Continuous and precise positioning in urban environments by tightly coupled integration of GNSS, INS and vision

Xingxing Li[1]*, Shengyu Li[1], Yuxuan Zhou[1], Zhiheng Shen[1], Xuanbin Wang[1], Xin Li[1] and Weisong Wen[2]

***Abstract*— Accurate, continuous and seamless state estimation is the fundamental module for intelligent navigation applications, such as self-driving cars and autonomous robots. However, it is often difficult for a standalone sensor to fulfill the demanding requirements of precise navigation in complex scenarios. To fill this gap, this paper proposes to exploit the complementariness of the GNSS, inertial measurement unit (IMU) and vision via a tightly coupled integration method, aiming to achieve continuous and accurate navigation in urban environments. Specifically, the raw GNSS carrier phase and pseudorange measurements, IMU data, and visual features are directly fused at the observation level through a centralized Extended Kalman Filter (EKF) to make full use of the multi-sensor information and reject potential outlier measurements. Furthermore, the widely used high-precision GNSS models including precise point positioning (PPP) and real-time kinematic (RTK) are unified in the proposed integrated system to increase usability and flexibility. We validate the performance of the proposed method on several challenging datasets collected in urban canyons and compare against the loosely coupled and state-of-the-art methods.**

I. INTRODUCTION AND RELATED WORKS

Accurate and continuous state estimation is of crucial importance for systems with navigation requirements, especially for autonomous driving [1]. However, each sensor has its inherent vulnerabilities and limitations, making it difficult to cover various complex scenarios in practical navigation applications. Thus, multi-sensor fusion, which can take advantage of complementary properties from heterogeneous sensors, is accepted as a feasible solution to this issue [2].

Nowadays, the integration of global navigation satellite system (GNSS) and inertial navigation system (INS) is the most widely used localization technology for onboard platforms [3]. For GNSS, differential GNSS (e.g., real-time kinematic, RTK) and precise point positioning (PPP) techniques have been widely applied in GNSS/INS integration for high-precision navigation tasks [4]. The GNSS can provide global position information to prevent inertial sensors from drifting, while INS could offer high-frequency and continuous navigation outputs to bridge GNSS data gaps and identify cycle slips, thereby improving GNSS reliability. Many researchers have presented detailed implementations and performance analyses of the loosely/tightly coupled GNSS/INS integration methods [4]-[6]. The loosely coupled systems process the measurements from each sensor separately and fuse the navigation results within an estimator, while tightly coupled systems jointly optimize the raw measurements in a single consistent estimator, which could fully exploit the available information and provide natural robustness to sensor degradation [7]. However, due to the vulnerability of GNSS to signal-degrading circumstances and poor satellite visibility, the main drawback of GNSS/INS integration is the rapid error accumulation during GNSS outages, especially when the low-cost microelectromechanical system (MEMS) IMU is equipped.

Generally, outdoor environments with poor GNSS availability are rich in environmental features, which could be used for state estimation. Accordingly, camera, as an exteroceptive sensor, could be integrated with GNSS and INS to augment the positioning performance in both suburban and dense urban areas. Due to the simplicity of implementation, the loosely coupled methods of GNSS, INS and vision have been well investigated in recent years. Angelino et al. (2012) proposed a loosely coupled GPS/INS/visual-odometry (VO) integration method [8], which fused the position and attitude results from each subsystem. Similarly, Chiang et al. (2020) developed a multi-sensor fusion framework using two-step Extended Kalman Filters (EKF) with non-holonomic constraint (NHC) constraints, in which meter-level positioning accuracy could be achieved in urban areas [9]. Instead of simply fusing navigation results from both the visual and GNSS submodules, the authors in [10] introduced the estimated parameters of visual feature position into a loosely coupled GPS/INS integration, whose performance was more stable than the VO/INS and GPS/INS methods in either GPS-denied or low-texture environments. To avoid information loss in the loosely coupled methods and provide better robustness against standalone-sensor outliers [11], Chu et al. (2012) tightly integrated the raw measurements from GPS pseudorange, MEMS-IMU and monocular camera for navigation in harsh signal environments [12], where the GPS differential positioning technique was utilized to resolve the scale ambiguity problem of the monocular camera. Based on a priori feature map, the geometric constraints constructed from extracted visual landmarks and lane lines were proven to effectively suppress INS divergence in the tightly coupled method during long-term GNSS outages [13].

Another implementation to integrate GNSS, INS and vision is to introduce GNSS information into a mature visual-inertial navigation system (VINS). Currently, substantial research efforts have been devoted to various VINS estimators,

This work was supported in part by the National Key Research and Development Program of China (2021YFB2501102), the National Natural Science Foundation of China (Grant 41974027), and the Sino-German mobility programme (Grant No. M-0054).

[1] School of Geodesy and Geomatics, Wuhan University, China

[2] Department of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, Hong Kong, China

* Corresponding author. Email: xxli@sgg.whu.edu.cn

such as VINS-Mono [14], MSCKF [15] and ORB-SLAM3 [16], which could achieve impressive performance in the local pose estimation. As the local VINS method commonly has drawbacks like drifting, the introduction of GNSS could provide nondrifting position information and transform the VINS from local pose estimation to global pose estimation. Following this trend, Qin et al. (2019) fused GPS position results and the relative poses of visual-inertial-odometry (VIO) in a global optimization thread [17]. This fusion was regarded as an alignment problem and the local-to-global transformation was estimated online, which is a commonly used method in some recent studies [18], [19]. Nevertheless, the estimation of the local-to-global transformation could be sensitive to noise in GNSS measurements and its observability would decrease in GNSS-degraded environments. Different from Qin et al. (2019), Liao et al. (2021) converted the VIO model from the local frame to the global frame and employed differential GNSS results to achieve meter-level positioning accuracy in GNSS-challenging environments [20]. Based on the GNSS on-and-off outage assumption, Won et al. (2014) demonstrated that VINS can provide highly available and lane-level vehicle navigation by fusing GNSS pseudorange observations [21]. A globally-referenced initialization method was proposed in [22] to solve solves the problem of relating the coordinate systems for position estimates based on two disparate sensors. Apart from filter-based methods, Cao et al. (2021) proposed an optimization-based GNSS/INS/Vision integrated system named GVINS, in which the GNSS pseudorange measurements, visual constraints, and inertial constraints were jointly optimized under a consistent factor graph framework [23].

The previous work on the integration of GNSS, INS and vision mainly focused on either the loosely coupled methods or using the GNSS pseudorange measurements. As a result, only meter-level positioning accuracy can be achieved. In this paper, we propose a tightly coupled GNSS/INS/Vision integration method to achieve accurate and continuous navigation in urban environments. To implement this integrated system, a centralized EKF is utilized to directly fuse the GNSS carrier phase measurements, MEMS-IMU data and stereo visual features at raw-data level. Instead of directly utilizing position results derived from the standalone GNSS module, the GNSS carrier phase observations are pre-processed and directly used for measurement updates. This tightly coupled method enables the system to gain information from GNSS even with only one satellite in view. To cover the cases with or without nearby base stations, both high-precision PPP and RTK models are unified to increase the usability and flexibility of the proposed method. Several vehicle experiments in different scenarios of urban environments were designed to evaluate the performance of the proposed method. We highlight the contributions of this paper as follows:

- A tightly coupled GNSS/INS/Vision integration method is proposed. Compared with the traditional decoupling methods, this tight integration method fully exploits the complementary properties from heterogeneous sensors, which makes it own the promising potential to ensure the navigation performance even in GNSS-challenging environments. Moreover, unlike current work that treats fusion as an alignment problem, no additional local-to-global transformation is included in the filter states and the obtained navigation outputs are under the Earth-centered Earth-fixed (ECEF) frame, rather than a self-defined local tangent plane frame, which could avoid overparameterization and possible numerical instability.
- To the best of our knowledge, this is the first work that unifies the PPP and RTK models in the GNSS/INS/Vision integration method, both of which use high-precision carrier phase measurements and have the potential to achieve decimeter-level positioning accuracy.
- The proposed method is extensively validated in real-world experiments with different GNSS observing conditions. The results indicate that the proposed tightly coupled integration method outperforms the corresponding loosely coupled method and state-of-the-art VINS-Fusion [17] in terms of positioning and attitude accuracy.

II. SENSOR MODELS

In this section, the sensor models involved in the proposed integrated system are introduced in detail.

A. INS Model

During the integration processing, the IMU outputs are used to propagate the system states, which is known as the INS mechanization. The simplified INS linearized dynamic model can be expressed as [7]:

$$\begin{bmatrix} \delta\dot{\boldsymbol{\theta}}_b^e \\ \delta\dot{\boldsymbol{v}}_b^e \\ \delta\dot{\boldsymbol{p}}_b^e \end{bmatrix} = \begin{bmatrix} -\lfloor\boldsymbol{\omega}_{ie}^e\rfloor_\times\delta\boldsymbol{\theta}_b^e - \boldsymbol{R}_b^e\delta\boldsymbol{\omega}_{ib}^b \\ \left\lfloor\boldsymbol{R}_b^e\boldsymbol{f}^b\right\rfloor_\times\delta\boldsymbol{\theta}_b^e - \lfloor 2\boldsymbol{\omega}_{ie}^e\rfloor_\times\delta\boldsymbol{v}_b^e + \boldsymbol{R}_b^e\delta\boldsymbol{f}^b + \delta\boldsymbol{g}^e \\ \delta\boldsymbol{v}_b^e \end{bmatrix} \quad (1)$$

where $\delta\dot{\boldsymbol{\theta}}_b^e$, $\delta\dot{\boldsymbol{v}}_b^e$ and $\delta\dot{\boldsymbol{p}}_b^e$ denote the derivatives of attitude, velocity and position errors, respectively; $\boldsymbol{R}_b^e$ corresponds to the rotation matrix from IMU body (b)-frame to e-frame; $\boldsymbol{\omega}_{ie}^e$ is the angular velocity of earth rotation; $\delta\boldsymbol{g}^e$ represents the error vector of gravity in the e-frame; $\delta\boldsymbol{f}^b$ and $\delta\boldsymbol{\omega}_{ib}^b$ are the errors of accelerometer and gyroscope measurements, respectively; The symbol $\lfloor\cdot\rfloor_\times$ is the cross-product. As to the IMU sensor bias model, the accelerometer bias vector $\boldsymbol{b}_a$ and gyroscope bias vector $\boldsymbol{b}_g$ are estimated. Their variations are modeled as random walks driven by white noise.

B. GNSS Measurement Model

The PPP and RTK models are considered in this work, both of which could be derived from the undifferenced observation equations of pseudorange and carrier phase [24]:

$$\begin{cases} P = \rho + c\,(dt_r - dt^s) + I + T + \varepsilon_P \\ L = \rho + c\,(dt_r - dt^s) - I + T + \lambda N + \varepsilon_L \end{cases} \quad (2)$$

where P and L are the pseudorange and carrier phase measurements, respectively; ρ denotes the geometric distance

between the satellite and receiver; c is the speed of light; dt_r and dt^s represent the offsets of the receiver clock and satellite clock, respectively; For simplicity, the index of the satellite s have been omitted; I and T denote ionospheric and tropospheric delays along the signal path, respectively; N and λ represent the carrier phase ambiguity and the corresponding wavelength, respectively; ε_P and ε_L are measurement noise and multipath errors of the pseudorange and carrier phase, respectively.

The accuracy of carrier phase observations could reach millimeter-level, which is much less noisy than the pseudorange observations. Thus, both PPP and RTK techniques have the potential to achieve decimeter-level positioning accuracy in practical navigation applications. In the PPP processing, the ionospheric-free (IF) combination is widely used to eliminate the first-order effects of the ionosphere, which can be written as:

$$\begin{cases} P_{IF} = \rho + c\,(dt_r - dt^s) + T + \varepsilon_{P,IF} \\ L_{IF} = \rho + c\,(dt_r - dt^s) + T + \lambda N_{IF} + \varepsilon_{L,IF} \end{cases} \tag{3}$$

Meanwhile, the precise orbit and clock products are required in the PPP processing to mitigate the satellite orbit error in ρ and satellite clock error dt^s. Generally, it takes about ten minutes for PPP to converge to sub-decimeter-level accuracy.

As for RTK processing, one or more base stations with precisely known coordinates are required. Different from PPP using precise products, the double-differenced operation is used to eliminate not only the satellite orbit and clock errors but the receiver clock error as well, which can be expressed as:

$$\begin{cases} \nabla\Delta P = \nabla\Delta\rho + \nabla\Delta I + \nabla\Delta T + \nabla\Delta\varepsilon_P \\ \nabla\Delta L = \nabla\Delta\rho - \nabla\Delta I + \nabla\Delta T + \lambda\nabla\Delta N + \nabla\Delta\varepsilon_L \end{cases} \tag{4}$$

where $\nabla\Delta$ is the double-differenced operator. For the case of short baselines within 10km, the tropospheric and ionospheric delays could also be eliminated and neglected. In this case, the parameters to be estimated are the receiver position and integer ambiguities. With the aiding of nearby base stations, the RTK could rapidly or even instantaneously achieve decimeter-level accuracy in open areas. The positioning accuracy could be further improved by recovering the double-differenced carrier phase ambiguities $\nabla\Delta N$ to integer values, which is known as the ambiguity resolution [25].

C. Visual Measurement Model

For the visual measurement model, we follow the well-known multi-state constraint Kalman Filter (MSCKF) framework [15], whose key is to maintain a sliding window of camera poses without including visual features in the state vector. On this basis, the visual geometric constraints could be constructed on multiple camera poses that observe the same feature. Consider the case that a single feature f_j is observed by the stereo camera at timestep i, the feature observations can be written as:

$$\boldsymbol{z}_i^j = \begin{bmatrix} u_{i,c1}^j \\ v_{i,c1}^j \\ u_{i,c2}^j \\ v_{i,c2}^j \end{bmatrix} = \begin{bmatrix} \frac{1}{Z_{i,c1}^j} & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \frac{1}{Z_{i,c2}^j} \end{bmatrix} \begin{bmatrix} X_{i,c1}^j \\ Y_{i,c1}^j \\ X_{i,c2}^j \\ Y_{i,c2}^j \end{bmatrix} \tag{5}$$

where $\boldsymbol{z}_i^j = [\ u_{i,c1}^j \quad v_{i,c1}^j \quad u_{i,c2}^j \quad v_{i,c2}^j\ ]^\top$ represents the feature coordinates on the normalized projective plane of the left and right cameras; $[X_{i,ck}^j, Y_{i,ck}^j, Z_{i,ck}^j]^\top k, \in \{1,2\}$ is the feature positions expressed in the left and right camera (c) frames, which are given by:

$$\begin{aligned} \begin{bmatrix} X_{i,c1}^j & Y_{i,c1}^j & Z_{i,c1}^j \end{bmatrix}^\top &= \left(\boldsymbol{R}_{c1}^e\right)^\top \left(\boldsymbol{p}_j^e - \boldsymbol{p}_{c1}^e\right) \\ \begin{bmatrix} X_{i,c2}^j & Y_{i,c2}^j & Z_{i,c2}^j \end{bmatrix}^\top &= \boldsymbol{R}_{c1}^{c2} \left(\boldsymbol{p}_j^{c1} - \boldsymbol{p}_{c1}^{c2}\right) \end{aligned} \tag{6}$$

where $\boldsymbol{R}_{c1}^e$ and $\boldsymbol{p}_{c1}^e$ denote the attitude and position of the left camera in the e-frame, respectively; $\boldsymbol{R}_{c1}^{c2}$ and $\boldsymbol{p}_{c1}^{c2}$ are the transformation from the left camera frame to the right camera frame, which can be accurately calibrated beforehand [26] and are assumed to be constants. It is worth noting that the feature position $\boldsymbol{p}_j^e$ in the e-frame is estimated through least-square multi-view triangulation and is then eliminated by the null space mapping to decorrelate it with the camera poses in the sliding window [27].

III. TIGHTLY COUPLED GNSS/INS/VISION INTEGRATED SYSTEM

In this section, the system overview is first given to present the tightly coupled GNSS/INS/Vision integrated system. Subsequently, the state description, state prediction, and measurement updates are introduced following the standard EKF routines.

A. System Overview

An overview of the proposed GNSS/INS/Vision integration method is presented in Fig. 1. After the initial dynamic alignment and initialization of the integration filter, the INS mechanization is used for state prediction and the system covariance would also be propagated. The high-frequency INS-predicted pose estimation will assist in the cycle slip [28] and outlier detection [7] in the GNSS pre-processing as well as the feature tracking in the image processing. When a new image is incoming, the feature extraction and tracking will be performed following [15]. Meanwhile, the current camera pose will be added into the state vector and the corresponding covariance will also be augmented. When the pre-processed GNSS or visual observations are available, the corresponding measurement updates will be carried out to correct the predicted states. Finally, the estimated accelerometer and gyroscope biases will be fed back to correct the next IMU observations [7].

B. State Description

Based on the sensor models illustrated above, the entire state vector of the proposed tightly coupled integrated system

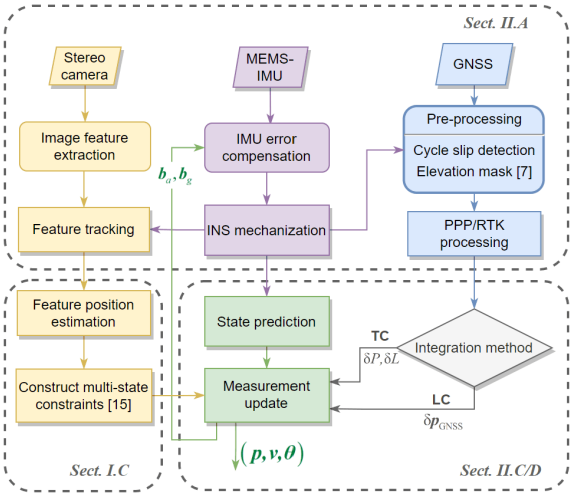

Fig. 1. Algorithm structure of the proposed GNSS/INS/Vision integrated system. The terms "LC" and "TC" represent the loosely coupled and tightly coupled integration, which use processed GNSS positioning results or raw GNSS measurements for measurement updates, respectively.

is composed of INS states ($\boldsymbol{x}_{INS}$), GNSS states ($\boldsymbol{x}_{GNSS}$) and camera states ($\boldsymbol{x}_{cam}$), which can be defined as:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_{INS} & \boldsymbol{x}_{GNSS} & \boldsymbol{x}_{cam} \end{bmatrix}^\top \tag{7}$$

Specifically, the INS states could be written as:

$$\boldsymbol{x}_{INS} = \begin{bmatrix} \delta\boldsymbol{\theta}_b^e & \delta\boldsymbol{v}_b^e & \delta\boldsymbol{p}_b^e & \delta\boldsymbol{b}_a & \delta\boldsymbol{b}_g \end{bmatrix}^\top \tag{8}$$

where $\delta\boldsymbol{\theta}_b^e$, $\delta\boldsymbol{v}_b^e$, $\delta\boldsymbol{p}_b^e$, $\delta\boldsymbol{b}_a$ and $\delta\boldsymbol{b}_g$ are the error states of attitude, velocity, position, accelerometer bias and gyroscope bias described in Section II-A.

The GNSS state vector for the PPP model can be written as:

$$\boldsymbol{x}_{GNSS,ppp} = \begin{bmatrix} \delta dt_r & \delta\boldsymbol{ISB} & \delta d_{zwd} & \delta\boldsymbol{N}_{IF} \end{bmatrix}^\top \tag{9}$$

where $\boldsymbol{ISB}$ denotes the inter-system biases due to different signal structures and hardware delays of each GNSS system [24]; d_{zwd} is the wet component of the zenith tropospheric delay; $\delta\boldsymbol{N}_{IF} = \begin{bmatrix} \delta\boldsymbol{N}_{IF}^G & \delta\boldsymbol{N}_{IF}^E & \delta\boldsymbol{N}_{IF}^C \end{bmatrix}^\top$ is the ambiguity vector of GPS (G), Galileo (E) and BDS (C) systems.

As for the RTK model, the corresponding GNSS state vector could be written as:

$$\boldsymbol{x}_{GNSS,RTK} = [\delta\nabla\Delta\boldsymbol{N}]^\top \tag{10}$$

where $\nabla\Delta\boldsymbol{N}$ is the double-differenced carrier phase ambiguity shown in (4). More details about the parameterization of PPP and RTK used in this paper can refer to [6] and [25], respectively.

As described in Section II, $\boldsymbol{x}_{cam}$ is a sliding window of historical camera poses for the construction of visual geometry constraints, which can be written as:

$$\boldsymbol{x}_{cam} = \begin{bmatrix} \delta\boldsymbol{\theta}_{c_1}^e & \delta\boldsymbol{p}_{c_1}^e & \delta\boldsymbol{\theta}_{c_2}^e & \delta\boldsymbol{p}_{c_2}^e & \cdots & \delta\boldsymbol{\theta}_{c_\eta}^e & \delta\boldsymbol{p}_{c_\eta}^e \end{bmatrix}^\top \tag{11}$$

where η is the pre-defined sliding window size.

C. State Prediction

In this step, the INS mechanization is used for state prediction and the corresponding covariance propagation can be calculated according to the process model [4], which can be described as:

$$\begin{bmatrix} \dot{\boldsymbol{x}}_{INS} \\ \dot{\boldsymbol{x}}_{GNSS} \\ \dot{\boldsymbol{x}}_{cam} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_{INS} & & \\ & \boldsymbol{F}_{GNSS} & \\ & & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{INS} \\ \boldsymbol{x}_{GNSS} \\ \boldsymbol{x}_{cam} \end{bmatrix} + \boldsymbol{w} \tag{12}$$

where $\boldsymbol{F}_{INS}$ and $\boldsymbol{F}_{GNSS}$ are the system matrices of INS and GNSS states, respectively; $\boldsymbol{w}$ is the process noise. Specifically, $\boldsymbol{F}_{INS}$ could be directly derived from (1), and the specific form of $\boldsymbol{F}_{GNSS}$ could refer to [6] for PPP and [25] for RTK. Moreover, the historical camera poses in the sliding window are seen as constants without process noise.

At each imaging moment, the state vector will be augmented with the current left camera pose for further construction of visual constraints. Its initial value can be obtained by:

$$\begin{cases} \boldsymbol{R}_c^e = \boldsymbol{R}_b^e \boldsymbol{R}_c^b \\ \boldsymbol{p}_c^e = \boldsymbol{R}_b^e \boldsymbol{p}_c^b + \boldsymbol{p}_b^e \end{cases} \tag{13}$$

where $\boldsymbol{R}_c^b$ and $\boldsymbol{p}_c^b$ are the extrinsic parameters between camera and IMU, which are calibrated offline following [26].

D. GNSS Measurement Update

Since the IMU center has a different reference point to the GNSS antenna phase center, the lever-arm correction should be considered following [29]. On this basis, the residual vector of PPP measurement update using GPS, Galileo and BDS observations can be unified as:

$$\boldsymbol{r}^{PPP} = \begin{bmatrix} P_{GNSS,IF} - \hat{P}_{INS,IF} \\ L_{GNSS,IF} - \hat{L}_{INS,IF} \end{bmatrix} = \boldsymbol{H}_{INS}^{PPP}\boldsymbol{x}_{INS} + \boldsymbol{H}_{GNSS}^{PPP}\boldsymbol{x}_{GNSS} + \boldsymbol{n}^{PPP} \tag{14}$$

where $\boldsymbol{H}_{INS}^{PPP}$ and $\boldsymbol{H}_{GNSS}^{PPP}$ are the Jacobians of INS states and GNSS states respectively, and the readers may refer to [4] for details; $\boldsymbol{n}^{PPP}$ is the noise vector of PPP measurements; The subscripts GNSS and INS are the original and INS-predicted GNSS measurements, respectively.

If the RTK model is applied, the corresponding residual vector could be expressed as:

$$\boldsymbol{r}^{RTK} = \begin{bmatrix} \nabla\Delta P_{GNSS} - \nabla\Delta\hat{P}_{INS} \\ \nabla\Delta L_{GNSS} - \nabla\Delta\hat{L}_{INS} \end{bmatrix} = \boldsymbol{H}_{INS}^{RTK}\boldsymbol{x}_{INS} + \boldsymbol{H}_{GNSS}^{RTK}\boldsymbol{x}_{GNSS} + \boldsymbol{n}^{RTK} \tag{15}$$

where $\boldsymbol{H}_{INS}^{RTK}$ and $\boldsymbol{H}_{GNSS}^{RTK}$ are respectively the Jacobians of INS states and GNSS states, whose specific forms could refer to [25]; $\boldsymbol{n}^{RTK}$ is the noise vector of RTK measurements. Moreover, considering that raw GNSS measurements can be pretty noisy in urban environments, a dynamic DIA test [30], which is based on the probabilistic features of the posteriori residuals, is used to filter out GNSS gross errors.

E. Visual Measurement Update

There are two different mechanisms to determine the execution of the visual measurement updates. Adhering to [15], all measurements of the same tracking-lost feature are used to update all involved camera poses that observe the same feature. Moreover, if the number of camera poses in

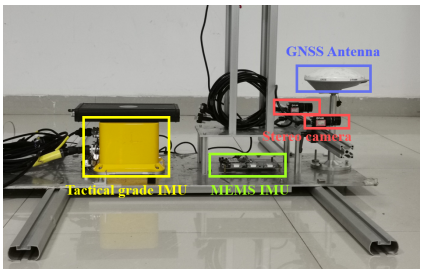

Fig. 2. Illustration of the experimental equipment.

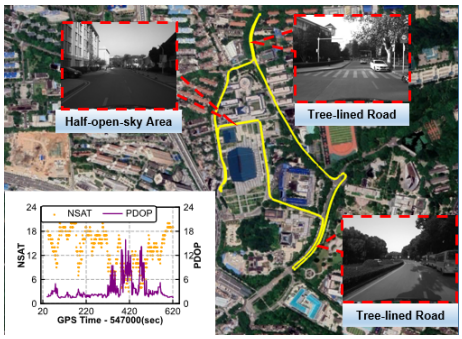

Fig. 3. The overview trajectory, typical scenarios and GNSS availability of the GNSS half-open-sky experiment. During the period from 50 s to 200 s, the vehicle ran into a GNSS open-sky area, in which the average number of available satellites (NSAT) and position dilution of precision (PDOP) are 16.13 and 2.09, respectively. During the period from 300 s to 500 s, the vehicle drove into the tree-lined road. The corresponding NSAT reduces to 9.23 and the corresponding PDOP degenerates to 5.30.

the sliding window reaches the predefined limit, two camera poses are selected by the parallax [27] and marginalized to perform delayed visual measurement updates. By stacking the feature's all measurements similar to (5) in a consecutive image sequence, the whole visual residual vector could be written as:

$$\boldsymbol{r}^j = \boldsymbol{z}^j - \hat{\boldsymbol{z}}^j = \boldsymbol{H}^j_{cam}\boldsymbol{x}_{cam} + \boldsymbol{H}^j_{f_j}\boldsymbol{p}^e_j + \boldsymbol{n}^j \tag{16}$$

where $\boldsymbol{H}^j_{cam}$ and $\boldsymbol{H}^j_{f_j}$ are the Jacobians of camera states and feature position, respectively. To eliminate the correlation between the feature position and camera poses, the expression in (16) can be projected to the left null space of $\boldsymbol{H}^j_{f_j}$. Thus, the modified residual vector could be re-written as:

$$\bar{\boldsymbol{r}}^j = \bar{\boldsymbol{H}}^j_{cam}\boldsymbol{x}_{cam} + \bar{\boldsymbol{n}}^j \tag{17}$$

Moreover, as the sampling rates of GNSS and camera may be different, the corresponding GNSS and visual measurement updates could be carried out asynchronously.

IV. REAL-WORLD EXPERIMENTS

There are altogether six different configurations considered in the evaluation, which are summarized in Table I. Since there are few open-sourced GNSS/INS/Vision integration methods, the optimization-based VINS-Fusion [17] is selected as the state-of-the-art method for comparison. To evaluate the effectiveness of using raw GNSS observations, a loosely coupled method [20] similar to our proposed method, but using processed GNSS positioning results for measurement updates, is also considered. In all RTK-related configurations, the ambiguity fixed rate is calculated following [30] and the classical AR ratio test [31] is used to ensure the accuracy of fixed ambiguity. To evaluate the navigation performance, the root mean square errors (RMSEs), maximum errors and standard deviations (STD) of position with respect to the reference are calculated.

TABLE I

DIFFERENT CONFIGURATIONS IN THE COMPARSION.

Method	Description
VINS-Fusion(PPP)	The open-sourced VINS-Fusion [17] using PPP positioning results for fusion
LC-GIV(PPP)	The loosely coupled (LC) GNSS/INS/Vision integration using PPP positioning results for fusion
TC-GIV(PPP)	The proposed tightly coupled (TC) GNSS/INS /Vision integration using PPP measurements for fusion
VINS-Fusion(RTK)	The open-sourced VINS-Fusion using RTK positioning results for fusion
LC-GIV(RTK)	The loosely coupled GNSS/INS/Vision integration using RTK positioning results for fusion, which is identical to the method proposed in [20]
TC-GIV(RTK)	The proposed tightly coupled GNSS/INS/Vision integration using RTK measurements for fusion

A. Setup

Fig. 2 shows the hardware platform for data acquisition. Apart from the rover GNSS receiver (Septentrio PolaRx5) mounted in the hardware platform, another GNSS receiver with precisely known coordinates was set up as a base station in the open area. The raw data from the two GNSS receivers, MEMS-IMU (ADIS-16470) and a stereo camera (FLIR BFS-PGE-31S4C) are processed to obtain the GNSS/INS/Vision integration following the proposed method. The sampling rates of GNSS, IMU and camera are 1 Hz, 200 Hz and 10 Hz, respectively. To generate the reference trajectory, a Tactical grade IMU (StarNeto XW-GI7660) is mounted on the hardware platform. On this basis, the raw data from the Tactical grade IMU and two GNSS receivers are used to obtain the smoothed solutions of the tightly coupled multi-GNSS post-processing kinematic (PPK)/INS through commercial software Inertial Explorer [32], which is taken as the reference trajectory in the experiments. The overall positioning accuracy of the reference trajectory can be maintained at the centimeter-level and the fixed rate is more than 90%, which can facilitate a fair and reliable performance comparison between different algorithms in the following evaluation.

B. GNSS Half-open-sky Experiment

The first real-world experiment was conducted around a stadium on the campus. According to Fig. 3, this scenario is a typical outdoor environment with a half-opened area on one side and some low-rise buildings or trees on the other side. In this case, the satellites with high elevations could still be tracked continuously. Compared with the open-sky environment during the period from 50 s to 200 s, the average NSAT and PDOP degrades to 9.23 and 5.30 in the period

TABLE II

POSITION AND ATTITUDE RESULTS OF DIFFERENT CONFIGURATIONS IN THE GNSS HALF-OPEN-SKY EXPERIMENT.

Method	Position RMSE (m)			Maximum Error (m)			STD (m)			Yaw RMSE	Fixed
	East	North	Up	East	North	Up	East	North	Up	(deg)	Rate
VINS-Fusion(PPP)	0.891	0.436	1.256	2.109	1.553	4.170	0.729	0.363	1.136	0.486	-
LC-GIV(PPP)	0.684	0.880	0.973	1.778	2.932	2.670	0.674	0.827	0.737	0.521	-
TC-GIV(PPP)	**0.176**	**0.281**	**0.433**	0.577	0.861	1.404	0.166	0.257	0.402	0.406	-
VINS-Fusion(RTK)	0.157	0.185	0.138	0.521	0.697	0.889	0.139	0.181	0.137	0.334	82.2%
LC-GIV(RTK)	0.193	0.155	0.112	0.881	0.616	0.375	0.192	0.150	0.107	0.302	82.2%
TC-GIV(RTK)	**0.081**	**0.078**	**0.089**	0.473	0.411	0.460	0.080	0.074	0.089	0.284	86.4%

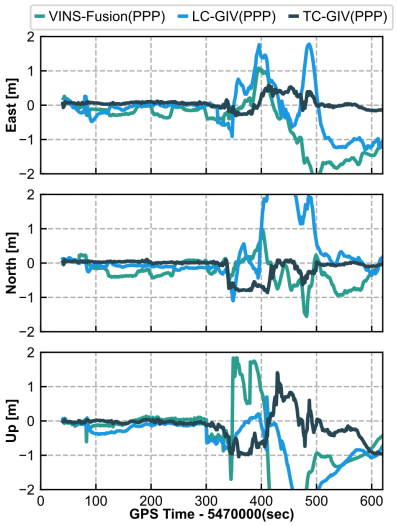

Fig. 4. Accuracy comparison of PPP-related VINS-Fusion, LC-GIV and TC-GIV configurations about the position in the GNSS half-open-sky experiment.

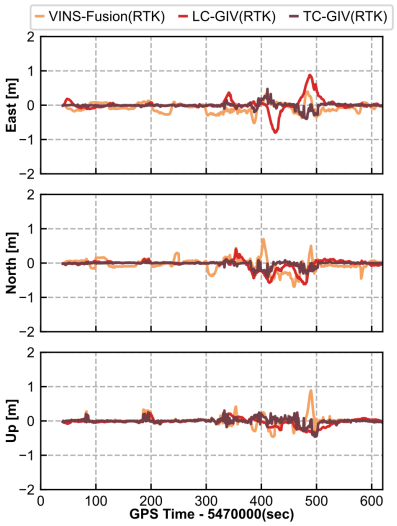

Fig. 5. Accuracy comparison of RTK-related VINS-Fusion, LC-GIV and TC-GIV configurations about the position in the GNSS half-open-sky experiment.

from 300 s to 500 s, which could pose a challenge for high-precision positioning.

Fig. 4 plots the time series of position errors for different PPP-related configurations and Table II reports the corresponding position RMSEs, maximum error and STD. It can be seen that the position errors of both VINS-Fusion(PPP) and LC-GIV(PPP) exceed 2 m when the GNSS visibility is poor. While the proposed TC-GIV(PPP) could maintain the positioning accuracy within 1 m during most of the periods. As for RTK-related results plotted in Fig. 5, the overall positioning accuracy is significantly improved with the aiding of the base station. Since both VINS-Fusion(RTK) and LC-GIV(RTK) directly utilize the RTK results from standalone GNSS module, they own the same ambiguity fixed rate of 82.2%. While the TC-GIV(RTK) improves the ambiguity fixed rate to 86.4% with INS-predicted pose estimations. Besides, the TC-GIV(RTK) configuration shows the highest positioning accuracy of 0.081, 0.078 and 0.089 m in the east, north, and vertical components, revealing improvements of (48.4%, 57.8%, 35.5%) and (58.0%, 49.7%, 20.5%) with respect to VINS-Fusion(RTK) and LC-GIV(RTK), respectively. These considerable improvements in the positioning accuracy mainly stem from the fact that the tightly coupled method could fully exploit the available information and provide better robustness against standalone-sensor outliers.

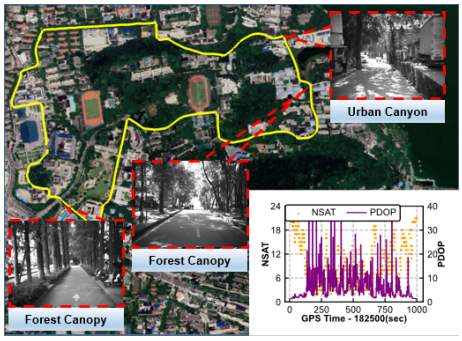

Fig. 6. The overview trajectory, typical scenarios and GNSS availability of the GNSS difficult experiment. The average NSAT and PDOP during the period from 250 s to 750 s are only 6.24 and 7.9, respectively.

C. GNSS Difficult Experiment

To further evaluate the effectiveness of the proposed integrated system, the second experiment in a GNSS-hostile environment was conducted. Compared to the GNSS half-open-sky condition in the first experiment, the road in this experiment is narrower with tall trees and buildings on both sides. As shown in Fig. 6, the number of available satellites decreases suddenly and is usually less than six, which may lead to frequent signal interruptions and severe multipath effects. Moreover, moving cars, pedestrians and

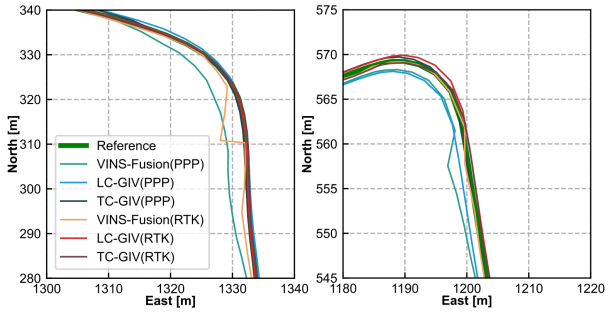

Fig. 7. Vehicle trajectories of different configurations in the urban canyon (left) and forest canopy (right).

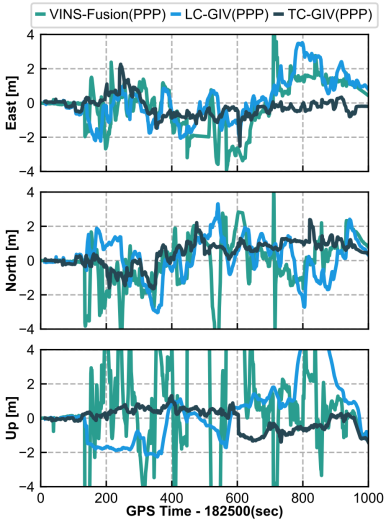

Fig. 8. Accuracy comparison of PPP-related VINS-Fusion, LC-GIV and TC-GIV configurations about the position in the GNSS difficult experiment.

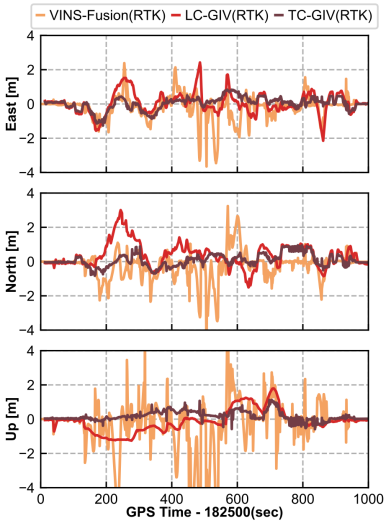

Fig. 9. Accuracy comparison of RTK-related VINS-Fusion, LC-GIV and TC-GIV configurations about the position in the GNSS difficult experiment.

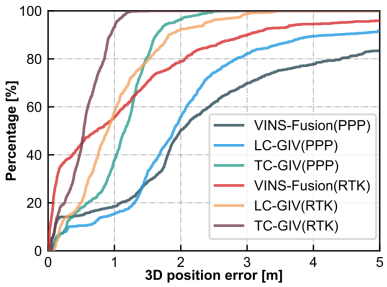

Fig. 10. The distribution of 3D position errors for all configurations in the GNSS difficult experiment.

dramatic lighting changes are also challenges for vision-based navigation.

Fig. 7 plots the vehicle trajectories of different configurations in the typical urban canyon and forest canopy. As can be seen, both VINS-Fusion(PPP) and VINS-Fusion(RTK) deviate from the reference due to severe signal interruptions. Comparatively, the proposed tightly integrated system using both PPP and RTK models can effectively suppress the position divergence and the corresponding trajectories are smoother and well-constrained within the reference.

Shown in Fig. 8, Fig. 9 and Table III are the positioning results of PPP-related and RTK-related configurations, respectively. Different from the similar performance of VINS-Fusion and LC-GIV in the GNSS half-open-sky experiment, VINS-Fusion performs much worse than LC-GIV in both PPP and RTK models with larger RMSE, maximum error and STD. The possible reason could be summarized as: (1) The VINS-Fusion is essentially based on the result-level integration, in which the VIO relative pose and GNSS position results are jointly fused in a global optimization thread. No information would be fed back to the VINS part to improve local pose estimation. (2) In LC-GIV, the periodic GNSS measurement updates are performed to correct the navigation states as well as IMU biases, which facilitates more accurate INS extrapolation results in the future. Similar to the GNSS half-open-sky experiment, the best positioning performance is achieved with the proposed TC-GIV(RTK) method. Such significant improvements could benefit from the optimal fusion of the tightly coupled method and better robustness to outlier measurements. Moreover, based on the PPP-related and RTK- related results shown in Fig. 8 and Fig. 9, we could infer that RTK is prioritized if there is an available base station nearby, which can facilitate more accurate and reliable location determination.

To further analyze the positioning performance of the proposed method, we also summarize the distribution of the position errors as depicted in Fig. 10. Intuitively, the TC-GIV(RTK) performs the best with 93.3% of position errors less than 1 m. Moreover, the percentage of the 3D position errors within 1 m is 18.5% and 55.8% for VINS-Fusion(PPP) and VINS-Fusion(RTK), which are higher than the corresponding LC-GIV(PPP) and LC-GIV(RTK) configurations. This could be attributed to the excellent smoothness of pose graph optimization.

TABLE III

POSITION AND ATTITUDE RESULTS OF DIFFERENT CONFIGURATIONS IN THE GNSS DIFFICULT EXPERIMENT.

Method	Position RMSE (m)			Maximum Error (m)			STD (m)			Yaw RMSE	Fixed
	East	North	Up	East	North	Up	East	North	Up	(deg)	Rate
VINS-Fusion(PPP)	1.373	1.453	3.797	6.343	6.538	8.491	1.373	1.435	3.781	0.772	-
LC-GIV(PPP)	1.285	1.171	1.862	3.514	3.312	4.914	1.272	1.158	1.829	0.593	-
TC-GIV(PPP)	**0.601**	**0.845**	**0.603**	2.517	2.391	1.407	0.557	0.730	0.601	0.459	-
VINS-Fusion(RTK)	0.744	0.864	1.629	3.648	5.814	6.845	0.713	0.823	1.600	0.615	46.2%
LC-GIV(RTK)	0.634	0.795	0.665	2.422	3.005	1.792	0.632	0.709	0.653	0.367	46.2%
TC-GIV(RTK)	**0.370**	**0.375**	**0.326**	1.264	0.953	1.121	0.368	0.366	0.272	0.307	52.0%

V. CONCLUSION

In this paper, we propose a tightly coupled GNSS/INS/Vision integration method to achieve accurate and continuous navigation in urban environments. Both widely used PPP and RTK models are unified in the proposed system to increase usability and flexibility. The large-scale real-world experiments indicate that our proposed tightly coupled method could achieve high-precision positioning accuracy in GNSS difficult environments and outperforms the corresponding loosely coupled method proposed in [20] as well as the state-of-the-art method in [17]. It is also worthwhile to notice that vision-based navigation solutions may fail in some degraded scenes, such as low-texture walls. In this case, incorporating observations from other sensors, such as lidar, may provide more credible results, which will be explored in our future work.

ACKNOWLEDGMENT

The algorithm implementation is based on the GREAT (GNSS+ REsearch, Application and Teaching) software [33] developed by the GREAT Group, School of Geodesy and Geomatics, Wuhan University.

REFERENCES

[1] M. Shalaby et al., "Relative Position Estimation in Multi-Agent Systems Using Attitude-Coupled Range Measurements," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 4955-4961, Jul. 2021.

[2] R. Jung and S. Weiss, "Modular Multi-Sensor Fusion: A Collaborative State Estimation Perspective," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 6891-6898, Oct. 2021.

[3] Y. Li et al., "Toward Location-Enabled IoT (LE-IoT): IoT Positioning Techniques, Error Sources, and Error Mitigation," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4035-4062, Mar. 2021.

[4] Z. Gao et al. "Evaluation on real-time dynamic performance of BDS in PPP, RTK, and INS tightly aided modes," *Advances in Space Research*, vol. 61, no. 9, pp. 2393-2405, May. 2018.

[5] X. Li, X. Li, J. Huang, Z. Shen, B. Wang, Y. Yuan and K. Zhang, "Improving PPP-RTK in urban environment by tightly coupled integration of GNSS and INS," *J. Geod.*, vol. 95, no. 132, Nov. 2021.

[6] M. Rabbou, A. El-Rabbany, "Tightly coupled integration of GPS precise point positioning and MEMS-based inertial systems," *GPS Solut.*, vol. 19, no. 4, pp. 601-609, 2014.

[7] P. Groves, "INS/GNSS Integration," in *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, 2nd ed., London, UK: Artech House, 2013, pp.363-406.

[8] C. Angelino, V. Baraniello and L. Cicala, "UAV position and attitude estimation using IMU, GNSS and camera," in *2012 15th International Conference on Information Fusion*, 2012, pp. 735-742.

[9] R. Sun et al., "Robust IMU/GPS/VO Integration for Vehicle Navigation in GNSS Degraded Urban Areas," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10110-10122, 2020.

[10] M. Nezhadshahbodaghi et al., "Fusing denoised stereo visual odometry, INS and GPS measurements for autonomous navigation in a tightly coupled approach," *GPS Solut.*, vol. 25, no. 47, 2021.

[11] J. Wang et al., "Evaluation on loosely and tightly coupled GNSS/INS vehicle navigation system," in *ICEAA*, 2017.

[12] T. Chu, N. Guo, S. Backén and D. Akos, "Monocular camera/IMU/GNSS integration for ground vehicle navigation in challenging GNSS environments," *Sensors*, vol. 12, no. 3, pp. 3162-3185, 2012.

[13] F. Zhu et al., "Fusing GNSS/INS/Vision With A Priori Feature Map for High-Precision and Continuous Navigation," *IEEE Sensors J.*, vol. 21, no. 20, pp. 23370-23381, Oct. 2021.

[14] T. Qin, P. Li and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004-1020, Aug. 2018.

[15] A. I. Mourikis, S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *ICRA*, 2007.

[16] C. Campos et al., "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874-1890, Dec. 2021.

[17] T. Qin et al., "A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors," 2019, arXiv:1901.03642.

[18] J. Liu, W. Gao and Z. Hu, "Optimization-Based Visual-Inertial SLAM Tightly Coupled with Raw GNSS Measurements," in *ICRA*, 2021.

[19] Z. Gong et al.,"Graph-Based Adaptive Fusion of GNSS and VIO Under Intermittent GNSS-Degraded Environment," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-16, 2021.

[20] J. Liao, X. Li, X. Wang, S. Li, and H. Wang, "Enhancing navigation performance through visual-inertial odometry in GNSS-degraded environment," *GPS Solut.*, vol. 25, no. 2, 2021.

[21] A. Vu, A. Ramanandan, A. Chen, J. A. Farrell and M. Barth, "Real-Time Computer Vision/DGPS-Aided Inertial Navigation System for Lane-Level Vehicle Navigation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 899-913, Jun. 2012.

[22] D. P. Shepard et al., "High-precision globally-referenced position and attitude via a fusion of visual SLAM, carrier-phase-based GPS, and inertial measurements," in *ION*, 2014.

[23] S. Cao et al., "GVINS: Tightly Coupled GNSS-Visual-Inertial Fusion for Smooth and Consistent State Estimation," *IEEE Transactions on Robotics*.

[24] X. Li et al., "Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo," *J. Geod.*, vol. 89, no. 6, pp. 607-635, Jun. 2015.

[25] T. Li et al., "High-Accuracy Positioning in Urban Environments Using Single-Frequency Multi- GNSS RTK/MEMS/IMU Integration", *Remote Sens.*, vol.10, no.2, pp.205, Jan. 2017.

[26] P. Furgale, J. Rehder and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc IEEE/RSJ Int. Conf. on Intell. Robots Syst.*, 2013, pp. 1280-1286.

[27] K. Sun, et al., "Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, 2018.

[28] D. Shuang, and Y. Gao, "Inertial Aided Cycle Slip Detection and Identification for Integrated PPP GPS and INS," *Sensors* vol. 12, no. 11, pp. 14344-14362, 2012.

[29] Z. Gao et al., "Odometer, low-cost inertial sensors, and four-GNSS data to enhance PPP and attitude determination," *GPS Solut.*, vol. 22, no. 3, 2018.

[30] P. J. G. Teunissen, "Quality control in integrated navigation systems," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 5, no. 7, pp. 35-41, Jul. 1990.

[31] P. J. G. Teunissen, "ADOP based upper bounds for the bootstrapped and the least-squares ambiguity success rates" *Artif. Satell.*, vol. 35, pp. 171-179, 2000.

[32] Novatel Inc., "Waypoint 8.90 User Manual", Available: https://novatel.com/support/waypoint-software/inertial-explorer.

[33] X. Li et al., "GREAT-UPD: An open-source software for uncalibrated phase delay estimation based on multi-GNSS and multi-frequency observations," *GPS Solut.*, vol. 25, no. 2, pp. 1-9, 2021.