A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation

Faraz M. Mirzaei, *Student Member, IEEE*, and Stergios I. Roumeliotis, *Member, IEEE*

Abstract—**Vision-aided inertial navigation systems (V-INSs) can provide precise state estimates for the 3-D motion of a vehicle when no external references (e.g., GPS) are available. This is achieved by combining inertial measurements from an inertial measurement unit (IMU) with visual observations from a camera under the assumption that the rigid transformation between the two sensors is known. Errors in the IMU-camera extrinsic calibration process cause biases that reduce the estimation accuracy and can even lead to divergence of any estimator processing the measurements from both sensors. In this paper, we present an extended Kalman filter for precisely determining the unknown transformation between a camera and an IMU. Contrary to previous approaches, we explicitly account for the time correlation of the IMU measurements and provide a figure of merit (covariance) for the estimated transformation. The proposed method does not require any special hardware (such as spin table or 3-D laser scanner) except a calibration target. Furthermore, we employ the observability rank criterion based on Lie derivatives and prove that the nonlinear system describing the IMU-camera calibration process is observable. Simulation and experimental results are presented that validate the proposed method and quantify its accuracy.**

Index Terms—**Extended Kalman filter, inertial measurement unit (IMU)-camera calibration, Lie derivatives, observability of nonlinear systems, vision-aided inertial navigation.**

I. INTRODUCTION

IN RECENT years, inertial navigation systems (INSs) have been widely used for estimating the motion of vehicles moving in a 3-D space such as airplanes, helicopters, automobiles, etc. [2]. At the core of most INS lies an inertial measurement unit (IMU) that measures linear accelerations and rotational velocities. By integrating these signals in real time, an INS is capable of tracking the position, velocity, and attitude of a vehicle. This deadreckoning process, however, cannot be used over extended periods of time because the errors in the computed estimates continuously increase. This is due to the noise and biases present in the inertial measurements. For this reason, current

INSs rely on GPS in order to receive periodic corrections. In most cases, a Kalman filter estimator is used for optimally combining the IMU and GPS measurements [3]. One of the main limitations of the *GPS-aided INS* configuration is that it cannot be used when the GPS signals are not available (e.g., indoors, underground, underwater, in space, etc.), or their reliability is limited (e.g., in the vicinity of tall buildings and structures due to specular reflections and multipath error). Furthermore, high-precision GPS receivers are prohibitively expensive, and often, the acquired level of accuracy is not sufficient for certain applications (e.g., parallel parking a car within a tight space).

An alternative approach to provide corrections to an INS is via the use of visual sensors such as cameras. Cameras are small-size, light-weight, passive sensors that provide rich information for the surroundings of a vehicle at low cost. When observing a known scene, both the position and attitude of the camera can be computed [4]. Furthermore, by tracking visual features through sequences of images, the motion of the camera can be estimated [5], [6]. Cameras and IMUs are complementary in terms of accuracy and frequency response. An IMU is ideal for tracking the state of a vehicle over short periods of time when it undergoes motions with high dynamic profile. On the other hand, a camera is best suited for state estimation over longer periods of time and smoother motion profiles. Combining these two sensors to form a *vision-aided INS* (V-INS) has recently become a popular topic of research [7].

In order to fuse measurements from an IMU and a camera in a V-INS, the 6-DOF *transformation* between these two devices must be known precisely (cf. Fig. 1). Inaccuracies in the values of the IMU-camera relative pose (position and attitude) will appear as biases that will reduce the accuracy of the estimation process or even cause the estimator to diverge. In most cases, in practice, this unknown transformation is computed manually (e.g., from computer-aided design (CAD) plots) or through the use of additional sensors. For example, for the Mars Exploration Rover (MER) mission [8], a high precision 3-D laser scanner was employed to measure the location of the four corners of the IMU housing with respect to a checkerboard placed in front of the camera for calibration purposes. Although this method achieved subdegree relative attitude accuracy and less than 1 cm relative position error [9], it is prohibitive for many applications due to the high cost of the equipment (3-D laser scanner) involved. Additionally, every time one of the two sensors is removed (e.g., for service) and repositioned, the same process needs to be repeated, which requires significant time and effort. Automating this procedure will reduce the cost of
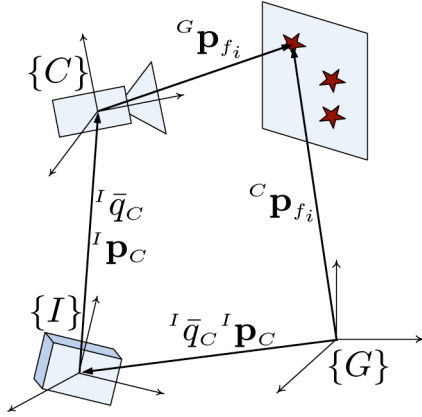
Fig. 1.   Geometric relation between the known landmarks f_i and the camera $\{C\}$, IMU $\{I\}$, and global $\{G\}$ frames of reference. The unknown IMU-camera transformation is denoted by the position and quaternion pair $({}^I\mathbf{p}_C, {}^I\bar{q}_C)$. This transformation is determined using estimates of the IMU motion, $({}^G\mathbf{p}_I, {}^I\bar{q}_G)$, the projections of the landmarks' positions, ${}^C\mathbf{p}_{f_i}$, on the camera frame (image observations), and the known positions of the landmarks, ${}^G\mathbf{p}_{f_i}$, expressed in the global frame of reference.

deploying a V-INS, increase the accuracy of the computed state estimates during regular operation, and minimize the probability of failure due to bias-induced divergence.

In this paper, we present an extended Kalman filter (EKF)-based algorithm for determining the 6-DOF transformation between a single camera and an IMU using measurements only from these two sensors [1]. Contrary to existing approaches [10], [11] that rely on modified hand–eye calibration processes (e.g., [12]–[14]), our method takes into account the *time corre-lation* of the IMU measurements by explicitly modeling them using an augmented-state EKF [15], [16]. Additionally, our algorithm computes the uncertainty in the estimated quantities, or equivalently, the *covariance*. Furthermore, we do *not* separate the task of translation estimation from rotation estimation that prevents potential error propagation. Moreover, unlike existing approaches, the described method does *not* require any special testbed except a calibration pattern that is also needed for es-timating the intrinsic parameters of the camera. Therefore, it offers the inherent capability of recalibrating the IMU-camera system as frequently as needed. Finally, a comprehensive ob-servability analysis based on Lie derivatives [17], [18] is per-formed to ensure that the sensor measurements provide suf-ficient information for accurately estimating the IMU-camera transformation.

The rest of this paper is structured as follows. Section II provides an overview of the related literature. Section III presents the proposed EKF-based algorithm, and Section IV investigates the observability of the nonlinear system describ-ing the IMU-camera calibration process. Simulation and ex-perimental results are provided in Section V, and finally, Section VI concludes the paper and suggests future research directions.

II. RELATED WORK

A well-known related process is the hand–eye calibration [12], whose objective is to estimate the 6-DOF transformation

between a camera and a robot manipulator. Recently, there have been some attempts to modify existing hand–eye calibration al-gorithms to determine the IMU-camera alignment [10], [11]. Specifically, the rotation part of the hand–eye equation is solved in [11] using nonlinear optimization software under the assump-tion that the translation between the IMU and the camera is neg-ligible. However, in most realistic situations, this assumption is *not* valid, and ignoring the translation introduces biases in estimation algorithms using these alignment parameters.

A different approach to this problem is proposed by Lobo and Dias [10], [19]. First, they obtain the vertical direction of the IMU and the camera frames by measuring the direction of gravity while viewing a vertically installed calibration pattern. Then, using Horn's method [20], they estimate the rotation be-tween the IMU and the camera. Finally, they use a spin table to rotate the system about the IMU's center and zero out the linear acceleration of the IMU due to rotation. This process al-lows one to compute the translation between the camera and the IMU based only on the camera measurements. The main drawback of this approach is that it ignores the *time correlation* between the inertial measurements due to the IMU biases. Ad-ditionally, it does *not* provide any figure of merit of the achieved level of accuracy (e.g., covariance of the estimated quantities). Furthermore, this two-stage process *decouples* the computation of rotation and translation, and hence, allows error propagation from the rotation estimates to the translation estimates. Finally, this method requires fine adjustment of the IMU-camera system on a *spin table* that limits its applicability when recalibration is frequently needed.

The IMU-camera and hand–eye calibration problems require separate treatments due to the different *noise characteristics* of the IMU and shaft-encoder signals. Specifically, while the shaft-encoder measurements at different time instants are un-correlated, consecutive IMU measurements are not. This is due to the IMU *biases*. Ignoring the time correlation of the inertial measurements limits the accuracy of the IMU-camera calibra-tion process and can lead to inconsistent estimates.

To the best of our knowledge, the presented EKF-based al-gorithm is the first approach to the IMU-camera calibration problem that does *not* ignore the correlations between the IMU measurements and requires no specialized hardware. Further-more, the uncertainty in the estimated alignment parameters is provided at every time step by computing their covariance. Fi-nally, it is shown that it suffices to rotate the camera in place in order for these parameters to become observable.

III. DESCRIPTION OF THE ALGORITHM

The IMU-camera calibration is achieved through a two-step process. First, camera images are processed in a *batch* algo-rithm to compute an *initial estimate* for the camera pose. Addi-tionally, the approximate value of the unknown transformation (e.g., hand-measured or from CAD plots) is combined with the camera-pose estimate to compute an *initial* estimate for the IMU pose (Section III-A). In the next step, both these estimates are used to initialize the corresponding variables in the EKF estima-tor. By *sequentially* processing additional measurements from

the camera and the IMU, the EKF is able to *refine* the initial estimate for the unknown transformation while simultaneously tracking the position, velocity, and attitude of the two sensors (Section III-B–III-E).

A. Filter Initialization

The purpose of this process is to determine the initial estimate for the *IMU pose* $(^G\mathbf{p}_I, {}^I\bar{q}_G)$, where $^G\mathbf{p}_I$ denotes the position of the IMU with respect to the global frame of reference and $^I\bar{q}_G$ is the rotation quaternion between the IMU and the global frames.

We first compute an estimate for the *camera pose* $(^G\mathbf{p}_C, {}^C\bar{q}_G)$ using visual features (corners of the squares in the calibration pattern) whose positions, $^G\mathbf{p}_{f_i}$, are known in global coordinates. Specifically, the initial estimates of the depth to these features are computed using Ansar's method [4], while the initial estimate for the camera pose is determined by employing Horn's method [20]. Finally, a least-squares algorithm refines the camera-pose estimate and additionally computes its covariance [21].

In the next step of the initialization process, we use an approximate estimate for the unknown IMU-camera transformation $(^I\mathbf{p}_C, {}^I\bar{q}_C)$. This was determined manually in our case but it can also be found using the CAD plots showing the IMU-camera placement. We should note that the requirement for an approximate estimate for the initial IMU-camera transformation is not limiting since it can also be determined by employing any hand–eye calibration algorithm. An initial estimate for the IMU pose is then computed from the following relations (cf. Fig. 1):

$$
\begin{aligned}
^G\mathbf{p}_I &= {}^G\mathbf{p}_C - \mathbf{C}^T(^C\bar{q}_G)\mathbf{C}^T(^I\bar{q}_C)^I\mathbf{p}_C \\
^I\bar{q}_G &= {}^I\bar{q}_C \otimes {}^C\bar{q}_G
\end{aligned}
\tag{1}
$$

where $\mathbf{C}(\bar{q})$ is the rotational matrix corresponding to quaternion $\bar{q}$ and $\otimes$ denotes quaternion multiplication. Finally, after computing the corresponding Jacobians [by linearizing (1)] and considering the uncertainty (covariance) in the estimates of $(^I\mathbf{p}_C, {}^I\bar{q}_C)$ and $(^G\mathbf{p}_C, {}^C\bar{q}_G)$, the covariance of the initial IMU pose estimate is readily found.

B. Filter Propagation

The EKF estimates the IMU pose and linear velocity as well as the unknown transformation (rotation and translation) between the camera and the IMU. Additionally, the filter estimates the biases in the IMU signals.

1) Continuous-Time System Model: We first derive the linearized continuous-time system model that describes the time evolution of the errors in the state estimates. Discretization of this model will allow us to employ the sampled measurements of the IMU for state propagation. The filter state is described by the vector

$$
\mathbf{x} = \begin{bmatrix} ^I\bar{q}_G^T & \mathbf{b}_g^T & {}^G\mathbf{v}_I^T & \mathbf{b}_a^T & {}^G\mathbf{p}_I^T & {}^I\bar{q}_C^T & {}^I\mathbf{p}_C^T \end{bmatrix}^T
\tag{2}
$$

where $^I\bar{q}_G(t)$ and $^I\bar{q}_C(t)$ are the quaternions that represent the orientation of the *global frame* and the *camera frame* in the IMU frame, respectively. The position and velocity of the IMU in the global frame are denoted by $^G\mathbf{p}_I(t)$ and $^G\mathbf{v}_I(t)$, respectively.

$^I\mathbf{p}_C(t)$ is the position of the camera in the IMU frame, and $\mathbf{b}_g$ and $\mathbf{b}_a$ are the 3×1 bias vectors affecting the gyroscope and accelerometer measurements, respectively. These biases are typically present in the signals of inertial sensors, and need to be modeled and estimated, in order to attain accurate state estimates. In our study, the IMU biases are modeled as random walk processes driven by the zero-mean white Gaussian noise vectors $\mathbf{n}_{wg}$ and $\mathbf{n}_{wa}$, respectively.

The system model describing the time evolution of the IMU state and of the IMU-camera transformation is given by the following equations [22], [23]:

$$
^I\dot{\bar{q}}_G(t) = \frac{1}{2}\mathbf{\Omega}(\boldsymbol{\omega}(t))^I\bar{q}_G(t)
\tag{3}
$$

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

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

$$
^I\dot{\bar{q}}_C(t) = \mathbf{0}_{3\times1} \qquad {}^I\dot{\mathbf{p}}_C(t) = \mathbf{0}_{3\times1}.
\tag{6}
$$

In these expressions, $\boldsymbol{\omega}(t) = [\omega_x \ \omega_y \ \omega_z]^T$ is the rotational velocity of the IMU, expressed in the IMU frame, and

$$
\mathbf{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor \omega \times \rfloor & \omega \\ -\omega^T & 0 \end{bmatrix} \quad \lfloor \omega \times \rfloor = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}
$$

Finally, $^G\mathbf{a}$ is the acceleration of the IMU, expressed in the global frame.

The gyroscope and accelerometer measurements $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$, respectively, which are employed for state propagation, are modeled as

$$
\boldsymbol{\omega}_m(t) = \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t)
\tag{7}
$$

$$
\mathbf{a}_m(t) = \mathbf{C}(^I\bar{q}_G(t))(^G\mathbf{a}(t) - {}^G\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t)
\tag{8}
$$

where $\mathbf{n}_g$ and $\mathbf{n}_a$ are zero-mean white Gaussian noise processes and $^G\mathbf{g}$ is the gravitational acceleration.

By applying the expectation operator on both sides of (3)–(6), we obtain the state estimates' propagation equations:

$$
^I\dot{\hat{\bar{q}}}_G(t) = \frac{1}{2}\mathbf{\Omega}(\hat{\boldsymbol{\omega}}(t))^I\hat{\bar{q}}_G(t)
\tag{9}
$$

$$
^G\dot{\hat{\mathbf{p}}}_I(t) = {}^G\hat{\mathbf{v}}_I(t) \qquad {}^G\dot{\hat{\mathbf{v}}}_I(t) = \mathbf{C}^T(^I\hat{\bar{q}}_G(t))\hat{\mathbf{a}}(t) + {}^G\mathbf{g}
\tag{10}
$$

$$
\dot{\hat{\mathbf{b}}}_g(t) = \mathbf{0}_{3\times1} \qquad \dot{\hat{\mathbf{b}}}_a(t) = \mathbf{0}_{3\times1}
\tag{11}
$$

$$
^I\dot{\hat{\bar{q}}}_C(t) = \mathbf{0}_{3\times1} \qquad {}^I\dot{\hat{\mathbf{p}}}_C(t) = \mathbf{0}_{3\times1}
\tag{12}
$$

with

$$
\hat{\mathbf{a}}(t) = \mathbf{a}_m(t) - \hat{\mathbf{b}}_a(t) \text{ and } \hat{\boldsymbol{\omega}}(t) = \boldsymbol{\omega}_m(t) - \hat{\mathbf{b}}_g(t).
\tag{13}
$$

The 21×1 filter error-state vector is defined as

$$
\widetilde{\mathbf{x}} = \begin{bmatrix} ^I\delta\boldsymbol{\theta}_G^T & \widetilde{\mathbf{b}}_g^T & {}^G\widetilde{\mathbf{v}}_I^T & \widetilde{\mathbf{b}}_a^T & {}^G\widetilde{\mathbf{p}}_I^T & {}^I\delta\boldsymbol{\theta}_C^T & {}^I\widetilde{\mathbf{p}}_C^T \end{bmatrix}^T
\tag{14}
$$

For the IMU and camera positions, and the IMU velocity and biases, the standard additive error definition is used (i.e., the error in the estimate $\hat{x}$ of a quantity x is $\widetilde{x} = x - \hat{x}$). However, for the quaternions, a different error definition is employed. In particular, if $\hat{\bar{q}}$ is is the estimated value of the quaternion $\bar{q}$, then

the attitude error is described by the *error quaternion*:

$$\delta \bar{q} = \bar{q} \otimes \hat{\bar{q}}^{-1} \simeq \left[\frac{1}{2} \boldsymbol{\delta\theta}^T \quad 1 \right]^T. \tag{15}$$

Intuitively, the quaternion $\delta\bar{q}$ describes the (small) rotation that causes the true and estimated attitude to coincide. The main advantage of this error definition is that it allows us to represent the attitude uncertainty by the 3×3 covariance matrix $E\{\boldsymbol{\delta\theta}\boldsymbol{\delta\theta}^T\}$. Since the attitude corresponds to 3 DOF, this is a minimal representation.

The linearized continuous-time error-state equation is

$$\dot{\tilde{\mathbf{x}}} = \mathbf{F}_c \tilde{\mathbf{x}} + \mathbf{G}_c \mathbf{n} \tag{16}$$

where

$$\mathbf{F}_c = \begin{bmatrix} -\lfloor \hat{\boldsymbol{\omega}} \times \rfloor & -\mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times9} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times9} \\ -\mathbf{C}^T({}^I\hat{\bar{q}}_G)\lfloor \hat{\mathbf{a}} \times \rfloor & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\mathbf{C}^T({}^I\hat{\bar{q}}_G) & \mathbf{0}_{3\times9} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times9} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times9} \\ \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times9} \end{bmatrix}$$

$$\mathbf{G}_c = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\mathbf{C}^T({}^I\hat{\bar{q}}_G) & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} \end{bmatrix} \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}_g \\ \mathbf{n}_{wg} \\ \mathbf{n}_a \\ \mathbf{n}_{wa} \end{bmatrix}$$

and $\mathbf{I}_3$ is the 3×3 identity matrix. The covariance $\mathbf{Q}_c$ of the system noise depends on the IMU noise characteristics and is computed offline according to [24] and [25].

2) Discrete-Time Implementation: The IMU signals $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$ are sampled at 100 Hz (i.e., T = 0.01 s). Every time a new IMU measurement is received, the state estimate is propagated using the fourth-order Runge–Kutta numerical integration of (9)–(12). In order to derive the covariance propagation equation, we evaluate the discrete-time state transition matrix

$$\boldsymbol{\Phi}_k = \boldsymbol{\Phi}(t_k + T, t_k) = \exp\left(\int_{t_k}^{t_k+T} \mathbf{F}_c(\tau)d\tau\right) \tag{17}$$

and the discrete-time system noise covariance matrix

$$\mathbf{Q}_d = \int_{t_k}^{t_k+T} \boldsymbol{\Phi}(t_{k+1}, \tau)\mathbf{G}_c\mathbf{Q}_c\mathbf{G}_c^T \boldsymbol{\Phi}^T(t_{k+1}, \tau)d\tau. \tag{18}$$

The propagated covariance is then computed as

$$\mathbf{P}_{k+1|k} = \boldsymbol{\Phi}_k\mathbf{P}_{k|k}\boldsymbol{\Phi}_k^T + \mathbf{Q}_d.$$

C. Measurement Model

The IMU camera moves continuously and records images of a calibration pattern. These are then processed to detect and identify point features whose positions ${}^G\mathbf{p}_{f_i}$ are known with respect to the global frame of reference (centered and aligned with the checker-board pattern of the calibration target). Once this process is completed for each image, a list of point features along with their measured image coordinates (u_i, v_i) is provided to the EKF, which uses them to update the state estimates.

The projective camera measurement model employed is

$$\mathbf{z}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} + \boldsymbol{\eta}_i = \begin{bmatrix} x_i/z_i \\ y_i/z_i \end{bmatrix} + \boldsymbol{\eta}_i = \mathbf{h}_i(\mathbf{x}, {}^G\mathbf{p}_{f_i}) + \boldsymbol{\eta}_i \tag{19}$$

where (cf. Fig. 1)

$${}^C\mathbf{p}_{f_i} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \mathbf{C}({}^C\bar{q}_I)\mathbf{C}({}^I\bar{q}_G)\left({}^G\mathbf{p}_{f_i} - {}^G\mathbf{p}_I\right) - \mathbf{C}({}^C\bar{q}_I)^I\mathbf{p}_C$$

and $\boldsymbol{\eta}_i$ is the feature-measurement noise with covariance $\mathbf{R}_i = \sigma_i^2\mathbf{I}_2$.

The measurement Jacobian matrix $\mathbf{H}_i$ is

$$\mathbf{H}_i = \mathbf{J}_{\text{cam}}^i \begin{bmatrix} \mathbf{J}_{\theta_G}^i & \mathbf{0}_{3\times9} & \mathbf{J}_{\mathbf{p}_I}^i & \mathbf{J}_{\theta_C}^i & \mathbf{J}_{\mathbf{p}_c}^i \end{bmatrix} \tag{20}$$

with

$$\mathbf{J}_{\text{cam}}^i = \frac{1}{\hat{z}_i^2}\begin{bmatrix} \hat{z}_i & 0 & -\hat{x}_i \\ 0 & \hat{z}_i & -\hat{y}_i \end{bmatrix} \tag{21}$$

$$\mathbf{J}_{\theta_G}^i = \mathbf{C}({}^C\hat{\bar{q}}_I)\lfloor\mathbf{C}({}^I\hat{\bar{q}}_G)({}^G\mathbf{p}_{f_i} - {}^G\hat{\mathbf{p}}_I)\times\rfloor$$

$$\mathbf{J}_{\theta_C}^i = -\mathbf{C}({}^C\hat{\bar{q}}_I)\lfloor\mathbf{C}({}^I\hat{\bar{q}}_G)({}^G\mathbf{p}_{f_i} - {}^G\hat{\mathbf{p}}_I) - {}^I\hat{\mathbf{p}}_C\times\rfloor$$

$$\mathbf{J}_{\mathbf{p}_I}^i = -\mathbf{C}({}^C\hat{\bar{q}}_I)\mathbf{C}({}^I\hat{\bar{q}}_G) \qquad \mathbf{J}_{\mathbf{p}_c}^i = -\mathbf{C}({}^C\hat{\bar{q}}_I)$$

$$\begin{bmatrix} \hat{x}_i \\ \hat{y}_i \\ \hat{z}_i \end{bmatrix} = \mathbf{C}({}^C\hat{\bar{q}}_I)\mathbf{C}({}^I\hat{\bar{q}}_G)\left({}^G\mathbf{p}_{f_i} - {}^G\hat{\mathbf{p}}_I\right) - \mathbf{C}({}^C\hat{\bar{q}}_I)^I\hat{\mathbf{p}}_C.$$

When observations to N features are available concurrently, we stack these in one measurement vector $\mathbf{z} = [\mathbf{z}_1^T \cdots \mathbf{z}_N^T]^T$ to form a single batch-form update equation. Similarly, the batch measurement Jacobian matrix is defined as $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1^T \cdots \mathbf{H}_N^T \end{bmatrix}^T$. Finally, the measurement residual is computed as

$$\mathbf{r} \stackrel{\triangle}{=} \mathbf{z} - \hat{\mathbf{z}} \simeq \mathbf{H}\tilde{\mathbf{x}} + \boldsymbol{\eta} \tag{22}$$

where $\boldsymbol{\eta} = [\boldsymbol{\eta}_1^T \cdots \boldsymbol{\eta}_N^T]^T$ is the measurement noise with covariance $\mathbf{R} = \mathbf{Diag}(\mathbf{R}_i), i = 1, \ldots, N$.

D. Iterated EKF Update

In order to increase the accuracy and numerical stability in the face of the highly nonlinear measurement model, we employ the *iterated EKF* [26], [27] to update the state. The iterative scheme proceeds as follows.

At each iteration step j, the following steps are done.
1) Compute $\hat{\mathbf{z}}^j = E\{\mathbf{z}\}$ as a function of the current jth iterate $\hat{\mathbf{x}}_{k+1|k+1}^j$ using the measurement function (19).
2) Evaluate the measurement Jacobian matrix $\mathbf{H}^j$ [cf. (20)] using the current iterate $\hat{\mathbf{x}}_{k+1|k+1}^j$.
3) Form the residual $\mathbf{r}^j = \mathbf{z} - \hat{\mathbf{z}}^j$, and compute its covariance $\mathbf{S}^j = \mathbf{H}^j\mathbf{P}_{k+1|k}\mathbf{H}^{jT} + \mathbf{R}$.
4) Using the Kalman gain $\mathbf{K}^j = \mathbf{P}_{k+1|k}\mathbf{H}^{jT}(\mathbf{S}^j)^{-1}$, compute the correction

$$\Delta\mathbf{x}^j = \mathbf{K}^j(\mathbf{r}^j + \mathbf{H}^j\Delta\mathbf{x}^{j-1}) \tag{23}$$

with $\Delta\mathbf{x}^0 = \mathbf{0}_{21\times1}$, necessary for determining the next iterate of the updated state estimate $\hat{\mathbf{x}}_{k+1|k+1}^{j+1}$.

The iteration begins using the propagated state estimate $\hat{\mathbf{x}}_{k+1|k+1}^0 = \hat{\mathbf{x}}_{k+1|k}$ as the zeroth iterate, which makes the first iteration equivalent to a regular EKF update. This process is repeated till the reduction in the cost function

$$J^j = \widetilde{\mathbf{x}}^{j^T} \mathbf{P}_{k+1|k}^{-1} \widetilde{\mathbf{x}}^j + \mathbf{r}^{j^T} \mathbf{R}^{-1} \mathbf{r}^j \qquad (24)$$

with $\widetilde{\mathbf{x}}^j \triangleq \hat{\mathbf{x}}_{k+1|k} - \hat{\mathbf{x}}_{k+1|k+1}^j$ falls below the threshold $\tau = \max(0.01, 0.001 \times J^{j-1})$, or when a maximum number of iterations is reached [28]. Finally, the covariance matrix for the current state is updated using the values for $\mathbf{K}$ and $\mathbf{S}$ from the last iteration:

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{KSK}^T. \qquad (25)$$

E. Outlier Rejection

Before using the detected features in the measurement update, we employ a Mahalanobis-distance test to detect and reject mismatches or very noisy observations. Every time a new measurement becomes available, we compute the following Mahalanobis distance:

$$\chi^2 = (\mathbf{z}_{i_k} - \hat{\mathbf{z}}_{i_k})^T \mathbf{S}_i^{-1} (\mathbf{z}_{i_k} - \hat{\mathbf{z}}_{i_k}). \qquad (26)$$

In this equation, $\mathbf{z}_{i_k}$ is the measurement of the ith landmark at time step k, $\hat{\mathbf{z}}_{i_k}$ is the expected measurement of the same landmark based on the latest state estimate, and $\mathbf{S}_i = \mathbf{H}_i \mathbf{P}_{k+1|k} \mathbf{H}_i + \mathbf{R}_i$ is the covariance of the corresponding measurement residual. A probabilistic threshold on χ^2 is used to specify whether the measurement is reliable or not. Measurements that pass this test are processed by the iterative update procedure as described before.

IV. OBSERVABILITY ANALYSIS

A system is observable if its state at a certain time instant can be uniquely determined given a finite sequence of its outputs [29]. Intuitively, this means that the measurements of an observable system provide sufficient information for estimating its state. In contrast, the state vector of unobservable systems cannot be recovered regardless of the duration of the estimation process. In dynamic systems that contain a number of *constant* but unknown quantities in their state vector (e.g., the IMU-camera 6-DOF transformation in our case), system observability results in an additional interesting property: given sufficient number of measurements, the unknown constants can be estimated with arbitrarily small uncertainty [30]. For this reason, it is important to study the observability of the system describing the IMU-camera calibration process.

The observability of *linear time-invariant* systems can be investigated by employing any of the well-known observability tests such as the rank of the Gramian matrix [15] or the Popov–Belevitch–Hautus (PBH) test [31]. However, due to the time-invariance requirement, the PBH test cannot be applied to (linearized) nonlinear systems since these are usually time varying (i.e., the current state used as the linearization point changes dynamically). On the other hand, application of the Gramian matrix criterion, which does not assume time invariance, to lin-earized systems in 2-D [32] often becomes intractable when extended to 3-D.

An alternative tool for studying the observability properties of *nonlinear* systems is the observability rank condition based on Lie derivatives [17]. Bonnifait and Garcia [33] were the first to employ this method for examining the observability of map-based bearing-only single-robot localization in 2-D. Later on, Martinelli and Siegwart [34] used Lie derivatives to analyze the observability of cooperative localization for pairs of mobile robots navigating in 2-D. In a related problem, Mariottini *et al.* [35] investigated the observability of 2-D leader-follower formations based on Lie derivatives and the observability rank condition. Recently, Lie derivatives were also used for examining the observability of the single-robot simultaneous localization and mapping (SLAM) in 2-D [36], and the camera-odometry extrinsic calibration process in 2-D [37]. However, to the best of our knowledge, there exists no similar analysis for sensors or robots navigating in 3-D. Furthermore, the observability of 3-D systems with additional unknown parameters, such as the 6-DOF IMU-camera transformation, has never been considered.

In this paper for the first time, we study the observability of the nonlinear system describing the IMU-camera calibration process. Specifically, after a brief review of the Lie derivatives and the observability rank condition (Section IV-A), in Section IV-B, we prove that the IMU-camera calibration system is *locally weakly observable* when at least two rotations about different axes are performed.

A. Nonlinear Observability

Consider the state-space representation of the following infinitely smooth nonlinear system:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \\ \mathbf{y} = \mathbf{h}(\mathbf{x}) \end{cases} \qquad (27)$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{u} = [u_1, \ldots, u_l]^T \in \mathbb{R}^l$ is the vector of control inputs, and $\mathbf{y} = [y_1, \ldots, y_m]^T \in \mathbb{R}^m$ is the measurement vector, with $y_k = h_k(\mathbf{x})$, $k = 1, \ldots, m$.

If the process function, $\mathbf{f}$, is input-linear [38], it can be separated into a summation of independent functions, each one corresponding to a different component of the control input vector. In this case, (27) can be written as

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}_0(\mathbf{x}) + \mathbf{f}_1(\mathbf{x})u_1 + \cdots + \mathbf{f}_l(\mathbf{x})u_l \\ \mathbf{y} = \mathbf{h}(\mathbf{x}) \end{cases} \qquad (28)$$

where $\mathbf{f}_0$ is the zero-input function of the process model.

The zeroth-order Lie derivative of any (scalar) function is the function itself, i.e., $\mathfrak{L}^0 h_k(\mathbf{x}) = h_k(\mathbf{x})$. The first-order Lie derivative of function $h_k(\mathbf{x})$ with respect to $\mathbf{f}_i$ is defined as

$$\mathfrak{L}_{\mathbf{f}_i}^1 h_k(\mathbf{x}) = \frac{\partial h_k(\mathbf{x})}{\partial x_1} f_{i1}(\mathbf{x}) + \cdots + \frac{\partial h_k(\mathbf{x})}{\partial x_n} f_{in}(\mathbf{x})$$
$$= \nabla h_k(\mathbf{x}) \cdot \mathbf{f}_i(\mathbf{x}) \qquad (29)$$

where $\mathbf{f}_i(\mathbf{x}) = [f_{i1}(\mathbf{x}), \ldots, f_{in}(\mathbf{x})]^T$, ∇ represents the gradient operator, and '$\cdot$' denotes the vector inner product. Considering that $\mathfrak{L}_{\mathbf{f}_i}^1 h_k(\mathbf{x})$ is a scalar function itself, the second-order Lie

derivative of $h_k(\mathbf{x})$ with respect to $\mathbf{f}_i$ is

$$\mathfrak{L}^2_{\mathbf{f}_i} h_k(\mathbf{x}) = \mathfrak{L}^1_{\mathbf{f}_i}\left(\mathfrak{L}^1_{\mathbf{f}_i} h_k(\mathbf{x})\right) = \nabla\mathfrak{L}^1_{\mathbf{f}_i} h_k(\mathbf{x}) \cdot \mathbf{f}_i(\mathbf{x}). \quad (30)$$

Higher order Lie derivatives are computed similarly. Additionally, it is possible to define mixed Lie derivatives, i.e., with respect to different functions of the process model. For example, the second-order Lie derivative of h_k with respect to $\mathbf{f}_j$ and $\mathbf{f}_i$, given its first derivative with respect to $\mathbf{f}_i$, is

$$\mathfrak{L}^2_{\mathbf{f}_j\mathbf{f}_i} h_k(\mathbf{x}) = \mathfrak{L}^1_{\mathbf{f}_j}\left(\mathfrak{L}^1_{\mathbf{f}_i} h_k(\mathbf{x})\right) = \nabla\mathfrak{L}^1_{\mathbf{f}_i} h_k(\mathbf{x}) \cdot \mathbf{f}_j(\mathbf{x}). \quad (31)$$

Based on the preceding expressions for the Lie derivatives, the observability matrix is defined as the matrix with rows:

$$\mathcal{O} \overset{\triangle}{=} \{\nabla\mathfrak{L}^\ell_{\mathbf{f}_i\cdots\mathbf{f}_j} h_k(\mathbf{x}) | i, j = 0, \ldots, l; k = 1, \ldots, m; \ell \in \mathbb{N}\}. \quad (32)$$

The important role of this matrix in the observability analysis of a nonlinear system is captured by the following proposition [17].

Proposition 1 (Observability rank condition): If the observability matrix [cf. (32)] of the nonlinear system defined in (27) is full rank, then the system is locally weakly observable.

Remark 1: Since the process and measurement functions [cf. (27)] are infinitely smooth, the observability matrix $\mathcal{O}$ can have infinite number of rows. However, to prove that $\mathcal{O}$ is full rank, it suffices to show that a subset of its rows are linearly independent.

In general, there exists no systematic method for selecting the suitable Lie derivatives and corresponding rows of $\mathcal{O}$ when examining the observability of a system. Instead, this selection is performed by sequentially considering the directions of the state space along which the gradient of each of the candidate Lie derivatives provides information.

B. Observability of IMU-Camera Calibration

In this section, we compute the observability matrix of the system describing the IMU-camera calibration process and prove that it is full rank.

First and in order to simplify the notation, we retain only few of the subscripts describing the variables in the system state vector [cf. (2)]:

$$\mathbf{x}(t) = [\,\bar{q}_I^T \quad \mathbf{b}_g^T \quad \mathbf{v}^T \quad \mathbf{b}_a^T \quad \mathbf{p}_I^T \quad \bar{q}_C^T \quad \mathbf{p}_C^T\,]^T. \quad (33)$$

Then, we rearrange the nonlinear kinematic equations (3)–(6) in a suitable format for computing the Lie derivatives:

$$\begin{bmatrix} \dot{\bar{q}}_I \\ \dot{\mathbf{b}}_g \\ \dot{\mathbf{v}} \\ \dot{\mathbf{b}}_a \\ \dot{\mathbf{p}}_I \\ \dot{\bar{q}}_C \\ \dot{\mathbf{p}}_C \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{1}{2}\mathbf{\Xi}(\bar{q}_I)\mathbf{b}_g \\ \mathbf{0}_{3\times1} \\ \mathbf{g} - \mathbf{C}^T(\bar{q}_I)\mathbf{b}_a \\ \mathbf{0}_{3\times1} \\ \mathbf{v} \\ \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times1} \end{bmatrix}}_{\mathbf{f}_0} + \underbrace{\begin{bmatrix} \frac{1}{2}\mathbf{\Xi}(\bar{q}_I) \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \end{bmatrix}}_{\underline{\mathbf{f}}_1} \boldsymbol{\omega}_m + \underbrace{\begin{bmatrix} \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{C}^T(\bar{q}_I) \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} \end{bmatrix}}_{\underline{\mathbf{f}}_2} \mathbf{a}_m \quad (34)$$

where $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$ are considered the control inputs, and

$$\mathbf{\Xi}(\bar{q}) = \begin{bmatrix} q_4\mathbf{I}_{3\times3} + \lfloor\mathbf{q}\times\rfloor \\ -\mathbf{q}^T \end{bmatrix} \text{ with } \bar{q} = \begin{bmatrix} \mathbf{q} \\ q_4 \end{bmatrix}. \quad (35)$$

Also, note that $\mathbf{f}_0$ is a 23×1 vector, while $\underline{\mathbf{f}}_1$ and $\underline{\mathbf{f}}_2$ are both compact representations of three vectors of dimension 23×1, i.e.,

$$\underline{\mathbf{f}}_1\boldsymbol{\omega}_m = \mathbf{f}_{11}\omega_{m1} + \mathbf{f}_{12}\omega_{m2} + \mathbf{f}_{13}\omega_{m3}$$

where, for $i = 1, \ldots, 3$, $\mathbf{f}_{1i}$ denotes the ith column vector comprising $\underline{\mathbf{f}}_1$ and ω_{mi} is the ith scalar component of the rotational velocity vector.

A well-known result that we will use in the observability analysis of (34) is the following: When four or more[1] known features are detected in each calibration image processed by the filter, the *camera pose* is observable [39] and can be computed in closed-form [4]. Based on this fact, we replace the measurement equation [cf. (19)] with the following pair of inferred measurements of the camera pose expressed with respect to the global frame of reference:

$$^G\bar{q}_C = \boldsymbol{\xi}_1(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) = \mathbf{h}_1^*(\mathbf{x}) = \mathbf{J}\bar{q}_I \otimes \bar{q}_C \quad (36)$$

$$^G\mathbf{p}_C = \boldsymbol{\xi}_2(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) = \mathbf{h}_2^*(\mathbf{x}) = \mathbf{p}_I + \mathbf{C}^T(\bar{q}_I)\mathbf{p}_C \quad (37)$$

where $\mathbf{C}(\bar{q}_I)$ is the rotational matrix corresponding to the quaternion $\bar{q}_I$, $\otimes$ denotes quaternion multiplication, and

$$\mathbf{J}\bar{q}_I = \bar{q}_I^{-1}, \mathbf{J} \overset{\triangle}{=} \begin{bmatrix} -\mathbf{I}_{3\times3} & 0 \\ 0 & 1 \end{bmatrix}. \quad (38)$$

At this point, we should note that the functions $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ in (36) and (37) need not to be known explicitly. Instead, what is required for the observability analysis is their functional relation with the random variables, $\bar{q}_I$ and $\mathbf{p}_I$, and the unknown parameters, $\bar{q}_C$ and $\mathbf{p}_C$, appearing in the system's state vector.

Furthermore, we enforce the unit-quaternion constraints by employing the following additional measurement equations:

$$h_3^*(\mathbf{x}) = \bar{q}_I^T \bar{q}_I - 1 = 0 \quad (39)$$

$$h_4^*(\mathbf{x}) = \bar{q}_C^T \bar{q}_C - 1 = 0. \quad (40)$$

According to Remark 1, it suffices to show that a subset of the rows of the observability matrix $\mathcal{O}$ [cf. (32)] are linearly independent. In the remaining of this section, we prove that the system described by (34) and (36)–(40) is observable by computing among the candidate zeroth-, first-, and second-order Lie derivatives of $\mathbf{h}_1^*$, $\mathbf{h}_2^*$, and h_3^*, the ones whose gradients ensure that $\mathcal{O}$ is full rank.

1) *Zeroth-Order Lie Derivatives* $(\mathfrak{L}^0\mathbf{h}_1^*, \mathfrak{L}^0\mathbf{h}_2^*, \mathfrak{L}^0 h_3^*)$: By definition, the zeroth-order Lie derivative of a function is the function itself, i.e.,

$$\mathfrak{L}^0\mathbf{h}_1^* = \mathbf{h}_1^* = \bar{q}_I^{-1} \otimes \bar{q}_C \quad (41)$$

$$\mathfrak{L}^0\mathbf{h}_2^* = \mathbf{h}_2^* = \mathbf{p}_I + \mathbf{C}^T(\bar{q}_I)\mathbf{p}_C \quad (42)$$

$$\mathfrak{L}^0 h_3^* = h_3^* = \bar{q}_I^T \bar{q}_I - 1. \quad (43)$$

---

[1]If an initial estimate of the pose is available, then observation of only three known features is sufficient for uniquely determining the camera pose [39].

Therefore, the gradients of the zeroth-order Lie derivatives are exactly the same as the Jacobians of the corresponding measurement functions:

$$\nabla \mathcal{L}^0 \mathbf{h}_1^* = [\, \mathcal{R}(\bar{q}_C)\mathbf{J} \quad \mathbf{0}_{4\times 12} \quad \mathcal{L}(\mathbf{J}\bar{q}_I) \quad \mathbf{0}_{4\times 3}\,] \tag{44}$$

$$\nabla \mathcal{L}^0 \mathbf{h}_2^* = [\, \boldsymbol{\Psi}(\bar{q}_I, \mathbf{p}_C) \quad \mathbf{0}_{3\times 9} \quad \mathbf{I}_{3\times 3} \quad \mathbf{0}_{3\times 4} \quad \mathbf{C}^T(\bar{q}_I)\,] \tag{45}$$

$$\nabla \mathcal{L}^0 h_3^* = [\, 2\bar{q}_I^T \quad \mathbf{0}_{1\times 19}\,] \tag{46}$$

where, for a quaternion $\bar{q}$ and a vector $\mathbf{p}$, we define

$$\mathcal{L}(\bar{q}) \triangleq \begin{bmatrix} q_4 \mathbf{I}_{3\times 3} - \lfloor \mathbf{q} \times \rfloor & \mathbf{q} \\ -\mathbf{q}^T & q_4 \end{bmatrix} \tag{47}$$

$$\mathcal{R}(\bar{q}) \triangleq \begin{bmatrix} q_4 \mathbf{I}_{3\times 3} + \lfloor \mathbf{q} \times \rfloor & \mathbf{q} \\ -\mathbf{q}^T & q_4 \end{bmatrix} \tag{48}$$

and

$$\boldsymbol{\Psi}(\bar{q}, \mathbf{p}) \triangleq \frac{\partial}{\partial \bar{q}} \mathbf{C}^T(\bar{q})\mathbf{p}. \tag{49}$$

Also, note that for deriving (44), we have used the following identities [23]:

$$\bar{q}_I^{-1} \otimes \bar{q}_C = \mathcal{R}(\bar{q}_C)\bar{q}_I^{-1} = \mathcal{R}(\bar{q}_C)\mathbf{J}\bar{q}_I$$
$$= \mathcal{L}(\mathbf{J}\bar{q}_I^{-1})\bar{q}_C = \mathcal{L}(\mathbf{J}\bar{q}_I)\bar{q}_C$$

2) *First-Order Lie Derivatives* ($\mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_1^*, \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_2^*, \mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^*$): The first-order Lie derivatives of $\mathbf{h}_1^*$ and $\mathbf{h}_2^*$ with respect to $\mathbf{f}_0$ are computed as (50), shown at the bottom of the page

$$\mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_1^* = \nabla \mathcal{L}^0 \mathbf{h}_1^* \cdot \mathbf{f}_0 = -\frac{1}{2}\mathcal{R}(\bar{q}_C)\mathbf{J}\boldsymbol{\Xi}(\bar{q}_I)\mathbf{b}_g \tag{51}$$

$$\mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_2^* = \nabla \mathcal{L}^0 \mathbf{h}_2^* \cdot \mathbf{f}_0 = -\frac{1}{2}\boldsymbol{\Psi}(\bar{q}_I, \mathbf{p}_C)\boldsymbol{\Xi}(\bar{q}_I)\mathbf{b}_g + \mathbf{v} \tag{52}$$

while their gradients are given by

$$\nabla \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_1^* = [\, \mathbf{X}_1 \quad -\frac{1}{2}\mathcal{R}(\bar{q}_C)\mathbf{J}\boldsymbol{\Xi}(\bar{q}_I) \quad \mathbf{0}_{4\times 9} \quad \mathbf{X}_2 \quad \mathbf{0}_{4\times 3}\,]$$

$$\nabla \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_2^* = [\, \mathbf{X}_3 \quad \mathbf{X}_4 \quad \mathbf{I}_{3\times 3} \quad \mathbf{0}_{3\times 10} \quad \mathbf{X}_5\,]. \tag{53}$$

In these last expressions, $\mathbf{X}_i, i = 1, \ldots, 5$, are matrices of appropriate dimensions (4×4 the first two, 3×4 the third one, and 3×3 the last two) which, regardless of their values, will be eliminated in the following derivations; hence, they need not be computed explicitly.

The next first-order Lie derivative of interest is that of $\mathbf{h}_2^*$ with respect to $\underline{\mathbf{f}}_1$, i.e., $\mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^*$. At this point, we remind the reader that $\underline{\mathbf{f}}_1$ as defined in (34) is a compact representation of three column vectors. Similarly, we can also write the resulting Lie derivative in a compact form (i.e., a 3×3 matrix):

$$\mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^* = \nabla \mathcal{L}^0 \mathbf{h}_2^* \cdot \underline{\mathbf{f}}_1 = \frac{1}{2}\boldsymbol{\Psi}(\bar{q}_I, \mathbf{p}_C)\boldsymbol{\Xi}(\bar{q}_I). \tag{54}$$

The gradients of the three columns of $\mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^*$ stacked together give

$$\nabla \mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^* = [\, \boldsymbol{\Gamma}(\bar{q}_I, \mathbf{p}_C) \quad \mathbf{0}_{9\times 16} \quad \boldsymbol{\Upsilon}(\bar{q}_I)\,] \tag{55}$$

where the matrices

$$\boldsymbol{\Gamma}(\bar{q}_I, \mathbf{p}_C) = \begin{bmatrix} \boldsymbol{\Gamma}_1(\bar{q}_I, \mathbf{p}_C) \\ \boldsymbol{\Gamma}_2(\bar{q}_I, \mathbf{p}_C) \\ \boldsymbol{\Gamma}_3(\bar{q}_I, \mathbf{p}_C) \end{bmatrix} \qquad \boldsymbol{\Upsilon}(\bar{q}_I) = \begin{bmatrix} \boldsymbol{\Upsilon}_1(\bar{q}_I) \\ \boldsymbol{\Upsilon}_2(\bar{q}_I) \\ \boldsymbol{\Upsilon}_3(\bar{q}_I) \end{bmatrix} \tag{56}$$

of dimensions 9×4 and 9×3, respectively, have block-row elements (for $i = 1, \ldots, 3$)

$$\boldsymbol{\Gamma}_i(\bar{q}_I, \mathbf{p}_C) = \frac{\partial}{\partial \bar{q}_I}\left[\left(\mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^*\right)\mathbf{e}_i\right]$$

$$\boldsymbol{\Upsilon}_i(\bar{q}_I) = \frac{\partial}{\partial \mathbf{p}_C}\left[\left(\mathfrak{L}_{\underline{\mathbf{f}}_1}^1 \mathbf{h}_2^*\right)\mathbf{e}_i\right]$$

with $\mathbf{e}_1 = [1 \quad 0 \quad 0]^T$, $\mathbf{e}_2 = [0 \quad 1 \quad 0]^T$, and $\mathbf{e}_3 = [0 \quad 0 \quad 1]^T$.

Note that inclusion of all the block-row elements of the gradient (55) in the observability matrix $\mathcal{O}$ [cf. (50)] implies that all components of $\boldsymbol{\omega}_m$ are nonzero. However, as it will become evident later on, in order to prove observability, *only two* of the elements of $\boldsymbol{\omega}_m$ need to be nonzero. In such case, matrix $\mathcal{O}$ will contain the block matrices

$$\boldsymbol{\Gamma}_{ij}(\bar{q}_I, \mathbf{p}_C) = \begin{bmatrix} \boldsymbol{\Gamma}_i(\bar{q}_I, \mathbf{p}_C) \\ \boldsymbol{\Gamma}_j(\bar{q}_I, \mathbf{p}_C) \end{bmatrix} \qquad \boldsymbol{\Upsilon}_{ij}(\bar{q}_I) = \begin{bmatrix} \boldsymbol{\Upsilon}_i(\bar{q}_I) \\ \boldsymbol{\Upsilon}_j(\bar{q}_I) \end{bmatrix} \tag{57}$$

with $i, j = 1, \ldots, 3$, $i \neq j$, instead of $\boldsymbol{\Gamma}(\bar{q}_I, \mathbf{p}_C)$ and $\boldsymbol{\Upsilon}(\bar{q}_I)$.

3) *Second-Order Lie Derivative* ($\mathfrak{L}_{\mathbf{f}_0}^2 \mathbf{h}_2^*$): Finally, we compute the second-order Lie derivative of $\mathbf{h}_2^*$ with respect to $\mathbf{f}_0$:

$$\mathfrak{L}_{\mathbf{f}_0}^2 \mathbf{h}_2^* = \mathfrak{L}_{\mathbf{f}_0}^1 \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_2^* = \nabla \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_2^* \cdot \mathbf{f}_0$$

$$= -\frac{1}{2}\mathbf{X}_3 \boldsymbol{\Xi}(\bar{q}_I)\mathbf{b}_g + \mathbf{g} - \mathbf{C}^T(\bar{q}_I)\mathbf{b}_a \tag{58}$$

and its gradient

$$\nabla \mathfrak{L}_{\mathbf{f}_0}^2 \mathbf{h}_2^* = [\, \mathbf{X}_6 \quad \mathbf{X}_7 \quad \mathbf{0}_{3\times 3} \quad -\mathbf{C}^T(\bar{q}_I) \quad \mathbf{0}_{3\times 7} \quad \mathbf{X}_8\,] \tag{59}$$

where the matrices $\mathbf{X}_5$, $\mathbf{X}_6$, and $\mathbf{X}_7$ (of dimensions 3×4 the first one and 3×3 the last two) will be eliminated in the ensuing derivations, and therefore, we do not need to compute them explicitly.

Stacking together all the previously computed gradients of the Lie derivatives, we form the observability matrix $\mathcal{O}$ shown in (50).

In order to prove that the system described by (34) and (36)–(40) is observable, we employ the result of Proposition 1 and

$$\mathcal{O} = \begin{bmatrix} \nabla \mathcal{L}^0 \mathbf{h}_1^* \\ \nabla \mathcal{L}^0 \mathbf{h}_2^* \\ \nabla \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_1^* \\ \nabla \mathfrak{L}_{\mathbf{f}_0}^1 \mathbf{h}_2^* \\ \nabla \mathfrak{L}_{\underline{\mathbf{f}}_{1ij}}^1 \mathbf{h}_2^* \\ \nabla \mathcal{L}^0 h_3^* \\ \nabla \mathfrak{L}_{\mathbf{f}_0}^2 \mathbf{h}_2^* \end{bmatrix} = \begin{bmatrix} \mathcal{R}(\bar{q}_C)\mathbf{J} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathcal{L}(\mathbf{J}\bar{q}_I) & \mathbf{0}_{4\times 3} \\ \boldsymbol{\Psi}(\bar{q}_I, \mathbf{p}_C) & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{C}^T(\bar{q}_I) \\ \mathbf{X}_1 & -\frac{1}{2}\mathcal{R}(\bar{q}_C)\mathbf{J}\boldsymbol{\Xi}(\bar{q}_I) & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{X}_2 & \mathbf{0}_{4\times 3} \\ \mathbf{X}_3 & \mathbf{X}_4 & \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{X}_5 \\ \boldsymbol{\Gamma}_{ij}(\bar{q}_I, \mathbf{p}_C) & \mathbf{0}_{6\times 3} & \mathbf{0}_{6\times 3} & \mathbf{0}_{6\times 3} & \mathbf{0}_{6\times 3} & \mathbf{0}_{6\times 4} & \boldsymbol{\Upsilon}_{ij}(\bar{q}_I) \\ 2\bar{q}_I^T & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 4} & \mathbf{0}_{1\times 3} \\ \mathbf{X}_6 & \mathbf{X}_7 & \mathbf{0}_{3\times 3} & -\mathbf{C}^T(\bar{q}_I) & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{X}_8 \end{bmatrix} \tag{50}$$

show that matrix $\mathcal{O}$ is full rank (i.e., the state space of the system is spanned by the gradients of the Lie derivatives of the measurement functions [17], [18]). Before presenting the main result of this section (cf. Lemma 3), we first state the following two lemmas whose proofs are detailed in Appendix I.

Lemma 1: The matrix

$$\mathbf{A}_{ij} = \begin{bmatrix} \boldsymbol{\Gamma}_{ij} & \boldsymbol{\Upsilon}_{ij} \\ 2\bar{q}_I^T & \mathbf{0}_{1\times 3} \end{bmatrix} \qquad (60)$$

formed by the fifth and sixth block-row elements of the first and last block-columns of the observability matrix $\mathcal{O}$ [cf. (50)], with $\boldsymbol{\Gamma}_{ij}$ and $\boldsymbol{\Upsilon}_{ij}$ defined in (57), is full rank.

Corollary 1: The matrix described by (60) is full rank if the IMU-camera rig is rotated about at least two different axes.

Note that only two block rows of $\boldsymbol{\Gamma}(\bar{q}_I, \mathbf{p}_C)$ and $\boldsymbol{\Upsilon}(\bar{q}_I)$ [cf. (56)]—the ones corresponding to two nonzero components of $\boldsymbol{\omega}_m$—are included in $\mathbf{A}_{ij}$ [cf. (60)]. Therefore, the third component of $\boldsymbol{\omega}_m$ can be zero (i.e., no rotation around the corresponding axis) without affecting the observability properties of the system.

Lemma 2: For any unit-quaternions $\bar{q}$ and $\bar{s}$, matrix $\mathbf{B} = \mathcal{R}(\bar{q})\mathbf{J}\boldsymbol{\Xi}(\bar{s})$ is full rank.

Lemma 3: The observability matrix $\mathcal{O}$ [cf. (50)] is full rank when the IMU-camera rig is rotated about at least two different axes.

Proof: Here, we provide a sketch of the proof based on block Gaussian elimination (for details please see [40]). We start by employing Lemma 1 and Corollary 1 to eliminate all the matrices in the first and last columns of $\mathcal{O}$. The next step is to eliminate $\mathbf{X}_2$ using $\mathcal{L}(\mathbf{J}\bar{q}_I)$, i.e., the (1,6) block element of $\mathcal{O}$ in (50). Note that $\bar{q}$ is unit quaternion and $\det(\mathcal{L}(\mathbf{J}\bar{q})) = ||\mathbf{J}\bar{q}|| = ||\bar{q}|| = 1$ [cf. (38), (40), and (47)]. Finally, since $-\frac{1}{2}\mathcal{R}(\bar{q}_C)\mathbf{J}\boldsymbol{\Xi}(\bar{q}_I)$ is full rank (cf. Lemma 2), it can be used to eliminate $\mathbf{X}_4$ and $\mathbf{X}_7$. Following these steps, $\mathcal{O}$ reduces to

$$\begin{bmatrix} \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 3} \\ \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 4} & \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{1\times 4} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 4} & \mathbf{0}_{1\times 3} \\ \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} \\ \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 3} \\ \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{I}_{3\times 3} \\ \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & -\mathbf{C}^T(\bar{q}_I) & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} & \mathbf{0}_{3\times 3} \end{bmatrix}. \qquad (61)$$

Considering that a property of a rotation matrix is that it is full rank ($\forall \bar{q}, \det(\mathbf{C}(\bar{q})) = 1$), it is easy to see that (61) is full rank, indicating that $\mathcal{O}$ is also full rank. ∎

Corollary 2: The system described by (34) and (36)–(40) is observable regardless of the linear motion of the IMU-camera rig.

This is evident from the fact that for proving Lemma 3, we did *not* use any Lie derivatives with respect to $\underline{\mathbf{f}}_2$ [cf. (34)]. Therefore, $\mathbf{a}_m$, the measured linear acceleration can take arbitrary values without compromising the observability of the system. This observation has important practical implications when no significant linear motion is possible due to physical constraints (e.g., calibration of an IMU-camera rig in an indoor
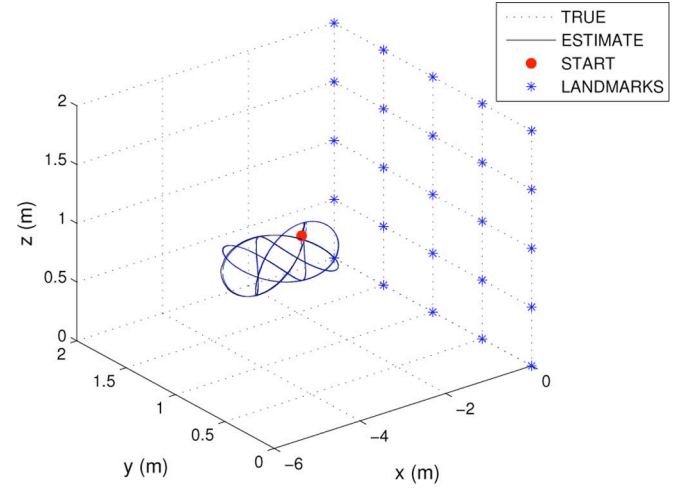

Fig. 2.   Trajectory of the IMU-camera system for 15 s.

laboratory): the IMU-camera transformation can be accurately estimated even if no linear acceleration is exerted.

Remark 2: Since no noise is injected into the system along the directions of the IMU-camera transformation [cf. (6)], regardless of the observability of the system, the uncertainty of the IMU-camera transformation will never increase.

When the linearization of the IMU-camera calibration system is sufficiently accurate, this remark has the following important implication: running the estimation algorithm during periods when the observability conditions are not met (e.g., as a result of stopping the IMU-camera rig) will not decrease the accuracy of the IMU-camera estimates, although it might not improve their quality either. However, it is advisable to excite at least two degrees of rotational freedom for sufficiently long time at the beginning of the calibration process, so as to significantly reduce the error in the IMU-camera transformation and ensure the validity of the linear approximation.

V. SIMULATION AND EXPERIMENTAL RESULTS

A. Simulation Results

In order to validate the proposed EKF algorithm for estimating the IMU-camera transformation when ground truth is available, we have performed a number of simulation experiments. In our simulation setup, an IMU-camera rig moves in front of a calibration target containing 25 known features. These correspond to the vertices of a rectangular grid with 50×50 cm cell size, which is aligned with the yz plane (cf. Fig. 2). The camera is assumed to have $50°$ field of view. Additionally, the image measurements received at a rate of 10 Hz are distorted with noise of $\sigma = 1$ pixel. The IMU noise characteristics are the same as those of the ISIS IMU used in the real-world experiments (cf. Section V-B). The IMU measurements are received at 100 Hz.

The initial alignment error for translation is set to ${}^I\widetilde{\mathbf{p}}_C = [5 - 5\ 6]^T$ cm with a standard deviation of 5 cm in each axis. The initial alignment error for rotation is set to $\delta\boldsymbol{\theta} = [4° - 4°\ 3°]^T$ [cf. (15)] with $3°$ standard deviation of uncertainty in
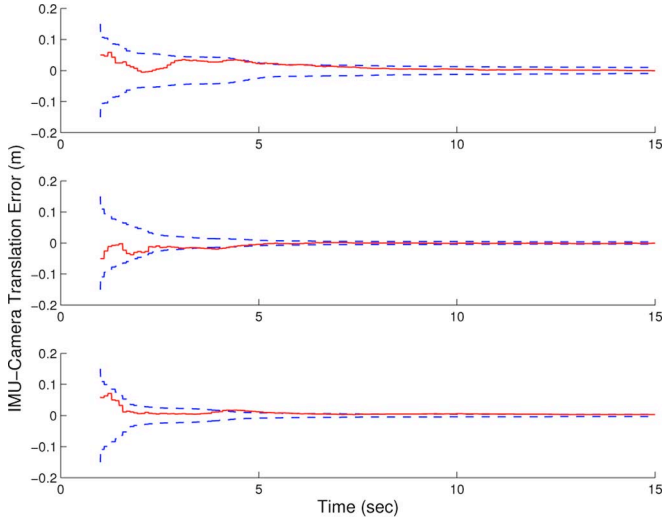
Fig. 3. State-estimate error and 3σ bounds for the IMU-camera transformation. Translation along axes x, y, and z. The initial error is $^I\widetilde{\mathbf{p}}_C = [5 \quad -5 \quad 6]^T$ cm.
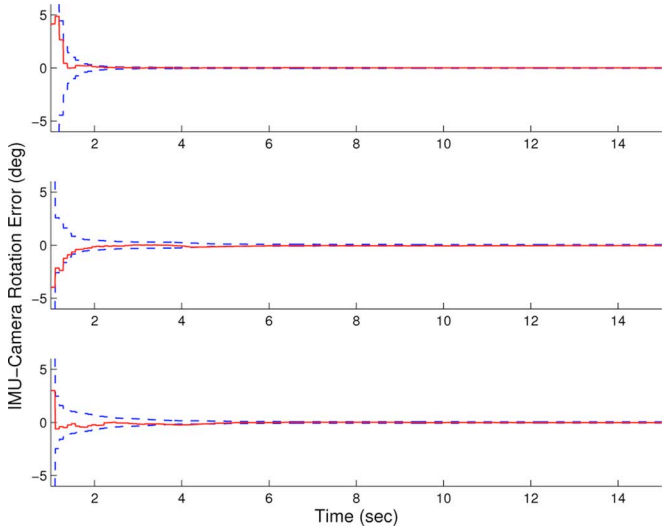

Fig. 4. State-estimate error and 3σ bounds for the IMU-camera transformation. Rotation about axes x (roll), y (pitch), and z (yaw). The initial alignment errors are $\boldsymbol{\delta\theta} = [4° \quad -4° \quad 3°]^T$.

each axis of rotation. Consequently, the filter state vector and error-state covariance matrix are initialized according to the process described in Section III-A.

Following the initialization step, the system performs a spiral motion within 3–5 m off the calibration pattern. The EKF processes the IMU and camera measurements, and concurrently estimates *all* the components of the state vector [cf. (2)]. The actual and estimated trajectories are shown in Fig. 2. For the duration of this simulation (only 15 s), 150 images were processed and, on the average, 21.7 landmarks were visible in each image. The state-estimate errors and their 3σ bounds for the 6-DOF transformation between the IMU and the camera in a typical simulation are shown in Figs. 3 and 4. As evident from these plots, even with a relatively large initial error for the IMU-camera transformation, the algorithm is still able to attain very accurate estimates of the calibration parameters. The final un-

3σ	x (cm)	y (cm)	z (cm)	r (°)	p (°)	y (°)
Initial	15	15	15	9	9	9
xyz-rpy	0.18	0.14	0.13	0.013	0.013	0.013
rpy	0.25	0.22	0.22	0.0082	0.024	0.024

xyz represents translation along the *x*-, *y*-, and *z*-axes, *rpy* indicates rotation about the local *x* (roll), *y* (pitch), and *z* (yaw) axes.

certainty (3σ) of the estimates is $[0.96 \quad 0.84 \quad 0.90]^T$ cm for translation and $[0.072° \quad 0.120° \quad 0.120°]^T$ for rotation.

1) General Motion Versus Rotation Only: In Section IV, we have shown that the system describing the IMU-camera calibration process is observable when the IMU-camera rig undergoes rotational motion even if no translation occurs. Hereafter, we examine the achievable accuracy for motions with and without translation after 100 s when the IMU-camera rig undergoes: 1) spiral motion (i.e., exciting all 6 DOF) and 2) pure rotation (i.e., exciting only the DOF corresponding to attitude). In all these simulations, the initial uncertainty of the IMU-camera translation and rotation are set to 15 cm and $9°$ deg (3σ) in each axis, respectively. A summary of these results is shown in Table I. The third row of Table I, (xyz–rpy), corresponds to motion with all 6 DOF excited. In this case, after sufficient time, the translation uncertainty is reduced to less than 2 mm (3σ) in each axis.

By comparing the results of Table I to those corresponding to Figs. 3 and 4, it is obvious that by allowing the EKF algorithm to run for longer period of time (i.e., 100 s instead of 15 s), we can estimate the calibration parameters more accurately. Additionally, as it can be seen in this particular example, the translational uncertainty along the x-axis is slightly higher than the uncertainty along the other two axes. This is a typical result observed in all simulations with similar setup. The main reason for this is the limited range of pitch and yaw rotations (i.e., about the y- and z-axes, respectively) required for keeping the landmarks within the field of view. On the other hand, the roll rotation (about the x-axis) is virtually unlimited, and it can span a complete circle without losing visual contact with the landmarks (note that the optical axis of the camera is aligned with the local x-axis).

The fourth row of Table I corresponds to a scenario where the motion is constrained to pure rotation. As expected, the system is still observable, and both the translation and the rotation between the IMU and the camera are accurately estimated [40]. The accuracy of the rotation estimation between the IMU and the camera in both scenarios (i.e., with or without translation) is shown in the last three columns of Table I. As evident, in all cases, the rotational parameters can be estimated extremely accurately, even when the system has not undergone any translation.

2) Monte Carlo Simulations: Finally, we have conducted Monte Carlo simulations to statistically evaluate the accuracy of the filter. We ran 100 simulations with a setup similar to the first simulation described in this section. The initial standard deviation of the IMU-camera transformation is set to 3 cm for translation and $3°$ for rotation. The initial values in each run are

TABLE II
MONTE CARLO SIMULATIONS: COMPARISON OF THE STANDARD DEVIATIONS
OF THE FINAL IMU-CAMERA TRANSFORMATION ERROR (σ_{err}) AND THE
AVERAGE COMPUTED UNCERTAINTY OF THE ESTIMATES (σ_{est})

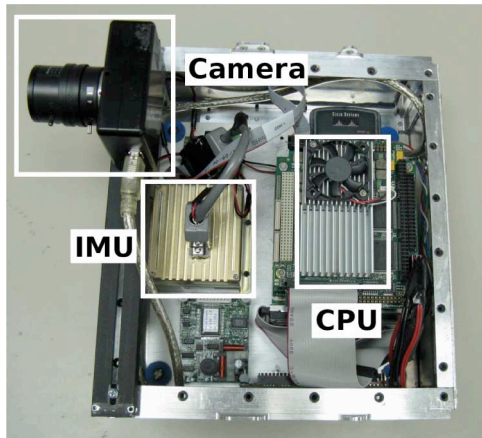|	x (cm)	y (cm)	z (cm)	r (°)	p (°)	y (°)
σ_{err}	0.29	0.23	0.28	0.019	0.036	0.039
σ_{est}	0.31	0.24	0.28	0.019	0.039	0.040

Fig. 5.   Testbed used for the experiments.

randomly generated according to a Gaussian probability distribution with these standard deviations. Each simulation is run for 15 s and the final calibration errors along with their estimated uncertainty are recorded. The ensemble mean of the recorded errors is $[0.058 \quad -0.002 \quad 0.044]^T$ cm for translation and $[-0.0038° \quad 0.0013° \quad -0.0009°]^T$ for rotation. It can be seen that the mean error is at least one order of magnitude smaller than the typical error, demonstrating that the filter is indeed unbiased.

The standard deviations of the recorded errors are shown in the second row of Table II. The third row of this table shows the average of the standard deviations computed by the EKF at each realization of the experiment. Comparison of these two rows indicates consistency of the filter as the standard deviation of the actual error is smaller than or equal to the standard deviations computed by the EKF (i.e., the filter estimates are not overconfident).

B. Experimental Results

In order to demonstrate the validity of our EKF algorithm in realistic situations, we have conducted experiments using a testbed that consists of an ISIS IMU, a firewire camera, and a PC104 computer for data acquisition (cf. Fig. 5). The IMU and the camera are rigidly mounted on the chassis and their relative pose does not change during the experiment. The intrinsic parameters of the camera were calibrated prior to the experiment [41] and are assumed constant. The camera's field of view is 60° with a focal length of 9 mm. The resolution of the images is 1024×768 pixels. Images are recorded at a rate of 3.75 Hz while the IMU provides measurements at 100 Hz. The PC104 stores the images and the IMU measurements for postprocessing using our EKF algorithm. Furthermore, considering that the exact values of the IMU-camera transformation
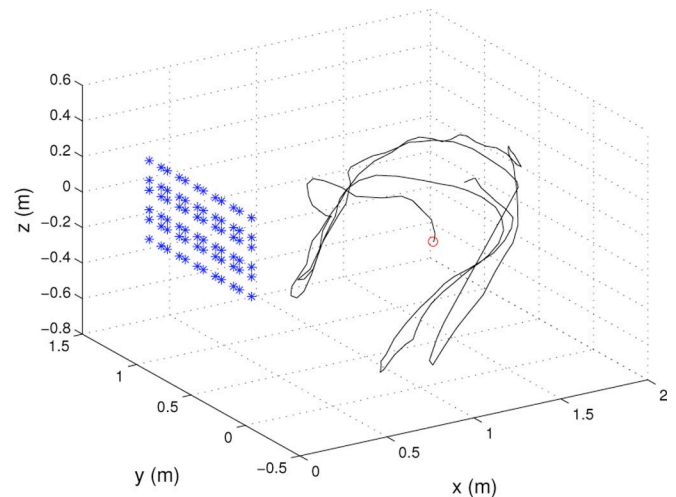

Fig. 6.   Estimated trajectory of the IMU for 50 s. The starting point is shown by a circle on the trajectory.

(ground truth) were not available in this experiment, a batch least squares (BLS) estimator was implemented to provide the *best possible estimates* of the alignment parameters by postprocessing all the collected IMU and image measurements concurrently (cf. Appendix II).

A calibration pattern (checker board) was used to provide 72 globally known landmarks that were placed 5.5–11 cm apart from each other. The bottom-left corner of this checker board was selected as the origin of the global reference frame, and the calibration pattern was aligned with the direction of the gravitational acceleration. The landmarks (i.e., the corners of the squares) were extracted using a least-squares corner detector. We have assumed that the camera measurements are corrupted by additive white Gaussian noise with standard deviation equal to 2 pixels.[2]

The hand-measured translation and rotation between the IMU and the camera was used as an initial guess for the unknown transformation. Additionally, the pose of the IMU was initialized as described in Section III-A. Finally, initialization of the gyro and the accelerometer biases was performed by placing the testbed in a static position for approximately 80 s. During this time, the EKF processed IMU and camera measurements while enforcing the static constraint (zero position and attitude displacement). The resulting state vector along with the error-state covariance matrix were then directly used to run the experiment.

Once the initialization process was complete, we started moving the testbed while the camera was facing the calibration pattern. For the duration of the experiment, the distance between the camera and the calibration pattern varied between 0.5 and 2.5 m in order to keep the corners of the checker board visible. Additionally, the testbed was moved in such a way so as to excite all degrees of freedom while at the same time keeping the landmarks within the camera's field of view.

During the motion of the testbed (~ 50 s), 180 images were recorded, of which 24 were not processed due to motion blur.

---

[2]The actual pixel noise is less than 2 pixels. However, in order to compensate for the existence of unmodeled nonlinearities and imperfect camera calibration, we have inflated the noise standard deviation to 2 pixels.
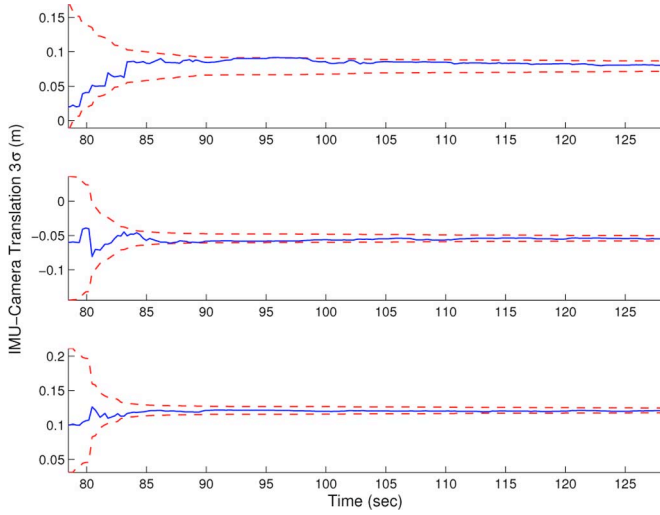
Fig. 7. Time evolution of the estimated IMU-camera translation along the x-, y-, and z-axes (solid blue lines) and the corresponding 3σ bounds centered around the BLS estimates (dashed red lines).
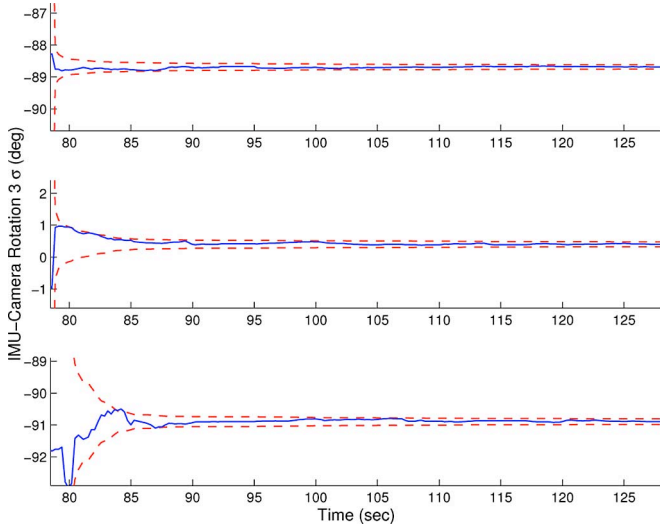

Fig. 8. Time evolution of the estimated IMU-camera rotation about the axes x, y, and z (solid blue lines), and the corresponding 3σ bounds centered around the BLS estimates (dashed red lines).

The EKF algorithm was able to estimate the IMU-camera transformation while keeping track of the IMU pose, velocity, and IMU biases. The estimated trajectory of the IMU is shown in Fig. 6.

The time evolution of the estimated calibration parameters along with their estimated 3σ bounds centered around the BLS estimates are depicted in Figs. 7 and 8. As evident from these plots, the calibration parameters converge to steady-state values after approximately 130 s (including the 80 s of the duration of the initialization process). The small inconsistencies observed during the initial transient period are due to the nonlinearities of the system and measurement models, and the imprecise initialization of the filter state vector. In particular, evaluating the Jacobians using the inaccurate state estimates available at the beginning of the calibration process causes the estimates to fluctuate significantly around their true values. As more feature

TABLE III
INITIAL, EKF, AND BLS ESTIMATES OF THE IMU-CAMERA PARAMETERS AND THEIR UNCERTAINTY FOR THE DESCRIBED EXPERIMENT

	x $\pm 3\sigma$ (cm)	y $\pm 3\sigma$ (cm)	z $\pm 3\sigma$ (cm)
Initial	2 ± 9	-6 ± 9	10 ± 9
EKF	7.93 ± 0.76	-5.49 ± 0.38	12.09 ± 0.35
BLS	7.43 ± 0.15	-5.19 ± 0.13	12.12 ± 0.10

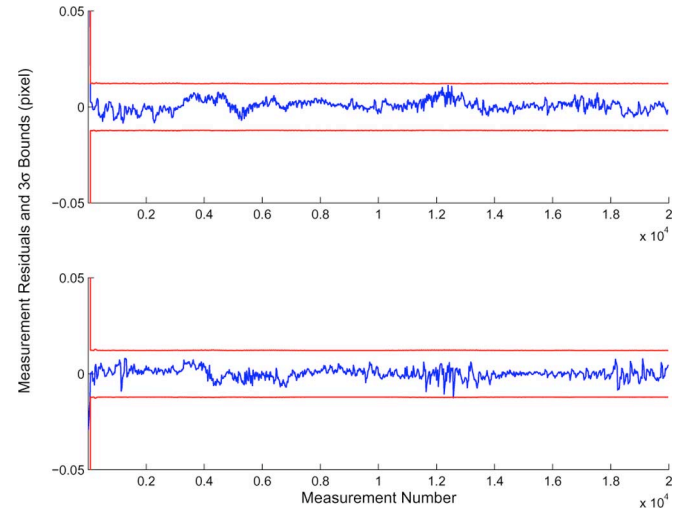|	roll $\pm 3\sigma$ (°)	pitch $\pm 3\sigma$ (°)	yaw $\pm 3\sigma$ (°)
Initial	-90 ± 6	0 ± 6	-90 ± 6
EKF	-88.69 ± 0.07	0.40 ± 0.08	-90.89 ± 0.08
BLS	-88.71 ± 0.02	0.38 ± 0.04	-90.93 ± 0.05

Fig. 9. [Calibrated IMU-Camera] Measurement residuals along with their 3σ bounds for the horizontal u (top plot) and vertical v (bottom plot) axes of the images.

observations become available, the accuracy of the state estimates improves, which subsequently increases the accuracy of the system Jacobian and eventually leads to convergence to the states' true value.

A summary of the results from this experiment is provided in Table III. It is worth mentioning that the initial uncertainty of 9 cm (3σ) in the translation parameters improves to less than 0.8 cm (3σ) for all axes. Additionally, the initial uncertainty of 6° (3σ) decreases to less than 0.1° (3σ) for each axis of rotation. Moreover, this table shows that the EKF estimator, *which can run in real time*, attains a level of accuracy close to that of the BLS. Also, note that the final accuracy of the EKF is consistent with that of the BLS, demonstrating that the EKF is not overconfident. A further indicator of the consistency of the EKF is provided in Fig. 9. As shown in these plots, the measurement residuals of the filter along the image axes (i.e., reprojection errors) lie within their estimated 3σ bounds.

In order to stress the importance of acquiring precise IMU-camera calibration estimates, we have also tested with the same experimental setup, an EKF-based estimator that does *not* estimate the calibration parameters online. Instead, this filter uses the initial guess for the unknown IMU-camera transformation to estimate the IMU pose, velocity, and biases. In this case, as
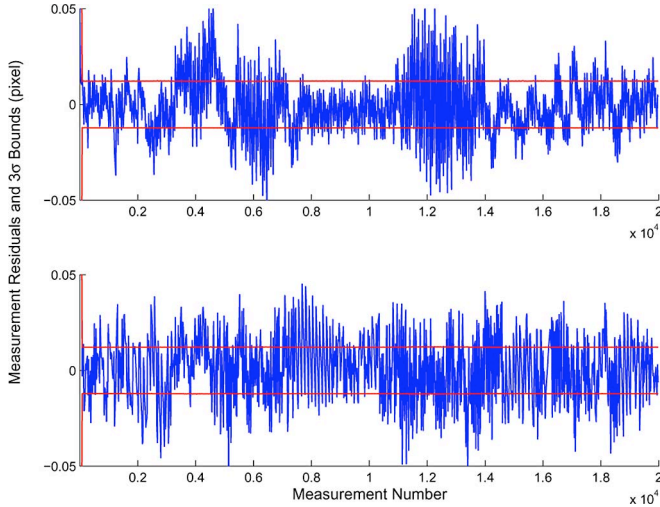
Fig. 10. [Uncalibrated IMU-Camera] Measurement residuals along with their 3σ bounds for the horizontal u (top plot) and vertical v (bottom plot) axes of the images.

evident from the camera measurement residuals shown in Fig. 10, the approximate values for the calibration parameters lead to large inconsistencies of the estimator.

VI. CONCLUSION

In this paper, we have presented an EKF-based algorithm for estimating the transformation between an IMU and a camera rigidly attached on a mobile platform. To the best of our knowledge, this is the first approach to the IMU-camera calibration problem that appropriately accounts for the *time correlation* between the IMU measurements. Additionally and contrary to previous work on this subject, we do *not* separate the task of translation estimation from rotation estimation, and hence, prevent error propagation. Moreover, by treating the problem within the Kalman filtering framework, we are also able to compute the *covariance* of the estimated quantities as an indicator of the achieved level of accuracy. Therefore, by accounting for this uncertainty in the consequent estimation algorithm, we are able to explicitly model their impact. Last but not the least, an important feature of this algorithm is the ability to perform the calibration process *without* requiring any specific testbed (such as rotating table [10] or high precision 3-D laser scanner [8]) except the calibration pattern that is also needed when calibrating the intrinsic parameters of the camera. The derived estimator was tested both in simulation and experimentally, and it was shown to achieve accuracy in the order of millimeters and sub-degrees, respectively, for the translational and rotational components of the IMU-camera transformation. Additionally and for the first time, the observability of the nonlinear system describing the IMU-camera calibration was investigated by employing the observability rank condition based on Lie derivatives. As presented, estimating the IMU-camera transformation requires exciting only two of the rotational DOF, while no translational motion is necessary.

Currently, we are investigating the possibility to extend this study to the most challenging case of the IMU-camera calibration, where instead of using a calibration pattern, we will

consider naturally occurring visual features whose positions are unknown.

APPENDIX I

PROOF OF LEMMAS 1 AND 2

Lemma 1: Matrix $\mathbf{A}_{ij}$ defined in (60) is full rank.

Proof: We prove this lemma for the case of practical interest when the elements of $\mathbf{p}_C = [p_1 \quad p_2 \quad p_3]^T$ (i.e., the vector denoting the position of the camera expressed with respect to the IMU frame) are nonzero.[3]

For $i, j = 1, \ldots, 3, i \neq j$, we expand $\mathbf{A}_{ij}$ as

$$\mathbf{A}_{ij} = \begin{bmatrix} \mathbf{\Gamma}_i & \mathbf{\Upsilon}_i \\ \mathbf{\Gamma}_j & \mathbf{\Upsilon}_j \\ 2\bar{q}_I^T & \mathbf{0}_{1\times3} \end{bmatrix} \begin{matrix} \} (1:3) \leftrightarrow \omega_{m\,i} \\ \} (4:6) \leftrightarrow \omega_{m\,j} \\ \} 7. \end{matrix} \quad (62)$$

The variables on the right side of the matrix next to the row numbers specify the component of $\boldsymbol{\omega}_m = [\omega_{m1} \ \omega_{m2} \ \omega_{m3}]^T$ that are excited in order for these rows to be included in the observability matrix $\mathcal{O}$ [cf. (50)]. After considerable algebra, it can be shown that [40]

$$\det(\mathbf{A}_{ij}) = 8(-1)^k p_k \left(p_j^2 + p_i^2\right) \quad (63)$$

where $i, j, k = 1, \ldots, 3, k \neq i, k \neq j$, and $i \neq j$. We conclude the proof by noting that since all elements of $\mathbf{p}_C$ are nonzero, the determinant of $\mathbf{A}_{ij}$ in (63) is nonzero; hence, $\mathbf{A}_{ij}$ is full rank. ∎

Lemma 2: For any unit quaternions $\bar{q} = [q_1 \ q_2 \ q_3 \ q_4]^T$ and $\bar{s} = [s_1 \ s_2 \ s_3 \ s_4]^T$, $\mathbf{B} = \mathcal{R}(\bar{q})\mathbf{J}\Xi(\bar{s})$ is full rank.

Proof: This can be readily proved by computing $\mathbf{B}^T\mathbf{B}$, which is a 3×3 matrix:

$$\mathbf{B}^T\mathbf{B} = \Xi^T(\bar{s})\mathbf{J}^T\mathcal{R}^T(\bar{q})\mathcal{R}(\bar{q})\mathbf{J}\Xi(\bar{s}) = \mathbf{I}_{3\times3}. \quad (64)$$

Therefore, matrix $\mathbf{B}$ is full rank. For computing (64), we used the identities $\mathcal{R}^T(\bar{q})\mathcal{R}(\bar{q}) = \mathbf{I}_{4\times4}$ and $\Xi^T(\bar{s})\Xi(\bar{s}) = \mathbf{I}_{3\times3}$ [23]. ∎

APPENDIX II

BUNDLE ADJUSTMENT

In order to compare the results of the proposed EKF algorithm for estimating the 6-DOF IMU-camera transformation with the best achievable (offline) estimates, we compute the batch least-squares estimate, also known as bundle adjustment [42]. For this purpose, we minimize the following linearized cost function:

$$\mathcal{C}_M^i = \frac{1}{2}\left(\bar{\mathbf{x}}_0 - \hat{\mathbf{x}}_0^i\right)^T \mathbf{P}_0^{-1}\left(\bar{\mathbf{x}}_0 - \hat{\mathbf{x}}_0^i\right)$$

$$+ \frac{1}{2}\sum_{k=0}^{M-1}\left(\widetilde{\mathbf{z}}_k - \mathbf{H}_k^i\widetilde{\mathbf{x}}_k^i\right)^T \mathbf{R}_k^{-1}\left(\widetilde{\mathbf{z}}_k - \mathbf{H}_k^i\widetilde{\mathbf{x}}_k^i\right)$$

$$+ \frac{1}{2}\sum_{k=0}^{M-1}\left(\widetilde{\mathbf{x}}_{k+1}^i - \mathbf{\Phi}_k^i\widetilde{\mathbf{x}}_k^i\right)^T \mathbf{Q}_k^{-1}\left(\widetilde{\mathbf{x}}_{k+1}^i - \mathbf{\Phi}_k^i\widetilde{\mathbf{x}}_k^i\right) \quad (65)$$

[3]Note that $\mathbf{p}_C = \mathbf{0}_{3\times1}$ is not physically realizable since it means that the centers of the IMU and the camera coincide. Also, the case when one or more elements of $\mathbf{p}_C$ are zero is extremely rare in practice, since it requires perfect position alignment of the camera and the IMU. However, the latter case is addressed in [40], where it is shown that the system is still observable when all three degrees of rotational freedom are excited.

where $\boldsymbol{\Phi}_k^i$ [cf. (17)], $\mathbf{H}_k^i$ [cf. (20)], and $\widetilde{\mathbf{z}}_k = \mathbf{r}_k$ [cf. (22)] are evaluated at $\hat{\mathbf{x}}_k^i = \hat{\mathbf{x}}_k^{i-1} + \widetilde{\mathbf{x}}_k^{i-1}$, the ith iterate of the system's state-vector estimates at time step k, $k = 0, \dots, M$ [cf. (2) and (14)]. Additionally, $\mathbf{R}_k$ represents the measurement noise covariance matrix (cf. Section III-C) and $\mathbf{Q}_k$ is the discrete-time system noise covariance matrix [cf. (18)]. Furthermore, $\bar{\mathbf{x}}_0$ and $\mathbf{P}_0$ represent the initial state estimate and its covariance, respectively. Minimizing this cost function requires solving the following system of equations iteratively:

$$\mathbf{M}\underline{\widetilde{\mathbf{x}}}^i = \boldsymbol{\epsilon} \qquad (66)$$

where $\mathbf{M}$ and $\boldsymbol{\epsilon}$ are computed as functions of $\hat{\mathbf{x}}_k^i, \mathbf{z}_k, \boldsymbol{\Phi}_k^i, \mathbf{H}_k^i, \mathbf{P}_0$, $\mathbf{R}_k, \mathbf{Q}_k$, and $\underline{\widetilde{\mathbf{x}}}^i = [\widetilde{\mathbf{x}}_0^{iT} \dots \widetilde{\mathbf{x}}_N^{iT}]^T$ [43]. In our implementation, we have initialized $\hat{\mathbf{x}}_k^0, k = 0, \dots, M$, with the results from the EKF proposed in this paper, and employed the sparse Cholesky factorization with symmetric approximate minimum degree permutation to solve (66) [44]. The resulting estimates are used as the EKF benchmarks in our experiments (cf. Section V).

REFERENCES

[1] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Diego, CA, Oct. 29–Nov. 2, 2007, pp. 2427–2434.

[2] A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*. Reston, VA: American Institute of Aeronautics and Astronautics, Inc., 1997.

[3] D. Titterton and J. Weston, Eds., *Strapdown Inertial Navigation Technology*. Piscataway, NJ: IEE, 2005.

[4] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 578–589, May 2003.

[5] A. Azarbayejani and A. Pentland, "Recursive estimation of motion, structure, and focal length," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 6, pp. 562–575, Jun. 1995.

[6] A. I. Mourikis, N. Trawny, S. I. Roumeliotis, A. E. Johnson, A. Ansar, and L. H. Matthies, "Vision-aided inertial navigation for spacecraft entry, descent, and landing," *IEEE Trans. Robot.*, in press.

[7] J. Dias, M. Vinzce, P. Corke, and J. Lobo, "Special issue: 2nd workshop on integration of vision and inertial sensors," in *The International Journal of Robotics Research*, vol. 26. Newbury Park, CA: SAGE Publications, Jun. 2007, no. 6, pp. 519–535.

[8] A. E. Johnson, R. Willson, J. Goguen, J. Alexander, and D. Meller, "Field testing of the mars exploration rovers descent image motion estimation system," in *Proc. IEEE Int. Conf. Robot. Autom.*, Barcelona, Spain, Apr. 18–22, 2005, pp. 4463–4469.

[9] A. E. Johnson, J. F. Montgomery, and L. H. Matthies, "Vision guided landing of an autonomous helicopter in hazardous terrain," in *Proc. IEEE Int. Conf. Robot. Autom.*, Barcelona, Spain, Apr. 18–22, 2005, pp. 3966–3971.

[10] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," presented at the InerVis, Barcelona, Spain, Apr. 2005.

[11] P. Lang and A. Pinz, "Callibration of hybrid vision/inertial tracking systems," presented at the InerVis, Barcelona, Spain, Apr. 2005.

[12] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3-D robotics hand/eye calibration," *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 345–358, Jun. 1989.

[13] J. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," *Int. J. Robot. Res.*, vol. 10, no. 3, pp. 240–254, Jun. 1991.

[14] K. Daniilidis, "Hand–eye calibration using dual quaternions," *Int. J. Robot. Res.*, vol. 18, no. 3, pp. 286–298, Mar. 1999.

[15] P. S. Maybeck, *Stochastic Models, Estimation and Control.* vol. 1. New York: Academic Press, 1979.

[16] E. M. Foxlin, "Generalized architecture for simultaneous localization, auto-calibration, and map-building," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Lauzanne, Switzerland, Sep. 30–Oct. 4, 2002, pp. 527–533.

[17] R. Hermann and A. Krener, "Nonlinear controlability and observability," *IEEE Trans. Autom. Control*, vol. AC-22, no. 5, pp. 728–740, Oct. 1977.

[18] S. Sastry, *Nonlinear Systems: Analysis, Stability, and Control.* New York: Springer-Verlag, 1999.

[19] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," *Int. J. Robot. Res.*, vol. 26, no. 6, pp. 561–575, Jun. 2007.

[20] B. K. P. Horn, "Closed-form solution of the absolute orientation using quaternions," *J. Opt. Soc. Amer. A*, vol. 4, pp. 629–642, Apr. 1987.

[21] R. Haralick, "Pose estimation from corresponding point data," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1426–1446, Nov./Dec. 1989.

[22] E. J. Lefferts, F. L. Markley, and M. D. Shuster, "Kalman filtering for spacecraft attitude estimation," *J. Guid., Control, Dyn.*, vol. 5, no. 5, pp. 417–429, Sep./Oct. 1982.

[23] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3-D pose estimation," Dept. Comput. Sci. & Eng., Univ. Minnesota, Minneapolis, MN, Tech. Rep., Mar. 2005.

[24] *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Laser Gyros*, The Institute of Electrical and Electronics Engineers, New York, IEEE Standard 647-2006.

[25] *IEEE Standard Specification Format Guide and Test Procedure for Linear, Singleaxis, Nongyroscopic Accelerometers*, The Institute of Electrical and Electronics Engineers, New York, IEEE Standard 1293-1998, 1999.

[26] P. S. Maybeck, *Stochastic Models, Estimation and Control, vol. 2.* New York: Academic Press, 1979.

[27] A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (series Mathematics in Science and Engineering). vol. 64, New York: Academic Press, 1970.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C.* Cambridge, MA: Cambridge Univ. Press, 1992.

[29] W. L. Brogan, *Modern Control Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1990.

[30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ: Prentice-Hall, 1993.

[31] W. J. Rugh, *Linear System Theory*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[32] J. Andrade-Cetto and A. Sanfeliu, "The effects of partial observability when building fully correlated maps," *IEEE Trans. Robot.*, vol. 21, no. 4, pp. 771–777, Aug. 2005.

[33] P. Bonnifait and G. Garcia, "Design and experimental validation of an odometric and goniometric localization system for outdoor robot vehicles," *IEEE Trans. Robot. Autom.*, vol. 14, no. 4, pp. 541–548, Aug. 1998.

[34] A. Martinelli and R. Siegwart, "Observability analysis for mobile robot localization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Edmonton, Canada, Aug. 2–6, 2005, pp. 1471–1476.

[35] G. L. Mariottini, G. Pappas, D. Prattichizzo, and K. Daniilidis, "Vision-based localization of leader follower formations," in *Proc. IEEE Conf. Decision and Control*, Seville, Spain, Dec. 12–15, 2005, pp. 635–640.

[36] K. W. Lee, W. S. Wijesoma, and J. I. Guzman, "On the observability and observability analysis of SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Beijing, China, Oct. 9–15, 2006, pp. 3569–3574.

[37] A. Martinelli, D. Scaramuzza, and R. Siegwart, "Automatic self-calibration of a vision system during robot motion," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 15–19, 2006, pp. 43–48.

[38] H. Nijmeijer and A. van der Schaft, *Nonlinear Dynamical Control Systems.* New York: Springer-Verlag, 1990.

[39] D. Sun and J. L. Crassidis, "Observability analysis of six-degree-of-freedom configuration determination using vector observations," *J. Guid., Control, Dyn.*, vol. 25, no. 6, pp. 1149–1157, Nov. 2002.

[40] F. M. Mirzaei and S. I. Roumeliotis. (2007, Jan.) IMU-camera calibration: Observability analysis. Dept. Comput. Sci. & Eng., Univ. Minnesota, Minneapolis, Tech. Rep. [Online]. Available: http://mars.cs.umn.edu/tr/reports/Mirzaei07.pdf.

[41] J.-Y. Bouguet. (2006). *Camera calibration toolbox for matlab.* [Online]. Available: http://www.vision.caltech.edu/bouguetj/calibdoc/

[42] B. Triggs, P. McLauchlan, R. Hartley, and Fitzgibbon, "Bundle adjustment for structure from motion," in *Vision Algorithms: Theory & Practice*, B. Triggs, A. Zisserman, and R. Szeliski, Eds. New York: Springer-Verlag, 2000.

[43] F. M. Mirzaei and S. I. Roumeliotis. (2007, Aug.). IMU-camera calibration: Bundle adjustment. Dept. Comput. Sci. & Eng., Univ. Minnesota, Minneapolis, Tech. Rep. [Online]. Available: http://mars.cs.umn.edu/tr/reports/Mirzaei07a.pdf.

[44] T. A. Davis, *Direct Methods for Sparse Linear Systems.* Philadelphia, PA: SIAM, 2006.

Faraz M. Mirzaei (S'05) received the B.Sc. degree in electrical engineering from Shiraz University, Shiraz, Iran, in 2004 and the M.Sc. degree in computer science and engineering (CSE) in 2007 from the University of Minnesota, Minneapolis, where he is currently working toward the Ph.D. degree with the CSE Department.

His current research interests include single- and multirobot systems, localization and tracking, structure from motion, and sensor calibration.

Mr. Mirzaei received the University of Minnesota 2007 Excellence in Research Award from the CSE Department and the 2008–2009 Doctoral Dissertation Fellowship.

Stergios I. Roumeliotis (M'03) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1995 and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1999 and 2000, respectively.

From 2000 to 2002, he was a Postdoctoral Fellow at California Institute of Technology, Pasadena. Between 2002 and 2008, he was an Assistant Professor with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, where he is currently an Associate Professor. His current research interests include inertial navigation of aerial and ground autonomous vehicles, fault detection and identification, sensor networks, distributed estimation under communication and processing constraints, and active sensing for reconfigurable networks of mobile sensors.

Dr. Roumeliotis is currently an Associate Editor for the IEEE TRANSACTIONS ON ROBOTICS. He received the National Science Foundation (NSF) CAREER Award (2006), the McKnight Land-Grant Professorship Award (2006–2008), the International Conference on Robotics and Automation (ICRA) Best Reviewer Award (2006), and was the co-recipient of the One National Aeronautics and Space Administration (NASA) Peer Award (2006), and the One NASA Center Best Award (2006). Papers he has coauthored have received the Robotics Society of Japan Best Journal Paper Award (2007), the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Best Student Paper Award (2006), the NASA Tech Briefs Award (2004), and one of them was the Finalist for the International Conference on Intelligent Robots and Systems (IROS) Best Paper Award (2006).