2022 IEEE International Conference on Robotics and Automation (ICRA)
May 23-27, 2022. Philadelphia, PA, USA

2022 IEEE International Conference on Robotics and Automation (ICRA) | 978-1-7281-9681-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRA46639.2022.9811362

Tightly-coupled GNSS-aided Visual-Inertial Localization

Woosik Lee, Patrick Geneva, Yulin Yang, and Guoquan Huang

***Abstract*— A navigation system which can output drift-free global trajectory estimation with local consistency holds great potential for autonomous vehicles and mobile devices. We propose a tightly-coupled GNSS-aided visual-inertial navigation system (GAINS) which is able to leverage the complementary sensing modality from a visual-inertial sensing pair, which provides high-frequency local information, and a Global Navigation Satellite System (GNSS) receiver with low-frequency global observations. Specifically, the raw GNSS measurements (including pseudorange, carrier phase changes, and Doppler frequency shift) are carefully leveraged and tightly fused within a visual-inertial framework. The proposed GAINS can accurately model the raw measurement uncertainties by canceling the atmospheric effects (e.g., ionospheric and tropospheric delays) which requires no prior model information. A robust state initialization procedure is presented to facilitate the fusion of global GNSS information with local visual-inertial odometry, and the spatiotemporal calibration between IMU-GNSS are also optimized in the estimator. The proposed GAINS is evaluated on extensive Monte-Carlo simulations on a trajectory generated from a large-scale urban driving dataset with specific verification for each component (i.e., online calibration and system initialization). GAINS also demonstrates competitive performance against existing state-of-the-art methods on a publicly available dataset with ground truth.**

I. INTRODUCTION

Simultaneous localization and mapping (SLAM) have been a fundamental technology for autonomous robots, such as micro-aerial vehicles (MAVs), autonomous vehicles, and mobile devices [1]. SLAM tries to recover the platform's pose (orientation and position) while also reconstructing the surrounding environment to enable perception or for the further benefit of localization. Of the many different sensor combinations available, e.g., RGB-D [2]–[4], event cameras [5], LiDARs [6]–[8], wheel odometry [9], [10], the monocular camera and inertial sensors, which can be leveraged to build visual-inertial navigation systems (VINS), have become increasingly prominent due to their small size, complimentary sensing nature, weight, and low cost [11]. VINS-based methods are still limited by their ability to only provide the relative pose change to an arbitrary local frame and inability to recover the 4 degree-of-freedom (dof) global yaw and position [12], [13]. This has motivated the fusion of additional sensors, and in particular, the coupling of VINS with a Global Navigation Satellite System (GNSS) has seen attraction due to the drift-free nature of GNSS in a global scale and the high-accuracy trajectory of VINS in the local scale. This combination can enable accurate drift-free large-scale localization, consistent local trajectories, and robustness to low GNSS visibility scenarios.

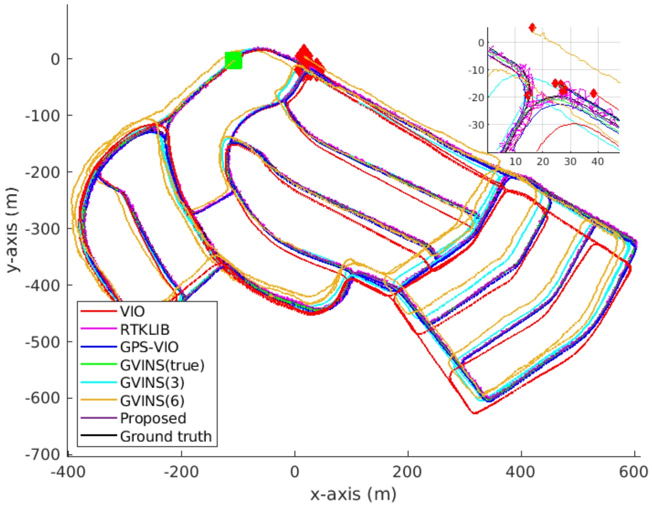

Fig. 1: Estimated trajectories of each algorithm with the 9.1 km simulation dataset. VIO [14] (red), RTKLIB [15] (magenta), GPS-VIO [16] (blue), GVINS [17] with true atmospheric model (green), GVINS with 3 m atmospheric model error (light blue), GVINS with 6m atmospheric model error (yellow), proposed GAINS (purple), and ground truth (black). The green and red squares correspond to the start and end of the trajectory, respectively.

There have been quite a few works which have investigated incorporating GNSS measurements in a "loosely-coupled" manor [18]–[22], along with many that have investigated a more "tightly-coupled" fusion of all sources of information into a single joint optimization problem [16], [23]–[26]. Of these, a few research efforts have looked to leverage the *raw* GNSS measurements (pseudorange, carrier phase changes, or Doppler frequency shift changes) with visual-inertial sensors [17], [26]–[32]. A major advantage of such effort is that with raw measurements the system can still gain information in scenarios with limited GNSS coverage (such as two or three satellites), which can be frequent in urban environments or with conservative elevation thresholding [27], [28], [31]. Soloviev and Venable [27] showed that by leveraging the raw measurements this limited GNSS coverage case could be handled. Won et al. [28] studied these "low GNSS visibility conditions" and additionally conducted an observability analysis to investigate the identifiability of the state with different numbers of GNSS satellites. Recently two preprints which have re-visited the tightly-coupled fusion of raw GNSS measurements [17], [32] have been released. Liu et al. [32] leveraged the raw GNSS measurements to enable au-

This work was partially supported by the University of Delaware (UD) College of Engineering, the NSF (IIS-1924897), the ARL (W911NF-19-2-0226, W911NF-20-2-0098), and Google ARCore. P. Geneva is also supported by the UD University Doctoral Fellowship.

The authors are with the Robot Perception and Navigation Group (RPNG), University of Delaware, Newark, DE 19716, USA. {woosik, pgeneva, yuyang, ghuang}@udel.edu

978-1-7281-9680-0/22/$31.00 ©2022 IEEE

tonomous driving in urban canyons where skyscrapers block a majority of the sky. They present an optimization-based approach which jointly fuses inertial, camera feature reprojection, GNSS pseudorange, and Doppler shift measurements in a joint sliding window. However, the atmospheric delays of the GNSS signal, such as ionospheric and tropospheric delays, were not considered which can yield very large errors in practice. Cao et al. [17] presented another optimization-based real-time estimator, termed GVINS, which accurately provides 6 dof estimates in complex indoor-outdoor environments by leveraging GNSS pseudorange and Doppler shift measurements. They modeled ionospheric delay using the ionospheric parameters that are included the signal, and used standard atmosphere model for troposphere delay [33]. As these models are approximating the possible delays, the accuracy of the models can vary from time to time.

As compared to these previous works we leverage a lightweight computationally efficient filter-based estimator. Raw measurements including GNSS pseudorange, carrier phase changes, and Doppler frequency are leveraged to constrain the visual-inertial state, without requirement of prior knowledge of the atmospheric delay model or external information. The spatiotemporal calibration between the receiver and IMU is also estimated and leverages efficient state interpolation to facilitate crucial sensor-to-sensor time-offset calibration. The system is able to leverage raw measurements from the constellation (e.g., GPS, Galileo, GLONASS, BeiDou) to improve measurement noise uncertainty modeling. Specifically, the main contributions are as follows:

- We propose *GAINS* which optimally and efficiently fuses raw GNSS pseudorange, carrier phase changes, and Doppler frequency measurements in a tightly coupled manner with visual and inertial information.
- A differential measurement model is leveraged to remove atmospheric effects (e.g., ionospheric and tropospheric delays) and enables the accurate measurement modeling without approximating the delays.
- The proposed initialization procedure enables robust recovery of all GNSS-related parameters required to fuse global GNSS information with the local VIO system. Additionally, all spatiotemporal calibrations between sensors are performed including the crucial temporal offset between the receiver and IMU.
- The system is compared against an existing state-of-the-art method and different baselines in both large-scale urban driving simulations and real-world experiments. The proposed GAINS is able to achieve high levels of accuracy and robustness to high GNSS noise during initialization.

II. VISUAL-INERTIAL LOCALIZATION

In this section, we review the standard multi-state constraint Kalman filter (MSCKF [34]) framework which is later expanded to handle raw GNSS measurements. Specifically, at time t_k, the state vector $\mathbf{x}_k$ consists of the current inertial state $\mathbf{x}_{I_k}$ and n historical IMU pose clones $\mathbf{x}_{C_k}$ represented in the local VIO world frame $\{W\}$:

$$\mathbf{x}_k = \begin{bmatrix}\mathbf{x}_{I_k}^\top & \mathbf{x}_{C_k}^\top\end{bmatrix}^\top \tag{1}$$

$$\mathbf{x}_{I_k} = \begin{bmatrix}{}^{I_k}_{W}\bar{q}^\top & {}^{W}\mathbf{p}_{I_k}^\top & {}^{W}\mathbf{v}_{I_k}^\top & \mathbf{b}_g^\top & \mathbf{b}_a^\top\end{bmatrix}^\top \tag{2}$$

$$\mathbf{x}_{C_k} = \begin{bmatrix}{}^{I_{k\text{-}1}}_{W}\bar{q}^\top & {}^{W}\mathbf{p}_{I_{k\text{-}1}}^\top & \cdots & {}^{I_{k\text{-}n}}_{W}\bar{q}^\top & {}^{W}\mathbf{p}_{I_{k\text{-}n}}^\top\end{bmatrix}^\top \tag{3}$$

where ${}^{I_k}_{W}\bar{q}$ is the JPL unit quaternion [35] corresponding to the rotation matrix ${}^{I_k}_{W}\mathbf{R}$ rotating from $\{W\}$ to IMU frame $\{I\}$; ${}^{W}\mathbf{p}_{I_k}$ and ${}^{W}\mathbf{v}_{I_k}$ are the position and velocity of IMU in $\{W\}$; $\mathbf{b}_g$ and $\mathbf{b}_a$ are the biases of the gyroscope and accelerometer.[1]

A. IMU Kinematic Model

The state is propagated forward in time using the IMU's linear acceleration $\mathbf{a}_m$ and angular velocity $\boldsymbol{\omega}_m$ measurements:

$$\mathbf{a}_m = \mathbf{a} + {}^{I}_{W}\mathbf{R}\mathbf{g} + \mathbf{b}_a + \mathbf{n}_a, \quad \boldsymbol{\omega}_m = \boldsymbol{\omega} + \mathbf{b}_g + \mathbf{n}_g \tag{4}$$

where $\mathbf{g} \approx [0\ 0\ 9.81]^\top$ is the gravity in $\{W\}$, and $\mathbf{n}_a$ and $\mathbf{n}_g$ are zero mean Gaussian noises. The state and covariance is propagated from time t_k to t_{k+1} using the inertial kinematics $\mathbf{f}(\cdot)$ [35]:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{a}_m, \boldsymbol{\omega}_m) \tag{5}$$

$$\mathbf{P}_{k+1|k} = \boldsymbol{\Phi}(t_{k+1}, t_k)\mathbf{P}_{k|k}\boldsymbol{\Phi}(t_{k+1}, t_k)^\top + \mathbf{Q}_k \tag{6}$$

where $\boldsymbol{\Phi}$ is the error state transition matrix and $\mathbf{Q}$ is the discrete noise covariance [34].

B. Camera Measurement Model

Sparse corner features are detected and tracked over a series of historical states $\mathbf{x}_{C_k}$. A bearing measurement, $\mathbf{z}_k$, collected at time t_k is:

$$\mathbf{z}_k = \boldsymbol{\Pi}({}^{C_k}\mathbf{p}_f) + \mathbf{n}_k \tag{7}$$

$${}^{C_k}\mathbf{p}_f = {}^{C}_{I}\mathbf{R}\,{}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_f - {}^{G}\mathbf{p}_{I_k}) + {}^{C}\mathbf{p}_I \tag{8}$$

where $\boldsymbol{\Pi}([x\ y\ z]^\top) = [x/z\ y/z]^\top$ is the perspective projection function; ${}^{G}\mathbf{p}_f$ is the 3D point feature; and $\{{}^{C}_{I}\mathbf{R}, {}^{C}\mathbf{p}_I\}$ are the camera-IMU extrinsics. We then stack all measurements and nullspace project to create a featureless residual (i.e., $\mathbf{N}^\top\mathbf{H}_f = \mathbf{0}$) [34], [37]:

$$\mathbf{N}^\top\tilde{\mathbf{z}} = \mathbf{N}^\top\mathbf{H}_{x_{C_k}}\tilde{\mathbf{x}}_{C_k} + \mathbf{N}^\top\mathbf{H}_f{}^{G}\tilde{\mathbf{p}}_f + \mathbf{N}^\top\mathbf{n}_f \tag{9}$$

$$\Rightarrow\ \tilde{\mathbf{z}}' = \mathbf{H}'_{x_{C_k}}\tilde{\mathbf{x}}_{C_k} + \mathbf{n}'_f \tag{10}$$

This then can be directly used in the EKF update.

III. FUNDAMENTALS OF GNSS

GNSS uses satellites to provide geospatial positioning and is composed of three distinct segments: space, control, and user [38]. The space segment consists of GNSS satellites, orbiting about 20,000km above the earth and broadcasting a signal that identifies it and provides its time, orbit, and status. The control segment is composed of ground stations that adjust the satellites' orbit parameters and clocks to maintain

[1] We define $\mathbf{x} = \hat{\mathbf{x}} \boxplus \tilde{\mathbf{x}}$, where $\mathbf{x}$ is the true state, $\hat{\mathbf{x}}$ is its estimate, $\tilde{\mathbf{x}}$ is the error state, and the operation $\boxplus$ which maps the error state vector to its corresponding manifold [36]. The state $\hat{\mathbf{x}}_{a|b}$ denotes the estimate at time t_a formed by processing the measurements up to time t_b. Camera and IMU spatiotemporal calibration is not included for clarity.

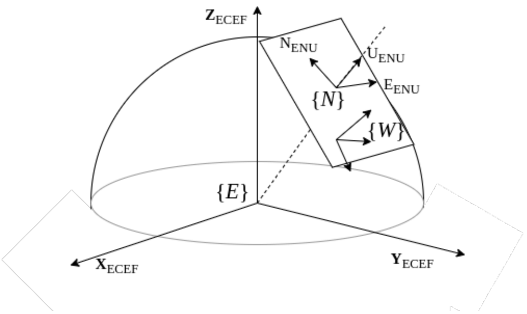

Fig. 2: The frame of the references: ECEF frame $\{E\}$, ENU frame $\{N\}$, and World frame $\{W\}$. $\{W\}$ is the local frame set up by VIO whose orientation is aligned with gravity along with the norm direction of $\{N\}$.

accuracy. The user segment consists of the signal receiver that performs signal processing. In general, three types of measurements can be extracted from the signal to estimate the receiver state: pseudorange, carrier phase, and Doppler shift. These signals are considered to be the *"raw"* measurements of GNSS and carry essential information which enables accurate global localization.

A. Pseudorange

The most basic GNSS measurement is the difference between the signal emission time from the satellite and the receiver reception time. After multiplying by the speed of light, this is now a *pseudorange* between the two systems. Consider that the receiver got the signal at ${}^{E}t_{r,m}$ and the decoded departure time from satellite is ${}^{E}t_{s,m}$ we have:

$$ {}^{E}t_{r,m} = {}^{E}t_r + {}^{E}b_r, \quad {}^{E}t_{s,m} = {}^{E}t_s + {}^{E}b_s \tag{11} $$

where the superscript E stands for the Earth-Centered, Earth-Fixed (ECEF) coordinate $\{E\}$ of the satellite system (see Fig. 2); ${}^{E}t_r$ and ${}^{E}t_s$ are the true signal reception and departure time, and ${}^{E}b_r$ and ${}^{E}b_s$ are the time bias of the receiver and the satellite, respectively. Hence, the pseudorange measurement can be computed by time differences:

$$ z_p = c({}^{E}t_{r,m} - {}^{E}t_{s,m}) \tag{12} $$

where $c \approx 3.0 \times 10^9 m/s$ is the speed of light. Considering the clock biases, the pseudorange measurement is modeled as:

$$ z_p = d_{r,s} + c({}^{E}b_r - {}^{E}b_s) + I + T + M + n_p \tag{13} $$

$$ d_{r,s} = ||{}^{E}\mathbf{p}_r({}^{E}t_r) - {}^{E}\mathbf{p}_s({}^{E}t_s)||_2 \tag{14} $$

where ${}^{E}\mathbf{p}_a(t)$ is the position of a in $\{E\}$ at time t_a with a denotes r or s, ionospheric delay I, tropospheric delay T, multi-path M, and n_p is a white Gaussian noise.

B. Carrier Phase

The carrier phase of the signal can also be used to obtain a range between the satellite and receiver. The carrier phase measurements are often more precise than pseudorange by typically two orders of magnitude. They are modeled as:

$$ z_c = d_{r,s} + c({}^{E}b_r - {}^{E}b_s) - I + T + M + \lambda N + n_c \tag{15} $$

$$ d_{r,s} = ||{}^{E}\mathbf{p}_r({}^{E}t_r) - {}^{E}\mathbf{p}_s({}^{E}t_s)||_2 \tag{16} $$

where λ is the GNSS carrier wavelength, N is the phase count error, and n_c is a white Gaussian noise. Note that the I has a negative sign due to the ionosphere producing an advance of the carrier phase measurement equal to the delay on the pseudorange measurements, and that N may change arbitrarily every time the receiver loses lock, producing measurement discontinuities.

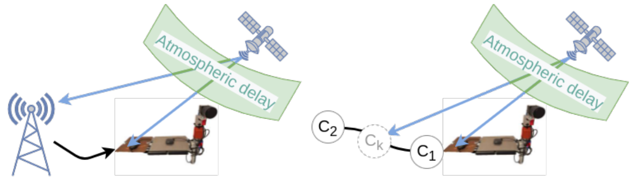

Fig. 3: In classical differential GNSS (left), the base interprets the GNSS signal (blue arrow) and provides correction information (black arrow) to the rover. The proposed system (right) stacks GNSS measurements and formulates a differential measurement model to perform update directly. To handle the asynchronicity, we interpolate two bounding IMU poses

C. Doppler Shift

The receiver and satellite range change as both move through space. This range change is reflected in the phase of the signal and is called the *Doppler shift*. The Doppler shift allows for recovery of the relative velocity between the receiver satellite. The Doppler shift is modeled as:

$$ z_d = -\frac{1}{\lambda}\left((\mathbf{k}^\top({}^{E}\mathbf{v}_s - {}^{E}\mathbf{v}_r) + c({}^{E}\dot{b}_r - {}^{E}\dot{b}_s)\right) + n_d \tag{17} $$

where $\mathbf{k}$ is a unit vector of the receiver to the satellite position, ${}^{E}\dot{b}_r$ and ${}^{E}\dot{b}_s$ are the receiver and the satellite time bias drift rate, and n_d is a white Gaussian noise.

D. Differential GNSS

Differential GNSS techniques are well-known enhancements that leverage a known base station and that certain parameters vary slowly with time (e.g., ephemeris prediction, ionospheric and tropospheric delays). The Differential GNSS algorithm is based on differences of pseudorange measurements [see Eq. (13) and Fig. 3] :

$$ z_D = z_{p,b} - z_{p,r} \tag{18} $$

Assuming no multi-path errors, we can model this measurement as (we drop the times for clarity):

$$ \begin{aligned} z_D = \Delta d_{r,s,b} &+ c({}^{E}b_b - {}^{E}b_s) - c({}^{E}b_r - {}^{E}b_s) \\ &+ I_b - I_r + T_b - T_r + n_{p,b} - n_{p,r} \end{aligned} \tag{19} $$

$$ \Delta d_{r,s,b} = ||{}^{E}\mathbf{p}_b - {}^{E}\mathbf{p}_s||_2^2 - ||{}^{E}\mathbf{p}_r - {}^{E}\mathbf{p}_s||_2 \tag{20} $$

where the subscript b stands for the base station. When the base station and receiver are close to each other, both signals travel through almost the same path and thus the slow varying ionospheric and tropospheric delays should approximately cancel out:

$$ z_D = \Delta d_{r,s,b} + c({}^{E}b_b - {}^{E}b_r) + n_{p,b} - n_{p,r} \tag{21} $$

This enables the use of GNSS measurements without modeling ionospheric and tropospheric delays. The resulting solutions can reach centimeter-level accuracy with a base station within 10km.

IV. SEQUENTIAL-DIFFERENTIAL GNSS

To minimize sources of noise and errors, we propose using differential measurements between sequentially received measurements, see Fig. 3. This differential measurement model is used to remove the affects of I and T [see Eq. (13)-(17)] which can have model errors on the level of 7m and 1m, respectively [39].

A. Differential Pseudorange

Consider two pseudorange measurements from a satellite at time t_k and t_{k+1}:

$$z_{Dp} = z_{p,k+1} - z_{p,k} \tag{22}$$

$$\begin{aligned} z_{Dp} = \Delta d_{Dr,s} + c({}^{E}b_{r,k+1} - {}^{E}b_{s,k+1}) - c({}^{E}b_{r,k} - {}^{E}b_{s,k}) \\ + I_{k+1} - I_k + T_{k+1} - T_k + M_{k+1} - M_k \\ + n_{p,k+1} - n_{p,k} \end{aligned}$$

$$\begin{aligned} \Delta d_{Dr,s} = ||{}^{E}\mathbf{p}_r({}^{E}t_{r,k+1}) - {}^{E}\mathbf{p}_s({}^{E}t_{s,k+1})||_2 \\ - ||{}^{E}\mathbf{p}_r({}^{E}t_{r,k}) - {}^{E}\mathbf{p}_s({}^{E}t_{s,k})||_2 \end{aligned} \tag{23}$$

The signal path and time remain nearly constant between consecutive timestamps. Hence, the atmospheric delays I and T remain constant, which indicates that $I_{k+1} - I_k \simeq 0$ and $T_{k+1} - T_k \simeq 0$. We also assume there is no multi-path delay M or that it is constant under the same logic. Thus, the following differential pseudorange measurement model is defined:

$$\begin{aligned} z_{Dp} = \Delta d_{Dr,s} + c({}^{E}b_{r,k+1} - {}^{E}b_{s,k+1}) \\ - c({}^{E}b_{r,k} - {}^{E}b_{s,k}) + n_{p,k+1} - n_{p,k} \end{aligned} \tag{24}$$

Note that satellite parameters ${}^{E}\mathbf{p}_s$, ${}^{E}t_s$ and ${}^{E}b_s$ can be predicted very accurately as the information is managed by the ground stations.

B. Differential Carrier Phase

The carrier phase measurement, Eq. (15), is similar to Eq. (13). Analogously, we define the differential carrier phase measurement model as:

$$z_{Dc} = \Delta d_{Dr,s} + c({}^{E}b_{r,k+1} - {}^{E}b_{s,k+1}) \tag{25}$$

$$- c({}^{E}b_{r,k} - {}^{E}b_{s,k}) + n_{c,k+1} - n_{c,k} \tag{26}$$

$$\begin{aligned} \Delta d_{Dr,s} = ||{}^{E}\mathbf{p}_r({}^{E}t_{r,k+1}) - {}^{E}\mathbf{p}_s({}^{E}t_{s,k+1})||_2 \\ - ||{}^{E}\mathbf{p}_r({}^{E}t_{r,k}) - {}^{E}\mathbf{p}_s({}^{E}t_{s,k})||_2 \end{aligned} \tag{27}$$

Note that we also cancel out the $\lambda N_{k+1} - \lambda N_k$ terms. In practice, we only choose to compute this model when the satellite lock is stable to ensure this is true.

C. Doppler Shift

Unlike the other measurement models, Eq. (17) can be used to directly update our state. We can define the receiver satellite bearing and relative velocity as:

$$\mathbf{k} = \frac{{}^{E}\mathbf{p}_s - {}^{E}\mathbf{p}_r}{\sqrt{{}^{E}\mathbf{p}_s - {}^{E}\mathbf{p}_r}} \tag{28}$$

$${}^{E}\mathbf{v}_r = {}^{E}_{W}\mathbf{R}({}^{W}\mathbf{v}_I + {}^{I}_{W}\mathbf{R}^{\top}\lfloor {}^{I}\boldsymbol{\omega}_I \rfloor {}^{I}\mathbf{p}_r) \tag{29}$$

where $\lfloor \cdot \rfloor$ is the skew-symmetric matrix.

D. Complete GNSS Measurement Model

Due to the delayed asynchronous nature of the GNSS receiver, the inertial state has likely advanced beyond the collection time and thus we need to be able to recover the pose at an arbitrary time. We leverage linear interpolation of the IMU pose (Fig. 3 right) to recover the receiver position in the world $\{W\}$ frame:

$${}^{E}\mathbf{p}_r(t_k) = {}^{E}\mathbf{p}_W + {}^{W}_{E}\mathbf{R}^{\top}\left({}^{W}\mathbf{p}_{I_k} + {}^{I_k}_{W}\mathbf{R}^{\top}{}^{I}\mathbf{p}_r\right) \tag{30}$$

$${}^{I_k}_{W}\mathbf{R} = \mathrm{Exp}\left(\lambda\ \mathrm{Log}\left({}^{I_b}_{W}\mathbf{R}{}^{I_a}_{W}\mathbf{R}^{\top}\right)\right){}^{I_a}_{W}\mathbf{R} \tag{31}$$

$${}^{W}\mathbf{p}_{I_k} = (1 - \lambda){}^{W}\mathbf{p}_{I_a} + \lambda{}^{W}\mathbf{p}_{I_b} \tag{32}$$

$$\lambda = (t_k + {}^{I}t_r - t_a)/(t_b - t_a) \tag{33}$$

where ${}^{I}t_r$ is the time offset between the GNSS and IMU clocks, the bounding poses have timestamps $t_a \leq (t_k + {}^{I}t_r) \leq t_b$, and $\mathrm{Exp}(\cdot)$, $\mathrm{Log}(\cdot)$ are the SO(3) matrix exponential and logarithmic functions [40]. We can see that the position of the receiver at time t_k is a function of, and thus all GNSS measurements, the IMU states, GNSS-IMU extrinsic, and time offset. All these can be directly estimated by finding their derivatives in the above functions.

Having now defining the complete measurement model in Eq. (24), (26), (17), and (30), we can now see that an initial estimate of the ECEF to World transform $\{{}^{E}_{W}\mathbf{R}, {}^{E}\mathbf{p}_W\}$ (see Fig. 2), receiver time bias ${}^{E}b_r$, and receiver time bias drift rate ${}^{E}\dot{b}_r$ is required.

V. ECEF-TO-WORLD GNSS INITIALIZATION

A. Step 1: Reference Frame Transform Initialization

To initialize the transform $\{{}^{E}_{W}\mathbf{R}, {}^{E}\mathbf{p}_W\}$, we follow Lee's method [16] which provides robust initialization given enough measurements. We collect Single Point Positioning (SPP) output of the GNSS receiver, which is the position in East-North-Up frame (ENU, $\{N\}$) when using the first measurement as the datum (see Fig. 2). Given a set of GNSS positions in the ENU frame $\{{}^{N}\mathbf{p}_{r_1}, \cdots, {}^{N}\mathbf{p}_{r_n}\}$ and the corresponding interpolated positions in the World frame $\{{}^{W}\mathbf{p}_{r_1}, \cdots {}^{W}\mathbf{p}_{r_n}\}$, we use the following geometric constraints to derive the frame initialization between $\{N\}$ and $\{W\}$:

$${}^{N}\mathbf{p}_{r_i} = {}^{N}\mathbf{p}_W + {}^{N}_{W}\mathbf{R}{}^{W}\mathbf{p}_{r_i},\ \forall i = 1 \cdots n \Rightarrow \tag{34}$$

$${}^{N}\mathbf{p}_{r_j} - {}^{N}\mathbf{p}_{r_1} = {}^{N}_{W}\mathbf{R}({}^{W}\mathbf{p}_{r_j} - {}^{W}\mathbf{p}_{r_1}),\ \forall j = 2 \cdots n \tag{35}$$

Note that, this is a 4 d.o.f (instead of 6 d.o.f) transformation including 3 d.o.f translation and 1 d.o.f for yaw between the $\{N\}$ and $\{W\}$ since both frames are gravity aligned. We solve Eq. (35) as the linear least-squares with quadratic constraint problem to get $\{{}^{N}_{W}\mathbf{R}, {}^{N}\mathbf{p}_W\}$. We then compute the following to find reference frame transform:

$${}^{E}_{W}\mathbf{R} = {}^{E}_{N}\mathbf{R}{}^{N}_{W}\mathbf{R} \tag{36}$$

$${}^{E}\mathbf{p}_W = {}^{E}\mathbf{p}_N + {}^{E}_{N}\mathbf{R}{}^{N}\mathbf{p}_W \tag{37}$$

where ${}^{E}_{N}\mathbf{R}$ and ${}^{E}\mathbf{p}_N$ can are computed from the datum [41]. The $\{{}^{E}_{W}\mathbf{R}, {}^{E}\mathbf{p}_W\}$ is further corrected during the delayed initialization into the state [42].

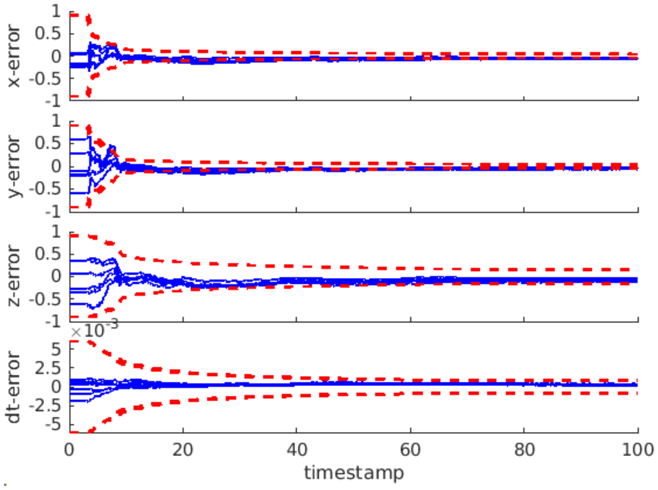

Fig. 4: Calibration errors of GNSS spatiotemporal calibration (solid) and 3σ bound (dotted) for five different runs. Each runs has a different realization of the measurement noise and perturbations.

TABLE I: Monte-Carlo simulation parameters

Parameter	Value	Parameter	Value
IMU Freq. (hz)	400	Max Cam Pts/Frame	100
Cam Freq. (hz)	10	Max SLAM Feats	50
GNSS Freq. (hz)	1	Feature Rep.	GLOBAL
Max. Clone Size	11	Pixel Noise (px)	1
Gyro White Noise	1.7e-4	Gyro Random Walk	1.9e-4
Accel. White Noise	2.0e-3	Accel. Random Walk	3.0-3
Ion. Random Walk	1.0e-2	Tro. Random Walk	1.0e-2
Satellite Time Drift	1.0e-6	Receiver Time Drift	1.0e-06
Pse. Meas. Noise	1.0e-0	Number of Satellites	30

B. Step 2: GNSS Receiver Bias Initialization

Once the reference frame transformation $\{{}^{E}_{W}\mathbf{R}, {}^{E}\mathbf{p}_{W}\}$ is initialized, we collect GNSS measurements and compute the initial guesses of ${}^{E}b_r$ and ${}^{E}\dot{b}_r$ as follows:

$$\mathbf{z}_R = \mathbf{t} + \mathbf{B}\mathbf{b} \tag{38}$$

where $\mathbf{z}_R$ is the stack of all the measurements [see z_{Dp} (24), z_{Dc} (26), and λz_d (17)], $\mathbf{t}$ is the stacked right side terms, $\mathbf{B}$ is a incident matrix [43] multiplied by the speed of light c, and $\mathbf{b}$ is a stack of GNSS receiver time biases ${}^{E}b_{r,k}$ $\forall k = 1 \cdots n$ and the bias drift rate ${}^{E}\dot{b}_r$. By solving the above least square problem, we can get the initial guesses of the time biases and the drift rate. After gaining the initial guess, we further perform delayed initialization for higher accuracy.

VI. Simulation

We verify our proposed system in a Monte-Carlo simulation on a large-scale trajectory, see Fig. 1. We simulate IMU readings, visual bearing tracks, and raw GNSS readings from 30 satellites based on the parameters in Table I [14], [44].

A. GNSS Spatiotemporal Calibration

We first verify the online spatiotemporal GNSS-IMU calibration (include the 3D translation and time offset) through 5 Monte-Carlo runs, with the different realization of measurement noises and initial calibration values. Fig. 4 shows the estimation errors and 3σ bounds. The estimation errors are encapsulated within the envelope of 3σ bounds and converge to near zero, thus showing the consistency and accuracy of the online calibration along with their ability to quickly converge in less than 20 seconds.

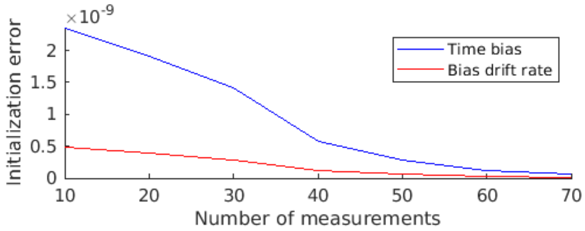

Fig. 5: The effects of time bias and bias drift rates on initialization errors.

TABLE II: Average yaw and position errors over 50 runs for different initialization distances and GNSS SPP noise values in units of degree/meter.

dist\\σ	0.1m	0.5m	1m	2m	5m
5m	1.57 / 0.58	6.25 / 2.91	14.52 / 6.79	30.66 / 71.75	69.26 / 88.42
10m	1.31 / 0.52	5.54 / 2.19	9.45 / 4.17	20.41 / 44.94	47.45 / 94.93
20m	0.79 / 0.27	2.47 / 0.99	4.84 / 2.01	10.24 / 4.10	26.96 / 51.54
50m	0.53 / 0.07	0.80 / 0.16	0.97 / 0.27	1.79 / 0.62	4.86 / 1.48
100m	0.45 / 0.09	0.49 / 0.06	0.50 / 0.12	0.78 / 0.24	2.11 / 0.65

B. Hyper-parameter Sensitivity of Initialization

To gain insight into how the initialization procedure is affected by GNSS measurement noise and trajectory length, we simulated 0.1-5m GNSS SPP noise and 5-100m initialization distance thresholds. To prevent biasing these results to the initial section of this particular trajectory, it is split into non-overlapping segments that the initalization procedure was performed on. The resulting statistics on the initialization accuracy are shown in Table II. In general, the initialization errors are smaller for larger distances and with smaller noise. The results indicate that reasonable accuracy for this transformation can be achieved after 50m for most realistic levels of GNSS SPP noise. In practice, these results can be used to determine the necessary distance threshold for system initialization with different sensors.

We also verify the proposed GNSS time bias and time bias drift rate initialization algorithm and its performances in Fig. 5. The time bias and time bias drift rate initialization errors highly relate to the number of available measurements used during the initialization. The proposed algorithm can achieve reasonable initialization performance using ≥ 40 measurements with corresponding range accuracy of 1.5m for time bias while 0.3m for time bias drift rates (error × speed of light).

Remarks: Intuitively, the shorter distance the robot travels, the less information about the transformation can be gained because the "true" trajectory is buried in the GNSS measurement noise. However, some systems may be required to be initialized before collecting the proper length of the trajectory. In this case, the transformation error can be taken into account by inflating the state covariance after the initialization step.

C. Navigation Performance Evaluation

We further evaluate the localization accuracy of the proposed GAINS (proposed) with regrading to GPS-VIO [16],

TABLE III: Relative pose error (RPE) plots for each estimator in simulation. Units are in degrees/meters.

Algorithms	8m	16m	24m	40m
VIO	0.042 / 0.112	0.058 / 0.163	0.071 / 0.199	0.082 / 0.265
GPS-VIO	0.039 / 0.092	0.053 / 0.131	0.065 / 0.161	0.075 / 0.211
RTKLIB	0.447 / 2.553	0.510 / 2.569	0.634 / 2.555	0.531 / 2.595
GVINS(true)	0.038 / 0.084	0.051 / 0.119	0.062 / 0.144	0.072 / 0.188
GVINS(3)	0.236 / 0.521	0.343 / 0.749	0.749 / 0.904	0.466 / 1.215
GVINS(6)	0.636 / 1.464	0.867 / 2.041	1.028 / 2.356	1.184 / 3.096
Proposed	**0.036 / 0.080**	**0.049 / 0.115**	**0.059 / 0.141**	**0.076 / 0.185**

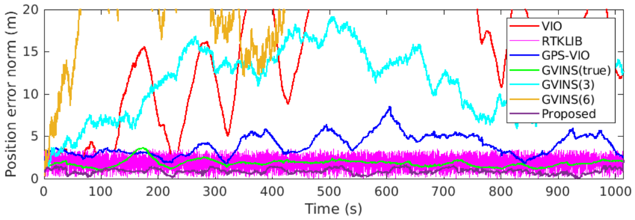

Fig. 6: The position error evaluation for different algorithms with simulated dataset. GAINS (proposed) achieves the best accuracy compared to GPS-VIO, VIO, and GVINS with true & perturbed atmospheric models.

GVINS [17], RTKLIB [15], and VIO [14] using the simulation data. Note that we tested GVINS providing perfect atmospheric model (GVINS(true)) and wrong atmospheric model equivalent to 3m & 6m pseudorange error (GVINS(3) & GVINS(6)). These errors are realistic as the tropospheric delay varies from 1-3m [45] and the ionospheric delay can be even larger. All methods leverage the same visual front-end to provide a fair comparison. The simulated trajectory is designed based a realistic dataset collected in drive through a neighbourhood with a total length of 9.1km (see Fig. 1). The position errors presented in Fig. 6 show that the proposed GAINS achieves the smallest position estimation errors compared to the other algorithms. Also, the estimation accuracy of GVINS(3) and GVINS(6) is sometimes worse than the baseline VIO system due to the GNSS atmospheric model error. The relative pose error (RPE) [46] for orientation and position are shown in Table III. The proposed GAINS outperforms the majority of other GNSS-aided algorithms in estimation accuracy, with slight performance gains when compared to GVINS (true) which requires a perfect GNSS model for the atmospheric affects.

VII. REAL-WORLD EXPERIMENT

We further evaluate the proposed system on the GVINS-Dataset `sports_field` [2] real-word dataset [17]. This dataset contains a monocular Aptina MT9V034 camera, ADIS16448 IMU, and a u-blox ZED-F9P GNSS receiver. The `sports_field` sequence from this dataset is collected in a typical outdoor opened area and buildings environment. To get reliable groundtruth, during the experiment, we ensured that the RTK GNSS locked onto a few satellites and outputted stable position estimates for the sensor. We ran the proposed GAINS on the dataset along with GVINS [17], and GPS-VIO [16]. OpenVINS [14] and VINS-Mono [47] were also run to provide a baseline and show the accuracy improvement of the proposed GNSS-aided system compared to visual-inertial only methods. The resulting trajectories from are shown in Fig 7 alongside the reference RTK trajectory. It can be seen that all GNSS-aided algorithms are close to the RTK while VIO systems diverge over time as they are the local estimators. Table IV further shows the Root Mean Squared Error (RMSE) results of each algorithm showing GAINS similar accuracy to GVINS even though GAINS do not explicitly model the atmospheric delays.

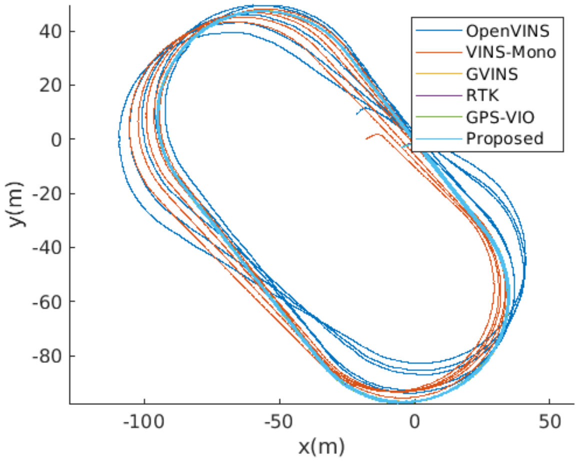

Fig. 7: Estimated trajectories of different VIO and GPS/GNSS aided algorithms with `sports_field` dataset.

TABLE IV: RMSE on the `sports_field` dataset (meters)

Proposed	GVINS [17]	GPS-VIO [16]	OpenVINS [14]	VINS-Mono [47]
0.319	0.327	0.374	11.265	9.189

VIII. Conclusion and Future Work

In this paper, we propose GAINS, a tightly coupled GNSS aided visual-inertial localization system, which can optimally and efficiently fuse raw GNSS pseudorange, Doppler frequency shift, and carrier phase measurements within a visual-inertial estimator framework. A differential measurement model is leveraged to accurately model the raw measurement uncertainty of GNSS measurements by removing the atmospheric effects (e.g., ionospheric and tropospheric delays). We also proposed a 2-step initialization algorithm which robustly recoveries all GNSS-related parameters needed to fuse global GNSS information with local visual-inertial odometry. In addition, all spatiotemporal calibration parameters between GNSS receiver and IMU are incorporated in state estimation, allowing for flexible sensor integration for GAINS. The system is compared against an existing state-of-the-art method and different baselines in both large-scale urban driving simulations and real-world experiments and is able to achieve high levels of accuracy and robustness to high GNSS noise during initialization. In the future, we will perform the observability analysis for GAINS to identify possible existing degeneration motions and explore integrating more sensors (multi-cameras, LiDAR, and wheel encoder) for improved localization accuracy.

[2] https://github.com/HKUST-Aerial-Robotics/GVINS-Dataset

REFERENCES

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. D. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2] A. Bachrach, S. Prentice, R. He, P. Henry, A. S. Huang, M. Krainin, D. Maturana, D. Fox, and N. Roy, “Estimation, planning, and mapping for autonomous flight using an RGB-D camera in GPS-denied environments,” *International Journal of Robotics Research*, vol. 31, no. 11, pp. 1320–1343, 2012.

[3] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, “3-d mapping with an rgb-d camera,” *IEEE Transactions on Robotics*, vol. PP, no. 99, pp. 1–11, 2013.

[4] T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. J. Leonard, and J. McDonald, “Real-time large-scale dense RGB-D SLAM with volumetric fusion,” *International Journal of Robotics Research*, Dec. 2014.

[5] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, “Continuous-time visual-inertial odometry for event cameras,” *IEEE Transactions on Robotics*, pp. 1–16, 2018.

[6] P. Geneva, K. Eckenhoff, Y. Yang, and G. Huang, “LIPS: Lidar-inertial 3d plane slam,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, Oct. 2018.

[7] X. Zuo, P. Geneva, W. Lee, Y. Liu, and G. Huang, “LIC-Fusion: Lidar-inertial-camera odometry,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Macau, China, Nov. 2019.

[8] X. Zuo, P. Geneva, Y. Yang, W. Ye, Y. Liu, and G. Huang, “Visual-inertial localization with prior lidar map constraints,” *IEEE Robotics and Automation Letters (RA-L)*, 2019.

[9] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, “VINS on wheels,” in *Proc. of the IEEE International Conference on Robotics and Automation*, May 2017, pp. 5155–5162.

[10] W. Lee, K. Eckenhoff, Y. Yang, P. Geneva, and G. Huang, “Visual-inertial-wheel odometry with online calibration,” in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV, 2020.

[11] G. Huang, “Visual-inertial navigation: A concise review,” in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[12] J. Kelly and G. S. Sukhatme, “Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration,” *International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, Jan. 2011.

[13] 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.

[14] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, “OpenVINS: A research platform for visual-inertial estimation,” in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020. [Online]. Available: https://github.com/rpng/open_vins

[15] T. Takasu and A. Yasuda, “Development of the low-cost rtk-gps receiver with an open source program package rtklib,” in *International symposium on GPS/GNSS*, vol. 1. International Convention Center Jeju Korea, 2009.

[16] W. Lee, K. Eckenhoff, P. Geneva, and G. Huang, “Intermittent gps-aided vio: Online initialization and calibration,” in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020.

[17] S. Cao, X. Lu, and S. Shen, “GVINS: Tightly coupled gnss-visual-inertial fusion for smooth and consistent state estimation,” *arXiv preprint arXiv:2103.07899*, 2021.

[18] 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*. IEEE, 2012, pp. 735–742.

[19] 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 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 3923–3929.

[20] R. Mascaro, L. Teixeira, T. Hinzmann, 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)*. IEEE, 2018, pp. 1421–1428.

[21] 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*, 2021.

[22] 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.

[23] D. P. Shepard and T. E. Humphreys, “High-precision globally-referenced position and attitude via a fusion of visual slam, carrier-phase-based gps, and inertial measurements,” in *2014 IEEE/ION Position, Location and Navigation Symposium-PLANS 2014*. IEEE, 2014, pp. 1309–1328.

[24] J. E. Yoder, P. A. Iannucci, L. Narula, and T. E. Humphreys, “Multi-antenna vision-and-inertial-aided cdgnss for micro aerial vehicle pose estimation,” in *Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020)*, 2020, pp. 2281–2298.

[25] G. Cioffi and D. Scaramuzza, “Tightly-coupled fusion of global positional measurements in optimization-based visual-inertial odometry,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5089–5095.

[26] X. Li, X. Wang, J. Liao, X. Li, S. Li, and H. Lyu, “Semi-tightly coupled integration of multi-gnss ppp and s-vins for precise positioning in gnss-challenged environments,” *Satellite Navigation*, vol. 2, no. 1, pp. 1–14, 2021.

[27] A. Soloviev and D. Venable, “Integration of gps and vision measurements for navigation in gps challenged environments,” in *IEEE/ION Position, Location and Navigation Symposium*, 2010, pp. 826–833.

[28] 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, no. 2, pp. 177–187, 2014.

[29] 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)*. IEEE, 2016, pp. 858–863.

[30] 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.

[31] P. V. Gakne and K. O’Keefe, “Tightly-coupled gnss/vision using a sky-pointing camera for vehicle navigation in urban areas,” *Sensors*, vol. 18, no. 4, p. 1244, 2018.

[32] J. Liu, W. Gao, and Z. Hu, “Optimization-based visual-inertial slam tightly coupled with raw gnss measurements,” *arXiv preprint arXiv:2010.11675*, 2020.

[33] J. Saastamoinen, “Contributions to the theory of atmospheric refraction,” *Bulletin Géodésique (1946-1975)*, vol. 105, no. 1, pp. 279–298, 1972.

[34] A. I. Mourikis and S. I. Roumeliotis, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” in *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy, Apr. 10–14, 2007, pp. 3565–3572.

[35] N. Trawny and S. I. Roumeliotis, “Indirect Kalman filter for 3D attitude estimation,” University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., Mar. 2005.

[36] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder, “Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds,” *Information Fusion*, vol. 14, no. 1, pp. 57–77, 2013.

[37] Y. Yang, J. Maley, and G. Huang, “Null-space-based marginalization: Analysis and algorithm,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vancouver, Canada, Sept. 2017, pp. 6749–6755.

[38] C. Jeffrey, “An introduction to gnss gps, glonass, galileo and other global navigation satellite systems.” NovAtel Inc, 2010.

[39] J. Farrell, *Aided navigation: GPS with high rate sensors*. McGraw-Hill, Inc., 2008.

[40] G. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Springer Science & Business Media, 2011, vol. 2.

[41] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global positioning system: theory and practice*. Springer Science &amp; Business Media, 2012.

[42] M. Li, “Visual-inertial odometry on resource-constrained systems,” Ph.D. dissertation, UC Riverside, 2014.

[43] P. Barooah and J. P. Hespanha, "Estimation on graphs from relative measurements," *IEEE Control Systems Magazine*, vol. 27, no. 4, pp. 57–74, 2007.

[44] W. Lee, Y. Yang, and G. Huang, "Efficient multi-sensor aided inertial navigation with online calibration," in *Proc. of the IEEE International Conference on Robotics and Automation*, Xi'an, China, 2021.

[45] S. N. Alojaiman, "Tropospheric delay modeling using gnss observations from continuously operating reference stations (cors)," Ph.D. dissertation, Ohio State University, 2019.

[46] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7244–7251.

[47] 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.