

Degenerate Motion Analysis for Aided INS with Online Spatial and Temporal Sensor Calibration

Yulin Yang¹, Patrick Geneva², Kevin Ekenhoff¹ and Guoquan Huang¹

Abstract—In this paper we perform in-depth observability analysis for both spatial and temporal calibration parameters of an aided inertial navigation system (INS) with global and/or local sensing modalities. In particular, we analytically show that both spatial and temporal calibration parameters are observable if the sensor platform undergoes random motion. More importantly, we identify four degenerate motion primitives that harm the calibration accuracy and thus should be avoided in reality whenever possible. Interestingly, we also prove that these degenerate motions would still hold even in the case where global pose measurements are available. Leveraging a particular multi-state constrained Kalman filter (MSCKF)-based vision-aided inertial navigation system (VINS) with online spatial and temporal calibration, we perform extensively both Monte-Carlo simulations and real-world experiments with the identified degenerate motions to validate our analysis.

Index Terms—Calibration and Identification, Visual-based Navigation, Inertial Navigation System, Observability Analysis

I. INTRODUCTION AND RELATED WORK

IN recent years, inertial measurement units (IMU) have emerged as a popular sensing modality for 6DOF motion estimation. IMUs are proprioceptive devices that provide readings of the local angular velocity and linear acceleration of a moving platform at a high frequency, thereby providing good motion estimates over short time intervals. However, consumer-class MEMS IMUs suffer from non-negligible time-varying biases and noises, preventing the direct integration of IMU measurements to obtain accurate velocity, position, and orientation estimates. To combat long-term drift due to integrating over solely noisy IMU measurements, fusion with information from lower frequency *exteroceptive* sensors (such as optical or event cameras [1]–[8], LiDAR [9] or sonars [10]) is often performed, resulting in an aided inertial navigation system (INS).

Manuscript received: September 9, 2018; Revised December 5, 2018; Accepted December 31, 2018; Date of publication January 17, 2019.

This paper was recommended for publication by Editor Dezhen Song upon evaluation of the Associate Editor and Reviewers' comments. This work was partially supported by the University of Delaware (UD) College of Engineering, the NSF (IIS-1566129), the DTRA (HDTRA1-16-1-0039), and Google Daydream.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors. This includes a video showing the implemented MSCKF-based online spatial and temporal calibration algorithm running on real world datasets. This material is 9.4 MB in size.

¹Y. Yang, K. Ekenhoff and G. Huang are with the Department of Mechanical Engineering, University of Delaware, Newark, DE 19716, USA. {yuyang, keck, ghuang}@udel.edu

²P. Geneva is with the Department of Computer and Information Sciences, University of Delaware, Newark, DE 19716, USA. pgeneva@udel.edu
Digital Object Identifier (DOI): see top of this page.

When fusing the measurements from external exteroceptive sensors, it is vital to determine in high precision the spatial and temporal sensor calibration parameters. In particular, we must know the *rigid-body transformation* from one sensor frame to another in order to correctly fuse motion information extracted from these sensor measurements. In addition, due to improper hardware triggering, transmission delays, and clock synchronization errors, the timestamped sensing data of each sensor may disagree and thus, a timeline misalignment (time offset) between different sensor measurements might occur, which will eventually lead to unstable or inaccurate state estimates. It is therefore critical that these *time offsets* should also be calibrated.

In particular, in vision-aided inertial navigation systems (VINS), sensor calibration of spatial and/or temporal parameters has attracted significant research efforts (e.g., [11]–[16] and references therein). Among them, notable work includes the (iterated) EKF-based camera-IMU extrinsic (i.e., spatial-only) calibration [11], in which nonlinear observability analysis shows that these transformation parameters are observable given sufficiently excited motions. Similar observability analysis is also performed in [13], where an UKF-based visual-inertial online extrinsic calibration that does not require external calibration aids is developed. Moreover, the observability/identifiability analysis [12] based on indistinguishable trajectories again shows that both the global gravity and camera-IMU extrinsic calibration are observable provided general motions. However, all these great works focus only on camera-IMU spatial calibration and perform observability analysis assuming general motions. Recently, temporal sensor calibration has been emerging. For example, in the widely-used calibration toolbox *Kalibr* [14], both sensor-sensor spatial transformation and time offset are estimated *offline* using calibration targets, within a continuous-time batch maximum likelihood estimation (MLE) framework. Online temporal calibration, however, is first rigorously addressed in [15], [17] based on the multi-state constrained Kalman filter (MSCKF) [1]. In this camera-IMU online calibration system, observability (identifiability) analysis is performed showing that the time offset is in general observable and provides the theoretical sufficient conditions for the observability of time offset *alone*, while practical degenerate motions are not fully examined. It should also be noted that very recently an optimization-based online temporal calibration of VINS is introduced in [16] which directly shifts visual feature observations in time to best match IMU constraints.

In this work, however, we focus on the degenerate motion observability analysis for any source-aided INS with both spa-

tial and temporal online calibration. In particular, leveraging our prior work [18], we perform an observability analysis for linearized aided INS with these spatial and temporal parameters for generic exteroceptive sensors, including both local and/or global measurements. Based on this analysis, we identify the degenerate motion primitives that can be encountered in practice and would degrade the calibration performance, thus providing insights about what motions a practitioner may consider when performing calibration. Specifically, the main contributions of this paper include:

- We perform observability analysis for the generalized any source-aided INS with both spatial and temporal calibration parameters, showing that under general motion these calibration parameters are observable.
- We identify four degenerate motion primitives causing online spatial/temporal calibration to partially fail, which hold great practical implications and should be avoided in real applications whenever possible.
- We study the effects of global sensor measurements on the spatial and temporal calibration, for the first time showing that the degenerate motions *still* hold even when global pose measurements are present.
- We implement the MSCKF-based vision-aided INS with online spatial and temporal calibration and validate extensively our degenerate motion analysis with both Monte-Carlo simulations and real-world experiments.

II. PROBLEM STATEMENT

During online navigation, we concurrently estimate the state variables together with spatial and temporal calibration parameters and update them as new measurements are available. The success of calibration heavily depends on the motion of the sensor platform, motivating the identification of *degenerate* motion profiles that negatively affect the estimation accuracy of aided INS.

A. State Vector

We consider the state vector with both spatial and temporal calibration parameters for aided INS: that is the 6DOF rigid-body transformation \mathbf{x}_{calib} and the temporal time offset t_d between the IMU and aiding sensor are included in the state vector. Thus, the total state vector that is being estimated can be written as:¹

$$\mathbf{x} = \left[\mathbf{x}_I^\top \quad \mathbf{x}_{calib}^\top \quad t_d \quad \mathbf{x}_f^\top \right]^\top \quad (1)$$

where \mathbf{x}_I and \mathbf{x}_f represents the IMU and feature states, respectively. For simplicity of the ensuing analysis, we consider one point feature in the state vector (which can be generalized to multiple features as in [19]), and express \mathbf{x}_I and \mathbf{x}_f as follows:

$$\mathbf{x}_I = \left[{}^I_G \bar{q}^\top \quad \mathbf{b}_g^\top \quad {}^G \mathbf{v}_I^\top \quad \mathbf{b}_a^\top \quad {}^G \mathbf{p}_I^\top \right]^\top, \quad \mathbf{x}_f = {}^G \mathbf{p}_f \quad (2)$$

¹Throughout this paper \hat{x} is used to denote the estimate of a random variable x , while $\tilde{x} = x - \hat{x}$ is the error in this estimate. $\mathbf{I}_{n \times m}$ and $\mathbf{0}_{n \times m}$ are the $n \times m$ identity and zero matrices, respectively. The left superscript denotes the frame of reference the vector is expressed with respect to.

where ${}^I_G \bar{q}$ represents the JPL quaternion [20] associated with the rotation matrix ${}^I_G \mathbf{R}$ which rotates vectors from the global frame into the local IMU frame, ${}^G \mathbf{v}_I$ and ${}^G \mathbf{p}_I$ refer to the velocity and position of the IMU as expressed in the global frame, and \mathbf{b}_g and \mathbf{b}_a represent the gyroscope and accelerometer biases, respectively. ${}^G \mathbf{p}_f$ is the position of an environmental feature point measured by the exteroceptive sensors as represented in the global frame. In addition, the spacial calibration state \mathbf{x}_{calib} contains the relative transform between the IMU frame and measurement sensor frame, i.e., $\mathbf{x}_{calib}^\top = [{}^C \bar{q}^\top \quad {}^C \mathbf{p}_I^\top]$.

Moreover, due to the inherited nature of electronic hardware (e.g., asynchronous clocks, data transmission delays and electronic triggering delays), the timestamps reported by each of the sensors will differ from the “true” time that the measurements were recorded. We treat the IMU clock as the true time and estimate the offset of the aiding sensor relative to this base clock [15], [16]. We model the time offset t_d as a constant value:

$$t_d = t_C - t_I \quad (3)$$

where t_C is the time recorded on the sensor measurements, and t_I is the corresponding true IMU time.

B. System Dynamical Model

Based on (1), the system dynamical model including IMU kinematics [20] can be described as:

$$\begin{aligned} {}^I_G \dot{\bar{q}}(t) &= \frac{1}{2} \boldsymbol{\Omega} \left({}^I \boldsymbol{\omega}(t) \right) {}^I_G \bar{q}(t) \\ {}^G \dot{\mathbf{p}}_I(t) &= {}^G \mathbf{v}_I(t), \quad {}^G \dot{\mathbf{v}}_I(t) = {}^G \mathbf{a}(t) \\ \dot{\mathbf{b}}_g(t) &= \mathbf{n}_{wg}(t), \quad \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t) \\ \dot{\mathbf{x}}_{calib}(t) &= \mathbf{0}_{6 \times 1}, \quad \dot{t}_d(t) = 0, \quad \dot{\mathbf{x}}_f(t) = \mathbf{0}_{3 \times 1} \end{aligned} \quad (4)$$

where $\boldsymbol{\omega}$ and \mathbf{a} denote the local angular velocity and linear acceleration, \mathbf{n}_{wg} and \mathbf{n}_{wa} are the white Gaussian noises driving the gyroscope and accelerometer biases. $\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -[\boldsymbol{\omega}] & \boldsymbol{\omega} \\ \boldsymbol{\omega}^\top & 0 \end{bmatrix}$ and $[\cdot]$ represents a skew symmetric matrix.

Based on the nonlinear system model (4), the linearized system state transition matrix $\Phi(k, 1)$ from time step 1 to k can be found analytically and is given by:

$$\Phi(k, 1) = \begin{bmatrix} \Phi_{I(k,1)} & \mathbf{0}_{15 \times 6} & \mathbf{0}_{15 \times 1} & \mathbf{0}_{15 \times 3} \\ \mathbf{0}_{6 \times 15} & \Phi_{Calib(k,1)} & \mathbf{0}_{5 \times 1} & \mathbf{0}_{6 \times 3} \\ \mathbf{0}_{1 \times 15} & \mathbf{0}_{1 \times 6} & \Phi_{t_d(k,1)} & \mathbf{0}_{1 \times 3} \\ \mathbf{0}_{3 \times 15} & \mathbf{0}_{3 \times 6} & \mathbf{0}_{3 \times 1} & \Phi_{f(k,1)} \end{bmatrix} \quad (5)$$

where $\Phi_{I(k,1)}$, $\Phi_{Calib(k,1)} = \mathbf{I}_6$, $\Phi_{t_d(k,1)} = 1$ and $\Phi_{f(k,1)} = \mathbf{I}_3$ represent the state transition matrix for \mathbf{x}_I , \mathbf{x}_{calib} , t_d and \mathbf{x}_f , respectively. Note that the linearized state transition matrix $\Phi_{I(k,1)}$ for IMU can be computed as [19]:

$$\Phi_{I(k,1)} = \begin{bmatrix} \Phi_{I11} & \Phi_{I12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \Phi_{I31} & \Phi_{I32} & \mathbf{I}_3 & \Phi_{I34} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \Phi_{I51} & \Phi_{I52} & \Phi_{I53} & \Phi_{I54} & \mathbf{I}_3 \end{bmatrix} \quad (6)$$

Note that the error state for quaternion is defined as:

$$\delta\bar{q} = \bar{q} \otimes \hat{q}^{-1} \simeq \left[\frac{1}{2}\delta\boldsymbol{\theta}^\top \quad 1 \right]^\top \quad (7)$$

where \otimes denotes the quaternion multiplication [20].

C. Generic Range/Bearing Measurement Model

While navigating through space, an aiding sensor provides range and/or bearing information to point features in the environment, which can be generically modeled as the following (linearized) range and/or bearing measurement model [18]:

$$\mathbf{z}_C = \underbrace{\begin{bmatrix} \lambda_r & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{2 \times 1} & \lambda_b \mathbf{I}_2 \end{bmatrix}}_{\mathbf{\Lambda}} \begin{bmatrix} z^{(r)} \\ \mathbf{z}^{(b)} \end{bmatrix} = \mathbf{\Lambda} \begin{bmatrix} \sqrt{{}^C\mathbf{p}_f^\top {}^C\mathbf{p}_f + n^{(r)}} \\ \mathbf{h}_b({}^C\mathbf{p}_f, \mathbf{n}^{(b)}) \end{bmatrix} \quad (8)$$

$$\simeq \mathbf{\Lambda} \begin{bmatrix} \mathbf{H}_r {}^C\tilde{\mathbf{p}}_f + n^{(r)} \\ \mathbf{H}_b {}^C\tilde{\mathbf{p}}_f + \mathbf{H}_n \mathbf{n}^{(b)} \end{bmatrix} \quad (9)$$

where $z^{(r)}$ and $\mathbf{z}^{(b)} = \mathbf{h}_b(\cdot)$ are the range and bearing measurements, respectively. $\mathbf{\Lambda}$ denotes the measurement selection matrix with scalars λ_r and λ_b ; for example, if $\lambda_b = 1$ and $\lambda_r = 1$, then \mathbf{z}_C contains both range and bearing measurements. ${}^C\mathbf{p}_f$ represents a 3D point in the sensor frame, \mathbf{H}_r and \mathbf{H}_b are the range and bearing measurement Jacobians with respect to ${}^C\mathbf{p}_f$. \mathbf{H}_n is the noise Jacobian, $n^{(r)}$ and $\mathbf{n}^{(b)}$ are zero-mean Gaussian noises for range and bearing measurements, respectively. To keep our presentation concise, we also define \mathbf{H}_{proj} as:

$$\mathbf{H}_{proj} = \frac{\partial \tilde{\mathbf{z}}_C}{\partial {}^C\tilde{\mathbf{p}}_f} = \mathbf{\Lambda} \begin{bmatrix} \mathbf{H}_r \\ \mathbf{H}_b \end{bmatrix} \quad (10)$$

Clearly, depending on the selection matrix $\mathbf{\Lambda}$, \mathbf{H}_{proj} may be the range-only measurement Jacobian \mathbf{H}_r ($\lambda_r = 1$ and $\lambda_b = 0$), bearing-only measurements Jacobian \mathbf{H}_b ($\lambda_r = 0$ and $\lambda_b = 1$), or both.

According to our time offset definition (3), the feature ${}^C\mathbf{p}_f$ in the sensor frame with reported time stamp t corresponds to the time $t - t_d$ in the IMU base clock. Hence, we have:

$${}^C\mathbf{p}_f = {}^C_I \mathbf{R}_G^I \mathbf{R}(t - t_d) \left({}^G\mathbf{p}_f - {}^G\mathbf{p}_I(t - t_d) \right) + {}^C\mathbf{p}_I \quad (11)$$

where ${}^I_G \mathbf{R}(t - t_d)$ and ${}^G\mathbf{p}_I(t - t_d)$ represent the IMU pose at time $t - t_d$, which will be denoted as time step k for simplicity in the ensuing derivations. Therefore, the measurement Jacobian of \mathbf{z}_C with respect to the state vector \mathbf{x} can be written as:

$$\mathbf{H}_C = \frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{x}}} = \frac{\partial \tilde{\mathbf{z}}_C}{\partial {}^C\tilde{\mathbf{p}}_f} \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \tilde{\mathbf{x}}} = \mathbf{H}_{proj} \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \tilde{\mathbf{x}}} \quad (12)$$

D. Global Pose Measurement Model

When calibrating an aided INS, we can limit the drift of the motion estimation by providing the estimator with direct measurements of the IMU pose (e.g., a motion capture system). The direct pose measurement is given by:

$$\mathbf{z}_v = \begin{bmatrix} \bar{q}_n \otimes {}^I_G \bar{q} \\ {}^G\mathbf{p}_I + \mathbf{n}_p \end{bmatrix} \quad (13)$$

where $\bar{q}_n \simeq [\frac{1}{2}\mathbf{n}_\theta^\top \quad 1]^\top$, and \mathbf{n}_θ and \mathbf{n}_p represent Gaussian noises for the orientation and position measurements.

E. Observability Analysis

We perform the observability analysis for the linearized system with both spatial and temporal calibration parameters. To this end, the observability matrix is defined as [21]:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_k \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \boldsymbol{\Phi}(1, 1) \\ \mathbf{H}_2 \boldsymbol{\Phi}(2, 1) \\ \vdots \\ \mathbf{H}_k \boldsymbol{\Phi}(k, 1) \end{bmatrix} \quad (14)$$

where \mathbf{M}_k represents the k -th block row of the observability matrix \mathbf{M} corresponding to time step k . By carefully examining this matrix, we are able to determine the degenerate motions for aided INS, which play important roles in keeping the state estimator healthy.

III. OBSERVABILITY ANALYSIS OF AIDED INS WITH CALIBRATION USING LOCAL MEASUREMENTS

Given a bearing and/or range measurement of a point feature [see (8)], we compute the measurement Jacobian at time step k with respect to the error state $\tilde{\mathbf{x}}_k$ as follows (note that we sometimes omit the time index to keep notation concise):

$$\mathbf{H}_C = \mathbf{H}_{proj} \begin{bmatrix} \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \tilde{\mathbf{x}}_I} & \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \tilde{\mathbf{x}}_{calib}} & \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial t_d} & \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \tilde{\mathbf{x}}_f} \end{bmatrix} \quad (15)$$

By noting that ${}^I_G \hat{\mathbf{R}} = -[{}^I_k \boldsymbol{\omega}] {}^I_k \hat{\mathbf{R}}$, we have:

$$\frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \delta \boldsymbol{\theta}} = {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}} \left[({}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_k}) \right] {}^I_k \hat{\mathbf{R}} \quad (16)$$

$$\frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial {}^C\tilde{\mathbf{p}}_{I_k}} = -{}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}}, \quad \frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial {}^G\tilde{\mathbf{p}}_f} = {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}} \quad (17)$$

$$\frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial \delta \boldsymbol{\theta}_c} = {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}} \left[({}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_k}) \right] {}^I_k \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} \quad (18)$$

$$\frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial {}^C\tilde{\mathbf{p}}_I} = {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} = \mathbf{I}_3 \quad (19)$$

$$\frac{\partial {}^C\tilde{\mathbf{p}}_f}{\partial t_d} = {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}} \left(-[({}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_k})] {}^I_k \hat{\mathbf{R}} \hat{\boldsymbol{\omega}} + {}^G\hat{\mathbf{v}}_{I_k} \right) \quad (20)$$

With these measurement Jacobians, we construct the k -th block of the observability matrix as follows [see (14)]:

$$\mathbf{M}_k = \underbrace{\mathbf{H}_{proj} {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}}_{\Xi_k}$$

where we have:

$$\Xi_{\Gamma k} = [\Gamma_1 \quad \Gamma_2 \quad -\mathbf{I}_3 \delta t_k \quad \Gamma_3 \quad -\mathbf{I}_3 \quad \Gamma_4 \quad {}^C_I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} \quad \Gamma_5 \quad \mathbf{I}_3] \quad (21)$$

$$\Gamma_1 = [{}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_1} - {}^G\hat{\mathbf{p}}_{I_1} \delta t_k - \frac{1}{2} {}^G\mathbf{g} \delta t_k^2] {}^G_{I_1} \hat{\mathbf{R}} \quad (22)$$

$$\Gamma_2 = [{}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_k}] {}^G_{I_k} \hat{\mathbf{R}} \boldsymbol{\Phi}_{I12} - \boldsymbol{\Phi}_{I52} \quad (23)$$

$$\Gamma_3 = -\boldsymbol{\Phi}_{I54}, \quad \Gamma_4 = [{}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_k}] {}^G_{I_k} \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} \quad (24)$$

$$\Gamma_5 = -[({}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_{I_k})] {}^G_{I_k} \hat{\mathbf{R}}^I \hat{\boldsymbol{\omega}} + {}^G\hat{\mathbf{v}}_{I_k} \quad (25)$$

Assuming the system undergoes general (random) motions, we can find the unobservable directions as the right null space of the observability matrix as follows:

$$\mathbf{M}_k \mathbf{N} = \Xi_k \Xi_{\Gamma k} \mathbf{N} = \mathbf{0} \Rightarrow \quad (26)$$

$$\mathbf{N} = [\mathbf{N}_r \quad \mathbf{N}_p] = \begin{bmatrix} I_1 \hat{\mathbf{R}}^G \mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \\ [{}^G \hat{\mathbf{v}}_{I_1}]^G \mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \\ -[{}^G \hat{\mathbf{p}}_{I_1}]^G \mathbf{g} & \mathbf{I}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \\ 0 & \mathbf{0}_{1 \times 3} \\ -[{}^G \hat{\mathbf{p}}_f]^G \mathbf{g} & \mathbf{I}_3 \end{bmatrix} \quad (27)$$

We have computed the above null space based on $\Xi_{\Gamma k} \mathbf{N} = \mathbf{0}$ and thus this *four-dimensional* unobservable subspace (i.e., $\text{span}(\mathbf{N})$) holds regardless range and/or bearing measurements. Close inspection of these null vectors shows that \mathbf{N}_r is the unobservable direction related to the global yaw while \mathbf{N}_p corresponds to the global IMU position. This, as expected, confirms the observability analysis of vision-aided INS [19], and importantly, reveals that given random motions both the spatial and temporal calibration parameters (\mathbf{x}_{calib} and t_d) are observable, which also agrees with the analysis results in [11]–[13], [15].

While these observability properties hold with *general* motion, which however may not always be the case in reality, identifying *degenerate* motion profiles that cause extra unobservable directions to appear in the system, becomes important. Since degenerate motions for the IMU state \mathbf{x}_I and the feature state \mathbf{x}_f have been studied in our previous work [18], in this work we will only focus on motions that cause the calibration parameters to become unobservable.

IV. DEGENERATE MOTION ANALYSIS

Based on our observability analysis, we now present four degenerate motions that cause unobservability of the spatial and/or temporal calibration. Note that due to the space limit, in the following we just sketch some of the proofs while the detailed derivations can be found in our companion technical report [22] which is available online.

A. Pure Translation (No Rotation)

Lemma 1. *If the aided INS undergoes pure translation (no rotation), the translation part ${}^C \mathbf{p}_I$ of the spatial calibration will be unobservable, with unobservable directions as:*

$$\mathbf{N}_{pt} = \begin{bmatrix} \mathbf{0}_{3 \times 15} & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & -({}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}})^{\top} \end{bmatrix}^{\top} \quad (28)$$

Proof. If performing pure translation, we have for any k :

$${}^G \hat{\mathbf{R}} = {}^G \hat{\mathbf{R}}_{I_1}^I \hat{\mathbf{R}} = {}^G \hat{\mathbf{R}}_{I_1} \mathbf{I}_3 = {}^G \hat{\mathbf{R}} \quad (29)$$

With this identity, we can verify that [see (21) and (28)]:

$$\mathbf{M}_k \mathbf{N}_{pt} = \Xi_k \left({}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} - {}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} \right) = \mathbf{0} \quad (30)$$

$$\Rightarrow \mathbf{M} \mathbf{N}_{pt} = \mathbf{0} \quad (31)$$

This completes the proof. \square

Note that as (30) holds no matter what \mathbf{H}_{proj} contains, the null space \mathbf{N}_{pt} is valid for either range or bearing measurements. To geometrically interpret this unobservable direction, we perturb the calibration translation with $\delta \mathbf{p}$:

$${}^C \hat{\mathbf{p}}'_I = {}^C \hat{\mathbf{p}}_I + \delta \mathbf{p} \quad (32)$$

where ($'$) is used to describe the perturbed parameter. The corresponding feature position change along this null space can be modeled as:

$${}^G \hat{\mathbf{p}}'_f = {}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} \delta \mathbf{p} \quad (33)$$

Substitution of (32) and (33) into (11) yields:

$$\begin{aligned} {}^C \hat{\mathbf{p}}'_f &= {}^C \hat{\mathbf{R}}_I^I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}} \left({}^G \hat{\mathbf{p}}'_f - {}^G \hat{\mathbf{p}}_{I_k} \right) + {}^C \hat{\mathbf{p}}'_I \\ &= {}^C \hat{\mathbf{R}}_I^I \hat{\mathbf{R}}_G^I \hat{\mathbf{R}} \left({}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k} \right) + {}^C \hat{\mathbf{p}}_I \\ &= {}^C \hat{\mathbf{p}}_f \end{aligned} \quad (34)$$

It is clear that the system cannot identify this perturbation $\delta \mathbf{p}$ to ${}^C \mathbf{p}_I$ given the feature measurements (11), and thus \mathbf{N}_{pt} is related to the IMU-sensor translation part [see (1)], which implies that if there is no rotation, the translation parameters of spatial calibration are unobservable.

B. One-Axis Rotation

Lemma 2. *If the aided INS undergoes random (general) translation but with only one-axis rotation, the translation calibration parameter ${}^C \mathbf{p}_I$ along the rotation axis will be unobservable, with the following unobservable direction:*

$$\mathbf{N}_{oa} = \begin{bmatrix} \mathbf{0}_{1 \times 18} & ({}^C \hat{\mathbf{R}}^I \hat{\mathbf{k}})^{\top} & 0_1 & -({}^G \hat{\mathbf{R}}_I^I \hat{\mathbf{k}})^{\top} \end{bmatrix}^{\top} \quad (35)$$

where ${}^I \mathbf{k}$ is the constant rotation axis in the IMU frame $\{I\}$.

Proof. See [22]. \square

By careful inspection of the structure of the null space \mathbf{N}_{oa} (35), we can conclude that if performing only one-axis rotation, the calibration translation along the rotation axis ${}^I \mathbf{k}$ is unobservable, while the orientation calibration and time offset are still observable.

C. Constant Local Angular and Linear Velocities

Lemma 3. *If the aided INS undergoes constant local angular velocity ${}^I \boldsymbol{\omega}$ and linear velocity ${}^I \mathbf{v}$, the time offset t_d will be unobservable with the following unobservable direction:*

$$\mathbf{N}_{t1} = \begin{bmatrix} \mathbf{0}_{1 \times 15} & ({}^C \hat{\mathbf{R}}^I \hat{\boldsymbol{\omega}})^{\top} & -({}^C \hat{\mathbf{R}}^I \hat{\mathbf{v}})^{\top} & 1 & \mathbf{0}_{1 \times 3} \end{bmatrix}^{\top} \quad (36)$$

Proof. Based on (21) and (36), we have for any k :

$$\mathbf{M}_k \mathbf{N}_{t1} = \Xi_k \left[\Gamma_{4I} {}^C \hat{\mathbf{R}}^I \hat{\boldsymbol{\omega}} - {}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}_I^I \hat{\mathbf{R}}^I \hat{\mathbf{v}} + \Gamma_5 \right] =$$

$$\Xi_k \left[[{}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k}]_I^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}_I^I \hat{\mathbf{R}}^I \hat{\boldsymbol{\omega}} - {}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}_I^I \hat{\mathbf{R}}^I \hat{\mathbf{v}} -$$

$$\lfloor ({}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k}) \rfloor_I^G \hat{\mathbf{R}}_I^I \hat{\boldsymbol{\omega}} + {}^G \hat{\mathbf{v}}_{I_k} \rfloor = \mathbf{0}$$

$$\Rightarrow \mathbf{M} \mathbf{N}_{t1} = \mathbf{0}$$

This completes the proof. \square

TABLE I: Observability of spatial and temporal calibration of aided INS with different motions

Motion	Unobservable	Observable
No motion	${}^C\mathbf{p}_I, {}^C\mathbf{R}$ and t_d	–
Pure Translation	${}^C\mathbf{p}_I$	${}^C\mathbf{R}$ and t_d
One-axis Rotation	${}^C\mathbf{p}_I$ along rotation axis	${}^C\mathbf{R}$ and t_d
Constant ${}^I\boldsymbol{\omega}$ Constant ${}^I\mathbf{v}$	t_d and ${}^C\mathbf{p}_I$ along rotation axis	${}^C\mathbf{R}$
Constant ${}^I\boldsymbol{\omega}$ Constant ${}^G\mathbf{a}$	t_d and ${}^C\mathbf{p}_I$ along rotation axis	${}^C\mathbf{R}$
One global axis translation Two-axis rotation	–	${}^C\mathbf{R}, {}^C\mathbf{p}_I, t_d$
Random motion	–	${}^C\mathbf{R}, {}^C\mathbf{p}_I, t_d$

D. Constant Local Angular Velocity and Global Linear Acceleration

Lemma 4. *If the aided INS undergoes constant local angular velocity ${}^I\boldsymbol{\omega}$ and global acceleration ${}^G\mathbf{a}$, the time offset t_d will be unobservable with the following unobservable direction:*

$$\mathbf{N}_{t2} = \begin{bmatrix} \mathbf{0}_{1 \times 6} & {}^G\hat{\mathbf{a}} & \mathbf{0}_{1 \times 6} & ({}^C\hat{\mathbf{R}}^I\hat{\boldsymbol{\omega}})^\top & \mathbf{0}_{1 \times 3} & 1 & -({}^G\hat{\mathbf{v}}_{I_1})^\top \end{bmatrix}^\top \quad (37)$$

Proof. See [22]. \square

E. Summary

Table I summarizes our analysis identifying the degenerate motions of temporal and spatial sensor calibration in aided INS. Clearly, as compared to the translation calibration, the observability of the orientation calibration is less sensitive to different motions, and in most cases it is observable. In general, pure three-axis rotation of the aiding sensor is *not* a degenerate motion for calibration parameters, but may be degenerate for other state variables; for example, we have shown before that pure rotation for monocular camera causes loss of feature scale [18]. Similar to the analysis in [11]–[13], in order to get reliable translation calibration, at least two-axis rotation (with random translation) is needed. However, here we have advanced our understanding even further by analytically showing that one-axis rotation will cause the translation calibration to become unobservable along the rotation axis. Therefore, the minimum motion requirement for a reliable spatial and temporal calibration is at least one global axis translation and two-axes rotation, and it is not recommended to perform online IMU-camera calibration under planar motions (e.g., planar VINS for autonomous driving). Interestingly, the time offset calibration generally is observable, except for very strict motions. We want to point out again that these findings are not only important to perform successful spatial and temporal sensor calibration but also useful to monitor the estimator health (if undergoing degenerate motions, the system should be alerted).

V. AIDED INS WITH GLOBAL SENSOR MEASUREMENTS

A straightforward way to improve the accuracy of the total estimation (including motion and calibration parameters) is to include additional global measurements of the state (e.g., pose

measurements from Vicon). In the following, we investigate the effects on the observability of the calibration parameters by the inclusion of *global* information .

A. Observability Analysis

If the system is also fed with global pose measurements, \mathbf{z}_v , we have the following measurement Jacobian [see (1)]:

$$\mathbf{H}_V = \frac{\partial \tilde{\mathbf{z}}_v}{\partial \tilde{\mathbf{x}}} = \begin{bmatrix} \frac{\partial \tilde{\mathbf{z}}_v}{\partial \tilde{\mathbf{x}}_I} & \frac{\partial \tilde{\mathbf{z}}_v}{\partial \tilde{\mathbf{x}}_{calib}} & \frac{\partial \tilde{\mathbf{z}}_v}{\partial t_d} & \frac{\partial \tilde{\mathbf{z}}_v}{\partial \tilde{\mathbf{x}}_f} \end{bmatrix} \quad (38)$$

where $\frac{\partial \tilde{\mathbf{z}}_v}{\partial \delta \boldsymbol{\theta}} = [\mathbf{I}_3 \quad \mathbf{0}_3]^\top$ and $\frac{\partial \tilde{\mathbf{z}}_v}{\partial {}^C\hat{\mathbf{p}}_I} = [\mathbf{0}_3 \quad \mathbf{I}_3]^\top$. Combined with measurements from the other exteroceptive aiding sensors, the overall measurement Jacobian is given by [see (15) and (38)]:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{C_k} \\ \mathbf{H}_{V_k} \end{bmatrix} \quad (39)$$

Thus, for this aided INS supplemented with additional global pose information, the k -th block of the observability matrix is given by (21):

$$\mathbf{M}_k^{(g)} = \underbrace{\begin{bmatrix} \Xi_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_6 \end{bmatrix}}_{\Xi_k^{(g)}} \times \quad (40)$$

$$\begin{bmatrix} \Gamma_1 & \Gamma_2 & -\mathbf{I}_3 \delta_k & \Gamma_3 & -\mathbf{I}_3 & \Gamma_4 & {}^C\hat{\mathbf{R}}_C^I \hat{\mathbf{R}} & \Gamma_5 & \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \end{bmatrix}$$

Based on this matrix (40) and thus the observability matrix (14), we can see that an aided INS undergoing random motion, with both generic range/bearing and global pose measurements, as expected, will have all calibration parameters observable. However, in the next section, we for the first time show that degenerate motions still hold even with global measurements, which, to our understanding, is interesting and important to report as it may counter one's intuition.

B. Degenerate Motion Analysis

If the system undergoes degenerate motions, based on the structure of the observability matrix (40), we have the following results:

Lemma 5. *If aided INS undergoes degenerate motion for calibration parameters, using extra global measurements will not cause the calibration to become observable, that is:*

- For the pure translation, the unobservable directions \mathbf{N}_{pt} w.r.t. translation calibration still holds.
- For the one-axis rotation, the unobservable direction \mathbf{N}_{oa} w.r.t. translation calibration along the rotation axis still holds.
- In the case of constant ${}^I\boldsymbol{\omega}$ and ${}^I\mathbf{v}$, the unobservable direction \mathbf{N}_{t1} w.r.t. time offset calibration still holds.
- In the case of constant ${}^I\boldsymbol{\omega}$ and ${}^G\mathbf{a}$, the unobservable direction \mathbf{N}_{t2} w.r.t. time offset calibration still holds.

Proof. We sketch the proof for each case as follows:

TABLE II: Simulation setup parameters

Parameters	Values
True ${}^C_I \bar{q}$	$[0, 0, 0, 1]^\top$
True ${}^C \mathbf{p}_I$	$[0.01, 0.02, 0.02]^\top$ (m)
True t_d	0.04 (sec)
Initial ${}^C_I \bar{q}$	$[0.0099, 0.0198, -0.0099, 0.9997]^\top$
Initial ${}^C \mathbf{p}_I$	$[0.05, 0.06, -0.02]^\top$ (m)
Initial t_d	0.02 (sec)
Initial σ for calib orientation	0.04 (rad)
Initial σ for calib translation	0.05 (m)
Initial σ for calib time offset	0.02 (sec)

- Given pure translation (no rotation), we verify \mathbf{N}_{pt} [see (40), (28) and Section IV-A]:

$$\mathbf{M}_k^{(g)} \mathbf{N}_{pt} = \Xi_k^{(g)} \begin{bmatrix} \left({}^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}} - {}^G \hat{\mathbf{R}}_I^C \hat{\mathbf{R}} \right) \\ \mathbf{0} \end{bmatrix} = \mathbf{0} \quad (41)$$

- Given one-axis rotation, we verify \mathbf{N}_{oa} [see (40), (35) and Section IV-B]:

$$\mathbf{M}_k^{(g)} \mathbf{N}_{oa} = \Xi_k^{(g)} \begin{bmatrix} \left({}^G \hat{\mathbf{R}}_I^I \hat{\mathbf{k}} - {}^G \hat{\mathbf{R}}_I^I \hat{\mathbf{k}} \right) \\ \mathbf{0} \end{bmatrix} = \mathbf{0} \quad (42)$$

- Given constant ${}^I \boldsymbol{\omega}$ and constant ${}^I \mathbf{v}$, we verify \mathbf{N}_{t1} [see (40), (36) and Section IV-C]:

$$\begin{aligned} \mathbf{M}_k^{(g)} \mathbf{N}_{t1} &= \Xi_k^{(g)} \times \\ &\begin{bmatrix} \left(\begin{array}{c} [{}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k}]_k^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}_I^C \hat{\mathbf{R}}^I \hat{\boldsymbol{\omega}} - {}^G \hat{\mathbf{R}}_I^C \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}^I \hat{\mathbf{v}} \\ -[({}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k})]_k^G \hat{\mathbf{R}}_I^I \hat{\boldsymbol{\omega}} + {}^G \hat{\mathbf{v}}_{I_k} \end{array} \right) \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{0} \end{aligned} \quad (43)$$

- Given constant ${}^I \boldsymbol{\omega}$ and constant ${}^G \mathbf{a}$, we verify \mathbf{N}_{t2} [see (40), (37) and Section IV-D]:

$$\begin{aligned} \mathbf{M}_k^{(g)} \mathbf{N}_{t2} &= \Xi_k^{(g)} \times \\ &\begin{bmatrix} \left(\begin{array}{c} [{}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k}]_k^G \hat{\mathbf{R}}_C^I \hat{\mathbf{R}}_I^C \hat{\mathbf{R}}^I \hat{\boldsymbol{\omega}} \\ -[({}^G \hat{\mathbf{p}}_f - {}^G \hat{\mathbf{p}}_{I_k})]_k^G \hat{\mathbf{R}}_I^I \hat{\boldsymbol{\omega}} + {}^G \hat{\mathbf{v}}_{I_k} - {}^G \hat{\mathbf{v}}_{I_1} - {}^G \hat{\mathbf{a}} \delta t_k \end{array} \right) \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{0} \end{aligned} \quad (44)$$

This completes the proof. \square

With these insightful findings, we stress that if the calibration parameters are observable (e.g., under random motion), the additional global measurements will help to improve the calibration accuracy, because they can make the navigation system observable. However, if the system undergoes degenerate motions for calibration, the additional global measurements will not make the parameters observable. These results are profound, in particular, for practitioners, as naively using global sensors may not necessarily guarantee good calibration.

VI. SIMULATION RESULTS

To validate our observability analysis and the identified degenerate motions for sensor calibration, We have implemented the MSCKF-based vision-aided INS (VINS) [1] with online spatial *and* temporal IMU/camera calibration similar to [15]. In our implementation [22], a fixed-size sliding window is

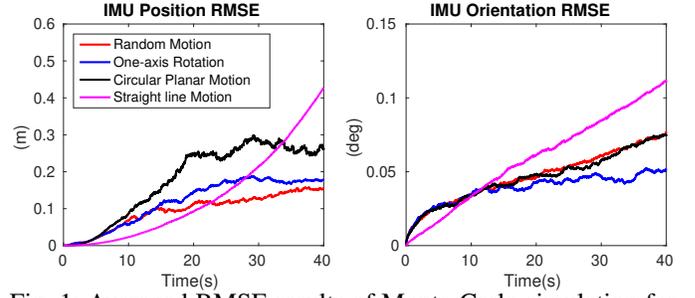


Fig. 1: Averaged RMSE results of Monte-Carlo simulation for the IMU pose estimates from MSCKF-based VINS with online spatial and temporal calibration in four considered motion cases.

built by stacking cloned IMU poses and corresponding visual feature measurements that are marginalized during update. The system has been tested in Monte-Carlo simulations with four motions: (i) *random motion* with varying linear acceleration and angular velocity, in which both spatial and temporal calibration parameters should be observable; (ii) *one-axis rotational motion* with varying linear acceleration and only varying angular motion around one axis, in which only the translation along the rotation axis should be unobservable; (iii) *circular planar motion* with constant local linear and angular velocity, in which both the translation and time offset calibration should be unobservable; (iv) *straight line motion* with constant velocity and no rotation, in which both the translation and time offset calibration should be unobservable. Table II shows the realistic parameters used in our simulation setup. The root mean squared errors (RMSE) [23] are used to quantify the accuracy of both calibration and localization.

Fig. 1 shows the averaged RMSE results of the IMU pose estimates (localization) in all four considered motion cases, while Fig. 2 depicts the RMSE values of both the spatial and temporal calibration parameters. Note that straight line motion will cause the orientation of VINS unobservable, therefore, it drifts faster than other motion patterns. Moreover, it is clear from Fig. 2 that, in the case of random motion, the RMSE of all calibration parameters converge to very small errors, indicating that these parameters converge to the true values. In the case of one-axis rotation, the accuracy of translation does not improve over time, as it is unobservable, while the orientation and t_d converge to reasonable values as they remain observable. In the case of circular planar motion, both the time offset t_d and translation calibration fail to improve, since they are unobservable given constant angular and linear velocities. Examining more carefully, in the case of random motion, each component of translation ${}^C \mathbf{p}_I$ converges close to the true value as shown in Fig. 2, while in the case of one-axis rotation, there is almost no improvement for the translation along y axis (which is the rotation axis in our tests) through online calibration. In the case of circular planar motion, both the time offset t_d and the translation along y axis are unobservable and thus their estimates do not improve or become even worse. Similarly, the time offset and translation calibration estimates are not improved in the case of straight line motion. All these results confirm what we found in our analysis. Due to limited space, more results regarding these

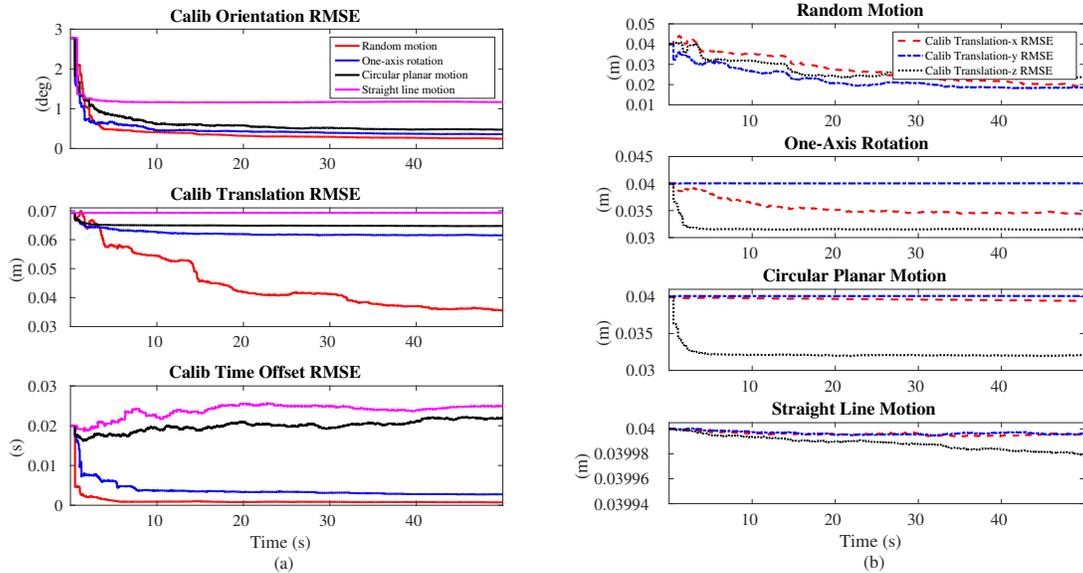


Fig. 2: Averaged RMSE of the spatial and temporal calibration in all four considered motion cases. (a) Calibration errors of both spatial and temporal parameters. (b) Three-axis translation error of spatial calibration.

degenerate motions with true $t_d = 0.0$ and with additional global position measurements can be found in [22].

VII. EXPERIMENTAL RESULTS

To further validate our analysis of these identified degenerate motions, we collected a series of real-world datasets that showcase these degenerate conditions. The left grayscale camera images and inertial readings from a MYNT EYE² sensor were used to provide a testbed for camera and IMU calibration. For quantitative comparison, the calibration results from the *Kalibr* calibration toolbox [14] were treated as reference values³ to which our results would be compared. The implemented system leveraged standard MSCKF features during updates, tracked a maximum of 400 features, and remained real-time during all evaluations. Three datasets with different motion profiles were collected and evaluated: (i) random motion, (ii) generic planar motion with one-axis rotation, and (iii) circular planar motion. The random motion trajectory captured hand-held motion while both the generic planar and circular planar motion were achieved by mounting the sensor testbed to a cart which was pushed in the desired motion profile. Fig. 3 shows the estimated trajectories of our VINS estimators in the three considered motion scenarios.

When evaluating, the initial guesses for both spatial and temporal calibration from the reference *Kalibr* calibration values were perturbed (see [22]). The main results are shown in Fig. 4. As evident, in the case of 3D random motion, all calibration parameters were observable as the calibration estimates were able to quickly converge to the reference *Kalibr* values. However, in the case of random planar motion with one-axis rotation, the orientation and t_d were able to converge while the translation was unable due to the unobservability

about the rotation axis. This confirms our analysis that single-axis rotation is a degenerate case for translation calibration. In the case of circular planar motion (constant local angular and linear velocities), the orientation was able to converge a few degrees before drifting, while both the translation and t_d were unable to converge, which again confirms our analysis. The increasing errors in rotation calibration after ten seconds might be caused by the incorrect t_d estimate which we found greatly impacts the rotation estimate if incorrect.

While for single-axis planar motion, a square trajectory with abrupt changes in velocity ensured good convergence of orientation and t_d . We found that in most cases semi-circular planar path produced poor results due to the large errors in t_d preventing the convergence of the rotation calibration.

VIII. CONCLUSION AND FUTURE WORK

We have performed observability analysis for the linearized aided INS with both spatial and temporal calibration parameters, and shown that both spatial and temporal calibration parameters are observable if the sensor platform undergoes general random motion. More importantly, we have identified four non-trivial degenerate motions that are often encountered in realistic scenarios and might cause the online spatial and temporal calibration to fail. In addition, we have also studied the effects of global pose measurements on the spatial and temporal calibration, and analytically shown that the identified degenerate motions for these calibration parameters still hold even with global pose measurements. Finally, we have implemented the MSCKF-based VINS with online temporal and spatial calibration, and extensively validated our analysis with both Monte-Carlo simulations and real-world experiments on the typical degenerate motion profiles. In the future, we will extend our work to multi-sensor (e.g., camera, IMU and LiDAR) spatial and temporal calibration, as well as investigate the case where the time offset is time-varying (e.g. [24]).

²<https://mynteyei.com/>

³Calib orientation ${}^C\hat{q} = [0.0001, 0.0011, 0.7108, 0.7034]^\top$, calib translation ${}^C\mathbf{p}_f = [0.0441, 0.0029, -0.0184]^\top$, and time offset $t_d = 0.0024$.

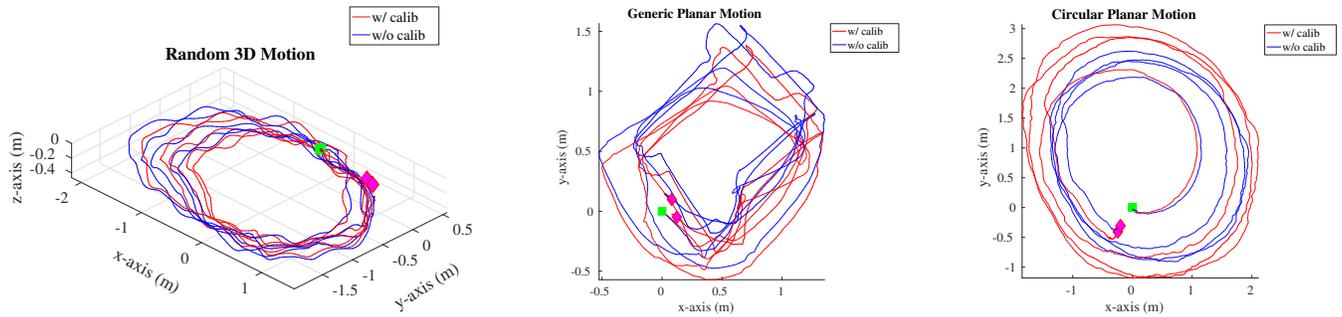


Fig. 3: Estimated trajectories in the real-world experiments. The starting location is denoted with a green square while the end is a magenta diamond. For the case of “w/o calib”, the reference *Kalibr* calibration parameters were used.

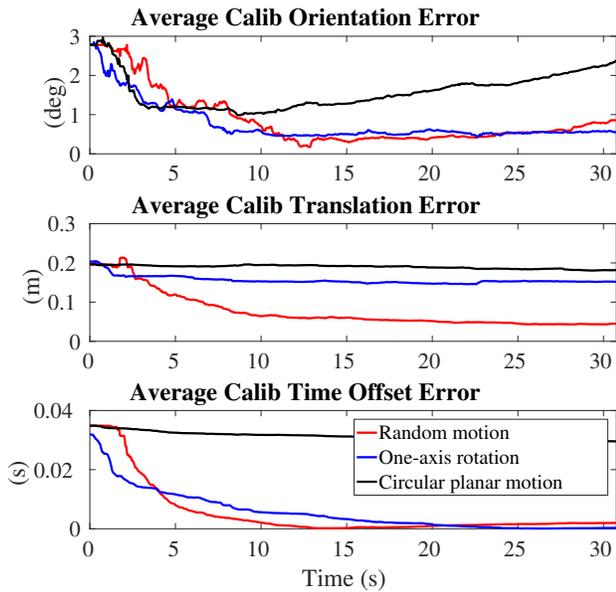


Fig. 4: Average calibration errors of both spatial and temporal parameters in the real-world experiments in respect to the *Kalibr* calibration parameters.

REFERENCES

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

[2] K. Eickenhoff, P. Geneva, and G. Huang, “Direct visual-inertial navigation with analytical preintegration,” in *Proc. of the IEEE International Conference on Robotics and Automation*, Singapore, May 29–Jun.3, 2017, pp. 1429–1435.

[3] 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, Aug 2018.

[4] 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, Feb 2017.

[5] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, “An invariant-ekf vins algorithm for improving consistency,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vancouver, Canada: IEEE, Sept. 24–28 2017, pp. 1578–1585.

[6] A. Zihao Zhu, N. Atanasov, and K. Daniilidis, “Event-based visual inertial odometry,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, Jul. 21–26 2017, pp. 5391–5399.

[7] M. Hsiao, E. Westman, and M. Kaess, “Dense planar-inertial slam with structural constraints,” in *Proc. of the IEEE International Conference on Robotics and Automation*, Brisbane, Australia, May. 21–25 2018, pp. 6521–6528.

[8] Z. Huai and G. Huang, “Robocentric visual-inertial odometry,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, Oct. 1–5 2018.

[9] P. Geneva, K. Eickenhoff, Y. Yang, and G. Huang, “LIPS: Lidar-inertial 3d plane slam,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, Oct. 1–5 2018.

[10] Y. Yang and G. Huang, “Acoustic-inertial underwater navigation,” in *Proc. of the IEEE International Conference on Robotics and Automation*, Singapore, May 29–Jun.3, 2017, pp. 4927–4933.

[11] 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, Oct. 2008.

[12] E. S. Jones and S. Soatto, “Visual-inertial navigation, mapping and localization: A scalable real-time causal approach,” *International Journal of Robotics Research*, vol. 30, no. 4, pp. 407–430, Apr. 2011.

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

[14] P. Furgale, J. Rehder, and R. Siegwart, “Unified temporal and spatial calibration for multi-sensor systems,” in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, Japan, Nov. 3–8, 2013, pp. 1280–1286.

[15] M. Li and A. I. Mourikis, “Online temporal calibration for Camera-IMU systems: Theory and algorithms,” *International Journal of Robotics Research*, vol. 33, no. 7, pp. 947–964, Jun. 2014.

[16] T. Qin and S. Shen, “Online temporal calibration for monocular visual-inertial systems,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, Oct. 1–5 2018.

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

[18] Y. Yang and G. Huang, “Aided inertial navigation with geometric features: Observability analysis,” in *Proc. of the IEEE International Conference on Robotics and Automation*, Brisbane, Australia, May 21–25, 2018, pp. 2334–2340.

[19] J. Hesch, D. Kottas, S. Bowman, and S. Roumeliotis, “Consistency analysis and improvement of vision-aided inertial navigation,” *IEEE Transactions on Robotics*, vol. PP, no. 99, pp. 1–19, 2013.

[20] 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.

[21] 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 Industrial Electronics Society*, Pacific Grove, CA, Nov. 27–30, 1990, pp. 100–103.

[22] Y. Yang, P. Geneva, K. Eickenhoff, and G. Huang, “Degenerate motion analysis for aided ins with online spatial and temporal calibration,” University of Delaware, Tech. Rep., 2018. [Online]. Available: http://udel.edu/~yuyang/downloads/tr_calib.pdf

[23] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. New York: Academic Press, 1988.

[24] C. Guo, D. Kottas, R. DuToit, A. Ahmed, R. Li, and S. Roumeliotis, “Efficient visual-inertial navigation using a rolling-shutter camera with inaccurate timestamps,” in *Proc. of the Robotics: Science and Systems Conference*, Berkeley, CA, Jul. 13–17, 2014.