

Square-Root Inverse Filter-based GNSS-Visual-Inertial Navigation

Jun Hu[†], Xiaoming Lang[†], Feng Zhang[†], Yinian Mao[†], and Guoquan Huang^{†‡}

Abstract—While Global Navigation Satellite System (GNSS) is often used to provide global positioning if available, its intermittency and/or inaccuracy calls for fusion with other sensors. In this paper, we develop a novel GNSS-Visual-Inertial Navigation System (GVINS) that fuses visual, inertial, and raw GNSS measurements within the square-root inverse sliding window filtering (SRI-SWF) framework in a tightly coupled fashion, which thus is termed SRI-GVINS. In particular, for the first time, we deeply fuse the GNSS pseudorange, Doppler shift, single-differenced pseudorange, and double-differenced carrier phase measurements, along with the visual-inertial measurements. Inherited from the SRI-SWF, the proposed SRI-GVINS gains significant numerical stability and computational efficiency over the start-of-the-art methods. Additionally, we propose to use a filter to sequentially initialize the reference frame transformation till converges, rather than collecting measurements for batch optimization. We also perform online calibration of GNSS-IMU extrinsic parameters to mitigate the possible extrinsic parameter degradation. The proposed SRI-GVINS is extensively evaluated on our own collected UAV datasets and the results demonstrate that the proposed method is able to suppress VIO drift in real-time and also show the effectiveness of online GNSS-IMU extrinsic calibration. The experimental validation on the public datasets further reveals that the proposed SRI-GVINS outperforms the state-of-the-art methods in terms of both accuracy and efficiency.

I. INTRODUCTION AND RELATED WORK

In recent years, visual-inertial odometry (VIO) or visual-inertial navigation system (VINS) has seen increasing popularity in various fields such as autonomous driving, augmented/virtual reality (AR/VR), and aerial vehicles [1]. VIO aims to fuse camera measurements and Inertial Measurement Unit (IMU) data to estimate 3D motion of the sensing platform including the orientation, position and velocity and its algorithms can be broadly categorized into optimization-based methods [2]–[5] and filter-based methods [6]–[8].

Despite its success, VIO can easily accumulate drifts and cannot provide global localization due to the four unobservable directions corresponding to the global position and yaw [9], [10]. On the other hand, Global Navigation Satellite System (GNSS) measurements can offer absolute measurements in the global frame if available and can be leveraged to correct the VIO drift. By fusing VIO and GNSS, it is possible to achieve drift-free global localization as well as consistent local trajectories.

There are two primary approaches for fusing VIO and GNSS: loosely-coupled and tightly-coupled. In the loosely-coupled approach, VIO and GNSS are processed inde-

pendently, as demonstrated in various studies [11]–[20]. Some researchers [11], [12] have developed state estimation systems that integrate GNSS solutions with visual and inertial data using the extended Kalman filter (EKF). [14] integrated visual odometry with GNSS pseudorange measurements using an EKF framework. Other researchers [15]–[18] employed an optimization framework to fuse the results from local VIO with GNSS solutions. Lee et al. [19] proposed a multi-state constraint Kalman filter (MSCKF)-based estimator to optimally fuse inertial, camera, and asynchronous GNSS position and perform online calibration of the GNSS-IMU extrinsic and time offset. However, most of the aforementioned works rely on GNSS solutions for estimation and are unable to provide information in scenarios where the number of tracked satellites is less than four.

In the tightly (or deeply)-coupled approach, VIO and GNSS are integrated within a joint optimization framework [21]–[28]. This framework simultaneously optimizes GNSS raw measurements (pseudorange, carrier phase, or Doppler shift), visual, and inertial information to minimize their errors. [23] and [24] have proposed optimization-based approaches that jointly fuse inertial, camera feature reprojection, GNSS pseudorange, and Doppler shift measurements in a sliding window. [25] employed the ionosphere-free (IF) model with dual-frequency observations and incorporated phase ambiguity into the estimated states to eliminate ionospheric effects and utilize carrier phase measurements. To enhance computational efficiency, some filter-based implementations [26]–[28] have been proposed. Lee et al. [26] model raw measurement uncertainties by canceling atmospheric effects. Liu et al. [28] proposed an invariant filter-based platform, which derives the degenerate motions and proves intrinsic consistency. However, it should be noted that [28] does not support online GNSS-IMU extrinsic calibration. Additionally, these filter-based methods do not interpolate velocity at the time of Doppler shift measurement, which can introduce unmodeled errors.

Note that both filtering-based [8] and optimization-based [2] VIO methods could have numerical issues [29], primarily due to the fact that if the condition number of the covariance/Hessian matrix is larger than 10^9 , it can result in numerical errors and lead to poor estimation accuracy or even divergence. In many embedded systems, single-precision arithmetic is preferred over double-precision arithmetic due to its lower hardware resource requirements and faster operations. Motivated by the superior numerical stability and potential speed gains of square-root methods [30], Wu et al. [31] proposed a square-root inverse sliding window filter (SRI-SWF) for VIO. Given the stringent resources available

[†] Meituan UAV, Beijing, China (e-mail: hujun11 | langxiaoming | zhangfeng53 | maoyinian | huanguoquan@meituan.com). [‡] Dept. of Mechanical Engineering, Computer and Information Sciences, University of Delaware, Newark, DE.

on unmanned aerial vehicles (UAVs) which we are interested in this work, the SRI-SWF becomes a promising solution.

In this paper, we propose a tightly-coupled GNSS-Visual-Inertial Navigation System (GVINS) that *deeply* fuses GNSS raw measurements within an efficient square-root filtering framework. In particular, our main contributions include:

- For the first time, we deeply fuse the raw GNSS data including pseudorange, Doppler shift, single-differenced pseudorange and double-differenced carrier phase, with IMU and camera measurements, within the SRI-SWF framework, in order to provide drift-free global localization. The proposed SRI-GVINS offers compelling numerical stability and computational efficiency.
- We develop a filter-based initialization approach to sequentially and adaptively converge the transformation between the reference frame and the VIO frame. Furthermore, we leverage IMU integration to update the state using delayed and asynchronous GNSS measurements. Additionally, we perform online GNSS-IMU extrinsic calibration to mitigate the adverse effects of extrinsic parameter degradation.
- We validate the proposed SRI-GVINS on UAV real-world datasets, demonstrating its ability to eliminate VIO drift in real-time and the effectiveness of online GNSS-IMU extrinsic calibration. Moreover, the experimental results on the public datasets reveal that the proposed SRI-GVINS achieves superior accuracy and efficiency compared to the state-of-the-art methods such as GVINS [23] and InGVIO [28].

II. SQUARE-ROOT INVERSE SLIDING-WINDOW FILTER

In this section, we briefly review the SRI-SWF [31], which serves as the foundation for the proposed SRI-GVINS whose architecture is shown in Fig. 1.

At each time step k , the state vector \mathbf{x}_k comprises the current SLAM features \mathbf{x}_S , historical IMU pose clones \mathbf{x}_C , parameter state vector \mathbf{x}_P , and IMU extra state \mathbf{x}_E .

$$\mathbf{x}_k = [\mathbf{x}_S^\top \ \mathbf{x}_C^\top \ \mathbf{x}_P^\top \ \mathbf{x}_E^\top]^\top, \quad \mathbf{x}_S = [\mathbf{p}_{f_1}^\top \ \dots \ \mathbf{p}_{f_n}^\top]^\top \quad (1)$$

$$\mathbf{x}_C = [\mathbf{x}_{C_{k-M+1}}^\top \ \dots \ \mathbf{x}_{C_k}^\top]^\top, \quad \mathbf{x}_{C_i} = [{}^W \mathbf{q}_{I_i}^\top \ {}^W \mathbf{p}_{I_i}^\top]^\top \quad (2)$$

$$\mathbf{x}_P = [{}^C \mathbf{q}_I^\top \ {}^C \mathbf{p}_I^\top \ {}^I \mathbf{p}_r^\top \ \delta \mathbf{t}_0^r \ \dots \ \delta \mathbf{t}_{N-1}^r \ \delta \dot{\mathbf{t}}^r]^\top \quad (3)$$

$$\mathbf{x}_E = [{}^W \mathbf{v}_{I_k}^\top \ \mathbf{b}_{a_k}^\top \ \mathbf{b}_{g_k}^\top]^\top \quad (4)$$

where \mathbf{p}_{f_j} for $j = 1, \dots, n$ denotes the state of point feature \mathbf{f}_j , represented by the inverse-depth parameterization [32] with respect to its first observing camera pose within the current sliding window. \mathbf{x}_{C_i} represents the state of cloned IMU pose at time step i , where ${}^W \mathbf{q}_{I_i}$ is the quaternion describing the rotation from the IMU frame $\{I_i\}$ to the local VIO world frame $\{W\}$, and ${}^W \mathbf{p}_{I_i}$ is the position of $\{I_i\}$ in $\{W\}$. $[{}^C \mathbf{q}_I \ {}^C \mathbf{p}_I]$ is the IMU to camera extrinsic parameters, ${}^I \mathbf{p}_r$ is the GNSS receiver to IMU extrinsic parameters. $\delta \mathbf{t}_j^r$ for $j = 0, \dots, N-1$ denotes the receiver clock bias with respect to GNSS constellation j . $\delta \dot{\mathbf{t}}^r$ is the receiver clock drift rate, which is the same for each constellation. ${}^W \mathbf{v}_{I_k}$ is the velocity of $\{I\}$ in $\{W\}$, while \mathbf{b}_{a_k} and \mathbf{b}_{g_k} are the accelerometer and gyroscope biases at time step k .

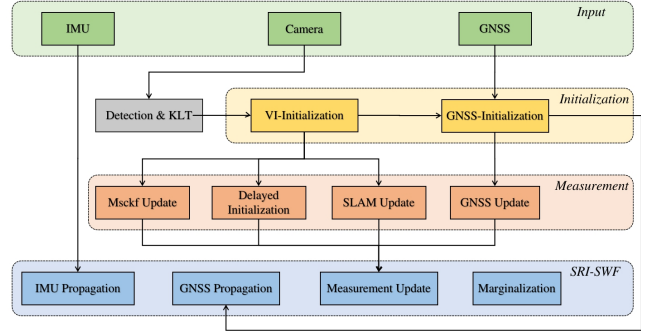


Fig. 1: Overview of the proposed SRI-GVINS system.

A. IMU Propagation

Given the inertial measurement $\mathbf{u}_{k,k+1} = [\boldsymbol{\omega}_{m_k}^T \ \mathbf{a}_{m_k}^T]^T$, we utilize the inertial kinematics [33] to propagate as:

$$\mathbf{x}_{I_{k+1}} = \mathbf{f}(\mathbf{x}_{I_k}, \mathbf{u}_{k,k+1} - \mathbf{w}_{k,k+1}) \quad (5)$$

where $\mathbf{x}_{I_k} = [\mathbf{x}_{C_k}^T \ \mathbf{x}_{E_k}^T]^T$, and $\mathbf{w}_{k,k+1}$ is the Gaussian noise affecting the IMU measurements with covariance \mathbf{Q}_k .

By linearizing Eq. (5) around the state estimates, we obtain the error-state propagation equation:

$$\tilde{\mathbf{x}}_{I_{k+1}} = \Phi_{k+1,k} \tilde{\mathbf{x}}_{I_k} + \mathbf{G}_{k+1,k} \mathbf{w}_{k,k+1} \quad (6)$$

where $\Phi_{k+1,k}$ and $\mathbf{G}_{k+1,k}$ are the corresponding Jacobians.

In the IMU propagation step, a new pose state $\mathbf{x}_{I_{k+1}}$ [see Eq. (5)] is added to the current state vector:

$$\mathbf{x}_{k+1}^\ominus = [\mathbf{x}_k^T \ \mathbf{x}_{I_{k+1}}^T]^T \quad (7)$$

The prior (upper-triangular) square-root information matrix and residual vector are defined as follows:

$$\mathbf{R}_{k+1}^\ominus = \begin{bmatrix} \mathbf{R}_k & \mathbf{0} \\ \mathbf{V}_1 & -\mathbf{Q}_k'^{-\frac{1}{2}} \end{bmatrix}, \quad \mathbf{r}_{k+1}^\ominus = \mathbf{0} \quad (8)$$

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{0} \cdots \mathbf{0} & \mathbf{Q}_k'^{-\frac{1}{2}} \Phi_{k+1,k}^{(C)} & \mathbf{0} \cdots \mathbf{0} & \mathbf{Q}_k'^{-\frac{1}{2}} \Phi_{k+1,k}^{(E)} \end{bmatrix}$$

where $\mathbf{Q}_k' = \mathbf{G}_{k+1,k} \mathbf{Q}_k \mathbf{G}_{k+1,k}^T$, $\Phi_{k+1,k}^{(C)}$ and $\Phi_{k+1,k}^{(E)}$ are the block columns of the Jacobian $\Phi_{k+1,k}$ with respect to the clone and extra IMU states, respectively. Note that \mathbf{R}_{k+1}^\ominus is not upper-triangular before the next marginalization step.

B. Marginalization

At each time step k , the past SLAM features $\tilde{\mathbf{x}}_{DS}$, the oldest clone $\tilde{\mathbf{x}}_{C_{k-M}}$, and the extra IMU state $\tilde{\mathbf{x}}_{E_{k-1}}$ are marginalized. The (error) state vector that includes all these states to be marginalized is defined as follows:

$$\tilde{\mathbf{x}}_k^M = [\tilde{\mathbf{x}}_{DS}^T \ \tilde{\mathbf{x}}_{C_{k-M}}^T \ \tilde{\mathbf{x}}_{E_{k-1}}^T]^T \quad (9)$$

We denote $\tilde{\mathbf{x}}_k^R$ as the remaining states, and through a permutation matrix, the state vector is transformed as follows:

$$\mathbf{P}_M \tilde{\mathbf{x}}_k^\ominus = \begin{bmatrix} \tilde{\mathbf{x}}_k^{M^T} & \tilde{\mathbf{x}}_k^{R^T} \end{bmatrix}^T \quad (10)$$

After performing a full QR decomposition, as described in [31], we obtain the prior(upper-triangular) square-root information matrix and residual vector, with only $\tilde{\mathbf{x}}_k^R$ retained.

C. Measurement Update

In general, a measurement can be modeled as follows:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{n}, \quad \text{with } \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (11)$$

where \mathbf{h} is a nonlinear function that describes the measurement model, \mathbf{z} is the obtained measurement (e.g., SLAM feature reobservations and MSCKF feature measurements), \mathbf{n} and Σ represent the white Gaussian noise of the measurement and its covariance, respectively.

By linearizing Eq. (11) around the state estimate $\hat{\mathbf{x}}$, we obtain the linearized measurement constraint $\mathbf{r} \simeq \mathbf{H}\tilde{\mathbf{x}} + \mathbf{n}$, where $\mathbf{H} = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}}$ and $\mathbf{r} = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})$ represent the measurement Jacobian matrix and residual, respectively. Finally, we can update the state by performing a QR decomposition:

$$\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \Sigma^{-\frac{1}{2}}\mathbf{H} & \Sigma^{-\frac{1}{2}}\mathbf{r} \end{bmatrix} \stackrel{\text{QR}}{=} \mathbf{Q} \begin{bmatrix} \mathbf{R}^\oplus & \mathbf{r}^\oplus \\ \mathbf{0} & \mathbf{e} \end{bmatrix} \quad (12)$$

$$\mathbf{R} \leftarrow \mathbf{R}^\oplus, \quad \Delta \mathbf{x} = \mathbf{R}^{\oplus^{-1}} \mathbf{r}^\oplus \quad (13)$$

III. SRI-GVINS: INITIALIZATION AND UPDATE

Within the SRI-SWF framework, as shown in Fig. 1, In addition to the visual-inertial measurements as in VIO, the proposed SRI-GVINS deeply fuses the GNSS measurements of pseudorange, Doppler shift, single-differenced pseudorange and double-differenced carrier phase. Therefore, we here focus on the measurement update with raw GNSS data.

A. Frame Initialization

The GNSS update requires the transformation from the local VIO World frame to the ECEF frame $\{^E\mathbf{R}_W \ ^E\mathbf{p}_W\}$, the receiver clock bias $[\delta t_0^r \dots \delta t_{N-1}^r]$, and the receiver clock drift rate $\delta \dot{\mathbf{t}}^r$ [see (3)]. To initialize the transform $\{^E\mathbf{R}_W \ ^E\mathbf{p}_W\}$, we insert the transform variables into the filter to adaptively converge the 4 degrees-of-freedom (DOF) transform. This approach differs from [19], which collects measurements of a desired trajectory length to solve a least squares estimation problem. The position and velocity of the receiver in VIO world frame can be expressed as:

$${}^W\mathbf{p}_r = {}^W\mathbf{p}_I + {}^W\mathbf{R}_I {}^I\mathbf{p}_r \quad (14)$$

$${}^W\mathbf{v}_r = {}^W\mathbf{v}_I + {}^W\mathbf{R}_I [{}^I\omega_I] {}^I\mathbf{p}_r \quad (15)$$

By utilizing the first GNSS measurement as the reference point, we employ GNSS Single Point Positioning (SPP) output to obtain the receiver's position and velocity in the North-East-Down frame (NED, $\{N\}$). At each time step k , we have the following geometric constraints:

$${}^N\mathbf{p}_{r_k} = {}^N\mathbf{p}_W + {}^N\mathbf{R}_W {}^W\mathbf{p}_{r_k}, \quad {}^N\mathbf{v}_{r_k} = {}^N\mathbf{R}_W {}^W\mathbf{v}_{r_k} \quad (16)$$

Note that the receiver's position and velocity at time step k can be predicted from the GNSS measurement at time step i through IMU integration. Since ${}^N\mathbf{R}_W$ is 1 DOF for the yaw angle, the transform states can be represented as $[\theta \ ^E\mathbf{p}_W]$.

By augmenting the state vector with the transform variables along with an infinite covariance prior, we perform the standard measurement update [see Eq. (16)]. We determine

that the transform has converged when the covariance matrix $P_{\theta\theta}$ for the yaw angle falls below a decision threshold of σ_θ^2 . After initialization, we marginalize the transform states. Given $\{{}^N\mathbf{R}_W \ ^N\mathbf{p}_W\}$ and $\{^E\mathbf{R}_N \ ^E\mathbf{p}_N\}$ computed from the reference point, the transform $\{^E\mathbf{R}_W \ ^E\mathbf{p}_W\}$ is obtained as:

$${}^E\mathbf{R}_W = {}^E\mathbf{R}_N {}^N\mathbf{R}_W, \quad {}^E\mathbf{p}_W = {}^E\mathbf{p}_N + {}^E\mathbf{R}_N {}^N\mathbf{p}_W \quad (17)$$

B. GNSS Receiver Clock Bias

We use the pseudorange and Doppler shift measurements to initialize the GNSS receiver clock bias $[\delta t_0^r \dots \delta t_{N-1}^r]$ and the receiver clock drift rate $\delta \dot{\mathbf{t}}^r$. To this end, we calculate the residual of the pseudorange measurement ${}^{s_j}p_{r_k}$ at time step k as follows:

$$\begin{aligned} \mathbf{z}_{p_{r_k}}{}^{s_j} = & \|{}^E\mathbf{p}_{s_j} - {}^E\mathbf{R}_W {}^W\mathbf{p}_{r_k} - {}^E\mathbf{p}_W\| - {}^{s_j}p_{r_k} \\ & + c(\delta t_{c_j}^r - \Delta t^{s_j}) + T_{r_k}^{s_j} + I_{r_k}^{s_j} + S_{r_k}^{s_j} + n_{pr} \end{aligned} \quad (18)$$

$c \approx 3.0 \times 10^8 \text{ m/s}$ denotes the speed of light. $\delta t_{c_j}^r$ and Δt^{s_j} refer to the clock bias of the receiver and satellite s_j , respectively. $T_{r_k}^{s_j}$, $I_{r_k}^{s_j}$ and $S_{r_k}^{s_j}$ represent the tropospheric delay, ionospheric delay and Sagnac term, respectively. n_{pr} represents the pseudorange measurement noise.

Similarly, the residual of the Doppler shift measurement ${}^{s_j}d_{r_k}$ is given by:

$$\begin{aligned} \mathbf{z}_{d_{r_k}}{}^{s_j} = & \frac{1}{\lambda} \kappa_{r_k}^{s_j T} ({}^E\mathbf{v}_{s_j} - {}^E\mathbf{R}_W {}^W\mathbf{v}_{r_k}) + \frac{c}{\lambda} (\delta \dot{\mathbf{t}}^r - \Delta \dot{t}^{s_j}) \\ & + {}^{s_j}d_{r_k} + n_d \end{aligned} \quad (19)$$

where λ represents the GNSS carrier wavelength. $\kappa_{r_k}^{s_j}$ is the unit direction vector from the receiver to satellite in $\{E\}$. ${}^E\mathbf{v}_{s_j}$ denotes the velocity of satellite s_j in $\{E\}$. $\delta \dot{\mathbf{t}}^r$ and $\Delta \dot{t}^{s_j}$ refer to the clock drift rate of the receiver and satellite s_j , respectively. n_d represents the measurement noise.

By using all the GNSS measurement residuals [see Eq. (18) and Eq. (19)], we solve the following linear constraint $\mathbf{A}\mathbf{X} = \mathbf{b}$, where \mathbf{X} is a stack of $[\delta t_0^r \dots \delta t_{N-1}^r]$ and $\delta \dot{\mathbf{t}}^r$. Initially, we apply the RANSAC algorithm [34] to eliminate any erroneous constraints. Subsequently, we solve the least squares problem mentioned above to obtain the initial guesses of the GNSS receiver clock bias and clock drift rate. Afterward, we further augment the initial values into the state through delayed initialization [35].

Once initialized, we propagate $[\delta t_0^r \dots \delta t_{N-1}^r]$ and $\delta \dot{\mathbf{t}}^r$ as follows:

$$\delta \mathbf{t}_k^r = \delta \mathbf{t}_{k-1}^r + \mathbf{1}_{N \times 1} \delta \dot{\mathbf{t}}^r (t_k - t_{k-1}), \quad \delta \dot{\mathbf{t}}_{t_k}^r = \delta \dot{\mathbf{t}}_{t_{k-1}}^r \quad (20)$$

The prior (upper-triangular) square-root information matrix [see Eq. (8)] is given by:

$$\begin{aligned} \mathbf{R}_{k+1}^\ominus = & \begin{bmatrix} \mathbf{R}_k & \mathbf{0} \\ \mathbf{V}_1 & -\mathbf{Q}_k'^{-\frac{1}{2}} \end{bmatrix}, \quad \mathbf{r}_{k+1}^\ominus = \mathbf{0} \\ \mathbf{V}_1 = & \begin{bmatrix} \mathbf{0} \dots \mathbf{0} & \mathbf{Q}_k'^{-\frac{1}{2}} \Phi_{k+1,k}^{(C)} & \mathbf{Q}_k'^{-\frac{1}{2}} \Phi_{k+1,k}^{(P)} & \mathbf{Q}_k'^{-\frac{1}{2}} \Phi_{k+1,k}^{(E)} \end{bmatrix} \end{aligned} \quad (21)$$

where $\Phi_{k+1,k}^{(P)}$ represents the block columns of the Jacobian $\Phi_{k+1,k}$ with respect to the parameter state.

C. GNSS Measurement Update

Due to the delay and asynchronous arrival of GNSS messages, the GNSS measurement time is typically earlier than the inertial state time. Since the velocity of each cloned pose is not preserved in the sliding window, it is not feasible to obtain the velocity state of the GNSS measurement time through interpolation [26]. Assuming the inertial state at time step i and the GNSS measurement at time step k , the following constraints hold through IMU integration:

$$\begin{aligned} {}^W\mathbf{R}_{I_k} &= {}^W\mathbf{R}_{I_i}(\gamma_{I_i}^{I_k})^T \\ {}^W\mathbf{v}_{I_k} &= {}^W\mathbf{v}_{I_i} - \mathbf{g}\Delta t - {}^W\mathbf{R}_{I_k}\beta_{I_i}^{I_k} \\ {}^W\mathbf{p}_{I_k} &= {}^W\mathbf{p}_{I_i} - {}^W\mathbf{v}_{I_k}\Delta t - \frac{1}{2}\mathbf{g}\Delta t^2 - {}^W\mathbf{R}_{I_k}\alpha_{I_i}^{I_k} \end{aligned} \quad (22)$$

where $\gamma_{I_i}^{I_k}$, $\beta_{I_i}^{I_k}$ and $\alpha_{I_i}^{I_k}$ are the pre-integration terms [36]. For inertial measurements within the time interval $[t_k, t_i]$, these terms are computed as follows:

$$\begin{aligned} \gamma_{I_i}^{I_k} &= \prod_{j=k}^{i-1} \text{Exp}((\boldsymbol{\omega}_j - \mathbf{b}_{w_j})\Delta t) \\ \beta_{I_i}^{I_k} &= \sum_{j=k}^{i-1} \gamma_{I_j}^{I_k}(\mathbf{a}_j - \mathbf{b}_{a_j})\Delta t \\ \alpha_{I_i}^{I_k} &= \sum_{j=k}^{i-1} \left[\beta_{I_j}^{I_k}\Delta t + \frac{1}{2}\gamma_{I_j}^{I_k}(\mathbf{a}_j - \mathbf{b}_{a_j})\Delta t^2 \right] \end{aligned} \quad (23)$$

1) *Pseudorange Measurement (Psr)*: We substitute Eq. (14) and Eq. (22) into Eq. (18) to define the complete residual of the pseudorange measurement.

2) *Doppler shift Measurement (Dopp)*: We substitute Eq. (15) and Eq. (22) into Eq. (19) to define the complete residual of the Doppler shift measurement.

3) *Single-differenced Pseudorange Measurement (Dpsr)*: Considering the presence of a base station, we can define the pseudorange measurement of the base station at time step k :

$$\begin{aligned} \mathbf{z}_{\mathbf{p}_{b_k}}^{s_j} &= \|\mathbf{p}_{s_j}^E - \mathbf{p}_{b_k}^E\| - s_j p_{b_k} + c(\delta t_{c_j}^b - \Delta t_b^{s_j}) \\ &\quad + T_{b_k}^{s_j} + I_{b_k}^{s_j} + S_{b_k}^{s_j} + n_{pb} \end{aligned} \quad (24)$$

When the base station and receiver are relatively close, the ionospheric and tropospheric delays should be approximately canceled by subtraction of Eq. [18] and Eq. [24]:

$$\begin{aligned} \mathbf{z}_{\mathbf{d}\mathbf{p}_{r_k}}^{s_j} &= \|\mathbf{p}_{s_j}^E - \mathbf{R}_W^E \mathbf{p}_{r_k}^W - \mathbf{p}_W^E\| - \|\mathbf{p}_{s_j}^E - \mathbf{p}_{b_k}^E\| \\ &\quad - s_j p_{r_k} + s_j p_{b_k} + c(\delta t_{c_j}^r - \Delta t^{s_j} - \delta t_{c_j}^b + \Delta t_b^{s_j}) \\ &\quad + T_{r_k}^{s_j} - T_{b_k}^{s_j} + I_{r_k}^{s_j} - I_{b_k}^{s_j} + S_{r_k}^{s_j} - S_{b_k}^{s_j} + n_{pr} - n_{pb} \end{aligned} \quad (25)$$

4) *Double-differenced Carrier Phase Measurement (Ddcp)*: The carrier phase measurements are often more precise than pseudorange measurements, the residual of the carrier phase measurement is given by:

$$\begin{aligned} \mathbf{z}_{\mathbf{c}\mathbf{p}_{r_k}}^{s_j} &= \|\mathbf{p}_{s_j}^E - \mathbf{R}_W^E \mathbf{p}_{r_k}^W - \mathbf{p}_W^E\| + \lambda N \\ &\quad - s_j c p_{r_k} + c(\delta t_{c_j}^r - \Delta t^{s_j}) + T_{r_k}^{s_j} - I_{r_k}^{s_j} + n_{cp} \end{aligned} \quad (26)$$

where N is the integer ambiguity, $s_j c p_{r_k}$ is the carrier phase measurement and n_{cp} is its noise. By using differential

measurements between base station and base satellite s_i , the residual of the double-differenced carrier phase measurement can be defined as:

$$\begin{aligned} \mathbf{z}_{d\mathbf{d}c\mathbf{p}}^k &= d_{s_j, r_k} - d_{s_j, b_k} - d_{s_i, r_k} + d_{s_i, b_k} + \lambda N_{dd} \\ &\quad - (s_j c p_{r_k} - s_j c p_{b_k} - s_i c p_{r_k} + s_i c p_{b_k}) + n_{d\mathbf{d}c\mathbf{p}} \end{aligned} \quad (27)$$

where $d_{s_j, r_k} = \|\mathbf{p}_{s_j}^E - \mathbf{p}_{r_k}^E\|$ and N_{dd} is the double-differenced integer ambiguity acquired from RTK algorithm.

Jacobians of these GNSS measurements can be computed:

$$\frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{I_i}} = \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{r_k}} \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{r_k}} \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{I_k}} \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{I_k}} \quad (28)$$

The overall measurement covariance matrix Σ is given by:

$$\Sigma = \Sigma_m + \mathbf{J}\Sigma_g\mathbf{J}^T, \mathbf{J} = \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{r_k}} \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{X}}_{I_k}} \quad (29)$$

where Σ_m is the GNSS raw measurement covariance matrix, and Σ_g is the noise from IMU integration measurements.

D. Online GNSS-IMU Calibration

The GNSS-IMU extrinsic is represented as \mathbf{p}_r in the state vector [see Eq. (3)]. We perform extrinsic estimation using the Doppler shift measurement and only update the extrinsic when the IMU excitation is sufficient. The Jacobian of the Doppler shift residual with respect to the extrinsic is:

$$\frac{\partial \tilde{\mathbf{z}}_{\mathbf{d}r_k}^{s_j}}{\partial \tilde{\mathbf{p}}_r} = \frac{\partial \tilde{\mathbf{z}}_{\mathbf{d}r_k}^{s_j}}{\partial \tilde{\mathbf{p}}_{r_k}} \frac{\partial \tilde{\mathbf{z}}_{\mathbf{d}r_k}^{s_j}}{\partial \tilde{\mathbf{p}}_{r_k}} + \frac{\partial \tilde{\mathbf{z}}_{\mathbf{d}r_k}^{s_j}}{\partial \tilde{\mathbf{v}}_{r_k}} \frac{\partial \tilde{\mathbf{z}}_{\mathbf{d}r_k}^{s_j}}{\partial \tilde{\mathbf{p}}_r} \quad (30)$$

IV. EXPERIMENTS ON UAV DATASETS

We first validate the proposed SRI-GVINS on our own collected UAV real-world datasets, the groundtruth is provided by an RTK-INS. The sensor platform consisted of two downward-looking cameras running at 20Hz with resolution 1920×1080, an IMU running at 250Hz and a Unicore UM982 GNSS receiver running at 5Hz. The real-time RTCM stream is broadcasted from a self-built base station located 3km away, which enables us to use single-differenced pseudorange and double-differenced carrier phase measurements.

A. GNSS Measurements

To investigate the effect of GNSS measurement update, we evaluated the performance of SRI-GVINS using different GNSS measurements. We selected 10 outdoor UAV datasets with varying heights ranging from 20m to 60m, trajectory lengths ranging from 400m to 3km, and flying speeds ranging from 4m/s to 12m/s. Fig. 2 shows the position errors of the 50m-3km-8m/s dataset.

1) *Psr and Dopp Measurements*: Since Psr [see Eq. (18)] and Dopp [see Eq. (19)] measurements do not require a base station, we conducted tests using the following configurations in the SRI-GVINS system: VIO no-fusion with GNSS measurements(VIO), VIO loosely-coupled fusion with GPS velocity(VIO+Gps_vel), VIO tightly-coupled fusion with Dopp(VIO+Dopp), and VIO tightly-coupled fusion with Dopp and Psr(VIO+Dopp+Psr). The root mean squared error (RMSE) [37] for translation and body velocity estimates on each dataset are shown in Table I.

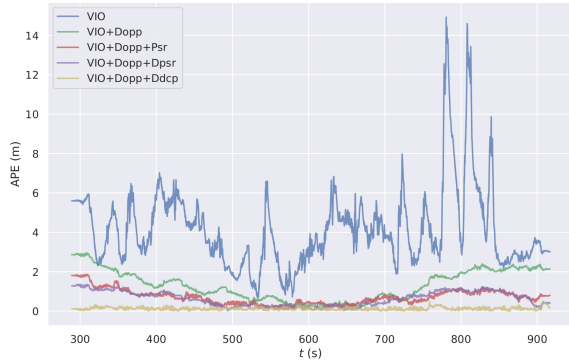


Fig. 2: Position errors of SRI-GVINS using different GNSS measurements in the 50m-3km-8m/s dataset.

TABLE I: RMSE with Psr and Dopp measurements

Height-Len -Vel	VIO +Dopp+Psr	VIO +Dopp	VIO +Gps_vel	VIO
60m-1km -6m/s	Tran. 0.652 Vel. 0.0813	Tran. 0.440 Vel. 0.0767	Tran. 0.656 Vel. 0.0978	Tran. 2.915 Vel. 0.2170
60m-1km -4m/s	Tran. 0.509 Vel. 0.0706	Tran. 0.492 Vel. 0.0709	Tran. 0.707 Vel. 0.0952	Tran. 1.758 Vel. 0.1842
60m-1km -8m/s	Tran. 0.522 Vel. 0.0933	Tran. 0.410 Vel. 0.0938	Tran. 0.595 Vel. 0.1043	Tran. 4.641 Vel. 0.3361
60m-1km -10m/s	Tran. 0.642 Vel. 0.0936	Tran. 0.4981 Vel. 0.0922	Tran. 0.854 Vel. 0.0983	Tran. 4.176 Vel. 0.3134
60m-400m -8m/s	Tran. 0.436 Vel. 0.0690	Tran. 0.276 Vel. 0.0684	Tran. 0.394 Vel. 0.0769	Tran. 2.628 Vel. 0.2527
50m-3km -4m/s	Tran. 0.707 Vel. 0.0756	Tran. 1.434 Vel. 0.0774	Tran. 0.902 Vel. 0.0868	Tran. 16.090 Vel. 0.2698
50m-3km -8m/s	Tran. 0.804 Vel. 0.0892	Tran. 1.522 Vel. 0.0869	Tran. 1.398 Vel. 0.0990	Tran. 4.634 Vel. 0.3017
50m-3km -12m/s	Tran. 0.694 Vel. 0.1006	Tran. 0.992 Vel. 0.0987	Tran. 1.171 Vel. 0.1207	Tran. 3.988 Vel. 0.3056
20m-1km -8m/s	Tran. 0.467 Vel. 0.1267	Tran. 0.381 Vel. 0.1401	Tran. 0.673 Vel. 0.1378	Tran. 8.614 Vel. 0.7148
40m-1km -8m/s	Tran. 0.516 Vel. 0.1050	Tran. 0.303 Vel. 0.1042	Tran. 0.493 Vel. 0.1006	Tran. 5.694 Vel. 0.4058
Average	Tran. 0.595 Vel. 0.0905	Tran. 0.675 Vel. 0.0909	Tran. 0.784 Vel. 0.1017	Tran. 5.514 Vel. 0.3301

From Table I, we can see that the VIO+Dopp+Psr method exhibited the highest average position and velocity accuracy, demonstrating its ability to eliminate VIO drift. It is also observed that the VIO+Dopp method has higher accuracy than the VIO+Gps_vel method. Additionally, it is interesting to note that fusing Psr on top of the VIO+Dopp fusion can improve position accuracy on long trajectory datasets, but may result in a slight decrease in position accuracy on short trajectory datasets.

2) *Dpsr and Ddcp Measurements*: Once a base station is available, we can perform Dpsr [see Eq. (25)] and Ddcp [see Eq. (27)] measurements update. We conducted experiments on the following configurations: VIO tightly-coupled fusion with Dopp and Dpsr(VIO+Dopp+Dpsr), as well as VIO tightly-coupled fusion with Dopp and Ddcp(VIO+Dopp+Ddcp).

The accuracy results are presented in Table II. It is evident that the VIO+Dopp+Ddcp method achieved the highest position accuracy when the integer ambiguity was successfully resolved by the RTK module. When the ambiguity resolution was unsuccessful, the VIO+Dopp+Dpsr method still exhibited higher position accuracy compared to the

TABLE II: RMSE with Dpsr and Ddcp measurements

Height-Len -Vel	VIO+Dopp +Ddcp	VIO+Dopp +Dpsr	VIO+Dopp +Psr	VIO+Dopp+ Psr wo.inte
60m-1km -6m/s	Tran. 0.118 Vel. 0.0771	Tran. 0.286 Vel. 0.0782	Tran. 0.652 Vel. 0.0813	Tran. 0.942 Vel. 0.1316
60m-1km -4m/s	Tran. 0.121 Vel. 0.0683	Tran. 0.306 Vel. 0.0677	Tran. 0.509 Vel. 0.0706	Tran. 0.750 Vel. 0.1136
60m-1km -8m/s	Tran. 0.156 Vel. 0.0910	Tran. 0.344 Vel. 0.0884	Tran. 0.522 Vel. 0.0933	Tran. 1.065 Vel. 0.1486
60m-1km -10m/s	Tran. 0.156 Vel. 0.1013	Tran. 0.419 Vel. 0.0945	Tran. 0.642 Vel. 0.0936	Tran. 1.208 Vel. 0.1646
60m-400m -8m/s	Tran. 0.123 Vel. 0.0672	Tran. 0.254 Vel. 0.0667	Tran. 0.436 Vel. 0.0690	Tran. 0.759 Vel. 0.1257
50m-3km -4m/s	Tran. 0.116 Vel. 0.0779	Tran. 0.387 Vel. 0.0764	Tran. 0.707 Vel. 0.0756	Tran. 0.767 Vel. 0.1027
50m-3km -8m/s	Tran. 0.149 Vel. 0.0942	Tran. 0.742 Vel. 0.0894	Tran. 0.804 Vel. 0.0892	Tran. 0.865 Vel. 0.1401
50m-3km -12m/s	Tran. 0.178 Vel. 0.0930	Tran. 0.440 Vel. 0.0982	Tran. 0.694 Vel. 0.1006	Tran. 0.994 Vel. 0.1808
20m-1km -8m/s	Tran. 0.162 Vel. 0.1355	Tran. 0.485 Vel. 0.1357	Tran. 0.467 Vel. 0.1267	Tran. 0.842 Vel. 0.1535
40m-1km -8m/s	Tran. 0.148 Vel. 0.1041	Tran. 0.441 Vel. 0.1067	Tran. 0.516 Vel. 0.1050	Tran. 0.878 Vel. 0.1458
Average	Tran. 0.143 Vel. 0.0910	Tran. 0.410 Vel. 0.0902	Tran. 0.595 Vel. 0.0905	Tran. 0.907 Vel. 0.1407

VIO+Dopp+Psr method. In terms of velocity accuracy, the fusion of Psr, Dpsr, and Ddcp measurements on top of the VIO+Dopp fusion resulted in comparable levels of accuracy.

We also performed an ablation experiment in which we excluded IMU integration for GNSS measurement update in the VIO+Dopp+Psr method. The results from Table II indicate a significant decrease in both position and velocity accuracy when IMU integration was not utilized.

B. Online GNSS-IMU Extrinsic Calibration

To investigate the impact of online GNSS-IMU extrinsic estimation, we conducted experiments using both accurate and poor initial GNSS-IMU extrinsic parameters. We also performed experiments with the GNSS-IMU extrinsic calibration enabled and disabled, separately.

Table III shows the accuracy results under different configurations. When given poor initial extrinsic parameters (with an error of 0.3m in the x-axis direction), enabling GNSS-IMU extrinsic estimation improved all velocity accuracy and almost all position accuracy. When given accurate initial extrinsic parameters, the effect of enabling GNSS-IMU extrinsic estimation was relatively minor, with only a slight improvement in average position and velocity accuracy.

Fig. 3 presents the estimation errors of GNSS-IMU extrinsic calibration when given poor initial GNSS-IMU extrinsic parameters. It can be observed that the estimation errors gradually converge to nearly zero. Additionally, the velocity errors with GNSS-IMU extrinsic calibration are significantly reduced compared to the case without extrinsic calibration. These results serve as validation that GNSS-IMU extrinsic estimation can effectively mitigate the impact of extrinsic parameter degradation and improve localization accuracy.

C. Real-Time Performance

Table IV presents the CPU usage and run time of the VIO+Dopp+Dpsr method. We replayed the datasets on a

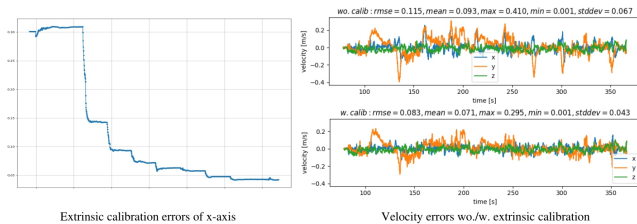


Fig. 3: Estimation errors of GNSS-IMU extrinsic calibration and velocity errors in the 60m-1km-6m/s dataset with poor initial GNSS-IMU extrinsic.

TABLE III: Online GNSS-IMU extrinsic estimation results

Height-Len	poor extr	poor extr	good extr	good extr
-Vel	+ w.calib	+ wo.calib	+ w.calib	+ wo.calib
60m-1km	Tran. 0.622	Tran. 0.888	Tran. 0.652	Tran. 0.650
-6m/s	Vel. 0.0833	Vel. 0.1150	Vel. 0.0813	Vel. 0.0859
60m-1km	Tran. 0.476	Tran. 0.690	Tran. 0.509	Tran. 0.518
-4m/s	Vel. 0.0718	Vel. 0.1069	Vel. 0.0706	Vel. 0.0665
60m-1km	Tran. 0.611	Tran. 0.994	Tran. 0.522	Tran. 0.483
-8m/s	Vel. 0.0958	Vel. 0.1328	Vel. 0.0933	Vel. 0.0913
60m-1km	Tran. 0.642	Tran. 1.083	Tran. 0.642	Tran. 0.652
-10m/s	Vel. 0.0938	Vel. 0.1337	Vel. 0.0936	Vel. 0.0965
60m-400m	Tran. 0.391	Tran. 0.479	Tran. 0.436	Tran. 0.426
-8m/s	Vel. 0.0751	Vel. 0.0963	Vel. 0.0690	Vel. 0.0695
50m-3km	Tran. 0.781	Tran. 0.654	Tran. 0.707	Tran. 0.717
-4m/s	Vel. 0.0769	Vel. 0.1104	Vel. 0.0756	Vel. 0.0749
50m-3km	Tran. 0.957	Tran. 0.810	Tran. 0.804	Tran. 0.945
-8m/s	Vel. 0.1120	Vel. 0.1403	Vel. 0.0892	Vel. 0.0991
50m-3km	Tran. 0.707	Tran. 0.585	Tran. 0.694	Tran. 0.646
-12m/s	Vel. 0.0983	Vel. 0.1417	Vel. 0.1006	Vel. 0.0953
20m-1km	Tran. 0.513	Tran. 1.486	Tran. 0.467	Tran. 0.633
-8m/s	Vel. 0.1296	Vel. 0.1671	Vel. 0.1267	Vel. 0.1239
40m-1km	Tran. 0.475	Tran. 0.777	Tran. 0.516	Tran. 0.536
-8m/s	Vel. 0.1032	Vel. 0.1394	Vel. 0.1050	Vel. 0.1040
Average	Tran. 0.618	Tran. 0.844	Tran. 0.595	Tran. 0.621
	Vel. 0.0940	Vel. 0.1284	Vel. 0.0905	Vel. 0.0907

Laptop with an Intel Core i7-10750U 2.60GHz \times 12 core CPU and 16GB memory, and conducted real-flight tests on a UAV platform featuring a Qualcomm RB5 processor. As shown in Table IV, the fusion of Dopp and Dpsr measurements increases the estimator time by 13.77% compared to the VIO method, accounting for 2.76% of the total processing time. The CPU usage only increased from 22.77% to 23.01%, indicating that the computational cost brought by GNSS measurement update is almost negligible. Furthermore, the real-flight tests demonstrated that SRI-GVINS can achieve real-time performance even on resource-constrained devices.

V. EXPERIMENTS ON PUBLIC DATASETS

SRI-GVINS was evaluated and compared with two other state-of-the-art methods, GVINS [23] and InGVIO [28], using three public datasets provided by [23], which contain IMU, stereo images, Dopp and Psr measurements. The comparison results are presented in Table V, which shows the accuracy and efficiency of the algorithms. Fig. 4 shows the trajectories and position errors of the three algorithms,

TABLE IV: Timing results of SRI-GVINS on laptop and Qualcomm RB5

	Laptop	Laptop	Qualcomm RB5
	VIO	SRI-GVINS	SRI-GVINS
CPU usage(%)	22.77	23.01	49.79
Estimator time(ms)	2.230	2.537	6.129
Total time(ms)	8.525	8.760	34.739

TABLE V: Translational RMSE(m) and average time(ms) comparison

	GVINS	InGVIO	SRI-GVINS
sports.field	Trans. 0.578	Trans. 0.566	Trans. 0.717
	Time. 24.858	Time. 5.136	Time. 3.372
complex_environment	Trans. 4.376	Trans. 3.063	Trans. 2.783
	Time. 19.340	Time. 4.363	Time. 3.703
urban_driving	Trans. -	Trans. 5.020	Trans. 4.548
	Time. 24.858	Time. 3.202	Time. 2.959

and it should be noted that any trajectory that is affected by unreliable RTK data has been excluded from the analysis.

From Table V, SRI-GVINS exhibits competitive accuracy owing to its superior numerical stability and the GNSS measurement update with IMU integration. Moreover, SRI-GVINS achieves the lowest computational time, thanks to the efficiency gained through the utilization of the square-root form.

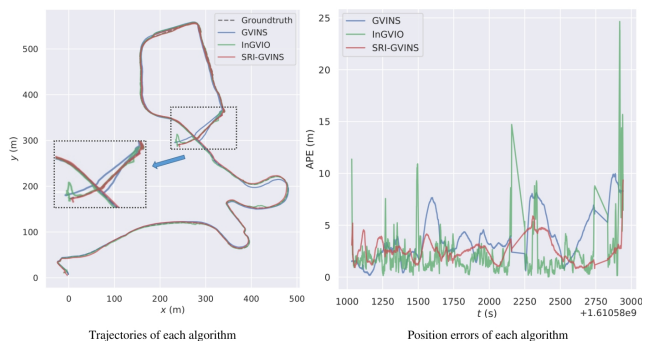


Fig. 4: Trajectories and position errors of each algorithm in the complex_environment dataset.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have developed a tightly-coupled GNSS-Visual-Inertial navigation system that deeply fuses visual, inertial, and raw GNSS data (including pseudorange, Doppler shift, single-differenced pseudorange and double-differenced carrier phase) within the efficient SRI-SWF framework. We have used a filter-based initialization approach to adaptively converge the reference frame transformation, leveraged IMU integration to update the state using delayed and asynchronous GNSS measurement, and performed online calibration of GNSS-IMU extrinsic parameters. Extensive experimental results demonstrate that the proposed SRI-GVINS outperforms other state-of-the-art methods in terms of accuracy and efficiency. In the future, we will use raw GNSS measurements such as Doppler shift to assist in the initialization of the reference frame transform. Additionally, we will directly employ the LAMBDA algorithm [38] to solve for the double-differenced ambiguity in SRI-GVINS. We will also perform an observability analysis for SRI-GVINS to identify any potential degenerate motions.

ACKNOWLEDGEMENT

The authors would like to thank Yunjun Shen, Dan Su and Jingqi Jiang for their contributions to the development of the VIO framework based on the SRI-SWF. Additionally, appreciation is extended to Qian Zhang for her efforts in preprocessing the raw GNSS data.

REFERENCES

- [1] G. Huang, "Visual-inertial navigation: A concise review," in *2019 international conference on robotics and automation (ICRA)*, pp. 9572–9582, IEEE, 2019.
- [2] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [3] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.
- [4] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.
- [5] L. Von Stumberg and D. Cremers, "Dm-vio: Delayed marginalization visual-inertial odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1408–1415, 2022.
- [6] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE international conference on robotics and automation*, pp. 3565–3572, IEEE, 2007.
- [7] M. Li and A. I. Mourikis, "High-precision, consistent ekf-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.
- [8] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang, and G. Huang, "Opencvins: A research platform for visual-inertial estimation," pp. 4666–4672, 2020.
- [9] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.
- [10] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Observability-constrained vision-aided inertial navigation," *University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Tech. Rep*, vol. 1, p. 6, 2012.
- [11] C. V. Angelino, V. R. Baraniello, and L. Cicala, "Uav position and attitude estimation using imu, gnss and camera," in *2012 15th International Conference on Information Fusion*, pp. 735–742, IEEE, 2012.
- [12] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to mav navigation," in *2013 IEEE/RSJ international conference on intelligent robots and systems*, pp. 3923–3929, IEEE, 2013.
- [13] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, "Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft mav," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4974–4981, IEEE, 2014.
- [14] M. Schreiber, H. Königshof, A.-M. Hellmund, and C. Stiller, "Vehicle localization with tightly coupled gnss and visual odometry," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, pp. 858–863, IEEE, 2016.
- [15] R. Mascaro, L. Teixeira, T. Hinzmam, R. Siegwart, and M. Chli, "Gomsf: Graph-optimization based multi-sensor fusion for robust uav pose estimation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1421–1428, IEEE, 2018.
- [16] Y. Yu, W. Gao, C. Liu, S. Shen, and M. Liu, "A gps-aided omnidirectional visual-inertial state estimator in ubiquitous environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7750–7755, IEEE, 2019.
- [17] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," *arXiv preprint arXiv:1901.03642*, 2019.
- [18] Z. Gong, P. Liu, F. Wen, R. Ying, X. Ji, R. Miao, and W. Xue, "Graph-based adaptive fusion of gnss and vio under intermittent gnss-degraded environment," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–16, 2020.
- [19] W. Lee, K. Eickenhoff, P. Geneva, and G. Huang, "Intermittent gps-aided vio: Online initialization and calibration," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5724–5731, IEEE, 2020.
- [20] L. Xiong, R. Kang, J. Zhao, P. Zhang, M. Xu, R. Ju, C. Ye, and T. Feng, "G-vido: A vehicle dynamics and intermittent gnss-aided visual-inertial state estimator for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11845–11861, 2021.
- [21] D. H. Won, E. Lee, M. Heo, S. Sung, J. Lee, and Y. J. Lee, "Gnss integration with vision-based navigation for low gnss visibility conditions," *GPS solutions*, vol. 18, pp. 177–187, 2014.
- [22] T. Li, H. Zhang, Z. Gao, X. Niu, and N. El-Sheimy, "Tight fusion of a monocular camera, mems-imu, and single-frequency multi-gnss rtk for precise navigation in gnss-challenged environments," *Remote Sensing*, vol. 11, no. 6, p. 610, 2019.
- [23] S. Cao, X. Lu, and S. Shen, "Gvins: Tightly coupled gnss-visual-inertial fusion for smooth and consistent state estimation," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2004–2021, 2022.
- [24] J. Liu, W. Gao, and Z. Hu, "Optimization-based visual-inertial slam tightly coupled with raw gnss measurements," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11612–11618, IEEE, 2021.
- [25] T. Li, L. Pei, Y. Xiang, W. Yu, and T.-K. Truong, "P³-vins: Tightly-coupled ppp/ins/visual slam based on optimization approach," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7021–7027, 2022.
- [26] W. Lee, P. Geneva, Y. Yang, and G. Huang, "Tightly-coupled gnss-aided visual-inertial localization," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 9484–9491, IEEE, 2022.
- [27] C. Liu, C. Jiang, and H. Wang, "Variable observability constrained visual-inertial-gnss ekf-based navigation," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6677–6684, 2022.
- [28] C. Liu, C. Jiang, and H. Wang, "Ingvio: A consistent invariant filter for fast and high-accuracy gnss-visual-inertial odometry," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1850–1857, 2023.
- [29] P. S. Maybeck, *Stochastic models, estimation, and control*. Academic press, 1982.
- [30] G. J. Bierman, *Factorization methods for discrete sequential estimation*. Courier Corporation, 2006.
- [31] K. Wu, A. M. Ahmed, G. A. Georgiou, and S. I. Roumeliotis, "A square root inverse filter for efficient vision-aided inertial navigation on mobile devices," in *Robotics: Science and Systems*, vol. 2, p. 2, Rome, Italy, 2015.
- [32] J. Civera, A. J. Davison, and J. M. Montiel, "Inverse depth parametrization for monocular slam," *IEEE transactions on robotics*, vol. 24, no. 5, pp. 932–945, 2008.
- [33] N. Trawny and S. I. Roumeliotis, "Indirect kalman filter for 3d attitude estimation," *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep*, vol. 2, p. 2005, 2005.
- [34] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [35] M. Li, *Visual-inertial odometry on resource-constrained systems*. University of California, Riverside, 2014.
- [36] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.
- [37] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ international conference on intelligent robots and systems*, pp. 573–580, IEEE, 2012.
- [38] J. Temiissen, "The least-squares ambiguity decorrelation adjustment: A method for fast gps integer ambiguity estimation," *Journal of Geodesy*, vol. 70, pp. 65–82, 1995.