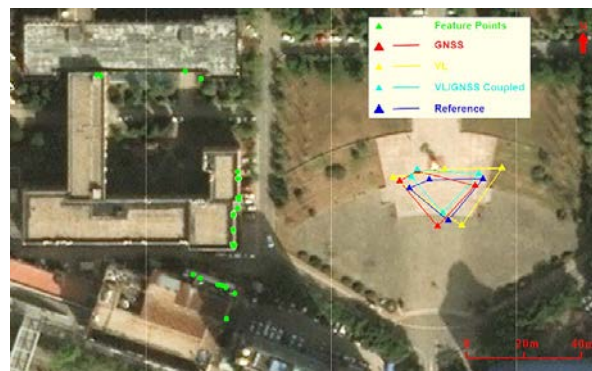

Figure 7 Positioning results of different positioning schemes

The root mean squares error (RMSE) of the three positioning schemes are listed in Table 1. The table indicates that smartphone-based GNSS positioning achieves several meter level accuracy and the up direction achieves the poorest positioning accuracy. With enough 'visual pseudolites', the visual-only positioning algorithm also achieves meter level accuracy as well and the visual-only positioning achieves the best accuracy in the up direction. In the tightly coupled solution, the positioning accuracy inherits the advantages of both techniques and thus achieving the best precision accuracy among all the three positioning schemes with its 3D RMSE 5.561m. The table also indicates that the integration of the visual range can achieve fairly good positioning results even in 3 visible GNSS satellite scenarios.

Table 1 Comparison of the position accuracy (RMSE) using different positioning schemes

Direction	GNSS Only	VL Only	GNSS/VL
North	2.463 m	4.269 m	2.758 m
East	4.176 m	6.350 m	4.641 m
Up	9.988 m	0.534 m	1.334 m
3D	11.103 m	7.671 m	5.561 m

An example of the positioning error time series

9

using the GPS-only and GPS/VL is presented in Figure 8. The figure explained how did the GPS/VL combined system improves the GPS-only positioning accuracy. Due to the poor quality of the smartphone's built-in GNSS antenna, the GPS-only positioning time series is quite noisy. By integrating with the visual range from images, the accuracy in the north direction and the up direction was dramatically improved, while the positioning accuracy in the east direction did not improve much, which may due to its poor observability in the visual ranging system.

For the tightly-coupled positioning method, when the number of observable satellites is insufficient, the visual ranging information will be involved in the positioning process as pseudo satellites. The positioning accuracy is improved (27.5%) compared with that of visual-only positioning and is more significantly improved (49.9%) subject to the GNSS-only solution. In detail, the method of tightly-coupled GNSS/vision ensures that the positioning error in each direction is within a certain range, to provide stable and high-precision positioning results in GNSS-hostile environments.
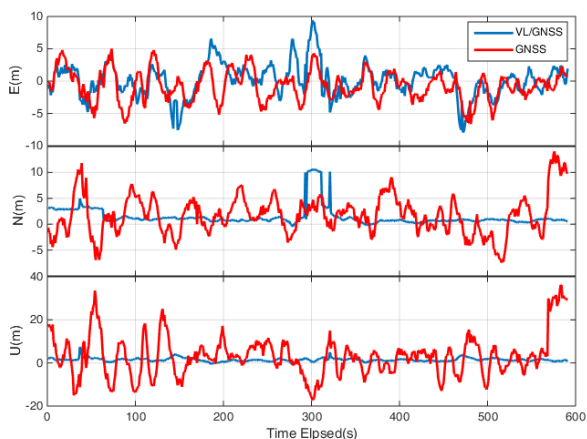

Figure 8 Positioning error time series comparison of the GNSS and GNSS/VL combined positioning.

5.3 Impact of visible GNSS satellite number on the positioning accuracy

In order to investigate the performance of the tightly-coupled approaches in the GNSS denied environment, we designed an experiment to investigate the positioning accuracy with different visible satellite numbers. Figure 9 shows the positioning error of the tightly coupled positioning

method with different visible GNSS satellite numbers. We selected two visible GNSS satellites and three visible GNSS satellite scenarios respectively and integrating the GNSS signals with the visual ranges from different images and then evaluate their positioning accuracy. Since different images have different feature numbers, so the positioning accuracy is different. The RMSE of positioning results from different numbers of satellites are summarized in Table 2. The results show that 3 visible satellite scenario generally achieves better accuracy than the 2 visible satellite scenarios. With 3 visible GNSS satellites, the precision improvement achieves about 33.5%~57.3%, which indicates the GNSS satellite has a substantial contribution to the integrated positioning results, particularly when the GNSS visible satellite number is small. The number and the geometry of the visual pseudolite also have an impact on the final positioning accuracy. Integration of one GNSS signal with the visual ranges is meaningless since the GNSS signals have no contribution to the positioning results due to the presence of the receiver clock parameter. For more visible satellite number scenario, GNSS can provide positioning service alone, but integrating with the visual ranges still benefit for the precision improvement in the up direction.
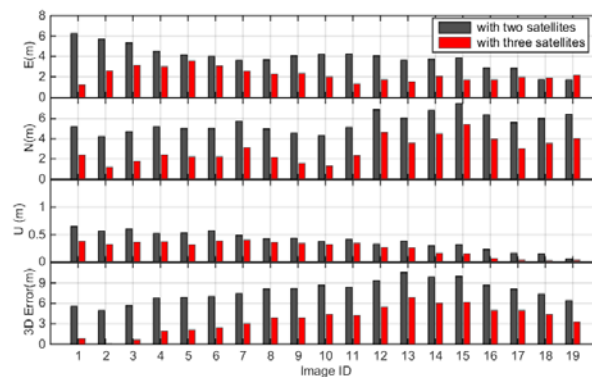

Figure 9 Positioning error of tightly-coupled method with different visible satellite number

Table 2 Comparison of the positional RMSE of the tightly-coupled method based on different visible satellite number

RMSE(m)	2 GNSS satellites	3 GNSS satellites	Improvement
North	6.703	4.458	33.49%

East	5.580	2.382	57.31%
Up	0.885	0.413	53.35%
3D	8.766	5.072	42.15%

5.4 Feature number impact of the positioning accuracy

There is no doubt that the matched feature number is an important factor for the tightly combined positioning system. An experiment was designed to investigate this problem. We selected 3 visible satellites in this experiment according to the discussion in section 5.3. We tested 3 visual GNSS satellites combined with 5, 7 and 9 visual pseudolite respectively and the results are presented in Figure 10. In this test, 11 images are used in the data processing. The figure shows that more visible pseudolite may improve the tightly coupled positioning accuracy, but it also depends on the geometry distribution of the visual pseudolite. For example, 7 visual pseudolite cases did not achieve precision improvement in the east direction comparing to the 5 visual pseudolite cases. 9 visual pseudolite case even achieves poorer accuracy in the up direction than 7 visual pseudolite cases. It concludes that more visual pseudolite may improve the positioning accuracy, but it also depends on the geometry distribution of the visual pseudolite. This conclusion may not hold for dense visual pseudolite cases since the positioning accuracy will convergence to the visual-only solution as the visual pseudolite number increases.
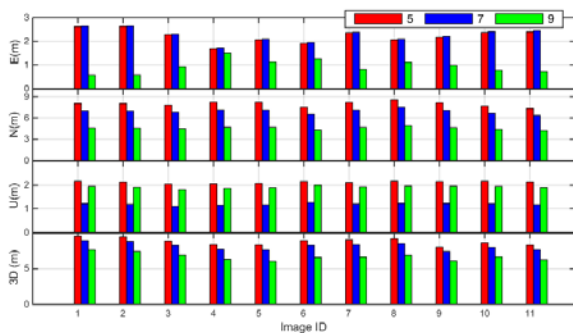

Figure 10 Positioning accuracy of tightly-coupled method with different feature number

6. Conclusion

Positioning in a GNSS denied environment has been a challenge for a long time. This paper proposed a smartphone-based GNSS/monocular camera tightly-coupled positioning method to solve the GNSS low positioning availability problem in GNSS-hostile environments. With this method, the feature points are viewed as 'visual pseudolites' and the visual ranges are obtained via an image processing procedure. Then, these visual ranges are combined with the GNSS ranges to calculate the user's positions. The performance of the tightly-coupled approach was verified by a field test using a smartphone.

The results showed that the tightly-coupled positioning method achieved about ±5.6-meter (1σ, 3D) positioning accuracy in GNSS-hostile environments, and its positioning accuracy is significantly higher than that of GNSS-only and vision-only solutions. The tightly-coupled solution delivered a resilient solution in the GNSS denied environment. In comparison with the GNSS only solution, the tightly-coupled solution has dramatically improved the positioning accuracy in the vertical direction. The tightly coupled positioning accuracy varies with the number of the feature points and the visible GNSS satellites.

This paper only builds a proof-of-concept system to verify the validness of the tightly coupled approach, many issues are not remained unconsidered in this study due to the time limit, e.g. the distribution form of feature points, the synchronization of the GNSS and the camera, the positioning performance in the mobile platform, etc. With an improved feature matching algorithm, it is possible to integrate the GNSS ranges with the real-time video to conduct kinematic positioning. A measurable 3D map can also be added to substantially improve the service coverage of the tightly coupled solution and these topics will be in our next step.

References

[1]. Jafarnia-Jahromi, A.; Broumandan, A.; Nielsen, J.; Lachapelle, G. GPS Vulnerability to Spoofing Threats and a Review of Antispoofing Techniques. International Journal of Navigation and Observation 2012, 2012, 1-16.

[2]. Cao, K.; Wang, L.; Li, B.; Ma, H. A Real-time Phase Center Variation Compensation Algorithm for The Anti-jamming GNSS Antennas. IEEE Access 2020, 2020, 128705-128715.

[3]. Zhao, X.; Zhan, X.; Yan, K. GNSS vulnerabilities: simulation, verification, and mitigation platform design. Geo-spatial Information Science 2013, 16, 149-154.

[4]. Han, S.; Gong, Z.; Meng, X.; Li, C.; Gu, X. Future alternative positioning, navigation, and timing techniques: A survey. IEEE Wireless Communications 2016, 23, 154-160.

[5]. Parkinson, B.W. Assured PNT for Our Future: PTA Actions Necessary to Reduce Vulnerability and Ensure Availability. GPS World 2014, 14, 1-10.

[6]. Montillet, J.-P.; Roberts, G.W.; Hancock, C.; Meng, X.; Ogundipe, O.; Barnes, J. Deploying a Locata network to enable precise positioning in urban canyons. Journal of Geodesy 2009, 83, 91-103.

[7]. Montillet, J.-P.; Bonenberg, L.K.; Hancock, C.M.; Roberts, G.W. On the improvements of the single point positioning accuracy with Locata technology. GPS Solutions 2013, 18, 273-282.

[8]. Wang, L.; Chen, R.; Li, D.; Zhang, G.; Shen, X.; Yu, B.; Wu, C.; Xie, S.; Zhang, P.; Li, M., et al. Initial Assessment of the LEO Based Navigation signal augmentation System from Luojia-1A Satellite. Sensors 2018, 18, 3919.

[9]. Wang, L.; Chen, R.; Xu, B.; Zhang, X.; Li, T.; Wu, C. The Challenges of LEO Based Navigation Augmentation System – Lessons Learned from Luojia-1A Satellite, In China Satellite Navigation Conference (CSNC) 2019 Proceedings, Beijing, China, 2019//, 2019; Sun, J.; Yang, C.; Yang, Y., Eds. Springer Singapore: Beijing, China, pp 298-310.

[10]. Nur, K.; Feng, S.; Ling, C.; Ochieng, W. Integration of GPS with a WiFi high accuracy ranging functionality. Geo-spatial Information Science 2013, 16, 155-168.

[11]. Li, Y.; Zhuang, Y.; Lan, H.; Zhou, Q.; Niu, X.; El-Sheimy, N. A Hybrid WiFi/Magnetic Matching/PDR Approach for Indoor Navigation With Smartphone Sensors. IEEE Communications Letters 2016, 20, 169-172.

[12]. Li, T.; Zhang, H.; Gao, Z.; Niu, X.; El-sheimy, N. Tight Fusion of a Monocular Camera, MEMS-IMU, and Single-Frequency Multi-GNSS RTK for Precise Navigation in GNSS-Challenged Environments. Remote Sensing 2019, 11, 610.

[13]. Hasnur-rabiain, A.; Kealy, A.; Morelande, M. INS stochastic error detection during kinematic tests and impacts on INS/GNSS performance. Geo-spatial Information Science 2013, 16, 169-176.

[14]. Ruotsalainen, L.; Kuusniemi, H.; Bhuiyan, M.Z.H.; Chen, L.; Chen, R. A two-dimensional pedestrian navigation solution aided with a visual gyroscope and a visual odometer. GPS Solutions 2012, 17, 575-586.

[15]. Durrant-Whyte, H.; Bailey, T. Simultaneous Localization and Mapping (SLAM): Part I. IEEE Robotics & Automation Magazine 2006, 13, 99-110.

[16]. Sneha, A.; Teja, V.; Mishra, T.; Satya Chitra, K. QR Code based Indoor Navigation system for Attender Robot. EAI Endorsed Transactions on Internet of Things 2020, 6, 165519.

[17]. Li, B.; Zou, D.; Sartori, D.; Pei, L.; Yu, W. TextSLAM: Visual SLAM with Planar Text Features, In IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020; Paris, France.

[18]. Liao, X.; Chen, R.; Li, M.; Guo, B.; Niu, X.; Zhang, W. Design of a Smartphone Indoor Positioning Dynamic Ground Truth Reference System Using Robust Visual Encoded Targets. Sensors (Basel) 2019, 19, 1261.

[19]. Werner, M.; Kessel, M.; Marouane, C. Indoor positioning using smartphone camera, In

International Conference on Indoor Positioning and Indoor Navigation (IPIN), Guimaraes, Portugal, 2011; IEEE: Guimaraes, Portugal.

[20]. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications To Image Analysis and Automated Cartography. Communications of the ACM 1981, 24, 381-395.

[21]. Liu, Y.; Huang, T.S.; Faugeras, O.D. Determination of Camera Location from 2-D to 3-D Line and Point Correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence 1990, 12, 28-37.

[22]. Lowe, D.G. Object recognition from local scale-invariant features, In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999; IEEE: Kerkyra, Greece, pp 1150-1157.

[23]. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 2004, 60, 91-110.

[24]. Yagi, Y.; Yachida, M. Real-time generation of environmental map and obstacle avoidance using omnidirectional image sensor with conic mirror, In Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, HI, USA, 1991; IEEE: Maui, HI, USA, pp 160-165.

[25]. Yang, Y.; Shen, Q.; Li, J.; Deng, Z.; Wang, H.; Gao, X. Position and Attitude Estimation Method Integrating Visual Odometer and GPS. Sensors (Basel) 2020, 20, 2121.

[26]. Shi, Y.; Ji, S.; Shi, Z.; Duan, Y.; Shibasaki, R. GPS-supported visual SLAM with a rigorous sensor model for a panoramic camera in outdoor environments. Sensors (Basel) 2012, 13, 119-136.

[27]. Dusha, D.; Mejias, L. Error analysis and attitude observability of a monocular GPS/visual odometry integrated navigation filter. The International Journal of Robotics Research 2012, 31, 714-737.

[28]. Chen, X.; Hu, W.; Zhang, L.; Shi, Z.; Li, M. Integration of Low-Cost GNSS and Monocular Cameras for Simultaneous Localization and Mapping. Sensors 2018, 18, 2193.

[29]. Chen, R.; Wang, L.; Li, D.; Chen, L.; Fu, W. A Survey on the Fusion of the Navigation and the Remote Sensing Techniques. Acta Geodaetica Et Cartographic Sinica 2019, 48, 1507-1522.

[30]. Gakne, P.V.; O'Keefe, K. Tightly-Coupled GNSS/Vision Using a Sky-Pointing Camera for Vehicle Navigation in Urban Areas. Sensors (Basel) 2018, 18, 1244.

[31]. Suzuki, T. GNSS Photo Matching: Positioning using GNSS and Camera in Urban Canyon, In ION GNSS+ 2015, Tampa, Florida, 2015; Tampa, Florida.

[32]. Schreiber, M.; Konigshof, H.; Hellmund, A.-M.; Stiller, C. Vehicle Localization with Tightly Coupled GNSS and Visual Odometry, In 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 2016; IEEE: Gothenburg, Sweden.

[33]. Soloviev, A.; Venable, D. Integration of GPS and Vision Measurements for Navigation in GPS Challenged Environments, In IEEE/ION Position, Location and Navigation Symposium, Indian Wells, CA, 2010; IEEE/ION: Indian Wells, CA.

[34]. Montenbruck, O.; Steigenberger, P.; Hauschild, A. Broadcast versus precise ephemerides: a multi-GNSS perspective. GPS Solutions 2015, 19, 321-333.

[35]. Yuan, Y.; Wang, N.; Li, Z.; Huo, X. The BeiDou global broadcast ionospheric delay correction model (BDGIM) and its preliminary performance evaluation results. Navigation 2019, 1-15.

[36]. Wang, L.; Feng, Y.; Guo, J. Reliability control of single-epoch RTK ambiguity resolution. GPS Solutions 2017, 21, 591-604.

[37]. Wang, L.; Feng, Y.; Wang, C. Real-Time Assessment of GNSS Observation Noise with Single Receivers. Journal of Global Positioning Systems 2013, 12, 73-82.

[38]. Li, W.; Huang, W.; Breier, M.; Merhof, D. Have we Underestimated the Power of Image Undistortion?, In IEEE International Conference on Image Processing (ICIP) Phoenix, AZ,, 2010;

Phoenix, AZ,, pp 2946-2950.

[39]. Zhang, Z. Camera Calibration with One-Dimensional Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 2004, 26, 892-899.

[40]. Wu, D.; Chen, R.; Chen, L. Visual Positioning Indoors: Human Eyes vs. Smartphone Cameras. Sensors 2017, 17, 2645.

[41]. Zhang, Z. A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000, 22, 1330-1334.

[42]. Yusefi, A.; Durdu, A.; Sungur, C. Görsel Odometride SIFT, SURF, FAST, STAR ve ORB özellik algılama algoritmalarının Performans ve Takas Değerlendirmesi. European Journal of Science and Technology 2020, 2020, 455-460.

[43]. Zhou, H.; Yuan, Y.; Shi, C. Object Tracking using SIFT Features and Mean Shift. Computer Vision and Image Understanding 2009, 113, 345-352.

[44]. Lillesand, T.M.; Kiefer, R.W. Remote sensing and image interpretation. John Wiley & Sons: New York, 1979.

Authors

Han Xu received the BSc. degree in National geoinformation monitoring from Wuhan University in 2018. She is currently pursuing the master degree at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. Her research interest includes surveying data processing and multi-system fusion positioning.

Lei Wang received Ph.D. degree in Electronical Engineer and Computer Science (EECS) from Queensland University of Technology, Australia in 2015. He is currently an associate research fellow in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. His research interest includes GNSS precise positioning, LEO orbit determination and LEO navigation augmentation and indoor positioning.

Ruizhi Chen is currently a Professor and the Director of the State Key Laboratory of Information Engineering in surveying, mapping, and remote sensing with Wuhan University. He was an Endowed Chair Professor with Texas A&M University Corpus Christi, USA, the Head and a Professor of the Department of Navigation and Positioning, Finnish Geodetic Institute, and the Engineering Manager of Nokia, Finland. He has published two books and more than 200 scientific papers. His current research interests include indoor positioning, satellite navigation, and location-based services.

Haitao Zhou received the M.S. degree in Geodesy and Surveying Engineering from Information Engineering University, Zhengzhou, China in 2016. He is pursuing the Ph.D. degree in Geodesy and Surveying Engineering from Wuhan University, Hubei, China. His research interests include surveying data processing and GNSS precise positioning.

Tao Li was born in Jiangxi, China in 1995. He received the B.Sc. degree in surveying and mapping engineering from China University of Mining and Technology in 2018. Now he is a Ph.D. student at Wuhan University. His research interests include troposphere delay modeling, ionosphere, and the LEO navigation augmentation technology.

Yi Han received M.sc degree in Geodesy and surveying engineering from Wuhan University in 2019, He is currently a Ph.D candidate at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. His current research interest includes the aplications of LEO constellation optimization and ionosphere.