Intermittent GPS-aided VIO: Online Initialization and Calibration

Woosik Lee, Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang

***Abstract*— In this paper, we present an efficient and robust GPS-aided visual inertial odometry (GPS-VIO) system that fuses IMU-camera data with intermittent GPS measurements. To perform sensor fusion, spatiotemporal sensor calibration and initialization of the transform between the sensor reference frames are required. We propose an online calibration method for both the GPS-IMU extrinsics and time offset as well as a reference frame initialization procedure that is robust to GPS sensor noise. In addition, we prove the existence of four unobservable directions of the GPS-VIO system when estimating in the VIO reference frame, and advocate a state transformation to the GPS reference frame for full observability. We extensively evaluate the proposed approach in Monte-Carlo simulations where we investigate the system's robustness to different levels of GPS noise and loss of GPS signal, and additionally study the hyper-parameters used in the initialization procedure. Finally, the proposed system is validated in a large-scale real-world experiment.**

I. Introduction and Related Work

For any autonomous robotic system, robust and accurate localization is a primary requirement. Localization is typically performed by estimating the robot's state using measurements from on-board sensors. Of many possible sensor deployments, cameras and inertial measurement units (IMUs) – which measure linear accelerations and angular velocities of the moving robot – are commonly used for 3D navigation [1] in both indoor and outdoor environments, as they are low-cost yet provide high-quality ego-motion estimation [2]–[5]. However, when only using these sensors it is difficult to provide long-term, drift-free estimation due to the accumulation of relative motion errors. A commonly used approach to bound navigation error is a simultaneous localization and mapping (SLAM) that exploits loop-closure constraints to correct the accumulated error [6], [7]. However, such methods have a major drawback of both increased computational complexity and memory requirements.

As compared to SLAM, global measurement sensors, such as those from Global Positioning System (GPS), directly provide absolute position information to reduce drift. However, the accuracy of GPS measurements is highly dependent on the surrounding environment and the availability of external correction data. Synchronous sensors have been particularly considered in prior works, of which many fused inertial and GPS readings [8]–[11], with others leveraging camera, inertial and GPS sensors fusion [12]–[17] with great success. The asynchronous inclusion of GPS measurements within a sensor fusion framework remains challenging due to their low rate, high noise, and intermittency.

This work was partially supported by the University of Delaware (UD) College of Engineering, the NSF (IIS-1924897), the ARL (W911NF-19-2-0226, JWS 10-051-003), and Google ARCore. Geneva was partially supported by the Delaware Space Grant College and Fellowship Program (NASA Grant NNX15AI19H).

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

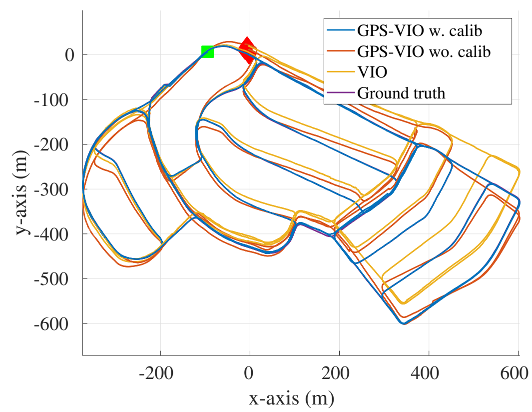

Fig. 1: Simulation results of GPS-VIO with calibration (blue), GPS-VIO without calibration (red), and VIO (yellow). The green and red squares correspond to the start and end of the 9.1km trajectory, respectively.

In order to optimally fuse multiple sensor measurements from different sensor frames, the transformation between the sensor frames and the time offset between the sensors must be known. An initial imperfect guess of the calibration between the sensor frames is often known beforehand, but if it is treated as perfect the state estimation can suffer, thus its refinement during online estimation is highly desirable. For a camera and IMU pair, the calibration of spatial and/or temporal parameters is well studied in [18]–[21]. While offline calibration of camera, IMU and GPS is often performed within an optimization framework [22], [23], online estimation of the transformation among the sensors was also investigated within a Kalman filter (KF) framework [13], [24]–[26]. However, to the best of our knowledge, no work to date has considered the estimation of the time offset between the GPS and IMU/camera, and their inherent asynchronous nature can greatly impact the estimation performance if ignored.

GPS provides latitude, longitude, and altitude readings in a geodetic coordinate frame, which is commonly converted to Cartesian coordinate East-North-Up (ENU) $\{E\}$, e.g., by setting the first GPS measurement as the datum. Conversely, a visual inertial odometry (VIO) system estimates its state relative to a starting VIO frame $\{V\}$, and is known to have four unobservable directions corresponding to the global position and yaw [27], [28]. In order to fuse GPS measurements in the global frame $\{E\}$ with VIO state estimates in the $\{V\}$ frame, the transformation between them must be computed, which is the "reference frame initialization" problem. Unlike sensor calibration, an initialization procedure is required to find this unknown transformation that varies from run-to-run.

This initialization problem can be formulated as a general 3D position trajectory alignment problem. For example, Horn [29] used singular value decomposition (SVD) of a covariance matrix to derive a closed-form solution. Shepard et al. [30] leveraged this method to compute a 7 degree-of-freedom (d.o.f) transformation between synchronized GPS and VIO trajectories. Umeyama [31] presented a method in the presence of large trajectory noises, which was used to find the transformation between two gravity aligned trajectories [32], [33]. Other works have employed additional information for initialization, including magnetic sensors [34]–[36], yaw calculation with a straight planar motion assumption [37], or a prior map constructed in the global frame [16].

Note that the closest to this work is VINS-Fusion [12], which is a loosely-coupled estimator that fuses GPS measurements and VIO's relative poses in a secondary optimization thread. While VINS-Fusion shows impressive performance in practice, the system (i) assumes synchronized measurements with perfectly known timestamps and an identity transformation between GPS and IMU, (ii) lacks support for online refinement of sensor calibration, and (iii) does not explicitly initialize the ENU to VIO frame transform while assuming that the estimates will converge in the ENU frame as more GPS measurements are collected.

In this paper, we develop a tightly-coupled VIO system aided by intermittent GPS measurements to provide persistent global localization results, while focusing on spatiotemporal sensor calibration and state initialization. In particular, the key contributions of this work are the following:

- We propose a tightly-coupled multi-state constraint Kalman filter (MSCKF)-based [38] estimator to optimally fuse inertial, camera, and asynchronous GPS measurements. The system can begin with VIO only (e.g., indoors) and convert the frame of reference to the ENU frame at an arbitrarily later timestep when GPS measurements become available for fusion. This ensures that the system provides seamless localization, and once global information is available the system is able to estimate in this frame of reference.
- To the best of our knowledge, this is the first work that models GPS-IMU time offset and performs online calibration of both the extrinsics and time offset. We also introduce a reference frame initialization procedure that is robust to high GPS noise which leverages the solution to a quadratic constraint least-squares problem [39]. We numerically analyze the choice of this procedure's hyper-parameters under different GPS measurement noise levels.
- We perform an observability analysis of the GPS-VIO system to show that there are four unobservable directions if the 4 d.o.f transformation between the ENU to VIO frames is kept in the state vector, while the system is fully observable if estimating in the ENU frame.
- We evaluate the proposed GPS-VIO extensively in simulations, showing the calibration convergence under different measurement noise levels and the robustness to loss of GPS. Moreover, the proposed method is also validated on a real-world, large-scale experiment with both indoor and outdoor portions which exhibits varying GPS noise levels.

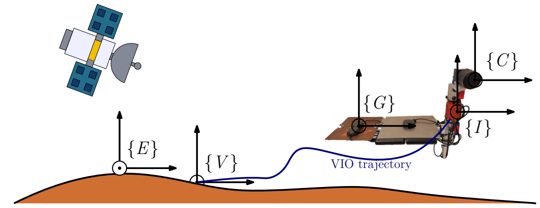

Fig. 2: Our integrated sensor system is composed of five different frames: ENU frame $\{E\}$, VIO frame $\{V\}$, IMU sensor frame $\{I\}$, camera sensor frame $\{C\}$, and GPS sensor frame $\{G\}$. $\{E\}$ is the frame of the reference of the GPS measurements and $\{V\}$ is the local frame set up by VIO whose orientation is aligned with gravity.

II. Preliminaries: MSCKF based VIO

The standard VIO state $\mathbf{x}_k$ at timestep k consists of the current inertial state $\mathbf{x}_{I_k}$ and n historical IMU pose clones $\mathbf{x}_{C_k}$ [38]. All states are represented in the arbitrarily chosen gravity aligned frame of reference, $\{V\}$, see Fig. 2:

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

$$ {}^V\mathbf{x}_{I_k} = \begin{bmatrix} {}^{I_k}_{V}\bar{q}^\top & {}^V\mathbf{p}_{I_k}^\top & {}^V\mathbf{v}_{I_k}^\top & \mathbf{b}_{\omega_k}^\top & \mathbf{b}_{a_k}^\top \end{bmatrix}^\top \tag{2} $$

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

where ${}^{I_k}_{V}\bar{q}$ is the JPL unit quaternion [40] corresponding to the rotation from $\{V\}$ to $\{I\}$ (i.e., rotation matrix ${}^{I_k}_{V}\mathbf{R}$), ${}^V\mathbf{p}_{I_k}$ and ${}^V\mathbf{v}_{I_k}$ are the position and velocity of $\{I\}$ in $\{V\}$, and $\mathbf{b}_{\omega_k}$ and $\mathbf{b}_{a_k}$ are the gyroscope and accelerometer biases, respectively. 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 a manifold element and its correction vector to an updated element on the same manifold [41].

A. State Propagation

The linear acceleration $\mathbf{a}_m$ and angular velocity $\boldsymbol{\omega}_m$ measurements of the IMU are used for propagation:

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

where $\mathbf{a}$ and $\boldsymbol{\omega}$ are true acceleration and angular velocity, $\mathbf{g} \approx [0 \;\, 0 \;\, 9.81]^\top$ is the global gravity, and $\mathbf{n}_a$ and $\mathbf{n}_\omega$ are zero mean Gaussian noises. These measurements are used to propagate the IMU state from timestep k to $k+1$ based on the following generic nonlinear kinematic model [40]:

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

where $\hat{\mathbf{x}}_{a|b}$ denotes the estimate at timestep a processing the measurements up to timestep b. We linearize (5) at the current estimate and propagate the covariance forward in time:

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

where $\boldsymbol{\Phi}$ and $\mathbf{Q}$ are the state transition matrix and discrete noise covariance [38].

B. Visual Measurement Update

We maintain a number of stochastic clones in ${}^V\mathbf{x}_{C_k}$, and perform visual feature tracking to obtain series of visual bearing measurements to 3D environmental features. A measurement $\mathbf{z}_i$ at timestep i is expressed as a function of a cloned pose and feature position ${}^V\mathbf{p}_f$:

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

$$ {}^{C_i}\mathbf{p}_f = {}^{C}_{I}\mathbf{R}\,{}^{I_i}_{V}\mathbf{R}\left({}^V\mathbf{p}_f - {}^V\mathbf{p}_{I_i}\right) + {}^C\mathbf{p}_I \tag{8} $$

where $\mathbf{\Pi}\left([x\ y\ z]^\top\right) = \left[\frac{x}{z} \quad \frac{y}{z}\right]^\top$ is the perspective projection, and ${}^C_I\mathbf{R}$ and ${}^C\mathbf{p}_I$ represent the camera to IMU extrinsics. By stacking all measurements for a given feature, the corresponding linearized residuals $\tilde{\mathbf{z}}_c$ is given by:

$$\tilde{\mathbf{z}}_{c_k} = \mathbf{H}_{x_k}{}^V\tilde{\mathbf{x}}_k + \mathbf{H}_{f_k}{}^V\tilde{\mathbf{p}}_{f_k} + \mathbf{n}_{f_k} \tag{9}$$

where $\mathbf{H}_x$ and $\mathbf{H}_f$ are the measurement Jacobians of the state and the feature. The key idea of the MSCKF is to find the left null space of $\mathbf{H}_f$ and left multiply (9) by it to infer a new measurement function that depends only on the state:

$$\tilde{\mathbf{z}}'_{c_k} = \mathbf{H}'_{x_k}{}^V\tilde{\mathbf{x}}_k + \mathbf{n}'_{f_k} \tag{10}$$

which can be directly used in an EKF update without storing features in the state, leading to substantial computational savings as the problem size remains bounded over time.

III. GPS Measurement Update and Calibration

Besides the visual measurement update as in the standard MSCKF, whenever a new GPS measurement in the ENU frame $\{E\}$ is available, we will update the state with it. This requires knowledge of the transform between the two frames $\{{}^E_V\mathbf{R}, {}^E\mathbf{p}_V\}$, which we will explain in the next section. In particular, the GPS measurement ${}^E\mathbf{p}_{G_k}$ at timestep k is:

$$\mathbf{z}_{g_k} := {}^E\mathbf{p}_{G_k} = {}^E\mathbf{p}_V + {}^E_V\mathbf{R}\,{}^V\mathbf{p}_{G_k} + \mathbf{n}_{g_k} =: \mathbf{h}({}^V\mathbf{x}_k) + \mathbf{n}_{g_k} \tag{11}$$

$${}^V\mathbf{p}_{G_k} = {}^V\mathbf{p}_{I_k} + {}^{I_k}_V\mathbf{R}^\top\,{}^I\mathbf{p}_G \tag{12}$$

where ${}^I\mathbf{p}_G$ is the GPS to IMU extrinsic calibration and $\mathbf{n}_{g_k}$ is a white Gaussian noise. We note that while here ${}^E_V\mathbf{R}$ is written as a full rotation matrix, we represent it as one that only rotates about the global gravity aligned z-axis. Due to the delayed asynchronous nature of the GPS sensor, the state has likely advanced beyond the collection time and thus we express the measurement as a function of the available stochastic clones. Using linear interpolation [42], the IMU pose in Eq. (12) can be expressed as:

$${}^{I_k}_V\mathbf{R} = \mathrm{Exp}\left(\lambda\ \mathrm{Log}\left({}^{I_b}_V\mathbf{R}\,{}^{I_a}_V\mathbf{R}^\top\right)\right){}^{I_a}_V\mathbf{R} \tag{13}$$

$${}^V\mathbf{p}_{I_k} = (1 - \lambda){}^V\mathbf{p}_{I_a} + \lambda{}^V\mathbf{p}_{I_b} \tag{14}$$

$$\lambda = (t_k + {}^It_G - t_a)/(t_b - t_a) \tag{15}$$

where ${}^It_G$ is the time offset between the GPS and IMU clocks, the bounding poses have timestamps $t_a \leq (t_k + {}^It_G) \leq t_b$, and $\mathrm{Exp}(\cdot)$, $\mathrm{Log}(\cdot)$ are the SO(3) matrix exponential and logarithmic functions [43].

As evident from Eqs. (11)-(15), the GPS measurement model depends on both the IMU states and the GPS-IMU extrinsic and time offset, thus enabling online spatiotemporal GPS-IMU calibration. To update with this measurement in the MSCKF, we linearize it at the current estimate and have the following measurement Jacobians:

$$\frac{\partial\tilde{\mathbf{z}}_{gk}}{\partial{}^{I_a}_V\tilde{\boldsymbol{\theta}}} = -{}^E_V\hat{\mathbf{R}}{}^{I_k}_V\hat{\mathbf{R}}^\top\lfloor{}^I\hat{\mathbf{p}}_G\times\rfloor\mathbf{J}_\mathbf{l}(\hat{\lambda}{}^b_a\hat{\boldsymbol{\theta}})(\mathbf{J}_\mathbf{r}(\hat{\lambda}{}^b_a\hat{\boldsymbol{\theta}})^{-1} - \hat{\lambda}\mathbf{J}_\mathbf{r}({}^b_a\hat{\boldsymbol{\theta}})^{-1}) \tag{16}$$

$$\frac{\partial\tilde{\mathbf{z}}_{gk}}{\partial{}^{I_b}_V\tilde{\boldsymbol{\theta}}} = -\hat{\lambda}{}^E_V\hat{\mathbf{R}}{}^{I_k}_V\hat{\mathbf{R}}^\top\lfloor{}^I\hat{\mathbf{p}}_G\times\rfloor\mathbf{J}_\mathbf{l}(\hat{\lambda}{}^b_a\hat{\boldsymbol{\theta}})\mathbf{J}_\mathbf{l}({}^b_a\hat{\boldsymbol{\theta}})^{-1} \tag{17}$$

$$\frac{\partial\tilde{\mathbf{z}}_{gk}}{\partial{}^V\tilde{\mathbf{p}}_{I_a}} = (1 - \hat{\lambda}){}^E_V\hat{\mathbf{R}}, \quad \frac{\partial\tilde{\mathbf{z}}_{gk}}{\partial{}^V\tilde{\mathbf{p}}_{I_b}} = \hat{\lambda}{}^E_V\hat{\mathbf{R}} \tag{18}$$

$$\frac{\partial\tilde{\mathbf{z}}_{gk}}{\partial{}^I\tilde{\mathbf{p}}_G} = {}^E_V\hat{\mathbf{R}}{}^{I_k}_V\hat{\mathbf{R}}^\top \tag{19}$$

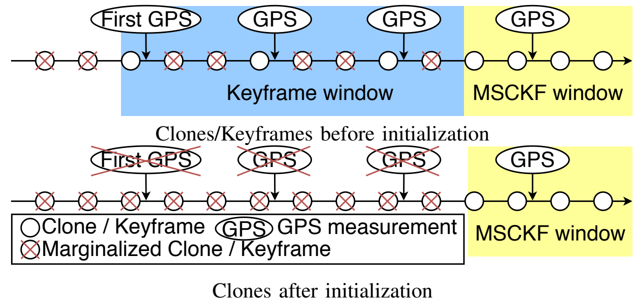

Fig. 3: The variation of windows during and after GPS-VIO initialization. GPS-VIO inserts keyframes after the first GPS measurement is received, and marginalize the keyframes after the initialization, leaving only the standard camera clones.

$$\frac{\partial\tilde{\mathbf{z}}_{gk}}{\partial{}^I\tilde{t}_G} = \frac{1}{t_b - t_a}{}^E_V\hat{\mathbf{R}}({}^V\hat{\mathbf{p}}_{I_b} - {}^V\hat{\mathbf{p}}_{I_a} - {}^{I_k}_V\hat{\mathbf{R}}^\top\lfloor{}^I\hat{\mathbf{p}}_G\times\rfloor\mathbf{J}_\mathbf{l}(\hat{\lambda}{}^b_a\hat{\boldsymbol{\theta}}){}^b_a\hat{\boldsymbol{\theta}}) \tag{20}$$

where $\lfloor\cdot\times\rfloor$ is the skew symmetric matrix, $\mathbf{J}_\mathbf{l}$ and $\mathbf{J}_\mathbf{r}$ are left and right Jacobians of SO(3) [43], and ${}^b_a\hat{\boldsymbol{\theta}} = \mathrm{Log}({}^{I_b}_V\hat{\mathbf{R}}{}^{I_a}_V\hat{\mathbf{R}}^\top)$. With these, we are ready to perform EKF update. The details can be found in the companion technical report [44].

IV. GPS-VIO Initialization

As mentioned earlier, when performing an EKF update with GPS measurements Eq. (11), the 4 d.o.f frame transformation $\{{}^E_V\mathbf{R}, {}^E\mathbf{p}_V\}$ must be known. To find this, we can collect two sets of position estimates of the GPS receiver in two different frames and formulate a non-linear optimization problem to align them. This process requires us to have estimates of the GPS receiver positions in both $\{E\}$ and $\{V\}$ frames. In the case of inaccurate GPS measurements, alignment using a short trajectory length may result in a poor transformation estimate due to the true trajectory being buried in the large measurement noise. As shown in simulations in Section VI, the smart use of longer trajectories allows for accurate alignment even with high noise.

In the standard MSCKF-VIO, the current sliding window typically contains a very short and most recent portion of the trajectory, which does not support reliable GPS-VIO initialization. Therefore, we augment our state by selectively keeping the clone poses (i.e., keyframes) that bound GPS measurements at a fixed temporal frequency. As illustrated in Fig. 3, once we reach the desired trajectory length, we perform interpolation for all GPS measurement times that fall within the keyframe window to find the corresponding position estimates in the VIO frame.

Given a set of GPS position measurements in the ENU frame $\{{}^E\mathbf{p}_{G_1}, \cdots, {}^E\mathbf{p}_{G_n}\}$ within the keyframe window and the corresponding interpolated positions in the VIO frame $\{{}^V\mathbf{p}_{G_1}, \cdots {}^V\mathbf{p}_{G_n}\}$, we use the following geometric constraints to derive the frame initialization:

$${}^E\mathbf{p}_{G_i} = {}^E\mathbf{p}_V + {}^E_V\mathbf{R}\,{}^V\mathbf{p}_{G_i}, \forall i = 1\cdots n \Rightarrow \tag{21}$$

$${}^E\mathbf{p}_{G_j} - {}^E\mathbf{p}_{G_1} = {}^E_V\mathbf{R}({}^V\mathbf{p}_{G_j} - {}^V\mathbf{p}_{G_1}), \forall j = 2\cdots n \tag{22}$$

As mentioned earlier, there 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 ENU and VIO frames due to the fact that both frames are gravity aligned, which entails that we can simply use the rotation about the global z-axis with yaw angle θ:

$$ {}^{E}_{V}\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{23} $$

With (23) we can re-write (22) as the following linear constraint:

$$ \mathbf{A}_j \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} := \mathbf{A}_j \mathbf{w} = \mathbf{b}_j, \forall j = 2 \cdots n \tag{24} $$

Stacking all these constraints yields the following linear least-squares with quadratic constraint, which can be solved for $\mathbf{w}$, e.g., by Lagrangian multipliers [39]:

$$ \min_{\mathbf{w}} \ ||\mathbf{A}\mathbf{w} - \mathbf{b}||^2, \quad \text{s.t. } ||\mathbf{w}||^2 = 1 \tag{25} $$

The solution of (25) immediately provides the sought rotation ${}^{E}_{V}\hat{\mathbf{R}}$. We substitute it into (21) and solve for ${}^{E}\mathbf{p}_V$ as:

$$ {}^{E}\hat{\mathbf{p}}_V = \frac{1}{n}\sum_{i=1}^{n}\left[{}^{E}\mathbf{p}_{G_i} - {}^{E}_{V}\hat{\mathbf{R}}\,{}^{V}\mathbf{p}_{G_i}\right] \tag{26} $$

The resulting $\{{}^{E}_{V}\hat{\mathbf{R}}, {}^{E}\hat{\mathbf{p}}_V\}$ initial guess of the GPS-VIO frame transformation is further corrected using delayed initialization [42], [45], which appends the transform to the state in a probabilistic fashion. Specifically, by augmenting the state vector with the transformation along with an *infinite* covariance prior for these new variables, we perform the standard EKF update using all collected GPS measurements. After initialization, we marginalize all the keyframes to reduce the state to the original state size (see Fig. 3).

V. Observability Analysis of GPS-VIO

As system observability plays an important role state estimation [27], [46], in this section we perform an observability analysis for the proposed GPS-aided VIO system to gain insights about state/parameter identifiability. For concise presentation, we consider a simplified case where the state does not contain biases or stochastic clones and assumes a single feature with perfectly synchronized and calibrated sensors, while the results can be extended to general cases:

$$ \mathbf{x}_k = \begin{bmatrix} {}^{I_k}_{V}\bar{q}^\top & {}^{V}\mathbf{p}_{I_k}^\top & {}^{V}\mathbf{v}_{I_k}^\top & {}^{V}\mathbf{p}_f^\top & {}^{E}_{V}\bar{q}^\top & {}^{E}\mathbf{p}_V^\top \end{bmatrix}^\top \tag{27} $$

The linearized error state evolution and residuals of both the GPS and visual measurement are generically given by (see (5), (7) and (11)):

$$ \tilde{\mathbf{x}}_k = \mathbf{\Phi}(t_k, t_0)\tilde{\mathbf{x}}_0 + \mathbf{w}_k \tag{28} $$

$$ \tilde{\mathbf{z}}_k = \mathbf{H}_k\tilde{\mathbf{x}}_k + \mathbf{n}_k \tag{29} $$

Given this linearized system, we have the following result:

Lemma 5.1: If estimating states in the VIO frame, even with global GPS measurements, the GPS-VIO system remains unobservable and has four unobservable directions.

Proof: We first compute the state transition matrix (6):

$$ \mathbf{\Phi}(t_k, t_0) = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times7} \\ \mathbf{\Phi}_2 & \mathbf{I}_3 & \Delta t\mathbf{I}_3 & \mathbf{0}_{3\times7} \\ \mathbf{\Phi}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3\times7} \\ \mathbf{0}_{7\times3} & \mathbf{0}_{7\times3} & \mathbf{0}_{7\times3} & \mathbf{I}_7 \end{bmatrix} \tag{30} $$

$$ \text{where } \mathbf{\Phi}_1 = {}^{I_k}_{V}\hat{\mathbf{R}}\,{}^{I_0}_{V}\hat{\mathbf{R}}^\top \tag{31} $$

$$ \mathbf{\Phi}_2 = \text{-}\lfloor {}^{V}\hat{\mathbf{p}}_{I_k} \text{ - } {}^{V}\hat{\mathbf{p}}_{I_0} \text{ - } {}^{V}\hat{\mathbf{v}}_{I_0}\Delta t + \frac{1}{2}\mathbf{g}\Delta t^2 \times\rfloor\,{}^{I_0}_{V}\hat{\mathbf{R}}^\top \tag{32} $$

$$ \mathbf{\Phi}_3 = \text{-}\lfloor {}^{V}\hat{\mathbf{v}}_{I_k} \text{ - } {}^{V}\hat{\mathbf{v}}_{I_0} + \mathbf{g}\Delta t \times\rfloor\,{}^{I_0}_{V}\hat{\mathbf{R}}^\top \tag{33} $$

$$ \Delta t = t_k \text{ - } t_0 \tag{34} $$

Linearization of (7) and (11) yields the following measurement Jacobians:

$$ \mathbf{H}_k = \begin{bmatrix} \mathbf{H}_\Pi\mathbf{H}_1 & \text{-}\mathbf{H}_\Pi{}^{I_k}_{V}\hat{\mathbf{R}} & \mathbf{0}_{2\times3} & \mathbf{H}_\Pi{}^{I_k}_{V}\hat{\mathbf{R}} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times3} \\ \mathbf{0}_3 & {}^{E}_{V}\hat{\mathbf{R}} & \mathbf{0}_3 & \mathbf{0}_3 & (\lfloor {}^{E}_{V}\hat{\mathbf{R}}\,{}^{V}\hat{\mathbf{p}}_{I_k}\times\rfloor)_z & \mathbf{I}_3 \end{bmatrix} \tag{35} $$

where $\mathbf{H}_1 = \lfloor {}^{I_k}_{V}\hat{\mathbf{R}}({}^{V}\hat{\mathbf{p}}_f \text{ - } {}^{V}\hat{\mathbf{p}}_{I_k})\times\rfloor$, $\mathbf{H}_\Pi = \begin{bmatrix} \frac{1}{z} & 0 & \frac{-x}{z^2} \\ 0 & \frac{1}{z} & \frac{-y}{z^2} \end{bmatrix}$, and $(\cdot)_z$ is the third column of the matrix. Note that the fifth column is 5 by 1, because we have 1 d.o.f for the $\{E\}$ to $\{V\}$ rotation. Now we can construct the observability matrix $\mathbf{M}$ (see [47]) and compute its null space as:

$$ \mathbf{M} = \begin{bmatrix} \vdots \\ \mathbf{H}_k\mathbf{\Phi}(t_k, t_0) \\ \vdots \end{bmatrix}, \ \mathbf{null}(\mathbf{M}) = \begin{bmatrix} \mathbf{0}_3 & \text{-}{}^{I_0}_{V}\hat{\mathbf{R}}\mathbf{g} \\ \mathbf{I}_3 & \lfloor {}^{V}\hat{\mathbf{p}}_{I_0}\times\rfloor\mathbf{g} \\ \mathbf{0}_3 & \lfloor {}^{V}\hat{\mathbf{v}}_{I_0}\times\rfloor\mathbf{g} \\ \mathbf{I}_3 & \lfloor {}^{V}\hat{\mathbf{p}}_f\times\rfloor\mathbf{g} \\ \mathbf{0}_{1\times3} & (\text{-}{}^{E}_{V}\hat{\mathbf{R}}\mathbf{g})_3 \\ \text{-}{}^{E}_{V}\hat{\mathbf{R}} & \mathbf{0}_{3\times1} \end{bmatrix}_{16\times4} \tag{36} $$

where $(\cdot)_3$ is the third element of the vector. The span of the columns of this matrix encodes the unobservable subspace. By inspection, the first block column corresponds to the translation of $\{V\}$ relative to $\{E\}$ and the second block column to the rotation of $\{V\}$ with respect to $\{E\}$ along the axis of gravity. It thus becomes clear that the GPS-VIO system in the VIO frame has these four unobservable directions which are essentially inherited from the standard VIO [27], [28]. ■

While the above results seem to be counter-intuitive given the availability of global GPS measurements, the root cause of this unobservability is the gauge freedom of the 4 d.o.f GPS-VIO frame transformation. Thus even though we utilize global measurements, the system maintains a non-trivial null space. Unobservable directions are known to cause inconsistency issues for linearized estimators as these null spaces falsely disappear due to numerical errors. Therefore the estimator gains information in spurious directions, hurting overall consistency and accuracy, unless special techniques are utilized [27], [48]. To address this issue, we perform state estimation directly in the ENU global frame of reference once initialized, which can be shown to make the system fully observable.

Lemma 5.2: If estimating states in the ENU frame, the GPS-VIO system is fully observable.

Proof: The simplified state in the ENU frame is:

$$ \mathbf{x}_k = \begin{bmatrix} {}^{I_k}_{E}\bar{q}^\top & {}^{E}\mathbf{p}_{I_k}^\top & {}^{E}\mathbf{v}_{I_k}^\top & {}^{E}\mathbf{p}_f^\top & {}^{E}_{V}\bar{q}^\top & {}^{E}\mathbf{p}_V^\top \end{bmatrix}^\top \tag{37} $$

Then the state transition matrix of the new state $\mathbf{\Phi}'(t_k, t_0)$ is equivalent to Eq. (30) with all parameters that are in $\{V\}$ are now in $\{E\}$. Also, the corresponding GPS measurement Jacobian is $\mathbf{H}_g = \begin{bmatrix} \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3\times10} \end{bmatrix}$ (see Eq. (11)). Clearly, the multiplication of $\mathbf{H}_g\mathbf{\Phi}'(t_k, t_0)$ with $\mathbf{null}(\mathbf{M})$ does not yield a zero matrix which means the four unobservable directions of Eq. (27) are now observable given GPS measurements. Since VIO is known to have four unobservable directions [27], [28], we can conclude that the state in the ENU, see Eq. (37), is fully observable. ■

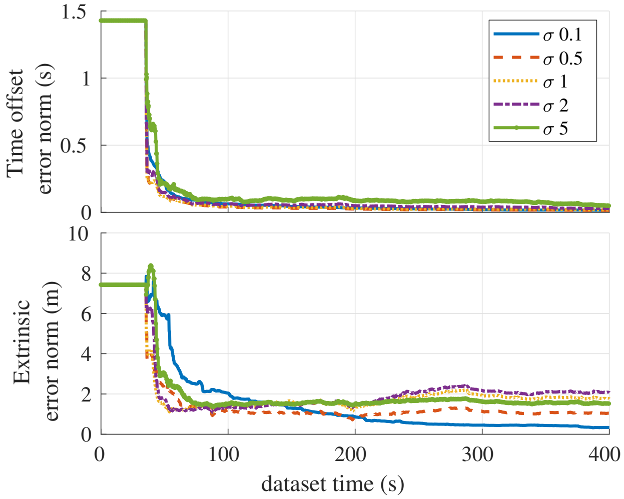

Fig. 4: The calibration errors respect to the size of GPS measurement noise.

TABLE I: Average position and orientation errors over ten runs for different initialization distances and GPS noise values in units of meters/degree.

dist\σ	0.1m	0.5m	1m	2m	5m
5m	1.59 / 0.65	7.08 / 3.20	14.32 / 6.56	29.37 / 69.84	69.17 / 92.37
10m	1.39 / 0.52	5.23 / 2.23	10.22 / 4.39	19.80 / 47.79	45.02 / 91.85
20m	0.90 / 0.29	2.68 / 1.08	5.02 / 2.07	9.78 / 4.08	25.49 / 49.75
50m	0.55 / 0.08	0.77 / 0.16	1.09 / 0.30	1.88 / 0.61	4.58 / 1.49
100m	0.51 / 0.09	0.49 / 0.06	0.55 / 0.12	0.85 / 0.24	2.18 / 0.63

As a final remark about the proposed GPS-VIO estimator, based on the above lemma, after GPS-VIO initialization, we therefore transform the state from $\{V\}$ to the $\{E\}$ and propagate the error state and covariance based on the linearization of this transform function $g(\cdot)$ as follows:

$$ {}^{E}\mathbf{x}_k = g({}^{V}\mathbf{x}_k, {}^{E}_{V}\mathbf{R}, {}^{E}\mathbf{p}_V) \tag{38} $$

$$ \Rightarrow\ {}^{E}\tilde{\mathbf{x}}_k = \mathbf{\Psi}\,{}^{V}\tilde{\mathbf{x}}_k\ ,\quad \mathbf{P}^{\oplus} = \mathbf{\Psi}\mathbf{P}^{\ominus}\mathbf{\Psi}^{\top} \tag{39} $$

where $\mathbf{\Psi}$ is the Jacobian matrix [44]. We note that the $\{E\}$ to $\{V\}$ transformation inserted into the state during initialization, see Section IV, has been marginalized since all measurements can now be written directly in terms of the remaining state variables.

VI. Simulation Results

The proposed GPS-VIO was implemented within *Open-VINS* [49] which provides both simulation and evaluation utilities. The key simulation parameters are: maximum of 15 clones, maximum of 100 actively tracked features with 1 pixel Gaussian noise, while the IMU was simulated using realistic noise from a real sensor. The camera was simulated at 5Hz, while the GPS sensor was simulated with a lower frequency of 2Hz with varying measurement noises ranging from 0.1m to 5m. As shown in Fig. 1, the trajectory of the dataset is 9.1km in length, following that of a planar vehicle motion with an average velocity of 9m/s. Except for the results in Section VI-C, Monte-Carlo simulation results are reported over 10 runs.

A. Initialization with Different Hyper-parameters

To gain insight into how the initialization procedure is affected by GPS measurement noise and trajectory length, we simulated 0.1, 0.5, 1, 2 and 5m GPS measurement noise and 5, 10, 20, 50 and 100m initialization distance thresholds. To prevent biasing these results to the initial section of this particular trajectory, it is split into non-overlapping segments for each distance threshold. The initialization procedure was independently performed on each segment and the resulting statistics on the accuracy of the initialized VIO to ENU transform are shown in Table. I.

In general, the initialization errors are smaller with a larger distance threshold and with smaller GPS noise. The results indicate that reasonable accuracy for this transformation can be achieved after 50m for most realistic levels of GPS measurement noise. In practice, these results can be used to determine the needed distance threshold for different sensor uncertainties.

B. Calibration with Different GPS Noise Levels

In order to study the calibration convergence of the proposed system, we performed extrinsic calibration and time offset between the GPS and IMU with poor initial guesses. The groundtruth and initial guess for the extrinsic were $[2.00\ 3.00\ 1.00]^{\top}$ and $[5.40\ 1.65\ 6.62]^{\top}$ meters, while for the time offset they were 0 and -1.3 seconds. The calibration results for the first 400 seconds are shown in Fig. 4, which clearly demonstrates that the time offset calibration converges to near zero. The final converged extrinsic calibration error follows that of the GPS measurement noise except in the 5m σ case. This shows that the convergence of the extrinsic is highly dependent on the measurement noise and whose final error is on the order of the GPS measurements.

A representative run is shown in Fig. 5, all calibration was able to converge within the first 100 seconds of the dataset while remaining consistent. The static lines at the beginning of the each are from before initialization in the ENU and thus no GPS measurements that are required to update these parameters have been used. As expected, the y-error, which is mostly aligned with gravity in this scenario, shows little decrease in state uncertainty due to this axis corresponding to the normal of the plane of motion [21].

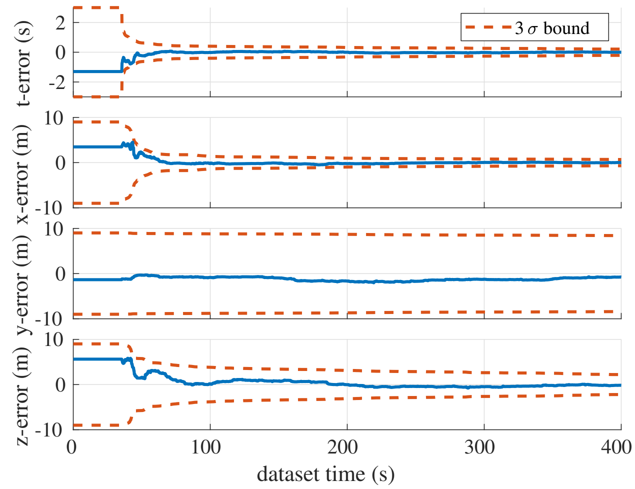

Fig. 5: Calibration errors of the proposed GPS-VIO with 5m GPS sensor noise and the corresponding consistency bounds. The blue lines are the errors and the red dotted lines are the 3 standard deviation bounds of each error

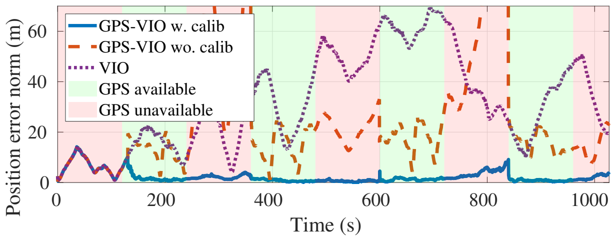

Fig. 6: The position error of each method in the case of intermittent GPS.

C. Robustness to Intermittent GPS Signals

To validate the robustness to intermittent GPS measurements, we simulated a series of GPS dropouts during which the GPS-VIO purely relied on visual and inertial information. The GPS dropouts lasted 120 seconds in length, the ENU to VIO transformation was set to identity to allow for comparison to pure VIO, and the calibration was perturbed as in the last simulation.

Fig. 6 shows the errors of GPS-VIO with online calibration, GPS-VIO without online calibration, and pure VIO. Before initialization in ENU all systems are purely VIO. After the initialization in the ENU at around 150 seconds, the proposed method begins fusing GPS data and thus quickly bounds its errors. It is clear that poor calibration can hurt the system's estimation even in the presence of global measurements. Finally, the relatively good accuracy of the VIO is able to "bridge the gap" between the GPS-available regions, providing high-quality navigation estimates over the entire trajectory. Note that reducing error of the pure VIO at time around 800 seconds is because the trajectory has loops, which brings the estimation and groundtruth close by chance.

VII. Experimental Results

We further evaluate the proposed GPS-VIO in a real-world scenario. The trajectory begins in an indoor parking garage during which the system does not have access to GPS measurements until a minute in when it exits the structure. During the outdoor segment the vehicle travels several kilometers before returning to the same GPS-denied structure. The total length of the data is about 4.9km, and we used a monocular camera-imu pair, alongside two GPS receivers all of which were mounted rearward on the trunk of the collection vehicle. One low-cost GPS sensor was used for GPS measurements while the second provided RTK data for groundtruth. The covariance of each GPS measurement was computed by RTKLIB library [50]. We compared our system against the open sourced VINS-Fusion [12] system. We used 30 clones and a max of 150 features for the proposed GPS-VIO while for VINS-Fusion a max of 200 features and a max solver time 0.04 were used. We note that VINS-Fusion does not take into account the GPS-IMU calibration.

Fig. 7 shows the result of the experiment. The RMSE of each trajectory compared with RTK groundtruth whenever it is available are: 3.57m (GPS-VIO w. calib), 7.03m (GPS-VIO wo. calib), 6.66m (VINS-Fusion) and 15.95m (VIO). We gave the initial hand measured extrinsic value $[0.06\ 0.11\ -0.03]^\top$ meters and time offset of zero for both

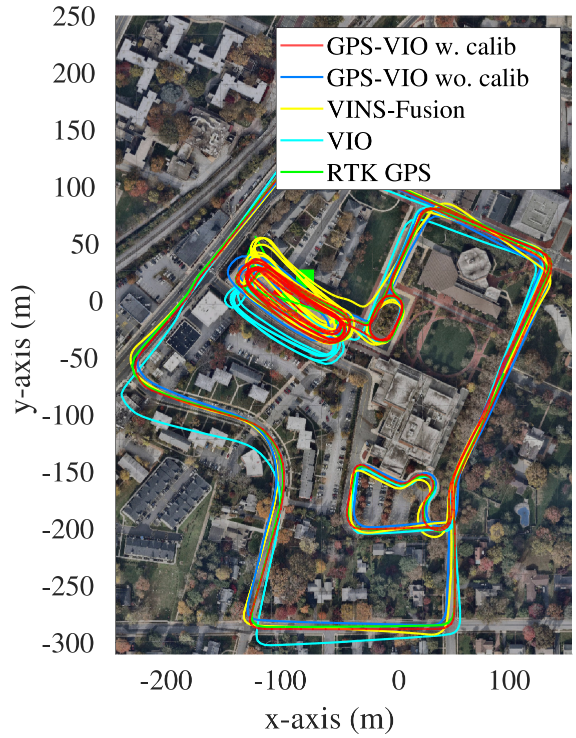

Fig. 7: The red line is GPS-VIO with online calibration, blue is GPS-VIO without online calibration, yellow is VINS-Fusion, light blue is VIO only, and green is RTK GPS. The green and red boxes denote the start end points of the trajectory.

GPS-VIO w. calib and GPS-VIO wo. calib. The extrinsic calibration converged to a value of $[-0.49\ 0.70\ -0.07]^\top$ at the end of the dataset which is within the expected convergence bounds associated with the 1-10 meters observed GPS noise. We found time offset calibration quickly converged to a nontrivial value of -0.85 seconds which given the 7.4m/s average vehicle velocity equates to 6.3m position error if not properly calibrated, and thus validates the need for online estimation of this parameter. As compared to VINS-Fusion our method can achieve higher accuracy while also being a light-weight single threaded estimator which runs in real time. We also note that since VINS-Fusion does not estimate VIO to ENU transform explicitly, its pose output may not suitable for real time applications, and Fig. 7 shows the final optimized trajectory after completion of the dataset.

VIII. Conclusions and Future Work

In this paper, we have developed an efficient and robust GPS-VIO system that fuses GPS, IMU, and camera measurements in a tightly-coupled estimator. In particular, to robustify the system, we have focused on the online GPS-VIO spatiotemporal sensor calibration and frame initialization. The observability analysis shows that if estimating states naively in the VIO frame, the system remains unobservable as the standard VIO; however, this can be mitigated by transforming the system to the global ENU frame after GPS-VIO frame initialization, which is exploited in the proposed GPS-VIO estimator. This system has been validated in both Monte-Carlo simulations and real-world experiments. In the future, we will integrate mapping capability into this GPS-VIO system.

REFERENCES

[1] G. Huang, "Visual-inertial navigation: A concise review," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019, pp. 9572–9582.

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

[3] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.

[4] G. Loianno, C. Brunner, G. McGrath, and V. Kumar, "Estimation, control, and planning for aggressive flight with a small quadrotor with a single camera and imu," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 404–411, 2016.

[5] G. Huang, M. Kaess, and J. Leonard, "Towards consistent visual-inertial navigation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Hong Kong, China, May 31-June 7 2014, pp. 4926–4933.

[6] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[7] P. Geneva, K. Eckenhoff, and G. Huang, "A linear-complexity EKF for visual-inertial navigation with loop closures," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019, pp. 3535–3541.

[8] A. Almagbile, J. Wang, and W. Ding, "Evaluating the performances of adaptive kalman filter methods in gps/ins integration," *Journal of Global Positioning Systems*, vol. 9, no. 1, pp. 33–40, 2010.

[9] C. Eling, L. Klingbeil, and H. Kuhlmann, "Real-time single-frequency gps/mems-imu attitude determination of lightweight uavs," *Sensors*, vol. 15, no. 10, pp. 26 212–26 235, 2015.

[10] R. Sun, K. Han, J. Hu, Y. Wang, M. Hu, and W. Y. Ochieng, "Integrated solution for anomalous driving detection based on beidou/gps/imu measurements," *Transportation research part C: emerging technologies*, vol. 69, pp. 193–207, 2016.

[11] T. Zhang and X. Xu, "A new method of seamless land navigation for gps/ins integrated system," *Measurement*, vol. 45, no. 4, pp. 691–701, 2012.

[12] 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.

[13] Y. Lee, J. Yoon, H. Yang, C. Kim, and D. Lee, "Camera-gps-imu sensor fusion for autonomous flying," in *Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2016, pp. 85–88.

[14] C. V. Angelino, V. R. Baraniello, and L. Cicala, "Uav position and attitude estimation using imu, gnss and camera," in *15th International Conference on Information Fusion*, 2012, pp. 735–742.

[15] T. Zsedrovits, P. Bauer, A. Zarandy, B. Vanek, J. Bokor, and T. Roska, "Error analysis of algorithms for camera rotation calculation in gps/imu/camera fusion for uav sense and avoid systems," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2014, pp. 864–875.

[16] T. Oskiper, S. Samarasekera, and R. Kumar, "Multi-sensor navigation algorithm using monocular camera, imu and gps for large scale augmented reality," in *IEEE international symposium on mixed and augmented reality (ISMAR)*, 2012, pp. 71–80.

[17] 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 *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 4974–4981.

[18] F. M. Mirzaei and S. I. Roumeliotis, "A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation," *IEEE transactions on robotics*, vol. 24, no. 5, pp. 1143–1156, 2008.

[19] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1280–1286.

[20] M. Li and A. I. Mourikis, "Online temporal calibration for camera–imu systems: Theory and algorithms," *The International Journal of Robotics Research*, vol. 33, no. 7, pp. 947–964, 2014.

[21] Y. Yang, P. Geneva, K. Eckenhoff, and G. Huang, "Degenerate motion analysis for aided INS with online spatial and temporal calibration," *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 2, pp. 2070–2077, 2019.

[22] H. Wegmann, "Image orientation by combined (a) at with gps and imu," *International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences*, vol. 34, no. 1, pp. 278–283, 2002.

[23] M. Bryson, M. Johnson-Roberson, and S. Sukkarieh, "Airborne smoothing and mapping using vision and inertial sensors," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 2037–2042.

[24] X. Kong, E. M. Nebot, and H. Durrant-Whyte, "Development of a nonlinear psi-angle model for large misalignment errors and its application in ins alignment and calibration," in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, vol. 2, 1999, pp. 1430–1435.

[25] K. Hausman, J. Preiss, G. S. Sukhatme, and S. Weiss, "Observability-aware trajectory optimization for self-calibration with application to uavs," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1770–1777, 2017.

[26] A. Ramanandan, M. Chari, and A. Joshi, "Systems and methods for using a global positioning system velocity in visual-inertial odometry," Aug. 6 2019, uS Patent 10,371,530.

[27] 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.

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

[29] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Josa a*, vol. 4, no. 4, pp. 629–642, 1987.

[30] 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 *IEEE/ION Position, Location and Navigation Symposium-PLANS*, 2014, pp. 1309–1328.

[31] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 4, pp. 376–380, 1991.

[32] I. Zendjebil, F. Ababsa, J.-Y. Didier, and M. Mallem, "A gps-imu-camera modelization and calibration for 3d localization dedicated to outdoor mobile applications," in *ICCAS 2010*, 2010, pp. 1580–1585.

[33] 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)*, 2018, pp. 7244–7251.

[34] A. Waegli and J. Skaloud, "Optimization of two gps/mems-imu integration strategies with application to sports," *GPS solutions*, vol. 13, no. 4, pp. 315–326, 2009.

[35] B. Khaleghi, A. El-Ghazal, A. R. Hilal, J. Toonstra, W. B. Miners, and O. A. Basir, "Opportunistic calibration of smartphone orientation in a vehicle," in *IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2015, pp. 1–6.

[36] S. A. Shaukat, K. Munawar, M. Arif, A. I. Bhatti, U. I. Bhatti, and U. M. Al-Saggaf, "Robust vehicle localization with gps dropouts," in *6th International Conference on Intelligent and Advanced Systems (ICIAS)*, 2016, pp. 1–6.

[37] J. Almazán, L. M. Bergasa, J. J. Yebes, R. Barea, and R. Arroyo, "Full auto-calibration of a smartphone on board a vehicle using imu and gps embedded sensors," in *IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 1374–1380.

[38] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3565–3572.

[39] C. X. Guo, K. Sartipi, R. C. DuToit, G. A. Georgiou, R. Li, J. OLeary, E. D. Nerurkar, J. A. Hesch, and S. I. Roumeliotis, "Resource-aware large-scale cooperative three-dimensional mapping using multiple mobile devices," *IEEE Transactions on Robotics*, vol. 34, no. 5, pp. 1349–1369, 2018.

[40] 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.

[41] 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.

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

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

[44] W. Lee, K. Eckenhoff, P. Geneva, and G. Huang, "Gps-aided visual-inertial navigation in large-scale environments," Robot Perception and Navigation Group (RPNG), University of Delaware, Tech. Rep., 2019. [Online]. Available: http://udel.edu/~ghuang/papers/tr_gps-vio.pdf

[45] R. C. DuToit, J. A. Hesch, E. D. Nerurkar, and S. I. Roumeliotis, "Consistent map-based 3d localization on mobile devices," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 6253–6260.

[46] G. Huang, "Improving the consistency of nonlinear estimators: Analysis, algorithms, and applications," Ph.D. dissertation, Department of Computer Science and Engineering, University of Minnesota, 2012. [Online]. Available: https://conservancy.umn.edu/handle/11299/146717

[47] Z. Chen, K. Jiang, and J. Hung, "Local observability matrix and its application to observability analyses," in *Proc. of the 16th Annual Conference of IEEE*, Pacific Grove, CA, Nov. 27–30, 1990, pp. 100–103.

[48] G. Huang, A. I. Mourikis, and S. I. Roumeliotis, "A first-estimates Jacobian EKF for improving SLAM consistency," in *Proc. of the 11th International Symposium on Experimental Robotics*, Athens, Greece, July 14–17, 2008.

[49] 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

[50] T.Takasu, "Rtklib: An open source program package for gnss positioning." [Online]. Available: https://http://www.rtklib.com/