Consistency Analysis and Improvement of Vision-aided Inertial Navigation

Joel A. Hesch, *Student Member, IEEE,* Dimitrios G. Kottas, *Student Member, IEEE,*
Sean L. Bowman, *Student Member, IEEE,* and Stergios I. Roumeliotis, *Member, IEEE*

***Abstract*—In this paper, we study estimator inconsistency in Vision-aided Inertial Navigation Systems (VINS) from a standpoint of system observability. We postulate that a leading cause of inconsistency is the gain of spurious information along unobservable directions, resulting in smaller uncertainties, larger estimation errors, and divergence. We develop an Observability-Constrained VINS (OC-VINS), which explicitly enforces the unobservable directions of the system, hence preventing spurious information gain and reducing inconsistency. This framework is applicable to several variants of the VINS problem such as Visual Simultaneous Localization and Mapping (V-SLAM) as well as visual-inertial odometry using the Multi-state Constraint Kalman Filter (MSC-KF). Our analysis, along with the proposed method for reducing inconsistency, are extensively validated with simulation trials and real-world experimentation.**

***Index Terms*—Vision-aided Inertial Navigation, Consistency, Nonlinear Estimation, Observability analysis**

I. Introduction

A Vision-aided Inertial Navigation System (VINS) fuses data from a camera and an Inertial Measurement Unit (IMU) to track the six-degrees-of-freedom (d.o.f.) position and orientation (pose) of a sensing platform. This sensor pair is ideal since it combines complementary sensing capabilities [1]. For example, an IMU can accurately track dynamic motions over short time durations, while visual data can be used to estimate the pose displacement (up to scale) between consecutive views. Within the robotics community, VINS has gained popularity as a method to address GPS-denied navigation for several reasons. First, contrary to approaches which utilize wheel odometry, VINS uses inertial sensing that can track general 3D motions of a vehicle. Hence, it is applicable to a variety of platforms such as aerial vehicles, legged robots, and even humans, which are not constrained to move along planar trajectories. Second, unlike laser-scanner-based methods that rely on the existence of structural planes [2] or height invariance in semi-structured environments [3], using vision as an exteroceptive sensor enables VINS methods to work in unstructured areas such as collapsed buildings or outdoors. Furthermore, both cameras and IMUs are light-weight and have low power-consumption requirements, which has lead to recent advances in onboard estimation for payload-constrained platforms such as Micro Aerial Vehicles (MAVs) (e.g., [4]–[7]).

J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis are with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, 55455 USA
e-mail: {joel|kottas|bowman|stergios@cs.umn.edu}.

Numerous VINS approaches have been presented in the literature, including methods based on the Extended Kalman Filter (EKF) [8]–[11], the Unscented Kalman Filter (UKF) [12], and Batch-least Squares (BLS) [13]. Non-parametric estimators, such as the Particle Filter (PF), have also been applied to visual odometry (e.g., [14], [15]). However, these have focused on the simplified problem of estimating the pose of a vehicle whose motion is constrained to 2D, since the number of particles required is exponential in the size of the state vector. Existing work has addressed a variety of issues in VINS, such as reducing its computational cost [4], [9], dealing with delayed observations [7], improving fault detection by processing the visual and inertial measurements in a loosely-coupled manner [5], increasing the accuracy of feature initialization and estimation [16], and improving the robustness to estimator initialization errors [17].

A fundamental issue that has not yet been addressed in the literature is how estimator inconsistency affects VINS. As defined in [18], a state estimator is consistent if the estimation errors are zero-mean and have covariance smaller than or equal to the one calculated by the filter. Estimator inconsistency can have a devastating affect, particularly in navigation applications, since both the current pose estimate and its uncertainty, must be accurate in order to address tasks that depend on the localization solution, such as path planning. For nonlinear systems, several potential sources of inconsistency exist (e.g., motion-model mismatch in target tracking), and great care must be taken when designing an estimator to improve consistency.

To the best of our knowledge, we provide the first report on VINS inconsistency.[1] We focus specifically on estimator inconsistency due to spurious information gain which arises from approximations incurred when applying linear estimation tools to nonlinear problems (i.e., when using linearized estimators such as the EKF). In summary, the main contributions of this work are:

- We analyze the structure of the true and estimated systems and show that for the true system four unobservable directions exist (i.e., 3-d.o.f. global translation and 1-d.o.f. rotation about the gravity vector), while the system

[1] A poster describing the main results of this work appeared at [19]. Moreover, a short version of this paper detailing the OC-VINS framework applied to V-SLAM will appear in [20], while the application to the MSC-KF is accepted to appear in [21]. In comparison to [20] and [21], in this paper we provide an extended theoretic analysis of the problem, including an observability analysis of the linearized VINS system model, and present extensive simulation and experimental results validating our approach.

employed for estimation purposes has only three unobservable directions (3-d.o.f. global translation). Moreover, we postulate that a main source of inconsistency in VINS is spurious information gained when orientation information is incorrectly projected along the direction corresponding to rotations about the gravity vector.
- We propose a simple, yet powerful, estimator modification that explicitly prohibits this incorrect information gain. Our approach is general enough to be applied in multiple VINS domains (e.g., V-SLAM and the MSC-KF [22]) when linearized estimators, such as the EKF, are used.
- We provide extensive evidence to demonstrate inconsistency in standard VINS approaches as well as validate our method with Monte-Carlo simulations to show that it improves consistency and reduces estimation errors as compared to standard VINS. In addition, we demonstrate the performance of our approach experimentally using a miniature IMU and a small-size camera.

The rest of this paper is organized as follows: We begin with an overview of the related work (Sect. II). In Sect. III, we describe the system and measurement models, followed by our analysis of VINS inconsistency in Sect. IV. The proposed estimator modification is presented in Sect. V, and subsequently validated both in simulations and experimentally (Sects. VI and VII). Finally, we provide our concluding remarks and outline our future research directions in Sect. VIII.

II. Related Work

Until recently, little attention was paid to the effect that the observability properties of a system can have on the consistency of a linearized estimator, employed to solve a nonlinear estimation problem. The work by Huang et al. [23]–[25] was the first to identify this connection for several 2D localization problems [i.e., Simultaneous Localization and Mapping (SLAM), and Cooperative Localization (CL)]. The authors showed that, for these problems, a mismatch exists between the number of unobservable directions of the true nonlinear system and the linearized system used for estimation purposes. In particular, the estimated (linearized) system has one-fewer unobservable direction than the true system, allowing the estimator to surreptitiously gain spurious information along the direction corresponding to global orientation (yaw). This increases the estimation errors while erroneously reducing the estimator uncertainty, and leads to inconsistency.

To date, no similar study exists linking the VINS observability properties to estimator inconsistency, despite the fact that several authors have studied VINS observability for various scenarios. For the task of IMU-camera extrinsic calibration, Mirzaei and Roumeliotis [26], as well as, Kelly and Sukhatme [27], have analyzed the system observability using Lie derivatives [28] to determine when the IMU-camera transformation is observable. Jones and Soatto [16] studied VINS observability by examining the indistinguishable trajectories of the system [29] under different sensor configurations (i.e., inertial only, vision only, vision and inertial). Recently, Martinelli [30] utilized the concept of continuous symmetries

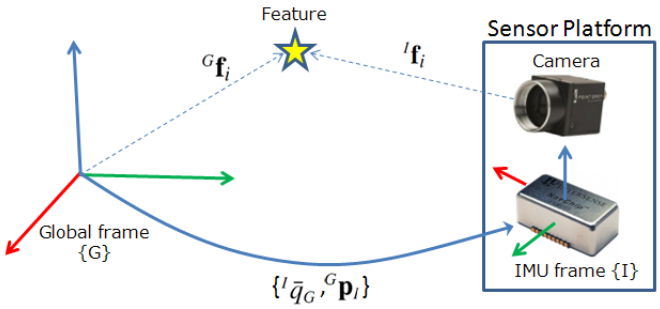


Fig. 1. Sensor platform comprising an IMU and a camera. $\{^I\bar{q}_G, {}^G\mathbf{p}_I\}$ are the quaternion of orientation and position vector describing the pose of the sensing frame $\{I\}$ with respect to the global frame $\{G\}$. The i-th feature's 3D coordinates are denoted as $^G\mathbf{f}_i$, and $^I\mathbf{f}_i$, with respect to $\{G\}$ and $\{I\}$, respectively.

to show that the IMU biases, 3D velocity, and absolute roll and pitch angles are observable for VINS.

In this work, we study the observability properties of the ideal linearized VINS model (i.e., the one whose Jacobians are evaluated at the true states), and show it has four unobservable d.o.f., corresponding to three-d.o.f. global translations and one-d.o.f. global rotation about the gravity vector. Moreover, we show that when the estimated states are used for evaluating the Jacobians, as is the case for the EKF, the number of unobservable directions is reduced by one. In particular, the global rotation about the gravity vector becomes (erroneously) observable, allowing the estimator to gain spurious information and leading to inconsistency. These results confirm the findings of [16] and [30] using a different approach (i.e., the observability matrix), while additionally specifying the exact mathematical structure of the unobservable directions necessary for assessing the EKF's inconsistency.[2]

To address this problem, we introduce a modification of the VINS EKF where its estimated Jacobians are updated so as to ensure that the number of unobservable directions is the same as when using the true Jacobians. In this manner, the global rotation about the gravity vector remains unobservable (as it should) and the consistency of the VINS EKF is significantly improved.

III. VINS Estimator Description

We begin with an overview of the propagation and measurement models which govern the VINS. In particular, we employ an EKF for fusing the camera and IMU measurements to estimate the state of the system including the pose, velocity, and IMU biases, as well as the 3D positions of visual landmarks observed by the camera. We operate in a previously unknown environment and utilize two types of visual features in our VINS framework. The first are opportunistic features (OFs) that can be accurately and efficiently tracked across short image sequences (e.g., using KLT [31]), but are not visually distinctive enough to be efficiently recognized when revisiting an area. OFs can be efficiently used to estimate the motion

[2]The analysis in [30] addresses the special case with one known feature at the origin, which in contrast to our approach, cannot be used to determine the nullspace directions when multiple features are considered.

of the camera over short time horizons (i.e., using the MSC-KF), but they are not included in the state vector. The second are Persistent Features (PFs), which are typically much fewer in number, and can be reliably redetected when revisiting an area (e.g., SIFT keys [32]). The 3D coordinates of the PFs are estimated to construct a map of the area.

A. System State and Propagation Model

The EKF estimates the 3D IMU pose and linear velocity together with the time-varying IMU biases and a map of visual features (see Fig. 1). In particular, the filter state is the $(16+3N)\times 1$ vector:

$$\begin{aligned}\mathbf{x} &= \begin{bmatrix} {}^I\bar{q}_G^T & \mathbf{b}_g^T & {}^G\mathbf{v}_I^T & \mathbf{b}_a^T & {}^G\mathbf{p}_I^T & | & {}^G\mathbf{f}_1^T & \cdots & {}^G\mathbf{f}_N^T \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{x}_s^T & | & \mathbf{x}_f^T \end{bmatrix}^T, \end{aligned} \tag{1}$$

where $\mathbf{x}_s(t)$ is the 16×1 sensor platform state, and $\mathbf{x}_f(t)$ is the $3N\times 1$ state of the feature map. The first component of the sensor platform state is ${}^I\bar{q}_G(t)$ which is the unit quaternion representing the orientation of the *global frame* $\{G\}$ in the IMU frame, $\{I\}$, at time t. The frame $\{I\}$ is attached to the IMU, while $\{G\}$ is a local-vertical reference frame whose origin coincides with the initial IMU position. The sensor platform state also includes the position and velocity of $\{I\}$ in $\{G\}$, denoted by the 3×1 vectors ${}^G\mathbf{p}_I(t)$ and ${}^G\mathbf{v}_I(t)$, respectively. The remaining components are the biases, $\mathbf{b}_g(t)$ and $\mathbf{b}_a(t)$, affecting the gyroscope and accelerometer measurements, which are modeled as random-walk processes driven by the zero-mean, white Gaussian noise $\mathbf{n}_{wg}(t)$ and $\mathbf{n}_{wa}(t)$, respectively.

The map state, $\mathbf{x}_f$, comprises the 3D coordinates of N PFs, ${}^G\mathbf{f}_i$, $i=1,\ldots,N$, and grows as new PFs are observed [33]. In contrast, we do not store OFs in the map. Instead, all OFs are processed and marginalized on-the-fly using the MSC-KF approach [22] (see Sect. III-B). With the state of the system now defined, we turn our attention to the continuous-time model which governs the state of the system.

1) Continuous-time model: The system model describing the time evolution of the state is (see [34], [35]):

$$ {}^I\dot{\bar{q}}_G(t) = \frac{1}{2}\mathbf{\Omega}(\boldsymbol{\omega}(t)){}^I\bar{q}_G(t) \tag{2}$$

$$ {}^G\dot{\mathbf{p}}_I(t) = {}^G\mathbf{v}_I(t) \tag{3}$$

$$ {}^G\dot{\mathbf{v}}_I(t) = {}^G\mathbf{a}(t) \tag{4}$$

$$ \dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg}(t) \tag{5}$$

$$ \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t) \tag{6}$$

$$ {}^G\dot{\mathbf{f}}_i(t) = \mathbf{0}_{3\times 1}\ ,\quad i=1,\ldots,N. \tag{7}$$

In these expressions, $\boldsymbol{\omega}(t)=[\omega_1(t)\ \ \omega_2(t)\ \ \omega_3(t)]^T$ is the rotational velocity of the IMU, expressed in $\{I\}$, ${}^G\mathbf{a}$ is the IMU acceleration expressed in $\{G\}$, and

$$\mathbf{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor\boldsymbol{\omega}\times\rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix},\quad \lfloor\boldsymbol{\omega}\times\rfloor \triangleq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}.$$

The gyroscope and accelerometer measurements, $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$, are modeled as

$$\boldsymbol{\omega}_m(t) = \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \tag{8}$$

$$\mathbf{a}_m(t) = \mathbf{C}({}^I\bar{q}_G(t))\,({}^G\mathbf{a}(t) - {}^G\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t), \tag{9}$$

where $\mathbf{n}_g$ and $\mathbf{n}_a$ are zero-mean, white Gaussian noise processes, and ${}^G\mathbf{g}$ is the gravitational acceleration. The matrix $\mathbf{C}(\bar{q})$ is the rotation matrix corresponding to $\bar{q}$. The PFs belong to the static scene, thus, their time derivatives are zero [see (7)].

Linearizing at the current estimates and applying the expectation operator on both sides of (2)-(7), we obtain the state estimate propagation model

$$ {}^I\dot{\hat{\bar{q}}}_G(t) = \frac{1}{2}\mathbf{\Omega}(\hat{\boldsymbol{\omega}}(t)){}^I\hat{\bar{q}}_G(t) \tag{10}$$

$$ {}^G\dot{\hat{\mathbf{p}}}_I(t) = {}^G\hat{\mathbf{v}}_I(t) \tag{11}$$

$$ {}^G\dot{\hat{\mathbf{v}}}_I(t) = \mathbf{C}^T({}^I\hat{\bar{q}}_G(t))\,\hat{\mathbf{a}}(t) + {}^G\mathbf{g} \tag{12}$$

$$ \dot{\hat{\mathbf{b}}}_g(t) = \mathbf{0}_{3\times 1} \tag{13}$$

$$ \dot{\hat{\mathbf{b}}}_a(t) = \mathbf{0}_{3\times 1} \tag{14}$$

$$ {}^G\dot{\hat{\mathbf{f}}}_i\ (t) = \mathbf{0}_{3\times 1}\ ,\quad i=1,\ldots,N, \tag{15}$$

where $\hat{\mathbf{a}}(t)=\mathbf{a}_m(t)-\hat{\mathbf{b}}_a(t)$, and $\hat{\boldsymbol{\omega}}(t)=\boldsymbol{\omega}_m(t)-\hat{\mathbf{b}}_g(t)$.

The $(15+3N)\times 1$ error-state vector is defined as

$$\begin{aligned}\widetilde{\mathbf{x}} &= \begin{bmatrix} {}^I\boldsymbol{\delta\theta}_G^T & \widetilde{\mathbf{b}}_g^T & {}^G\widetilde{\mathbf{v}}_I^T & \widetilde{\mathbf{b}}_a^T & {}^G\widetilde{\mathbf{p}}_I^T & | & {}^G\widetilde{\mathbf{f}}_1^T & \cdots & {}^G\widetilde{\mathbf{f}}_N^T \end{bmatrix}^T \\ &= \begin{bmatrix} \widetilde{\mathbf{x}}_s^T & | & \widetilde{\mathbf{x}}_f^T \end{bmatrix}^T, \end{aligned} \tag{16}$$

where $\widetilde{\mathbf{x}}_s(t)$ is the 15×1 error state corresponding to the sensing platform, and $\widetilde{\mathbf{x}}_f(t)$ is the $3N\times 1$ error state of the map. For the IMU position, velocity, biases, and the map, an additive error model is utilized (i.e., $\widetilde{x}=x-\hat{x}$ is the error in the estimate $\hat{x}$ of a quantity x). However, for the quaternion we employ a multiplicative error model [35]. Specifically, the error between the quaternion $\bar{q}$ and its estimate $\hat{\bar{q}}$ is the 3×1 angle-error vector, $\boldsymbol{\delta\theta}$, implicitly defined by the error quaternion

$$\delta\bar{q} = \bar{q}\otimes\hat{\bar{q}}^{-1} \simeq \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}^T & 1 \end{bmatrix}^T, \tag{17}$$

where $\delta\bar{q}$ describes the small rotation that causes the true and estimated attitude to coincide. This allows us to represent the attitude uncertainty by the 3×3 covariance matrix $\mathbb{E}[\boldsymbol{\delta\theta}\boldsymbol{\delta\theta}^T]$, which is a minimal representation.

The linearized continuous-time error-state equation is

$$\begin{aligned}\dot{\widetilde{\mathbf{x}}} &= \begin{bmatrix} \mathbf{F}_s & \mathbf{0}_{15\times 3N} \\ \mathbf{0}_{3N\times 15} & \mathbf{0}_{3N} \end{bmatrix}\widetilde{\mathbf{x}} + \begin{bmatrix} \mathbf{G}_s \\ \mathbf{0}_{3N\times 12} \end{bmatrix}\mathbf{n} \\ &= \mathbf{F}\,\widetilde{\mathbf{x}} + \mathbf{G}\,\mathbf{n}, \end{aligned} \tag{18}$$

where $\mathbf{0}_{3N}$ denotes the $3N\times 3N$ matrix of zeros. Here, $\mathbf{n}$ is the vector comprising the IMU measurement noise terms as well as the process noise driving the IMU biases, i.e.,

$$\mathbf{n} = \begin{bmatrix} \mathbf{n}_g^T & \mathbf{n}_{wg}^T & \mathbf{n}_a^T & \mathbf{n}_{wa}^T \end{bmatrix}^T \tag{19}$$

while $\mathbf{F}_s$ is the continuous-time error-state transition matrix corresponding to the sensor platform state, and $\mathbf{G}_s$ is the

continuous-time input noise matrix, i.e.,

$$\mathbf{F}_s = \begin{bmatrix} -\lfloor \hat{\boldsymbol{\omega}} \times \rfloor & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{C}^T({}^I\hat{\bar{q}}_G)\lfloor \hat{\mathbf{a}} \times \rfloor & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C}^T({}^I\hat{\bar{q}}_G) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (20)$$

$$\mathbf{G}_s = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C}^T({}^I\hat{\bar{q}}_G) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (21)$$

where $\mathbf{I}_3$ is the 3×3 identity matrix. The system noise is modeled as a zero-mean white Gaussian process with autocorrelation $\mathbb{E}[\mathbf{n}(t)\mathbf{n}^T(\tau)] = \mathbf{Q}_c\delta(t-\tau)$, where $\mathbf{Q}_c$ depends on the IMU noise characteristics and is computed off-line [35].

2) Discrete-time implementation: The IMU signals $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$ are sampled at a constant rate $1/\delta t$, where $\delta t \triangleq t_{k+1} - t_k$. Every time a new IMU measurement is received, the state estimate is propagated using 4th-order Runge-Kutta numerical integration of (10)–(15). In order to derive the covariance propagation equation, we compute the discrete-time state transition matrix, $\boldsymbol{\Phi}_k$, and the discrete-time system noise covariance matrix, $\mathbf{Q}_k$, as

$$\boldsymbol{\Phi}_k = \boldsymbol{\Phi}(t_{k+1}, t_k) = \exp\left(\int_{t_k}^{t_{k+1}} \mathbf{F}(\tau)\mathrm{d}\tau\right) \quad (22)$$

$$\mathbf{Q}_k = \int_{t_k}^{t_{k+1}} \boldsymbol{\Phi}(t_{k+1}, \tau)\mathbf{G}\mathbf{Q}_c\mathbf{G}^T\boldsymbol{\Phi}^T(t_{k+1}, \tau)\mathrm{d}\tau.$$

The covariance is then propagated as

$$\mathbf{P}_{k+1|k} = \boldsymbol{\Phi}_k\mathbf{P}_{k|k}\boldsymbol{\Phi}_k^T + \mathbf{Q}_k. \quad (23)$$

We note that in the above expression, and throughout the paper, $\mathbf{P}_{i|j}$ and $\hat{\mathbf{x}}_{i|j}$ denote the estimates of the error-state covariance and state, respectively, at time-step i computed using measurements up to time-step j.

B. Measurement Update Model

As the camera-IMU platform moves, the camera observes both opportunistic and persistent visual features. These measurements are utilized to concurrently estimate the motion of the sensing platform and the map of PFs. We distinguish three types of filter updates: (i) PF updates of features already in the map, (ii) initialization of PFs not yet in the map, and (iii) OF updates. We first describe the feature measurement model, and subsequently detail how it is employed in each case.

To simplify the discussion, we consider the observation of a single PF point $\mathbf{f}_i$. The camera measures $\mathbf{z}_i$, which is the perspective projection of the 3D point ${}^I\mathbf{f}_i$, expressed in the current IMU frame $\{I\}$, onto the image plane[3], i.e.,

$$\mathbf{z}_i = \frac{1}{z}\begin{bmatrix} x \\ y \end{bmatrix} + \boldsymbol{\eta}_i \quad (24)$$

$$\text{where} \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = {}^I\mathbf{f}_i = \mathbf{C}\left({}^I\bar{q}_G\right)\left({}^G\mathbf{f}_i - {}^G\mathbf{p}_I\right). \quad (25)$$

The measurement noise, $\boldsymbol{\eta}_i$, is modeled as zero mean, white Gaussian with covariance $\mathbf{R}_i$. The linearized error model is

$$\tilde{\mathbf{z}}_i = \mathbf{z}_i - \hat{\mathbf{z}}_i \simeq \mathbf{H}_i\tilde{\mathbf{x}} + \eta_i \quad (26)$$

where $\hat{\mathbf{z}}$ is the expected measurement computed by evaluating (25) at the current state estimate, and the measurement Jacobian, $\mathbf{H}_i$, is

$$\mathbf{H}_i = \mathbf{H}_c\begin{bmatrix}\mathbf{H}_{\bar{q}} & \mathbf{0}_{3\times 9} & \mathbf{H}_{\mathbf{p}} \mid \mathbf{0}_3 \cdots \mathbf{H}_{\mathbf{f}_i} \cdots \mathbf{0}_3\end{bmatrix} \quad (27)$$

with

$$\mathbf{H}_c = \frac{1}{z^2}\begin{bmatrix} z & 0 & -x \\ 0 & z & -y \end{bmatrix} \quad (28)$$

$$\mathbf{H}_{\bar{q}} = \lfloor \mathbf{C}\left({}^I\bar{q}_G\right)\left({}^G\mathbf{f}_i - {}^G\mathbf{p}_I\right) \times \rfloor \quad (29)$$

$$\mathbf{H}_{\mathbf{p}} = -\mathbf{C}\left({}^I\bar{q}_G\right) \quad (30)$$

$$\mathbf{H}_{\mathbf{f}_i} = \mathbf{C}\left({}^I\bar{q}_G\right). \quad (31)$$

evaluated at the current state estimate. Here, $\mathbf{H}_c$, is the Jacobian of the camera's perspective projection with respect to ${}^I\mathbf{f}_i$, while $\mathbf{H}_{\bar{q}}$, $\mathbf{H}_{\mathbf{p}}$, and $\mathbf{H}_{\mathbf{f}_i}$, are the Jacobians of ${}^I\mathbf{f}_i$ with respect to ${}^I\bar{q}_G$, ${}^G\mathbf{p}_I$, and ${}^G\mathbf{f}_i$.

This measurement model is utilized in each of the three update methods. For PFs that are already in the map, we directly apply the measurement model (25)-(27) to update the filter. In particular, we compute the measurement residual $\mathbf{r}_i$, along with its covariance $\mathbf{S}_i$, and the Kalman gain $\mathbf{K}_i$, i.e.,

$$\mathbf{r}_i = \mathbf{z}_i - \hat{\mathbf{z}}_i \quad (32)$$

$$\mathbf{S}_i = \mathbf{H}_i\mathbf{P}_{k+1|k}\mathbf{H}_i^T + \mathbf{R}_i \quad (33)$$

$$\mathbf{K}_i = \mathbf{P}_{k+1|k}\mathbf{H}_i^T\mathbf{S}_i^{-1}. \quad (34)$$

and update the EKF state and covariance as

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_i\mathbf{r}_i \quad (35)$$

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{P}_{k+1|k}\mathbf{H}_i^T\mathbf{S}_i^{-1}\mathbf{H}_i\mathbf{P}_{k+1|k} \quad (36)$$

For previously unseen (new) PFs, we compute an initial estimate, along with covariance and cross-correlations by solving a bundle-adjustment problem over a short time window [37] (see Appendix A). Finally, for OFs, we employ the MSC-KF to impose an efficient (linear complexity) pose update constraining all the views from which a set of features was seen. For a detailed description of the MSC-KF algorithm, we refer the interested reader to [22].

[3]Without loss of generality, we express the image measurement in normalized pixel coordinates, and consider the camera frame to be coincident with the IMU. In practice, we perform both intrinsic and extrinsic IMU-camera calibration off-line [26], [36].

IV. VINS Observability Analysis

In this section, we examine the observability properties of the linearized VINS model. Specifically, we first study and analytically determine the four unobservable directions of the *ideal* linearized VINS (i.e., the system whose Jacobians are evaluated at the true states). Subsequently, we show that the linearized VINS used by the EKF, whose Jacobians are evaluated using the current state estimates, has only three unobservable directions (i.e., the ones corresponding to global translation), while the one corresponding to global rotation about the gravity vector becomes (erroneously) observable. The key findings of this analysis are then employed in Sect. V for improving the consistency of the EKF-based VINS.

A. Observability analysis of the ideal linearized VINS model

The Observability matrix [38] is defined as a function of the linearized measurement model, $\mathbf{H}$, and the discrete-time state transition matrix, $\mathbf{\Phi}$, which are in turn functions of the linearization point, $\mathbf{x}^\star$, i.e.,

$$\mathbf{M}\left(\mathbf{x}^\star\right) = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2\mathbf{\Phi}_{2,1} \\ \vdots \\ \mathbf{H}_k\mathbf{\Phi}_{k,1} \end{bmatrix} \tag{37}$$

where $\mathbf{\Phi}_{k,1} = \mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_1$ is the state transition matrix from time step 1 to k. We start by first considering the case where we use the true state values as linearization point $\mathbf{x}^\star$ for evaluating the system and measurement Jacobians. Moreover, and in order to preserve the clarity of presentation, we focus on the case where only a single feature point is visible.[4] The first block-row of $\mathbf{M}$ is written as [see (27)] (for $k=1$):

$$\mathbf{H}_k = \mathbf{\Psi}_1\begin{bmatrix}\mathbf{\Psi}_2 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{I}_3\end{bmatrix} \tag{38}$$

where

$$\mathbf{\Psi}_1 = \mathbf{H}_{c,k}\mathbf{C}\left({}^{I_k}\bar{q}_G\right) \tag{39}$$

$$\mathbf{\Psi}_2 = \lfloor {}^G\mathbf{f} - {}^G\mathbf{p}_{I_k}\times\rfloor\mathbf{C}\left({}^{I_k}\bar{q}_G\right)^T \tag{40}$$

and ${}^{I_k}\bar{q}_G$, denotes the rotation of $\{G\}$ with respect to frame $\{I_k\}$ at time step $k=1$.

To compute the remaining block rows of the observability matrix, we require $\mathbf{\Phi}_{k,1}$, which we determine analytically by solving the matrix differential equation [33]

$$\dot{\mathbf{\Phi}}_{k,1} = \mathbf{F}\mathbf{\Phi}_{k,1}, \quad \text{i.c. } \mathbf{\Phi}_{1,1} = \mathbf{I}_{18}. \tag{41}$$

with $\mathbf{F}$ detailed in (18). The solution has the following structure

$$\mathbf{\Phi}_{k,1} = \begin{bmatrix} \mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{\Phi}_{31} & \mathbf{\Phi}_{32} & \mathbf{I}_3 & \mathbf{\Phi}_{34} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{\Phi}_{51} & \mathbf{\Phi}_{52} & \delta t(k-1)\mathbf{I}_3 & \mathbf{\Phi}_{54} & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \tag{42}$$

[4]Note that the case of multiple features can be easily captured by our analysis by appropriately augmenting the corresponding matrices. Also, the derived nullspace directions remain the same, in terms of the number, with an identity matrix ($-\lfloor {}^G\mathbf{f}_i\times\rfloor {}^G\mathbf{g}$) appended to the ones corresponding to global translation (rotation) for each new feature [see also (56)].

where among the different block elements $\mathbf{\Phi}_{ij}$, we list below the ones necessary in our analysis and refer to [33] for the remaining ones, as well as the details of the approach followed for solving (41):

$$\mathbf{\Phi}_{11} = \mathbf{C}\left({}^{I_k}\bar{q}_{I_1}\right) \tag{43}$$

$$\mathbf{\Phi}_{31} = -\lfloor\left({}^G\mathbf{v}_{I_k} - {}^G\mathbf{v}_{I_1}\right) + {}^G\mathbf{g}\,(k-1)\,\delta t\times\rfloor\mathbf{C}\left({}^G\bar{q}_{I_1}\right) \tag{44}$$

$$\mathbf{\Phi}_{51} = \lfloor {}^G\mathbf{p}_{I_1} + {}^G\mathbf{v}_{I_1}(k-1)\delta t - \frac{1}{2}{}^G\mathbf{g}((k-1)\delta t)^2 - {}^G\mathbf{p}_{I_k}\times\rfloor\mathbf{C}({}^G\bar{q}_{I_1}) \tag{45}$$

By multiplying (38) at time-step k and (42) we obtain the k-th block row of $\mathbf{M}$, for $k>1$:

$$\begin{aligned}\mathbf{M}_k &= \mathbf{H}_k\mathbf{\Phi}_{k,1} \\ &= \mathbf{\Gamma}_1\begin{bmatrix}\mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & -\delta t(k-1)\mathbf{I}_3 & \mathbf{\Gamma}_4 & -\mathbf{I}_3 & \mathbf{I}_3\end{bmatrix}\end{aligned} \tag{46}$$

where

$$\mathbf{\Gamma}_1 = \mathbf{H}_{c,k}\mathbf{C}\left({}^{I_k}\bar{q}_G\right) \tag{47}$$

$$\begin{aligned}\mathbf{\Gamma}_2 = \lfloor {}^G\mathbf{f} - {}^G\mathbf{p}_{I_1} - {}^G\mathbf{v}_{I_1}\,(k-1)\,\delta t + \frac{1}{2}{}^G\mathbf{g}\,((k-1)\,\delta t)^2\times\rfloor \\ \cdot\,\mathbf{C}\left({}^{I_1}\bar{q}_G\right)^T\end{aligned} \tag{48}$$

$$\mathbf{\Gamma}_3 = \lfloor {}^G\mathbf{f} - {}^G\mathbf{p}_{I_k}\times\rfloor\mathbf{C}^T({}^{I_k}\bar{q}_G)\mathbf{\Phi}_{12} - \mathbf{\Phi}_{52} \tag{49}$$

$$\mathbf{\Gamma}_4 = -\mathbf{\Phi}_{54} \tag{50}$$

At this point, we state the main result of our analysis

Theorem 1. *The right nullspace $\mathbf{N}_1$ of the observability matrix $\mathbf{M}(\mathbf{x})$ [see (37)] of the linearized VINS*

$$\mathbf{M}(\mathbf{x})\mathbf{N}_1 = \mathbf{0} \tag{51}$$

spans the following four directions:

$$\mathbf{N}_1 = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^{I_1}\bar{q}_G\right){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & -\lfloor {}^G\mathbf{v}_{I_1}\times\rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{p}_{I_1}\times\rfloor {}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{f}\times\rfloor {}^G\mathbf{g} \end{bmatrix} = \begin{bmatrix}\mathbf{N}_{t,1} & | & \mathbf{N}_{r,1}\end{bmatrix}. \tag{52}$$

Proof: The fact that $\mathbf{N}_1$ is indeed the right nullspace of $\mathbf{M}(\mathbf{x})$ can be easily verified by multiplying each block row of $\mathbf{M}$ [see (46)] with $\mathbf{N}_{t,1}$ and $\mathbf{N}_{r,1}$ in (52). Since $\mathbf{M}_k\mathbf{N}_{t,1} = \mathbf{0}$ and $\mathbf{M}_k\mathbf{N}_{r,1} = \mathbf{0}$, it follows that $\mathbf{M}\mathbf{N}_1 = \mathbf{0}$. ■

Remark 1. The 18×3 block column $\mathbf{N}_{t,1}$ corresponds to global translations, i.e., translating both the sensing platform and the landmark by the same amount.

Remark 2. The 18×1 column $\mathbf{N}_{r,1}$ corresponds to global rotations of the sensing platform and the landmark about the gravity vector.

B. Observability analysis of the EKF linearized VINS model

Ideally, any VINS estimator should employ a linearized system with an unobservable subspace that matches the true unobservable directions (52), both in number and structure. However, when linearizing about the estimated state $\hat{\mathbf{x}}$, $\widehat{\mathbf{M}} = \mathbf{M}\left(\hat{\mathbf{x}}\right)$ gains rank due to errors in the state estimates across time [33]. In particular, the last two block columns of $\mathbf{M}_k$ in (46) remain the same when computing $\widehat{\mathbf{M}}_k = \widehat{\mathbf{H}}_k\widehat{\mathbf{\Phi}}_{k,1}$ from

Algorithm 1 OC-VINS Algorithm Overview

```
Initialization: Compute initial nullspace from (55)
while running do
  Propagation:
    Integrate state equations [see (10)-(15)]
    Compute nullspace at current time-step from (56)
    Compute Φ_k from (22)
    Modify Φ_k using (60)-(62)
    Propagate covariance [see (23)]
  Update:
  for all observed features do
    Compute measurement Jacobian from (27)
    Modify H using (69)-(74)
    Apply filter update [see (32)-(36)]
  end for
  New landmark initialization:
  for all new PFs observed do
    Initialize ^Gf̂_i, using approach in Appendix A
    Create nullspace block, N_{f_i}, for ^Gf̂_i [see (52)]
    Augment N_k with the new sub-block N_{f_i}
  end for
end while
```

the Jacobians $\widehat{\mathbf{H}}_k$ and $\widehat{\mathbf{\Phi}}_{k,1}$ evaluated at the current state estimates and thus the global translation remains unobservable. In contrast, the rest of the block elements of (46), and specifically $\mathbf{\Gamma}_2$ do not adhere to the structure shown in (48) and as a result the rank of the observability matrix $\widehat{\mathbf{M}}$ corresponding to the EKF linearized VINS model increases by one. In particular, it can be easily verified that the right nullspace $\widehat{\mathbf{N}}_1$ of $\widehat{\mathbf{M}}$ does not contain the direction corresponding to the global rotation about the $\mathbf{g}$ vector, which becomes (erroneously) observable. We conjecture that this causes the EKF estimator to become inconsistent and propose a formal approach for addressing this issue in the following section.

V. OC-VINS: Algorithm Description

In order to address the EKF VINS inconsistency problem, we must ensure that (51) is satisfied for every block row of $\widehat{\mathbf{M}}$ when the state estimates are used for computing $\widehat{\mathbf{H}}_k$, and $\widehat{\mathbf{\Phi}}_{k,1}$, $\forall k > 0$, i.e., we must ensure that

$$\widehat{\mathbf{H}}_k \widehat{\mathbf{\Phi}}_{k,1} \widehat{\mathbf{N}}_1 = \mathbf{0}, \quad \forall k > 0 \tag{53}$$

One way to enforce this is by requiring that at each time step, $\widehat{\mathbf{\Phi}}_k$ and $\widehat{\mathbf{H}}_k$ satisfy the following constraints:

$$\widehat{\mathbf{N}}_{k+1} = \widehat{\mathbf{\Phi}}_k \widehat{\mathbf{N}}_k \tag{54a}$$

$$\widehat{\mathbf{H}}_k \widehat{\mathbf{N}}_k = \mathbf{0}, \quad \forall k > 0 \tag{54b}$$

where $\widehat{\mathbf{N}}_k$, $k > 0$ is computed analytically (see (56) and [33]). This can be accomplished by appropriately modifying $\widehat{\mathbf{\Phi}}_k$ and $\widehat{\mathbf{H}}_k$.

In particular, rather than changing the linearization points explicitly (e.g., as in [23]), we maintain the nullspace, $\widehat{\mathbf{N}}_k$, at each time step, and use it to enforce the unobservable directions. This has the benefit of allowing us to linearize with the most accurate state estimates, hence reducing the linearization error, while still explicitly adhering to the system observability properties.

A. Nullspace initialization

The initial nullspace is analytically defined as [see (52)]:

$$\widehat{\mathbf{N}}_1 = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^I\hat{\bar{q}}_{G,0|0}\right){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & -\lfloor {}^G\hat{\mathbf{v}}_{I,0|0} \times \rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{p}}_{I,0|0} \times \rfloor {}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{f}}_{1,0|0} \times \rfloor {}^G\mathbf{g} \end{bmatrix}. \tag{55}$$

At subsequent time steps, the nullspace is augmented to include sub-blocks corresponding to each new PF in the filter state, i.e.,

$$\widehat{\mathbf{N}}_k = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & -\lfloor {}^G\hat{\mathbf{v}}_{I,k|k-1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{p}}_{I,k|k-1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{f}}_{1,k|k-\ell} \times \rfloor {}^G\mathbf{g} \\ \vdots & \vdots \\ \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{f}}_{N,k|k-\ell'} \times \rfloor {}^G\mathbf{g} \end{bmatrix}. \tag{56}$$

where the sub-blocks $\widehat{\mathbf{N}}_{\mathbf{f}_i} = \begin{bmatrix} \mathbf{I}_3 & -\lfloor {}^G\hat{\mathbf{f}}_{i,k|k-\ell} \times \rfloor {}^G\mathbf{g} \end{bmatrix}$, are the rows corresponding to the i-th feature in the map, which are a function of the feature estimate at the time-step when it was initialized ($k - \ell$).

B. Modification of the state transition matrix $\mathbf{\Phi}$

During the covariance propagation step, we must ensure that $\widehat{\mathbf{N}}_{k+1} = \widehat{\mathbf{\Phi}}_k \widehat{\mathbf{N}}_k$. We note that the constraint on $\widehat{\mathbf{N}}_{t,k}$ is automatically satisfied due to the structure of $\widehat{\mathbf{\Phi}}_k$ (see [33]), so we focus on $\widehat{\mathbf{N}}_{r,k}$. Specifically, we rewrite (54a) element-wise as:[5]

$$\widehat{\mathbf{N}}_{r,k+1} = \widehat{\mathbf{\Phi}}_k \widehat{\mathbf{N}}_{r,k} \quad \Rightarrow \tag{57}$$

$$\begin{bmatrix} \mathbf{C}\left({}^I\hat{\bar{q}}_{G,k+1|k}\right){}^G\mathbf{g} \\ \mathbf{0}_3 \\ -\lfloor {}^G\hat{\mathbf{v}}_{I,k+1|k} \times \rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 \\ -\lfloor {}^G\hat{\mathbf{p}}_{I,k+1|k} \times \rfloor {}^G\mathbf{g} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{\Phi}}_{11} & \widehat{\mathbf{\Phi}}_{12} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \widehat{\mathbf{\Phi}}_{31} & \widehat{\mathbf{\Phi}}_{32} & \mathbf{I}_3 & \widehat{\mathbf{\Phi}}_{34} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \widehat{\mathbf{\Phi}}_{51} & \widehat{\mathbf{\Phi}}_{52} & \delta t \mathbf{I}_3 & \widehat{\mathbf{\Phi}}_{54} & \mathbf{I}_3 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \\ \mathbf{0}_3 \\ -\lfloor {}^G\hat{\mathbf{v}}_{I,k|k-1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 \\ -\lfloor {}^G\hat{\mathbf{p}}_{I,k|k-1} \times \rfloor {}^G\mathbf{g} \end{bmatrix} \tag{58}$$

and collect the constraints resulting from each block row of the above vector. Specifically, from the first block row we have

$$\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k+1|k}\right){}^G\mathbf{g} = \widehat{\mathbf{\Phi}}_{11}\mathbf{C}\left({}^I\hat{\bar{q}}_{G,k|k-1}\right){}^G\mathbf{g} \tag{59}$$

$$\Rightarrow \widehat{\mathbf{\Phi}}_{11}^{\star} = \mathbf{C}\left({}^{I,k+1|k}_{\quad\ I,k|k-1}\hat{\bar{q}}\right). \tag{60}$$

[5] Note that due to the structure of the matrices $\mathbf{\Phi}_k$ [see (42)] and $\mathbf{N}_{r,k}$ [see (56)], we only need to consider the first five block elements of (57) while the equality for the remaining ones, i.e., the elements corresponding to the features, are automatically satisfied.

The requirements for the third and fifth block rows are

$$\widehat{\mathbf{\Phi}}_{31}\mathbf{C}\big({}^{I}\hat{\bar{q}}_{G,k|k-1}\big)\,{}^{G}\mathbf{g} = \lfloor {}^{G}\hat{\mathbf{v}}_{I,k|k-1} - {}^{G}\hat{\mathbf{v}}_{I,k+1|k}\times\rfloor\,{}^{G}\mathbf{g} \quad (61)$$

$$\widehat{\mathbf{\Phi}}_{51}\mathbf{C}\big({}^{I}\hat{\bar{q}}_{G,k|k-1}\big)\,{}^{G}\mathbf{g} = \lfloor \delta t\,{}^{G}\hat{\mathbf{v}}_{I,k|k-1} - {}^{G}\hat{\mathbf{p}}_{I,k+1|k}\times\rfloor\,{}^{G}\mathbf{g} \quad (62)$$

both of which are in the form $\mathbf{Au}=\mathbf{w}$, where $\mathbf{u}$ and $\mathbf{w}$ are nullspace vector elements that are fixed. In order to ensure that (61) and (62) are satisfied, we seek to find a perturbed $\mathbf{A}^\star$, for $\mathbf{A}=\mathbf{\Phi}_{31}$ and $\mathbf{A}=\mathbf{\Phi}_{51}$ that fulfills the constraint. To compute the minimum perturbation, $\mathbf{A}^\star$, of $\mathbf{A}$, we formulate the following minimization problem

$$\min_{\mathbf{A}^\star} ||\mathbf{A}^\star-\mathbf{A}||_{\mathcal{F}}^2 \quad , \quad \text{s.t. } \mathbf{A}^\star\mathbf{u}=\mathbf{w} \quad (63)$$

where $||\cdot||_{\mathcal{F}}$ denotes the Frobenius matrix norm. After employing the method of Lagrange multipliers, and solving the corresponding KKT optimality conditions, the optimal $\mathbf{A}^\star$ that fulfills (63) is

$$\mathbf{A}^\star = \mathbf{A} - (\mathbf{Au}-\mathbf{w})(\mathbf{u}^T\mathbf{u})^{-1}\mathbf{u}^T. \quad (64)$$

Once we have computed the modified $\widehat{\mathbf{\Phi}}_{11}^\star$ from (60), and $\widehat{\mathbf{\Phi}}_{31}^\star$ and $\widehat{\mathbf{\Phi}}_{51}^\star$ from (63) and (64), we update the corresponding elements of $\widehat{\mathbf{\Phi}}_k$, and proceed with the covariance propagation (see Sect. III-A).

C. Modification of $\mathbf{H}$

During each update step, we seek to satisfy $\widehat{\mathbf{H}}_k\widehat{\mathbf{N}}_k=\mathbf{0}$ [see (54b)]. In turn, this means that

$$\widehat{\mathbf{H}}_k\widehat{\mathbf{N}}_{t,k}=\mathbf{0} \quad (65)$$

$$\widehat{\mathbf{H}}_k\widehat{\mathbf{N}}_{r,k}=\mathbf{0} \quad (66)$$

must both hold. Expressing (65) for a single point we have [see (27) and (52)]

$$\widehat{\mathbf{H}}_c\begin{bmatrix}\widehat{\mathbf{H}}_{\bar{q}} & \mathbf{0}_{3\times9} & \widehat{\mathbf{H}}_{\mathbf{p}} & | & \widehat{\mathbf{H}}_{\mathbf{f}}\end{bmatrix}\begin{bmatrix}\mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{0}_3\\ \mathbf{I}_3\\ \mathbf{I}_3\end{bmatrix}=\mathbf{0} \quad (67)$$

which is satisfied automatically, since $\widehat{\mathbf{H}}_{\mathbf{p}}=-\widehat{\mathbf{H}}_{\mathbf{f}}$ [see (30) and (31)]. Hence, the nullspace direction corresponding to translation is not violated.

Expanding the second constraint (66), we have

$$\widehat{\mathbf{H}}_c\begin{bmatrix}\widehat{\mathbf{H}}_{\bar{q}} & \mathbf{0}_{3\times9} & \widehat{\mathbf{H}}_{\mathbf{p}} & | & \widehat{\mathbf{H}}_{\mathbf{f}}\end{bmatrix}\begin{bmatrix}\mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g}\\ \mathbf{0}_3\\ -\lfloor {}^{G}\hat{\mathbf{v}}_{I,k|k-1}\times\rfloor\,{}^{G}\mathbf{g}\\ \mathbf{0}_3\\ -\lfloor {}^{G}\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor\,{}^{G}\mathbf{g}\\ -\lfloor {}^{G}\hat{\mathbf{f}}_{k|k-\ell}\times\rfloor\,{}^{G}\mathbf{g}\end{bmatrix}=\mathbf{0} \quad (68)$$

Since $\widehat{\mathbf{H}}_{\mathbf{p}}=-\widehat{\mathbf{H}}_{\mathbf{f}}$, (68) is equivalent to satisfying the following relationship

$$\begin{bmatrix}\widehat{\mathbf{H}}_c\widehat{\mathbf{H}}_{\bar{q}} & \widehat{\mathbf{H}}_c\widehat{\mathbf{H}}_{\mathbf{p}}\end{bmatrix}\begin{bmatrix}\mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g}\\ \lfloor {}^{G}\hat{\mathbf{f}}_{k|k-\ell}-{}^{G}\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor\,{}^{G}\mathbf{g}\end{bmatrix}=0$$

$$\begin{bmatrix}\widehat{\mathbf{H}}_{c\bar{q}} & \widehat{\mathbf{H}}_{c\mathbf{p}}\end{bmatrix}\begin{bmatrix}\mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g}\\ \lfloor {}^{G}\hat{\mathbf{f}}_{k|k-\ell}-{}^{G}\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor\,{}^{G}\mathbf{g}\end{bmatrix}=0 \quad (69)$$

where we have implicitly defined $\widehat{\mathbf{H}}_{c\bar{q}}$ and $\widehat{\mathbf{H}}_{c\mathbf{p}}$ as elements of the Jacobian. This is a constraint of the form $\mathbf{Au}=\mathbf{0}$, where $\mathbf{u}$ is a function of the nullspace elements, and hence is fixed, while $\mathbf{A}$ comprises block elements of the measurement Jacobian. We compute the optimal $\mathbf{A}^\star$ that satisfies (69) using (63) and (64). By direct application of (63), for

$$\mathbf{A}=\begin{bmatrix}\widehat{\mathbf{H}}_{c\bar{q}} & \widehat{\mathbf{H}}_{c\mathbf{p}}\end{bmatrix}$$

$$\mathbf{u}=\begin{bmatrix}\mathbf{C}\left({}^{I}\hat{\bar{q}}_{G,k|k-1}\right){}^{G}\mathbf{g}\\ \lfloor {}^{G}\hat{\mathbf{f}}_{k|k-\ell}-{}^{G}\hat{\mathbf{p}}_{I,k|k-1}\times\rfloor\,{}^{G}\mathbf{g}\end{bmatrix}$$

$$\mathbf{w}=\mathbf{0} \quad (70)$$

$\mathbf{A}^\star$ is computed as

$$\mathbf{A}^\star=\mathbf{A}-\mathbf{Au}(\mathbf{u}^T\mathbf{u})^{-1}\mathbf{u}^T \quad (71)$$

After computing the optimal $\mathbf{A}^\star$, we recover the Jacobian elements as

$$\widehat{\mathbf{H}}_{c\bar{q}}^\star=\mathbf{A}^\star_{1:2,1:3} \quad (72)$$

$$\widehat{\mathbf{H}}_{c\mathbf{p}}^\star=\mathbf{A}^\star_{1:2,4:6} \quad (73)$$

$$\widehat{\mathbf{H}}_{c\mathbf{f}}^\star=-\widehat{\mathbf{H}}_{c\mathbf{p}}^\star \quad (74)$$

where the subscripts (i:j, m:n) denote the submatrix spanning rows i to j, and columns m to n. Hence the modified observation matrix is

$$\widehat{\mathbf{H}}_k^\star=\begin{bmatrix}\widehat{\mathbf{H}}_{c\bar{q}}^\star & \mathbf{0}_{2\times9} & \widehat{\mathbf{H}}_{c\mathbf{p}}^\star & \widehat{\mathbf{H}}_{c\mathbf{f}}^\star\end{bmatrix} \quad (75)$$

Having computed the modified measurement Jacobian, we proceed with the filter update as described in Sect. III-B. By following this process, we ensure that the EKF does not gain information along the unobservable directions of the system. An overview of the OC-VINS modified EKF estimator is presented in Alg. 1.

VI. Simulations

We conducted Monte-Carlo simulations to evaluate the impact of the proposed Observability-Constrained VINS (OC-VINS) method on estimator consistency. We applied the proposed methodology to two VINS systems: (i) Visual Simultaneous Localization and Mapping (V-SLAM) (see Sect. VI-A), and (ii) the Multi-state Constraint Kalman Filter (MSC-KF), which performs visual-inertial localization without constructing a map (see Sect. VI-B).

A. Simulation 1: Application of the proposed framework to V-SLAM

In this section, we present the results of applying our proposed OC-VINS to V-SLAM, which we term OC-V-SLAM. We compared its performance to the standard V-SLAM (Std-V-SLAM), as well as the ideal V-SLAM that linearizes about the true state[6]. Specifically, we computed the Root Mean Squared Error (RMSE) and Normalized Estimation Error Squared (NEES) over 20 trials in which the camera-IMU platform traversed a circular trajectory of radius 5 m at an average

[6]Since the ideal V-SLAM has access to the true state, it is not realizable in practice, but we include it here as a baseline comparison.

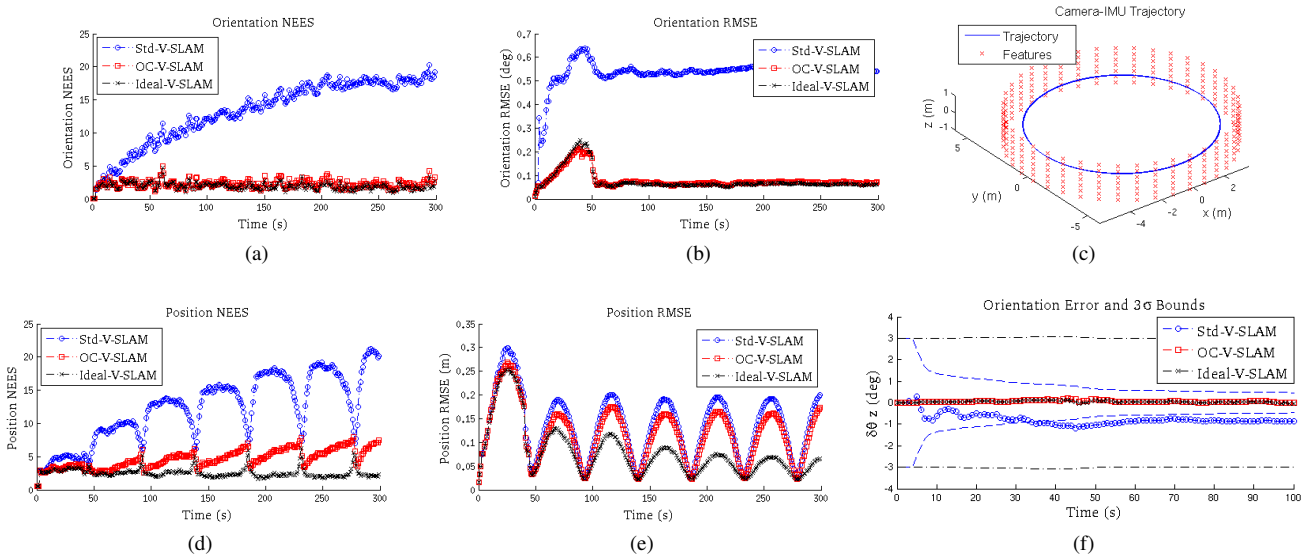


Fig. 2. Simulation 1: The RMSE and NEES errors for orientation (a)-(b) and position (d)-(e) plotted for all three filters, averaged per time step over 20 Monte Carlo trials. (c) Camera-IMU trajectory and 3D features. (f) Error and 3σ bounds for the rotation about the gravity vector, plotted for the first 100 sec of a representative run.

velocity of 60 cm/s. The camera[7] observed visual features distributed on the interior wall of a circumscribing cylinder with radius 6 m and height 2 m (see Fig. 2c). The effect of inconsistency during a single run is depicted in Fig. 2f. The error and corresponding 3σ bounds of uncertainty are plotted for the rotation about the gravity vector. It is clear that the Std-V-SLAM gains spurious information, hence reducing its 3σ bounds of uncertainty, while the Ideal-V-SLAM and the OC-V-SLAM do not. The Std-V-SLAM becomes inconsistent on this run as the orientation errors fall outside of the uncertainty bounds, while both the Ideal-V-SLAM and the OC-V-SLAM remain consistent. Figure 2 also displays the RMSE and NEES plots, in which we observe that the OC-V-SLAM attains orientation accuracy and consistency levels similar to the Ideal-V-SLAM, while significantly outperforming the Std-V-SLAM. Similarly, the OC-V-SLAM obtains better positioning accuracy compared to the Std-V-SLAM.

B. Simulation 2: Application of the proposed framework to MSC-KF

We applied our OC-VINS methodology to the MSC-KF, which we term the OC-MSC-KF. In the MSC-KF framework, all the measurements to a given OF are incorporated during a single update step of the filter, after which each OF is marginalized. Hence, in the OC-MSC-KF, we do not maintain the sub-blocks of the nullspace corresponding to the features [i.e., $\mathbf{N}_{\mathbf{f}_i}$, $i = 1, \ldots, N$, see (56)]. Instead, we propagate forward only the portion of the nullspace corresponding to the sensor platform state, and we form the feature nullspace block for each feature, only when it is processed in an update.

We conducted Monte-Carlo simulations to evaluate the consistency of the proposed method applied to the MSC-KF [11]. Specifically, we compared the standard MSC-KF (Std-MSC-KF), with the Observability-Constrained MSC-KF (OC-MSC-KF), which is obtained by applying the methodology described in Sect. V, as well as the Ideal-MSC-KF, whose Jacobians are linearized at the true states, which we use as a benchmark. We evaluated the RMSE and NEES over 30 trials (see Fig. 3) in which the camera-IMU platform traversed a circular trajectory of radius 5 m at an average speed of 60 cm/s, and observed 50 randomly distributed features per image. As evident, the OC-MSC-KF outperforms the Std-MSC-KF and attains performance almost indistinguishable from the Ideal-MSC-KF in terms of RMSE and NEES.

[7]The camera had 45 deg field of view, with $\sigma_{px} = 1px$, while the IMU was modeled after MEMS quality sensors.

VII. Experimental Results

The proposed OC-VINS framework has been validated experimentally and compared with standard VINS approaches. Specifically, we evaluated the performance of OC-V-SLAM (Sect. VII-B) and OC-MSC-KF (Sect. VII-C and Sect. VII-D) on both indoor and outdoor datasets. In our experimental setup, we utilized a light-weight sensing platform comprised of an InterSense NavChip IMU and a PointGrey Chameleon camera (see Fig. 4). During the indoor experimental tests (see Sect. VII-B and Sect. VII-C), the sensing platform was mounted on an Ascending Technologies Pelican quadrotor equipped with a VersaLogic Core 2 Duo single board computer. For the outdoor dataset, the sensing platform was head-mounted on a bicycle helmet (see Sect. VII-D), and interfaced to a handheld Sony Vaio. We hereafter provide an overview of the system implementation, along with a discussion of the experimental setup and results.

A. Implementation remarks

The image processing is separated into two components: one for extracting and tracking short-term opportunistic features (OFs), and one for extracting persistent features (PFs) to use in V-SLAM.

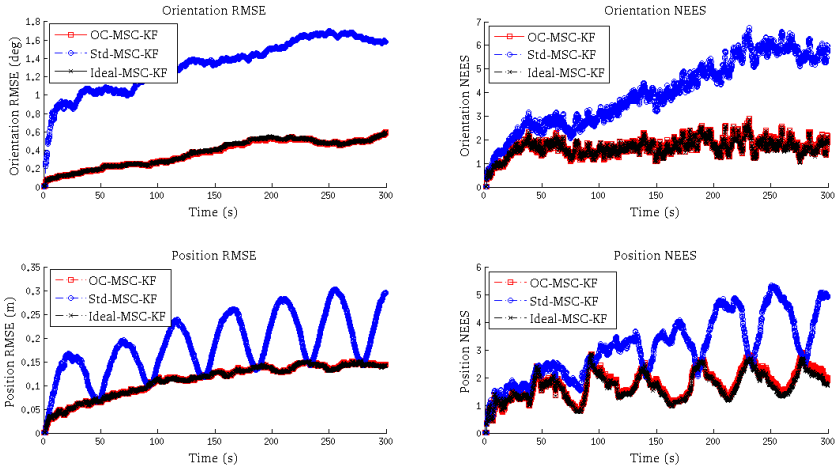
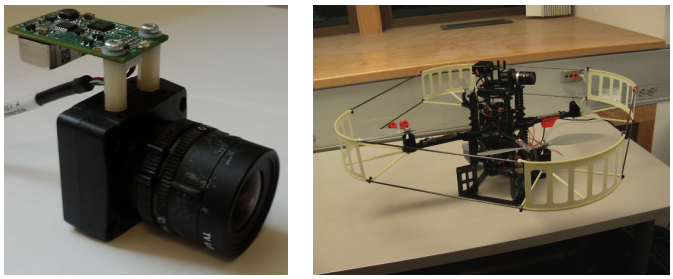


Fig. 3. Simulation 2: The average RMSE and NEES over 30 Monte-Carlo simulation trials for orientation (above) and position (below). Note that the OC-MSC-KF attains performance almost indistinguishable to the Ideal-MSC-KF.



(a) (b)

Fig. 4. (a) The experimental testbed comprises a light-weight InterSense NavChip IMU and a Point Grey Chameleon Camera. IMU signals are sampled at a frequency of 100Hz while camera images are acquired at 7.5Hz. The dimensions of the sensing package are approximately 6 cm tall, by 5 cm wide, by 8 cm deep. (b) An AscTech Pelican on which the camera-IMU package was mounted during the indoor experiments (see Sect. VII-B and Sect. VII-C).

OFs are extracted from images using the Shi-Tomasi corner detector [39]. After acquiring image k, it is inserted into a sliding window buffer of m images, $\{k-m+1, k-m+2, \ldots, k\}$. We then extract features from the first image in the window and track them pairwise through the window using the KLT tracking algorithm [31]. To remove outliers from the resulting tracks, we use a two-point algorithm to find the essential matrix between successive frames. Specifically, given the filter's estimated rotation (frome the gyroscopes' measurements) between image i and j, ${}^i\hat{\bar{q}}_j$, we estimate the essential matrix from only two feature correspondences. This approach is more robust than the five-point algorithm [40] because it provides two solutions for the essential matrix rather than up to ten. Moreover, it requires only two data points, and thus it reaches a consensus with fewer hypotheses when used in a RANSAC framework.

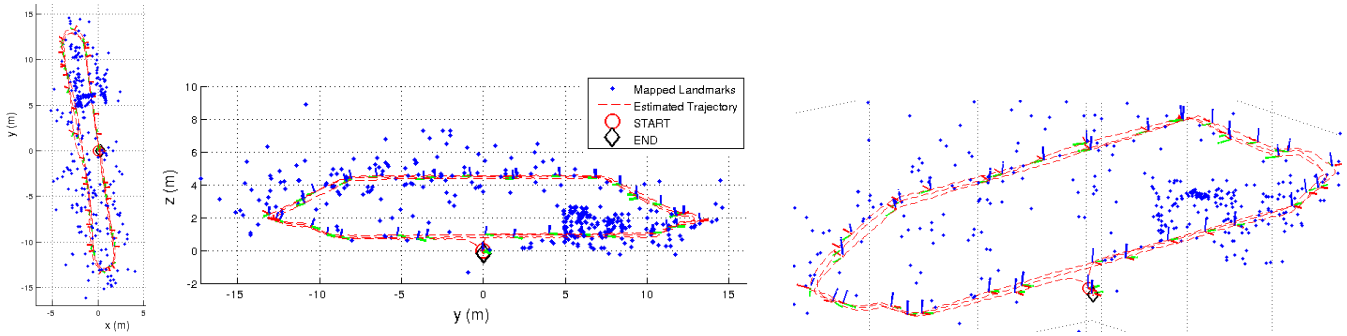
The PFs are extracted using SIFT descriptors [32]. To identify global features observed from several different images, we first utilize a vocabulary tree (VT) structure for image matching [41]. Specifically, for an image taken at time k, the VT is used to select which image(s) taken at times $1, 2, \ldots, k-1$ correspond to the same physical scene. Among those images that the VT reports as potential matches, the SIFT descriptors from each of them are compared to those from image k to create tentative feature correspondences. The epipolar constraint is then enforced using RANSAC and Nister's five-point algorithm [40] to eliminate outliers. It is important to note that the images used to construct the VT (offline) are not taken along our experimental trajectory, but rather are randomly selected from a set of representative images.

B. Experiment 1: Indoor validation of OC-V-SLAM

In the first experimental trial, we compared the performance of OC-V-SLAM to that of Std-V-SLAM on an indoor trajectory. The sensing platform traveled a total distance of 172.5 m, covering three loops over two floors in Walter Library at the University of Minnesota. The quadrotor was returned to its starting location at the end of the trajectory, to provide a quantitative characterization of the achieved accuracy.

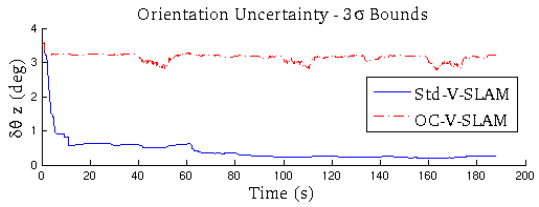
Opportunistic features were tracked using a window of $m = 10$ images. Every m camera frames, up to 30 features from all available PFs are initialized and the state vector is augmented with their 3D coordinates. The process of initializing PFs [33] is continued until the occurrence of the first loop closure; from that point on, no new PFs are considered and the filter relies upon the re-observation of previously initialized PFs and the processing of OFs.

For both the Std-V-SLAM and the OC-V-SLAM, the final position error was approximately 34 cm, which is less than 0.2% of the total distance traveled (see Fig. 5). However, the estimated covariances from the Std-V-SLAM are smaller than those from the OC-V-SLAM (see Fig. 6). Furthermore, uncertainty estimates from the Std-V-SLAM decreased in directions that are unobservable (i.e., rotations about the gravity vector); this violates the observability properties of the system and demonstrates that spurious information is injected to the filter.

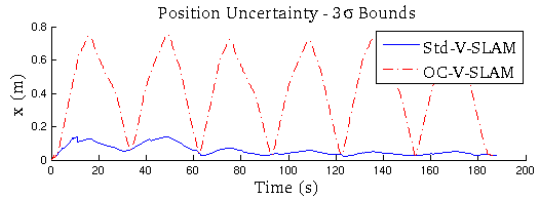


(a) (b) (c)

Fig. 5. Experiment 1: The estimated 3D trajectory over the three traversals of the two floors of the building, along with the estimated positions of the persistent features. (a) projection on the x and y axis, (b) projection on the y and z axis, (c) 3D view of the overall trajectory and the estimated features.

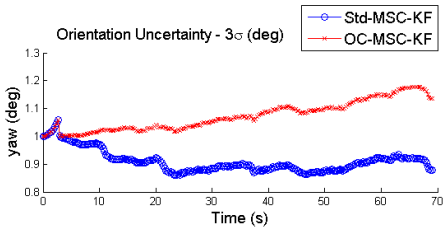


(a)

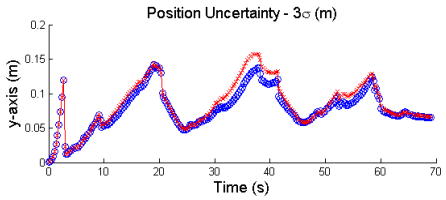


(b)

Fig. 6. Experiment 1: Comparison of the estimated 3σ error bounds for attitude and position between Std-V-SLAM and OC-V-SLAM.



(a)



(b)

Fig. 7. Experiment 2: The position (a) and orientation (b) uncertainties (3σ bounds) for the yaw angle and the y-axis, which demonstrate that the Std-MSC-KF gains spurious information about its orientation.

Figure 6(a) highlights the difference in estimated yaw uncertainty between the OC-V-SLAM and the Std-V-SLAM. In contrast to the OC-V-SLAM, the Std-V-SLAM covariance rapidly decreases, violating the observability properties of the system. Similarly, large differences can be seen in the covariance estimates for the x-axis position estimates (see Fig. 6(b)). The Std-V-SLAM estimates a much smaller uncertainty than the OC-V-SLAM, supporting the claim that the Std-V-SLAM tends to be inconsistent.

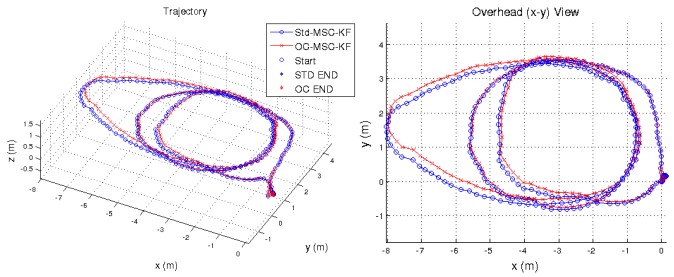
C. Experiment 2: Indoor validation of OC-MSC-KF

We validated the proposed OC-MSC-KF on real-world data. The first test comprised a trajectory 50 m in length that covered three loops in an indoor area, after which the testbed was returned to its initial position. At the end of the trajectory, the Std-MSC-KF had a position error of 18.73 cm, while the final error for the OC-MSC-KF was 16.39 cm (approx. 0.38% and 0.33% of the distance traveled, respectively). In order to assess the impact of inconsistency on the orientation estimates of both methods, we used as ground truth the rotation between the first and last images computed independently using Batch Least-Squares (BLS) and feature point matches. The Std-MSC-KF had final orientation error $\begin{bmatrix}0.15 & -0.23 & -5.13\end{bmatrix}$ deg for roll, pitch, and yaw (rpy), while the rpy errors for the OC-MSC-KF were $\begin{bmatrix}0.19 & -0.20 & -1.32\end{bmatrix}$ deg, respectively.

In addition to achieving higher accuracy, for yaw in particular, the OC-MSC-KF is more conservative since it strictly adheres to the unobservable directions of the system. This is evident in both the position and orientation uncertainties. We plot the y-axis position and yaw angle uncertainties in Fig. 7, as representative results. Most notably, the yaw uncertainty of the OC-MSC-KF remains approximately 1.13 deg (3σ), while for the Std-MSC-KF it reduces to 0.87 deg (3σ). This indicates that the Std-MSC-KF gains spurious orientation information, which leads to inconsistency. Lastly, in Fig. 8 we show the 3D trajectory along with an overhead (x-y) view. It is evident that the Std-MSC-KF yaw error impacts the position accuracy, as the Std-MSC-KF trajectory exhibits a rotation with respect to the OC-MSC-KF.

D. Experiment 3: Outdoor validation of OC-MSC-KF

In our final experimental trial, we tested the OC-MSC-KF on a large outdoor dataset (approx. 1.5 km in length). Fig. 9a depicts the OC-MSC-KF (red) and the Std-MSC-KF (blue) trajectory estimates, along with position markers from



(a) (b)

Fig. 8. Experiment 2: The 3D trajectory (a) and corresponding overhead (x-y) view (b).

a low-grade onboard GPS receiver (green). In order to assess the accuracy of both filters, the estimates are overlaid on an overhead image taken from Google-Earth.

Fig. 9b depicts a zoomed-in plot of the starting location (center) for both filters, along with the final position estimates. In order to evaluate the accuracy of the proposed method, the sensing platform was returned to its starting location at the end of the trajectory. The OC-MSC-KF obtains a final position error of 4.38 m (approx. 0.3% of the distance travelled), while the Std-MSC-KF obtains a final position error of 10.97 m. This represents an improvement in performance of approximately 60%.

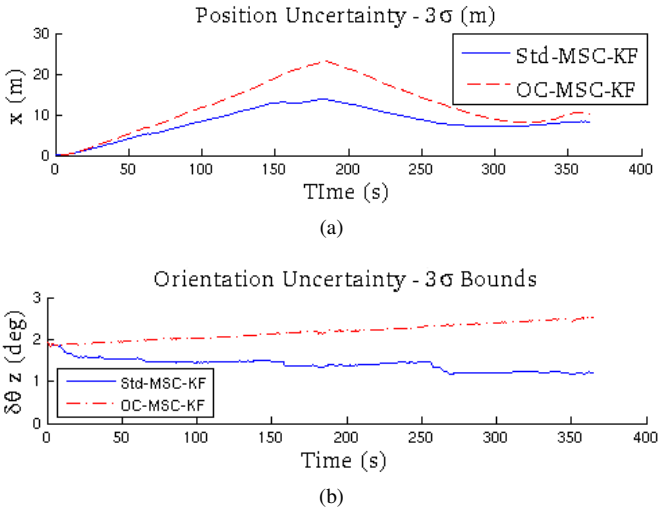
The filters' performance is also illustrated visually in Fig. 9c which shows a zoomed-in plot of the turn-around point. The OC-MSC-KF estimates remain on the light-brown portion of the ground (which is the sidewalk), which coincides with the true trajectory. In contrast, the Std-MSC-KF estimates drift over the dark triangles in the image, which are wading pools filled with water. This shifting of the trajectory represents a slight rotation around the vertical axis, indicating a violation of the rotation nullspace direction $\mathbf{N}_r$.

Figure 10 depicts the uncertainty in the position estimates along the x-axis (perpendicular to the direction of motion), along with the uncertainty in yaw (corresponding to rotations about the gravity vector). It is clear that the Std-MSC-KF reduces its uncertainty in its heading direction, indicating that the filter gains spurious information, while the OC-MSC-KF does not gain information for the rotation around the gravity vector.

VIII. Conclusion and Future Work

In this paper, we analyzed the inconsistency of VINS from the standpoint of observability. Specifically, we showed that standard EKF-based filtering approaches lead to spurious information gain since they do not adhere to the unobservable directions of the true system. Furthermore, we introduced an observability-constrained VINS approach to mitigate estimator inconsistency by enforcing the nullspace explicitly. We presented extensive simulation and experimental results to support our claims and validated the proposed estimator, by applying it to both V-SLAM and the MSC-KF.

In our future work, we are interested in analyzing the additional sources of estimator inconsistency in V-INS such as the existence of multiple local minima.



(a)

(b)

Fig. 10. Experiment 3: (a) Position uncertainty along the x-axis (perpendicular to the direction of motion) for the Std-MSC-KF, and OC-MSC-KF respectively. The OC-MSC-KF maintains more conservative estimates for position, indicating that the Std-MSC-KF may be inconsistent. (b) Orientation uncertainty about the vertical axis (z). Since rotations about gravity are unobservable, the Std-MSC-KF should not gain any information in this direction. However, as evident from this plot, the Std-MSC-KF uncertainty reduces, indicating inconsistency. For the OC-MSC-KF, the uncertainty does not decrease, indicating that the OC-MSC-KF respects the unobservable system directions.

Appendix A

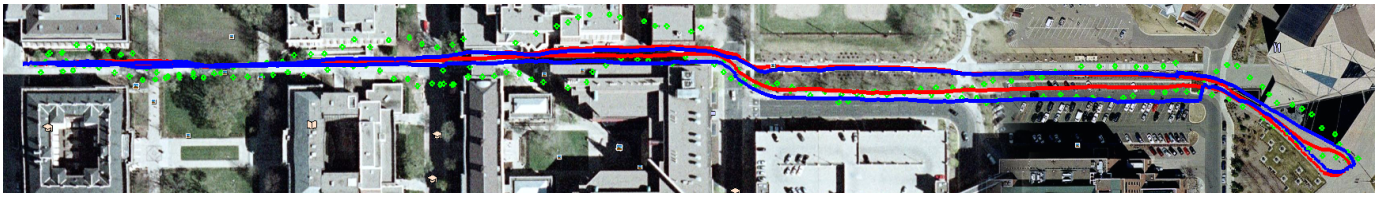
As the camera-IMU platform moves into new environments, new features must be added into the map. This entails intersecting the bearing measurements from multiple camera observations to obtain an initial estimate of each new feature's 3D location, as well as computing the initial covariance and cross-correlation between the new landmark estimate and the state. We solve this as a minimization problem over a parameter vector $\mathbf{x} = \begin{bmatrix}\mathbf{x}_{s,1}^T & \cdots & \mathbf{x}_{s,m}^T & | & \mathbf{f}^T\end{bmatrix}^T$, where $\mathbf{x}_{s,i}$, $i = 1 \ldots m$, are the m camera poses which the new landmark, $\mathbf{f}$, was observed from. Specifically, we minimize

$$C(\mathbf{x}) = \frac{1}{2}\{(\mathbf{x}-\hat{\mathbf{x}})^T \begin{bmatrix}\mathbf{P}_{ss}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} (\mathbf{x}-\hat{\mathbf{x}}) + \sum_i (\mathbf{z}_i - h(\mathbf{x}))^T \mathbf{R}_i^{-1} (\mathbf{z}_i - h(\mathbf{x}))\} \quad (76)$$

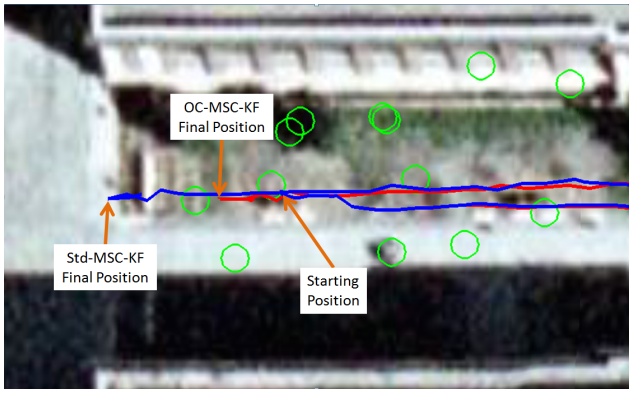
where $\mathbf{P}_{ss}^{-1}$ is the information matrix (prior) of the state estimates across all poses obtained from the filter[8], and we have no initial information about the feature location (denoted by the block (2,2) element of the prior information being equal to zero). The m measurements $\mathbf{z}_i$, $i = 1 \ldots m$ are the perspective projection observations of the point [see (25)].

We obtain an initial guess for the landmark location using any intersection method, and then we iteratively minimize (76). At each iteration, we need to solve the following

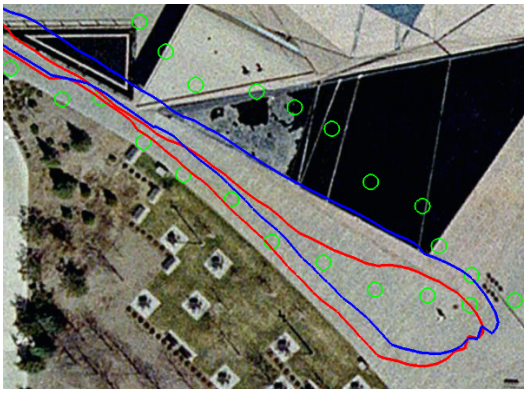
[8]We employ stochastic cloning over m time steps to ensure that the cross-correlation between the camera poses are properly accounted for [42].



(a)



(b)



(c)

Fig. 9. Experiment 3: (a) An outdoor experimental trajectory covering 1.5 km across the University of Minnesota campus. The red (blue) line denotes the OC-MSC-KF (Std-MSC-KF) estimated trajectory. The green circles denote a low-quality GPS-based estimate of the position across the trajectory. (b) A zoom-in view of the beginning / end of the run. Both filters start with the same initial pose estimate, however, the error for the Std-MSC-KF at the end of the run is 10.97 m, while for the OC-MSC-KF the final error is 4.38 m (an improvement of approx. 60%). Furthermore, the final error for the OC-MSC-KF is approximately 0.3% of the distance traveled. (c) A zoomed-in view of the turn-around point. The Std-MSC-KF trajectory is shifted compared to the OC-MSC-KF, which remains on the path (light-brown region).

linear system of equations

$$\begin{bmatrix} \mathbf{P}_{ss}^{-1} + \mathbf{H}_s^T\mathbf{R}^{-1}\mathbf{H}_s & \mathbf{H}_s^T\mathbf{R}^{-1}\mathbf{H}_f \\ \mathbf{H}_f^T\mathbf{R}^{-1}\mathbf{H}_s & \mathbf{H}_f^T\mathbf{R}^{-1}\mathbf{H}_f \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_s \\ \tilde{\mathbf{x}}_f \end{bmatrix} = \begin{bmatrix} \mathbf{H}_s^T\mathbf{R}^{-1} \\ \mathbf{H}_f^T\mathbf{R}^{-1} \end{bmatrix} \tilde{\mathbf{z}}$$
$$\leftrightarrow \quad \begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{C} \end{bmatrix} \tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} \tilde{\mathbf{z}} \qquad (77)$$

Applying the Sherman-Morrison-Woodbury matrix identity, we solve the system by inverting the matrix on the left-hand side as

$$\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Upsilon}_1 & \boldsymbol{\Upsilon}_2 \\ \boldsymbol{\Upsilon}_3 & \boldsymbol{\Upsilon}_4 \end{bmatrix} \qquad (78)$$

where

$$\begin{aligned} \boldsymbol{\Upsilon}_1 &= \left(\mathbf{A} - \mathbf{U}\mathbf{C}^{-1}\mathbf{V}\right)^{-1} \\ &= \mathbf{P}_{ss} - \mathbf{P}_{ss}\mathbf{H}_s^T \\ &\quad \cdot \{\mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{H}_f \left(\mathbf{H}_f^T\mathbf{M}^{-1}\mathbf{H}_f\right)^{-1} \mathbf{H}_f^T\mathbf{M}^{-1}\}\mathbf{H}_s\mathbf{P}_{ss} \end{aligned} \qquad (79)$$

$$\begin{aligned} \boldsymbol{\Upsilon}_2 = \boldsymbol{\Upsilon}_3^T &= -\left(\mathbf{A} - \mathbf{U}\mathbf{C}^{-1}\mathbf{V}\right)^{-1}\mathbf{U}\mathbf{C}^{-1} \\ &= -\mathbf{P}_{ss}\mathbf{H}_s^T\mathbf{M}^{-1}\mathbf{H}_f \left(\mathbf{H}_f^T\mathbf{M}^{-1}\mathbf{H}_f\right)^{-1} \end{aligned} \qquad (80)$$

$$\begin{aligned} \boldsymbol{\Upsilon}_4 &= \mathbf{C}^{-1}\mathbf{V}\left(\mathbf{A} - \mathbf{U}\mathbf{C}^{-1}\mathbf{V}\right)^{-1}\mathbf{U}\mathbf{C}^{-1} + \mathbf{C}^{-1} \\ &= \left(\mathbf{H}_f^T\mathbf{M}^{-1}\mathbf{H}_f\right)^{-1}. \end{aligned} \qquad (81)$$

Here, $\mathbf{M} = \mathbf{H}_s\mathbf{P}_{ss}\mathbf{H}_s^T + \mathbf{R}$. During each iteration, the parameter vector is updated as

$$\mathbf{x}^{\oplus} = \mathbf{x}^{\ominus} + \begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{C} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} \tilde{\mathbf{z}}. \qquad (82)$$

After the minimization process converges, we compute the posterior covariance of the new state (including the initialized feature) as

$$\mathbf{P}^{\oplus} = \begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{C} \end{bmatrix}^{-1} \qquad (83)$$

where each element is defined from (79)-(80).

ACKNOWLEDGMENT

This work was supported by the University of Minnesota through the Digital Technology Center (DTC), and the Air Force Office of Scientific Research MURI FA9550- 10-1-0567. J. A. Hesch was supported by the UMN Doctoral Dissertation Fellowship.

REFERENCES

[1] P. Corke, J. Lobo, and J. Dias, "An introduction to inertial and visual sensing," *Int. Journal of Robotics Research*, vol. 26, no. 6, pp. 519–535, Jun. 2007.

[2] J. A. Hesch, F. M. Mirzaei, G. L. Mariottini, and S. I. Roumeliotis, "A Laser-aided Inertial Navigation System (L-INS) for human localization in unknown indoor environments," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Anchorage, AK, May 3–8, 2010, pp. 5376–5382.

[3] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained MAV," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 20–25.

[4] B. Williams, N. Hudson, B. Tweddle, R. Brockers, and L. Matthies, "Feature and pose constrained visual aided inertial navigation for computationally constrained aerial vehicles," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 431–438.

[5] S. Weiss and R. Siegwart, "Real-time metric state estimation for modular vision-inertial systems," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 4531–4537.

[6] S. Weiss, D. Scaramuzza, and R. Siegwart, "Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments," *Journal of Field Robotics*, vol. 28, no. 6, pp. 854–874, Nov./Dec. 2011.

[7] S. Weiss, M. W. Achtelik, M. Chli, and R. Siegwart, "Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, St. Paul, MN, May 14–18, 2012.

[8] J. Kim and S. Sukkarieh, "Real-time implementation of airborne inertial-SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 62–71, Jan. 2007.

[9] A. I. Mourikis and S. I. Roumeliotis, "A dual-layer estimator architecture for long-term localization," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, Jun. 2008, pp. 1–8.

[10] M. Bryson and S. Sukkarieh, "Observability analysis and active control for airborne SLAM," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 261–280, Jan. 2008.

[11] A. I. Mourikis, N. Trawny, S. I. Roumeliotis, A. E. Johnson, A. Ansar, and L. Matthies, "Vision-aided inertial navigation for spacecraft entry, descent, and landing," *IEEE Trans. on Robotics*, vol. 25, no. 2, pp. 264–280, Apr. 2009.

[12] S. Ebcin and M. Veth, "Tightly-coupled image-aided inertial navigation using the unscented Kalman filter," Air Force Institute of Technology, Dayton, OH, Tech. Rep., 2007.

[13] D. W. Strelow, "Motion estimation from image and inertial measurements," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, Nov. 2004.

[14] J. Durrie, T. Gerritsen, E. W. Frew, and S. Pledgie, "Vision-aided inertial navigation on an uncertain map using a particle filter," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Kobe, Japan, May 12–17, 2009, pp. 4189–4194.

[15] J. Teddy Yap, M. Li, A. I. Mourikis, and C. R. Shelton, "A particle filter for monocular vision-aided odometry," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 5663–5669.

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

[17] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Trans. on Robotics*, vol. 28, no. 1, pp. 61–76, Feb. 2012.

[18] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY: John Wiley & Sons, 2001.

[19] D. G. Kottas, J. A. Hesch, S. L. Bowman, and S. I. Roumeliotis, "On the consistency of vision-aided inertial navigation," in *Proc. of the National Academy of Engineering: German-American Frontiers of Engineering Symposium*, Potsdam, Germany, Mar. 28–31, 2012.

[20] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Towards consistent vision-aided inertial navigation," in *Proc. of the Int. Workshop on the Algorithmic Foundations of Robotics*, Cambridge, MA, Jun. 13–15, 2012.

[21] D. G. Kottas, J. A. Hesch, S. L. Bowman, and S. I. Roumeliotis, "On the consistency of vision-aided inertial navigation," in *Proc. of the Int. Symposium on Experimental Robotics*, Quebec City, Canada, Jun. 17–21, 2012.

[22] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Rome, Italy, Apr. 10–14, 2007, pp. 3565–3572.

[23] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "A first-estimates Jacobian EKF for improving SLAM consistency," in *Proc. of the Int. Symposium on Experimental Robotics*, Athens, Greece, Jul. 14–17, 2008, pp. 373–382.

[24] ——, "Observability-based rules for designing consistent EKF SLAM estimators," *Int. Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, Apr. 2010.

[25] G. P. Huang, N. Trawny, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based consistent EKF estimators for multi-robot cooperative localization," *Autonomous Robots*, vol. 30, no. 1, pp. 99–122, Jan. 2011.

[26] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 1143–1156, Oct. 2008.

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

[28] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Trans. on Automatic Control*, vol. 22, no. 5, pp. 728–740, Oct. 1977.

[29] A. Isidori, *Nonlinear Control Systems*. Springer-Verlag, 1989.

[30] A. Martinelli, "Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *IEEE Trans. on Robotics*, vol. 28, no. 1, pp. 44–60, Feb. 2012.

[31] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the Int. Joint Conf. on Artificial Intelligence*, Vancouver, B.C., Canada, Aug. 24–28, 1981, pp. 674–679.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[33] 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. 2012-001, Feb. 2012.

[34] A. Chatfield, *Fundamentals of high accuracy inertial navigation*. AIAA (American Institute of Aeronautics & Astronautics), 1997.

[35] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Tech. Rep. 2005-002, Mar. 2005.

[36] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2006. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calibdoc/

[37] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science, B. Triggs, A. Zisserman, and R. Szeliski, Eds., vol. 1883. Springer-Verlag, 2000, pp. 298–372.

[38] P. S. Maybeck, *Stochastic models, estimation, and control*. New York, NY: Academic Press, 1979, vol. I.

[39] J. Shi and C. Tomasi, "Good features to track," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Washington, DC, Jun. 27–Jul. 2, 1994, pp. 593–600.

[40] D. Nistér, "An efficient solution to the five-point relative pose problem," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Madison, WI, Jun. 16–22, 2003, pp. 195–202.

[41] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, New York, NY, Jun. 17–22, 2006, pp. 2161–2168.

[42] S. I. Roumeliotis and J. W. Burdick, "Stochastic cloning: A generalized framework for processing relative state measurements," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Washington D.C., May 11-15, 2002, pp. 1788–1795.