

# 3D Vision Aided GNSS Real-Time Kinematic Positioning for Autonomous Systems in Urban Canyons

Weisong Wen | Xiwei Bai | Li-Ta Hsu

Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong

## Correspondence

Li-Ta Hsu  
Department of Aeronautical and Aviation Engineering  
The Hong Kong Polytechnic University  
Hong Kong  
Email: [lt.hsu@polyu.edu.hk](mailto:lt.hsu@polyu.edu.hk)

## Abstract

In this paper, a three-dimensional vision-aided method is proposed to improve global navigation satellite system (GNSS) real-time kinematic (RTK) positioning. To mitigate the impact of reflected non-line-of-sight (NLOS) reception, a sky-pointing camera with a deep neural network was employed to exclude these measurements. However, NLOS exclusion results in distorted satellite geometry. To fill this gap, complementarity between the low-lying visual landmarks and the healthy but high-elevation satellite measurements was explored to improve the geometric constraints. Specifically, inertial measurement units, visual landmarks captured by a forward-looking camera, and healthy GNSS measurements were tightly integrated via sliding window optimization to estimate the GNSS-RTK float solution. The integer ambiguities and the fixed GNSS-RTK solution were then resolved. The effectiveness of the proposed method was verified using several challenging data sets collected in urban canyons in Hong Kong.

## Keywords

3D vision, autonomous system, GNSS-RTK, NLOS, urban canyons

## 1 | INTRODUCTION

Accurate and globally-referenced positioning is one of the crucial components of autonomous navigation systems, including autonomous driving (Geiger et al., 2013), unmanned aerial vehicles (Cho et al., 2011; Zhang et al., 2011), and unmanned ground vehicles (T. Liu et al., 2021). During the past several decades, there has been an extensive exploration of these approaches using onboard local sensors, such as cameras, inertial measurement units (IMUs), or light detection and ranging (lidar). Lidar-based simultaneous localization and mapping (SLAM) (Cadena et al., 2016) has attracted much attention due to its superior robustness and accuracy, for example, the lidar odometry and mapping (LOAM) pipeline (Zhang & Singh, 2017) as well as variants of LOAM (Shan & Englot, 2018; Wang et al., 2021), and its integration with IMU (Li, et al., 2021; Qin et al., 2020; Shan et al., 2020; Ye et al., 2019) to enhance its robustness in challenging environments. However, the high cost of the three-dimensional (3D) lidar sensor is one of the major problems preventing its full deployment in autonomous systems. Moreover, lidar odometry-based positioning is subjected to inevitable drift in long-term operations, even with the assistance of IMU. In contrast to lidar-based positioning solutions, the visual/inertial system

(VINS) is a complementary and cost-effective method for robotics navigation applications. The VINS exhibits several advantages, including size, power assumption, weight, and availability. Many state-of-the-art VINS pipelines have been developed in the past several decades, including filtering-based methods such as Multi-state Constraint Kalman Filter (MSCKF), robust visual inertial odometry (ROVIO) (Bloesch et al., 2015), and Openvins (Geneva et al., 2020). Other research directions include optimization-based VINS pipelines such as the Open Keyframe-based Visual-Inertial SLAM (OKVIS) (Leutenegger et al., 2015), VINS-Mono (Qin et al., 2018), and ORB-SLAM3 (Campos et al., 2021). However, the VINS shares a similar drawback with lidar, namely, that the estimation state is subject to drift over time. Moreover, results from previous studies (Bai et al., 2020a; Bai et al., 2020c) revealed that the urban canyon scenarios introduce additional challenges to the VINS due to excessive movement of objects in the surroundings, unstable illumination, and even stronger motion blur, all of which would enhance drift. To mitigate this problem, some studies (Gong et al., 2020; X. Li et al., 2021) proposed a method to integrate the globally-referenced global navigation satellite system (GNSS) single point positioning (SPP) with VINS in a loosely (Gong et al., 2020; Li et al., 2021; Qin et al., 2019) or even a tightly coupled manner (Cao et al., 2022). In each of these cases, the absolute positioning accuracy will depend highly on the GNSS solution. Unfortunately, the accuracy of GNSS readings based on data from a single receiver is limited to several meters because of atmospheric and receiver clock errors (Enge, 1994). According to one evaluation of the *urbanLoco* data set (Wen et al., 2020), and an overall absolute mean error of 4.5 meters is obtained with tightly coupled integration of visual/inertial/GNSS in urban scenarios. This cannot satisfy the navigation requirements of a typical autonomous system.

### 1.1 | Problems Using GNSS-Real-Time Kinematic Positioning (RTK) in Urban Canyons

Fortunately, GNSS real-time kinematic (GNSS-RTK) positioning (Counselman & Gourevitch, 1981) introduces a significantly higher degree of absolute positioning accuracy with the assistance of carrier-phase measurements and corrections from the reference station. Centimeter-level accuracy can be achieved in open areas using GNSS-RTK positioning. Therefore, the combination of the VINS and GNSS-RTK may be a promising solution to achieve globally-referenced and locally-accurate positioning. GNSS-RTK positioning typically involves two steps: (1) the float solution is estimated based on the double-differenced (DD) GNSS carrier-phase and pseudorange measurements (Enge, 1994); (2) the integer ambiguity is subsequently resolved via integer least-squares algorithms (e.g., LAMBDA) (Teunissen, 2000) based on the derived float solution. Centimeter-level positioning accuracy can be achieved in an open area when the fixed solution is achieved. Unfortunately, the accuracy of GNSS-RTK positioning is significantly degraded in urban canyons due to the non-line-of-sight (NLOS) reception caused by GNSS signal reflection, blockage from surrounding buildings, and even dynamic objects (Wen, et al. 2019a), for example, double-decker buses. Theoretically, the significantly degraded GNSS-RTK positioning accuracy observed in urban canyons is induced by two major problems.

- (1) Problem 1:** According to our previous findings (Wen et al., 2019b), most of the GNSS signals received in highly urbanized areas are NLOS receptions. These typically emerge under conditions in which direct light-of-sight (LOS) signal transmission is blocked and reflected signals from the same satellite

are received. In this case, the accuracy of the float solution estimation based on DD carrier-phase and pseudorange measurements is severely degraded.

**(2) Problem 2:** The number of available satellites may be limited in urban canyons due to signal blockage caused by surrounding buildings. In these situations, the geometry of the satellite distribution is distorted, which would cause a large ambiguity dilution of precision (ADOP) (Teunissen, 1997). The ADOP can be calculated based on the covariance of the integer ambiguity. Mathematically, a smaller ADOP could lead to an increased rate of ambiguity resolution. A similar phenomenon describing poor LOS satellite geometry was described by Marais et al. (2014; 2015). As a result, it may be difficult to obtain a fixed solution due to the poor satellite geometry with larger ambiguity in the searching space.

In summary, the poor quality of the GNSS measurements and the satellite geometry distributions are the major problems limiting the performance of GNSS-RTK positioning systems.

## 1.2 | Related Work on NLOS Mitigation for GNSSs

Numerous scientific endeavors have attempted to eliminate these two bottlenecks. One method was proposed by Li et al. (2017) to integrate a high-cost IMU with GNSS-RTK positioning to enhance the robustness against the impact of outlier measurements (e.g., NLOS reception). Similarly, potential GNSS NLOS measurements were also partially excluded based on a GNSS measurement residual-based consistency check (Hsu et al., 2017). However, these enhancements are limited by the cost of the employed IMU sensor and the percentage contributions of the GNSS outlier measurements. Similar work by T. Li et al. (2019) went one step further by integrating the GNSS-RTK/visual/inertial based on an Extended Kalman Filter (EKF) estimator together with a similar outlier rejection scheme based on the residual. Sub-meter positioning accuracy was achieved using the evaluated data sets. However, the GNSS NLOS exclusion (Li et al., 2019; Li et al., 2017) can only partially exclude these signals and can even aggravate the problem of degraded satellite geometry distribution. Considering that the GNSS measurements are highly environmentally-dependent and time-correlated, we recently proposed exchanging a filtering-based estimator with a factor graph optimization (FGO)-based GNSS-RTK positioning method (Wen & Hsu, 2021). This method permitted us to explore the time correlation between multiple epoch measurements while simultaneously enhancing the robustness against the GNSS outlier measurements. However, the performance of this method was dependent on the percentage of the GNSS outlier measurements involved in the factor graph. In a recent study, Fan et al. (2019), employed multiple antennas to enhance the robustness of GNSS-RTK against interference from outlier measurements. However, this method relied on expensive geodetic antennas. In another recent study, Furukawa et al. (2020) improved GNSS-RTK positioning in urban canyons by excluding polluted GNSS signals with the assistance of 3D building models and a pre-defined initial guess of the position of the GNSS receiver. An increased fixed rate was obtained after selecting healthy line-of-sight (LOS) measurements. Recently, Niu et al. (2021) explored the loosely-coupled integration of visual-inertial odometry and GNSS-RTK positioning based on signals from smartphones. However, the performance of this integrated system remains limited in urban areas because of the GNSS NLOS receptions. Similarly, Ng and Hsu (2021) proposed the use of multiple hypothesis-based

3D building models to aid GNSS-RTK positioning and improve the accuracy of the geodetic level GNSS receivers. However, the NLOS satellite exclusion relies on the availability of accurate 3D building models and a good initial guess of the GNSS receiver position. In particular, an initial guess of the receiver position is required in each epoch. This initial guess is typically based on GNSS single point positioning using pseudorange measurements which are not accurate in urban canyons.

Additional research was performed that focused on reconstructing (Wen, 2020) or sensing (Suzuki & Kubo, 2013; Wen et al., 2019b) the surrounding environmental features in real-time to improve GNSS positioning. In our previous work (Wen et al., 2018; Wen et al., 2019a; Wen et al., 2019b), we proposed methods to promote continuous improvements in the GNSS positioning in urban canyons using onboard 3D lidar sensors and excluding (Wen et al., 2019a) or correcting (Wen, 2020; Wen et al., 2019b) potential NLOS measurements. However, we found that the performance of NLOS detection based on these methods was limited by the field of view (FOV) of the applied 3D lidar sensor. Moreover, 3D lidar was still too expensive for low-cost autonomous system applications. Instead of relying on the expensive 3D lidar sensor, we proposed a method to detect the GNSS NLOS in dense urban canyons using a skyward-pointing fish-eye camera that resulted in improved GNSS SPP (Bai et al., 2020b). These methods contributed to the alleviation of the adverse impacts of GNSS NLOS receptions and would thus improve GNSS-RTK positioning in urban canyons. However, this can lead to a secondary problem, i.e., the geometric distributions can be significantly degraded due to the exclusion of NLOS receptions from satellites at low elevation angles. In other words, only the satellites with high elevation angles can survive the NLOS exclusion. In contrast, to GNSS, the visual features from a forward-looking camera are mainly detected from the low-lying environmental structures with low elevation angles with respect to vehicles, for example, trees, roadside pillars, and poles. Inspired by the observed complementarity of the low-lying visual landmarks and the healthy but high-elevation satellite measurements, we proposed to employ 3D vision to assist with GNSS-RTK positioning in urban canyons.

### 1.3 | Related Work on GNSS-Visual Initialization

Given the fact that the raw measurements from the visual and the GNSS are in different frames, efficient and reliable initialization of the GNSS-visual will be of great significance. The raw visual measurement is located in the body frame of the camera. By contrast, the raw GNSS measurement is in the global frame, i.e., the Earth-centered Earth fixed (ECEF) frame. Therefore, the objective of the initialization was to estimate the extrinsic parameters linking the camera body and the ECEF frames at the time that the system starts. Qin et al. (2019) proposed a loosely-coupled integration of the GNSS and the VINS using factor graph optimization. Of particular note, the extrinsic parameters connecting the VINS local frame and the GNSS global frame are estimated iteratively based on the observation from the GNSS and VINS. However, the initialization can only be achieved in a batch optimization mode which does not apply to a real-time application. A similar initialization strategy was described by Niu et al. (2021). As an extension, Liu et al. (2021) proposed to conduct experiments on the GNSS-visual initialization using the raw GNSS pseudorange and Doppler measurements. However, initialization in this case relied heavily on the accuracy of the pseudorange measurements. In addition, the high-accuracy carrier-phase measurements were not exploited. The carrier-phase measurements differ from the GNSS pseudorange measurements and can provide high-accuracy motion constraints

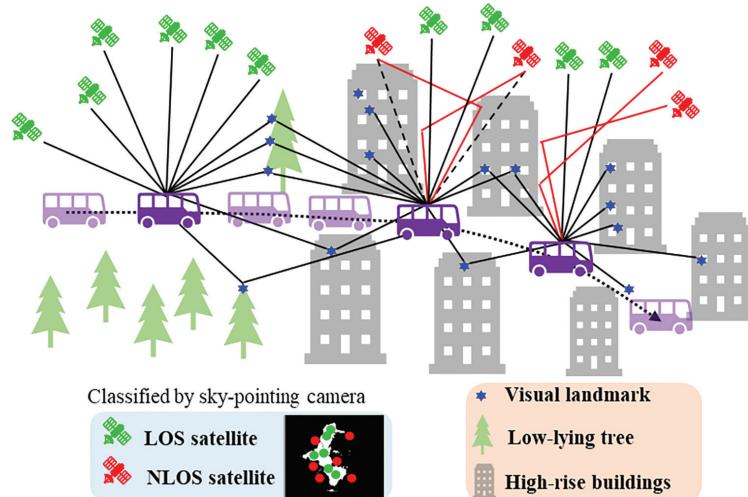
between multiple epochs of state once their unknown integer ambiguities have been removed. Thus, this paper proposed a carrier-phase ambiguity-free algorithm that can reliably initialize the extrinsic parameters connecting the VINS local frame and the GNSS global frame, by exploiting the GNSS pseudorange, Doppler, and carrier-phase measurements.

#### 1.4 | Contributions of This Paper

A demonstration of the complementarity is shown in Figure 1. To the best of our knowledge, this is the first effort directed at solving the major problem of GNSS-RTK positioning for autonomous systems via a deep exploration of the complementarity between GNSS satellite measurements and visual landmarks in urban canyons.

The key contributions of this paper are as follows:

- (1) We plan to address the problem of poor geometry resulting from the exclusion of GNSS NLOS receptions by employing low-lying visual landmarks to improve the geometric constraints of GNSS-RTK. To achieve this, a tightly coupled fusion estimator was proposed that would integrate measurements from IMU pre-integration, low-lying visual landmarks reprojection, healthy but high-elevation GNSS DD pseudorange, and carrier-phase measurements together with the Doppler frequency measurements via a factor graph-based sliding window optimization method. This will facilitate the exploration of the time-correlation of GNSS measurements between multiple epochs. With assistance from low-lying visual constraints, the ADOP was significantly decreased (i.e., better geometric constraints for GNSS-RTK). The estimated state of the system was then employed as the float solution for the ambiguity resolution of GNSS-RTK, which could be used to achieve a fixed solution. It is notable that the integration of the visual measurements also contributes to a more accurate float position estimate and is also significant for achieving a fixed solution.
- (2) The proposed estimator involves multiple data sources with different observation models that could induce additional non-linearity. As a result, the performance of the estimator relies heavily on the initial guess of the



**FIGURE 1** Demonstration of the strong complementarity between the low-lying visual landmarks and healthy but high-elevation satellite measurements

system state. Thus, a carrier-phase ambiguity-free algorithm was proposed that would permit careful initialization of the extrinsic parameters between the VINS local frame and the GNSS global frame as well as the integer ambiguity of the DD carrier-phase measurements.

- (3) We will verify the proposed method through several challenging data sets collected in urban canyons of Hong Kong using low-cost automobile-level GNSS receivers together with an automobile visual/inertial sensor suite.

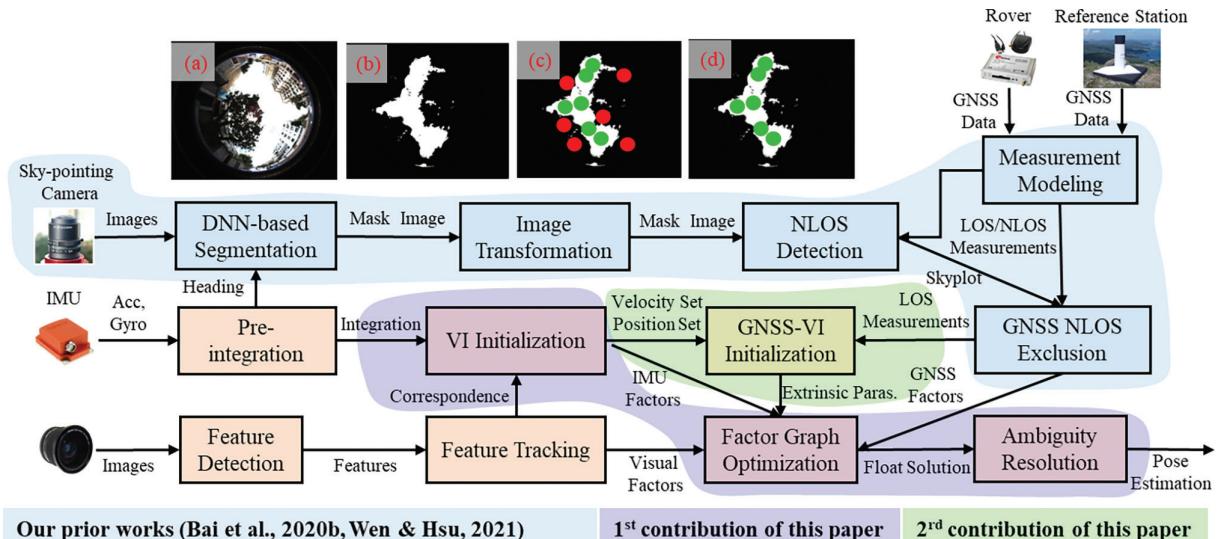
The remaining sections of this paper are organized as follows: Section 2 includes an overview of the proposed method. GNSS measurement modeling is discussed in Section 3. Tightly-coupled integration of GNSS-RTK/visual/inertial systems is presented in Section 4 followed by an interpretation of the system initialization in Section 5. Several real-life experiments conducted to evaluate the effectiveness of the proposed method are presented in Section 6. Finally, the conclusions and future work will be discussed in Section 7.

## 2 | SYSTEM OVERVIEW AND NOTATIONS

### 2.1 | System Overview

An overview of the proposed method is shown in Figure 2. The system consists of two parts:

- (1) **GNSS NLOS outlier removal.** The real-time environmental understanding aided by a sky-pointing fish-eye camera and its application in GNSS NLOS exclusion is shown in the light blue shaded boxes in Figure 2. The outcomes presented in this section include satellite visibility, line-of-sight (LOS), or NLOS receptions. This was based on our previous work and improved by using the U-net (Ronneberger et al., 2015) to separate the sky from the non-sky area. Additionally, the high-cost 3D lidar sensor to aid in correcting the detected NLOS delay was not employed in this study. Instead, the NLOS measurements



**FIGURE 2** Overview of the proposed method. DNN and VI denote deep neural networks and visual/inertial, respectively. Acc and Gyro denote the accelerometer and gyroscope measurements, respectively.

would be directly excluded from signals received from the satellites based on the segmented image. An illustration of the sky-view image is shown in Figure 2(a) with a sky view that includes the building roofs, trees, and in some cases even the roofs of surrounding moving vehicles. The segmented sky view image prepared using U-net is shown in Figure 2(b), with white and black pixels denoting the sky and non-sky areas, respectively. Subsequently, the satellites can be projected on the segmented image, based on their elevation, azimuth angles, and the heading (provided by the high-frequency pose estimation from IMU, which will be discussed in Section 4) from north of the earth of the sky-pointing camera. GNSS NLOS measurements can then be detected (Figure 2[c]); only healthy LOS measurements are utilized (Figure 2[d]). Additional details regarding GNSS NLOS detection via the sky-pointing camera can be found in our earlier work (Bai et al., 2020b).

- (2) **Improving the geometry with visual landmark measurements.** The tightly coupled fusion estimator generated by integrating the measurements from IMU pre-integration, low-lying visual landmarks reprojection from the forward-looking camera, and healthy (high elevation angle) GNSS LOS measurements from a factor graph-based sliding window optimization are shown in the lavender-shaded boxes in Figure 2. The outcomes presented in this section include the improved pose estimation of the system. A concept illustration of the proposed geometric improvement is shown in Figure 1. When the vehicle starts moving from the left side of Figure 1 which denotes an open-sky and GNSS-friendly scenario, all the GNSS measurements are LOS and a fixed solution can be obtained by GNSS-RTK with centimeter-level accuracy. The surroundings are mainly composed of low-lying trees which can provide low-lying visual features that will further smooth the state estimation. When the vehicle has traveled to the right side of Figure 1 which includes a GNSS-challenging urban canyon scenario, the GNSS NLOS reception would be detected (the red satellite icon in Figure 1). Fortunately, the low-lying blue visual landmarks and the remaining few LOS satellites (the green satellite icon in Figure 1) are highly complementary, thus providing the main inspiration for this study.

## 2.2 | Frame Definitions

Because these multiple data sources arise from different spatial frames, it will be important to define the associated coordinates. The following frames are defined as described by Liu et al. (2021) and will provide a reference for the remaining sections of this paper.

- (1) **Earth-centred, earth-fixed (ECEF) frames** (Enge, 1994). This refers to a global frame typically used to define the position of the satellites and associated measurements.  $(\cdot)^{WE}$  has been adopted to denote the variable represented in the ECEF frame.
- (2) **East, north, and up (ENU) frames** (Enge, 1994): This refers to another commonly used global frame of GNSS with the  $x$ ,  $y$  and  $z$  axis of ENU frame pointing to the east, north, and up directions, respectively.  $(\cdot)^{WG}$  has been adopted to denote the variable represented in the ENU frame.

- (3) **Local world frames** (Liu et al., 2021): This refers to a local world frame of pose estimation from local sensors (such as visual and inertial). This is the reference frame of the VINS in this study.  $(\cdot)^{WL}$  has been adopted to denote the variable represented in the local world frame.
- (4) **Sensor frames**: The sensor frame is fixed on the sensor body. In this paper, sensor frames include the IMU frame, the camera frame, and the GNSS receiver frame, which are denoted as  $(\cdot)^b$ ,  $(\cdot)^c$  and  $(\cdot)^r$ , respectively. Similar to the work describing VINS-Mono (Qin et al., 2018), the IMU frame was selected as the target frame for state estimation in this study.

### 2.3 | Notations

In this paper, matrices are denoted in uppercase with bold letters. Vectors are denoted in lowercase with bold letters. Variable scalars are denoted with italicized letters. Constant scalars are denoted as lowercase letters. To ensure that the proposed pipeline has been presented clearly, the following major notations are defined here and used consistently in the remaining sections of this paper.

- a) The pseudorange measurement received from a satellite  $s$  at a given GNSS epoch  $r_t$  is expressed as  $\rho_{r_t}^s$ . The subscripts  $r$  and  $t$  denote the GNSS receiver and the time index, respectively. The superscript  $s$  denotes the index of the satellite at the given epoch  $r_t$ .
- b) The Doppler measurement received from satellite  $s$  at a given epoch  $r_t$  is expressed as  $d_{r_t}^s$ .
- c) The carrier-phase measurement received from a satellite  $s$  at a given epoch  $r_t$  is expressed as  $\psi_{r_t}^s$ .
- d) Let  $\mathbf{R}_A^B$  denote the rotation matrix that rotates a vector in the frame {A} to frame {B}, and the  $\mathbf{q}_A^B$  is its quaternion form. A similar definition is applied to the definition of translation, e.g.,  $\mathbf{p}_A^B$  denotes the translation from the frame {A} to frame {B}. The transformation between both frames is expressed as  $\mathbf{T}_A^B = [\mathbf{R}_A^B, \mathbf{p}_A^B]$ .
- e) The position of the satellite  $s$  at a given epoch  $r_t$  is expressed as  $\mathbf{p}_{r_t}^{WE,s} = (p_{r_t,x}^{WE,s}, p_{r_t,y}^{WE,s}, p_{r_t,z}^{WE,s})^T$ . The “T” denotes the transpose operator.
- f) The velocity of the satellite  $s$  at a given epoch  $r_t$  is expressed as  $\mathbf{v}_{r_t}^{WE,s} = (v_{r_t,x}^{WE,s}, v_{r_t,y}^{WE,s}, v_{r_t,z}^{WE,s})^T$ .
- g) The position of the GNSS receiver at a given epoch  $r_t$  is expressed as  $\mathbf{p}_{r_t}^{WE} = (p_{r_t,x}^{WE}, p_{r_t,y}^{WE}, p_{r_t,z}^{WE})^T$ .
- h) The velocity of the GNSS receiver at a given epoch  $r_t$  is expressed as  $\mathbf{v}_{r_t}^{WE} = (v_{r_t,x}^{WE}, v_{r_t,y}^{WE}, v_{r_t,z}^{WE})^T$ .
- i) The clock bias of the GNSS receiver at a given epoch  $r_t$  is expressed as  $\delta_{r_t}^{r,j}$ , with the unit in meters. The variable  $j$  denotes the associated satellite systems, such as GPS and BeiDou satellite systems. The receiver clock bias drift is expressed as  $\dot{\delta}_{r_t}^r$ . Note that all the satellite systems share the same clock bias drift.
- j) The clock bias of satellite  $s$  at epoch  $r_t$  is expressed as  $\delta_{r_t}^s$ . The satellite clock bias is expressed as  $\dot{\delta}_{r_t}^s$ .
- k) The position of the base(reference)station is expressed as  $\mathbf{p}_e = (p_{e,x}, p_{e,y}, p_{e,z})^T$ . The variables  $\rho_{e,r_t}^s$  and  $\psi_{e,r_t}^s$  denote the pseudorange and range measurements of carrier-phase from satellite  $s$  received by the reference station at epoch  $r_t$ .

The subscript “e” denotes the reference station.

- l) The extrinsic parameter between the camera and IMU sensor frame is expressed as  $\mathbf{T}_c^b = [\mathbf{R}_c^b, \mathbf{p}_c^b]$ .

The extrinsic parameter between the GNSS receiver and IMU sensor frame is expressed as  $\mathbf{p}_r^b$ .

## 2.4 | GNSS Measurement Modeling

In this Section, the modeling of the GNSS measurements will be described, including the DD pseudorange and carrier-phase measurements together with the Doppler measurements.

### 2.4.1 | DD pseudorange/carrier-phase measurements modeling

In terms of the measurements from the GNSS receiver, each pseudorange measurement,  $\rho_{r_t}^s$ , can be expressed as shown in Equation (1) (Takasu & Yasuda, 2009):

$$\rho_{r_t}^s = r_{r_t}^s + c(\delta_{r_t}^{r,j} - \delta_{r_t}^s) + I_{r_t}^s + T_{r_t}^s + \varepsilon_{r_t}^s \quad (1)$$

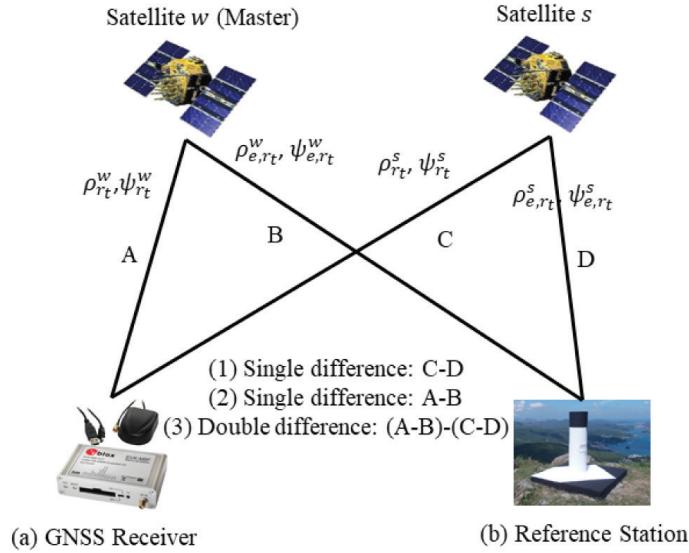
where  $r_{r_t}^s$  represents the geometric range between the satellite and the GNSS receiver.  $I_{r_t}^s$  denotes the ionospheric delay distance;  $T_{r_t}^s$  represents the tropospheric delay distance, and  $\varepsilon_{r_t}^s$  represents the errors caused by the multipath effects, NLOS receptions, receiver noise, and antenna phase-related noise. Due to the potential clock-related and atmosphere-related errors involved in the generation of raw pseudorange measurements, the accuracy of the conventional pseudorange-based SPP (Rycroft, 1997) is limited to several meters. Similarly, the carrier-phase measurement,  $\psi_{r_t}^s$ , is expressed as shown in Equation (2) (Takasu & Yasuda, 2009):

$$\lambda_j \psi_{r_t}^s = r_{r_t}^s + c_L(\delta_{r_t}^{r,j} - \delta_{r_t}^s) - I_{r_t}^s + T_{r_t}^s + \lambda B_{r_t}^s + d\psi_{r_t}^s + \epsilon_{r_t}^s \quad (2)$$

$$\text{with } B_{r_t}^s = \psi_{r_t,0} - \psi_{0,r_t}^s + N_{r_t}^s$$

where  $B_{r_t}^s$  refers to the carrier-phase bias, the variable  $\lambda_j$  represents the carrier wavelength of the satellite system  $j$ , and the variable  $d\psi_{r_t}^s$  represents the carrier-phase correction terms, including antenna phase offsets and variations, station displacement by earth tides, phase windup effect, and the relativity correction on the satellite clock. A detailed formulation of the carrier-phase corrections can be found in Takasu and Yasuda (2009). The variable  $\psi_{r_t,0}$  represents the initial phase of the receiver local oscillator. Similarly,  $\psi_{0,r_t}^s$  is the initial phase of the transmitted navigation signal from the satellite. The variable  $N_{r_t}^s$  represents the carrier-phase integer ambiguity which is expected to be an integer value.  $\epsilon_{r_t}^s$  represents the errors caused by the multipath effects, NLOS receptions, receiver noise, and antenna delay.

To improve the GNSS positioning accuracy, the DD method was adopted in GNSS-RTK positioning to remove systematic errors including atmospheric and clock-related errors. The illustration in Figure 3 presents the DD operation. Given the pseudorange and carrier-phase measurements received from a satellite  $s$  and a master satellite  $w$ , the DD pseudorange measurement ( $\rho_{DD,t}^s$ ) can be expressed as follows in Equation (3):



**FIGURE 3** Demonstration of the double-difference (DD) technique (a) GNSS receiver installed in the autonomous system, and (b) Reference station installed in open sky area with geodetic level GNSS receiver.

$$\rho_{DD,r_t}^s = (\rho_{r_t}^s - \rho_{e,r_t}^s) - (\rho_{r_t}^w - \rho_{e,r_t}^w) \quad (3)$$

The variables  $\rho_{e,r_t}^w$  and  $\rho_{e,r_t}^s$  represent the pseudorange measurements received by the reference station denoted by the subscript “e”.

Generally, the satellite with the highest elevation angle tends to encounter the fewest multipath and NLOS errors. Therefore, the satellite  $w$  with the highest elevation angle is selected as the master satellite. By stacking Equations (1) through (3), the following formula can be obtained as shown in Equation (4):

$$\rho_{DD,r_t}^s = (r_{r_t}^s - r_{e,r_t}^s) - (r_{r_t}^w - r_{e,r_t}^w) + \varepsilon_{DD,r_t}^s \quad (4)$$

where  $\varepsilon_{DD,r_t}^s$  denotes the noise related to the DD pseudorange measurements. After applying the DD process to the pseudorange measurements, the derived  $\rho_{DD,r_t}^s$  will be free of the clock bias and atmospheric effects (Takasu & Yasuda, 2009), based on the premise that the GNSS measurements received from the user-end receiver and the station receiver share the same atmospheric error.

Similarly, the DD carrier-phase measurement ( $\psi_{DD,r_t}^s$ ) of satellite  $s$  is expressed as indicated in Equation (5) (Takasu & Yasuda, 2009):

$$\psi_{DD,r_t}^s = (\psi_{r_t}^s - \psi_{e,r_t}^s) - (\psi_{r_t}^w - \psi_{e,r_t}^w) \quad (5)$$

The variables  $\psi_{e,r_t}^s$  and  $\psi_{e,r_t}^w$  represent the carrier-phase measurements received by the reference station. Similarly, the clock bias and atmospheric effects are waived from  $\psi_{DD,r_t}^s$ . By stacking Equations (2) through (5), the following formula can be obtained as shown in Equation (6):

$$\lambda_j \psi_{DD,r_t}^s = (r_{r_t}^s - r_{e,r_t}^s) - (r_{r_t}^w - r_{e,r_t}^w) + \lambda_j N_{DD,r_t}^s + \varepsilon_{DD,r_t}^s \quad (6)$$

where  $\varepsilon_{DD,r_t}^s$  is the noise related to the DD carrier-phase measurements and  $N_{DD,r_t}^s$  is the DD integer ambiguity to be estimated.

### 2.4.2 | Doppler measurements modeling

In terms of the measurements from the GNSS receiver, each Doppler measurement,  $d_{r_t}^s$ , is expressed as indicated in Equation (7) (Takasu & Yasuda, 2009):

$$d_{r_t}^s = \frac{1}{\lambda_j} \left( \mathbf{e}_{r_t}^s (\mathbf{v}_{r_t}^{WE,s} - \mathbf{v}_{r_t}^{WE}) + c(\dot{\delta}_{r_t}^{r,j} - \dot{\delta}_{r_t}^s) \right) + \zeta_{r_t}^s \quad (7)$$

where  $\zeta_{r_t}^s$  refers to the noise related to the received Doppler measurement and  $\mathbf{e}_{r_t}^s$  represents the LOS vector connecting the GNSS receiver and the satellite  $s$  which can be calculated as indicated in Equation (8):

$$\mathbf{e}_{r_t}^s = \left( \frac{\mathbf{p}_{r_t}^{WE,s} - \mathbf{p}_{r_t}^{WE}}{\|\mathbf{p}_{r_t}^{WE,s} - \mathbf{p}_{r_t}^{WE}\|} \right)^T \quad (8)$$

where the operator  $\|*\|$  is employed to obtain the norm of the associated vector. Because it encodes the motion difference between the satellite and the GNSS receiver, the Doppler measurement is employed to constrain the relative motion of the GNSS receiver between two consecutive epochs. The Doppler measurements are incorporated into the factor graph in Section 4.

## 2.5 | GNSS-RTK Positioning By Weighted Least Square Estimation

### 2.5.1 | Float solution estimation

Given  $S+1$  satellite measurements at a given epoch  $r_t$ , the DD pseudorange and carrier-phase measurements can be structured as  $\rho_{DD,r_t} = [\rho_{DD,r_t}^0, \rho_{DD,r_t}^1, \dots, \rho_{DD,r_t}^{S-1}]^T$  and  $\psi_{DD,r_t} = [\psi_{DD,r_t}^0, \psi_{DD,r_t}^1, \dots, \psi_{DD,r_t}^{S-1}]^T$ , respectively. The DD integer ambiguities related to the  $S$  DD measurements can be structured as  $\mathbf{N}_{DD,r_t} = [N_{DD,r_t}^0, N_{DD,r_t}^1, \dots, N_{DD,r_t}^{S-1}]^T$ . The linearized DD observations can be expressed as indicated in Equation (9):

$$\begin{bmatrix} \Psi_{DD,r_t} \\ \rho_{DD,r_t} \end{bmatrix} = \begin{bmatrix} \lambda_j^{-1} \mathbf{G}_{r_t} & \mathbf{I}_{S \times S} \\ \mathbf{G}_{r_t} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_{r_t}^{WE} \\ \mathbf{N}_{DD,r_t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{DD,r_t} \\ \boldsymbol{\epsilon}_{DD,r_t} \end{bmatrix} \quad (9)$$

The vectors  $\boldsymbol{\varepsilon}_{DD,r_t}$  and  $\boldsymbol{\epsilon}_{DD,r_t}$  represent the stacked noise related to the  $S$  DD carrier-phase and pseudorange measurements, respectively, while  $\mathbf{p}_{r_t}^{WE}$  and the  $\mathbf{N}_{DD,r_t}$  represent the states to be estimated.  $\mathbf{G}_{r_t}$  represents the geometric matrix related to the GNSS receiver position and the received  $S$  satellites which be expressed as indicated in Equation (10):

$$\mathbf{G}_t^d = \begin{bmatrix} \frac{\mathbf{p}_{r_t}^{WE,0} - \mathbf{p}_{r_t}^{WE}}{\|\mathbf{p}_{r_t}^{WE,0} - \mathbf{p}_{r_t}^{WE}\|} \\ \dots \\ \frac{\mathbf{p}_{r_t}^{WE,S-1} - \mathbf{p}_{r_t}^{WE}}{\|\mathbf{p}_{r_t}^{WE,S-1} - \mathbf{p}_{r_t}^{WE}\|} \end{bmatrix} \quad (10)$$

where the size of the geometric matrix  $\mathbf{G}_t^d$  is  $S \times 3$ . Therefore, the float solution of  $[\mathbf{p}_{r_t}^{WE} \mathbf{N}_{DD,r_t}]^T$  can be calculated iteratively via weighted least-square estimations as indicated in Equation (11):

$$\begin{bmatrix} \mathbf{p}_{r_t}^{WE} \\ \mathbf{N}_{DD,r_t} \end{bmatrix} = (\mathcal{M}_{r_t}^T \mathbf{W}_{r_t} \mathcal{M}_{r_t})^{-1} \mathcal{M}_{r_t}^T \mathbf{W}_{r_t} \begin{bmatrix} \Psi_{DD,r_t} \\ \rho_{DD,r_t} \end{bmatrix}$$

With  $\mathcal{M}_{r_t} = \begin{bmatrix} \lambda_j^{-1} \mathbf{G}_{r_t} & \mathbf{I}_{S \times S} \\ \mathbf{G}_{r_t} & 0 \end{bmatrix}$

(11)

where  $\mathbf{W}_{r_t}$  represents the weighting matrix related to  $\epsilon_{DD,r_t}$  and  $\epsilon_{DD,r_t}$  that can be calculated based on the satellite elevation angle and the signal-to-noise ratio (SNR) (Herrera, et al. 2015). After solving Equation (11), it is not guaranteed that the estimated ambiguity in  $\mathbf{N}_{DD,r_t}$  will be an integer value due to the noise associated with the DD measurements. As a result, the accuracy of the estimated  $\mathbf{p}_{r_t}^{WE}$  is usually within 1 meter in a sparsely-populated area. Because the DD ambiguity in  $\mathbf{N}_{DD,r_t}$  should be an integer value, it should be resolved to obtain a fixed solution for GNSS-RTK, thereby achieving centimeter-level positioning accuracy that will be explained in the sub-section to follow.

### 2.5.2 | Integer ambiguity resolution

Assuming the covariance matrix  $\mathbf{Q}_{r_t}$  associated with the estimated  $\mathbf{N}_{DD,r_t}$  and  $\mathbf{p}_{r_t}^{WE}$  is denoted as indicated in Equation (12):

$$\mathbf{Q}_{r_t} = \begin{bmatrix} \mathbf{Q}_{pp} & \mathbf{Q}_{pn} \\ \mathbf{Q}_{np} & \mathbf{Q}_{nn} \end{bmatrix} \quad (12)$$

where  $\mathbf{Q}_{pn} = \mathbf{Q}_{np}^T$  are the float covariance matrices, and the subscripts  $p$  and  $n$  represent the position and DD ambiguity, respectively. Meanwhile,  $\mathbf{Q}_{r_t}$  can be estimated as indicated in Equation (13):

$$\mathbf{Q}_{r_t} = \begin{bmatrix} \lambda_j^{-2} \mathbf{G}_{r_t}^T \mathbf{W}_{r_t}^\psi \mathbf{G}_{r_t} + \mathbf{G}_{r_t}^T \mathbf{W}_{r_t}^\rho \mathbf{G}_{r_t} & \lambda_j^{-1} \mathbf{G}_{r_t}^T \mathbf{W}_{r_t}^\psi \\ \lambda_j^{-1} \mathbf{W}_{r_t}^\psi \mathbf{G}_{r_t} & \mathbf{W}_{r_t}^\psi \end{bmatrix}^{-1} \quad (13)$$

$\mathbf{W}_{r_t}^\rho$  and  $\mathbf{W}_{r_t}^\psi$  denote the covariance matrices associated with the DD carrier-phase and pseudorange measurements. Therefore, the integer ambiguity resolution problem may be solved by finding a set of integer DD ambiguities denoted by  $\bar{\mathbf{N}}_{DD,r_t}$  and associated fixed position denoted by  $\bar{\mathbf{p}}_{r_t}^{WE}$  based on the float position solution  $[\mathbf{p}_{r_t}^{WE} \ \mathbf{N}_{DD,r_t}]^T$  and associated covariance  $\mathbf{Q}_{r_t}$ . The LAMBDA algorithm (Teunissen, 2003) can be adopted as an efficient solution to this problem. As described in a previous study, (Teunissen, 2003), the first step will be to decorrelate the ambiguities through a z-transform (Teunissen, 2003)  $\hat{\mathbf{Z}}_t = \mathbf{Z}^T \mathbf{N}_{DD,r_t}$  to obtain a near-diagonal form of the covariance matrix  $\mathbf{Q}_{nn}$  as shown in Equation (14):

$$\mathbf{Q}_{\hat{\mathbf{z}}_t \hat{\mathbf{z}}_t} = \mathbf{Z}^T \mathbf{Q}_{nn} \mathbf{Z} \quad (14)$$

Subsequently, the optimal candidate ( $\bar{\mathbf{z}}_t$ ) of the integer set in  $\mathbf{z}$  space can be found by minimizing the following weighted square norm as shown in Equation (15):

$$\arg \min_{\mathbf{z}_t \in \mathbf{Z}^{S-1}} \|\hat{\mathbf{z}} - \mathbf{z}_t\|_{\mathbf{Q}_{\hat{\mathbf{z}}_t \hat{\mathbf{z}}_t}}^2 \quad (15)$$

The search for the minimization is performed by testing the potential grid points inside the known integer candidates. During the searching process, multiple candidates will be found within the ellipsoid. The candidate that leads to the minimum

square norm is selected as the optimal integer solution (denoted  $\check{\mathbf{z}}_t$ ). Then,  $\check{\mathbf{z}}_t$  is transformed back to obtain the fixed positioning solution  $\check{\mathbf{p}}_{r_t}^{WE}$ . Upon validation by the simple ratio-test (Teunissen, 2003), the fixed solution ( $\check{\mathbf{p}}_{r_t}^{WE}$ ) can be expressed as indicated in Equation (16):

$$\check{\mathbf{p}}_{r_t}^{WE} = \mathbf{p}_{r_t}^{WE} - \mathbf{Q}_{pn} \mathbf{Q}_{nn}^{-1} (\mathbf{N}_{DD,r_t} - \check{\mathbf{N}}_{DD,r_t}) \quad (16)$$

Thus, both the fixed solution and integer ambiguity can be estimated using this method. In an open-sky area, the fixed solution can be achieved with centimeter-level accuracy (Takasu & Yasuda, 2009).

### 2.5.3 | Effects of GNSS NLOS reception on GNSS-RTK positioning

In contrast to open-sky conditions, numerous NLOS receptions are introduced in an urban canyon environment with dense and tall building structures. A typical urban canyon with 12 satellites, six of which transmit NLOS receptions (red circles) is shown in Figure 2(c). For this model, we assume that the first  $m$  satellites transmit NLOS measurements with certain bias in pseudorange and carrier-phase measurements denoted by  $\boldsymbol{\rho}_{DD,b,r_t} = [\rho_{DD,b,r_t}^0, \rho_{DD,b,r_t}^1, \dots, \rho_{DD,b,r_t}^{m-1}, 0, \dots, 0]^T$  and  $\boldsymbol{\psi}_{DD,b,r_t} = [\psi_{DD,b,r_t}^0, \psi_{DD,b,r_t}^1, \dots, \psi_{DD,b,r_t}^{m-1}, 0, \dots, 0]^T$ , respectively. The subscript  $b$  represents the bias. Therefore, the consequent bias in the float solution can be estimated as shown in Equation (17):

$$\begin{bmatrix} \Delta \mathbf{p}_{r_t}^{WE} \\ \Delta \mathbf{N}_{DD,r_t} \end{bmatrix} = (\mathcal{M}_{r_t}^T \mathbf{W}_{r_t} \mathcal{M}_{r_t})^{-1} \mathcal{M}_{r_t}^T \mathbf{W}_{r_t} \begin{bmatrix} \boldsymbol{\psi}_{DD,b,r_t} \\ \boldsymbol{\rho}_{DD,b,r_t} \end{bmatrix} \quad (17)$$

Note that this formula is based on the premise that the positioning error caused by the GNSS NLOS can only exert only a minor impact on the geometric matrix  $\mathcal{M}_{r_t}$ , as the satellites are about 20,000 kilometers away from Earth (Zampieri et al., 2020). As a result, the accuracy of the float solution estimated from Equation (11) is degraded with bias terms  $\Delta \mathbf{p}_{r_t}^{WE}$  and  $\Delta \mathbf{N}_{DD,r_t}$ . According to Teunissen (2001), the success rate of the integer ambiguity resolution in the presence of bias can be derived using Equation (18):

$$P_b(\mathbf{z} = \check{\mathbf{z}}_t) = \prod_{s=0}^{S-1} \left( \Phi \left( \frac{1 + 2\mathbf{c}_s \mathbf{L}_z^{-1} \mathbf{Z}^T \Delta \mathbf{N}_{DD,r_t}}{2\mathbf{c}_s \mathbf{D}_z \mathbf{c}_s^T} \right) + \Phi \left( \frac{1 - 2\mathbf{c}_s \mathbf{L}_z^{-1} \mathbf{Z}^T \Delta \mathbf{N}_{DD,r_t}}{2\mathbf{c}_s \mathbf{D}_z \mathbf{c}_s^T} \right) - 1 \right) \quad (18)$$

with triangular factorization  $\mathbf{Z}^T \mathbf{Q}_{nn} \mathbf{Z} = \mathbf{L}_z \mathbf{D}_z \mathbf{L}_z^T$  where  $P_b(\mathbf{z} = \check{\mathbf{z}}_t)$  represents the success rate of correct integer ambiguity resolution. The subscript  $s$  represents the index of the DD measurements while  $\mathbf{c}_s$  represents a  $(S-1) \times 1$  vector with its  $s$ -th element equaling 1 for entries equaling 0. The function  $\Phi(*)$  can be expressed by Equation (19) as follows:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}y^2)} dy \quad (19)$$

Therefore,  $\Delta \mathbf{N}_{DD,r_t}$  caused by biased terms from the NLOS receptions can also decrease the success rate of the integer ambiguity resolution. In this paper, the sky-pointing camera is employed to detect the exclusion of NLOS, as will be discussed in Section 3A.

### 2.5.4 | Effects of satellite geometry on GNSS-RTK positioning

Assuming that the GNSS NLOS receptions can be effectively detected and removed, the success rate of the integer ambiguity resolution can be re-written as shown in Equation (20):

$$P_b(\mathbf{z} = \check{\mathbf{z}}_t) = \prod_{s=0}^{s-1} \left( 2\Phi\left(\frac{1}{2\mathbf{c}_s \mathbf{D}_z \mathbf{c}_s^T}\right) - 1 \right) \quad (20)$$

According to Wang et al. (2020), the aforementioned formula related to the success rate can be further simplified as shown in Equation (21):

$$P_b(\mathbf{z} = \check{\mathbf{z}}_t) = \prod_{s=0}^{s-1} \left( 2\Phi\left(\frac{1}{2\sqrt[s-1]{\det(\mathbf{Q}_{nn})}}\right) - 1 \right) \quad (21)$$

where the operator  $\det(*)$  is employed to calculate the determinant of the  $\mathbf{Q}_{nn}$ . Given the certain weighting matrix, the success rate of the integer ambiguity resolution is directly determined by  $\sqrt[s-1]{\det(\mathbf{Q}_{nn})}$  which is also denoted as ADOP and is further dependent on the satellite geometry matrix  $\mathbf{G}_{r_t}$  (K. Wang et al., 2020). Unfortunately, the GNSS NLOS exclusion can significantly decrease the number of available satellites. For example, only six satellites with high-elevation angles remain after the GNSS NLOS exclusion, as shown in Figure 2(c). Therefore, the LOS of the received satellite is deterministic and uncorrelated since different satellites have different elevations and azimuth angles. As a result, the rank of  $\mathbf{G}_{r_t}$  decreases and induces an increase in  $\det(\mathbf{Q}_{nn})$  (K. Wang et al., 2020), thus reducing the success rate of integer ambiguity resolution. This is one of the major problems limiting the performance of GNSS-RTK positioning in urban canyons even when the potential GNSS NLOS receptions are removed (Furukawa et al., 2020). To fill this gap, the low-lying visual landmarks are adopted here for the first time to compensate for the removal of low-elevation angle NLOS satellites and thus to improve the geometric constraints. This will be discussed in Section 4.

## 3 | IMPROVING GEOMETRY: TIGHTLY-COUPLED INTEGRATION OF GNSS-RTK/VISUAL/INERTIAL

In this section, we will discuss the use of low-lying visual landmarks to improve geometric constraints.

### 3.1 | System State and Factor Graph Structure

To make use of the visual measurements, we will refer to the work of Qin et al. (2018). The visual measurements and IMU measurements are tightly integrated with the GNSS DD pseudorange and carrier-phase measurements, together with the Doppler measurements using the sliding window factor graph optimization method. The GNSS measurements are reported in the ECEF frame and are thus different from the VINS system. The extrinsic parameters between the world local frame ( $\cdot^{WL}$ ) of the conventional VINS and the GNSS ECEF frame remain unknown and will need to be estimated simultaneously. To achieve this, the states of the tightly-coupled system include the following:

- a) The position and orientation of the IMU body related to the world local frame are denoted as  $\mathbf{p}_{b_k}^{WL}$  and  $\mathbf{q}_{b_k}^{WL}$ , respectively, with  $b_k$  representing the body frame, in which the  $k$ -th image is captured.
- b) The velocity of the IMU body related to the world local frame is denoted as  $\mathbf{v}_{b_k}^{WL}$  together with the bias of the gyroscope  $\mathbf{b}_{w_k}$  and accelerometer  $\mathbf{b}_{a_k}$ .
- c) The inverse depth of the visual landmark is denoted as  $f_m$ .
- d) The extrinsic parameter between the world local frame  $(\cdot)^{WL}$  and the ENU frame is denoted as  $\mathbf{T}_{WL}^{WG} = [\mathbf{R}_{WL}^{WG}, \mathbf{t}_{WL}^{WG}]$ . Because the directions of the  $z$ -axis represent both the ENU frame and the world local frame of VINS (Qin et al., 2018), the  $\mathbf{R}_{WL}^{WG}$  will include only one unknown degree, i.e., the yaw offset between both frames.
- e) The receiver clock drift is denoted as  $\dot{\delta}_{r_t}^r$ .

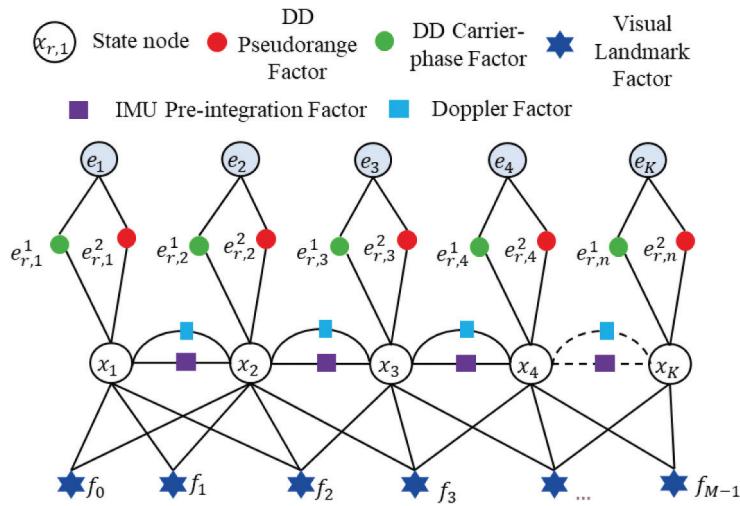
Therefore, the state set of the system inside a sliding window can be expressed as indicated in Equation (22):

$$\chi = [\mathbf{x}_0, \dots, \mathbf{x}_{K-1}, f_0, \dots, f_{M-1}, \mathbf{T}_{WL}^{WG}] \quad (22)$$

$$\text{With } \mathbf{x}_k = [\mathbf{p}_{b_k}^{WL}, \mathbf{q}_{b_k}^{WL}, \mathbf{b}_{a_k}, \mathbf{b}_{w_k}, \dot{\delta}_{r_t}^r]$$

where  $K$  represents the size of the sliding window and  $M$  represents the total number of features involved in the sliding window. Since the data frequency of the GNSS raw measurements is different from the image, we interpolate (Ch'ng et al., 2019) the intermediate GNSS measurement at time  $r_t$  towards the keyframe at time  $b_k$ .

The factor graph structure of the proposed tightly-coupled integration via a sliding window factor graph optimization is shown in Figure 4. Five types of factors are involved in generating the factor graph. Specifically, Doppler and IMU pre-integration factors provide a relative constraint between two consecutive epochs. The DD pseudorange and carrier-phase measurements provide absolute constraints related to the ECEF frame that are free of drift. The visual



**FIGURE 4** Factor graph structure of the proposed tightly coupled integration. Five types of factors contribute to a sliding window. The red and green circles denote the DD pseudorange and carrier-phase factors, respectively. Note that in this example the DD factor only connects a single epoch for simplicity. However, the DD factor is correlated with two consecutive epochs due to the interpolation. The purple and light blue rectangles denote the IMU pre-integration and Doppler factors, respectively. The blue six-pointed stars denote the visual landmarks.

provides the multi-epoch constraints across the factor graph. The remaining text in this Section presents formulations of these factors.

### 3.2 | IMU Pre-integration Factor

Similar to the findings presented by Qin et al. (2018), the IMU pre-integration technique (Forster et al., 2016) was employed to integrate several measurements (inertial measurements between time intervals  $t \in [t_k, t_{k+1}]$ ) into a single factor between two consecutive frames  $b_k$  and  $b_{k+1}$ ; the residual  $r_B(\cdot)$  is expressed as  $r_B(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \boldsymbol{\chi})$ . Because the IMU pre-integration employed in this study is identical to that presented by Qin et al. (2018), the reader may refer to Equations (14), (16), and (17) from this paper for additional details on IMU pre-integration.

### 3.3 | Visual Landmark Factor

The visual measurement employed in this study is a set of features identified by the Shi-Tomasi corner algorithm (Shi, 1994). After feature detection, the optical-flow feature tracking algorithm (Lucas & Kanade, 1981) was performed to track them between consecutive epochs. For the details of the feature detection and tracking, readers are referred to Bai et al. (2020a). Considering that the  $l$ -th feature was first observed in the  $i$ -th image and again in  $j$ -th image, let  $(\hat{u}_l^{c_i}, \hat{v}_l^{c_i})$  represent the pixel position of  $l$ -th feature in the  $i$ -th image, and let  $(\hat{u}_l^{c_j}, \hat{v}_l^{c_j})$  represent the pixel position of  $l$ -th feature in the  $j$ -th image. Therefore, the reprojection model of  $l$ -th feature from the  $i$ -th image to the  $j$ -th image can be further expressed as indicated in Equation (23):

$$\mathbf{p}_l^{c_j} = \mathbf{R}_b^c (\mathbf{R}_w^{b_j} (\mathbf{R}_{b_i}^w (\mathbf{R}_c^b \frac{1}{\lambda_l} \pi_c^{-1} \begin{bmatrix} \hat{u}_l^{c_i} \\ \hat{v}_l^{c_i} \end{bmatrix} + \mathbf{p}_c^b) + \mathbf{p}_{b_i}^w - \mathbf{p}_{b_j}^w) - \mathbf{p}_c^b) \quad (23)$$

Let  $\mathbf{p}_l^{c_j}$  represent the 3D coordinates of the  $l$ -th feature with:

$$\bar{\mathbf{p}}_l^{c_j} = \frac{\mathbf{p}_l^{c_j}}{\|\mathbf{p}_l^{c_j}\|} \quad (24)$$

where  $\bar{\mathbf{p}}_l^{c_j}$  represents the expected observation in the normalized plane. Let the observation measurement of  $l$ -th feature in the  $j$ -th image be  $\hat{\mathbf{p}}_l^{c_j}$  as shown in Equation (25):

$$\hat{\mathbf{p}}_l^{c_j} = \pi_c^{-1} \begin{bmatrix} \hat{u}_l^{c_j} \\ \hat{v}_l^{c_j} \end{bmatrix} \quad (25)$$

hence, the reprojection residual model can be derived as follows (Equation (26)):

$$r_c(\hat{\mathbf{z}}_l^{c_j}, \boldsymbol{\chi}) = (\hat{\mathbf{p}}_l^{c_j} - \bar{\mathbf{p}}_l^{c_j}) \quad (26)$$

where  $\mathcal{C}$  represents the set of features that have been observed at least twice and  $r_c(\cdot)$  represents the residual of the  $l$ -th feature measurement in the  $j$ -th image.  $\hat{\mathbf{z}}_l^{c_j}$  represents the observation measurement.

### 3.4 | GNSS DD Pseudorange/Carrier-Phase Factor

Because the GNSS measurements are represented in the ECEF frame, the extrinsic parameter between the ECEF frame and the world local frame can be expressed as  $\mathbf{T}_{WL}^{WE} = \mathbf{T}_{WG}^{WE} \mathbf{T}_{WL}^{WG}$ . In this paper, the first point is selected as the reference position for the ENU frame, which can be expressed as  $\mathbf{p}_{ref}^{WE}$  with  $\mathbf{p}_{WG}^{WE} = \mathbf{p}_{ref}^{WE}$ .

#### 3.4.1 | DD pseudorange factor

Based on a given DD pseudorange measurement received at epoch  $r_t$ , the residual can be expressed as indicated in Equation (27):

$$\mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL}) = (r_{r_t}^s - r_{e,r_t}^s) - (r_{r_t}^w - r_{e,r_t}^w) - \rho_{DD,r_t}^s \quad (27)$$

where the variable  $\mathbf{p}_{r_t,b}^{WL}$  represents the position of the IMU body frame (IMU sensor frame) at the given epoch  $r_t$ .  $r_{r_t}^s$  and  $r_{r_t}^w$  represent the range distances between the GNSS receiver and the satellite  $s$  and  $w$ . They can be expressed as indicated in Equation (28):

$$\begin{aligned} r_{r_t}^s &= \left\| \mathbf{p}_{r_t}^{WE,s} - \left( \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{p}_{r_t}^{WL} + \mathbf{R}_{WG}^{WE} \mathbf{p}_{WL}^{WG} + \mathbf{p}_{WG}^{WE} \right) \right\| \\ r_{r_t}^w &= \left\| \mathbf{p}_{r_t}^{WE,w} - \left( \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{p}_{r_t}^{WL} + \mathbf{R}_{WG}^{WE} \mathbf{p}_{WL}^{WG} + \mathbf{p}_{WG}^{WE} \right) \right\| \end{aligned} \quad (28)$$

Similarly,  $r_{e,r_t}^s$  and  $r_{e,r_t}^w$  represent the range distance between the reference station and satellite  $s$  and the master satellite  $w$  (as shown in Figure 3). They can be expressed as shown in Equation (29):

$$\begin{aligned} r_{e,r_t}^s &= \left\| \mathbf{p}_{r_t}^{WE,s} - \mathbf{p}_e \right\| \\ r_{e,r_t}^w &= \left\| \mathbf{p}_{r_t}^{WE,w} - \mathbf{p}_e \right\| \end{aligned} \quad (29)$$

Likewise,  $\mathbf{p}_{r_t}^{WL}$  in (28) can be expressed as shown in Equation (30):

$$\mathbf{p}_{r_t}^{WL} = \mathbf{p}_{r_t,b}^{WL} + \mathbf{R}_{r_t,b}^{WL} \mathbf{p}_r^b \quad (30)$$

where  $\mathbf{p}_r^b$  represents the translation between the GNSS receiver and the IMU sensor frame. However, the position estimation of the IMU body frame  $\mathbf{p}_{r_t,b}^{WL}$  at GNSS epoch  $r_t$  cannot be obtained directly. Assuming that the GNSS epoch  $r_t$  locates between two image keyframes  $[t_k, t_{k+1}]$  with  $r_t \in [t_k, t_{k+1}]$ , the following linear interpolation (Gong et al., 2020) can be adopted as shown in Equation (31):

$$\mathbf{p}_{r_t,b}^{WL} = \frac{(t_{k+1} - r_t)}{(t_{k+1} - t_k)} \mathbf{p}_{b_k}^{WL} + \left( 1 - \frac{(t_{k+1} - r_t)}{(t_{k+1} - t_k)} \right) \mathbf{p}_{b_{k+1}}^{WL} \quad (31)$$

#### 3.4.2 | DD carrier-phase factor

Given a DD carrier-phase measurement received at epoch  $r_t$ , the residual can be expressed as indicated in Equation (32):

$$\mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL}) = (r_t^s - r_{e,r_t}^s) - (r_{r_t}^w - r_{e,r_t}^w) + \lambda_j N_{DD,r_t}^s - \lambda_j \psi_{DD,r_t}^s \quad (32)$$

Similarly, the uncertainty related to the DD carrier-phase factor  $\mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})$  is calculated based on the SNR and elevation angle (Herrera et al., 2015).

### 3.4.3 | GNSS Doppler factor

Given a Doppler measurement received at epoch  $r_t$ , the residual can be expressed as indicated in Equation (33):

$$\mathbf{r}_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL}) = \frac{1}{\lambda_j} \left( \mathbf{e}_{r_t}^s \left( \mathbf{v}_{r_t}^{WE,s} - \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{v}_{r_t}^{WL} \right) + c \left( \dot{\delta}_{r_t}^{r,j} - \dot{\delta}_{r_t}^s \right) \right) - d_{r_t}^s \quad (33)$$

$$\text{With } \mathbf{v}_{r_t}^{WL} = \mathbf{v}_{r_t,b}^{WL} + \mathbf{R}_{r_t,b}^{WL} [\mathbf{p}_r^b] \times (\boldsymbol{\omega}_{r_t} - \mathbf{b}_{\omega_r})$$

where the  $\mathbf{v}_{r_t}^{WL}$  represents the velocity of the GNSS receiver at epoch  $r_t$  related to the local world frame. The operator  $[*]_x$  is employed to obtain the skew-symmetric matrix corresponding to a vector. Similar time interpolation is employed for the velocity constraint.  $\mathbf{e}_{r_t}^s$  which represents the LOS vector connecting the GNSS receiver and the satellite  $s$ . This can be expressed as shown in Equation (34):

$$\mathbf{e}_{r_t}^s = \frac{\mathbf{p}_{r_t}^{WE,s} - (\mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{p}_{r_t}^{WL} + \mathbf{R}_{WG}^{WE} \mathbf{p}_{WL}^{WG} + \mathbf{p}_{WG}^{WE})}{\left\| \mathbf{p}_{r_t}^{WE,s} - (\mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{p}_{r_t}^{WL} + \mathbf{R}_{WG}^{WE} \mathbf{p}_{WL}^{WG} + \mathbf{p}_{WG}^{WE}) \right\|} \quad (34)$$

### 3.5 | Marginalization Factor

To maintain the efficiency and consistency of the state estimation inside of the sliding window, the oldest keyframe is marginalized and the associated factors are added as a prior, including an IMU pre-integration factor, visual factor, and GNSS-related factors, by applying the Schur complement (Sibley et al., 2010). Marginalization plays an important role in sliding window optimization as it incorporates historical information into a prior. The information from marginalization is denoted as  $\{\mathbf{r}_p, \mathbf{H}_p\}$  with the  $\mathbf{r}_p$  and  $\mathbf{H}_p$  representing the prior residual and information matrix, respectively.

### 3.6 | Tightly-Coupled Optimization

Based on the factors derived above, the states inside the sliding window can be estimated by solving the following objective function as shown in Equation (35):

$$\begin{aligned} \chi^* = \arg \min_{\chi} \sum_{s, r_t, t, b_k} & \left( \left\| \mathbf{r}_p - \mathbf{H}_p \chi \right\|^2 + \left\| \mathbf{r}_B(\hat{\mathbf{z}}_{b_k}^{b_{k+1}}, \chi) \right\|_{\Sigma_B}^2 + \left\| \mathbf{r}_C(\hat{\mathbf{z}}_l^{c_j}, \chi) \right\|_{\Sigma_C}^2 \right. \\ & \left. + \left\| \mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL}) \right\|_{\sigma_{\rho,r_t}^s}^2 + \left\| \mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL}) \right\|_{\sigma_{\psi,r_t}^s}^2 + \left\| \mathbf{r}_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL}) \right\|_{\sigma_{d,r_t}^s}^2 \right) \end{aligned} \quad (35)$$

where  $\Sigma_B$  and  $\Sigma_C$  represents the covariance matrix related to the IMU pre-integration and visual reprojection, respectively. The uncertainty ( $\sigma_{\rho,r_t}^s$ ) related to the DD pseudorange factor  $\mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})$  is calculated based on the

SNR and elevation angle (Herrera et al., 2015). A similar method is applied for the uncertainty estimation of the DD carrier-phase ( $\sigma_{\psi, r_t}^s$ ) and  $\sigma_{\psi, r_t}^s = k_\psi \sigma_{\rho, r_t}^s$  with a scaling factor  $k_\psi = 100$  (Takasu & Yasuda, 2009). Similarly, the uncertainty of the Doppler measurements ( $\sigma_{d, r_t}^s$ ) is  $\sigma_{d, r_t}^s = k_d \sigma_{\rho, r_t}^s$  with a scaling factor  $k_d = 3$ .

### 3.7 | Covariance Estimation and Resolution of Integer Ambiguity

The float solution is obtained by solving the tightly coupled cost function as shown in Equation (35). To resolve the integer ambiguity, the covariance related to the position and float ambiguity is required. With the assistance of the visual constraint and IMU pre-integration constraint, the combined linearized model connecting the float solution and observation can be expressed as shown in Equation (36):

$$\begin{bmatrix} \Psi_{DD, r_t} \\ \boldsymbol{\rho}_{DD, r_t} \\ \mathcal{B} \\ \mathcal{C} \end{bmatrix} = \begin{bmatrix} \lambda_j^{-1} \mathbf{G}_{r_t}^\psi & \mathbf{I}_{S \times S} \\ \mathbf{G}_{r_t}^\rho & 0 \\ \mathbf{G}_{r_t}^d & 0 \\ \mathbf{G}_{r_t}^B & 0 \\ \mathbf{G}_{r_t}^C & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_{r_t}^{WE} \\ \lambda_j \mathbf{N}_{DD, r_t} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{DD, r_t} \\ \boldsymbol{\epsilon}_{DD, r_t} \end{bmatrix} \quad (36)$$

where  $\mathcal{B}$  and  $\mathcal{C}$  represent the set of IMU and visual measurements involved. Therefore, the covariance can be estimated as shown in Equation (37):

$$\mathbf{Q}_{r_t}^{3DVA} = \left( \begin{bmatrix} \lambda_j^{-1} \mathbf{G}_{r_t}^\psi & \mathbf{I}_{S \times S} \\ \mathbf{G}_{r_t}^\rho & 0 \\ \mathbf{G}_{r_t}^d & 0 \\ \mathbf{G}_{r_t}^B & 0 \\ \mathbf{G}_{r_t}^C & 0 \end{bmatrix}^T \mathbf{W}_{r_t}^{3DVA} \begin{bmatrix} \lambda_j^{-1} \mathbf{G}_{r_t}^\psi & \mathbf{I}_{S \times S} \\ \mathbf{G}_{r_t}^\rho & 0 \\ \mathbf{G}_{r_t}^d & 0 \\ \mathbf{G}_{r_t}^B & 0 \\ \mathbf{G}_{r_t}^C & 0 \end{bmatrix} \right)^{-1} \quad (37)$$

where  $\mathbf{W}_{r_t}^{3DVA}$  represents the weightings involving the uncertainty of the GNSS DD carrier-phase ( $\mathbf{W}_{r_t}^\psi$ ), DD pseudorange ( $\mathbf{W}_{r_t}^\rho$ ) and Doppler ( $\mathbf{W}_{r_t}^d$ ) measurements, and visual ( $\mathbf{W}_{r_t}^C$ ) and inertial pre-integration ( $\mathbf{W}_{r_t}^B$ ) measurements, which can be further expressed as indicated in Equation (38):

$$\mathbf{W}_{r_t}^{3DVA} = \begin{bmatrix} \mathbf{W}_{r_t}^\psi & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{r_t}^\rho & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_{r_t}^d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}_{r_t}^B & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}_{r_t}^C \end{bmatrix} \quad (38)$$

In contrast to the conventional GNSS-RTK positioning, in which the covariance of the float ambiguity is dominated by the geometry of surviving satellites after the GNSS NLOS exclusion, the position of the GNSS receiver is constrained by both DD satellite and visual measurements. To present the  $\mathbf{Q}_{r_t}^{3DVA}$  in a more representative form, it can be simplified as shown in Equation (39):

$$\mathbf{Q}_{r_t}^{3DVA} = \begin{pmatrix} \lambda_j^{-1} \mathbf{A} & \mathbf{G}_{r_t}^{\psi^T} \mathbf{W}_{r_t}^{\psi} \\ \mathbf{W}_{r_t}^{\psi} \mathbf{G}_{r_t}^{\psi} & \lambda_j^{-1} \mathbf{W}_{r_t}^{\psi} \end{pmatrix}^{-1} \quad (39)$$

With  $\mathbf{A} = \lambda_j^{-2} \mathbf{G}_{r_t}^{\psi^T} \mathbf{W}_{r_t}^{\psi} \mathbf{G}_{r_t}^{\psi} + \mathbf{G}_{r_t}^{\rho^T} \mathbf{W}_{r_t}^{\rho} \mathbf{G}_{r_t}^{\rho} + \mathbf{G}_{r_t}^{d^T} \mathbf{W}_{r_t}^d \mathbf{G}_{r_t}^d + \mathbf{G}_{r_t}^{\mathcal{B}^T} \mathbf{W}_{r_t}^{\mathcal{B}} \mathbf{G}_{r_t}^{\mathcal{B}} + \mathbf{G}_{r_t}^{\mathcal{C}^T} \mathbf{W}_{r_t}^{\mathcal{C}} \mathbf{G}_{r_t}^{\mathcal{C}}$

Therefore, the covariance of the float ambiguity is further constrained by the addition of  $\mathbf{G}_{r_t}^{d^T} \mathbf{W}_{r_t}^d \mathbf{G}_{r_t}^d + \mathbf{G}_{r_t}^{\mathcal{B}^T} \mathbf{W}_{r_t}^{\mathcal{B}} \mathbf{G}_{r_t}^{\mathcal{B}} + \mathbf{G}_{r_t}^{\mathcal{C}^T} \mathbf{W}_{r_t}^{\mathcal{C}} \mathbf{G}_{r_t}^{\mathcal{C}}$ . Specifically, given the assumptions that: (1) the visual landmark measurements have an accuracy that is similar to the DD pseudorange measurements and (2) the geometry matrix  $\mathcal{M}_{r_t}^{3DVA}$  is a full rank, which means that the satellite and the visual measurements are uncorrelated and the uncertainty associated with the float ambiguity will be decreased (Wang et al., 2020). This would result in a decrease in the determinant of  $\mathbf{Q}_{nn}$ , thereby increasing the success rate ( $P_b[\mathbf{z} = \check{\mathbf{z}}_t]$ ) of integer ambiguity resolution. This will be evaluated in the experimental validation presented in Section 6.

## 4 | SYSTEM INITIALIZATION

The initialization of the system state is significant for an estimator with high nonlinearity related as shown in Equation (35). This only needs to be done once when the system starts up; this is one of the major differences between the 3D building model-aided GNSS positioning method and the initialization of the system state discussed in this study. The initialization presented here involves primarily (1) the visual/inertial (VI) initialization and (2) GNSS-VI initialization which refers to the extrinsic parameters connecting GNSS-RTK and VINS. These will be presented in the following Section.

### 4.1 | VI Initialization

The VI initialization includes the recovery of the scale of the features as well as the bias of the gyroscope, accelerometers, and other elements. We used the work of Qin et al. (2018) to complete the visual/inertial initialization. Meanwhile, the outcome of the VI initialization includes a set position and velocity of the keyframes inside the sliding window, denoted by  $\{\mathbf{p}_{b_k}^{WL} | k = 0, \dots, K - 1\}$  and  $\{\mathbf{v}_{b_k}^{WL} | k = 0, \dots, K - 1\}$ , which will be used subsequently for the GNSS-VI initialization.

### 4.2 | GNSS-VI Initialization

The GNSS-VI initialization aims to generate an initial guess for various GNSS-related states, including integer ambiguity, receiver clock bias drift rate, and the extrinsic parameter ( $\mathbf{T}_{WL}^{WG}$ ) between the ENU frame and the world local frame. Conventionally, the initial guess of the GNSS-related states is estimated using an epoch-by-epoch-based weighted least squares (WLS) approach (Takasu & Yasuda, 2009). Subsequently, the initial guess of the extrinsic parameters  $\mathbf{T}_{WL}^{WG}$  can be optimized by aligning the two trajectory sets estimated from the VI initialization and GNSS WLS, respectively (Liu et al., 2021). However, the potential of the Doppler measurements is neglected when using this approach. By contrast, Doppler measurements are considered in the second stage of initialization after VI initialization

and GNSS SPP. The velocity is then estimated from GNSS Doppler measurements with VI used to align the local and global frames (Cao et al., 2022). However, the carrier-phase is still not adapted for this work due to the unknown integer ambiguity. To benefit from the Doppler and carrier-phase measurements with very high precision, we propose a method in which we estimate the velocity and trajectory sets inside a sliding window that is based simultaneously on the Doppler, DD pseudorange, and carrier-phase measurements.

Given the fact that the carrier-phase measurements received from the same satellite  $s$  inside the sliding window involve the same integer ambiguity value, a left null space matrix is employed to eliminate the shared integer ambiguity within the sliding window. This facilitates the utilization of the DD carrier-phase measurement without resolving the known integer ambiguity.

Based on a set of carrier-phase measurements from satellite  $s$  received in multiple discrete epochs, they can be stacked into the following form as indicated in Equation (40):

$$\begin{bmatrix} \lambda_j \psi_{DD,0}^s \\ \lambda_j \psi_{DD,1}^s \\ \dots \\ \lambda_j \psi_{DD,K-1}^s \end{bmatrix} = \begin{bmatrix} m_{r,0}^s \\ m_{r,1}^s \\ \dots \\ m_{r,K-1}^s \end{bmatrix} + \begin{bmatrix} \lambda_j N_{DD,0}^s \\ \lambda_j N_{DD,1}^s \\ \dots \\ \lambda_j N_{DD,K-1}^s \end{bmatrix} + \begin{bmatrix} \epsilon_{DD,0}^s \\ \epsilon_{DD,1}^s \\ \dots \\ \epsilon_{DD,K-1}^s \end{bmatrix} \quad (40)$$

with  $m_{r,t}^s = (r_{r_t}^s - r_{e,r_t}^s) - (r_{r,r_t}^w - r_{e,r_t}^w)$

where  $K$  represents the size of the sliding window. This means that satellite  $s$  is tracked by the GNSS receiver continuously for  $K$  epochs. The  $m_{r,t}^s$  is defined for simplicity. Given the fact that the carrier-phase measurements inside the window share the same integer ambiguity, Equation (40) can be rewritten as follows:

$$\begin{bmatrix} \lambda_j \psi_{DD,0}^s \\ \lambda_j \psi_{DD,1}^s \\ \dots \\ \lambda_j \psi_{DD,K-1}^s \end{bmatrix} = \begin{bmatrix} m_{r,0}^s \\ m_{r,1}^s \\ \dots \\ m_{r,K-1}^s \end{bmatrix} + \lambda_j N_{DD,K-1}^s \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_{DD,0}^s \\ \epsilon_{DD,1}^s \\ \dots \\ \epsilon_{DD,K-1}^s \end{bmatrix} \quad (41)$$

The compact form of Equation (41) can be further organized by multiplying a left null space matrix  $\mathbf{G}_{r,k}^s$  on both sides of Equation (40). Hence, the following form can be obtained as shown in Equation (42):

$$\mathbf{G}_{DD,K}^s \begin{bmatrix} \lambda_j \psi_{DD,0}^s \\ \lambda_j \psi_{DD,1}^s \\ \dots \\ \lambda_j \psi_{DD,K-1}^s \end{bmatrix} = \mathbf{G}_{DD,K}^s \begin{bmatrix} m_{r,0}^s \\ m_{r,1}^s \\ \dots \\ m_{r,K-1}^s \end{bmatrix} + \mathbf{G}_{DD,K}^s \begin{bmatrix} \epsilon_{DD,0}^s \\ \epsilon_{DD,1}^s \\ \dots \\ \epsilon_{DD,K-1}^s \end{bmatrix} \quad (42)$$

with  $\mathbf{G}_{DD,K}^s \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$

Therefore, Equation (42) is free of ambiguous variables which are eliminated by a left multiplication of the left null space matrix  $\mathbf{G}_{DD,K}^s$ . The elimination of the integer ambiguity was based on a study published by Li and Mourikis (2013) in which a similar elimination was employed to avoid the repetitive estimation of the position of landmarks. The feasibility of the integer ambiguity elimination via the left null space matrix was presented in our recent study

(Bai et al., 2022), in which the GNSS Doppler/pseudorange fusion is smoothed by the ambiguity-free carrier-phase measurements. Therefore, the residual of the ambiguity-free DD carrier-phase measurement can be derived as indicated in Equation (43):

$$\mathbf{r}_{\psi_s}(\psi_{DD,0}^s, \dots, \psi_{DD,K-1}^s, \mathbf{p}_{0,b}^{WL}, \dots, \mathbf{p}_{K-1,b}^{WL}) = \mathbf{G}_{DD,K}^s \begin{bmatrix} \lambda_j \psi_{DD,0}^s \\ \lambda_j \psi_{DD,1}^s \\ \vdots \\ \lambda_j \psi_{DD,K-1}^s \end{bmatrix} - \mathbf{G}_{DD,K}^s \begin{bmatrix} m_{r,0}^s \\ m_{r,1}^s \\ \vdots \\ m_{r,K-1}^s \end{bmatrix} \quad (43)$$

where  $\psi_s$  represents a set of DD carrier-phase measurements for satellite  $s$  inside a sliding window with the same integer ambiguity. Therefore, the trajectory and the velocity sets of the sliding window can be estimated based on Doppler, DD pseudorange, and ambiguity-free carrier-phase measurements as shown in Equation (44):

$$\boxed{\boldsymbol{\Sigma}^* = \arg \min_{\boldsymbol{\Sigma}} \sum_{s,r_t,t,b_k} \left( \left\| \mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL}) \right\|_{\sigma_{\rho,r_t}^s}^2 + \left\| \mathbf{r}_{\psi_s}(\psi_{DD,0}^s, \dots, \psi_{DD,K-1}^s, \mathbf{p}_{0,b}^{WL}, \dots, \mathbf{p}_{K-1,b}^{WL}) \right\|_{\Sigma_{\psi_s}^s}^2 \right. \right. \\ \left. \left. + \left\| \mathbf{r}_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL}) \right\|_{\sigma_{d,r_t}^s}^2 \right) } \quad (44)$$

where  $\boldsymbol{\Sigma}$  represents a set of states, including the position, velocity, and the receiver clock bias drift rate.  $\Sigma_{\psi_s}^s$  represents the covariance matrix related to a set of DD carrier-phase measurements  $\Sigma_{\psi_s}^s$  and can be expressed as shown in Equation (45):

$$\Sigma_{r,k}^{WCP,s} = \mathbf{G}_{r,k}^s \begin{bmatrix} \sigma_{\psi,r_t}^s & 0 & 0 & 0 \\ 0 & \sigma_{\psi,r_t}^s & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \sigma_{\psi,r_t}^s \end{bmatrix} \mathbf{G}_{r,k}^s {}^T \quad (45)$$

Therefore, the position and velocity set expressed in the ECEF frame can be estimated by solving Equation (44) denoted by  $\{\mathbf{p}_{r_t}^{WE} | k=0, \dots, K-1\}$  and  $\{\mathbf{v}_{r_t}^{WE} | k=0, \dots, K-1\}$  with  $r_t \in [t_k, t_{k+1}]$ , respectively. Based on the estimated trajectory and velocity sets from VI initialization and optimization of Equation (43), the extrinsic parameter  $\mathbf{T}_{WL}^{WG}$  between the ENU and the world local frame can be estimated by solving the aligning objective function shown in Equation (46):

$$\mathbf{T}_{WL}^{WG*} = \arg \min_{\mathbf{T}_{WL}^{WG}} \sum_{s,r_t,t,b_k} \left( \left( \left\| \mathbf{R}_{WG}^{WE} (\mathbf{p}_{r_t}^{WE} - \mathbf{p}_{WG}^{WE}) - \mathbf{R}_{WL}^{WG} (\mathbf{p}_{r_t,b}^{WL} + \mathbf{R}_{r_t,b}^{WL} \mathbf{p}_r^b) - \mathbf{p}_{WL}^{WG} \right\|^2 \right) \right. \\ \left. + \left( \left\| \mathbf{R}_{WG}^{WE} {}^T \mathbf{v}_{r_t}^{WE} - \mathbf{R}_{WL}^{WG} (\mathbf{v}_{r_t,b}^{WL} + \mathbf{R}_{r_t,b}^{WL} [\mathbf{p}_r^b] \times (\omega_{r_t} - \mathbf{b}_{\omega_r})) \right\|^2 \right) \right) \quad (46)$$

where  $\mathbf{T}_{WL}^{WG*}$  represents the optimal extrinsic parameter to be estimated. Meanwhile, similar time interpolation logic is employed to obtain  $\mathbf{v}_{r_t,b}^{WL}$  from  $\{\mathbf{v}_{b_k}^{WL} | k=0, \dots, K-1\}$ . The extrinsic parameter estimation takes advantage of both the Doppler and DD pseudorange measurements. Meanwhile, the elimination operation facilitates the utilization of the high-accuracy carrier-phase measurement without resolving the integer ambiguity. In a previous (Bai et al., 2022), we demonstrated that the multi-epochs carrier-phase constraint corresponding to Equation (43) can generate a significantly smoother trajectory compared with a single-epoch DD carrier-phase constraint. Therefore, the GNSS-VI initialization is completed with the extrinsic parameter ( $\mathbf{T}_{WL}^{WG}$ ) estimated as shown in Equation (46).



**FIGURE 5** Demonstration of the experimental vehicle and the evaluated urban canyons 1 and 2

## 5 | EXPERIMENTAL VALIDATION

### 5.1 | Experiment Setup

#### 5.1.1 | *Experimental scenarios*

To verify the effectiveness of the proposed method, two experiments were conducted in urban canyons in Hong Kong (Figure 5). The top-left and bottom-left figures present the urban canyon scenarios evaluated in this study. Both urban scenarios contain static buildings, trees, and dynamic objects, such as double-decker buses, all of which have the potential to generate GNSS NLOS receptions. To determine how the proposed method works in scenarios with various degrees of urbanization, the first experiment was performed in a typical urban canyon scenario in Hong Kong (urban canyon 1 in Figure 5[a]). The second experiment was conducted in a more urbanized scenario (urban canyon 2 in Fig. 5b) with denser and taller building structures; this type of scenario is more challenging for GNSS-RTK positioning. Of note, we do not know the ground truth of the fixed solution even when the GNSS-RTK passed the ratio test with a threshold of 3.0. During the evaluation of this paper, we labeled a fixed solution based on two criteria: (1) the ratio test should be passed with a threshold of 3.0 and (2) the 3D positioning error should not exceed 20 centimeters.

#### 5.1.2 | *Sensor setups*

The details of the data collection vehicle can be found in our recently open-sourced UrbanNav data set (Hsu et al., 2021)<sup>1</sup>. This data set includes multi-sensor data collected in typical urban canyons in Hong Kong and Tokyo. In both experiments, a low-cost automobile-level u-blox M8T GNSS receiver was used to collect raw single-frequency GPS/BeiDou measurements at a frequency of 10 Hz. The Xsens Ti-10 IMU was employed to collect data at the frequency of 400 Hz. A forward-looking monocular camera was used to capture the visual features with a FOV of 90° (horizontal) x 60° (vertical) at a frequency of 10 Hz together with a sky-pointing fish-eye camera that captured the sky-view image with a FOV of 185° at a frequency of 10 Hz. Note that the fish-eye camera can help to capture the wide sky view image which benefits the detection of GNSS NLOS satellites. However, the significant geometrical distortion caused

<sup>1</sup><https://github.com/weisongwen/UrbanNavDataset>

by the fish-eye lens will damage the image quality which can limit feature detection. Therefore, a forward-looking monocular camera is used to detect the features. The sensor setup is shown in Figure 5(c). The NovAtel SPAN-CPT (Kennedy et al., 2006), GNSS (GPS, GLONASS, and BeiDou), and RTK/INS (fiber-optic gyroscopes, FOG) integrated navigation system were employed to provide ground truth of positioning. The gyro bias in-run stability of the FOG is one degree per hour and its random walk is 0.067 degrees per hour. According to the specification of the NovAtel SPAN-CPT, a centimeter-level accuracy can be obtained when the RTK correction is available with the correct fixed solution. Moreover, the ground truth output from SPAN-CPT was post-processed using the *Inertial Explorer* software from NovAtel to guarantee the accuracy of the trajectory by processing forward and reverse in time and backward smoothing followed by combining the results. All data were post-processed with a desktop (Intel Core i7-9700K CPU, 3.60 GHz) computer.

### 5.1.3 | Evaluation methods

We verified the effectiveness of the proposed method step-by-step by comparing the following five methods:

- (a) **RTK**: GNSS-RTK positioning results using the conventional method (Herrera et al., 2015). This was to show how the conventional GNSS real-time kinematic positioning performs in the evaluated urban canyon data sets.
- (b) **RTK-NE**: GNSS RTK positioning aided by the GNSS NLOS exclusion (NE) based on the outlier removal presented in Section 2A. This was performed in addition to **RTK**. This was performed to verify the contribution of the GNSS NLOS removal.
- (c) **RTK-NE-VINS**: This loosely integrates the results from **RTK-NE** with the pose estimation from VINS (Qin et al., 2018) in a conventional manner via factor graph optimization as shown previously (Qin et al., 2019). This was performed to show how the loosely-coupled integration of VINS and GNSS-RTK operates even after the GNSS NLOS receptions are removed.
- (d) **RTK-IG**: Used when the geometric constraints are improved by tightly coupling the visual measurements with the GNSS measurements as shown in Section 4. This was performed to demonstrate the improvement resulting from the tightly-coupled visual measurements with GNSS-RTK, which is the main contribution of this work. We note that the GNSS NLOS receptions are not detected using this method. “**IG**” denotes improved geometry.
- (e) **RTK-NE-IG**: Used once the GNSS NLOS receptions are removed and after the geometric constraints are improved by tightly coupling the visual measurements with the healthy GNSS measurements as shown in Section 4. This was performed to demonstrate the outcomes resulting from both the NLOS exclusion and geometric improvements.

## 5.2 | Experimental Evaluation in Urban Canyon 1

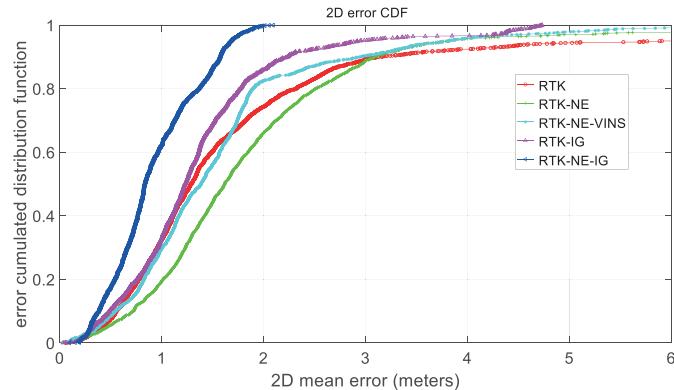
### 5.2.1 | Evaluation of positioning performance

The positioning performance observed using each of the five aforementioned methods is shown in Table 1. The error cumulative distribution function (ECDF) and trajectories are shown in Figure 6 and Figure 7, respectively. A mean error

**TABLE 1**  
Positioning Performance of the Five Methods in Urban Canyon 1

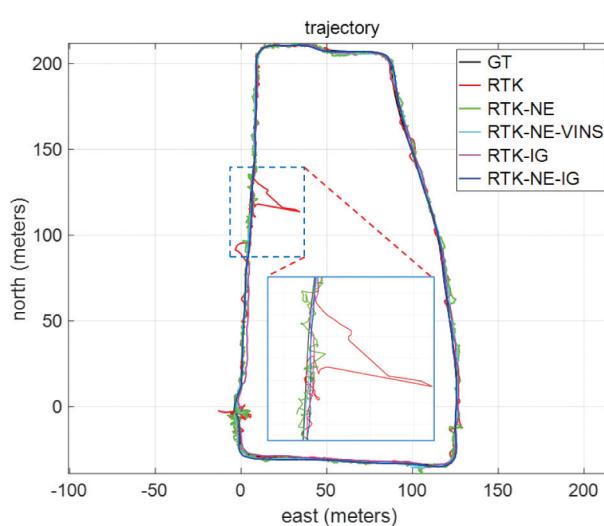
All Data	RTK	RTK-NE	RTK-NE-VINS	RTK-IG	RTK-NE-IG
<b>2D MEAN</b>	1.83	1.78	1.59	1.36	0.84
<b>2D STD</b>	2.01	1.25	1.06	0.84	0.39
<b>2D Max</b>	30.33	9.06	6.53	4.7	1.94
<b>2D Impr.</b>		2.7%	13.11%	25.46%	54.09%
<b>3D MEAN</b>	4.26	5.12	4.21	2.44	2.27
<b>3D STD</b>	6.27	3.76	2.15	0.81	0.85
<b>3D Max</b>	37.62	25.56	9.85	4.8	3.68
<b>3D Impr.</b>		-20%	1.17%	42.72%	46.70%
<b>Fixed Rate</b>	0%	0%	0%	0.49%	1.67%
<b>Avail.</b>	99.1%	72%	100%	100%	100%
<b>Time Consu. (ms)</b>	5.2	5.2	50.7	67.9	65.4

Note: Improvement (Impr.) was calculated based on the RTK method. 2D/3D, horizontal/3D positioning; STD, standard deviation; Avail., availability. Time Consu., the average time consumption of the method.



**FIGURE 6** The cumulative distribution function for 2D positioning errors of the five methods evaluated in urban canyon 1

The red, green, cyan, magenta, and blue denote RTK, RTK-NE, RTK-NE-VINS, RTK-IG, and RTK-NE-IG, respectively.



**FIGURE 7** 2D positioning trajectories of the five methods evaluated in urban canyon 1  
The red, green, cyan, magenta, and blue curves denote the RTK, RTK-NE, RTK-NE-VINS, RTK-IG, and RTK-NE-IG, respectively. The black curve denotes the ground truth trajectory.

of 1.83 meters was obtained using the conventional GNSS-RTK positioning method (Herrera et al., 2016) with a standard deviation of 2.01 meters. Moreover, the maximum error reached 30 meters as shown in the annotated area of Figure 7 by a dashed rectangle. This is very far from the positioning requirements of autonomous systems and was attributed to the GNSS NLOS receptions from tall surrounding building structures. After excluding the detected GNSS NLOS receptions using the RTK-NE method, the error decreased to 1.78 meters, representing only a 2.7% improvement compared with the conventional RTK method. This results from the distortion of the satellite geometry due to the NLOS exclusion. In other words, the GNSS NLOS exclusion can increase the dilution of precision (DOP) (Enge, 1994) and thus increase the uncertainty of GNSS positioning. The average DOP value for the satellite distribution is about 3.1 and 3.5 before and after the GNSS NLOS exclusion, respectively. A more detailed analysis of the geometry of the GNSS-RTK will be presented in the following Section. Moreover, the availability decreases from 99.1% to 72% for RTK-NE because excessive GNSS NLOS may result in an insufficient number of satellite measurements for GNSS positioning. We observed a similar phenomenon in our previous work (Wen et al., 2019c; Wen et al., 2019b). However, the standard deviation (STD) and the maximum errors decrease significantly after excluding the outlier GNSS NLOS receptions, which documents the effectiveness of the GNSS NLOS exclusion in mitigating the adverse impact of gross outlier measurements.

We then integrated the improved GNSS-RTK positioning result from the RTK-NE method with the pose estimation from the VINS as originally described by Qin et al. (2019) to determine how the VINS can help the GNSS-RTK in a loosely-coupled manner. The mean error decreases to 1.59 meters with the assistance of the VINS and results in an improvement of 13.11%. Meanwhile, both the STD and the maximum error decrease slightly and the availability increases to 100% because the state estimation from VINS is continuously available. The improved results demonstrate that the integration of the VINS contributes to resisting the outlier GNSS measurements. However, the improvement is still limited because the loosely coupled integration cannot fully exploit the complementarity between GNSS and visual measurements. Furthermore, no fixed solution can be estimated from RTK, RTK-NE, or RTK-NE-VINS methods with a fixed rate of 0% (He et al., 2014). The fixed rate is defined as a ratio of the number of ambiguity-fixed epochs over the number of total epochs and documents the availability of epochs for positioning.

Interestingly, further improvements were obtained by tightly integrating the visual measurements with the GNSS DD pseudorange, carrier-phase, and Doppler measurements based on the RTK-IG method presented in Section 4. The mean error decreased to 1.36 meters with an improvement of 25.46% and the STD decreased significantly to 0.84 meters. These results demonstrate that tightly-coupled integration can result in better performance compared with loosely-coupled integration. Improvements from tightly-coupled integration result in two main benefits: (1) the tightly coupled visual measurements can improve the geometric constraints as discussed in Section 4, and (2) both the visual and GNSS measurements can be modeled more precisely. With the visual constraint improving the geometry, a fixed rate of 0.49% can be obtained using the RTK-IG method. This achievement documents the effectiveness of the proposed method in improving the geometry of conventional GNSS-RTK. However, the maximum error still reaches 4.7 meters due to unexpected GNSS NLOS receptions. To determine how the GNSS NLOS exclusion might help to improve the RTK-IG method, we removed the GNSS NLOS receptions during the tightly-coupled integration leading to the RTK-NE-IG method. In this case, the mean error decreased to 0.84 meters with an improvement of 54.09%. This method also led to a significantly decreased STD of 0.39 meters and a maximum error that decreased to only 1.94 meters. These results demonstrate that the

GNSS NLOS exclusion can significantly improve the performance of GNSS-IG. Moreover, an increased fixed rate of 1.67% was obtained, which further documents the effectiveness of NLOS exclusion compared with RTK-IG. Of note, the fixed rate remains low even when using the proposed method. This is because the proposed method also resolves the integer ambiguity of the carrier-phase epoch-by-epoch, i.e., so-called instantaneous ambiguity resolution. We plan to examine how the visual measurements can help improve the geometric constraints of the satellite measurements and evaluate the performance improvement epoch-by-epoch. In the future, we plan to add the fixed-and-hold mode to the integer ambiguity resolution together with cycle slip detection. This has the potential to lead to a significantly higher fixed rate.

Although navigation requirements mainly focus on 2D (horizontal) positioning, we also present the results of 3D positioning performance in the bottom section of Table 1. These findings may reveal whether the proposed method can also improve vertical positioning compared with conventional RTK positioning. While the 3D mean error of the conventional RTK method reaches more than 4 meters, the 3D mean error increases to 5.12 meters using the RTK-NE. This is mainly because of the GNSS NLOS exclusion, which can lead to larger vertical dilutions of precision (VDOPs) (Enge, 1994) and thus larger vertical positioning errors. The 3D mean error decreases slightly with the assistance of the loosely coupled integration of VINS, which led to an improvement of 1.17%. A significantly decreased 3D mean error was observed after applying the RTK-IG method which led to an improvement of 42.72%. Further improved 3D positioning was obtained by the GNSS NLOS exclusion with an improvement of 46.70% using RTK-NE-IG.

In summary, the gradual decreases in positioning error demonstrate the contributions of both the GNSS NLOS exclusion and the geometric improvement; these findings were also demonstrated by ECDF as shown in Figure 6. Availability of 100% was obtained using the proposed method (RTK-NE-VINS). While the remaining positioning error still reaches 0.84 meters, mainly because the GNSS multipath effect (Enge, 1994), another major error source, was not considered in this paper. Because the multipath phenomenon involves both direct and reflected signals, it cannot be detected directly via sky view visibility.

### 5.2.2 | Evaluation of rotation performance

Conventionally, the GNSS receiver provides only positioning estimates. The attitude of the system cannot be estimated directly by the GNSS measurements. However, visual and inertial measurements can be used to determine the attitude of the system. Therefore, it will be interesting to determine whether the proposed method can be used for attitude estimation. Because the RTK and RTK-NE methods do not provide attitude estimation, we limited the comparison to results from RTK-NE-VINS, RTK-IG, and RTK-NE-IG methods as shown in Table 2.

**TABLE 2**  
Absolute Rotational Accuracy of the Three Methods Evaluated in Urban Canyon 1

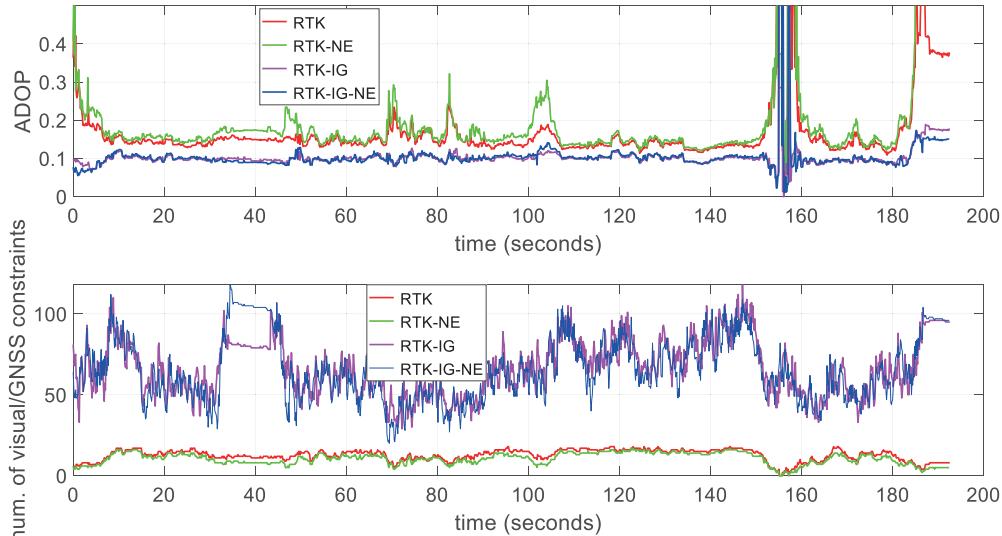
All Data (degree)	RTK-NE-VINS	RTK-IG	RTK-NE-IG
<b>MEAN</b>	5.18	3.01	2.73
<b>STD</b>	3.57	1.71	1.56
<b>Max</b>	6.18	3.84	3.75
<b>Min</b>	1.13	1.08	0.91

The absolute rotation accuracy was evaluated using the popular EVO tool kit (Grupp, 2017). A mean error of 5.18 degrees was obtained using the RTK-NE-VINS with a standard deviation of 3.57 degrees. With the assistance of the tightly coupled integration of visual measurements, the mean error of attitude decreases to 3.01 degrees with a decreased standard deviation of 1.71 degrees. Interestingly, the mean attitude error decreased slightly to 2.73 degrees using GNSS NLOS exclusion by RTK-NE-IG. In summary, a slightly improved attitude estimation was obtained using the proposed method (RTK-NE-IG) compared with conventional loosely coupled integration strategies (RTK-NE-VINS).

### 5.2.3 | Analysis of the geometric constraints

To confirm that the geometric improvements were caused mainly caused by the visual measurements, we present ADOP measurements (T. Liu et al., 2021; Teunissen, 1997) which are effective indicators of the geometric distribution. The findings shown in Figure 8 present the geometry using four of the aforementioned methods, including the RTK, RTK-NE, RTK-IG, and RTK-NE-IG. Of note, the geometry of the RTK-NE-VINS is the same as that of RTK-NE. The ADOP values are presented in the top panel of Figure 8. The green curve (RTK-NE) has the largest ADOP value compared to conventional RTK because of the GNSS NLOS exclusion. Based on visual measurements, the ADOP decreases dramatically to about 0.1. Interestingly, the ADOP resulting from RTK-NE-IG increases only slightly compared to RTK-IG even with the GNSS NLOS exclusion. This demonstrates that the visual measurements can effectively guarantee the appropriate geometry even when the GNSS NLOS exclusion has been performed. Moreover, the ADOP value resulting from the RTK-NE-IG is smaller than the RTK even with the GNSS NLOS exclusion.

To determine the number of constraints involved, the number of visual measurements and GNSS measurements are presented (bottom panel of Figure 8). The total of GPS/BeiDou satellites remains below 20 throughout the entire experiment.



**FIGURE 8** Illustration of the geometric analysis of the four methods evaluated in urban canyon 1

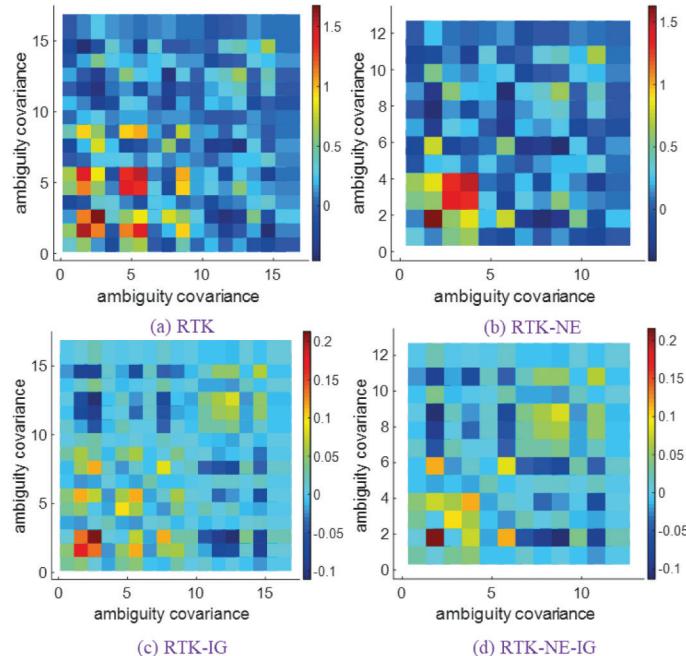
The red, green, magenta, and blue curves denote the RTK, RTK-NE, RTK-IG, and RTK-NE-IG, respectively. The top panel denotes the ADOP during the experiment with the x-axis denoting time and the y-axis denoting the value of ADOP. The bottom panel denotes the number of constraints including the visual and GNSS measurements.

Fortunately, the number of visual measurements exceeds 40 at most of the epochs. Of note, visual measurements are counted if they are tracked at the latest epoch. The statistical number of constraints contributed by visual and GNSS are listed in Table 3. Overall, signals from a mean number of 12 satellites were received during the experiment. After the GNSS NLOS exclusion, the mean number decreased to 10. The mean number of visual constraints was approximately 53 during the experiment in urban canyon 1, leading to a total of more than 60 by combining both the visual and GNSS measurements. However, because the optical flow-based visual feature tracking algorithm is sensitive to noise and illumination conditions, the tracked features may vary in different implementations. That is why the number of visual constraints looks different in the NLOS exclusion case, as shown in Figure 8. However, the change in the GNSS satellite before and after the NLOS exclusion can be still observed by comparing the methods RTK and RTK-NE.

To conduct an explicit comparison of the covariance matrix of the float ambiguity ( $\mathbf{Q}_{nn}$ ) that dominates the success rate of integer ambiguity resolution, each component of the matrix was visualized at an experimentally-selected epoch corresponding to the four methods listed in Table 3 and as shown in Figure 9. The top left figure presents the  $\mathbf{Q}_{nn}$  of the RTK method, with some of the diagonal elements reaching 0.5 with a total of 16 satellites. The deeper color denotes a larger value of the component in the matrix. After GNSS NLOS exclusion in the RTK-NE method (top right), the number of satellites decreases to 12 (the NLOS signals are from satellites 2, 3, 10, and 12). As a result, some of the diagonal components include deeper colors, which indicates larger

**TABLE 3**  
Comparison of Geometric Constraints Using Four Methods

Number of Constraints	RTK	RTK-NE	RTK-IG	RTK-NE-IG
<b>Visual</b>	0	0	53.92	55.16
<b>GNSS</b>	12.64	10.59	12.64	10.59
<b>Total</b>	12.64	10.59	66.56	65.75



**FIGURE 9** Illustration of the covariance matrix ( $\mathbf{Q}_{nn}$ ) of float ambiguities at a selected epoch for the four methods Shown are (a) RTK, (b) RTK-NE, (c) RTK-IG, and (d) RTK-NE-IG.

uncertainty in the corresponding DD ambiguity and thus leads to a lower success rate of integer ambiguity resolution. After the visual measurements are applied, the values of the diagonal components of both RTK-IG and RTK-NE-IG are reduced to below 0.2.

In summary, visual measurements can have a significant impact on improvements in geometry. However, the fixed rate of the proposed method (RTK-NE-IG) remains limited at 1.67%. Apart from the potential contributions of multipath effects as discussed previously, another dominant factor may be the poor quality of the GNSS measurements received by the low-cost commercial GNSS receiver (u-blox M8T) used in this study. The accuracy of this device is far below that of geodetic-level GNSS receivers. Moreover, the visual measurements may also locally constrain the multiple epochs that do not directly provide a globally referenced constraint. However, these findings reveal that at least several healthy and high-quality GNSS measurements are received that increase the fixed rate of GNSS-RTK positioning.

#### 5.2.4 | Analysis of system initialization

The initialization of the extrinsic parameters between the ENU and local world frames is significant for the performance of the state estimation due to the high nonlinearity of the system; this is especially the case for the initialization of the yaw angle offset between both frames. To verify the effectiveness of the proposed initialization method in which Doppler and high-accuracy DD carrier-phase measurements are applied, we compare the following two methods:

- (1) **Weighted least squares (conventional):** Initialize  $T_{WL}^{WG}$  only by aligning two sets of estimated trajectories and the one from VI-initialization described in Section 5A.
- (2) **Doppler/Carrier Smoothing (proposed):** Initialize the  $T_{WL}^{WG}$  only by solving Equation (29) which is free of ambiguity resolution.

Results from this comparison are shown in Table 4.

As shown in Table 4, an error of 15.70 degrees of the yaw offset was obtained using the conventional WLS-based method. Using the proposed method, the error of the yaw offset decreased to 5.93 degrees. Several factors contributed to this improved result compared with the conventional WLS-based method. First, the proposed method makes use of multiple epoch DD carrier-phase measurements to estimate a set of positions. This can significantly increase the smoothness compared with the traditional utilization of the DD carrier phase. Furthermore, the unknown integer ambiguity terms are eliminated via the left null space matrix. The velocity estimated from the Doppler measurement also contributes to the initialization of the yaw offset. The translation part ( $\mathbf{p}_{WL}^{WG}$ ) of  $T_{WL}^{WG}$  was also slightly

**TABLE 4**  
Comparison of Errors/Time in the Initialization of the Extrinsic Parameters

Items	Weighted Least Squares-based	Doppler/Carrier Smoothing
Yaw (degree)	15.70	5.93
East (meter)	1.05	0.92
North (meter)	0.86	0.80
Up (meter)	0.73	0.70
Ave. Time (milliseconds)	3097	4129

Note: Data for the ENU and local world frames were obtained using different methods.

improved with the assistance of the proposed method. Meanwhile, the time utilized for the whole initialization process increased slightly.

To initialize the entire system, one must provide a satisfactory initial guess of the  $\mathbf{T}_{WL}^{WG}$  to the estimator. However, if the measurements begin in a highly dense urbanized area where both the carrier phase and Doppler measurements are noisy, it will be difficult to obtain a good initial guess of  $\mathbf{T}_{WL}^{WG}$  even using the proposed method. One possible solution is to apply an electronic compass to provide the yaw offset between the ENU and local world frames and refine  $\mathbf{T}_{WL}^{WG}$  when the receiver is transported into less congested areas.

### 5.3 | Experimental Evaluation in Urban Canyon 2

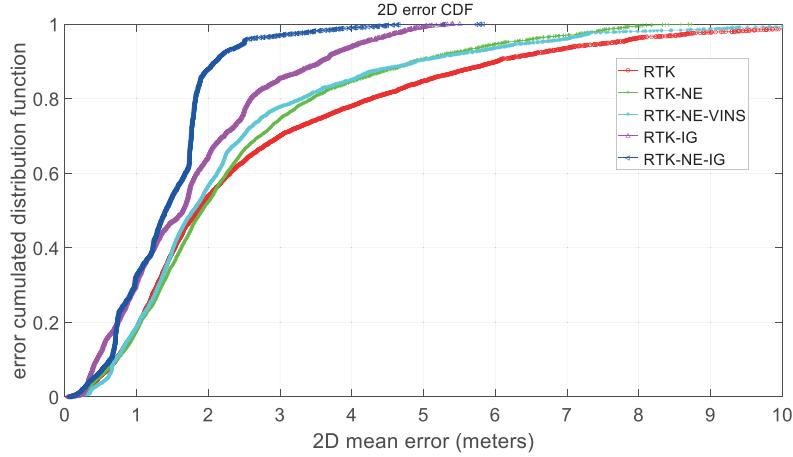
To challenge its effectiveness, the proposed method was verified using a data set collected in a denser urban canyon scenario as shown in Figure 5(b). The positioning performance using this data set is shown in Table 5. The ECDF and trajectory are shown in Figure 10 and Figure 11, respectively. A mean error of 2.68 meters was obtained using the conventional GNSS-RTK positioning method (Herrera et al., 2015) with an STD of 2.23 meters. In this scenario, the mean error was doubled because of the taller building structures in urban canyon 2 compared with the results in urban canyon 1. The 3D error reached 9.94 meters with a maximum error of 34.89 meters. Even after excluding the potential GNSS NLOS receptions, the improvement in the mean error (11.56%) remained limited. Similar improvements were also obtained in 3D positioning. Moreover, the maximum errors decreased in both 2D and 3D positioning because of the exclusion of gross GNSS outlier measurements. However, similar to what was observed for urban canyon 1, the availability in urban canyon 2 decreased from 99.05% to 71.5% due to excessive GNSS NLOS exclusion.

The positioning accuracy was further improved after loosely integrating the GNSS positioning from the RTK-NE with the VINS as originally described by Qin et al. (2019). This resulted in improvements of 13.81% for 2D and 29.53% for 3D positioning, respectively. Moreover, the availability increased to 100% with the assistance of pose estimation from VINS. Interestingly, the 2D maximum error increased from 8.73 meters (RTK-NE) to 12.24 meters. A closer look at these data

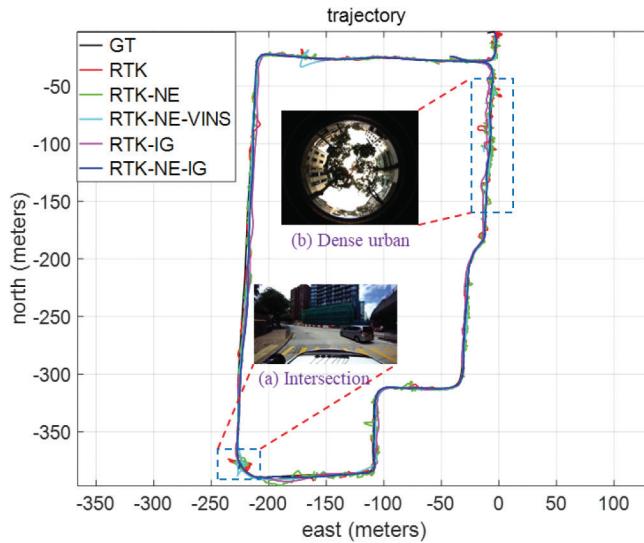
**TABLE 5**  
Positioning Performance of the Five Methods Evaluated in Urban Canyon 2

All Data	RTK	RTK-NE	RTK-NE-VINS	RTK-IG	RTK-NE-IG
<b>2D MEAN</b>	2.68	2.37	2.31	1.77	1.37
<b>2D STD</b>	2.23	1.69	1.86	1.15	0.72
<b>2D Max</b>	11.38	8.73	12.24	5.50	5.83
<b>2D Impr.</b>		11.56%	13.81%	33.95%	48.90%
<b>3D MEAN</b>	6.94	5.70	4.89	2.74	2.64
<b>3D STD</b>	7.32	3.99	2.83	0.96	1.24
<b>3D Max</b>	34.89	30.68	14.56	5.88	6.89
<b>3D Impr.</b>		11.57%	29.53%	60.51%	61.95%
<b>Fixed Rate</b>	0%	0%	0%	0.72%	5.19%
<b>Avail.</b>	99.05%	71.5%	100%	100%	100%
<b>Time Consu. (ms)</b>	4.9	4.9	49.2	66.2	65.1

Note: Improvement is calculated with respect to the RTK method; (2D/3D, horizontal/3D positioning; Time Consu., the average time consumption of the method.



**FIGURE 10** The error cumulative distribution function (ECDF) for 2D positioning errors of the five methods evaluated in urban canyon 2. The red, green, cyan, magenta, and blue denote RTK, RTK-NE, RTK-NE-VINS, RTK-IG, and RTK-NE-IG, respectively.



**FIGURE 11** 2D positioning trajectories of the five methods evaluated in urban canyon 2. The red, green, cyan, magenta, and blue curves denote the RTK, RTK-NE, RTK-NE-VINS, RTK-IG, and RTK-NE-IG, respectively. The black curve denotes the ground truth trajectory.

is presented in Figure 11(a) and the case of a vehicle that has stopped and remains in one position at an intersection. The pose estimation from the VINS drifts gradually because of the random walk of the IMU and the impact of the surrounding dynamic objects on the VINS. As a result, the RTK-NE-VINS drifted significantly at the road intersection. Fortunately, this problem is solved with tightly coupled integration of visual measurements using RTK-IG. In this case, the maximum error decreased to 5.50 for 2D positioning. More importantly, the 2D mean error decreased to 1.77 meters with an improvement of 33.95%. However, an improvement of 60.51% was obtained for 3D positioning. Furthermore, a fixed rate of 0.72% was obtained even in this heavily urbanized scenario and the availability was also guaranteed. The 2D mean error decreased to 1.37 meters with an improvement of 48.90% after applying the proposed RTK-NE-IG method. Collectively, these findings highlight the effectiveness of the proposed method. Meanwhile, the STD also decreased to 0.72 meters. Similar improvements (61.95%) were also obtained in the 3D positioning. Moreover, the fixed rate of GNSS-RTK increased to 5.19%, which

further demonstrates the effectiveness of the proposed method in both excluding the GNSS NLOS receptions and improving the geometric constraint.

In summary, the use of the proposed method led to improvements in positioning with a 2D mean error of 1.37 in this heavily-urbanized scenario (as per the sky view image in Figure 11[b]). As discussed in the experimental validation in urban canyon 1, the unconsidered multipath effects are likely to be the major factor limiting the absolute positioning accuracy even when potential GNSS NLOS are excluded.

## 6 | CONCLUSIONS

The integration of VINS and GNSS-RTK is a promising potential solution that may provide accurate, cost-effective, and drift-free positioning services for autonomous systems with specific navigation requirements. Unfortunately, the performance of the GNSS-RTK/VINS is significantly challenged in urban canyons due to the poor quality of GNSS measurements and satellite geometric distributions caused by signal blockage and reflections from surrounding buildings. In this paper, a 3D vision-aided method was proposed as a means to improve the GNSS-RTK positioning by detecting the potential reflected outlier GNSS signals and improving the geometric distribution of the satellites via low-lying visual landmarks. The effectiveness of the proposed method was verified through several challenging data sets collected in urban canyons of Hong Kong using low-cost automobile-level GNSS receivers together with an automobile visual/inertial sensor suite.

There are several directions for future work on this subject. Because visual measurements can be tracked repetitively, one solution might be to employ already-explored visual landmarks in order to constrain the system once their positions have been correctly estimated, for example, by fixed GNSS-RTK positioning in a previous epoch. In this way, the visual landmark can be applied as an additional auxiliary GNSS satellite that might be used to improve GNSS-RTK positioning in urban canyons. This will be the focus of some of our future work. The multipath effect is another major factor limiting the overall performance of the proposed method (for example, the 2D mean error of 0.84 identified in urban canyon 1). The modeling of multipath effects will also be considered in future investigations.

We note that vegetation (e.g., trees) may not directly block GNSS signal transmission but may lead to signal diffraction (Marais et al., 2020). We are also interested in determining how the GNSS transmits signals in areas of significant vegetation. Moreover, as the visual measurements introduce significant measurement redundancy, will also be important to determine how to estimate bias in the multipath-affected GNSS signals through the sparse estimation technique (Lesouple et al. 2018).

## ACKNOWLEDGMENTS

This research is supported by the University Grants Committee of Hong Kong under the scheme Research Impact Fund on project R5009-21 “Reliable Multiagent Collaborative Global Navigation Satellite System Positioning for Intelligent Transportation Systems.” The authors appreciate the kind comments and suggestions from Dr. Jinxu Liu from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences and Dr. Di Wang from Lab of Visual Cognitive Computing and Intelligent Vehicle, Xi'an Jiaotong University, Xi'an, China. The authors also would like to thank members of the Intelligent Positioning and Navigation Laboratory at The Hong Kong Polytechnic University for their help and suggestions on algorithm development, data collection, and synchronization.

## REFERENCES

- Bai, X., Wen, W., & Hsu, L.-T. (2020a). Robust visual-inertial integrated navigation system aided by online sensor model adaption for autonomous ground vehicles in urban areas. *Remote Sensing*, 12(10), 1686. <https://doi.org/10.3390/rs12101686>
- Bai, X., Wen, W., & Hsu, L. T. (2020b). Using sky-pointing fish-eye camera and Lidar to aid GNSS single-point positioning in urban canyons. *IET Intelligent Transport Systems*, 14(8), 908–914. <http://dx.doi.org/10.1049/iet-its.2019.0587>
- Bai, X., Wen, W., & Hsu, L.-T. (2022). Time-correlated window-carrier-phase-aided GNSS positioning using factor graph optimization for urban positioning. *IEEE Transactions on Aerospace and Electronic Systems*, 58(4), 3370–3384. <https://doi.org/10.1109/TAES.2022.3149730>
- Bai, X., Zhang, B., Wen, W., Hsu, L.-T., & Li, H. (2020c). Perception-aided visual-inertial integrated positioning in dynamic urban areas. *Proc. of the IEEE/ION Position, Location and Navigation Symposium (PLANS 2020)*. Portland, OR. 1563–1571. <https://doi.org/10.1109/PLANS46316.2020.9109963>
- Bloesch, M., Omari, S., Hutter, M., & Siegwart, R. (2015). Robust visual inertial odometry using a direct EKF-based approach. *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015)*, Hamburg, Germany. 298–304. <https://doi.org/10.1109/IROS.2015.7353389>
- Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J. J. (2016). Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6), 1309–1332. <https://doi.org/10.1109/TRO.2016.2624754>
- Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). Orb-SLAM3: An accurate open-source library for visual, visual-inertial, and multimodal SLAM. *IEEE Transactions on Robotics*, 37(6), 1874–1890. <https://doi.org/10.1109/TRO.2021.3075644>
- Cao, S., Lu, X., & Shen, S. (2022). GVINS: Tightly coupled GNSS-visual-inertial fusion for smooth and consistent state estimation. *IEEE Transactions on Robotics*, 38(4), 2004–2021. <https://doi.org/10.1109/TRO.2021.3133730>
- Ch'ng, S.-F., Khosravian, A., Doan, A.-D., & Chin, T.-J. (2019). Outlier-robust manifold pre-integration for INS/GPS fusion. *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019)*, Macau, China. 7489–7496. <https://doi.org/10.1109/IROS40897.2019.8967643>
- Cho, A., Kim, J., Lee, S., & Kee, C. (2011). Wind estimation and airspeed calibration using a UAV with a single-antenna GPS receiver and pitot tube. *IEEE Transactions on Aerospace and Electronic Systems*, 47(1), 109–117. <https://doi.org/10.1109/TAES.2011.5705663>
- Counselman, C. C., & Gourevitch, S. A. (1981). Miniature interferometer terminals for earth surveying: ambiguity and multipath with global positioning system. *IEEE Transactions on Geoscience and Remote Sensing GE-19(4)*, 244–252. <https://doi.org/10.1109/TGRS.1981.350379>
- Enge, P. K. (1994). The global positioning system: Signals, measurements, and performance. *International Journal of Wireless Information Networks*, 1(2), 83–105. <https://doi.org/10.1007/BF02106512>
- Fan, P., Li, W., Cui, X., & Lu, M. (2019). Precise and robust RTK-GNSS positioning in urban environments with dual-antenna configuration. *Sensors*, 19(16), 3586. <https://doi.org/10.3390/s19163586>
- Forster, C., Carlone, L., Dellaert, F., & Scaramuzza, D. (2016). On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1), 1–21. <https://doi.org/10.1109/TRO.2016.2597321>
- Furukawa, R., Kubo, N., & El-Mowafy, A. (2020). Prediction of RTK-GNSS performance in urban environments using a 3D model and continuous LoS method. *Proc. of the International Technical Meeting of the Institute of Navigation, (ION GNSS+ 2020)*, San Diego, CA. 763–771. <https://doi.org/10.33012/2020.17176>
- Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11), 1231–1237. <https://doi.org/10.1177/0278364913491297>
- Geneva, P., Eckenhoff, K., Lee, W., Yang, Y., & Huang, G. (2020). Openvins: A research platform for visual-inertial estimation. *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2020)*, Paris, France. 4666–4672. <https://doi.org/10.1109/ICRA40945.2020.9196524>
- Gong, Z., Liu, P., Wen, F., Ying, R., Ji, X., Miao, R., & Xue, W. (2020). Graph-based adaptive fusion of GNSS and VIO under intermittent GNSS-degraded environment. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–16. <https://doi.org/10.1109/TIM.2020.3039640>
- Grupp, M. (2017). evo: Python package for the evaluation of odometry and slam. <https://github.com/MichaelGrupp/evo>
- He, H., Li, J., Yang, Y., Xu, J., Guo, H., & Wang, A. (2014). Performance assessment of single-and dual-frequency BeiDou/GPS single-epoch kinematic positioning. *GPS Solutions*, 18, 393–403. <https://doi.org/10.1007/s10291-013-0339-3>

- Herrera, A. M., Suhandri, H. F., Realini, E., Reguzzoni, M., & de Lacy, M. C. (2015). goGPS: open-source MATLAB software. *GPS Solutions*, 20(3), 595–603. <https://doi.org/10.1007/s10291-015-0469-x>
- Herrera, A. M., Suhandri, H. F., Realini, E., Reguzzoni, M., & de Lacy, M. C. (2016). goGPS: open-source MATLAB software. *GPS Solutions*, 20, 595–603.
- Hsu, L.-T., Kubo, N., Wen, W., Chen, W., Liu, Z., Suzuki, T., & Meguro, J. (2021). UrbanNav: An open-sourced multisensory dataset for benchmarking positioning algorithms designed for urban areas. *Proc. of the 34th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2021)*, St. Louis, MO.226–256. <https://doi.org/10.33012/2021.17895>
- Hsu, L.-T., Tokura, H., Kubo, N., Gu, Y., & Kamijo, S. (2017). Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons. *IEEE Sensors Journal*, 17(6), 1909–1917. <https://doi.org/10.1109/JSEN.2017.2654359>
- Kennedy, S., Hamilton, J., & Martell, H. (2006). Architecture and system performance of SPAN-NovAtel's GPS/INS solution. *Proc. of the IEEE/ION Position, Location, And Navigation Symposium (PLANS 2006)*, Coronado, CA. 216. <https://doi.org/10.1109/PLANS.2006.1650612>
- Lesouple, J., Robert, T., Sahmoudi, M., Tourneret, J.-Y., & Vigneau, W. (2018). Multipath mitigation for GNSS positioning in an urban environment using sparse estimation. *IEEE Transactions on Intelligent Transportation Systems*, 20(4), 1316–1328. <https://doi.org/10.1109/TITS.2018.2848461>
- Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., & Furgale, P. (2015). Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3), 314–334. <https://doi.org/10.1177/0278364914554813>
- Li, K., Li, M., & Hanebeck, U. D. (2021). Towards high-performance solid-state-Lidar-inertial odometry and mapping. *IEEE Robotics and Automation Letters*, 6(3), 5167–5174. <https://doi.org/10.1109/LRA.2021.3070251>
- Li, M., & Mourikis, A. I. (2013). High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research*, 32(6), 690–711. <https://doi.org/10.1177/0278364913481251>
- Li, T., Zhang, H., Gao, Z., Niu, X., & El-Sheimy, N. (2019). Tight fusion of a monocular camera, MEMS-IMU, and single-frequency multi-GNSS RTK for precise navigation in GNSS-challenged environments. *Remote Sensing*, 11(6), 610. <https://doi.org/10.3390/rs11060610>
- Li, T., Zhang, H., Niu, X., & Gao, Z. (2017). Tightly-coupled integration of multi-GNSS single-frequency RTK and MEMS-IMU for enhanced positioning performance. *Sensors*, 17(11), 2462. <https://doi.org/10.3390/s17112462>
- Li, X., Wang, X., Liao, J., Li, X., Li, S., & Lyu, H. (2021). Semi-tightly coupled integration of multi-GNSS PPP and S-VINS for precise positioning in GNSS-challenged environments. *Satellite Navigation*, 2(1), 1–14. <https://doi.org/10.1186/s43020-020-00033-9>
- Liu, J., Gao, W., & Hu, Z. (2021). Optimization-based visual-inertial SLAM tightly coupled with raw GNSS measurements. *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2021)*, Xi'an, China. 11612–11618. <https://doi.org/10.1109/ICRA48506.2021.9562013>
- Liu, T., hai Liao, Q., Gan, L., Ma, F., Cheng, J., Xie, X., Wang, Z., Chen Y., Zhu, Y., Zhang, S., Chen, Z., Liu, Y., Xie, M., Y, Y., Guo, Z., Li, G., Yuan, P., Han, D., Chen, Y., ... & Liu, M. (2021). The role of the Hercules autonomous vehicle during the COVID-19 pandemic: An autonomous logistic vehicle for contactless goods transportation. *IEEE Robotics & Automation Magazine*, 28(1), 48–58. <https://doi.org/10.1109/MRA.2020.3045040>
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proc. of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81)*, Vancouver, Canada. 674–679. <https://hal.science/hal-03697340/document>
- Marais, J., Ambellouis, S., Meurie, C., Moreau, J., Flancquart, A., & Ruichek, Y. (2015). Image processing for a more accurate GNSS-based positioning in urban environment. *Proc. of the 22nd ITS World Congress (2015)*. Bordeaux, France. <http://dx.doi.org/10.13140/RG.2.2.10493.97760>
- Marais, J., Kazim, S. A., Cocheril, Y., & Meurie, C. (2020). Multipath and NLOS detection based on the combination of CN0 values and a fish-eye camera. *Proc. of the European Navigation Conference (ENC 2020)*, Dresden, Germany. 1–13. <https://doi.org/10.23919/ENC48637.2020.9317408>
- Marais, J., Meurie, C., Attia, D., Ruichek, Y., & Flancquart, A. (2014). Toward accurate localization in guided transport: Combining GNSS data and imaging information. *Transportation Research Part C: Emerging Technologies*, 43, 188–197. <https://doi.org/10.1016/j.trc.2013.11.008>
- Ng, H.-F., & Hsu, L.-T. (2021). 3D mapping database-aided GNSS RTK and its assessments in urban canyons. *IEEE Transactions on Aerospace and Electronic Systems*, 57(5), 3150–3166. <https://doi.org/10.1109/TAES.2021.3069271>
- Niu, Z., Guo, F., Shuai, Q., Li, G., & Zhu, B. (2021). The integration of GPS/BDS real-time kinematic positioning and visual-inertial odometry based on smartphones. *ISPRS International Journal of Geo-Information*, 10(10), 699. <https://doi.org/10.3390/ijgi10100699>

- Qin, C., Ye, H., Pranata, C. E., Han, J., Zhang, S., & Liu, M. (2020). Lins: A Lidar-inertial state estimator for robust and efficient navigation. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2020)*. <https://doi.org/10.48550/arXiv.1907.02233>
- Qin, T., Cao, S., Pan, J., & Shen, S. (2019). A general optimization-based framework for global pose estimation with multiple sensors. Pre-print; *arXiv:1901.03642* <https://doi.org/10.48550/arXiv.1901.03642>
- Qin, T., Li, P., & Shen, S. (2018). Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020. <https://doi.org/10.1109/TRO.2018.2853729>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Proc. (Part III) of the 18th International Conference Medical Image Computing and Computer-Assisted Intervention(MICCAI 2015)*, Munich, Germany. 18 <https://doi.org/10.48550/arXiv.1505.04597>
- Rycroft, M. J. (1997). Understanding GPS. Principles and applications. *Journal of Atmospheric and Solar-Terrestrial Physics*, 5(59), 598–599. [https://doi.org/10.1016/s1364-6826\(97\)83337-8](https://doi.org/10.1016/s1364-6826(97)83337-8)
- Shan, T., & Englot, B. (2018). Lego-loam: Lightweight and ground-optimized LiDAR odometry and mapping on variable terrain. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain (pp. 4758–4765). <https://doi.org/10.1109/IROS.2018.8594299>
- Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., & Rus, D. (2020). Lio-sam: Tightly-coupled LiDAR inertial odometry via smoothing and mapping. Proc.of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020), Las Vegas, NV. <https://doi.org/10.48550/arXiv.2007.00258>
- Shi, J. (1994). Good features to track. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* (1994), Seattle, WA. 593–600. <https://doi.org/10.1109/CVPR.1994.323794>
- Sibley, G., Matthies, L., & Sukhatme, G. (2010). Sliding window filter with application to planetary landing. *Journal of Field Robotics*, 27(5), 587–608. <https://doi.org/10.1002/rob.20360>
- Suzuki, T., & Kubo, N. (2013). Correcting GNSS multipath errors using a 3D surface model and particle filter. *Proc. of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*, Nashville, TN. 1583–1595.
- Takasu, T., & Yasuda, A. (2009). Development of the low-cost RTK-GPS receiver with an open-source program package RTKLIB. Proc. of the International Symposium on GPS/GNSS (2009), Seogwipo-si, Korea.
- Teunissen, P. (1997). A canonical theory for short GPS baselines. Part IV: Precision versus reliability. *Journal of Geodesy*, 71(9), 513–525. <https://doi.org/10.1007/s001900050119>
- Teunissen, P. (2000). The GPS integer least-squares statistics. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 25(9–11), 673–677. [https://doi.org/10.1016/S1464-1895\(00\)00104-6](https://doi.org/10.1016/S1464-1895(00)00104-6)
- Teunissen, P. (2001). Integer estimation in the presence of biases. *Journal of Geodesy*, 75, 399–407. <https://doi.org/10.1007/s001900100191>
- Teunissen, P. (2003). Theory of integer equivariant estimation with application to GNSS. *Journal of Geodesy*, 77(7–8), 402–410. <https://doi.org/10.1007/s00190-003-0344-3>
- Wang, H., Wang, C., & Xie, L. (2021). Intensity-slam: Intensity assisted localization and mapping for large scale environment. *IEEE Robotics and Automation Letters*, 6(2), 1715–1721. <https://doi.org/10.1109/LRA.2021.3059567>
- Wang, K., Teunissen, P. J., & El-Mowafy, A. (2020). The ADOP and PDOP: two complementary diagnostics for GNSS positioning. *Journal of Surveying Engineering*, 146(2), 04020008. [https://doi.org/10.1061/\(ASCE\)SU.1943-5428.0000313](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000313)
- Wen, W. (2020). 3D LiDAR aided GNSS and its tightly coupled integration with INS via factor graph optimization. In *Proc. of the 33rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2020)*, St. Louis, MO. 1649–1672 <https://doi.org/10.33012/2020.17557>
- Wen, W., Bai, X., Kan, Y.C., & Hsu, L.-T. (2019c). Tightly-coupled GNSS/INS integration via factor graph and aided by fish-eye camera. *IEEE Transactions on Vehicular Technology*, 68(11), 10651–10662. <https://doi.org/10.1109/TVT.2019.2944680>
- Wen, W., & Hsu, L.-T. (2021). Towards robust GNSS positioning and real-time kinematic using factor graph optimization. *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2021)*, Xi'an, China. 5884–5890. <https://doi.org/10.1109/ICRA48506.2021.9562037>
- Wen, W., Zhang, G., & Hsu, L.-T. (201). Exclusion of GNSS NLOS receptions caused by dynamic objects in heavy traffic urban scenarios using real-time 3D point cloud: An approach without 3D maps. *Proc. of the IEEE/ION Position, Location and Navigation Symposium (PLANS 2018)*, Monterey, CA, 158–165. <https://doi.org/10.1109/PLANS.2018.8373377>
- Wen, W., Zhang, G., & Hsu, L.-T. (2019a). GNSS NLOS exclusion based on dynamic object detection using LiDAR point cloud. *IEEE Transactions on Intelligent Transportation Systems*, 22(2), 853–862. <https://doi.org/10.1109/TITS.2019.2961128>
- Wen, W., Zhang, G., & Hsu, L. T. (2019b). Correcting NLOS by 3D Lidar and building height to improve GNSS single point positioning. *NAVIGATION*, 66(4), 705–718 <https://doi.org/10.1002/navi.335>

- Wen, W., Zhou, Y., Zhang, G., Fahandezh-Saadi, S., Bai, X., Zhan, W., Tomizuka, M., and Hsu, L. T. (2020). Urbanloco: A full sensor suite dataset for mapping and localization in urban scenes. *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2020)*, Paris, France. 2310–2316. <http://dx.doi.org/10.1109/ICRA40945.2020.9196526>
- Ye, H., Chen, Y., & Liu, M. (2019). Tightly coupled 3D LiDAR inertial odometry and mapping. *Proc. of the International Conference on Robotics and Automation (ICRA 2019)*, Montreal, QC, Canada. pp. 3144–3150. <https://doi.org/10.1109/ICRA.2019.8793511>
- Zampieri, G., Narayanan, S., Crespillo, O. G., & Osechas, O. (2020). A regularized least squares estimator for pseudorange-based terrestrial positioning under degraded geometries. *Proc. of the 33rd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2020)*, St. Louis, MO. 698–707 <https://doi.org/10.33012/2020.17690>
- Zhang, J., Liu, W., & Wu, Y. (2011). Novel technique for vision-based UAV navigation. *IEEE Transactions on Aerospace and Electronic Systems*, 47(4), 2731–2741. <https://doi.org/10.1109/TAES.2011.6034661>
- Zhang, J., & Singh, S. (2017). Low-drift and real-time Lidar odometry and mapping. *Autonomous Robots*, 41(2), 401–416. <https://link.springer.com/article/10.1007/s10514-016-9548-2>

**How to cite this article:** Wen, W., Bai, X., & Hsu, L. (2023). 3D Vision aided GNSS real-time kinematic positioning for autonomous systems in urban canyons. *NAVIGATION*, 70(3). <https://doi.org/10.33012/navi.590>

## 7 | APPENDIX

### 7.2 | Jacobians of Residuals

We adopted the same Jacobian matrix derivation for the IMU pre-integration and the visual reprojection residual as described in Qin et al. (2018). Therefore, we mainly derive the Jacobians related to the DD pseudorange and carrier-phase residuals together with the Doppler measurement residuals as described in this section.

#### 7.2.3 | Jacobian of DD pseudorange residual

The DD pseudorange residual correlates with the position of the system at two key-frames ( $\mathbf{p}_{b_k}^{WL}$  and  $\mathbf{p}_{b_{k+1}}^{WL}$ ) as well as with the extrinsic parameter of  $\mathbf{T}_{WL}^{WG}$ . The Jacobian of  $\mathbf{p}_{b_k}^{WL}$  with respect to the DD pseudorange residuals can be derived as follows from Equation (A-1):

$$\frac{\partial \mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{p}_{b_k}^{WL}} = \left( -\sigma_{\rho, r_t}^s \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) - \left( -\sigma_{\rho, r_t}^s \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-1})$$

where the  $\mathbf{e}_{r_t}^w$  denotes the LOS vector connecting the GNSS receiver and the master satellite  $w$ . Similarly, the Jacobian of  $\mathbf{p}_{b_{k+1}}^{WL}$  concerning the DD pseudorange residual can be derived from Equation (A-2) as follows:

$$\frac{\partial \mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{p}_{b_{k+1}}^{WL}} = \left( -\sigma_{\rho, r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) - \left( -\sigma_{\rho, r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-2})$$

Because the  $\mathbf{R}_{WL}^{WG}$  involves only the yaw angle which is denoted as the Euler angle  $\varepsilon_{WL,z}^{WG}$ , the Jacobian of  $\mathbf{R}_{WL}^{WG}$  with respect to the DD pseudorange residual can be derived from Equation (A-3):

$$\frac{\partial \mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{e}_{WL,z}^{WG}} = \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_z \mathbf{p}_{b_k}^{WL} \right) - \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \mathbf{R}_z \mathbf{p}_{b_k}^{WL} \right) \quad (\text{A-3})$$

with  $\mathbf{R}_z = \begin{bmatrix} -\sin(\varepsilon_{WL,z}^{WG}) & -\cos(\varepsilon_{WL,z}^{WG}) & 0 \\ \cos(\varepsilon_{WL,z}^{WG}) & -\sin(\varepsilon_{WL,z}^{WG}) & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The Jacobian of  $\mathbf{p}_{WL}^{WG}$  concerning the DD pseudorange residual can be derived from Equation (A-4) as follows:

$$\frac{\partial \mathbf{r}_\rho(\rho_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{p}_{WL}^{WG}} = \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \right) - \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \right) \quad (\text{A-4})$$

### 7.2.4 | Jacobian of DD carrier-phase residual

The DD carrier-phase residual correlates with the position of the system at two keyframes ( $\mathbf{p}_{b_k}^{WL}$  and  $\mathbf{p}_{b_{k+1}}^{WL}$ ) as well as the extrinsic parameter of  $\mathbf{T}_{WL}^{WG}$  and the integer ambiguity  $N_{DD,r_t}^s$ . The Jacobian of  $\mathbf{p}_{b_k}^{WL}$  for the DD carrier-phase residual can be derived from Equation (A-5) as follows:

$$\frac{\partial \mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{p}_{b_k}^{WL}} = \left( -\sigma_{\rho,r_t}^s \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) - \left( -\sigma_{\rho,r_t}^s \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-5})$$

where the  $\mathbf{e}_{r_t}^w$  denotes the LOS vector connecting the GNSS receiver and the master satellite  $w$ . Similarly, the Jacobian of  $\mathbf{p}_{b_{k+1}}^{WL}$  concerning the DD carrier-phase residual can be derived from Equation (A-6) as follows:

$$\frac{\partial \mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{p}_{b_{k+1}}^{WL}} = \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) - \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-6})$$

Similar to the DD pseudorange residuals, the Jacobian of  $\mathbf{R}_{WL}^{WG}$  for the DD carrier-phase residual can be derived from Equation (A-7) as follows:

$$\frac{\partial \mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{e}_{WL,z}^{WG}} = \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_z \mathbf{p}_{b_k}^{WL} \right) - \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \mathbf{R}_z \mathbf{p}_{b_k}^{WL} \right) \quad (\text{A-7})$$

With  $\mathbf{R}_z = \begin{bmatrix} -\sin(\varepsilon_{WL,z}^{WG}) & -\cos(\varepsilon_{WL,z}^{WG}) & 0 \\ \cos(\varepsilon_{WL,z}^{WG}) & -\sin(\varepsilon_{WL,z}^{WG}) & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The Jacobian of  $\mathbf{p}_{WL}^{WG}$  for the DD carrier-phase residual can be derived from Equation (A-8) as follows:

$$\frac{\partial \mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial \mathbf{p}_{WL}^{WG}} = \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \right) - \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^w \mathbf{R}_{WG}^{WE} \right) \quad (\text{A-8})$$

Different from the DD pseudorange residual, The integer ambiguity differs from the DD pseudorange residual as it includes an additional term to be estimated. The Jacobian is derived from Equation (A-9) as follows:

$$\frac{\partial \mathbf{r}_\psi(\psi_{DD,r_t}^s, \mathbf{p}_{r_t,b}^{WL})}{\partial N_{DD,r_t}^s} = \sigma_{\rho,r_t}^s \quad (\text{A-9})$$

### 7.2.5 | Jacobian of the Doppler measurement residual

The Doppler measurement residual correlates with the position and velocity of the system in two keyframes ( $\mathbf{p}_{b_k}^{WL}, \mathbf{v}_{b_k}^{WL}$  and  $\mathbf{p}_{b_{k+1}}^{WL}, \mathbf{v}_{b_{k+1}}^{WL}$ ) as well as with the extrinsic parameter of  $\mathbf{R}_{WL}^{WG}$ . The Jacobian of  $\mathbf{p}_{b_k}^{WL}$  for the Doppler measurement residual can be derived from Equation (A-10) as follows:

$$\frac{\partial r_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL})}{\partial \mathbf{p}_{b_k}^{WL}} = \left( \sigma_{d,r_t}^s \frac{(r_t - t_k)}{(t_{k+1} - t_k)} (\mathbf{v}_{r_t}^{WE,s} - \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{v}_{r_t}^{WL})^T \mathbf{K} \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-10})$$

$$\text{With } \mathbf{K} = \begin{bmatrix} \frac{\|\mathbf{E}_{r_t}^s\|^2 - 1 * \mathbf{E}_{r_t}^s(0)^2}{\|\mathbf{E}_{r_t}^s\|^3} & \frac{-1 * \mathbf{E}_{r_t}^s(0) * \mathbf{E}_{r_t}^s(1)}{\|\mathbf{E}_{r_t}^s\|^3} & \frac{-1 * \mathbf{E}_{r_t}^s(0) * \mathbf{E}_{r_t}^s(2)}{\|\mathbf{E}_{r_t}^s\|^3} \\ \frac{-1 * \mathbf{E}_{r_t}^s(1) * \mathbf{E}_{r_t}^s(0)}{\|\mathbf{E}_{r_t}^s\|^3} & \frac{\|\mathbf{E}_{r_t}^s\|^2 - 1 * \mathbf{E}_{r_t}^s(1)^2}{\|\mathbf{E}_{r_t}^s\|^3} & \frac{-1 * \mathbf{E}_{r_t}^s(0) * \mathbf{E}_{r_t}^s(2)}{\|\mathbf{E}_{r_t}^s\|^3} \\ \frac{-1 * \mathbf{E}_{r_t}^s(2) * \mathbf{E}_{r_t}^s(0)}{\|\mathbf{E}_{r_t}^s\|^3} & \frac{-1 * \mathbf{E}_{r_t}^s(2) * \mathbf{E}_{r_t}^s(1)}{\|\mathbf{E}_{r_t}^s\|^3} & \frac{\|\mathbf{E}_{r_t}^s\|^2 - 1 * \mathbf{E}_{r_t}^s(2)^2}{\|\mathbf{E}_{r_t}^s\|^3} \end{bmatrix}$$

and  $\mathbf{E}_{r_t}^s = \mathbf{p}_{r_t}^{WE,s} - (\mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{p}_{r_t}^{WL} + \mathbf{R}_{WG}^{WE} \mathbf{p}_{WL}^{WG} + \mathbf{p}_{WG}^{WE})$

where the  $\mathbf{K}$  is a matrix related to the  $\mathbf{E}_{r_t}^s$ .  $\mathbf{E}_{r_t}^s(i)$  is employed to obtain the  $i$ -th element of the vector  $\mathbf{E}_{r_t}^s$ . Similarly, the Jacobian of  $\mathbf{p}_{b_{k+1}}^{WL}$  for the Doppler measurement residual can be derived from Equation (A-11) as follows:

$$\frac{\partial r_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL})}{\partial \mathbf{p}_{b_{k+1}}^{WL}} = \left( \sigma_{d,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) (\mathbf{v}_{r_t}^{WE,s} - \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \mathbf{v}_{r_t}^{WL})^T \mathbf{K} \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-11})$$

The Jacobian of  $\mathbf{v}_{b_k}^{WL}$  for the Doppler measurement residual can be derived from Equation (A-12) as follows:

$$\frac{\partial r_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL})}{\partial \mathbf{v}_{b_k}^{WL}} = \left( -\sigma_{\rho,r_t}^s \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-12})$$

Similarly, the Jacobian of  $\mathbf{v}_{b_{k+1}}^{WL}$  for the Doppler measurement residual can be derived from Equation (A-13) as follows:

$$\frac{\partial r_d(d_{r_t}^s, \mathbf{v}_{r_t,b}^{WL})}{\partial \mathbf{v}_{b_{k+1}}^{WL}} = \left( -\sigma_{\rho,r_t}^s \left( 1 - \frac{(r_t - t_k)}{(t_{k+1} - t_k)} \right) \mathbf{e}_{r_t}^s \mathbf{R}_{WG}^{WE} \mathbf{R}_{WL}^{WG} \right) \quad (\text{A-13})$$

The Jacobians of  $\mathbf{p}_{WL}^{WG}$  and  $\mathbf{R}_{WL}^{WG}$  for the Doppler measurement residual are the same as shown in Equations (A-8) and (A-7), respectively.