



Centimeter-accurate vehicle navigation in urban environments with a tightly integrated PPP-RTK/MEMS/vision system

Xin Li¹ · Xingxing Li¹ · Shengyu Li¹ · Yuxuan Zhou¹ · Meng Sun¹ · Qi Xu¹ · Zhili Xu¹

Received: 26 April 2022 / Accepted: 19 July 2022 / Published online: 12 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Emerging applications like autonomous cars and unmanned aerial vehicles demand accurate, continuous and reliable positioning. The PPP-RTK technique, which could provide a rapid centimeter-accurate positioning service using a single GNSS receiver, has been recognized as a promising tool for mass-market and automotive applications. Nevertheless, the positioning performance of PPP-RTK degrades in urban environments because of the frequent signal deteriorating and blocking. Inertial navigation system (INS) is commonly integrated with GNSS to improve the continuity, accuracy and reliability of precise positioning as it has several advantages of all-environment operability and high temporal resolution, but it is limited by rapid error accumulation in long-term operation, especially when a microelectromechanical system inertial measurement unit (MEMS-IMU) is employed. Fortunately, the camera, another low-cost sensor, which provides rich information about the surrounding environment, is expected to improve the navigation performance further. This contribution develops a tightly integrated PPP-RTK/MEMS/vision model to achieve continuous and accurate positioning in urban environments. The raw data of MEMS-IMU and a stereo camera, as well as the high-precision GNSS phase measurements, are fused based on a multistate constraint Kalman filter to fully exploit the complementary properties from heterogeneous sensors. On this basis, a fast ambiguity resolution is achievable with the augmentation of the high-precision INS/vision information and the precise atmospheric corrections. The proposed integrated system is validated by several vehicle experiments conducted in urban areas. Results indicate that a centimeter-level accuracy of 4.1, 2.2 and 7.3 cm in the east, north and up components, respectively, and a high fixing percentage of 96.8% can be achieved for PPP-RTK/MEMS/vision in an urban environment, which exhibits comparable performance with respect to the tight integration of PPP-RTK and a tactical IMU. Besides, ambiguity refixing can be implemented instantaneously for the integrated system when going under three consecutive overpasses in 25 s.

Keywords PPP-RTK · INS · Vision · Rapid ambiguity resolution · Tightly coupled integration

Introduction

High accuracy and reliable positioning are increasingly significant in emerging technologies such as self-driving cars and mobile robots. The PPP-RTK technique (Wubbena et al. 2005), which is proposed to achieve centimeter-level positioning using a single GNSS receiver by rapid ambiguity resolution (AR), is well recognized as a promising tool for high-precision and time-critical applications. Generally, the implementation of PPP-RTK includes two sequential steps. The state space representation (SSR) corrections,

such as precise orbits and clocks, phase and code biases, as well as atmospheric delays, are estimated first and then disseminated to users to enable the precise point positioning (PPP) rapid AR for the area covered by reference stations. In this way, PPP-RTK shows the same flexibility as PPP and achieves higher accuracy by applying ambiguity resolution. Moreover, compared with the traditional real-time kinematic (RTK) technique using the observational state representation (OSR), the use of SSR corrections significantly releases the real-time communication burden and enables the further improvement of positioning performance by precise atmospheric modeling (Li et al. 2014; Yang et al. 2022). In recent years, several studies concerning atmospheric modeling (Teunissen and Khodabandeh 2015; de Oliveira et al. 2017; Psychas and Verhagen 2020; Ren et al. 2021), ambiguity resolution (Ge et al. 2008; Laurichesse et al. 2009; Zhang et al.

✉ Xingxing Li
xxli@sgg.whu.edu.cn

¹ School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

2019) and multi-GNSS fusion (Li et al. 2021a, 2022) have been conducted to investigate and improve the performance of PPP-RTK. These results indicate that under the open-sky, centimeter-accurate positioning can be easily achieved by PPP-RTK with the atmospheric corrections augmentation.

However, in an urban environment where signals are frequently blocked, the performance of PPP-RTK degrades rapidly due to the limited satellite number and poor observation quality. Fortunately, as an autonomic and spontaneous system, the inertial navigation system (INS) is free from signal deteriorating and blocking, which is commonly integrated with the GNSS to enhance positioning performance. Currently, the INS is widely used to integrate with RTK or PPP processing. The significant improvement in accuracy, continuity and reliability that a combining utilization of GNSS/INS brings to precise positioning is carefully analyzed and evaluated (Rabbou and El-Rabbany 2014, 2015; Liu et al. 2016; Gao et al. 2016). Recently, more attention has been paid to PPP-RTK/INS, attributing to the superiorities of PPP-RTK. Gu et al. (2021) developed a multi-GNSS PPP/INS tightly coupled integration with atmospheric augmentation, reaching an accuracy of 0.22 m and 0.45 m for dual- and single-frequency users, respectively. Li et al. (2021b) proposed a tightly coupled PPP-RTK/INS integration method with a tactical and a microelectromechanical system inertial measurement unit (MEMS-IMU) employed to verify the positioning performance, respectively. Results indicate that, for a tactical IMU, the horizontal positioning accuracy of 5–6 cm can be achieved in an urban environment, whereas for a MEMS-IMU, positioning accuracy of decimeter level or even worse is obtained because of the rapid error accumulation during the GNSS outage.

To mitigate the navigation error drifts of INS and further improve the positioning performance in complex urban environments, some low-cost sensors, such as vehicle odometer and visual sensors, have been applied in GNSS/INS integration to provide redundant measurements for the filter update (Gao et al. 2018; Reuper et al. 2018). Compared to the vehicle's internal odometry only providing forward velocity information in one dimension, the visual sensor, which is capable of capturing abundant surrounding features, has the potential to present a better augmentation performance for the robust estimation of positioning and velocity parameters. Angelino et al. (2012) proposed a loosely coupled GNSS/INS/vision integration method in order to avoid the divergence of position estimates in a GNSS-denied environment. It shows that using visual measurements allows continuous meter-level positioning accuracy with a 120-s outage of GPS. Lynen et al. (2013) constructed the multi-sensor fusion extended Kalman filter (MSF-EKF) framework that achieved the seamless sensor-feed loose integration of GNSS, INS and visual sensors. Besides, a graph optimization-based multi-sensor fusion approach was also employed to fuse the

local poses from visual inertial odometry and GPS position measurements (Mascaro et al. 2018). To avoid information loss in the loosely coupled model, Cao et al. (2021) proposed a nonlinear optimization-based approach to tightly fuse GNSS pseudorange measurements with visual and IMU data, which leads to a majority of meter-level and decimeter-level position results in complex environments. Li et al. (2019) integrated the single-frequency RTK, MEMS-IMU and the monocular camera to improve positioning availability in GNSS-challenged environments where ambiguity resolution is still conducted on the basis of GNSS observations with the INS-derived position constraint.

In the previous studies, the low-cost MEMS-IMU and camera were commonly integrated to GNSS in a loosely coupled mode, or only pseudorange observations of GNSS were employed for filter updates. Mostly, only meter-level or decimeter-level positioning accuracy can be achieved with a single GNSS receiver. This contribution develops a low-cost navigation system with a tightly coupled PPP-RTK/MEMS/vision integration method for centimeter-accurate vehicle navigation. The GNSS carrier phase observations and the raw data of MEMS-IMU and stereo camera are processed in a centralized EKF estimator to exploit the complementary sensors. In addition, rapid AR is achieved with the augmentation of the high-precision INS/vision information and precise atmospheric corrections. Several experiments in different situations of urban environments are designed to evaluate the performance of the proposed method.

After this introduction, the method of tightly coupled integration of PPP-RTK/MEMS/vision is described. Then the experimental sets and the processing strategies in the vehicle-borne experiments are detailed. Hereafter, the experimental results in typical urban situations are investigated and the positioning performance of PPP-RTK/MEMS/vision tight integration is validated. Finally, the conclusions and perspectives are illustrated.

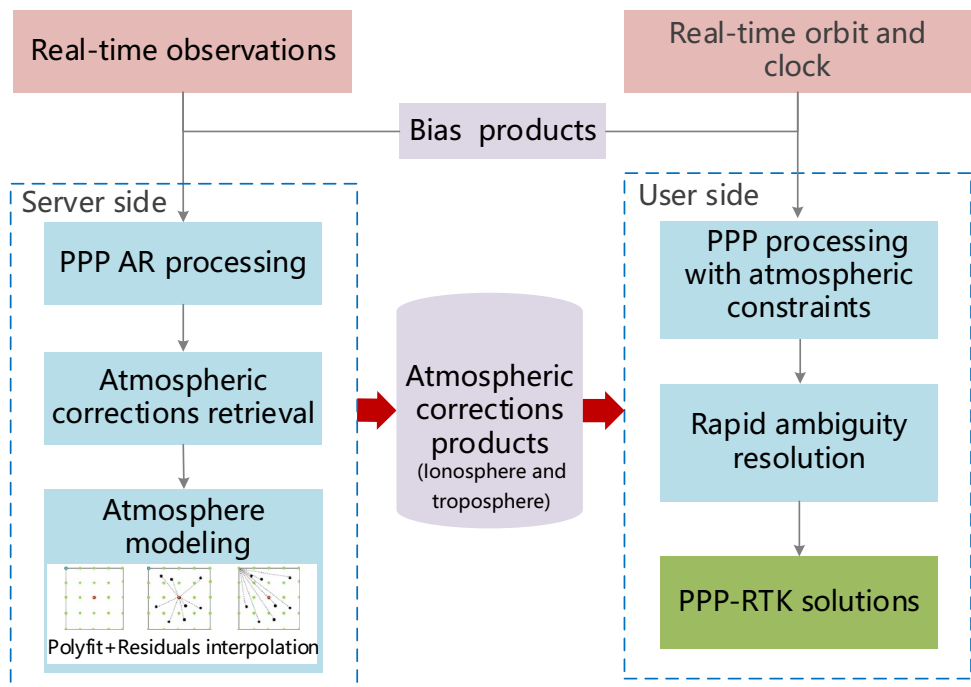
Method

In this section, the implementation of a PPP-RTK method is introduced first. Then, with an overview of the PPP-RTK/MEMS/vision tightly integrated system, the error model of all involved sensors in the integrated system is introduced, respectively; then, the state model and measurement model of the integrated filter are detailed. Finally, a theoretical analysis of the benefits of the tightly coupled integrated system is presented.

PPP-RTK method

In our PPP-RTK model shown in Fig. 1, the precise atmospheric corrections, including the ionospheric and the

Fig. 1 Implementation of PPP-RTK processing



tropospheric delays, are derived from the PPP fixed solutions with precise orbit and clock, as well as uncalibrated phase delay (UPD) products prepared at the server, and then disseminated to and interpolated at the user stations to enable the PPP rapid AR. For a specific description of the proposed PPP-RTK model, the methods of atmospheric corrections retrieval and modeling at the server, and PPP rapid ambiguity resolution with precise atmospheric constraints at the user will be detailed, respectively.

Atmosphere retrieval and modeling at the server

With the real-time orbits, clock and bias corrections determined using observation streams from the global network in advance, PPP AR can be implemented to estimate the precise atmospheric delays (including the slant ionospheric and zenith tropospheric delays). The derived ionospheric and tropospheric corrections (δT_r and δI_r^s) can be expressed as:

$$\begin{cases} \delta T_r = Z_{r,w} \\ \delta I_r^s = \bar{I}_{r,1}^s \end{cases} \quad (1)$$

with $\bar{I}_{r,1}^s = I_{r,1}^s - f_2^2 (b_{r,1}^s - b_{r,2}^s) / (f_1^2 - f_2^2)$ (2)

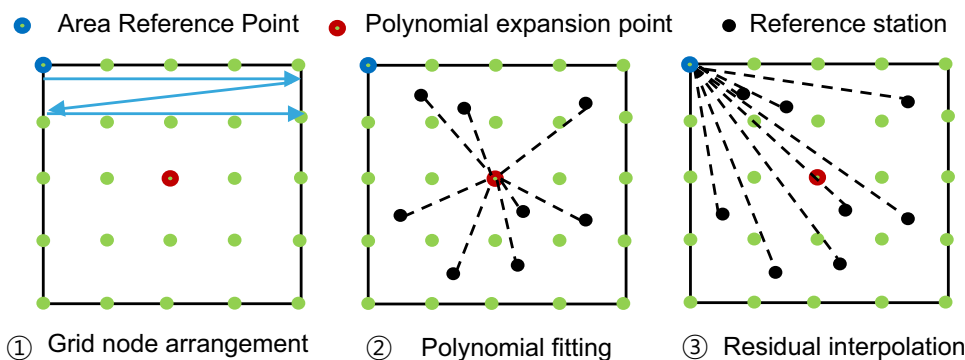
where $Z_{r,w}$ and $\bar{I}_{r,1}^s$ are estimated zenith tropospheric wet delay and slant ionospheric delay, respectively, and b refers to the code hardware delay, which in the combined forms are absorbed by ionospheric delays in the estimation system.

After the precise atmospheric corrections at each server station are derived, the interpolation method such as the modified linear combination method (MLCM) is commonly applied to obtain the precise atmospheric information at the user station according to the distance between the reference and user stations (Li et al. 2011). We employ a grid-based atmospheric correction model in our study to improve communication efficiency. Figure 2 shows the implementation procedure. After determining the area reference point and grid node spacing, the grid node arrangement is implemented following in order from west to east and from north to south. With the precise atmospheric corrections available at the reference stations, the low-order polynomial functions as (3) are used to compute the coefficients of ionospheric and tropospheric corrections with the polynomial expansion point at the center of the grid. A least-squares method is employed for atmospheric coefficients estimation.

$$\begin{cases} k = 3: \delta A_{r_i} = C_{00} + C_{01}(\phi_{r_i} - \phi_0) + C_{10}(\lambda_{r_i} - \lambda_0) \\ k \geq 3: \delta A_{r_i} = C_{00} + C_{01}(\phi_{r_i} - \phi_0) + C_{10}(\lambda_{r_i} - \lambda_0) \\ \quad \quad \quad + C_{11}(\phi_{r_i} - \phi_0)(\lambda_{r_i} - \lambda_0) \end{cases} \quad (3)$$

where δA_{r_i} denotes the retrieved atmospheric correction at the reference station i ; k is the number of reference stations; ϕ and λ refer to the geographic latitude and longitude of the regional stations, respectively, whereas ϕ_0 and λ_0 are counterparts of the polynomial expansion point; and C_{00} , C_{01} , C_{02} and C_{03} are polynomial coefficients to be estimated. Once the polynomial coefficients are obtained, the residual

Fig. 2 Procedure of the grid-based atmospheric correction modeling



between the retrieved and fitting values is computed and then interpolated at each grid node based on the MCLM method. In this way, the sum of polynomial fitting values and interpolated residuals from the surrounding grid nodes are the corrections users required.

PPP AR with precise atmospheric constraints at the user

At the user, the PPP model with atmospheric constraints can be expressed as:

$$\begin{cases} p_{r,j}^s = \mathbf{g}_r^s \cdot \delta \mathbf{p} + \bar{t}_r + m_{r,w}^s \cdot Z_{r,w} + \gamma_n \cdot \bar{I}_{r,1}^s + e_{r,n}^s \cdot \sigma_{p_{r,n}^s}^2 \\ l_{r,j}^s = \mathbf{g}_r^s \cdot \delta \mathbf{p} + \bar{t}_r + m_{r,w}^s \cdot Z_{r,w} - \gamma_n \cdot \bar{I}_{r,1}^s + \lambda_n \cdot \bar{N}_{r,n}^s + \varepsilon_{r,n}^s \cdot \sigma_{l_{r,n}^s}^2 \\ Z_{r,w} - \tilde{Z}_{r_1, r_2, \dots, r_k} = 0, \sigma_{Z_{r,w}}^2 \\ \bar{I}_{r,1}^s - \tilde{I}_{r_1, r_2, \dots, r_k} = 0, \sigma_{I_{r,1}^s}^2 \end{cases} \quad (4)$$

where $p_{r,j}^s$ and $l_{r,j}^s$ represent the observed minus computed values of the pseudorange and carrier phase observations, respectively; \mathbf{g}_r^s symbolizes the unit vector of the component from the receiver to the satellite; $\delta \mathbf{p}$ ($\delta \mathbf{p} = [dx, dy, dz]$) indicates the vector of the receiver position increments relative to the a priori position; $m_{r,w}^s$ stands for the mapping function of zenith tropospheric delay; and $\bar{N}_{r,n}^s$ denotes the float ambiguity at frequency n , which is biased by the receiver- and satellite-specific hardware delays. The precise atmospheric corrections ($\tilde{Z}_{r_1, r_2, \dots, r_k}$ and $\tilde{I}_{r_1, r_2, \dots, r_k}$) will be taken as the virtual observations to impose atmospheric constraints with variance of $\sigma_{Z_{r,w}}^2$ and $\sigma_{I_{r,1}^s}^2$ (Zhang et al. 2022). It is worth noting that the ionospheric corrections absorb the frequency-dependent code biases from reference stations, which will introduce an additional bias for the observations. To avoid the effect of this code bias, a frequency-dependent receiver code bias ($\tilde{b}_{r,u}$, Psychas and Verhagen 2020) should be estimated for a PPP-RTK user. Overall, the parameters X to be estimated in the PPP-RTK system can be written as:

$$\mathbf{X} = \left[\delta \mathbf{p}, \bar{t}_r^G, \mathbf{ISB}_r^{\text{sys}}, Z_{r,w}, \bar{I}_r^s, \tilde{b}_{r,u}, \mathbf{N}_{r,1}^s, \mathbf{N}_{r,2}^s \right]^T \quad (5)$$

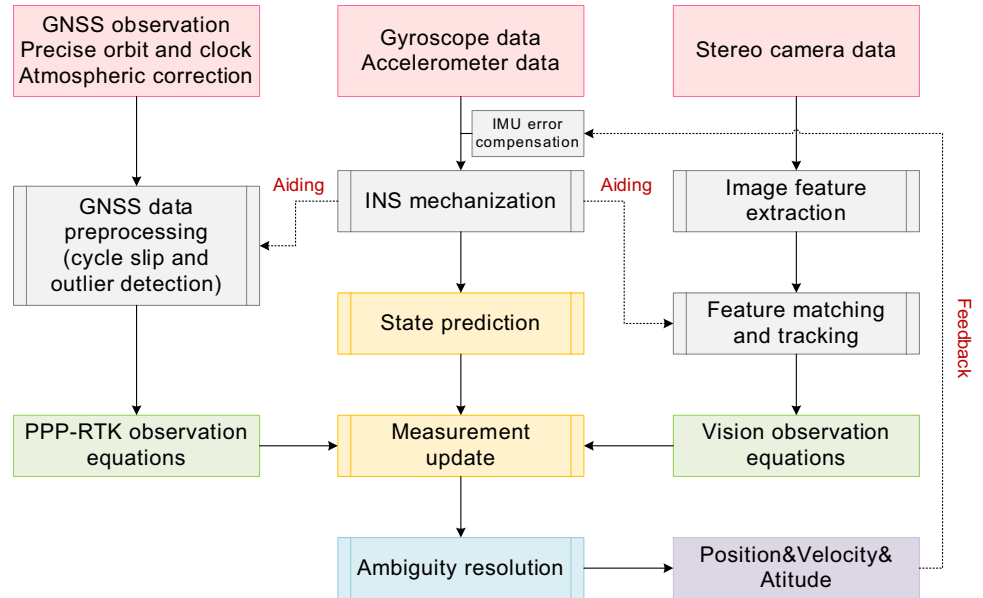
where **ISB** denotes inter-system bias, which needs to be estimated for multi-GNSS processing with the assumption that all systems share the GPS receiver clock (\bar{t}_r^G).

With the precise atmospheric corrections and bias corrections, the PPP rapid AR is achieved based on a cascade ambiguity fixing strategy using the wide-lane ($\mathbf{N}_{r,1}^s - \mathbf{N}_{r,2}^s$) and L1 ($\mathbf{N}_{r,1}^s$) ambiguities. The WL ambiguities are fixed using the round strategy, and the integer WL ambiguities will be taken as the virtual observation to improve the precision of L1 ambiguity. Finally, the L1 ambiguity will be fixed by the LAMBDA method (Teunissen 1995).

Tightly integrated PPP-RTK/INS/vision system

The system architecture of the PPP-RTK/INS/vision tight integration is presented in Fig. 3. In the integrated system, a multistate constraint Kalman filter (MSCKF) estimator is employed to fuse the high-precision GNSS carrier phase observations, as well as the raw data of MEMS-IMU and stereo camera. After the INS dynamic alignment and initialization of the integration filter, the state propagation is performed by the INS mechanization. The high-precision INS-predicted positions will assist in the cycle slip and outlier detection in the PPP-RTK processing as well as the feature matching and tracking in the image processing. If a new image is obtained, the camera pose will be added to the state vector with the corresponding covariance augmented. When the GNSS or vision observations are available, the measurement update will be applied to update the state parameters and the corresponding covariance matrix. Finally, the estimated gyroscope and accelerometer biases will be fed back to correct the next IMU data to restrain the INS divergence. Since the PPP-RTK method has been detailed above, this section first gives a brief view of the INS and vision sensor model and then focuses on the tight integration filter algorithm of PPP-RTK, INS and vision.

Fig. 3 System architecture of the PPP-RTK/INS/vision tight integration



INS error model

The linearized INS dynamic model can be written as (Shin 2005):

$$\begin{cases} \delta\dot{\theta} = -\omega_{ie}^e \times \delta\theta - \mathbf{R}_b^e \delta\epsilon^b + \xi_\theta \\ \delta\dot{\mathbf{v}}^e = -2\omega_{ie}^e \times \delta\mathbf{v}^e + (\mathbf{R}_b^e \mathbf{f}_{ib}^b) \times \delta\theta + \mathbf{R}_b^e \delta\mathbf{a}^b + \xi_v \\ \delta\dot{\mathbf{p}}^e = \delta\mathbf{v}^e + \xi_r \end{cases} \quad (6)$$

where e , i and b are the earth-centered earth-fixed (ECEF) frame, earth-centered inertial and inertial sensor body frames, respectively; $\delta\theta^e$, $\delta\mathbf{v}^e$ and $\delta\mathbf{p}^e$ stand for attitude, velocity and position error in e -frame, respectively; \mathbf{R}_b^e is defined as a rotation matrix from b -frame to e -frame; \mathbf{f}_{ib}^b is the specific force output of accelerometer in b -frame; ω_{ie}^e is rotation angular rate vector of the e -frame with respect to the i -frame expressed in e -frame; ξ_θ , ξ_v and ξ_r demonstrate random walk process in attitude, velocity and position; and $\delta\epsilon^b$ and $\delta\mathbf{a}^b$ are the errors of gyroscope and accelerometer measurements in b -frame, respectively, which usually can be described by the first-order Gauss–Markov procedures and expressed as follows:

$$\begin{cases} \delta\dot{\epsilon}^b = -\frac{1}{\tau_\epsilon} \delta\epsilon^b + \xi_\epsilon \\ \delta\dot{\mathbf{a}}^b = -\frac{1}{\tau_a} \delta\mathbf{a}^b + \xi_a \end{cases} \quad (7)$$

where ξ_ϵ and ξ_a denote the power spectral density (PSD) of accelerometer and gyroscope, respectively, while τ_ϵ and τ_a are corresponding correlation time of the random process. It should be noted that the coarse values of these parameters

can be obtained using the Allan variance analysis technique. Therefore, the INS error state vector can be written as:

$$\mathbf{X}_{\text{ins}} = [\delta\theta^e \ \delta\mathbf{v}^e \ \delta\mathbf{p}^e \ \delta\epsilon^b \ \delta\mathbf{a}^b]^T \quad (8)$$

where the errors of attitude, velocity, position, gyroscope and accelerator measurements are estimated as random walk process with a proper power density to exhibit their temporal characteristics.

Visual measurement model

For a single static feature f_j observed at the time step i by a stereo camera, the visual measurement $\mathbf{z}_{\text{vis},i}^j$ on the normalized projective plane of the left and right cameras can be expressed as:

$$\mathbf{z}_{\text{vis},i}^j = \begin{bmatrix} u_{i,l}^j \\ v_{i,l}^j \\ u_{i,r}^j \\ v_{i,r}^j \end{bmatrix} = \begin{bmatrix} X_j^{C_{i,l}} / Z_j^{C_{i,l}} \\ Y_j^{C_{i,l}} / Z_j^{C_{i,l}} \\ X_j^{C_{i,r}} / Z_j^{C_{i,r}} \\ Y_j^{C_{i,r}} / Z_j^{C_{i,r}} \end{bmatrix} + \mathbf{n}_{\text{vis},i}^j \quad (9)$$

where the subscripts l and r denote left camera and right camera, respectively; u_i^j and v_i^j denote the normalized pixel coordinates that are provided by the camera; $X_j^{C_{i,n}}$, $Y_j^{C_{i,n}}$ and $Z_j^{C_{i,n}}$ ($n \in l, r$) are feature positions in the camera frame; and $\mathbf{n}_{\text{vis},i}^j$ refers to the measurement noise. Here, $X_j^{C_{i,n}}$, $Y_j^{C_{i,n}}$, $Z_j^{C_{i,n}}$ can be calculated by:

$$\begin{aligned} \begin{bmatrix} X_j^{C_{i,l}} & Y_j^{C_{i,l}} & Z_j^{C_{i,l}} \end{bmatrix}^T &= (\mathbf{R}_{C_{i,l}}^e)^T (\mathbf{p}_j^e - \mathbf{p}_{C_{i,l}}^e) \\ \begin{bmatrix} X_j^{C_{i,r}} & Y_j^{C_{i,r}} & Z_j^{C_{i,r}} \end{bmatrix}^T &= \mathbf{R}_{C_{i,l}}^{C_{i,r}} (\mathbf{p}_j^{C_{i,l}} - \mathbf{p}_{C_{i,l}}^{C_{i,r}}) \end{aligned} \tag{10}$$

where $\mathbf{R}_{C_{i,l}}^e$ and $\mathbf{p}_{C_{i,l}}^e$ represent the attitude rotation matrix and position of the left camera frame with respect to the ECEF frame, which can be initialized by position, velocity and attitude information from the INS mechanization; $\mathbf{R}_{C_{i,l}}^{C_{i,r}}$ and $\mathbf{p}_{C_{i,l}}^{C_{i,r}}$ are the attitude rotation matrix and position of left camera frame with respect to right camera frame, which can be accurately calibrated offline (Furgale et al. 2013); and \mathbf{p}_j^e and $\mathbf{p}_j^{C_{i,l}}$ represent feature positions in e-frame and left camera frame, respectively.

In a typical visual method, the positions of the features are included in the state vector and estimated together with camera poses (Mur-Artal and Tardós 2017; Qin et al. 2018). However, in a complex environment, a large number of features will be tracked, which will make the dimension of the state vector become very large. We compute the feature positions are computed through least-squares multi-view triangulation based on the currently estimated camera poses (Sun et al. 2018). Moreover, we adopt the MSCKF algorithm that maintains several previous camera poses in the state vector

and uses visual measurements of the same feature across multiple camera views to form a multi-constraint update, which is efficient in avoiding dimension explosion. In this model, the visual state vector is described as:

$$\mathbf{X}_{\text{vis}} = \begin{bmatrix} \delta\theta_{c_1}^e & \delta\mathbf{p}_{c_1}^e & \delta\theta_{c_2}^e & \delta\mathbf{p}_{c_2}^e & \dots & \delta\theta_{c_k}^e & \delta\mathbf{p}_{c_k}^e \end{bmatrix}^T \tag{11}$$

where $\delta\theta_{c_i}^e$ and $\delta\mathbf{p}_{c_i}^e$ are the attitude and position error of the left camera in different moments, respectively; and k denotes the total number of poses in the sliding window.

The corresponding visual observation model can be expressed as:

$$\mathbf{r}_{\text{vis}} = \mathbf{z}_{\text{vis}} - \hat{\mathbf{z}}_{\text{vis}} = \mathbf{H}_{\text{vis}}(\mathbf{X}_{\text{vis}}) + \mathbf{n}_{\text{vis}} \tag{12}$$

where \mathbf{z}_{vis} and $\hat{\mathbf{z}}_{\text{vis}}$ denote the observed and estimated image measurements on the normalized projective plane of camera; and \mathbf{H}_{vis} is the Jacobian of involved camera states which are given in ‘‘Appendix 1’’.

Tight integration filter of PPP-RTK/INS/vision

As mentioned previously, an MSCKF estimator is employed to fuse the high-precision GNSS carrier phase observations and the raw data of MEMS-IMU and stereo camera. Here, the state vector of the filter is defined as:

$$\mathbf{X}_i = \begin{bmatrix} \delta\theta_i^e & \delta\mathbf{v}_i^e & \delta\mathbf{p}_i^e & \delta\boldsymbol{\varepsilon}_i^b & \delta\mathbf{a}_i^b & \delta\theta_{i-k+1}^e & \delta\mathbf{p}_{i-k+1}^e & \dots & \delta\theta_{i-1}^e & \delta\mathbf{p}_{i-1}^e & \underbrace{\begin{bmatrix} \tilde{t}_r^G, \mathbf{ISB}_r^{\text{sys}}, Z_{r,w}, \tilde{\mathbf{I}}_r^s, \tilde{\mathbf{b}}_{r,u}, \mathbf{N}_{r,1}^s, \mathbf{N}_{r,2}^s \end{bmatrix}^T}_{\mathbf{X}_{\text{ppp-rtk}}} \end{bmatrix}^T \tag{13}$$

The optimal state vector estimates from the Kalman filter can be obtained through processing time updates and measurement updates. The time update process of the PPP-RTK/INS/vision system is written as:

$$\begin{bmatrix} \delta\dot{\theta}_i^e \\ \delta\dot{\mathbf{p}}^e \\ \delta\dot{\theta}_{i-k+1}^e \\ \delta\dot{\mathbf{p}}_{i-k+1}^e \\ \vdots \\ \delta\dot{\theta}_{i-1}^e \\ \delta\dot{\mathbf{p}}_{i-1}^e \\ \delta\dot{\mathbf{v}}_i^e \\ \delta\dot{\boldsymbol{\varepsilon}}^b \\ \delta\dot{\mathbf{a}}^b \\ \dot{\mathbf{X}}_{\text{ppp-rtk}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\omega}_{ie}^e \times & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{R}_b^e & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \\ \mathbf{R}_b^e \mathbf{f}^b \times & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & -2\boldsymbol{\omega}_{ie}^e \times & \mathbf{R}_b^e & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \\ \dots & \mathbf{0}_{\gamma,15+6(k-1)} & \dots & \dots & \dots & \mathbf{0}_{\gamma,\gamma} & \dots & \dots & \dots & \dots & \vdots \end{bmatrix} \begin{bmatrix} \delta\theta_i^e \\ \delta\mathbf{p}^e \\ \delta\theta_{i-k+1}^e \\ \delta\mathbf{p}_{i-k+1}^e \\ \vdots \\ \delta\theta_{i-1}^e \\ \delta\mathbf{p}_{i-1}^e \\ \delta\mathbf{v}_i^e \\ \delta\boldsymbol{\varepsilon}^b \\ \delta\mathbf{a}^b \\ \mathbf{X}_{\text{ppp-rtk}} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{6k \times 6} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{0}_{\gamma \times 6} \end{bmatrix} \begin{bmatrix} \zeta_\varepsilon \\ \zeta_a \end{bmatrix} \tag{14}$$

where γ indicates the number of parameters of PPP-RTK model; and \mathbf{I} and $\mathbf{0}$ denote the 3×3 identity and 3×3 zero matrices, respectively.

It is worth noting that every time a new image is recorded, the state and covariance matrix will be augmented with a copy of the current camera pose estimate. The initial value of camera pose is derived from the INS mechanization and the augmented covariance can be expressed as:

$$\mathbf{Q}'_{ili} = \begin{pmatrix} \mathbf{I}_{15+\gamma+6k} \\ \mathbf{H}' \end{pmatrix} \mathbf{Q}_{ili} \begin{pmatrix} \mathbf{I}_{15+\gamma+6k} \\ \mathbf{H}' \end{pmatrix}^T \tag{15}$$

$$\text{with } \mathbf{H}' = \begin{pmatrix} -(\hat{\mathbf{R}}_b^e \mathbf{R}_c^b)^T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}_{3 \times (\gamma+6\eta)} \\ (\hat{\mathbf{R}}_b^e \mathbf{P}_c^b \times) & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0}_{3 \times (\gamma+6\eta)} \end{pmatrix} \tag{16}$$

where \mathbf{Q}_{ili} and \mathbf{Q}'_{ili} are the covariance matrices before and after the camera state augmentation, respectively.

When there are GNSS or visual observations available, the corresponding measurement update shown as follows is employed to update the state estimates:

$$\begin{bmatrix} \mathbf{r}_{vis} \\ \mathbf{l}_{r,n}^s \\ \mathbf{p}_{r,n}^s \end{bmatrix} = \begin{bmatrix} \mathbf{z}_{vis} - \hat{\mathbf{z}}_{vis} \\ \mathbf{L}_{r,n}^s - \hat{\mathbf{L}}_{ins,n}^s \\ \mathbf{P}_{r,n}^s - \hat{\mathbf{P}}_{ins,n}^s \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{vis} \cdot \mathbf{H}' & \mathbf{0}_{6k,3} \\ \mu(\mathbf{R}_b^e \mathbf{l}^b \times) & \mathbf{g}_r^s & \mathbf{0}_{1,2(k-1)+3} \\ \mu(\mathbf{R}_b^e \mathbf{l}^b \times) & \mathbf{g}_r^s & \mathbf{0}_{1,2(k-1)+3} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{6k,\gamma} \\ \delta\theta_i^e \\ \delta\mathbf{p}^e \\ \delta\theta_{i-k+1}^e \\ \delta\mathbf{p}_{i-k+1}^e \\ \vdots \\ \delta\theta_{i-1}^e \\ \delta\mathbf{p}_{i-1}^e \\ \delta\mathbf{v}_i^e \\ \delta\boldsymbol{\varepsilon}^b \\ \delta\mathbf{a}^b \\ \mathbf{X}_{ppp-rtk} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{vis} \\ \boldsymbol{\varepsilon}_{r,n}^s \\ \mathbf{e}_{r,n}^s \end{bmatrix} \tag{17}$$

$$\text{with } \mathbf{H}_{ppp-rtk} = \begin{bmatrix} I & 0 & m_{r,w}^s & -\gamma_n & 0 & \lambda_1 & \lambda_2 \\ I & I & m_{r,w}^s & \gamma_n & I & 0 & 0 \end{bmatrix} \tag{18}$$

where the $\hat{\mathbf{L}}_{ins,n}^s$ and $\hat{\mathbf{P}}_{ins,n}^s$ denote the INS-predicted GNSS carrier phase and pseudorange measurements, respectively. Since the IMU center has a different reference point to the GNSS antenna phase center, the lever-arm correction (\mathbf{l}^b) should be considered.

Theoretical analysis of the benefits of the tightly coupled integrated system

This section addresses the benefits of the tightly coupled integrated PPP-RTK/INS/vision system, which can be

summarized as follows: (1) The high-frequency visual features and IMU data assist to GNSS cycle slip detection and outlier detection, which will alleviate the effects of non-line of sight (NLOS) and multi-path in urban areas (Li et al. 2021b). (2) The position and attitude information derived from INS mechanization offer accurate initial guesses for visual feature tracking and outlier detection. Accordingly, the visual features observed consecutively are expected to provide periodic corrections for reducing the error accumulation of INS bias, which makes it possible for the low-cost MEMS/vision system to obtain comparable performance to a tactical IMU (Qin et al. 2019). (3) The tightly coupled integration filter fully uses each sensor's complementariness and avoids the possible information loss in loosely coupled integration. The GNSS measurement still works in a tight integration system and contributes to providing continuous and accurate solutions when the available satellites are less than four. In addition, both PPP-RTK and visual measurement updates contribute to the estimation of IMU biases and restraining INS divergence in a tight integration filter. (4) The inclusion of vision and INS (VINS) is capable of

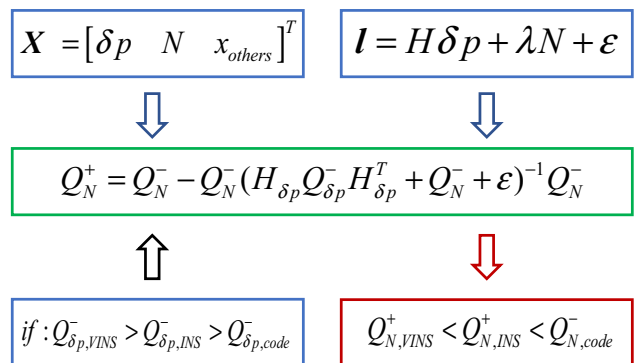


Fig. 4 Aiding effects of VINS in the ambiguity parameter estimation during GNSS outages. H is a matrix describing the satellite geometry; Q^+ and Q^- refer to the posterior and priori covariance, respectively

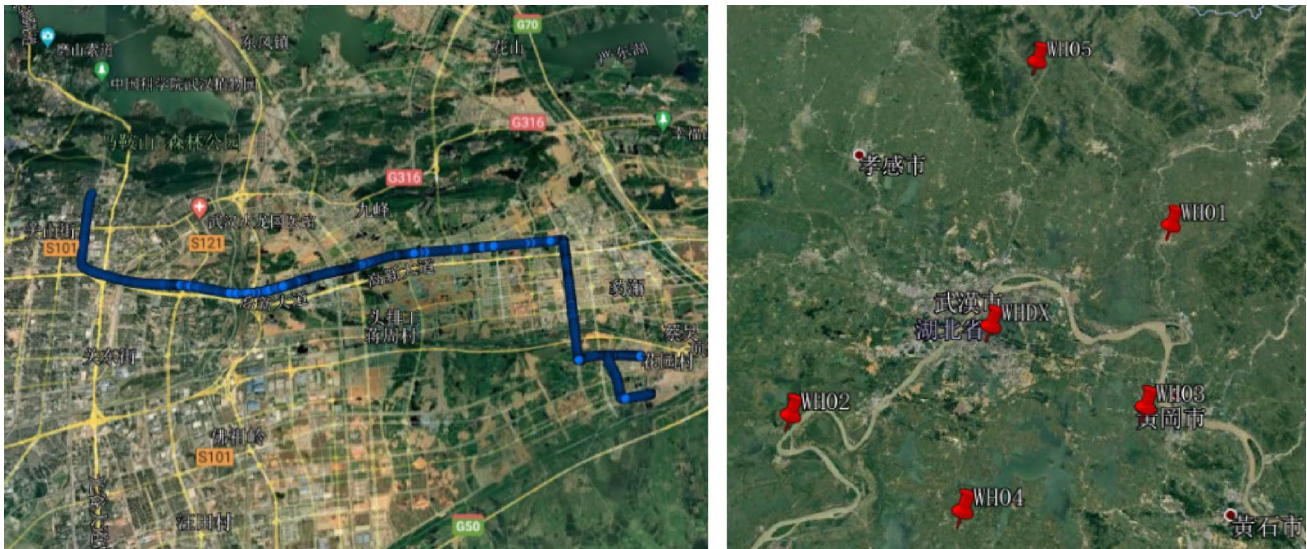


Fig. 5 Trajectory of the vehicle experiment (left panel) and regional augmentation stations employed to generate atmospheric corrections (right panel)



Fig. 6 Illustration of the experimental equipment

providing initial float ambiguity with superior accuracy than GNSS-only and enables fast ambiguity fixing when GNSS signals are recaptured. As shown in Fig. 4, for the ambiguity parameter, the higher the precision of initial float ambiguity (Q^-) is, the better the precision of the estimated float

ambiguity (Q^+). Based on the good quality of float ambiguities provided by VINS, the ambiguity dilution precision (ADOP) and the size of integer ambiguity search space are reduced, which will facilitate the rapid identification of the correct integer candidates (Geng and Shi 2017; Zhang et al. 2018).

Experimental validations

In order to validate the proposed integrated system and evaluate its performance in typical urban situations, several road vehicular tests were conducted in Wuhan, China. For a detailed assessment, a typical experiment from 8:22:51 to 8:59:31 on October 13, 2021, including both open-sky and complex urban environments, is presented. The vehicle's trajectory and the distribution of six reference stations are presented in Fig. 5, respectively.

The experimental setup consists of a Septentrio PolaRx5 GNSS receiver together with a GNSS antenna (NovAtel GPS-702-GG), a tactical grade IMU (StarNeto XW-GI7660), a MEMS-IMU (ADIS-16470) and a stereo camera (FLIR BFS-PGE-31S4C), which is placed on the top of a car (Fig. 6). The raw GNSS observations collected at 1 Hz, the registering images of resolution 1024×768 pixels at 20 Hz from a camera and an inertial measurement at a rate of 100 Hz from MEMS-IMU are employed for tightly integrated system validation. The technical specifications of IMU sensors are provided in Table 1. The reference trajectory is calculated by a tightly coupled multi-GNSS RTK/INS solution with a bidirectional smoothing

Table 1 Technical specifications of the IMU sensors used in the experiment

IMU sensors	Grade	Sampling rates (Hz)	Bias		Random walk	
			Gyro. ($^{\circ}/h$)	Acc. (mGal)	Angular ($^{\circ}/\sqrt{h}$)	Velocity ($m/s/\sqrt{h}$)
XW-GI7660	Tactical	200	0.3	100	0.01	–
ADIS-16470	MEMS	100	8	1300	0.34	0.18

Table 2 Processing strategies of multi-GNSS PPP-RTK

Item	Model
GNSS systems	GPS, Galileo and BDS
Signal selection	GPS: L1/L2; Galileo: E1/E5a; BDS: B1I/B3I
Sampling rate	1 s
Elevation cutoff angle	7°
Weight for observations	Elevation-dependent weight
Ionospheric delay	Corrected by atmospheric corrections
Tropospheric delay	Dry component corrected by Saastamoinen model; wet component corrected by atmospheric corrections
Satellite antenna phase center	igs14.atx
Receiver antenna phase center	igs14.atx
Receiver clock	Epoch-wise estimated
Phase ambiguities	Partial fixing

Table 3 Vision parameters and models applied in the experiment

Item	Model
Sliding window size	20
Feature observation noise	0.01 m
Resolutions	1024*768 pixels
Sampling rate	20 Hz
Feature detection	FAST corner
Feature tracking	KLT optical flow
Maximum tracking features	200

algorithm based on commercial Inertial Explorer (IE) 8.9 software.

Presented in Table 2 is the detail of the multi-GNSS PPP-RTK processing strategy. The satellite orbit and clock errors are corrected by the precise products of the GeoForschungZentrum (GFZ). To enable the ambiguity resolution as well as the precise atmospheric correction retrieval, the satellite UPDs are computed by an open-source software called GREAT-UPD (<https://geodesy.noaa.gov/gps-toolbox/>) in advance. Table 3 shows the vision parameters and models applied in the experiment. For visual data processing, the maximum number of features in tracking is set as 200 to reduce the computational complexity and memory requirements. In addition, a

two-point RANSAC and spatiotemporal loop matching mechanism are employed to reject the outliers in the stereo and temporal feature matching (Sun et al. 2018), and the Chi-square test is applied for conducting the outlier detection before constructing the visual residual vectors (Zuo et al. 2019).

Results

To verify the effectiveness of the tightly coupled PPP-RTK/MEMS/vision method, four different schemes, including PPP-RTK, PPP-RTK/MEMS, PPP-RTK/Tactical grade INS (PPP-RTK/TINS) and PPP-RTK/MEMS/vision, are used to processing the experimental observations. The performance of four solutions was analyzed and compared in terms of positioning accuracy, fixing percentage and refixing time. In our study, the fixing percentage is defined as the percentage of the correctly fixed epochs over the total epochs. A correctly fixed solution is identified when: (1) the requirement of the ratio test and the success rate has been satisfied; and (2) the ambiguity fixed position agrees well with the reference coordinate (positioning errors below 0.1 m), and its accuracy is better than that of ambiguity-float solutions (Zhang and Li 2013).

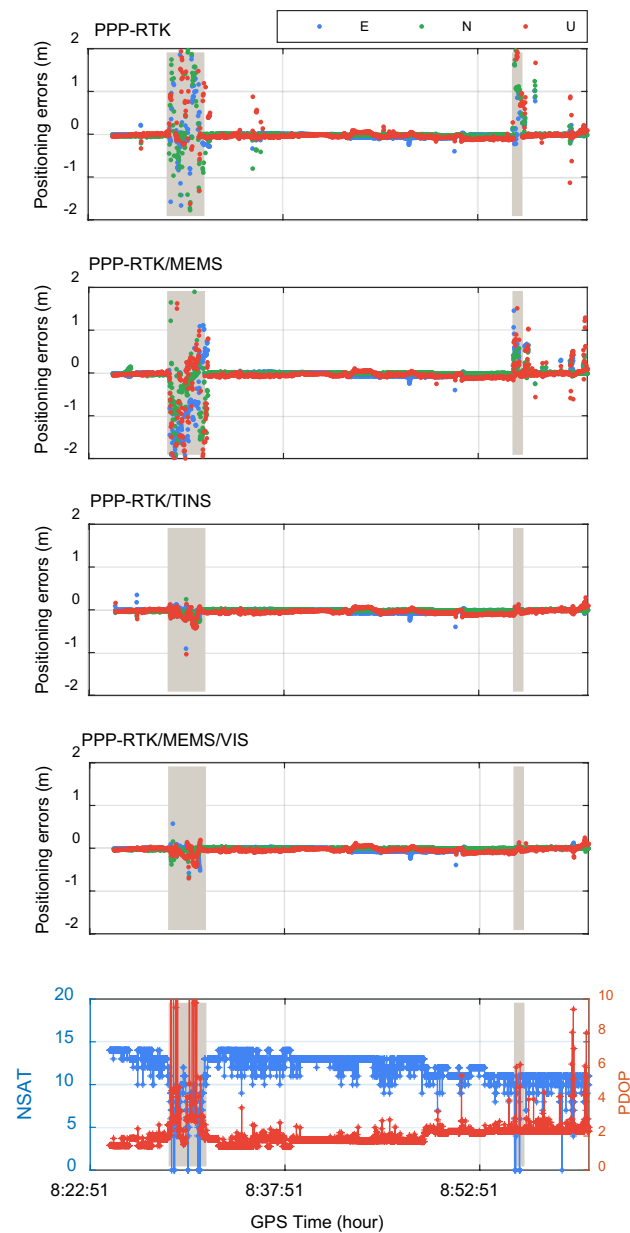


Fig. 7 Positioning series of PPP-RTK, PPP-RTK/MEMS, PPP-RTK/TINS and PPP-RTK/MEMS/vision

Table 4 Positioning accuracy and fixing percentage of PPP-RTK, PPP-RTK/MEMS, PPP-RTK/TINS and PPP-RTK/MEMS/vision

Modes	Positioning accuracy (m) under 95% confidence level			Fixing percentage (%)
	<i>E</i>	<i>N</i>	<i>U</i>	
PPP-RTK	6.16	2.31	15.73	89.7
PPP-RTK/MEMS	0.08	0.12	0.15	91.9
PPP-RTK/TINS	0.04	0.02	0.07	95.6
PPP-RTK/MEMS/VIS	0.04	0.02	0.07	96.8

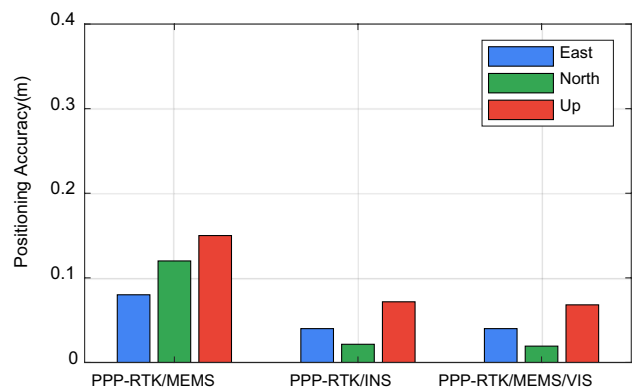


Fig. 8 Positioning accuracy of PPP-RTK/MEMS, PPP-RTK/TINS and PPP-RTK/MEMS/vision

Performance of PPP-RTK/MEMS/vision tight integration

Figure 7 shows the positioning errors series of four processing schemes during the experimental period, respectively. The number of visible satellites (NSAT) and the position dilution of precision (PDOP) are also given in the figure. It can be clearly seen that the total number of satellites of multi-GNSS reaches 10–15 in most epochs and the PDOP is generally less than 2. With the augmentation of precise atmospheric corrections, PPP-RTK can achieve centimeter-level positioning when the available satellite observations are sufficient. The inclusion of other sensors (either INS or vision) further improves the positioning performance, with a positioning series showing fewer outliers. Outliers generally result from the failure of ambiguity resolution when GNSS signals are degraded. Fortunately, the high-frequency IMU data and the abundant visual observation information will both contribute to more reliable ambiguity estimates for ambiguity resolution when tight integration is employed. Two shaded areas in Fig. 7 indicate two typical situations in the urban area where the NSAT decreases sharply. The positioning error of PPP-RTK diverges obviously and reaches several meters. With the aid of a tactical IMU, the positioning solutions can still maintain decimeter-level and even centimeter-level accuracy. However, the improvement is limited with a MEMS-IMU, which could be attributed to its high noise level and the significant bias instability. After applying visual information to PPP-RTK/MEMS, a comparable performance is achieved with respect to the PPP-RTK/TINS in terms of accuracy and reliability.

The statistical results of positioning accuracy and fixing percentage in four processing schemes are given in Table 4 and Fig. 8. Because of the filter divergence in the signal-blocked area, the overall positioning accuracy of PPP-RTK is degraded to the meter level. Significant improvement can be achieved with INS/vision aiding. The positioning

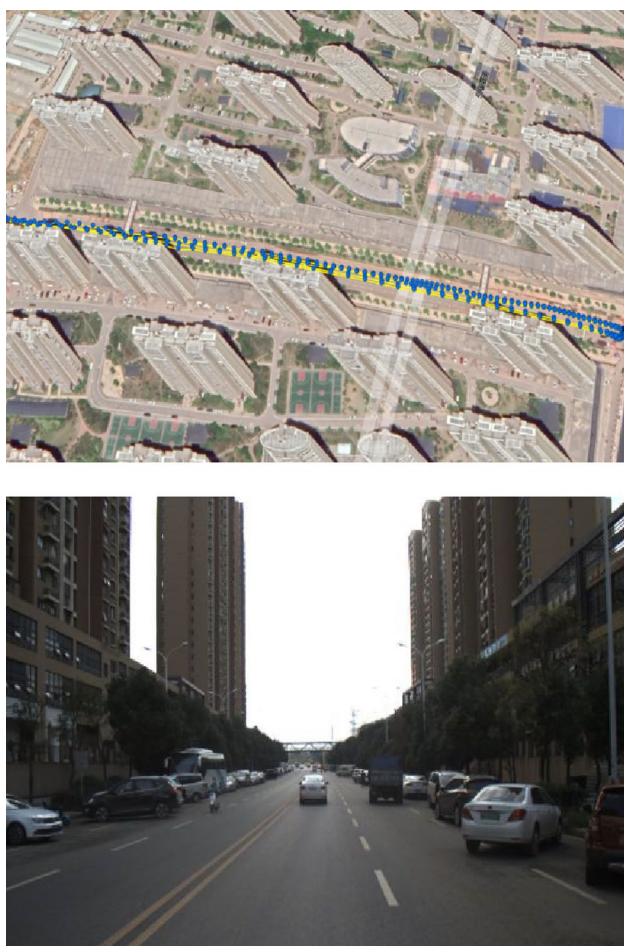


Fig. 9 Scene 1: vehicle trajectory overlaid on Google Maps (top panel) and the typical environment (bottom panel)

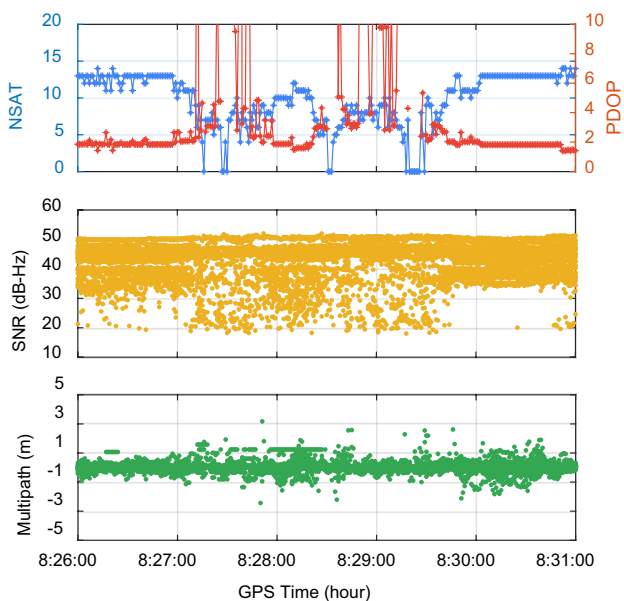


Fig. 10 NSAT, PDOP, SNR and multi-path series in Scene 1

accuracy of 0.08, 0.12 and 0.15 m in the east, north and up components, respectively, can be achieved for PPP-RTK/MEMS and is further improved to centimeter level of 0.04, 0.02 and 0.07 m for PPP-RTK/MEMS/vision integration. Moreover, we also found that PPP-RTK/MEMS/vision solutions exhibit the same level of accuracy as the PPP-RTK/TINS. In addition, PPP-RTK/MEMS/vision integration has the highest fixing percentage of 96.8%, while that of PPP-RTK, PPP-RTK/MEMS and PPP-RTK/TINS is 89.7%, 91.9% and 95.6%, respectively. Note that the PPP-RTK can obtain a relatively high fixing percentage with the augmentation of precise atmospheric corrections; however, the positioning accuracy is still at the meter level due to the rapid divergence and significant deviation of the filter solution with the frequent signal interruptions in the urban canyon area (Shadow area in Fig. 7). By contrast, the tight integration of PPP-RTK, MEMS and vision exhibits a slightly better fixing percentage, but gives a much higher positioning accuracy at the centimeter level, which demonstrates that PPP-RTK/MEMS/vision tight integration can obtain continuous and accurate positions in an urban environment and is expected to be widely applied in massive market applications because of its significant advantages of low expense and portability.

Results validations in typical urban environments

To fully investigate the positioning performance of PPP-RTK/MEMS/vision tight integration in GNSS-challenged environments and analyze the benefits of the inclusion of INS and vision sensors, two typical urban situations (denoted by the shaded area in Fig. 7) are selected for detailed analysis. The vehicle trajectory of Scene 1 overlaid on a Google Maps and its observational environments are presented in Fig. 9. This situation is a typical urban canyon with high-rise buildings along the road, where GNSS signals are badly deteriorated and blocked.

Figure 10 shows the series of NSAT, PDOP, SNR and multi-path in Scene I. We can find that both NSAT and PDOP series change rapidly between 8:27:00 and 8:30:00. At the same time, SNR values of most satellites are decreased to the range of 20–30 dB Hz, while those gather between 30 and 50 dB Hz under an open-sky environment. Meanwhile, a few outliers up to several meters can be found in the third subgraph from the multi-path series. All these factors, including the lack of satellites, the deterioration of signal and severe multi-path, may lead to the failure of ambiguity fixing that impacts the accuracy and reliability of PPP-RTK.

In Fig. 11, the positioning series of PPP-RTK, PPP-RTK/MEMS and PPP-RTK/MEMS/vision in Scene 1 are shown from top to bottom, respectively. The blue, green and red lines represent the east, north and vertical components,

Fig. 11 Positioning series of PPP-RTK, PPP-RTK/MEMS and PPP-RTK/MEMS/VIS in Scene 1. The left and right panels present positioning solutions in different scales

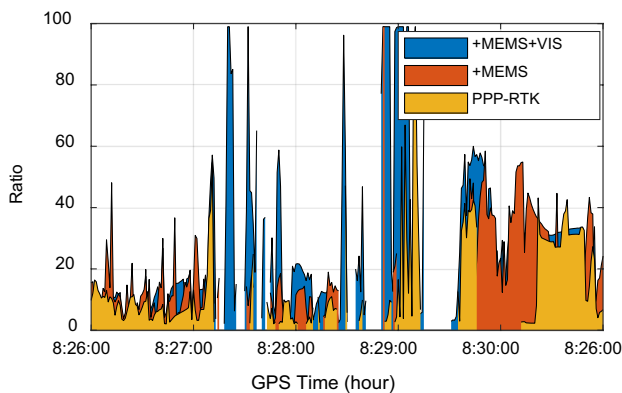
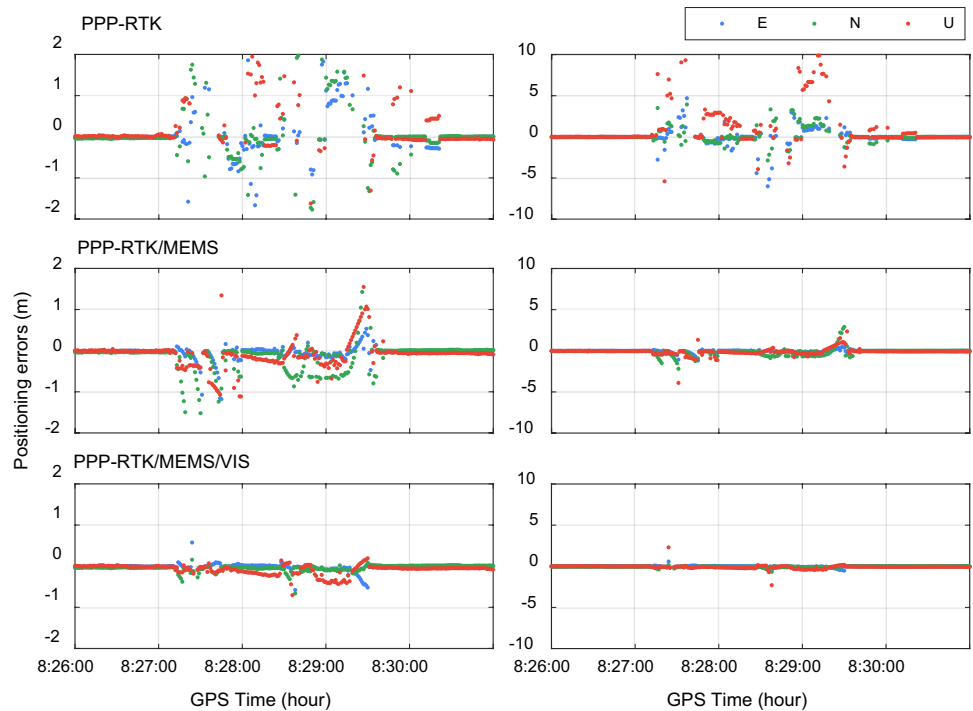


Fig. 12 Ratio values of PPP-RTK, PPP-RTK/MEMS and PPP-RTK/MEMS/VIS in Scene 1

respectively. We can find that the PPP-RTK series diverges rapidly, with positioning accuracy falling to several meters (sometimes larger than 10 m). It indicates that the PPP-RTK method could hardly provide continuous and accurate positions in such complex urban environments, which cannot satisfy the accuracy requirement of vehicle navigation. The tight integration of PPP-RTK and INS (MEMS) can improve the positioning performance with an average accuracy of 0.16, 0.47 and 0.41 m. However, the solutions of PPP-RTK/MEMS are limited by the high noise level and significant instability of MEMS-IMU and thereby cannot enable the ambiguity resolution for a centimeter-level positioning. By contrast, the PPP-RTK/MEMS/vision tight integration exhibits a more stable and accurate positioning series since



Fig. 13 Scene 2: vehicle trajectory overlaid on Google Maps (top) and the observational environment (bottom)

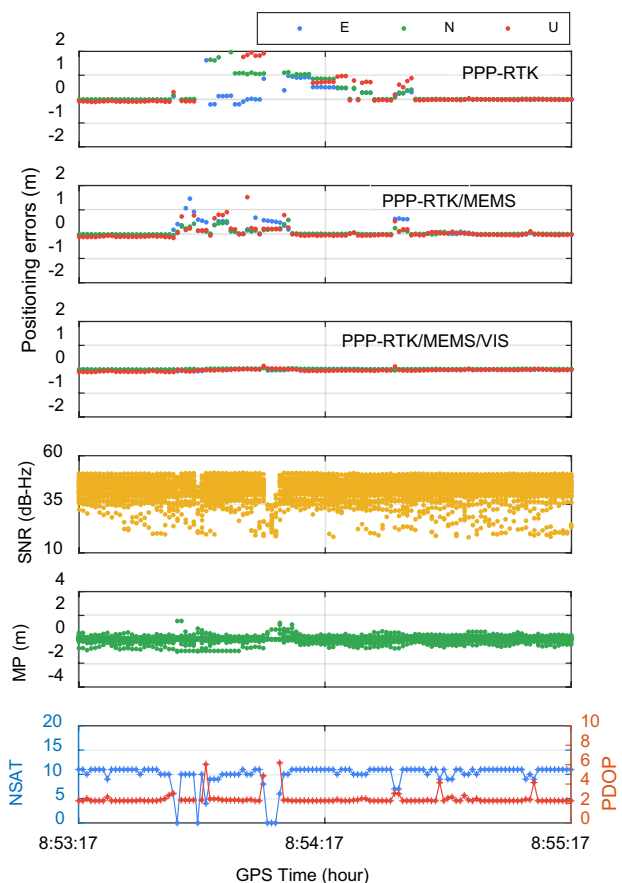


Fig. 14 Positioning series of PPP-RTK, PPP-RTK/MEMS and PPP-RTK/MEMS/VIS in Scene 2. NSAT, PDOP, SNR and multi-path series are also shown in the figure

Table 5 Refixing time of PPP-RTK, PPP-RTK/MEMS and PPP-RTK/MEMS/VIS in Scene 2

Modes	Refixing time (s)		
	①	②	③
PPP-RTK	Failed	Failed	17
PPP-RTK/MEMS	Failed	Failed	3
PPP-RTK/MEMS/VIS	< 1	< 1	< 1

visual measurement updates contribute to the estimation of the IMU biases and restraining INS divergence. The positioning accuracy of PPP-RTK/MEMS/vision in this urban canyon environment is 0.07, 0.08 and 0.14 m in the east, north and vertical components, respectively.

The ratio values in ambiguity resolution in three processing modes are presented in Fig. 12. The ratio is generally defined as the ratio of the second minimum quadratic form of the residuals to the minimum quadratic form of the residuals (Han et al. 1997). In principle, the ambiguity set corresponding to the minimum quadratic form of the

residuals should be the correct one. Therefore, the ratio value can be considered an index to denote the reliability of ambiguity resolution, and larger ratio values generally refer to more reliable ambiguity resolution. As shown in this figure, the PPP-RTK/MEMS exhibits higher ratio values than PPP-RTK solutions. The average ratio values of PPP-RTK and PPP-RTK/MEMS are 15.0 and 20.9, respectively. It also can be seen that the PPP-RTK/MEMS/vision integration brings significant improvements in terms of fixing numbers and ratio values in the complex signal conditions (8:27:00–8:29:30). The average ratio value for PPP-RTK/MEMS/vision is improved to 27.7. It is reasonable since the tightly integrated PPP-RTK/MEMS/vision solution could enhance the float ambiguity estimation, reducing the search space of integer ambiguity and making it easier to identify the correct ambiguity.

The second scene includes three overpasses (8:53:17–8:55:17). Shown in the top panel of Fig. 13 presents the vehicle trajectory overlaid on Google Maps in which ①–③ represent three different overpasses that the vehicle crosses consecutively. The specific observational environment can be seen in the right panel. This scene can be regarded as a situation where GNSS is frequently disturbed by obstacles.

The positioning series and the NSAT, PDOP, SNR and multi-path series are shown in Fig. 14. The NSAT during the period of Scene 2 is approximately 10. The corresponding SNR values are in the range of 20–50 dB Hz, and the multi-path errors are within 1 m mostly. As the vehicle crosses the overpass, the number of available GNSS satellites will decrease to zero immediately. The obvious deterioration can be observed for both the SNR and multi-path when crossing overpasses. At the same time, PPP-RTK solutions exhibit low availability, with positioning errors increasing rapidly to meter level. Ambiguity resolution cannot be implemented in the circumstance and still fails even if the GNSS signals are available after the outage because of unfavorable tracking geometry and poor signal quality. PPP-RTK/MEMS integration also faces restrictions because of the frequent signal interruptions and the corresponding positioning accuracy is decreased to decimeter level. Nevertheless, the positioning performance of PPP-RTK/MEMS tight integration is still much better than GNSS-only solutions in terms of accuracy, reliability and reconvergence. Note that the best performance can be obtained for PPP-RTK/MEMS/vision tight integration, which shows a continuous centimeter-level positioning series over time even in a GNSS-challenged environment.

The frequent interference of GNSS will lead to the reinitialization of ambiguities. Rapid reinitialization attaches great importance to the continuous and accurate positioning for vehicle navigation. Table 5 provides the required time of ambiguity refixing in three processing modes after each interruption. The marks ①, ② and ③ denote three interruption

processes when the vehicle crossed three overpasses consecutively. The first and second interruptions last one second and occur in 5 s. After that, continuous tracking of 8–10 available satellites for a period of 12 s is achieved until the third outage occurs. Results indicate that both PPP-RTK and PPP-RTK/MEMS fail to achieve ambiguity refixing during two consecutive interruptions. In addition, it takes 17 s for PPP-RTK to reobtain ambiguity resolution when the GNSS signals can be continuously tracked, whereas PPP-RTK/MEMS could achieve ambiguity recovery within 3 s. For a PPP-RTK/MEMS/vision tight integration system, an instantaneous ambiguity refixing can always be obtained for these three interruptions. The results are reasonable once we acknowledge that MEMS and vision sensors are immune to such signal interruptions and maintain high-accuracy positioning for a certain period of time. Such accurate location information will facilitate the rapid convergence of ambiguity parameters for rapid ambiguity refixing.

Conclusions

This study developed a tightly integrated PPP-RTK/MEMS/vision model aiming to achieve centimeter-level vehicle navigation in urban environments. The raw GNSS measurements (i.e., pseudorange and carrier phase measurements), MEMS-IMU data and raw stereo camera data are tightly integrated with a centralized EKF framework. The navigation performance of the proposed system is investigated in urban environments with different observation conditions and carefully compared with the PPP-RTK and PPP-RTK/INS integration modes, respectively.

The vehicular experiments indicate that PPP-RTK can achieve stable centimeter-level positioning for more than visible satellites, but faces restrictions in case of GNSS signal blockage, where the positioning errors increase to several meters. The tight integration of PPP-RTK and MEMS could improve the positioning performance in terms of accuracy and availability. However, the accuracy is still at the level of meters or decimeters, which hardly meets the accuracy requirement of vehicle navigation. By contrast, the proposed PPP-RTK/MEMS/vision tight integration system can obtain continuous and accurate positions in an urban environment with an accuracy of 4.1, 2.2 and 7.3 cm in the east, north

and up components, respectively, and fixing percentage of 96.8%, showing comparable performance respect to the tight integration of PPP-RTK and a tactical IMU. It demonstrates that using this multi-sensor system allows for high-precision positioning with a relatively low expense.

The detailed analysis in GNSS-challenged scenes validates the contribution of the inclusion the raw data of MEMS-IMU and cameras for continuous and accurate positioning. In an urban canyon environment where GNSS signals are deteriorated and blocked, the tight integration of PPP-RTK, MEMS and vision exploits the complementarity of each sensor and avoids information loss in the loosely coupled mode, which enables a horizontal accuracy of 7–8 cm with high fixing percentage. Moreover, the inclusion of the INS and vision plays an important role in ambiguity refixing after an outage. An instantaneous ambiguity refixing can be obtained when crossing three consecutive overpasses during a period of 25 s while the PPP-RTK solution requires 17 s.

Appendix 1: Jacobians of the estimated camera states

This section introduces the specific expression of Jacobians of estimated camera states. The linearized visual observation equations for a single feature f_j observed at the time step i can be written as:

$$\mathbf{r}_{vis,i}^j = \mathbf{z}_{vis,i}^j - \hat{\mathbf{z}}_{vis,i}^j = \mathbf{H}_{C_i}^j \delta \mathbf{x}_{C_i} + \mathbf{H}_{f_i}^j \delta \mathbf{p}_j^e + \mathbf{n}_{vis,i}^j \tag{19}$$

where $\delta \mathbf{x}_{C_i} = [\delta \theta_{C_i}^e, \delta \mathbf{p}_{C_i}^e]^T$ denotes the error state vector of the camera pose; $\delta \mathbf{p}_j^e$ is the feature position error; and $\mathbf{H}_{C_i}^j$ and $\mathbf{H}_{f_i}^j$ represent the corresponding Jacobians, respectively, which can be written as:

$$\mathbf{H}_{C_i}^j = \begin{pmatrix} \mathbf{H}_1(\hat{\mathbf{p}}_j^{C_{i,l}} \times) & -\mathbf{H}_1(\hat{\mathbf{R}}_{C_{i,l}}^e)^T \\ (\mathbf{H}_2 \cdot \mathbf{R}_{C_{i,l}}^{C_{i,r}})(\hat{\mathbf{p}}_j^{C_{i,l}} \times) & -(\mathbf{H}_2 \cdot \mathbf{R}_{C_{i,l}}^{C_{i,r}})(\hat{\mathbf{R}}_{C_{i,l}}^e)^T \end{pmatrix}, \tag{20}$$

$$\mathbf{H}_{f_i}^j = \begin{pmatrix} \mathbf{H}_1(\hat{\mathbf{R}}_{C_{i,l}}^e)^T \\ \mathbf{H}_2 \mathbf{R}_{C_{i,l}}^{C_{i,r}}(\hat{\mathbf{R}}_{C_{i,l}}^e)^T \end{pmatrix}$$

with
$$\mathbf{H}_1 = \frac{1}{(\hat{z}_j^{C_{i,l}})^2} \begin{pmatrix} \hat{z}_j^{C_{i,l}} & 0 & -\hat{x}_j^{C_{i,l}} \\ 0 & \hat{z}_j^{C_{i,l}} & -\hat{y}_j^{C_{i,l}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_2 = \frac{1}{(\hat{z}_j^{C_{i,r}})^2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \hat{z}_j^{C_{i,r}} & 0 & -\hat{x}_j^{C_{i,r}} \\ 0 & \hat{z}_j^{C_{i,r}} & -\hat{y}_j^{C_{i,r}} \end{pmatrix} \tag{21}$$

Considering that the same feature will be tracked by multiple consecutive camera poses, therefore the linearized visual observation vector for this feature can be obtained by stacking all the individual equations together:

$$\begin{bmatrix} u_{1,l}^j \\ \vdots \\ v_{k,r}^j \end{bmatrix} = \begin{bmatrix} \hat{X}_j^{C_{1,l}} / \hat{Z}_j^{C_{1,l}} \\ \vdots \\ \hat{Y}_j^{C_{k,r}} / \hat{Z}_j^{C_{k,r}} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{C_1}^j \\ \vdots \\ \mathbf{H}_{C_k}^j \end{bmatrix} \delta \mathbf{x}_C + \begin{bmatrix} \mathbf{H}_{f,1}^j \\ \vdots \\ \mathbf{H}_{f,k}^j \end{bmatrix} \delta \mathbf{p}_f^e + \mathbf{n}_{vis}^j \tag{22}$$

where k denotes the k -th camera pose in the sliding window.

Since the feature positions are computed with the camera poses, the uncertainty of feature position is thereby correlated with the camera pose in the estimator. To eliminate the correlation between the feature position and camera poses, the linearized observation vector is reformulated by projecting it on the left null space (\mathbf{A}_j^T) of the matrix \mathbf{H}_f^j , which can be rewritten as:

$$\mathbf{A}_j^T \mathbf{z}_{vis}^j = \mathbf{A}_j^T \hat{\mathbf{z}}_{vis}^j + \mathbf{A}_j^T \mathbf{H}_C^j \delta \mathbf{x}_C + \mathbf{n}_{vis}^j \tag{23}$$

In this way, the visual observation equation is independent of the errors of the estimated feature position; therefore, the EKF update can be performed.

Acknowledgements This work has been supported by the National Key Research and Development Program of China (2021YFB2501102).

Data availability The datasets collected in the vehicular test campaign are available on reasonable request from the corresponding author.

References

Angelino CV, Baraniello VR, Cicala L (2012) UAV position and attitude estimation using IMU, GNSS and camera. In: Proceedings of the IEEE ICIF 2012, Singapore, July 9–112, pp 735–742

Cao S, Lu X, Shen S (2021) GVINS: tightly coupled GNSS-visual-inertial for smooth and consistent state estimation. <http://arxiv.org/abs/2103.07899>

de Oliveira PS, Morel L, Fund F, Legros R, Monico JFG, Durand S, Durand F (2017) Modeling tropospheric wet delays with dense and sparse network configurations for PPP-RTK. *GPS Solut* 21(1):237–250

Furgale P, Rehder J, Siegwart R (2013) Unified temporal and spatial calibration for multi-sensor systems. In: Proceedings of the IEEE/RSJ ICIRS, Tokyo, Japan, November 3–7, pp 1280–1286

Gao Z, Shen W, Zhang H, Niu X, Ge M (2016) Real-time kinematic positioning of ins tightly aided multi-GNSS ionospheric constrained PPP. *Sci Rep* 6(1):30488

Gao Z, Ge M, Li Y, Chen Q, Zhang Q, Niu X, Zhang H, Shen W, Harald S (2018) Odometer, low-cost inertial sensors, and four-GNSS data to enhance PPP and attitude determination. *GPS Solut* 22(3):57

Ge M, Gendt G, Rothacher M, Shi C, Liu J (2008) Resolution of GPS carrier-phase ambiguities in precise point positioning with daily observations. *J Geod* 82(7):389–399

Geng J, Shi C (2017) Rapid initialization of real-time PPP by resolving undifferenced GPS and GLONASS ambiguities simultaneously. *J Geod* 91(4):361–374

Gu S, Dai C, Fang W, Zhang Z, Wang Y, Zhang Q, Lou Y, Niu X (2021) Multi-GNSS PPP/INS tightly coupled integration with atmospheric augmentation and its application in urban vehicle navigation. *J Geod* 95(6):64

Han S (1997) Quality-control issues relating to instantaneous ambiguity resolution for real-time GPS kinematic positioning. *J Geod* 71(6):351–361

Laurichesse D, Mercier F, Berthias JP, Broca B, Cerri L (2009) Integer ambiguity resolution on undifferenced GPS phase measurements and its application to PPP and satellite precise orbit determination. *Navigation* 56(2):135–149

Li X, Zhang X, Ge M (2011) Regional reference network augmented precise point positioning for instantaneous ambiguity resolution. *J Geod* 85(3):151–158

Li X, Ge M, Douša J, Wickert J (2014) Real-time precise point positioning regional augmentation for large GPS reference networks. *GPS Solut* 18(1):61–71

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 Sens* 11(6):610

Li X, Huang J, Li X, Lyu H, Wang B, Xiong Y, Xie W (2021a) Multi-constellation GNSS PPP instantaneous ambiguity resolution with precise atmospheric corrections augmentation. *GPS Solut*. <https://doi.org/10.1007/s10291-021-01123-0>

Li X, Li X, Huang J, Shen Z, Wang B, Yuan Y, Zhang K (2021b) Improving PPP-RTK in urban environment by tightly coupled integration of GNSS and INS. *J Geod* 95(12):132

Li X, Wang B, Li X, Huang J, Lyu H, Han X (2022) Principle and performance of multi-frequency and multi GNSS PPP RTK. *Satell Navig*. <https://doi.org/10.1186/s43020-022-00068-0>

Liu S, Sun F, Zhang L, Li W, Zhu X (2016) Tight integration of ambiguity-fixed PPP and INS: model description and initial results. *GPS Solut* 20(1):39–49

Lynen S, Achtelik MW, Weiss S, Chli M, Siegwart R (2013) A robust and modular multi-sensor fusion approach applied to MAV navigation. In: Proceedings of the IEEE/RSJ ICIRS, Tokyo, Japan, November 3–7, pp 3923–3929

Mascaro R, Teixeira L, Hinzmann T, Siegwart R, Chli M (2018) GOMSF: graph-optimization based multi-sensor fusion for robust UAV pose estimation. In: Proceedings of the IEEE ICRA 2018, Brisbane, Australia, May 21–25, pp 1421–1428

Mur-Artal R, Tardós JD (2017) ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Trans Robot* 33(5):1255–1262

Psychas D, Verhagen S (2020) Real-time PPP-RTK performance analysis using ionospheric corrections from multi-scale network configurations. *Sensors* 20(11):3012

Qin T, Li P, Shen S (2018) VINS-mono: a robust and versatile monocular visual-inertial state estimator. *IEEE Trans Robot* 34(4):1004–1020

Qin T, Cao S, Pan J, Shen S (2019) A general optimization-based framework for global pose estimation with multiple sensors. <http://arxiv.org/1901.03642>

Rabbou MA, El-Rabbany A (2014) Tightly coupled integration of GPS precise point positioning and MEMS-based inertial systems. *GPS Solut* 19(4):601–609

Rabbou MA, El-Rabbany A (2015) Integration of multi-constellation GNSS ISE point positioning and mems-based inertial systems using tightly coupled mechanization. *Positioning* 6(4):81–95

Ren X, Chen J, Li X, Zhang X (2021) Ionospheric total electron content estimation using GNSS carrier phase observations based on

zero-difference integer ambiguity: methodology and assessment. *IEEE Trans Geosci Remote Sens* 59(1):1–14

Reuper B, Becker M, Leinen S (2018) Benefits of multi-constellation/multi-frequency GNSS in a tightly coupled GNSS/IMU/odometry integration algorithm. *Sensors* 18(9):3052

Shin EH (2005) Estimation techniques for low-cost inertial navigation. UCGE report, p 20219

Sun K, Mohta K, Pfrommer B, Watterson M, Liu S, Mulgaonkar Y, Taylor CJ, Kumar V (2018) Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robot Autom Lett* 3(2):965–972

Teunissen PJG (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. *J Geod* 70(1–2):65–82

Teunissen PJG, Khodabandeh A (2015) Review and principles of PPP-RTK methods. *J Geod* 89(3):217–240

Wübbena G, Schmitz M, Bagg A (2005) PPP-RTK: precise point positioning using state-space representation in RTK networks. In: Proceedings of the ION GNSS 2005, Long Beach, CA, September 13–16, pp 2584–2594

Yang Y, Ding Q, Gao W, Li J, Xu Y, Sun B (2022) Principle and performance of BDSBAS and PPP-B2b of BDS-3. *Satell Navig*. <https://doi.org/10.1186/s43020-022-00066-2>

Zhang X, Li P (2013) Assessment of correct fixing rate for precise point positioning ambiguity resolution on a global scale. *J Geod* 87(6):579–589

Zhang X, Zhu F, Zhang Y, Mohamed F, Zhou W (2018) The improvement in integer ambiguity resolution with ins aiding for kinematic precise point positioning. *J Geod* 93(7):993–1010

Zhang B, Chen Y, Yuan Y (2019) PPP-RTK based on undifferenced and uncombined observations: theoretical and practical aspects. *J Geod* 93(7):1011–1024

Zhang X, Ren X, Chen J, Zuo X, Mei D, Liu W (2022) Investigating GNSS PPP-RTK with external ionospheric constraints. *Satell Navig* 3(6):1–13

Zuo X, Geneva P, Lee W, Liu Y, Huang G (2019) LIC-fusion: lidar-inertial-camera odometry. In: Proceedings of the IEEE/RSJ IROS 2019, Macau, China, November 04–08, pp 5848–5854

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Xin Li is currently a postdoctoral at Wuhan University. She received a Master's and Ph.D. degree at the School of Geodesy and Geomatics at Wuhan University in 2015 and 2021. Her area of research currently focuses on GNSS precise positioning and multi-sensor navigation.



Xingxing Li has completed his B.Sc. degree at the School of Geodesy and Geomatics in Wuhan University and obtained his Ph.D. degree at the Department of Geodesy and Remote Sensing of the German Research Center for Geosciences (GFZ). He is currently a professor at Wuhan University. His current research mainly involves precise GNSS data processing and geosciences application.



Shengyu Li received a B.S. degree in geodesy and survey engineering from Wuhan University, Wuhan, China, in 2019. He is currently a master's candidate at the school of Geodesy and Geomatics, Wuhan University. His current research focuses on integrated navigation systems and multi-sensor fusion.



Yuxuan Zhou received a B.S. degree in geodesy and survey engineering from Wuhan University in 2019, and is currently a master's candidate at the School of Geodesy and Geomatics, Wuhan University. His current research focuses on integrated navigation systems and multi-sensor fusion.



Meng Sun is currently a bachelor's candidate at Wuhan University. Her area of research currently focuses on integrated navigation systems and multi-sensor fusion.



Zhili Xu is currently a bachelor's candidate at Wuhan University. His area of research currently focuses on GNSS precise positioning.



Qi Xu is currently a bachelor's candidate at Wuhan University. His area of research currently focuses on GNSS precise positioning.