GVINS: Tightly Coupled GNSS-Visual-Inertial Fusion for Smooth and Consistent State Estimation

Shaozu Cao, Xiuyuan Lu, and Shaojie Shen

***Abstract*—Visual-Inertial odometry (VIO) is known to suffer from drifting especially over long-term runs. In this paper, we present GVINS, a non-linear optimization based system that tightly fuses GNSS raw measurements with visual and inertial information for real-time and drift-free state estimation. Our system is aiming to provide accurate global 6-DoF estimation under complex indoor-outdoor environment where GNSS signals may be largely intercepted or even totally unavailable. To connect global measurements with local states, a coarse-to-fine initialization procedure is proposed to efficiently online calibrate the transformation and initialize GNSS states from only a short window of measurements. The GNSS pseudorange and Doppler shift measurements are then modelled and optimized under a factor graph framework along with visual and inertial constraints. For complex and GNSS-unfriendly areas, the degenerate cases are discussed and carefully handled to ensure robustness. The engineering challenges involved in the system are also included to facilitate relevant GNSS fusion researches. Thanks to the tightly-coupled multi-sensor approach and system design, our system fully exploits the merits of three types of sensors and is capable to seamlessly cope with the transition between indoor and outdoor environments, where satellites are lost and recaptured again. We extensively evaluate the proposed system by both simulation and real-world experiments, and the result demonstrates that our system substantially eliminates the drift of VIO and preserves the local accuracy in spite of noisy GNSS measurements. In addition, experiments also show that our system can gain from even a single satellite while conventional GNSS algorithms need four at lease.**

I. Introduction

LOCALIZATION is an essential functionality for many spatial-aware applications, such as autonomous driving, UAV navigation and augmented reality (AR). Estimating system states with various sensors has been widely studied for decades. Among these, the sensor fusion approach has been more and more popular in recent years. Due to the complementary properties provided by heterogeneous sensors, sensor fusion algorithms can significantly improve the accuracy and robustness of the state estimation system.

The camera provides rich visual information with only a low cost and small footprint, thus attracts much attention from both computer vision and robotics area. Combining with a MEMS IMU, which offers high frequency and outlier-free inertial measurement, Visual-Inertial Navigation (VIN) algorithms can often achieve high accuracy and be more robust in complex environments. Nevertheless, both camera and IMU operate in the local frame and it has been proven that the VIN system has four unobservable directions [1], namely x, y, z and yaw. Thus the odometry drift is inevitable for any VIN system. On the other hand, Global Navigation Satellite System (GNSS) provides a drift-free and global-aware solution for localization tasks, thus has been extensively used in various scenarios. GNSS signal is freely available and conveys the range information between the receiver and satellites. With at least 4 satellites being tracked simultaneously, the receiver is able to obtain its unique coordinate under the global Earth frame. Considering the complementary characteristics between VIN and GNSS system, it seems natural that improvements can be made by fusing information from both systems together.

However, many challenges exist during the fusion of two systems. Firstly, a stable initialization from noisy GNSS measurement is indispensable. Among quantities need to be initialized, the 4-DoF transformation between the local VIN frame and the global GNSS frame, which is necessary to associate measurements from multiple sources together, is important. Unlike the extrinsic transformation between camera and IMU, this transformation cannot be offline calibrated because each time VIN system starts such transformation will vary. In addition, one-shot alignment using a portion of sequence does not work well as the drift of the fusion system makes such alignment invalid during GNSS outage situations. Thus, an online initialization and calibration between local frame and global frame is necessary to fuse heterogeneous measurements and cope with complex indoor-outdoor environment. Secondly, the precision of the GNSS measurement does not match with that of VIN system, and various error sources exist during the GNSS signal propagation. In practice, the pseudorange measurement, which is used for global localization in GNSS system, can only achieve a meter-level precision while VIN system is capable to provide centimeter-level estimation over a short range. As a result, the fusion system will be susceptible to the unstable GNSS measurement if not formulated carefully. Thirdly, degeneration happens when the fusion system experiences certain movements such as pure rotation or the number of locked satellites is insufficient. Normally the GNSS-visual-inertial fusion system can offer a drift-free 6-DoF global estimation, but the conclusion no longer holds under degenerate cases. In addition, the transition between indoor and outdoor environments, during which all satellites are lost and gradually recaptured again, also poses challenges to the system design.

To address the above-mentioned issues, we propose a non-linear optimization-based system to tightly fuse GNSS raw measurements (pseudorange and Doppler frequency shift) with visual and inertial data for accurate and drift-free state esti-

All authors are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China. {shaozu.cao, xluaj}@connect.ust.hk, eeshaojie@ust.hk.

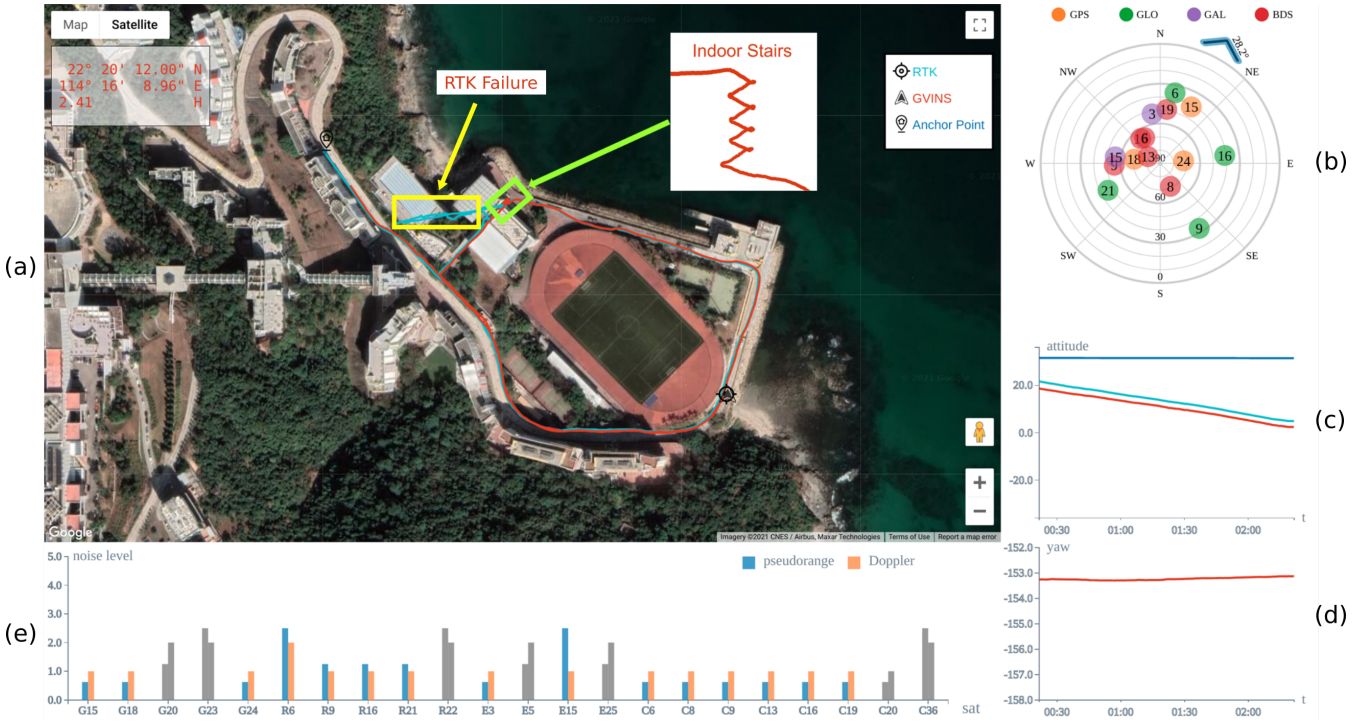

Fig. 1. A snapshot of our system in a complex indoor-outdoor environment. The global estimation result is directly plotted on Google map and aligns well with the ground truth RTK trajectory as shown in part (a). Part (b) depicts the distribution of satellites with tangential direction representing the azimuth and radial direction being the elevation angle. The blue arrow is a compass-like application which indicts the global yaw orientation of the camera. Subplot (c) and (d) illustrate the attitude information and the local-ENU yaw offset respectively. The measurement noise level of each tracked satellite is shown in part (e). Note that there is an obvious failure on the RTK trajectory when we walk the indoor stairs, while our system can still perform global estimation even in indoor environment.

mation. The 4-DoF transformation between local and global frames is recovered via a coarse-to-fine approach during initialization phase and is further optimized subsequently. To incorporate noisy GNSS raw measurements, all GNSS constraints are formulated under a uniform probabilistic framework in which all states are jointly optimized. In addition, degenerate cases are discussed and carefully handled to ensure robustness. Thanks to the tightly-coupled approach and system design, our system fully exploits the complementary properties among GNSS, visual and inertial measurements and is able to provide locally smooth and globally consistent estimation even in complex environments, as shown in Fig. 1. We highlight the contributions of this paper as follows:

- an online coarse-to-fine approach to initialize GNSS-visual-inertial states.
- an optimization-based, tightly-coupled approach to fuse visual-inertial data with multi-constellation GNSS raw measurements under the probabilistic framework.
- a real-time estimator which is capable to provide drift-free 6-DoF global estimation in complex environment where GNSS signals may be largely intercepted or even totally unavailable.
- an evaluation of the proposed system in both simulation and real-world environment.

For the benefit of the research community, the source code and datasets in this work will be made public[1].

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

The rest of this work is structured as follows: in Section II we discuss the existing relevant literature. Section III describes the notation and coordinate system involved in the system. In Section IV we briefly introduce relevant background knowledge of GNSS. Section V shows the structure and workflow of the proposed system. The problem formulation and methodology are illustrated in Section VI. In Section VII we address the GNSS initialization issues and discuss several degenerate cases that degrade the performance of our system. As we solve the problem from a system level, engineering challenges during the development of the system are listed in VIII for the benefit of the community. The experiment setup and evaluation are given in Section IX. Finally Section X concludes this paper.

II. Related Work

State estimation via multiple sensors fusion approach has been proven to be effective and robust, and there is extensive literature on this area. Among those, we are particularly interested in the combination of small size and low cost sensors such as camera, IMU and GNSS receiver, to produce a real-time accurate estimation in the unknown environment.

The fusion of visual and inertial measurement in a tightly-coupled manner can be classified into either filter-based method or optimization-based method. MSCKF [2] is an excellent filter-based state estimation which utilizes the geometric constraints between multiple camera poses to efficiently optimize the system states. Based on MSCKF, [3] makes improvements on its accuracy and consistency, and [4] aims

to overcome its numerical stability issue especially on mobile devices. Compared with the filter based approach, nonlinear batch optimization method can achieve better performance by re-linearization at the expense of computational cost. OKVIS [5] utilizes keyframe-based sliding window optimization approach for state estimation. VINS-Mono [6] also optimizes system states within the sliding window but is more complete with online relocalization and pose graph optimization. Since camera and IMU only impose relative constraint among states, accumulated drift is a critical issue in the VIN system, especially over long-term operation.

As GNSS provides absolute measurement in the global Earth frame, incorporating GNSS information is a natural way to reduce accumulated drift. In terms of loosely-coupled manner, [7] [8] describe state estimation systems which fuse GNSS solution with visual and inertial data under the EKF framework. [9] proposes a UKF algorithm that fuses visual, inertial, LiDAR and GNSS solution to produce a smooth and consistent trajectory under different environments. [10], [11] and our previous work VINS-Fusion [12] fuse the result from local VIO with GNSS solution under the optimization framework. All aforementioned works rely on the GNSS solution to perform estimation so system failure will occur once the GNSS solution is highly corrupted or unavailable under the situation where the number of tracked satellites is below than 4.

There are also some works on tightly fusing GNSS raw measurement with visual and inertial information. [13] and [14] combine camera, IMU and GNSS RTK measurement under the EKF framework for localization, but a static GNSS reference station is required to for the centimeter-level RTK solution. [15] and [16] investigates the performance of the fusion system under cluttered urban environment where less than 4 satellites are being tracked. However, the transformation between local and global frames is not handled and the scale of their real-world experiments is limited. In addition, the performance of the underlying VIN system in [16] is not good to fuse with GNSS. [17] tightly fuses GNSS pseudorange data and a sky-pointing camera measurement in the EKF manner. The upward-facing camera is used to filter multipath GNSS signal as well as tracking features from top buildings, thus is only suitable for the urban environment. The transformation between the local vehicle frame and the global frame is assumed known in their work. Recently we found a similar work [18] that tightly fuse GNSS raw measurements with visual-inertial SLAM. While it works well in outdoor environments, it is not able to handle indoor scenarios such as tunnels which limits the potential of the tightly multi-sensor fusion approach.

To this end, we aim to build a robust and accurate state estimator with GNSS raw measurements, visual and inertial data tightly fused. By leveraging the global measurement from GNSS, the accumulated error from visual-inertial system will be eliminated. The transformation between the local and global frame will be estimated without any offline calibration. The system is capable to work in complex indoor and outdoor environments and achieves local smoothness and global consistency.

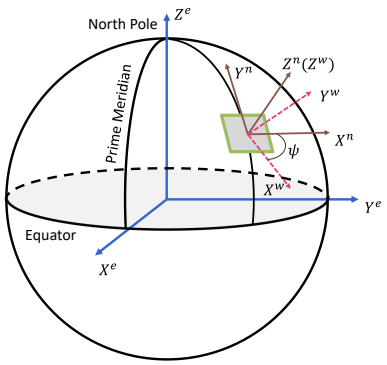

Fig. 2. An illustration of the local world, ECEF and ENU frames.

III. Notation and Definitions

A. Frames

The spatial frames involved in our system consist of:

1) Sensor Frame: Sensor frame is attached to the sensor and is a local frame in which sensor reports its reading. In our system, sensor frames include the camera frame $(\cdot)^c$ and the IMU frame $(\cdot)^i$, and we choose IMU frame as our estimation target frame and denote it as body frame $(\cdot)^b$.

2) Local World Frame: we represent the conventional frame in which visual-inertial system operates as the local world frame $(\cdot)^w$. In VIN system, the origin of the local world frame is arbitrarily set and the z axis is often chosen to be orthogonal to the local ground plane, as illustrated in Fig. 2.

3) ECEF Frame: The ECEF (Earth-Centered, Earth-Fixed) frame $(\cdot)^e$ is a Cartesian coordinate system that is fixed with respect to Earth. As shown in Fig. 2, the origin of ECEF frame is attached to the center of mass of Earth. The z axis is perpendicular to Earth's Equator and points to the true north. The x-y plane coincides with Earth's Equator with x axis points to the prime meridian.

4) ENU Frame: In order to connect the local world and global ECEF frames, a semi-global frame, ENU, is introduced. The x, y, z axis of the ENU frame $(\cdot)^n$ point to the east, north, and up direction respectively (Fig. 2). Given a point in ECEF frame, a unique ENU frame can be determined with its origin sitting on that point. Note that the both the z axis of the local world frame and ENU frame are parallel to the gravity direction.

In terms of temporal frames, GNSS data is tagged in GNSS time system (for example, GPST), while visual and inertial measurements are marked in the local time system. We assume that these two time systems are aligned beforehand and do not distinguish them accordingly[2].

B. States

The system states to be estimated include:

- the position $\mathbf{p}_b^w$ and orientation $\mathbf{q}_b^w$ of the body frame with respect to the local world frame,
- the velocity $\mathbf{v}_b^w$, accelerometer bias $\mathbf{b}_a$ and gyroscope bias $\mathbf{b}_w$,

[2]The temporal frame alignment is discussed later in Section. VIII-A.

- the inverse depth ρ for each feature,
- the yaw offset ψ between the local world frame and ENU frame, receiver clock bias $\delta\mathbf{t}$ and receiver clock drifting rate $\delta\dot{t}$. Because our system support all four constellations, the clock biases for GPS, GLONASS, Galileo and BeiDou are estimated separately. Note that the receiver clock drifting rate for each constellation is the same.

Our system adopt a sliding window optimization manner and states $\mathcal{X}$ inside the window can be summarized as:

$$\begin{aligned}\mathcal{X} &= [\mathbf{x}_0,\, \mathbf{x}_1,\, \cdots \mathbf{x}_n,\, \rho_0,\, \rho_1,\, \cdots \rho_m,\, \psi\,] \\ \mathbf{x}_k &= \left[\mathbf{p}_{b_k}^w,\, \mathbf{v}_{b_k}^w,\, \mathbf{q}_{b_k}^w,\, \mathbf{b}_a,\, \mathbf{b}_w,\, \delta\mathbf{t},\, \dot{\delta t}\right], k \in [0, n] \\ \delta\mathbf{t} &= [\delta t_G,\, \delta t_R,\, \delta t_E,\, \delta t_C]\,,\end{aligned} \tag{1}$$

Where n is the window size and m is the number of feature points in the window.

IV. GNSS Fundamentals

Since our system requires GNSS raw measurement processing, background knowledge about GNSS is necessary. In this section, we first give an overview about GNSS. Then the raw measurements, namely pseudorange and Doppler shift, are introduced and modelled. Finally the principle of SPP algorithm for global localization is described in the end of this section.

A. GNSS Overview

Global Navigation Satellite System (GNSS), as its name suggests, is a satellite-based system which is capable to provide global localization service. Currently there are four independent and fully operational systems, namely GPS, GLONASS, Galileo and BeiDou. Each GNSS system consists of control segment, satellite segment and user segment, with user segment consisting of an unlimited number of receivers. The satellite segment is a constellation of satellites orbiting Earth at an altitude of about 20,000 kilometers (except for BeiDou's GSO/IGSO satellites), with their status being monitored and updated by the control segment. The navigation satellite continuously transmits specific radio signal from which the receiver can uniquely identify the satellite and retrieve the navigation message. The typical structure of the GPS L1 signal is illustrated in Fig. 3. Each satellite has a unique Pseudo Random Number (PRN) code which repeats every 1ms. The navigation message (ephemeris), which contains satellite orbit and GNSS time parameters, is firstly combined with the PRN code and then used to modulate the high frequency carrier signal. After receiving the signal, the receiver obtains the Doppler shift (Section. IV-C) by measuring the frequency difference between the received one and designed one. The pseudorange measurement (Section. IV-B) is inferred from the PRN code shift which indicts the propagation time. Finally the navigation message is uncovered by a reverse demodulation process.

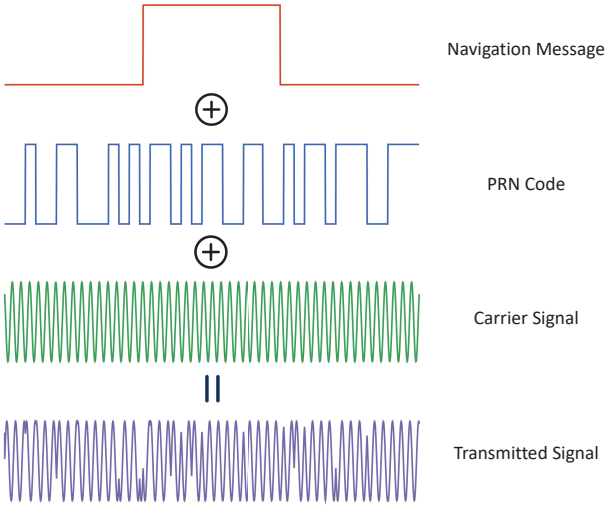

Fig. 3. The hierarchical structure of the GPS signal. The navigation message first mixes with the satellite-specific PRN code, and then the resulting sequence is used to modulate the high frequency carrier signal. The final signal is transmitted by the satellite and captured by the receiver, which applies a reverse process to obtain the measurement and retrieve the message.

B. Pseudorange Measurement

Upon the reception of the signal, the time of flight (ToF) of the signal is measured from the PRN code shift. By multiplying with the light of speed, the receiver obtains pseudorange measurement. The pseudorange is called "pseudo" because it not only contains the geometric distance between the satellite and the receiver, but also includes various errors during the signal generation, propagation and processing.

The error source on the satellite side mainly consists of satellite orbit and clock error. The orbit error comes from the influence of other celestial objects which are not precisely modelled by the ephemeris, and the clock error is the result of imperfect satellite onboard atomic clock with respect to the standard system time. The orbit and clock errors are monitored and constantly corrected by the system control segment. During the signal propagation from satellite to receiver, it goes through the ionosphere and troposphere, where the speed of the electromagnetic signal is no longer as same as that in vacuum and the signal gets delayed according to the atmosphere components and propagation path. The phenomenon that signal reaches the receiver with different ways, known as multipath effect, may occur and add extra delay especially for low elevation satellites. When the signal arrivals, the ToF is calculated by comparing the signal transmission time, which is marked by the satellite's atomic clock, with the receiver's less accurate local clock time. Thus the range information is also offset by the receiver clock bias with respect to the GNSS system time. In conclusion, the pseudorange measurement can be modelled as:

$$\begin{aligned}\tilde{P}_r^s =& \|\mathbf{p}_s^e - \mathbf{p}_r^e\| + c\,(\boldsymbol{\zeta}_s \delta\mathbf{t} - \Delta t^s) \\ &+ T_r^s + I_r^s + M_r^s + S_r^s + \epsilon_r^s,\end{aligned} \tag{2}$$

where $\mathbf{p}_s^e$ and $\mathbf{p}_r^e$ is the ECEF coordinate of the satellite s and receiver r, respectively. c represents the speed of light in vacuum. $\boldsymbol{\zeta}_s$ is designed to be a 4×1 indictor vector with the corresponding satellite constellation entity being 1 and other three entities being 0. Δt^s is the satellite clock error, which can be calculated from the ephemeris. T_r^s and I_r^s stand for the tropospheric and the ionospheric delay respectively. We use M_r^s to denote the delay caused by multipath effect and ϵ_r^s for the measurement noise. Because the GNSS reference frame, namely ECEF, is rotating along with the Earth, the propagation of GNSS signal must account for Sagnac effect [19]. The Sagnac term S_r^s is given by:

$$S_r^s = \frac{\omega_E}{c}\left(\left[\mathbf{p}_s^e\right]_x \left[\mathbf{p}_r^e\right]_y - \left[\mathbf{p}_s^e\right]_y \left[\mathbf{p}_r^e\right]_x\right), \tag{3}$$

where ω_E represents the angular velocity of the Earth and $\left[\cdot\right]_x$ and $\left[\cdot\right]_y$ extract the x and y component of a vector respectively.

C. Doppler Measurement

The Doppler frequency shift is measured from the difference between the received carrier signal and the designed one, and it reflects the receiver-satellite relative motion along the signal propagation path. Due to the characteristic of the GNSS signal structure, the accuracy of the Doppler measurement is usually an order of magnitude higher than that of pseudorange. The Doppler shift is modelled as:

$$\widetilde{\Delta \dot{f}}_r^s = -\frac{1}{\lambda}\left[\boldsymbol{\kappa}_r^{sT}(\mathbf{v}_s^e - \mathbf{v}_r^e) + c(\dot{\delta t} - \dot{\Delta t}^s)\right] + \eta_r^s, \tag{4}$$

where $\mathbf{v}_r^e$ and $\mathbf{v}_s^e$ represent the receiver and satellite velocity in ECEF frame respectively. We use λ to denote the wavelength of the carrier signal, and $\boldsymbol{\kappa}_r^s$ for the unit direction vector from receiver to satellite in ECEF frame. $\dot{\Delta t}^s$ is the drift rate of the satellite clock error which is reported in the ephemeris, and finally η_r^s represents the Doppler measurement noise.

D. SPP Algorithm

The Single Point Positioning (SPP) algorithm utilizes pseudorange measurement to determine the 3-DOF global position of the GNSS receiver via trilateration. Thus in theory the coordinate of the receiver can be obtained by the aid of 3 different satellites. However, as mentioned in Section IV-B, pseudorange measurement is offset by the receiver clock bias. Because the receiver clock bias can cause an error of hundreds of kilometers, it must be estimated along with the location in order to get a reasonable result. To this end, at least 4 pseudorange measurements are required to fully constraint the 3-DOF global position and receiver clock bias. Because different navigation systems use different time references, there exists clock offset between different systems. Additional measurements are necessary in order to estimate the inter-system clock offset if the satellites are from multiple constellations. To summarize, at least $(N+3)$ satellites are required to be simultaneously tracked in order to obtain the uniquely localize the receiver, where N is the number of constellations among the tracked satellites.

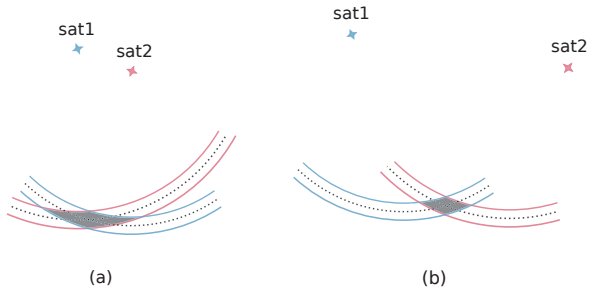

Fig. 4. A simplified 2D illustration of how satellites distribution affects the uncertainty of SPP solution. Here we assume the receiver-satellite are synchronized thus two satellites are enough for localization. The dash line represents ground truth pseudorange while the area in between the two solid lines denotes the possible noisy measurement. The uncertainty of final SPP solution is represented by the shadow area.

After collecting enough measurements, the constraints from Eq. 2 are stacked together to form a series of equations with $\mathbf{p}_r^e$ and $\delta\mathbf{t}$ unknown. Corrections are applied to pseudorange measurement making it only a function of $\mathbf{p}_r^e$ and $\delta\mathbf{t}$. In our system the tropospheric delay T_r^s is estimated by Saastamoinen model [20], and ionospheric delay I_r^s is computed using Klobuchar model [21] and parameters in ephemeris. By excluding the low elevation satellites, we ignore the delay M_r^s caused by multipath effect. In practice, more than $(N+3)$ measurements will be used in order to get a reliable solution and the problem is optimized by minimizing the sum of the squared residuals. As is shown in [22], the noise of the SPP solution not only depends on the measurement noise but also has a relationship with the geometric distribution of satellites, and such a relationship is also revelled later in Eq. 17. A simplified 2D case in Fig. 4 shows the effect of satellites distribution on the noise characteristic of the final solution. Thus the performance of SPP algorithm will be better with an evenly satellites distribution over the sky, even with the measurement noise unchanged.

V. System Overview

The structure of our proposed system is illustrated in Fig. 5. The estimator takes raw GNSS, IMU and camera measurements as input, and applies necessary preprocessing on each type of measurement afterwards. As in [6], the IMU measurements are pre-integrated and the whole image is summarized as a series of sparse feature points. For GNSS raw data, we first filter out low-elevation and unhealthy satellites which are prone to be erroneous. In order to reject unstable satellite signal, only satellites which are continuously locked for a certain amount of epochs are allowed to enter the system. Because the ephemeris data is acquired via the slow satellite-receiver wireless link (50 bit/s), a GNSS measurement is unusable until its corresponding ephemeris is fully transmitted. After the preprocessing phase, all measurements are ready for the estimator, but before stepping into the optimization part,

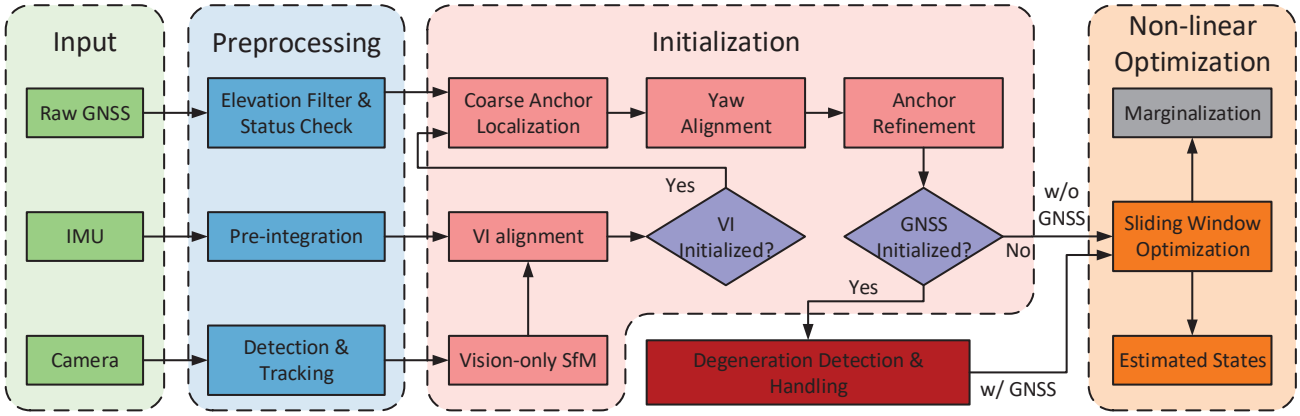

Fig. 5. The diagram above shows the workflow of our proposed system. At first measurements from all sensors are preprocessed before going into follow-up procedures. In the initialization stage, visual-inertial initialization is accomplished by aligning the inertial information with the result of vision-only SfM. If visual-inertial successfully gets aligned, a coarse-to-fine process is performed in order to initialize the GNSS states. The system monitors and handles GNSS degeneration cases once GNSS gets involved. Finally constraints from the measurements within the sliding window are optimized under the non-linear optimization framework. Note that if GNSS cannot get initialized, our system can still work in visual-inertial mode. The marginalization strategy is also adopted to ensure real-time estimation.

an initialization phase is necessary to properly initialize the system states of the non-linear estimator.

The initialization starts with a vision-only SfM, from which an up-to-similarity motion and structure are jointly estimated, then the trajectory from IMU is aligned to the SfM result in order to recover the scale, velocity, gravity and IMU bias. After VI initialization is finished, a coarse-to-fine GNSS initialization process is conducted. At first a coarse anchor localization result is obtained by the SPP algorithm, then the local and global frames are associated in the yaw alignment stage using the local velocity from VI initialization and GNSS Doppler measurement. Finally the initialization phase ends with the anchor refinement, which utilizes accurate local trajectory and imposes clock constraints to further refine the anchor's global position.

After the initialization phase, the GNSS degeneration cases are checked and carefully handled to ensure robust performance. Then constraints from all measurements are formulated to jointly estimate system states within the sliding window under the non-linear optimization framework. Note that our system is naturally degraded to a VIO if GNSS is not available or cannot be properly initialized. To ensure the real-time performance and handle visual-inertial degenerate motions, the two-way marginalization strategy is also applied after each optimization.

VI. Probabilistic Formulation

In this section, we first formulate and derive our state estimation problem under the probabilistic framework. As shown later, the whole problem is organized as a factor graph and measurements from sensors form a series of factors which in turn constraint the system states. Each type of factor in the probabilistic graph will be discussed in detail through this section. Note that the formulation of visual and inertial factors are inherited from [6] [23] [24] thus not the contribution of this work. The relevant content is listed only for the completeness of this literature.

A. MAP Estimation

We define the optimum system state as the one that maximizes a posterior (MAP) given all the measurements. Assuming that all measurements are independent to each other and the noise with each measurement is zero-mean Gaussian distributed, the MAP problem can be further transformed to the one that minimize the sum of a series of costs, with each cost corresponding to one specific measurement.

$$\begin{aligned}\mathcal{X}^{\star} &= \arg\max_{\mathcal{X}} p(\mathcal{X}|\mathbf{z}) \\ &= \arg\max_{\mathcal{X}} p(\mathcal{X})p(\mathbf{z}|\mathcal{X}) \\ &= \arg\max_{\mathcal{X}} p(\mathcal{X}) \prod_{i=0}^{n} p(\mathbf{z}_i|\mathcal{X}) \\ &= \arg\min_{\mathcal{X}} \left\{ \|\mathbf{r}_p - \mathbf{H}_p\mathcal{X}\|^2 + \sum_{i=0}^{n} \|\mathbf{r}(\mathbf{z}_i, \mathcal{X})\|^2_{\mathbf{P}_i} \right\},\end{aligned} \tag{5}$$

where $\mathbf{z}$ stands for the aggregation of n independent sensor measurements and $\{\mathbf{r}_p, \mathbf{H}_p\}$ encapsulates the prior information of the system state. $\mathbf{e}(\cdot)$ denotes the residual function of each each measurement and $\|\cdot\|_{\mathbf{P}}$ is the Mahalanobis norm.

Note that such formulation naturally fits with the factor graph representation [25], thus we decompose our optimization problem as individual factors that relate states and measurements as shown in Fig. 6. In the following we will discuss each factor in details.

B. Inertial Factor

The measurements involved in the inertial factor consist of the biased, noisy linear acceleration and angular velocity of the platform. As the accelerometer operates near the Earth's

surface, the linear acceleration measurement also contains the gravity component, as is listed below:

$$\begin{aligned}\tilde{\mathbf{a}}_t &= \mathbf{a}_t + \mathbf{b}_{a_t} + \mathbf{R}_w^t \mathbf{g}^w + \mathbf{n}_a \\ \tilde{\boldsymbol{\omega}}_t &= \boldsymbol{\omega}_t + \mathbf{b}_{w_t} + \mathbf{n}_w,\end{aligned} \tag{6}$$

where $\{\tilde{\mathbf{a}}_t,\ \tilde{\boldsymbol{\omega}}_t\}$ is the output of the IMU, and $\{\mathbf{a}_t,\ \boldsymbol{\omega}_t\}$ stands for the linear acceleration and angular velocity of the platform in IMU sensor frame. The additive noise $\mathbf{n}_a$ and $\mathbf{n}_w$ are assumed to be zero-mean Gaussian distributed, e.g., $\mathbf{n}_a \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_a)$, $\mathbf{n}_w \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_w)$. The slowly varying biases associated with the accelerometer and gyroscope are modelled as a random walk as follows:

$$\dot{\mathbf{b}}_{a_t} = \mathbf{n}_{b_a}, \quad \dot{\mathbf{b}}_{w_t} = \mathbf{n}_{b_w}, \tag{7}$$

with $\mathbf{n}_{b_a} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{b_a})$, $\mathbf{n}_{b_w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{b_w})$.

In practice, the frequency of IMU is often an oder of magnitude higher than that of camera, thus it is computationally intractable to estimate each state of the IMU measurements. To this end, IMU pre-integration approach [23] is adopted to aggregate multiple measurements into a single one. For inertial measurements within the time interval $[t_k, t_{k+1}]$, the derived measurements are computed as follows:

$$\begin{aligned}\boldsymbol{\alpha}_{b_{k+1}}^{b_k} &= \iint_{t\in[t_k,t_{k+1}]} \mathbf{R}_t^{b_k}(\tilde{\mathbf{a}}_t - \mathbf{b}_{a_t})dt^2 \\ \boldsymbol{\beta}_{b_{k+1}}^{b_k} &= \int_{t\in[t_k,t_{k+1}]} \mathbf{R}_t^{b_k}(\tilde{\mathbf{a}}_t - \mathbf{b}_{a_t})dt \\ \boldsymbol{\gamma}_{b_{k+1}}^{b_k} &= \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\boldsymbol{\Omega}(\tilde{\boldsymbol{\omega}}_t - \mathbf{b}_{w_t})\boldsymbol{\gamma}_t^k dt,\end{aligned} \tag{8}$$

with

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor\boldsymbol{\omega}\rfloor_\times & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix}, \lfloor\boldsymbol{\omega}\rfloor_\times = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{9}$$

Here b_k stands for the body frame in time t_k. $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ encapsulates the relative position, velocity and rotation information between frame b_k and b_{k+1}, and can be constructed without the initial position, velocity and rotation profiles given IMU biases. Finally the residual relates the system states and pre-integrated IMU measurements can be formulated as:

$$\begin{aligned}\mathbf{r}_{\mathcal{B}}(\tilde{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) &= \begin{bmatrix} \delta\boldsymbol{\alpha}_{b_{k+1}}^{b_k} \\ \delta\boldsymbol{\beta}_{b_{k+1}}^{b_k} \\ \delta\boldsymbol{\theta}_{b_{k+1}}^{b_k} \\ \delta\mathbf{b}_a \\ \delta\mathbf{b}_g \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_w^{b_k}(\mathbf{p}_{b_{k+1}}^w - \mathbf{p}_{b_k}^w + \frac{1}{2}\mathbf{g}^w\Delta t_k^2 - \mathbf{v}_{b_k}^w\Delta t_k) - \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} \\ \mathbf{R}_w^{b_k}(\mathbf{v}_{b_{k+1}}^w + \mathbf{g}^w\Delta t_k - \mathbf{v}_{b_k}^w) - \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \\ 2\left[\mathbf{q}_{b_k}^{w^{-1}} \otimes \mathbf{q}_{b_{k+1}}^w \otimes (\hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k})^{-1}\right]_{xyz} \\ \mathbf{b}_{ab_{k+1}} - \mathbf{b}_{ab_k} \\ \mathbf{b}_{wb_{k+1}} - \mathbf{b}_{wb_k} \end{bmatrix},\end{aligned} \tag{10}$$

where $\delta\boldsymbol{\theta}_{b_{k+1}}^{b_k}$ models the relative rotation error in 3D Euclidean space, and the $[\cdot]_{xyz}$ operation returns the imaginary part of a quaternion.

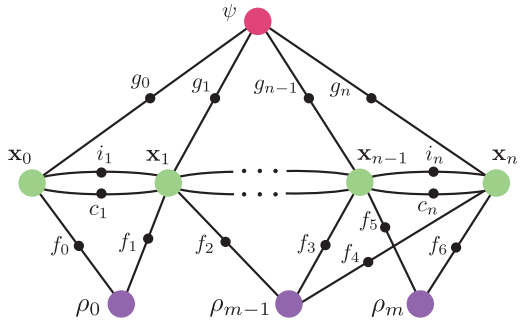

Fig. 6. Factor graph representation of the optimization problem in our system, where system states are denoted by large colored circles and factors are represented by small black circles. The factors from various measurements consist of inertial factor i, visual factor f, pseudorange and Doppler factor g and clock factor c.

C. Visual Factor

The visual measurement used in our system is a bunch of sparse feature points extracted from image frames. The strong corners [26] within the image are detected as feature points and are further tracked by the iterative Lucas-Kanade method [27]. After distortion correction [28] being applied to feature points, the projection process can be modelled as:

$$\widetilde{\mathcal{P}} = \pi_c(\mathbf{R}_b^c(\mathbf{R}_w^b\,\mathbf{x}^w + \mathbf{p}_w^b) + \mathbf{p}_b^c) + \mathbf{n}_c, \tag{11}$$

where $\widetilde{\mathcal{P}} = [u,\ v]^T$ is the feature coordinate in image plane, and $\mathbf{x}^w$ is its corresponding 3D landmark position in local world frame. $\pi_c(\cdot)$ represents the camera projection function and $\mathbf{n}_c$ is the measurement noise. Thus for a feature l with inverse depth ρ_l in frame i, if it is observed again in frame j, the residual that relates two frames can be expressed as:

$$\begin{aligned}\mathbf{r}_{\mathcal{C}}(\tilde{\mathbf{z}}_l,\ \mathcal{X}) &= \widetilde{\mathcal{P}}_l^{c_j} - \pi_c(\hat{\mathbf{x}}_l^{c_j}) \\ \hat{\mathbf{x}}_l^{c_j} &= \mathbf{R}_b^c(\mathbf{R}_w^{b_j}(\mathbf{R}_{b_i}^w(\mathbf{R}_c^b\frac{1}{\rho_l}{\pi_c}^{-1}(\widetilde{\mathcal{P}}_l^{c_i}) + \mathbf{p}_c^b) + \\ &\quad \mathbf{p}_{b_i}^w) + \mathbf{p}_w^{b_j}) + \mathbf{p}_b^c,\end{aligned} \tag{12}$$

with $\{\mathbf{R}_c^b, \mathbf{t}_c^b\}$ the transformation between IMU and camera.

D. Pseudorange Factor

Consider a GNSS receiver r which locks a navigation satellite s, it measures the code shift to obtain the pseudorange information as illustrated in Eq. (2). The satellite clock error and atmospheric delay are compensated using the models described in Section IV-D. In our system, the pseudorange noise ϵ_r^s is assumed to be zero-mean Gaussian distributed such as $\epsilon_r^s \sim N(0, \sigma_{r,pr}^s)$, where the variance $\sigma_{r,pr}^s$ is modelled as:

$$\sigma_{r,pr}^s = \frac{n_s \times n_{pr}}{\sin^2\theta_{el}}\ . \tag{13}$$

Here n_s is the broadcast satellite space accuracy index, and n_{pr} is the pseudorange measurement noise index reported by the receiver. θ_{el} represents the satellite elevation angle at the view of the receiver, and there are two reasons for this denominator term. Firstly it can suppress the noise caused by

GNSS multiple path effect that usually occurs on low elevation satellites. Furthermore, the ionospheric delay obtained by Klobuchar model, which is widely adopted by navigation system, still contains an error up to 50% [21]. As the low elevation satellites will experience a significant ionospheric delay, the denominator term can also reduce the error come with ionospheric compensation.

The receiver's ECEF coordinate can be mapped to the corresponding coordinate in local world frame via an anchor point, at which the ENU frame is built. Given the ECEF coordinate of the anchor point, the rotation from ENU frame to ECEF frame is:

$$\mathbf{R}_n^e = \begin{bmatrix} -\sin\lambda & -\sin\phi\cos\lambda & \cos\phi\cos\lambda \\ \cos\lambda & -\sin\phi\sin\lambda & \cos\phi\sin\lambda \\ 0 & \cos\phi & \sin\phi \end{bmatrix} , \tag{14}$$

where ϕ and λ is the latitude and longitude of the reference point in geographic coordinate system. The 1-DOF rotation between ENU and local world frame $\mathbf{R}_w^n$ is given by the yaw offset ψ. Then the relationship between ECEF coordinate and local world coordinates can be expressed as:

$$\mathbf{p}_r^e = \mathbf{R}_n^e \mathbf{R}_w^n (\mathbf{p}_r^w - \mathbf{p}_{anc}^w) + \mathbf{p}_{anc}^e \ . \tag{15}$$

In our implementation we set the anchor point to the origin of the local world frame, that is, the origin of the local world frame coincides with the origin of the ENU frame, as illustrated in figure 2. In addition, the offset between the receiver's antenna and IMU, which is only several centimeters on our platform, is omitted so we do not distinguish between $\mathbf{p}_r^w$ and $\mathbf{p}_b^w$. Then the residual of pseudorange measurement in t_k, which connects body b_k and satellite s_j, can be formulated as:

$$\begin{aligned} r_{\mathcal{P}}(\tilde{\mathbf{z}}_{r_k}^{s_j}, \mathcal{X}) = & \|\mathbf{p}_{s_j}^e - \mathbf{R}_n^e \mathbf{R}_w^n \mathbf{p}_{b_k}^w - \mathbf{p}_{anc}^e\| + \\ & c(\boldsymbol{\zeta}_s^T \delta\mathbf{t}_k - \Delta t^{s_j}) + T_{r_k}^{s_j} + \\ & I_{r_k}^{s_j} + S_{r_k}^{s_j} - \tilde{P}_{r_k}^{s_j} \ . \end{aligned} \tag{16}$$

The Jacobian with respect to states $\begin{bmatrix}\mathbf{p}_{b_k}^w & \delta\mathbf{t}_k & \psi\end{bmatrix}^T$ is given by:

$$\mathbf{J}_{r_k,pr}^{s_j} = \begin{bmatrix} -\boldsymbol{\kappa}_{r_k}^{s_j\,T} \mathbf{R}_n^e \mathbf{R}_w^n & c\boldsymbol{\zeta}_s^T & -\boldsymbol{\kappa}_{r_k}^{s_j\,T} \mathbf{R}_n^e \mathbf{G} \mathbf{p}_{b_k}^w \end{bmatrix} , \tag{17}$$

where $\mathbf{G}$ is:

$$\mathbf{G} = \begin{bmatrix} -\sin\psi & -\cos\psi & 0 \\ \cos\psi & -\sin\psi & 0 \\ 0 & 0 & 0 \end{bmatrix} . \tag{18}$$

E. Doppler Factor

The Doppler frequency shift, as shown in Eq. (4), reflects the relative velocity along the line of the signal propagation path between receiver and satellite. Similar to pseudorange noise, the Doppler measurement noise $\eta_{r,dp}^s$ is assumed to be Gaussian distributed and the corresponding variance is modelled as:

$$\sigma_{r,dp}^s = \frac{n_s \times n_{dp}}{\sin^2\theta_{el}} \ , \tag{19}$$

where n_{dp} is the measurement noise index reported by the receiver. The receiver's velocity in ECEF frame can be obtained from the local world velocity via:

$$\mathbf{v}_r^e = \mathbf{R}_n^e \mathbf{R}_w^n \mathbf{v}_r^w \ . \tag{20}$$

Here $\mathbf{v}_r^w$ is regarded as $\mathbf{v}_b^w$. Finally the residual with related to Doppler measurement in t_k, which connects body b_k and satellite s_j, can be formulated as:

$$\begin{aligned} r_{\mathcal{D}}(\tilde{\mathbf{z}}_{r_k}^{s_j}, \mathcal{X}) = & \frac{1}{\lambda} \boldsymbol{\kappa}_{r_k}^{s_j\,T} (\mathbf{v}_{s_j}^e - \mathbf{R}_n^e \mathbf{R}_w^n \mathbf{v}_{b_k}^w) + \\ & \frac{c}{\lambda}(\dot{\delta t}_k - \dot{\Delta t}^{s_j}) + \widetilde{\Delta f}_{r_k}^{s_j} \ . \end{aligned} \tag{21}$$

The Jacobian with respect to states $\begin{bmatrix}\mathbf{v}_{b_k}^w & \dot{\delta t}_k & \psi\end{bmatrix}^T$ is listed as:

$$\mathbf{J}_{r_k,dp}^{s_j} = \frac{1}{\lambda} \begin{bmatrix} -\boldsymbol{\kappa}_{r_k}^{s_j\,T} \mathbf{R}_n^e \mathbf{R}_w^n & c & -\boldsymbol{\kappa}_{r_k}^{s_j\,T} \mathbf{R}_n^e \mathbf{G} \mathbf{v}_{b_k}^w \end{bmatrix} . \tag{22}$$

F. Receiver clock factors

The receiver clock biases in t_k and t_{k+1} relate the clock drift rate by:

$$\delta\mathbf{t}_k = \delta\mathbf{t}_{k-1} + \mathbf{1}_{4\times1} \int_{t_{k-1}}^{t_k} \dot{\delta t}\, dt \ , \tag{23}$$

where $\mathbf{1}_{n\times m}$ stands for n by m all-ones matrix, and the residual in discrete case is:

$$\mathbf{r}_{\mathcal{T}}(\tilde{\mathbf{z}}_{k-1}^k, \mathcal{X}) = \delta\mathbf{t}_k - \delta\mathbf{t}_{k-1} - \mathbf{1}_{4\times1} \dot{\delta t}_{k-1} \tau_{k-1}^k \ , \tag{24}$$

where τ_{k-1}^k is the time difference between measurement $k-1$ and k. The covariance matrix associate with this residual is defined as a 4 by 4 diagonal matrix $\mathbf{D}_{t,k}$ with its elements describe the discretization error.

The GNSS receiver clock drift rate, on the other hand, is determined by the frequency stability of the receiver clock. TCXO is often chosen as the clock source on low-cost GNSS receivers. Due to the noise characteristic of TCXO, the receiver clock drift rate is modelled as a random walk process, thus the residual becomes:

$$r_{\mathcal{W}}(\tilde{\mathbf{z}}_{k-1}^k, \mathcal{X}) = \dot{\delta t}_k - \dot{\delta t}_{k-1} \ , \tag{25}$$

The corresponding variance $\sigma_{dt,k}$ is determined by the stability of the clock frequency drift.

VII. GNSS Initialization and Degeneration

The state estimation process described in the last section is non-linear with respect to the system states thus its performance relies heavily on the initial values. With online initialization, the initial states can be well recovered from an unknown situation without any assumption or manual intervention. During the system operation, the estimator may also encounter imperfect situations where some of sensors experience failure or degeneration. As there is already extensive literature on the topics of initialization and degeneration with respect to the visual-inertial system, in this section we limit the scope to the GNSS part. In the following we first introduce the proposed coarse-to-fine GNSS initialization approach, then we discuss several scenarios that degrade the performance of our system.

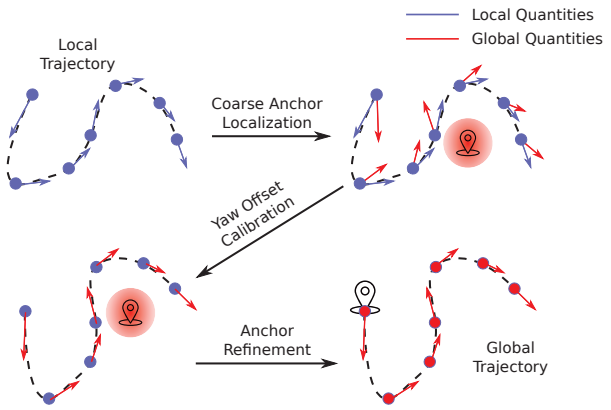

Fig. 7. An illustration of the proposed coarse-to-fine initialization process. The module takes the local position and velocity result from VIO and outputs the corresponding trajectory in global ECEF frame.

A. Initialization

As mentioned before, an anchor point with known global and local coordinate is necessary to fuse the global GNSS measurement with the local visual and inertial information. As the anchor point is already set to the origin of the local world frame, the ECEF coordinate of the local world origin need to be calibrated beforehand. In this paper, we propose a multi-stage GNSS-VI initialization procedure to online calibrate the anchor point and the yaw offset ψ between ENU and local world frame. Before the GNSS-VI initialization, we assume that the VIO has been successfully initialized, i.e. the gravity vector, initial velocity, initial IMU bias and scale have obtained an initial value [29]. After that, a smooth trajectory in the local world frame is formed and is ready to be used in the GNSS-VI initialization phase. As illustrated in Fig. 7, the online GNSS-VI initialization is conducted in a coarse-to-fine manner and consists of following three steps:

1) Coarse Anchor Point Localization: At first a coarse ECEF coordinate is generated by the GNSS SPP algorithm without any prior information. The SPP algorithm takes all pseudorange measurements from the sliding window as input. Since the measurements within the sliding window are time and spatial spanned, the result from SPP is an average location of the current window.

2) Yaw Offset Calibration: In the second step, we calibrate the yaw offset between the ENU frame and the local world frame using the less noisy Doppler measurement. The initial yaw offset and receiver clock drift rate are obtained through the following optimization problem:

$$\underset{\dot{\delta t},\psi}{\text{minimize}} \sum_{k=1}^{n}\sum_{j=1}^{p_k} \left\| r_{\mathcal{D}}(\tilde{\mathbf{z}}_{r_k}^{s_j},\, \mathcal{X}) \right\|^2_{\sigma^{s_j}_{\dot{r}_k,dp}} \tag{26}$$

where n is the sliding window size and p_k is the number of satellites observed in k-th epoch inside the window. Here we fix the velocity $\mathbf{v}_b^w$ to the result of VIO and assume that $\dot{\delta t}_k$ is constant within the window. The coarse anchor coordinate obtained from the first step is used to calculated the direction vector $\boldsymbol{\kappa}_r^s$ and rotation $\mathbf{R}_n^e$. $\boldsymbol{\kappa}_r^s$ and $\mathbf{R}_n^e$ are not sensitive to the receiver's location thus a coarse anchor point coordinate is sufficient. The parameters to be estimated only include the yaw offset ψ and the average clock bias drift rate $\dot{\delta t}$ over the entire window measurements. After that, the transformation between the ENU frame and local world frame is fully calibrated.

3) Anchor Point Refinement: Finally we are ready to refine the previous coarse anchor point and align the local world trajectory with that in ECEF frame. Different from the first step, the position result from VIO is used as prior information. The following problem is optimized over the sliding window measurements:

$$\begin{aligned}\underset{\delta\mathbf{t},\,\mathbf{p}^e_{anc}}{\text{minimize}} & \sum_{k=1}^{n}\sum_{j=1}^{p_k} \left\| r_{\mathcal{P}}(\tilde{\mathbf{z}}_{r_k}^{s_j},\, \mathcal{X}) \right\|^2_{\sigma^{s_j}_{r_k,pr}} + \\ & \sum_{k=1}^{n} \left\| \mathbf{r}_{\mathcal{T}}(\tilde{\mathbf{z}}_{k-1}^{k},\, \mathcal{X}) \right\|^2_{\mathbf{D}_{t,k}}\end{aligned} \tag{27}$$

The anchor point coordinate and the receiver clock biases associate with each GNSS epoch are refined through the optimization of the above problem. After this step, the anchor point, origin of the ENU frame, is set to the origin of the local world frame. Finally the initialization phase of the entire estimator is finished and all necessary initial quantities have been generated to boot the system up.

B. Degenerate Cases

There is no doubt that our fusion system will perform best in an open-area where GNSS signal is stable and satellites are well-distributed. In the following we will discuss several situations which may degrade the performance of our system.

1) Low speed movement: Since the noise level of Doppler shift measurement is an order of magnitude lower than that of pseudorange, the yaw offset between the local world frame and ENU frame can be well constrained by a short window of Doppler shift measurements. Once the velocity of the GNSS receiver is below the noise level of the Doppler shift, the estimated yaw offset may be corrupted by the measurement noise. In addition, low speed movement also implies that the translational distance within the window is short, thus the yaw estimation may be affected by pseudorange as well. In an extreme case where the platform experiences a rotation-only movement, GNSS cannot provide any information on the rotational directions and in turn the yaw component will drift as that in VIO. Thus we fix the yaw offset variable if the average velocity inside the window is below the threshold v_{ths}. In our system, v_{ths} is set to $0.5\ m/s$ which can be easily satisfied even by a pedestrian.

2) Less than 4 satellites being tracked: If the number of satellites being tracked is less than 4, the SPP or loosely-coupled approaches will fail to resolve the receiver's location. However, with the help of the tightly-coupled structure, our system is still able to make use of available satellites and subsequently update the states vector. Later in Section IX-B we will investigate the performance degradations under various satellite configurations.

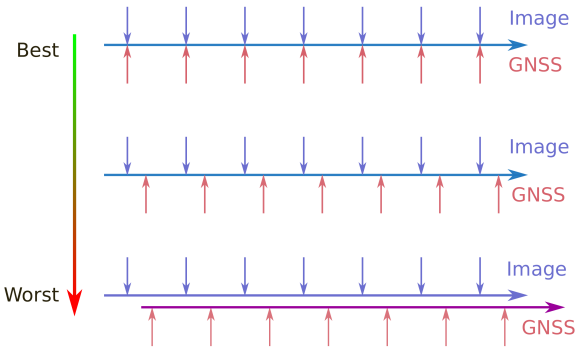

Fig. 8. An illustration of different synchronization levels between visual and GNSS measurements.

3) No GNSS signal: In indoor or cluttered environments where GNSS signal is totally unavailable, the states related to global information, namely the yaw offset ψ, receiver clock bias $\delta\mathbf{t}$ and drift rate $\dot{\delta t}$ are no longer observable. However, constraints from Eq. (24) and (25) are still kept during the optimization. The clock drift rate of low-cost receivers is quite stable as we found in the receiver stand-still analysis, thus the (near-)optimum clock drift rate is maintained by the constraint (25). Similarly, the receiver bias is propagated by constraint (24), which in turn provides a good initial value when the GNSS signal is recaptured. This mechanism improves the stability of our fusion system when the GNSS signal is fickle and eliminates the need for re-initialization when signal is lost-and-recaptured.

VIII. Engineering Challenges

The integration of multiple heterogeneous sensors often imposes a lot of challenges, such as time synchronization, calibration etc., which may bring significant errors into the estimator if not properly considered. In this section we will discuss several engineering challenges during the development of the proposed multi-sensor fusion system.

A. Time Synchronization

It is already shown in [30] that a millisecond-level misalignment between visual and inertial measurement may severely impact the performance of the VINS system. Unfortunately, the situation gets even worse for a GNSS involved multi-sensor system because time information is crucial to GNSS and the error may be magnified by the speed of light if time is misused. As mentioned earlier, GNSS system operates in its own time system while the local sensors such as camera and IMU work in the local time system, thus properly alignment between GNSS and local time system is necessary for system integration.

According to the impact on the system performance, the alignment between GNSS and visual inertial measurements can be classified into three levels as shown in Fig. 8. In the ideal case the measurements from different sensors are marked in a unified time system and are hardware-triggered via a pulse signal. The situation becomes worse if measurements are captured at different times but still under the same time system, thus interpolation must be applied to properly fuse measurements of different timestamps. The worst case for synchronization is that the two time systems are not aligned thus the time relationship between GNSS and visual inertial measurements is totally unknown. In this case, the estimation result is unreliable and the system behavior is unpredictable.

In our system, the two time systems are aligned via hardware trigger. The PPS signal generated by the GNSS receiver, which has an accuracy at nanosecond level, is fed into VI-Sensor. Once the VI-Sensor detects the edge of the time pulse, it reports the corresponding timestamp in its local time system. Meanwhile, the time information associated with the PPS signal is offered by the receiver, thus the global GNSS time system can be well aligned with the local one. Nevertheless, GNSS and VI measurements are still captured at different time in our current system and the gap between two types of measurement depends on the boot time of VI-Sensor. To this end, we apply linear interpolation on system states to minimize the effect of different capture times.

B. Electromagnetic Interference

The power of the GNSS signal perceived by receiver's antenna is very low and totally drowns in the thermal noise. As a result, the performance of the receiver may be affected by the inference of nearby RF sources. Note that although the specific band has been exclusively kept for GNSS usage, many devices still emit strong RF power into GNSS band unintentionally. As pointed out by [31], many modern digital equipments such as computers or cameras tend to generate a broad frequency spectrum which in turn overlaps with GNSS band. In addition, the I/O cables of peripheral devices may serve as an antenna and transfer noisy RF signal to the receiver subsequently. Unfortunately, the power level of such inference RF signals is usually much greater than that of the near-surface GNSS signal, so Electromagnetic Interference (EMI) must be seriously considered. During our earlier experiments we found that some of USB3 cameras severely degraded the performance of the receiver during image data transfer. With proper grounding design and shielding to cables and connectors, the interference can be mitigated to a certain extent. Later when we switch to VI-Sensor for the time synchronization purpose, the interference issue no longer exists as it deliveries visual and inertial data via Ethernet cable.

C. Receiver Clock Jump

The clock of the low-cost receiver often suffer from accumulated drifting due to the frequency stability of the underlying cheap oscillator. When the drift reaches a certain level, some manufacturers will choose to reset the clock to prevent a large deviation from GNSS time. Such operation will cause a clock jumps at the magnitude of microsecond or millisecond level, and the corresponding discontinuity can be observed on the pseudorange measurement as is obtained via signal's ToF [32]. By magnifying by the speed of light, such discontinuity may

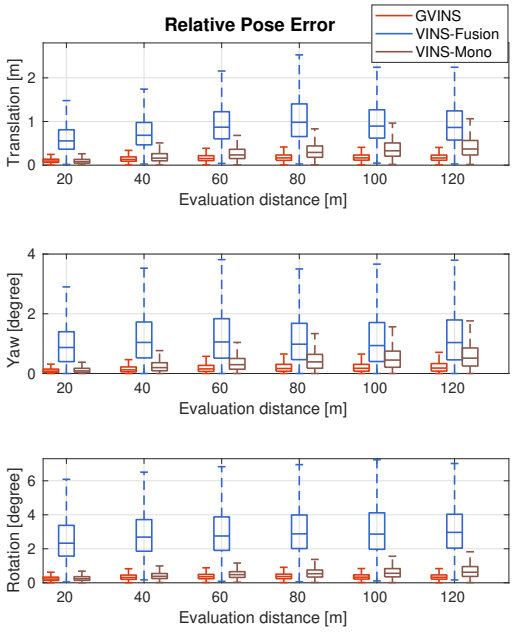

Fig. 9. Relative pose error of GVINS, VINS-Fusion and VINS-Mono with respect to the evaluation distance on the simulation environment. The top two figures correspond to the four unobservable directions (x, y, z and yaw) of VIO and the bottom figure is the overall relative rotation error.

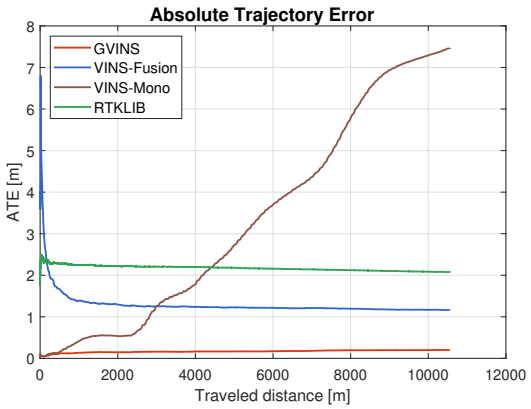

Fig. 10. Absolute trajectory error of GVINS, VINS-Fusion, VINS-Mono and RTKLIB with respect to the traveled distance on the simulation environment.

reach the level of $10^6\,m$, which in turn brings the system into failure. Note that as we introduce constraints between receiver clocks within the sliding window, the situation will get even worse. With our platform we observed a $20\,ms$ clock jump along with a pseudorange gap of 2×10^6 meters during a long-duration static experiment. To this end the data sequence with clock jumps happened must be avoided or carefully handled.

IX. Experimental Results

We conduct both simulation and real-world experiments to verify the performance of our proposed system. In this section, we compare our system with VINS-Mono [6], VINS-Fusion [12] (Monocular+IMU+GNSS) and RTKLIB [33]. Since we are only interested in the real-time estimation result, the loop function of VINS-Mono and VINS-Fusion is disabled. We use RTKLIB[3] to compute the GNSS SPP solution and feed the obtained GNSS location to VINS-Fusion for a loose-coupled result.

A. Simulation

1) Setup: The simulation environment is a $30m \times 30m \times 30m$ cube with random generated 3D landmarks. The landmarks are projected to a 10-Hz virtual camera with 75 degree horizontal FOV and 55 degree vertical FOV, which in turn generates around 100 visible features per frame. An additional white noise term with a standard deviation of 0.5 pixel is added to all feature points. A virtual 200-Hz IMU is rigidly connected to the camera and moves along a pre-designed 3D path. The standard deviation associate with the white noise of the accelerometer and gyroscope is set to $0.05m/s^2$ and $0.005rad/s$ respectively, and the standard deviation of the accelerometer and gyroscope bias random walk is set to $3.5\times10^{-4}m/s^2$ and $3.5\times10^{-5}rad/s$ respectively. In the meantime, a 10-Hz virtual GNSS receiver generates pseudorange and Doppler shift measurements using the past or real-time broadcast ephemeris data. The standard deviation of pseudorange and Doppler white noise shift is set to $1m$ and 0.5 Hz ($\sim 0.1m/s$ equivalent) respectively. The simulation experiment lasts for 30 minutes, with a trajectory over 10 kilometers. The maximum velocity and acceleration during the experiment is $10m/s$ and $6m/s^2$ respectively.

2) result: Fig. 9 shows the relative pose error (RPE) [34] with respect to the evaluation distance. As can be seen from the figure, The relative error of VINS-Mono increases with the evaluation distance in both translational and rotational directions. Among those the rotational error mainly comes from yaw component. This indicts that VINS-Mono suffers from accumulated drift in the four unobservable directions, namely x, y, z and yaw. The error of VINS-Fusion exhibits similar tendency when the evaluation distance is short, and remains at a constant level when the distance increases further. This implies that VINS-Fusion is able to bound the accumulated drift by loosely incorporating the GNSS solution. However, the magnitude of its relative error is much larger compared with the result of VINS-Mono and GVINS, thus the smoothness of the estimator is highly affected by the noisy GNSS measurement. Thanks to the tightly-coupled approach we adopted, our proposed system combines advantages of both VINS-Mono and VINS-Fusion. On the one hand, the relative error is comparable to that of VINS-Mono for short range thus the smoothness is preserved. On the other hand, the error no longer accumulates in all directions and the global consistency is also guaranteed.

Fig. 10 depicts the absolute trajectory error (ATE) along with the traveled distance. The error plot of VINS-Mono keeps increasing as a result of accumulated drift, while it remains constant for all other three approaches. The ATE of RTKLIB SPP algorithm shows the noise level of the

[3]https://github.com/tomojitakasu/RTKLIB/tree/rtklib_2.4.3

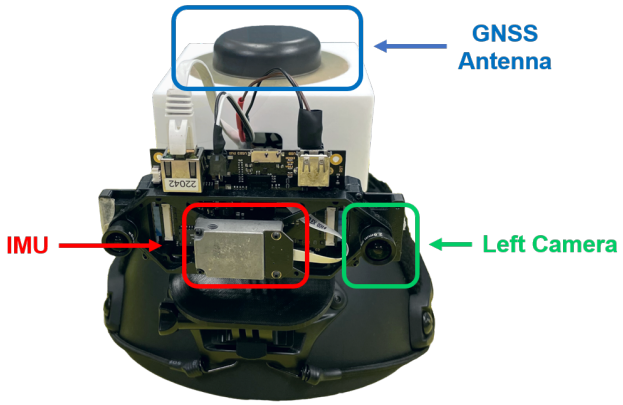

Fig. 11. The equipment used in our real-world experiments is a helmet with a VI-Sensor and a u-blox ZED-F9P attached. The camera and IMU measurements are well synchronized by VI-Sensor itself. The PPS signal from the GNSS receiver is used to trigger the VI-Sensor to align the global time with the local time.

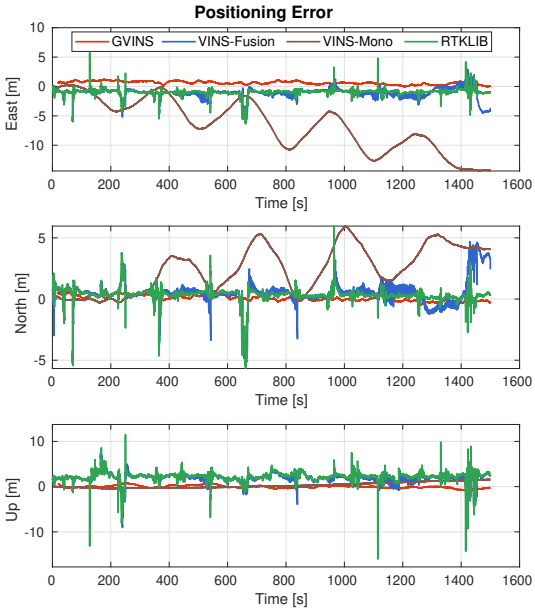

Fig. 12. Positioning error of GVINS, VINS-Fusion, VINS-Mono and RTKLIB at the sports field experiment. The three sub-figures correspond to the three directions of ENU frame. The result from GVINS, VINS-Fusion and RTKLIB are compared directly against the RTK ground truth without any alignment, while the result from VINS-Mono is aligned to the ground truth trajectory beforehand.

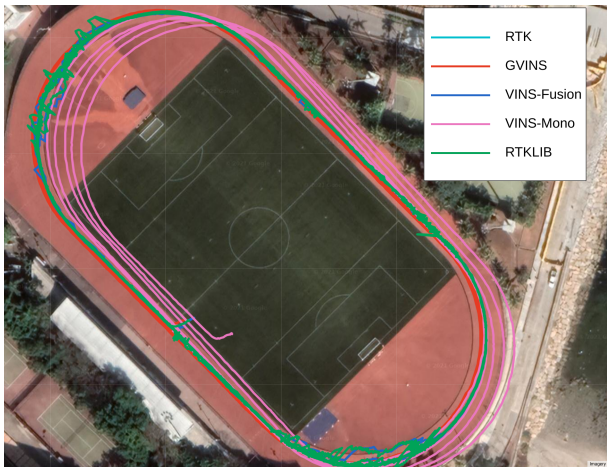

Fig. 13. The trajectory of RTK, GVINS, VINS-Fusion, VINS-Mono and RTKLIB in the sports field experiment. The resulting trajectory of our proposed system is smooth and aligns well with that of the RTK.

GNSS pseudorange measurement, and VINS-Fusion is able to reduce the magnitude of ATE by combine the result of VIO in a loosely-coupled manner. By tightly fusing GNSS raw measurements and visual inertial data in a unified framework, our algorithm effectively suppresses the noise of GNSS signal and keeps the ATE at a low level. The final root mean squared error (RMSE) of each approach is shown as the end point of each error curve.

B. Real-world Experiments

1) Setup: As illustrated in Fig. 11, the device used in our real-world experiments is a helmet with a VI-Sensor [35] and an u-blox ZED-F9P GNSS receiver [4] attached. In terms of image sensor, only the left camera of VI-Sensor is used during experiments. The u-blox ZED-F9P is a low-cost multi-band receiver with multiple constellations support. In addition, ZED-F9P owns an internal RTK engine which is capable to provide receiver's location at an accuracy of 1cm in open area. The real-time RTCM stream from a 3km away base station is fed to to the ZED-F9P receiver for the ground truth RTK solution. The local time is aligned with the global GNSS time via the method described in Section VIII-A.

2) Sports Field Experiment: This experiment is conducted on a sports field at our campus where we follow an athletic track for 5 laps. The sports field is a typical outdoor environment with an opened area on one side and some buildings the other side. During the experiment most of the satellites are well locked and the status of RTK remains fix throughout the whole path. In this experiment the global consistency of our estimator is examined against the repeated trajectory and the unstable signal near buildings also poses challenges to the local smoothness of the result.

The positioning error of this experiment is plotted against ENU axes as depicted in Fig. 12. A reference point, which is used to transform the ECEF result to a ENU frame, is arbitrarily selected on the sports field. Since VINS-Fusion, RTKLIB and our system can directly output estimation results in ECEF frame, we do not apply any alignment for their trajectories. For VINS-Mono which only gives results in local frame, we perform a 4-DOF alignment between its trajectory

[4]https://www.u-blox.com/en/product/zed-f9p-module

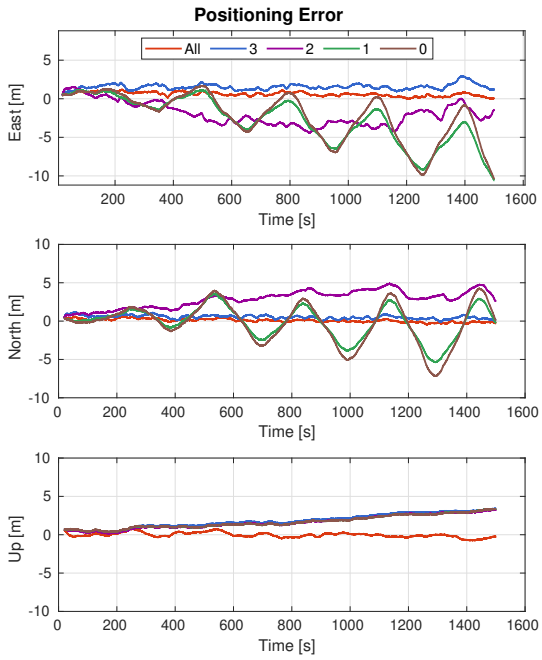

Fig. 14. Positioning error of our proposed system under situations where the number of locked satellites is insufficient.

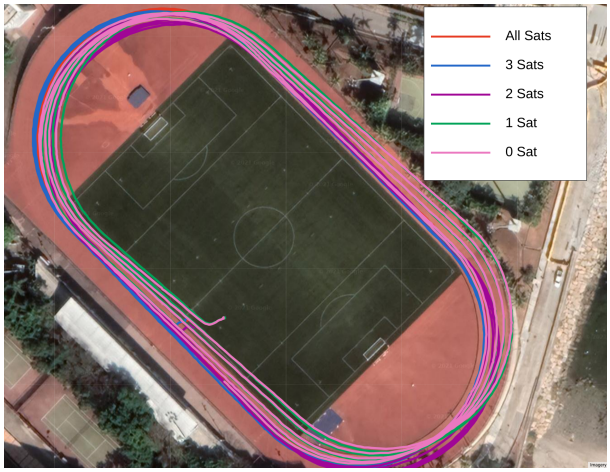

Fig. 15. The trajectories of our proposed system under different satellite configurations. GVINS performs best by utilizing all available satellites, and degrades to VIO with zero satellite configuration. A small bias occurs when only 3 satellites are employed, and translational drift emerges when the satellite number is further reduced to 2. If there is only 1 satellite available, yaw estimation starts to drift as well, but with a smaller magnitude compared with VIO (0 satellite).

and the ENU path of RTK using the first 2000 poses. Note that the global positioning results from VINS-Fusion, RTKLIB and our system suffer from a certain bias due to satellite ephemeris error, inaccurate atmospheric delay modeling and multi-path effect, while the aligned VINS-Mono result does not have such issue because of the pre-alignment.

TABLE I
RMSE[M] STATISTICS

	GVINS	VINS-Fusion	VINS-Mono	RTKLIB
Sports field	0.776	2.861	8.400	2.842
Indoor-outdoor	3.540	5.134	36.734	6.995

From Fig. 12 we see that VINS-Mono suffers from drifting among all three directions. In addition, the periodic fluctuations on horizontal directions (east and north) implies an obvious drift on the yaw estimation. On the other hand, the SPP solution from RTKLIB does not drift at all, but is highly affected by the noisy GNSS measurement. The error of VINS-Fusion is bounded as a result of combining the global information from SPP result. However, the local accuracy oscillates a lot and the local smoothness is ruined in the meantime. As a comparison, the positioning error of our proposed system does not grow with the traveled distance and is always maintained at a low level. Meanwhile, the error varies slowly and continuously, which also indicates our system effectively suppresses the noise from unstable GNSS signals. Table. I lists the RMSE statistics of this experiment and Fig. 13 shows the final trajectory on Google map. The resulting paths of 5 trips from our system overlap with each other and align well with those of RTK. Through this experiment, we state that our system is able to achieve global consistency to eliminate drifts of VIO and also preserve the local smoothness under noisy GNSS conditions.

3) Insufficient Satellites Experiment: Based on the data sequence of sports field experiment, we further investigated the degenerate case where the number of tracked satellites is less than 4. Normally there are about 20 satellites being locked in this sequence, and we intentionally remove most of the satellites in the non-linear optimization phase in order to test the system behavior. Starting from no-satellite setting, we sequentially add satellite G2, G13 and G5 to the system to emulate the one, two, and three-satellite situations respectively. It is worth to mention that our system naturally degrades to a VIO when there is no satellite available.

The positioning error with 5 different settings is illustrated in Fig. 14. Obviously our system performs best under the normal setting where all satellites are used for estimation. In the up direction, the errors of all other 4 configurations accumulate in a similar manner. This indicts that the drift in the up direction can no longer be eliminated with 3 satellites or less. In terms of horizontal directions, no accumulated error but only a small bias occurs for three-satellite setting, which means our system is still able to suppress drifts in east, north and yaw directions. If the number of satellites is further reduced to 2, the horizontal positioning error starts growing with the traveled distance, and we observed small periodic fluctuations in north direction which coincides with that of VIO. This implies that the drift in horizontal plane occurs and yaw error also emerges although the magnitude is very small. Finally with the one-satellite configuration, accumulated errors occur on all four unobservable directions of VIO. However, the error of the yaw component is still smaller compared to that of VIO, which can be inferred from the amplitude of the sine-wave-like error curve. The final trajectories with different satellite settings is shown in Fig. 15. Through this experiment, we claim that our system degrades to different extents when

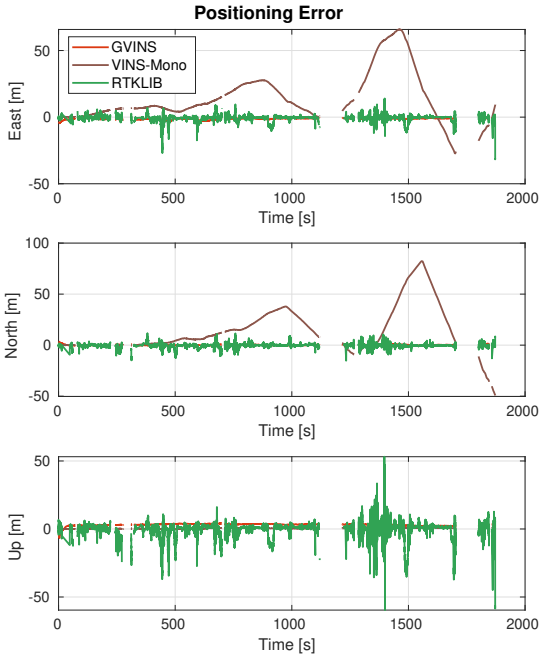

Fig. 16. Positioning error of GVINS, VINS-Mono and RTKLIB in the complex indoor-outdoor experiment. We only compare with the RTK fix solutions, so the gaps in the figure corresponding to the situations where ground truth is not available. The result of VINS-Fusion is not shown because of huge errors and oscillations.

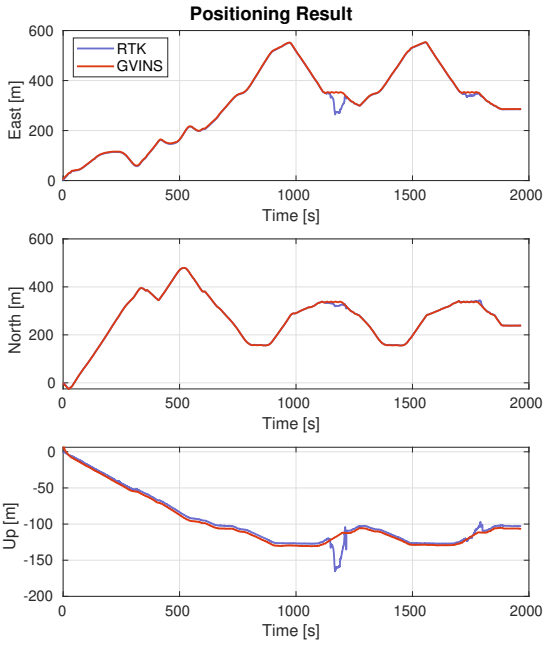

Fig. 17. Positioning result of RTK and GVINS in the complex indoor-outdoor experiment.

the number of locked satellites varies from 3 to 0. However, the proposed system outperforms a pure VIO in all different settings which indicts that our tightly fusion approach can still gain information from limited satellites.

4) Indoor-outdoor Experiment: This experiment, through which we aim to test the robustness of our system, is performed under a complex indoor-outdoor environment. The path of this experiment goes through many challenging scenarios which may bring a single-sensor-based system to failure. For example, no features are detected and tracked in dim or bright area, and the GNSS signal is highly corrupted or totally unavailable in cluttered or indoor environment. In addition, the path is similar to the one in a typical exploration task where no large loops exist, thus drifting is inevitable for any visual-inertial SLAM system. The overall distance of the resulting trajectory is over 3 kilometers and the attitude change is around 130 meters.

Fig. 16 shows the ENU positioning error on the indoor-outdoor sequence. During this experiment the RTK ground truth is no longer always available because of the GNSS-unfriendly environment. Thus we only compare with segments where RTK is in fix status. The gaps around 300s occurs when we were under a bridge and passing through the woods which blocked most of the sky, and the blanks around 1200s and 1800s correspond to the situation where we were going up the indoor stairs. The result of VINS-Fusion is not shown in the figure because of huge errors and oscillations. It can be observed from the figure that VINS-Mono still experiences large accumulated errors on horizontal and yaw directions, while the error in the up direction is smaller than the previous experiment because of the attitude excitation on this sequence. The result of RTKLIB, although does not drift, varies a lot around the ground truth value. Those oscillations indict the condition of GNSS signal and severely affect the performance of VINS-Fusion. Our proposed system outperforms other three approaches in terms of positioning error and overcomes the harsh condition brought by the noisy GNSS measurement. The result of our system still has a bias on the up direction because of imperfect GNSS modelling and various error sources, while the up error of VINS-Mono starts from zero because of pre-alignment. The final trajectories of RTK, aligned VINS-Mono and our system is shown in Fig. 18. The figure shows that both VINS-Mono and our proposed system work well across the whole sequence, although obvious drift occurs on the result of VINS-Mono. The discontinuities on the trajectory of RTK is the result of cluttered and indoor environment. The trajectory of our system follows the RTK result well, and the positioning result, even in GNSS-unfriendly area, can be effectively recovered. Although the duration where RTK fails is short in the whole sequence, the impact can be significant. As shown in Fig. 17, the result of RTK is smooth and aligns well with that of GVINS when GNSS is reliable. However, the solution reported by RTK results in an error of up to 80 meters during GNSS outage, and such behavior is catastrophic for any location-based tasks. The final RMSE of all four approaches is shown in table. I.

5) GNSS Factor Experiment: Based on the previous indoor-outdoor sequence, we further investigate the role of each GNSS measurement(i.e. pseudorange, Doppler shift) on the

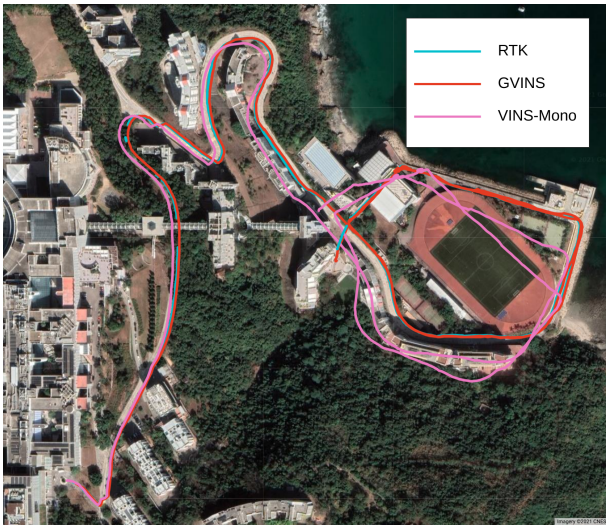

Fig. 18. Final Trajectories in the complex indoor-outdoor experiment. The result of RTKLIB and VINS-Fusion are not plotted because of large noise and jitters. The discontinuities of the RTK path is the result of bad GNSS signal and fix-lost events.

performance of our proposed system. By removing the corresponding graph factor after initialization phase, we obtain the positioning error on pseudorange-only and Doppler-only configurations as depicted in Fig. 19. In the situation where we only employ Doppler shift measurement, an obvious drift occurs as the system no longer has global position constraints. In addition, the initialization error, which is inevitable because we initialize from only a short window of measurements, cannot be eliminated and acts like a bias subsequently. If we instead conduct the pseudorange-only optimization, the system behaves like a normal GVINS, e.g. the system does not drift any more and the initialization error can be eliminated after a short period. However, as the pseudorange measurement tends to be noisy and receiver clock biases are no longer constrained by Doppler shift, the smoothness of the estimation result is affected by the unstable signal, as shown in the magnified portion of Fig. 19. Through this experiment, we show that the pseudorange measurement is the key to eliminating the accumulated drift of VIO. However, with the constraint of Doppler shift, the estimation result tends to be smoother under unstable GNSS conditions.

X. Conclusion

In this paper, we propose a tightly-coupled system to fuse measurements from camera, IMU and GNSS receiver under a non-linear optimization-based framework. Our system starts with an initialization phase, during which a coarse-to-fine procedure is employed to online calibrate the transformation between the local and global frames. In the optimization phase, GNSS raw measurements are modelled and formulated under the probabilistic factor graph framework. The degenerate cases are considered and carefully handled to keep the system robust in the complex environment. In addition, engineering challenges during the integration of the system is discussed to facilitate other GNSS fusion researches. We conduct experiments on both simulation and real-world environment to evaluate the performance of our system, and the results show

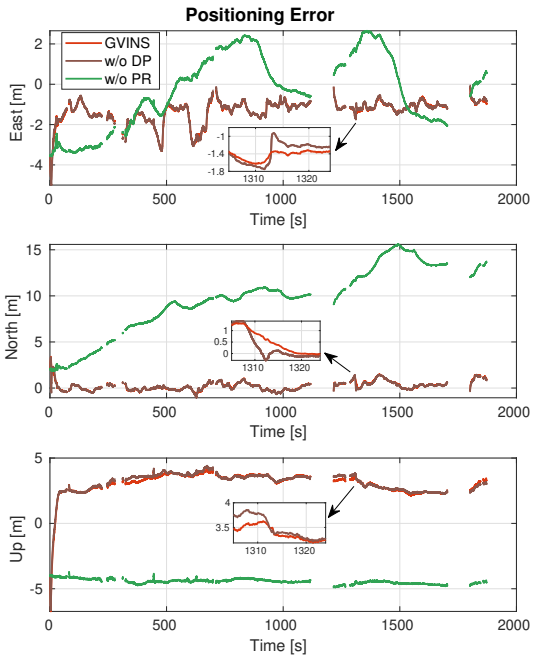

Fig. 19. Positioning error of normal GVINS, GVINS w/o Doppler factor and GVINS w/o pseudorange factor.

that our system effectively eliminates the accumulated drift and preserves the local accuracy of a typical VIO system. To this end, we state that our system can achieve both local smoothness and global consistency.

In future work, the theoretical observability analysis will be conducted under various degenerate situations and we aim to build an online observability-aware state estimator to deal with complex environments and possible sensor failures. In addition, we are also interested in reducing the absolute positioning error by GNSS measurements combination [36] or Precise Point Positining (PPP) [37] techniques to handle distributed localization tasks in swarm system.

References

[1] G. Huang, M. Kaess, and J. J. Leonard, "Towards consistent visual-inertial navigation," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 4926–4933.

[2] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. of the IEEE Int. Conf. on Robot. and Autom.*, Roma, Italy, Apr. 2007, pp. 3565–3572.

[3] M. Li and A. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *Int. J. Robot. Research*, vol. 32, no. 6, pp. 690–711, May 2013.

[4] K. Wu, A. 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*, 2015.

[5] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Research*, vol. 34, no. 3, pp. 314–334, Mar. 2014.

[6] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[7] 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*, 2012, pp. 735–742.

[8] 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 *Proc. of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst.* IEEE, 2013, pp. 3923–3929.

[9] 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 *Proc. of the IEEE Int. Conf. on Robot. and Autom.*, Hong Kong, China, May 2014, pp. 4974–4981.

[10] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart, and M. Chli, "Gomsf: Graph-optimization based multi-sensor fusion for robust uav pose estimation," in *International Conference on Robotics and Automation (ICRA 2018)*. IEEE, 2018.

[11] 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)*, 2019, pp. 7750–7755.

[12] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," 2019.

[13] 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*, 2014, pp. 1309–1328.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] J. Liu, W. Gao, and Z. Hu, "Optimization-based visual-inertial SLAM tightly coupled with raw GNSS measurements," *CoRR*, vol. abs/2010.11675, 2020. [Online]. Available: https://arxiv.org/abs/2010.11675

[19] N. Ashby, "The sagnac effect in the global positioning system," in *Relativity in Rotating Frames*. Springer, 2004, pp. 11–28.

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

[21] J. A. Klobuchar, "Ionospheric time-delay algorithm for single-frequency gps users," *IEEE Transactions on aerospace and electronic systems*, no. 3, pp. 325–331, 1987.

[22] E. Kaplan and C. Hegarty, *Understanding GPS: principles and applications*. Artech house, 2005.

[23] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, 2017.

[24] S. Shen, N. Michael, and V. Kumar, "Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft mavs," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 5303–5310.

[25] F. R. Kschischang, B. J. Frey, and H. . Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[26] Jianbo Shi and Tomasi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[27] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the Intl. Joint Conf. on Artificial Intelligence*, Vancouver, Canada, Aug. 1981, pp. 24–28.

[28] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1793–1800.

[29] T. Qin and S. Shen, "Robust initialization of monocular visual-inertial estimation on aerial robots." in *Proc. of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Vancouver, Canada, 2017.

[30] T. Qin and S. Shen, "Online temporal calibration for monocular visual-inertial systems," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 3662–3669.

[31] U-blox, *Integration manual of u-blox F9 high precision GNSS module*, U-blox.

[32] F. Guo and X. Zhang, "Real-time clock jump compensation for precise point positioning," *GPS solutions*, vol. 18, no. 1, pp. 41–50, 2014.

[33] 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.

[34] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proc. of the IEEE Int. Conf. on Pattern Recognition*, 2012, pp. 3354–3361.

[35] J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. T. Furgale, and R. Siegwart, "A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 431–437.

[36] G. Blewitt, "An automatic editing algorithm for gps data," *Geophysical research letters*, vol. 17, no. 3, pp. 199–202, 1990.

[37] J. Zumberge, M. Heflin, D. Jefferson, M. Watkins, and F. Webb, "Precise point positioning for the efficient and robust analysis of gps data from large networks," *Journal of geophysical research: solid earth*, vol. 102, no. B3, pp. 5005–5017, 1997.