Tightly Coupled Integration of MVIO and UC-PPP based on factor graph optimization for complex urban environments

Cheng Pan  
China University of Mining and Technology

Fangchao Li  
Northwest A&F University

Yuanxin Pan  
ETH Zürich

Yonghui Wang  
China University of Mining and Technology

Benedikt Soja  
ETH Zürich

Zengke Li (✉ zengkeli@cumt.edu.cn)  
China University of Mining and Technology

Jingxiang Gao  
China University of Mining and Technology

Research Article

Keywords: monocular visual-inertial odometry, factor graph optimization, uncombined PPP, tightly coupled system

Posted Date: May 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2907885/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Additional Declarations: No competing interests reported.
Tightly Coupled Integration of MVIO and UC-PPP based on factor graph optimization for complex urban environments
Cheng Pan1,2, Fangchao Li3, Yuanxin Pan2, Yonghui Wang1, Benedikt Soja2, Zengke Li1,*, Jingxiang Gao1
1 School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, China
2 Institute of Geodesy and Photogrammetry, ETH Zürich, Zürich, Switzerland
3 College of Forestry, Northwest A&F University, Yangling, China
* Corresponding author. E-mail addresses: zengkeli@cumt.edu.cn

Abstract: The emergence of low-cost micro-electro-mechanical system inertial measurement units, cameras and Global Navigation Satellite System (GNSS) receivers has promoted the research of multi-sensor fusion for positioning. With the rapid development of chip technology, it has become a trend for low-cost GNSS chips to provide multi-frequency carrier phase observations while supporting multiple constellations. To enhance the positioning performance of multi-frequency and multi-system precise point positioning (PPP) in the complex urban environment, this paper proposes a tightly coupled system of monocular visual-inertial odometry (MVIO) and uncombined PPP (UC-PPP) based on factor graph optimization. The initialization of the integrated system includes two parts: one is for monocular vision and inertial navigation systems, and another is for MVIO and UC-PPP. The initialization of MVIO and UC-PPP adopts a coarse-to-fine approach to correct online the transformation of local and global frames. In addition, the sliding window and marginalization methods are adopted to retain the constraints between adjacent observations and eliminate useless observations in the window. The pedestrian and vehicle tests in the complex urban environment verify the performance of the proposed method. Compared with open-source software GVINS, the positioning accuracy of the proposed method has been further improved due to the use of carrier phase observations with higher measurement accuracy. Compared with PPP alone, the improvement of the proposed method for the low-speed and short-distance pedestrian test in the east, north, and up directions is 73.3, 54.8 and 62.7%, respectively, while the improvement for the high-speed and long-distance vehicle test is 63.0, 59.3 and 70.5%, respectively. Experiment results show that the proposed method has better positioning accuracy and performance in complex urban environments.

Keywords: monocular visual-inertial odometry; factor graph optimization; uncombined PPP; tightly coupled system

1 Introduction
In recent years, high-precision navigation and positioning in the complex urban environment have become a general concern (Zhai and Yuan 2022; Li et al. 2023). The rapid urbanization and dense construction have led to many urban canyons, tunnels, overpasses, boulevards, and other environments in highly developed cities. Global Navigation Satellite Systems (GNSS) cannot provide users with accurate, continuous, and reliable absolute positions in complex environments (Li et al. 2022b; Li et al. 2023). In addition, even multi-GNSS cannot independently realize location services in complex environments such as tunnels and boulevards. Precise point positioning (PPP) is a technology that can realize dynamic positioning at the decimeter level with only a single GNSS receiver (Pan et al. 2021), which can fully meet the conventional positioning service requirements of vehicles and pedestrians. Compared with real-time kinematic (RTK) and Standard Point Positioning (SPP), PPP has the advantages of low cost and high precision and has broad application prospects.

To overcome the problem that PPP is challenging in the complex urban environment, scholars try to fuse other sensors with PPP to achieve high-precision positioning. As an independent system that
does not need external signals, the inertial navigation system (INS) can provide a predicted high-precision relative position in a short time so that it is gradually used to make up for the deficiencies of PPP in complex environments (Li et al. 2017). Le and Lorga introduced the concept of loose coupling between GPS PPP and INS, and verified that PPP assisted by INS could achieve more precise positioning through flight experiments (Le et al. 2006). Zhang and Gao implemented an undifferenced GPS/INS tightly coupled system for the first time (Zhang and Gao 2008). The test results show that the proposed system can provide the position and velocity solutions equivalent to the double-differential GPS/INS integrated system. Subsequently, the promotion and application of low-cost micro-electromechnical system inertial measurement unit (MEMS-IMU) further facilitated the research of INS-aided PPP, such as cycle slip detection, ambiguity fixation, precision positioning, and attitude determination.

Furthermore, the gradual maturity of camera hardware technology and image processing technology provides a good opportunity for the development of visual simultaneous localization and mapping (SLAM). In 2007, Klein and Murray proposed the Parallel Tracking and Mapping Framework (PTAM) (Klein and Murray 2007), which laid the foundation for visual SLAM. Whereafter, the well-known open solutions such as MonoSLAM (A. et al. 2007), DTAM (Newcombe et al. 2011), and ORB-SLAM2 (Mur-Artal and Tardós 2017) significantly promoted the research and application of visual SLAM. With the gradual deepening of the research related to pure vision SLAM, the shortcomings of cameras have also been a concern. Pure vision has poor dynamic adaptability under weak light conditions, and inertial guidance is unaffected by moving scenes and speed. Gradually, visual-inertial odometry (VIO), which complements the advantages of vision and INS, has become a topic of great interest. Bloesch et al. proposed a robust vision-inertial odometer (ROVIO) based on an extended Kalman filter (Bloesch et al. 2015). The results of highly dynamic handheld experiments show that this method has high tracking accuracy and robustness. Leutenegger et al. proposed a nonlinear optimized tightly coupled binocular visual-inertial odometer (OKVIS) based on a sliding window (Leutenegger et al. 2015). Compared with the tightly coupled visual-inertial system based on filtering with a sliding window, which requires more costly computations but can obtain higher accuracy. Based on a multi-state constrained Kalman filter (MSCKF), Ke et al. proposed a binocular vision-inertial odometer (MSCKF-VIO), which is equivalent to the most advanced monocular solution in terms of computational cost and has stronger robustness (Sun et al. 2018). In 2020, Campos et al. proposed the vision-inertial tightly coupled system ORB-SLAM3 for monocular, binocular, and RGB-D cameras, which can provide accurate positioning in the environment of lack of visual information for a long time (Campos et al. 2021). Currently, the mainstream VIO schemes all use a tight-coupling approach to realize the fusion of vision and inertial information. Due to the introduction of inertial information, the problem of scale recovery of monocular SLAM has been solved. Based on monocular VIO (MVIO), the position and attitude with the same accuracy as that of binocular VIO can be obtained. On the premise of considering the equipment cost, MVIO is typically the first choice in the industry.

Since VIO can provide more robust local attitude estimation than INS, GNSS/INS integration is gradually replaced by GNSS/INS/Vision integration. Tong proposed VINS-Fusion, which supports monocular VIO and binocular VIO. At the same time, it also gives an example of the fusion of VINS and GPS original observations (Qin et al. 2019). Its essence is a multi-sensor state estimator based on optimization. Li et al. proposed a semi-tightly coupled scheme of multi-GNSS PPP and stereo VIO based on an optimization method, in which PPP adopts the ionosphere-free (IF) model (Li et al. 2021). Xu et al. investigated the mitigation method of cycle slip and multipath for a MSCKF-based stereo
VIO/PPP semi-tightly coupled system, providing an idea for improving the performance of the VIO/PPP integrated system (Xu et al. 2023). However, these works belong to the loosely coupled scheme and its variants. To study the performance of the GNSS/VIO tightly coupled system, Lee et al. proposed an intermittent GPS-assisted VIO scheme to achieve online initialization and correction (Lee et al. 2020), which is based on MSCKF. Li et al. proposed a tightly coupled system of a monocular camera, MEMS-IMU, and single-frequency RTK based on an extended Kalman filter (Li et al. 2019). The integration of RTK technology and VIO has been further studied and applied. Subsequently, GVINS realized the tightly coupled integration of multi-system SPP and VIO based on the optimization method (Cao et al. 2022). The VIO-assisted SPP has good robustness even in complex environments. However, because carrier phase observations with higher measurement accuracy are not used, the accuracy of the position provided by the fusion system needs to be improved. Li et al. established a PPP/INS/visual tightly coupled system based on GVINS (Li et al. 2022a). The carrier phase observation improved the positioning performance of the integrated system, but this method kept the IF model and its performance in complex environments needs to be verified.

Nowadays, GNSS, IMU, and vision in the civil field are developing in the direction of low cost. With the progress in chip technology, it has become a trend for low-cost GNSS chips to provide multi-frequency carrier phase observations while supporting multi-GNSS. Compared with the IF model, uncombined PPP (UC-PPP) has unique advantages in multi-frequency data processing. To enhance the positioning performance of UC-PPP in complex environments, we propose a low-cost MVIO/UC-PPP tightly coupled system based on factor graph optimization, which can achieve lane-level positioning accuracy. In this work, the UC-PPP is utilized to process the dual-frequency pseudorange and carrier phase observations, and the factor graph diagram of the proposed method is established. Two open datasets collected in complex urban environments are used to evaluate the performance of the proposed method.

The rest of the article is organized as follows: First, the coordinate frames and observation models involved in the MVIO/UC-PPP tightly coupled system are introduced. Then, the state vector is constructed, and the expression of the nonlinear problem based on factor graph optimization is given. In addition, UC-PPP-related factors are derived, and the workflow of the VIO/UC-PPP tightly coupled system is further revealed. Next, the test equipment and schemes are introduced in detail. Based on two open datasets, the performance of the proposed method in the complex urban environment is tested, and the experimental results are analyzed and discussed. Finally, the main work of this paper is summarized.

2 Observation model of MVIO/UC-PPP integrated system
2.1 Definition of coordinate frame

The absolute position provided by PPP for vehicle navigation and positioning is usually expressed in the Earth-Centered, Earth-Fixed (ECEF) coordinate system (e-frame). To compare the positioning error of different methods relative to the ground truth, we introduce the East-North-Up coordinate system centered on the vehicle, which is referred to as the n-frame. The X, Y and Z axes of the n-frame point to the east, north and up directions along the ellipsoidal normal, respectively. The VIO provides the position and attitude of the vehicle in the local world coordinate system, which is marked as the v-frame. The origin of the v-frame is the same as that of the n-frame. Since VIO can estimate the pitch and roll angles in the n-frame, the v-frame and n-frame only have one degree of freedom (DOF) rotation (Cao et al. 2022). The relationship between the three frames is shown in Fig. 1. Due to the difference in physical dimensions of the camera, IMU and GNSS receiver, the lever arm needs to be
corrected during installation. This correction is usually carried out under the \( n \)-frame. The coordinate system of the camera takes the center of the camera as the origin, which is the \( e \)-frame. When the camera is placed horizontally, the X axis points to the right along the camera’s horizontal axis, the Y axis points to the top along the camera’s vertical axis, and the Z axis is perpendicular to the X-Y plane (Li et al. 2022a). The coordinate system of the INS takes the center of the IMU as the origin, which is abbreviated as the \( b \)-frame. Although the IMU can be installed in any direction, to facilitate deviation compensation, the Y direction of the \( b \)-frame is consistent with the forward direction, and the Z direction is aligned with the gravity of the earth. We take the IMU frame as the target framework for state estimation.

![Fig. 1 Schematic diagram of coordinate systems related to GNSS, camera and IMU](image)

### 2.2 Multi-GNSS UC-PPP observation model

Compared with a single constellation, multi-GNSS has more visible satellites in the complex urban environment, so it has better navigation performance. Uncombined pseudorange \( P_{s,j}^{\text{GNSS}} \) and carrier phase \( L_{s,j}^{\text{GNSS}} \) can be expressed as follows:

\[
P_{s,j}^{\text{GNSS}} = \rho + c \left( \delta t_r^{\text{GNSS}} - \delta t_s^{\text{GNSS}} \right) + M \cdot \text{ZTD} + \gamma_{s,j}^{\text{GNSS}} \cdot \tilde{T}_{s,j}^{\text{GNSS}} + \varepsilon_{s,j}^{\text{GNSS}} \tag{1}
\]

\[
L_{s,j}^{\text{GNSS}} = \rho + c \left( \delta t_r^{\text{GNSS}} - \delta t_s^{\text{GNSS}} \right) + M \cdot \text{ZTD} - \gamma_{s,j}^{\text{GNSS}} \cdot \tilde{T}_{s,j}^{\text{GNSS}} + \lambda_{s,j}^{\text{GNSS}} N_{s,j}^{\text{GNSS}} + \varphi_{s,j}^{\text{GNSS}} \tag{2}
\]

where GNSS can be GPS (G), GLONASS (R), Galileo (E) and BDS (C) respectively. \( s \) and \( r \) represent the satellite and receiver ID. \( j \) represents the frequency of GNSS observations. \( \rho \) indicates the geometric distance from the satellite phase center to the receiver antenna. \( c \) indicates the speed of satellite signal transmission, namely the speed of light. \( \delta t_r^{\text{GNSS}} \) and \( \delta t_s^{\text{GNSS}} \) denote receiver and satellite clock compensation, respectively. \( M \) and \( \text{ZTD} \) represent the projection function and zenith delay of the tropospheric delay separately. \( \gamma_{s,j}^{\text{GNSS}} = f_i^2 / f_j^2 \) is the ionospheric delay amplification factor and \( f_j \) is the \( j \)-th frequency. \( \tilde{T}_{s,j}^{\text{GNSS}} \) indicates the slant ionospheric delay of the first frequency. \( \lambda_{s,j}^{\text{GNSS}} \) and \( N_{s,j}^{\text{GNSS}} \) are the wavelength and float ambiguity of carrier phase observation, respectively. \( \varepsilon_{s,j}^{\text{GNSS}} \) and \( \varphi_{s,j}^{\text{GNSS}} \) represent the sum of noise and multipath error of pseudorange and carrier phase, respectively. In addition, due to the
different signal systems and hardware delays of various constellations, multi-GNSS PPP needs to consider the inter-system bias (ISB). Based on the GPS receiver clock, the following constraints are adopted to estimate the ISBs from other constellations:

\[
\begin{bmatrix}
ISB_{G,R} \\
ISB_{G,E} \\
ISB_{G,C}
\end{bmatrix} = 
\begin{bmatrix}
-1 & 1 & 0 \\
-1 & 0 & 1 \\
-1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\delta t_G^r \\
\delta t_E^r \\
\delta t_C^r
\end{bmatrix}
\]

(3)

where, \( ISB_{G,R} \), \( ISB_{G,E} \) and \( ISB_{G,C} \) represent the ISBs of GLONASS, Galileo and BDS with GPS, respectively.

Compared with the SPP algorithm, UC-PPP needs to estimate more parameters. SPP only needs four visible satellites to calculate the 3-DoF global position of the GNSS receiver. However, UC-PPP needs to estimate unknown parameters such as receiver clock error, tropospheric delay, ionospheric delay and carrier phase ambiguity at different frequencies. More visible satellites are required to establish observation equations. Take multi-GNSS and dual-frequency UC-PPP as an example. If the current number of visible satellites and constellations are \( n \) and \( N \), respectively, the number of parameters to be estimated will be \( 4 + N + 3n \). Since the number of observation equations is \( 4n \), the receiver position can be calculated only when \( n \geq 4 + N \). This condition is very challenging in a complex urban environment, so enhancing the PPP positioning performance with other sensors is desired.

2.3 INS observation model

Generally, the INS consists of accelerometers and gyroscopes. The raw observations that can be provided by INS are acceleration and angular velocity, which can be defined as (Li et al. 2023):

\[
\begin{align*}
\ddot{a}_i &= a_i + b_{a_i} + R_{b}^a g^n + n_a \\
\ddot{w}_i &= w_i + b_{w_i} + n_w
\end{align*}
\]

(4)

where, \( \ddot{a}_i \) and \( a_i \) represent the raw observation and its linearization of the accelerometer, respectively. \( \ddot{w}_i \) and \( w_i \) represent the raw observation and its linearization of the gyroscope, respectively. Since land vehicles loaded with INS will be affected by the gravity of the earth, the raw observation of the accelerometer needs to consider eliminating the gravity acceleration deviation. \( R_{b}^a \) represents the rotation matrix from the \( n \)-frame to the \( b \)-frame at the current time. We assume that the noise of the IMU observation \( n_a \) and \( n_w \) obey the zero-mean Gaussian distribution, for example, \( n_a \sim \mathcal{N}(0, \Sigma_a) \), \( n_w \sim \mathcal{N}(0, \Sigma_w) \). In addition, the slowly varying biases of the accelerometer and gyroscope also need to be estimated to correct the error of INS. The variations of biases are usually modeled as the first-order Gauss-Markov process and are expressed as follows (Chen 1992):

\[
\begin{bmatrix}
b_{a,t} \\
b_{w,t}
\end{bmatrix} = 
\begin{bmatrix}
\frac{1}{\tau_a} b_{a,t} + n_{a,t} \\
\frac{1}{\tau_w} b_{w,t} + n_{w,t}
\end{bmatrix}
\]

(5)

where, \( \tau_a \) and \( \tau_w \) represent the correlation time of the Gauss-Markov process for the accelerometer and
gyroscope, respectively. The clean linear acceleration and angular velocity can be used in IMU preintegration only after the observation noise and sensor bias are compensated.

2.4 Monocular visual observation model

Visual SLAM extracts and matches feature points as observations from images obtained by cameras. The monocular vision observation model can be simplified from the binocular vision observation model. For the m-th visual point feature \( f^m \) observed by the monocular camera at the time stamp \( t_i \), the visual observation \( z_{\text{vision}}^m \) of the camera on the normalization projection plane can be expressed as (Xu et al. 2023):

\[
    z_{\text{vision},i}^m = \begin{bmatrix} u_i^m \\ v_i^m \end{bmatrix} = \frac{1}{Z_{vi}} \begin{bmatrix} X_i^m \\ Y_i^m \end{bmatrix} + n_{\text{vision},i}^m
\]  

(6)

where, \( u_i^m \) and \( v_i^m \) represent the pixel coordinates in the normalization plane. \( [X_i^m, Y_i^m, Z_i^m] \) represents a visual landmark in the e-frame. \( n_{\text{vision},i}^m \) indicates visual observation noise. The landmarks in the e-frame can be calculated as follows:

\[
    \begin{bmatrix} X_i^m \\ Y_i^m \end{bmatrix} = \left( R_{i,e}^n \right)^T \left( p_{i,e}^n - p_{i,e}^e \right)
\]  

(7)

where, \( R_{i,e}^n \) and \( p_{i,e}^n \) respectively represent the rotation matrix and position of the monocular camera from the e-frame to the n-frame at the time stamp \( t_i \); \( p_{i,e}^e \) indicates the position of the m-th visual landmark in the n-frame. The position of visual point features in the e-frame is obtained by the Levenberg-Marquardt method (Mourikis and Roumeliotis 2007), where the inverse depth is parameterized by the Gaussian noise attribute (Civera et al. 2008).

3 State vector and factor graph optimization for the nonlinear problem

3.1 MAP estimation

Our integrated system employs a sliding window optimization method to estimate the attitude of land vehicles in the visual keyframe and the depth of visual features. Based on the parameters to be estimated of the multi-sensor observation models described above, we assume that all sensors are well calibrated to IMU, and the state vector to be optimized for MVIO/UC-PPP tightly coupled system is established as follows:

\[
    X = \begin{bmatrix} x_o, x_1, \ldots, x_n, d_1, \ldots, d_n, \psi \end{bmatrix}
\]  

(8)

with

\[
    x_k = \begin{bmatrix} p_{b_k}^e, v_{b_k}^e, q_{b_k}^e, b_{a_k}, \delta t_{G_k}, \delta t_N \end{bmatrix}, k \in [0, n] \\
    \text{ISB} = \begin{bmatrix} ISB_{G,b_k}, ISB_{G,e}, ISB_{b,G} \end{bmatrix} \\
    N = \begin{bmatrix} N_1, \ldots, N_{i_k}, \ldots, N_{i_e}, \ldots, N_n \end{bmatrix}
\]  

(9)

where, \( x_k \) represents the state vector of the \( k \)-th window. \( d_n \) indicates the inverse depth of each point feature at the current time. \( \psi \) indicates the heading compensation between the v-frame and the e-frame. \( p_{b_k}^e, v_{b_k}^e \) and \( q_{b_k}^e \) are the positions, velocities and orientations from the b-frame to the v-frame. \( b_a \) and
\( b_a \) represent the biases of acceleration and gyroscope, respectively. \( \delta t^G \) indicates clock error of GPS receiver; \( \text{ISB} = [\text{ISB}_{G,R}, \text{ISB}_{G,E}, \text{ISB}_{G,C}] \) represents the ISBs corresponding to GLONASS, Galileo and BDS. \( \delta i \) is the receiver clock drift rate. \( N \) represents the float ambiguities of all satellites in different frequencies.

This paper defines the optimal system state as one that maximizes a posterior (MAP) given all the observations. The integrated system establishes an objective equation combining IMU pre-integration, visual projection and UC-PPP constraints for nonlinear optimization. To formulate the optimization problem properly representing the factor graph, we assume that IMU, camera and GNSS observations are independent and their noise follows the zero-mean Gaussian distribution. The MAP estimation can be transformed into one that minimizes the sum of a series of costs, where each cost corresponds to an observation (Gong et al. 2021; Cao et al. 2022). The nonlinear problem can be represented as:

\[
X^* = \underset{X}{\text{arg max}} \, p(X \mid z) = \underset{X}{\text{arg max}} \, p(X) \cdot p(z \mid X) = \underset{X}{\text{arg max}} \, p(X) \prod_{i=1}^{n} p(z_i \mid X) = \underset{X}{\text{arg min}} \begin{pmatrix}
\Gamma \left( \left\| r_{\text{pri}} - H_{\text{pri}} X \right\|^2 \right) + \sum_{k=\text{imu}}^{\text{imu}} \Gamma \left( \left\| r_{\text{imu}} (\hat{z}_k^{\text{imu}}, X) \right\|^2 \right) \\
+ \sum_{k=\text{vision}} \Gamma \left( \left\| r_{\text{vision}} (\hat{z}_k^{\text{vision}}, X) \right\|^2 \right) + \sum_{j=\text{gnss}} \Gamma \left( \left\| r_{\text{gnss}} (\hat{z}_j^{\text{gnss}}, X) \right\|^2 \right)
\end{pmatrix}
\]  

(10)

where, \( \Gamma(\cdot) \) represents the robust loss function. \( \{r_{\text{pri}}, H_{\text{pri}}\} \) is a priori information from marginalization. \( \left\| \cdot \right\| \) represents the square of the Mahalanobis distance with covariance \( \Sigma \). \( r_{\text{imu}} (\hat{z}_k^{\text{imu}}, X), \, r_{\text{vision}} (\hat{z}_k^{\text{imu}}, X) \) and \( r_{\text{gnss}} (\hat{z}_j^{\text{gnss}}, X) \) represent the residuals of IMU, vision and UC-PPP observations \( \hat{z}_k^{\text{imu}}, \hat{z}_m^{\text{imu}} \) and \( \hat{z}_j^{\text{gnss}} \) respectively. This transformation makes the optimization problem decompose into independent factors. The factor graph diagram of the MVIO/UC-PPP tightly coupled system is shown in Fig. 2. Colored circles represent different optimization states, and colored rectangles represent different factors. Concerning the construction of visual factor \( f \), inertial factor \( i \), Doppler factor \( D \) and receiver clock drift factor \( i \), we refer to (Cao et al. 2022). The following will introduce the UC-PPP-related main factors.
3.2 Pseudorange and carrier phase factors

After error correction, the observations of pseudorange and carrier phase factors can be obtained from equations (1) and (2), respectively. We assume that the observation noises of the pseudorange and carrier phase obey the zero mean Gaussian distribution, for example, $\varepsilon_j^{\text{GNSS}} \sim \mathcal{N}(0, \sigma_{\varepsilon})$ and $\xi_j^{\text{GNSS}} \sim \mathcal{N}(0, \sigma_{\xi})$. The stochastic model of GNSS has been widely studied, and it is usually based on satellite elevation angle, signal-to-noise ratio and other information that can reflect satellite visibility and measurement accuracy. Given the analysis in a previous paper (Pan et al. 2021), we choose the sine function model based on the elevation angle to model the carrier phase observation noise:

$$\sigma_j^2 = a^2 + b^2 \sin^2 E$$  \hspace{1cm} (11)

where, $a$ and $b$ are empirical parameters related to the receiver’s observation performance. In this paper, $a = 3 \text{ mm}$, $b = 4 \text{ mm}$. In practice, the pseudorange and carrier phase noise ratio is set as 100 to obtain pseudorange observation noise.

To enhance UC-PPP with MVIO, the position of MVIO in the $n$-frame needs to be converted to the $e$-frame. Once the origin of the $n$-frame in the $e$-frame is given, the conversion can be completed through the following rotation matrix:

$$R_e^n = \begin{bmatrix}
-sin \lambda & -sin \phi \cos \lambda & \cos \phi \cos \lambda \\
\cos \lambda & -sin \phi \sin \lambda & \cos \phi \sin \lambda \\
0 & \cos \phi & \sin \phi
\end{bmatrix}$$  \hspace{1cm} (12)

where, $\phi$ and $\lambda$ are the latitude and longitude of the selected origin in the geodetic coordinate system. The 1-DOF rotation from $v$-frame to $n$-frame is given by heading angle compensation $\psi$. Therefore, the position of MVIO in the $e$-frame can be expressed as

$$\mathbf{p}_e^r = R_e^n R_v^n \left( \mathbf{p}_v^r - \mathbf{p}_{n, \text{origin}}^r \right) + \mathbf{p}_{e, \text{origin}}^r$$  \hspace{1cm} (13)

where $\mathbf{p}_v^r$ is the receiver position in the $v$-frame. $\mathbf{p}_{n, \text{origin}}^r$ and $\mathbf{p}_{e, \text{origin}}^r$ represent the origin positions of the $n$-frame in the $v$-frame and $e$-frame, respectively. To facilitate the conversion, we make the $v$-frame and $n$-frame have the same origin. In addition, $\mathbf{p}_v^r$ is usually obtained after lever arm correction. The
residuals of pseudorange and carrier phase at time stamps $t_i$ can be expressed as

$$r_p = \left( z_{p,i}^t, X \right) = \left\| \mathbf{p}^s_{t_i} - \mathbf{R}^s_{r_i} \mathbf{r}^s_{b} - \mathbf{p}^s_{\text{orig}} \right\| + c \left( \mathbf{t}_r^G - \mathbf{t}_s^G \right) + M \cdot \mathbf{ZTD} + \gamma_{t_i} \mathbf{f}_{s,i}^G - \mathbf{p}_{s,j}^G \right)$$

(14)

Where, $\| \|$ represents the modulus of the vector. $\mathbf{p}^s_{t_i}$ represents the coordinates of satellite $s$ in the e-frame.

### 3.3 Receiver clock and ISB factor

According to the equation (3), the ISBs corresponding to GLONASS, Galileo and BDS can be equivalent to the difference between their respective receiver clocks and GPS receiver clocks. For simplicity, we estimate the receiver clock errors of the four constellations simultaneously in the optimization. First of all, there is the following relationship between the deviation of the receiver clock and the clock drift rate from time stamps $t_i$ to $t_{i-1}$ (Cao et al. 2022):

$$\mathbf{t}_s = \mathbf{t}_s - \mathbf{t}_{s,i} + \mathbf{1}_{4 \times 1} \int_{t_{i-1}}^{t_i} \mathbf{d} t$$

(16)

where, $\mathbf{t}_s = \left[ \mathbf{t}_s^G, \mathbf{t}_s^E, \mathbf{t}_s^F \right]^T$, $\mathbf{1}_{4 \times 1}$ represents a matrix with $n$ rows and $m$ columns of the value 1.

Discretizes equation (16) to obtain the residual of the receiver clock error as

$$r_r = \left( z_r^t, X \right) = \mathbf{t}_s - \mathbf{t}_{s,i} - \mathbf{1}_{4 \times 1} \int_{t_{i-1}}^{t_i} \mathbf{d} t$$

(17)

where $\mathbf{t}_s^4$ is the time difference between adjacent timestamps; The receiver clock drift rate $\mathbf{t}_s^\delta$ is modeled as a random walk process.

### 4.4 Ambiguity factor

Generally, the integer cycle count of the carrier phase observation will be interrupted when GNSS signals are affected by factors such as low satellite elevation angle, multipath error, receiver motion state, and environmental occlusion, creating a cycle slip. Cycle slips are very frequent in complex urban environments. In this work, the cycle slip detection is no longer performed separately because the float ambiguity of the carrier phase on each frequency is solved in real-time through optimization (Wen and Hsu 2021). The residual of ambiguity at the time stamp $t_i$ is expressed as

$$r_a = \left( z_a^t, X \right) = \mathbf{N}_s - \mathbf{N}_{s,i}^t$$

(18)

where, $\mathbf{N}_{s,i}^t$ represents the float ambiguity of the satellite $s$ at the current time on the $j$-th frequency.

The structure of the MVIO/UC-PPP tightly coupled system based on factor graph optimization is shown in Fig. 3. After the system is started, it is necessary to initialize monocular vision and INS first, and then MVIO/UC-PPP. Since single constellation UC-PPP needs five satellites to realize positioning, this paper sets the minimum number of satellites required for MVIO/PPP initialization to 5. The initialization task of MVIO and PPP is completed with a coarse-to-fine method. The marginalization method is used to select and reject the observations for the sliding window to retain the constraints between the observations as much as possible and maintain the computation costs at a fixed level.
Fig. 3 Structure of MVIO/UC-PPP tightly coupled integration based on factor graph optimization

4 Experiment and analysis

4.1 Equipment and scheme

The MVIO/UC-PPP tightly coupled system is verified using two open datasets collected by pedestrian and car in the complex environment of Hong Kong. The sensors of the two datasets and their installation methods are the same, including a VI_Sensor (Nikolic et al. 2014) and a u-blox ZED-F9P GNSS receiver. The VI_Sensor consists of two Aptina MT9V034 cameras and one tactical ADIS16488 IMU. The technical indicators of low-cost IMU are shown in Table 1. In the following experiment, we only use the visual observations obtained by the left camera of the VI_Sensor. The technical indicators and processing strategies of monocular vision are shown in Table 2. The ground truth is obtained by the internal RTK engine of ZED-F9P, which is capable of providing the receiver’s location at an accuracy of 1 cm in an open area (Cao et al. 2022).

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Bias</th>
<th>Random Walk</th>
<th>Dynamic Range</th>
<th>Sampling Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gyroscope</td>
<td>6 °/hr</td>
<td>0.3 °/hr</td>
<td>±450 °/s</td>
<td>200Hz</td>
</tr>
<tr>
<td>Accelerometer</td>
<td>0.1 mg</td>
<td>0.029 m/s/√h</td>
<td>±18 g</td>
<td>200Hz</td>
</tr>
</tbody>
</table>

Table 1. Performance indicators of ADIS16488 IMU.

<table>
<thead>
<tr>
<th>Item</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling rate</td>
<td>20 Hz</td>
</tr>
<tr>
<td>Sliding window size</td>
<td>10</td>
</tr>
<tr>
<td>Feature observation noise</td>
<td>0.035m</td>
</tr>
<tr>
<td>Resolutions</td>
<td>752*480 pixels</td>
</tr>
<tr>
<td>Maximum tracking features</td>
<td>150</td>
</tr>
<tr>
<td>Feature detection</td>
<td>FAST corner</td>
</tr>
<tr>
<td>Feature tracking</td>
<td>KLT optical flow</td>
</tr>
</tbody>
</table>

Table 2. Processing strategy of monocular vision.

We design four schemes to demonstrate the performance of the proposed method. Scheme 1: standard point positioning with Inertial Explorer version 8.80 (SPP); Scheme 2: precise point positioning with Inertial Explorer (PPP); Scheme 3: VIO/SPP tightly coupled positioning of GVINS (GVINS) (Cao et al. 2022); Scheme 4: factor graph optimization-based MVIO/UC-PPP tightly coupled positioning proposed in this paper (MVIO-PPP). The multiple GNSS UC-PPP processing strategy used

1 https://github.com/HKUST-Aerial-Robotics/GVINS-Dataset
in Scheme 4 is shown in Table 3.

### Table 3. Multi-GNSS UC-PPP processing strategy.

<table>
<thead>
<tr>
<th>Item</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constellation and signal</td>
<td>GPS, L1+L2; GLONASS, G1+G2; Galileo, E1+E5b; BDS, B1+B2</td>
</tr>
<tr>
<td>Positioning model</td>
<td>Uncombined model</td>
</tr>
<tr>
<td>Sampling rate</td>
<td>10 Hz</td>
</tr>
<tr>
<td>Elevation cutoff angle</td>
<td>10°</td>
</tr>
<tr>
<td>Minimum number of satellites</td>
<td>5</td>
</tr>
<tr>
<td>Zenith tropospheric delay</td>
<td>Saastamoinen model (Boehm et al. 2006)</td>
</tr>
<tr>
<td>Ionospheric delay</td>
<td>Klobuchar model</td>
</tr>
<tr>
<td>Stochastic model of observation noise</td>
<td>Elevation-based sine function</td>
</tr>
<tr>
<td>Phase ambiguity</td>
<td>Float</td>
</tr>
<tr>
<td>Sagnac effect</td>
<td>Corrected (Ashby 2004)</td>
</tr>
</tbody>
</table>

#### 4.2 Low-speed and short-distance movement of pedestrian in complex urban environment

To test the performance of the proposed MVIO/UC-PPP tightly coupled system, we adopt a dataset collected by a pedestrian to carry out the test (Pedestrian test). The sensors are installed on a helmet, and the pedestrian wears the helmet to walk on the city road. The feature of this dataset is that the pedestrian has a slower moving speed and a shorter moving distance. The maximum speed is about 1.9 m/s, and the track is about 1.6 km. Fig. 4 shows the positioning trajectories of four schemes and gives six representative challenging scenes, including overpasses, fully-shaded avenues, and half-shaded high-rise buildings. It can be seen from Fig. 4 that the positioning performance of SPP and PPP without VIO assistance has declined sharply in challenging environments. Fig. 5 shows the visible satellites, the Position Dilution of Precision (PDOP) of multi-GNSS and the geodetic height of the track during the whole movement. Since the solution of UC-PPP requires at least five satellites, we only consider PDOP when the number of visible satellites is larger than five. Fig. 5 shows that the number of visible satellites in complex environments varies significantly, and even with multi-GNSS there is a shortage of satellites. PDOP is an indicator that reflects the geometric distribution of satellites. Generally, the number of satellites decreases while the PDOP increases, indicating the satellite distribution is poor, and the reliability of multi-GNSS positioning deteriorates. In addition, the geodetic height of the dataset is 0.5~130.5 m, and the trajectory height changes significantly. Therefore, UC-PPP positioning should focus on the three-dimensional (3D) position accuracy of pedestrian or vehicles rather than just the plane position accuracy.
Pedestrians usually walk on the sidewalk near trees or buildings in urban areas, which undoubtedly increases the difficulty of GNSS navigation and positioning. To more clearly and intuitively display the challenges of the six scenes marked in Fig. 4 and the positioning performance of the four schemes, Fig. 6 shows the visual images obtained by the camera in this scenario and the tracks of the four schemes displayed in the image map. The environment described in Fig. 6 (a) - (c) and (f) includes an overpass or high-rise, and the GNSS signal captured by the receiver has a low intensity and is prone to cycle slips. The sudden drop or complete interruption of the number of visible satellites led to a significant deviation in the positioning of SPP and PPP. Compared to PPP, the positioning accuracy of GVINS and MVIO-PPP is more reliable and stable. The environment shown in Fig. 6 (d) and (e) corresponds to an architectural canyon and a semi-sheltered situation, respectively. Here, GNSS signals are usually affected by multipath. The PPP scheme using carrier phase observations has higher positioning accuracy than SPP. In all challenging environments, the positioning performances of the GVINS scheme and MVIO-PPP scheme are close.

Based on the ground truths mentioned in Section 4.1, Table 4 gives the root mean square errors (RMSEs) and positioning continuity of the four schemes in the corresponding time intervals of representative scenes (a) to (f). As shown in Table 4, in scenarios (d) or (e) with a sufficient number of visible satellites, the positioning accuracy of the PPP scheme is higher than that of the SPP scheme and even better than that of the GVINS scheme. At the same time, their positioning continuities are all 100%. In scenarios (a) or (b), the continuity of the PPP scheme is lower than that of the SPP scheme, mainly due to the insufficient number of visible satellites to provide dual-frequency observations. Based on Fig. 6, it can be seen that the positioning performance of the PPP scheme in scenario (f) is inferior to that of the SPP scheme due to the significant positioning errors in a few epochs. With the aid of VIO, the GVINS scheme achieves better positioning accuracy than the SPP scheme, and the MVIO-
PPP scheme achieves better positioning accuracy than the PPP scheme. Compared to other schemes, the 3D positioning accuracy of MVIO-PPP in each scene is the highest and the positioning continuity is 100%.

Fig. 6 The images collected by the camera and the trajectories of the four schemes under the six challenging scenes of the pedestrian test. The yellow, pink, cyan and red track points are the position results of SPP, PPP, GVINS and MVIO-PPP, respectively.

Table 4 RMSEs and positioning continuity of the four schemes in the corresponding time intervals of representative scenes (a) to (f) (unit: meter). The positioning is considered effective if the scheme provides positioning results at the current time. The positioning continuity of the scheme in this time interval is determined by the ratio of the effective positioning time to the total time.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPP</td>
<td>East</td>
<td>5.260</td>
<td>3.228</td>
<td>4.473</td>
<td>5.436</td>
<td>2.808</td>
<td>2.414</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>3.113</td>
<td>3.366</td>
<td>4.559</td>
<td>3.268</td>
<td>1.897</td>
<td>2.575</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>11.073</td>
<td>4.959</td>
<td>11.079</td>
<td>8.137</td>
<td>4.102</td>
<td>6.087</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>PPP</td>
<td>East</td>
<td>2.403</td>
<td>2.887</td>
<td>4.595</td>
<td>1.057</td>
<td>1.035</td>
<td>6.296</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>2.285</td>
<td>2.362</td>
<td>4.435</td>
<td>0.206</td>
<td>0.464</td>
<td>6.497</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>6.773</td>
<td>5.756</td>
<td>10.476</td>
<td>2.818</td>
<td>3.150</td>
<td>10.726</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>7.541</td>
<td>6.859</td>
<td>12.269</td>
<td>3.017</td>
<td>3.348</td>
<td>14.032</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>99.71%</td>
<td>92%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>GVINS</td>
<td>East</td>
<td>0.987</td>
<td>1.609</td>
<td>2.678</td>
<td>2.233</td>
<td>1.716</td>
<td>0.435</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>1.379</td>
<td>1.524</td>
<td>1.197</td>
<td>1.465</td>
<td>1.348</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>2.386</td>
<td>2.766</td>
<td>3.093</td>
<td>3.651</td>
<td>3.705</td>
<td>3.906</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>2.927</td>
<td>3.544</td>
<td>4.263</td>
<td>4.524</td>
<td>4.300</td>
<td>3.962</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>MVIO-</td>
<td>East</td>
<td>0.898</td>
<td>0.092</td>
<td>0.363</td>
<td>0.868</td>
<td>0.669</td>
<td>0.409</td>
</tr>
<tr>
<td>PPP</td>
<td>North</td>
<td>1.471</td>
<td>0.654</td>
<td>0.740</td>
<td>0.330</td>
<td>1.289</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>0.801</td>
<td>0.992</td>
<td>0.908</td>
<td>2.494</td>
<td>2.386</td>
<td>1.899</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>1.900</td>
<td>1.192</td>
<td>1.226</td>
<td>2.661</td>
<td>2.793</td>
<td>2.176</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

In order to further analyze the positioning performance of the MVIO-PPP scheme for the whole duration of the test, the positioning errors of the four schemes are shown in Fig. 7. According to Fig. 7 and Fig. 5, it is difficult for SPP to provide reliable positioning under multipath effects and other
factors. The positioning performance of PPP decreases significantly when the number of visible satellites is less than 8. Table 5 shows the RMSE of the positioning results of the four schemes in the east, north, up and 3D directions for the whole duration of the test. The 3D position accuracy of the four schemes is ranked from high to low as follow: MVIO-PPP, GVINS, PPP, and SPP. The RMSEs of the MVIO-PPP scheme are the smallest in the east, north and up directions, 0.612, 0.978 and 1.880 m, respectively. Compared with the PPP scheme, the positioning accuracies of the MVIO-PPP scheme in the east, north and up directions are improved by 73.3, 54.8 and 62.7 %, respectively.

![Fig. 7](image)

**Fig. 7** Positioning error of the four schemes in the pedestrian test. The gray rectangular areas correspond to the six representative scenes (a)-(f) mentioned in Fig. 4

**Table 5** RMSEs of positioning errors and positioning continuity from the four schemes in the pedestrian test (unit: meter)

<table>
<thead>
<tr>
<th>Scheme</th>
<th>East</th>
<th>North</th>
<th>Up</th>
<th>3D</th>
<th>Continuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPP</td>
<td>3.223</td>
<td>3.424</td>
<td>7.194</td>
<td>8.594</td>
<td>100%</td>
</tr>
<tr>
<td>PPP</td>
<td>2.288</td>
<td>2.166</td>
<td>5.037</td>
<td>5.941</td>
<td>99.48%</td>
</tr>
<tr>
<td>GVINS</td>
<td>1.526</td>
<td>1.103</td>
<td>3.343</td>
<td>3.837</td>
<td>100%</td>
</tr>
<tr>
<td>MVIO-PPP</td>
<td>0.612</td>
<td>0.978</td>
<td>1.880</td>
<td>2.206</td>
<td>100%</td>
</tr>
</tbody>
</table>

4.3 High-speed and long-distance movements of vehicles in complex urban environments

The vehicle speed changes significantly due to continuous acceleration and deceleration during driving in the city. The intense movement will make it difficult for GNSS signal acquisition, so the cycle slips of carrier phase observations are significantly more prevalent than in static mode. In addition, frequent ambiguity initialization will reduce the performance of GNSS navigation and positioning. To test the performance of the MVIO-PPP scheme under high-speed and long-distance movement, a dataset collected by vehicle is used to carry out the test (Vehicle test). The sensors described in section 3.1 are installed on the vehicle in a similar manner. The maximum speed of the vehicle is 21.2 m/s, and the track is about 23 km. This dataset contains many challenging environments, such as boulevards, overpasses and architectural canyons. Fig. 8 shows the positioning trajectories of the four schemes and marks for six representative road sections. In the three urban canyon environments (d), (e) and (f) shown in Fig. 8, the SPP scheme is barely able to provide uninterrupted positioning but has significant errors due to multipath effects and fewer satellites. The PPP scheme cannot provide continuous location and is very unreliable. The MVIO-PPP scheme is better than
GVINS and has good positioning accuracy and continuity in these challenging environments. Fig. 9 shows the number of visible satellites, PDOP and vehicle geodetic height in the vehicle test. The gray rectangles represent time intervals for each scene. The vehicle passed through 13 overpasses similar to Fig. 10 (a) from 1800 s to 2100 s. Therefore, the number of visible satellites in this time frame changes rapidly. When the vehicle moves between different urban environments, the number of visible satellites changes, and there are fewer than five satellites for the PPP scheme on many occasions. In addition, the significant geodetic height change of the vehicle further emphasizes the importance of accurate 3D position in navigation and positioning.

![Positioning trajectories of the four schemes in the Google Earth and six representative challenging scenes of the vehicle test marked with the white rectangular frames](image)

**Fig. 8** Positioning trajectories of the four schemes in the Google Earth and six representative challenging scenes of the vehicle test marked with the white rectangular frames
Fig. 9 The number of visible satellites and PDOP from PPP scheme and the pedestrian geodetic height of vehicular test. The gray rectangular areas correspond to representative scenes (a), (b), (c) and (d) - (f) mentioned in Fig. 8. The vehicle passes through scene (b) and scenes (d) - (f) twice, respectively.

Vehicles are faced with highly complex traffic conditions when traveling in urban environments. Fig. 10 shows the images captured by the camera in the six representative complex scenes mentioned in Fig. 8 and the trajectories of the four schemes displayed in the image maps. The areas shown in Figs. 10 (a) - (d) are relatively common in urban environments, whereas the areas shown in Figs. 10 (e) - (f) are extreme architectural canyons. The satellite elevation mask angles of scenes (e) and (f) are about 75 ° and 60 °, respectively. In a heavily shaded scenario, even multi-GNSS cannot provide enough visible satellites for PPP schemes. Combining the images collected by the camera and positioning trajectories, we can find that compared with the PPP scheme, the positions of the GVINS scheme and the MVIO-PPP scheme are more reliable. The vehicle passed twice through the areas (d) - (f) shown in Fig. 10, but GVINS only provided part of the track due to the instability of the system. On the contrary, the MVIO-PPP scheme provides reliable and continuous positioning for vehicles. Table 6 shows the RMSEs and positioning continuity of the four schemes in the corresponding time intervals of representative scenes (a), (b), (c), and (d) - (f). Since the vehicle passes through scenes (b) and (d) - (f) twice in sequence, their corresponding time intervals are longer. The positioning accuracy of GVINS is improved by introducing carrier phase observation in MVIO-PPP. Compared with other schemes, MVIO-PPP has the highest positioning accuracy and good positioning continuity in all time intervals.
Fig. 10 The images collected by the camera and the positioning trajectories of the four schemes in the six challenging scenes of the vehicle test. The yellow, pink, cyan and red track points are the position results of SPP, PPP, GVINS and MVIO-PPP, respectively.

Table 6 RMSEs and positioning continuity of the four schemes in the corresponding time intervals of representative scenes (a), (b), (c), and (d) - (f) (unit: meter)

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Time interval (s)</th>
<th>(a): 241-301</th>
<th>(b): 326-536</th>
<th>(c): 564-644</th>
<th>(d) - (f): 1219-1619</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPP</td>
<td>East</td>
<td>8.755</td>
<td>14.537</td>
<td>4.366</td>
<td>15.793</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>8.101</td>
<td>15.504</td>
<td>2.683</td>
<td>11.690</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>21.368</td>
<td>61.695</td>
<td>16.152</td>
<td>44.210</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>24.472</td>
<td>65.253</td>
<td>16.945</td>
<td>48.380</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>100.00%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>PPP</td>
<td>East</td>
<td>3.949</td>
<td>1.797</td>
<td>2.571</td>
<td>8.211</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>3.660</td>
<td>2.806</td>
<td>2.456</td>
<td>7.128</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>8.335</td>
<td>11.291</td>
<td>5.184</td>
<td>20.323</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>9.923</td>
<td>11.772</td>
<td>6.286</td>
<td>23.049</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>92.33%</td>
<td>98.95%</td>
<td>89.13%</td>
<td>93.88%</td>
</tr>
<tr>
<td>GVINS</td>
<td>East</td>
<td>1.507</td>
<td>1.302</td>
<td>1.059</td>
<td>2.364</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>2.729</td>
<td>1.995</td>
<td>1.746</td>
<td>2.150</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>4.359</td>
<td>5.037</td>
<td>4.906</td>
<td>10.069</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>82.47%</td>
</tr>
<tr>
<td>MVIO-PPP</td>
<td>East</td>
<td>1.247</td>
<td>1.430</td>
<td>1.698</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>North</td>
<td>2.225</td>
<td>1.254</td>
<td>1.311</td>
<td>1.171</td>
</tr>
<tr>
<td></td>
<td>Up</td>
<td>3.239</td>
<td>2.757</td>
<td>2.614</td>
<td>2.940</td>
</tr>
<tr>
<td></td>
<td>3D</td>
<td>4.123</td>
<td>3.349</td>
<td>3.382</td>
<td>3.308</td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>100.00%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Fig. 11 shows the positioning error of the four schemes. Compared with the SPP and PPP schemes, the positioning results of the GVINS and MVIO-PPP schemes are closer to the ground truth. However, the positioning of the GVINS scheme was unexpectedly interrupted at about 1327s. Due to the dark sky and many vehicles passing through the section, VIO faced challenges. The positioning of GVINS did not have enough satellites to complete initialization within a short time after the outage. Table 7 shows the RMSE of the positioning results from the four schemes for the whole duration of the vehicle test. The RMSEs of the MVIO-PPP scheme are 1.309, 1.301 and 2.745 m in the east, north and up directions, respectively, which are the smallest among the four schemes. Compared with the PPP
scheme, the improvement rate of MVIO-PPP positioning accuracies in the east, north and up directions are 63.0, 59.3 and 70.5 %, respectively. In addition, we can see that the MVIO-PPP scheme has the best positioning performance based on the continuity and RMSEs of the 3D position from the four schemes.

![Graph showing positioning error of four schemes in vehicle test.](image)

**Table 7** RMSEs of positioning errors from four schemes in vehicle test (unit: meter)

<table>
<thead>
<tr>
<th>Schemes</th>
<th>East</th>
<th>North</th>
<th>Up</th>
<th>3D</th>
<th>Continuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPP</td>
<td>8.442</td>
<td>7.006</td>
<td>27.988</td>
<td>30.061</td>
<td>100%</td>
</tr>
<tr>
<td>PPP</td>
<td>3.542</td>
<td>3.200</td>
<td>9.307</td>
<td>10.460</td>
<td>97.65%</td>
</tr>
<tr>
<td>GVINS</td>
<td>1.527</td>
<td>1.536</td>
<td>5.349</td>
<td>5.771</td>
<td>97.23%</td>
</tr>
<tr>
<td>MVIO-PPP</td>
<td>1.309</td>
<td>1.301</td>
<td>2.745</td>
<td>3.308</td>
<td>100%</td>
</tr>
</tbody>
</table>

### 5 Conclusion

The low-cost MEMS-IMU, camera and GNSS receiver have significantly promoted the research of multi-sensor fusion positioning. This paper proposes a MVIO/UC-PPP tightly coupled system based on factor graph optimization to enhance the positioning performance of PPP in complex urban environments. The system employs the uncombined model as the fusion object to satisfy the application requirements of multi-system and multi-frequency GNSS. We transform the nonlinear optimization problem of multi-sensor fusion into a set of problems that minimize the sum of costs, namely, factor graph optimization. First, the alignment and initialization of the MVIO are completed through the constructed monocular vision and inertial factors. Secondly, the UC-PPP-related factors are introduced to complete the initialization process with the MVIO under the condition of enough visible satellites. From then on, the tightly coupled system starts to provide global position information. Similar to GVINS, the initialization of UC-PPP and MVIO uses a coarse-to-fine method to correct the conversion of local and global frames online. The sliding window and the marginalization methods are used to preserve constraints between different observations and eliminate bootless observations within the window, respectively. Pedestrian and vehicle tests verify the performance of the proposed method. Both tests contain a large number of complex environments. Compared to GVINS, the positioning accuracy of our proposed method has been further improved due to the use of carrier phase
observations with higher measurement accuracy. In the low-speed and short-distance pedestrian test, compared with PPP alone, the improvement rate of the proposed method in the east, north and up directions is 73.3, 54.8 and 62.7%, respectively, while the improvement rate in the high-speed and long-distance vehicle test is 63.0, 59.3 and 70.5% respectively. The proposed MVIO-PPP has higher positioning accuracy and outstanding performance in complex environments.

This paper only completed the tightly coupled integration of MVIO and UC-PPP and then verified the performance of the integrated system. In the future, the contribution of VIO in PPP cycle slip detection or ambiguity fixing ought to be explored. In addition, with the support of regional reference stations, VIO can assist PPP-RTK in achieving high-precision navigation and positioning.

**Authorship contribution statement**

Cheng Pan: Conceptualization, Methodology, Software, Writing - original draft. Fangchao Li, Yuanxin Pan, and Yonghui Wang: Investigation, Writing - review & editing. Benedikt Soja and Jingxiang Gao: Supervision, Writing - review & editing. Zengke Li: Supervision, Project administration, Funding acquisition. All authors read and approved the final manuscript.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgement**

Thanks to the HKUST Aerial Robotics Group for the open datasets. This paper was supported by the National Natural Science Foundation of China (41974026), the Graduate Innovation Program of China University of Mining and Technology (2022WLXJ108) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCXZ22_2591). Cheng Pan would like to thank the China Scholarship Council for sponsoring his visit to ETH Zurich.

**References**


