

# Planar-Based Visual Inertial Navigation: Observability Analysis and Motion Estimation

Ghazaleh Panahandeh · Seth Hutchinson ·  
Peter Händel · Magnus Jansson

Received: 7 January 2015 / Accepted: 6 August 2015  
© Springer Science+Business Media Dordrecht 2015

**Abstract** In this paper, we address the problem of ego-motion estimation by fusing visual and inertial information. The hardware consists of an inertial measurement unit (IMU) and a monocular camera. The camera provides visual observations in the form of features on a horizontal plane. Exploiting the geometric constraint of features on the plane into visual and inertial data, we propose a novel closed form measurement model for this system. Our first contribution in this paper is an observability analysis of the proposed planar-based visual inertial navigation system (VINS). In particular, we prove that the system has only three unobservable states corresponding to global translations parallel to the plane, and rotation around

the gravity vector. Hence, compared to general VINS, an advantage of using features on the horizontal plane is that the vertical translation along the normal of the plane becomes observable. As the second contribution, we present a state-space formulation for the pose estimation in the analyzed system and solve it via a modified unscented Kalman filter (UKF). Finally, the findings of the theoretical analysis and 6-DoF motion estimation are validated by simulations as well as using experimental data.

**Keywords** Visual-inertial navigation · Motion estimation · Observability analysis

## 1 Introduction

Integration of vision and inertial navigation systems (VINSs) has generated wide research interest in developing more accurate and simple, with low computational complexity, positioning systems in different areas such as robotics, traffic applications, as well as personal navigation systems. This is mainly due to the development of lightweight and cheap inertial sensors that are becoming a standard feature of smart-phones and personal digital assistants. Hence, finding a reliable solution for merging the inertial data with visual information is a crucial challenge in the current visual inertial systems.

---

G. Panahandeh (✉) · P. Händel · M. Jansson  
ACCESS Linnaeus Centre, School of Electrical  
Engineering, KTH Royal Institute of Technology,  
Stockholm, Sweden  
e-mail: ghpa@kth.se

P. Händel  
e-mail: ph@kth.se

M. Jansson  
e-mail: janssonm@kth.se

S. Hutchinson  
University of Illinois at Urbana-Champaign,  
Urbana, IL, USA  
e-mail: seth@illinois.edu

In this paper, we look at the problem of VINS motion estimation for a system that is navigating in an unknown environment, termed as *ego-motion estimation*. In particular, we are interested in estimating the 3D trajectory of a moving monocular camera mounted rigidly to an inertial measurement unit (IMU), without estimating a map of the environment. More specifically, we investigate the problem of IMU-camera ego-motion estimation when the visual observations are located on a horizontal plane. For this nonlinear system, we first study its observability properties and then we propose a novel and accurate 6-DoF motion estimation approach for the analysed system.

To the best of our knowledge, the work presented in this paper is the first to examine the observability properties of the 6-DoF VINS when observing features on a horizontal plane. In this system, we perform the observability analysis and show that while the key results of the previous observability analyses (e.g., [8, 13, 15, 16]) are valid (the robot's global position and its orientation around the normal of the plane are unobservable), by constraining visual observations to be on a horizontal plane, the orthogonal translation of the camera with respect to the plane becomes observable. More specifically, we prove that by observing unknown feature points on a horizontal plane, the navigation system has only three unobservable directions corresponding to the global translations parallel to the plane, and the rotation around the gravity vector.

Then for the presented VINS, we develop a motion estimation approach, which is an extension of our previous work published in [30]. In this paper, we relax the assumption of using a downward looking camera in which the camera optical axis is orthogonal to the horizontal plane and propose an accurate positioning system for 6-DoF motion estimation.

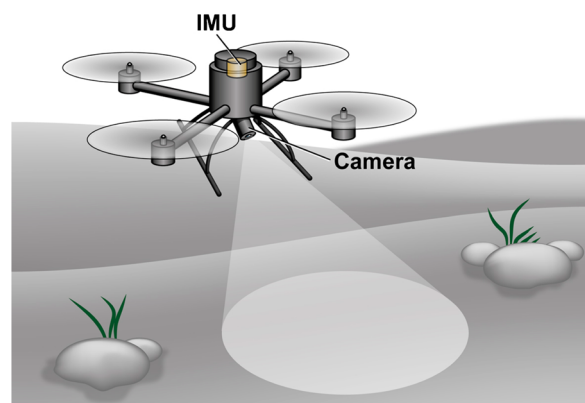
We study the performance of the proposed solution both with simulation and experimental data. The results show that it can be used as a promising positioning system in consumer products like visual inertial based applications in smartphones for localization, or 3D reconstruction without having access to external tracking setups, for indoor or GPS-denied environments. Moreover, estimating the metric

distance to the plane can provide useful information for take-off and landing without using any markers or pre-built maps. Currently, the main application for the considered work is the close-to-landing maneuvers of quadcopters and other UAVs, as illustrated in Fig. 1.

The key contributions of this paper are summarized in the following:

- The first contribution is the observability analysis of a VINS in which the visual observations are features that are located on a horizontal plane whose normal is assumed to be known. Our analysis is based on the Lie derivatives for finding the observable and unobservable modes of the time varying nonlinear VINS. In our study, the full INS parameter vector (including position, velocity, rotation, and inertial sensor biases) as well as the 3D position of the observed feature points are considered as state variables for analyzing the observability properties of the system. In particular, we prove that the system has (only) three unobservable directions corresponding to global translations along the horizontal plane, and rotations around the gravity vector. It is worth noting that only a single feature point is used for the analysis (given that the feature point is on the desired horizontal plane).

In conclusion, compared to general VINS in which there is no assumption on the geometrical properties of visual observations, an



**Fig. 1** Quadcopter equipped with IMU and camera

advantage of using features on a horizontal plane is that the vertical translation of the camera with respect to this plane becomes observable.

- As the second contribution of this paper, we propose an ego-motion estimation approach for the analyzed VINS. The system maintains a set of feature points that are observed on a horizontal plane. Based on matched feature points between the current and previous images, a novel measurement model is introduced that imposes constraints based on the geometrical properties of the visual observations to perform the motion estimation. Pose estimation is formulated implicitly in a state-space framework and is performed by an unscented Kalman filter (UKF). In contrast to the motion estimation approaches that rely on structure from motion or homography-based motion estimation methods, which need a sufficient baseline between the images for camera pose estimation, our proposed solution is independent of the motion of the system. Our proposed measurement model holds even in the absence of translational or rotational motion. Moreover, the IMU-camera sensor fusion system is free to have 6-DoF motion while observing visual features on the plane.
- Finally, the findings of the theoretical analysis for observability analysis and motion estimation are validated through extensive simulations and experiments.

The structure of this paper is as follows. A summary of the related works on VINS observability analysis and motion estimation is given in Section 1.1. Notations are introduced in Section 1.2. The system model is described in Section 2. In Section 3, we provide an overview of the nonlinear observability analysis, and briefly describe the approach in [5], which is used for analysis. In Section 4, we specifically study the observability properties of a VINS in which the visual observations are on a horizontal plane. Our proposed motion estimation approach is presented in Section 5.

Experimental results are presented in Section 6. Discussion is given in Section 7. Finally, the conclusion of the study is summarized in Section 8.

## 1.1 Related Works

### 1.1.1 VINS Observability Analysis

Recently, there has been a growing interest in studying observability properties of VINSs [8, 12, 13, 15, 16, 20, 23, 27, 38]. Observability provides an understanding of how well states of a system can be inferred from the system output measurements [7].

The observability properties of a time invariant linear system can be derived using the Kalman canonical decomposition. However, the problem becomes more complex for a nonlinear, time-varying system, such as the VINS. In this case, the study of observability properties is restricted to *locally weakly observability analysis* [7].

In the VINS, basic state variables are the INS unknown parameters including the position, velocity, and rotation of the camera with respect to a fixed reference frame, and the biases in the IMU sensor. However depending on the problem, the system state vector can be augmented with additional variables, such as IMU-camera calibration parameters [15, 23], landmark positions [8, 16], or gravity vector [15].

The VINS observability properties were first studied in [12, 13, 15, 23]. Specifically in [15, 23], the authors prove that all the quantities of the IMU-camera sensor fusion system (i.e., the robot pose, velocity, IMU biases, feature positions and the transformation between the IMU and the camera) are observable given observations of known feature points. Both of these analyses in [15, 23] are established by studying the observability rank condition based on Lie derivatives [7].

In [12, 13], the observability properties of VINS have been studied by analyzing indistinguishable trajectories of the system for different sensor configurations including: inertial only, vision only, vision and inertial where it includes the effects of unknown gravity and IMU-camera calibration parameters, and it is shown that under general position conditions both gravity and the IMU-camera calibration parameters are observable.

The presented VINS observability analyses in [8–10, 16, 19, 38] are among the most recent related works, which specifically study observability properties of the INS state variables for motion estimation

in unknown environments. For instance, the analyses in [8, 16] result in four unobservable directions, corresponding to global translations and global rotation about the gravity vector. The analytical studies in [8, 16] are done by utilizing the concept of observability Gramian [21].

In [19] a closed form solution to the VINS problem is proposed together with the analytical derivation of the system's observable modes, where for the observability analysis the concept of the continuous symmetric method is used [18].

Of particular relevance to our work is the system introduced in [38] in which the system is equipped with a laser pointer rigidly mounted into the IMU-camera frame. In this system, the visual navigation is based on observing a single laser spot on a planar surface, while the pose of the laser pointer is partially known with respect to the camera. Although the IMU biases are neither considered in the observability analysis nor in the motion estimation, the system reveals similar observability properties in the observable modes as our system, where for the analysis the proposed method in [18] is used. In [27], we provide the observability analysis for a downward looking camera in a VINS. However, the analysis was performed for a special case where the camera optical axis is orthogonal to the ground plane.

To date no study exists on the observability analysis of a 6-DoF VINS system by considering geometrical properties of the observed planar features. For the observability analysis, we employ the method introduced in [5], which is based on the Lie derivatives [7] and significantly reduces the complexity of finding the system's unobservable directions. In this paper, we prove that by observing unknown feature points on a horizontal plane, the navigation system has only three unobservable directions corresponding to the global translations parallel to the defined horizontal plane, and the rotation around the gravity vector. It should be mentioned that, our analysis, which is based on using the exact measurement equation of the system, results in a rigorous observability analysis and exactly reveals all the unobservable directions of our system. Compared to the methods that are based on using the inferred camera measurement model for the observability analysis [15, 23], our proposed method has

the advantage of analysing all the system parameters through the measurement equation simultaneously.

### 1.1.2 VINS Motion Estimation

Navigating in structured environments, e.g., man-made buildings, can provide significant information and the consequent constraints in the visual inertial solution. In the literature, various structures have been investigated for vision-based motion estimation, e.g., horizontal or vertical lines, vanishing points and lines [6], edges of buildings or fixed objects, and artificial visual tags [39]. However, there are only a few approaches that take the advantage of planar features for motion estimation, which is a salient structure in indoor environments, e.g., planar features on the ground, wall, or roof.

Within all different planar structures in environments, in this paper we focus our study on navigating over the horizontal plane of the ground. The choice of observing planar features on the ground is the most feasible scenario in many positioning applications both for personal and robotic navigation systems. More specifically, the importance of fusing ground planar features for motion estimation is vital where there are many moving objects (such as pedestrians and vehicles) in front of the camera. Furthermore, in certain cases, such as a micro air vehicle (MAV) or spacecraft landing, most of the features observed by camera lie on the ground, and one should take advantage of this fact to improve the estimation accuracy [25].

Among the VINS motion estimation approaches that are based on using ground facing camera, we can refer to the following methods: To date, the best ego-motion estimation results is obtained by a multi-state constraint Kalman filter introduced in [25]. The key advantage of this method is the use of a new measurement model that employs the geometric constraints of observing static features from multiple camera poses [24]. In [37] a method is proposed for estimating the velocity of mobile robots based on the Kalman filter integration using an optical flow method for a downward-looking camera.

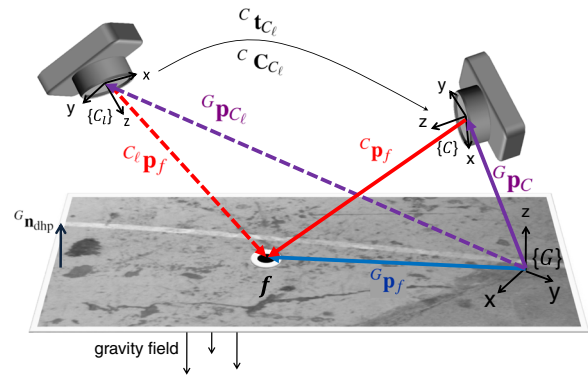
In [11], an IMU-camera GPS aided system is designed for pedestrian navigation in which the user is carrying a ground facing camera in the sensor fusion

system. To restrict the IMU error drift in the system, the planar homography of the ground plane features is constructed and it provides the velocity of the camera in the measurement model. However, the use of homography for motion estimation include these issues: 1) the sufficient number of detected image feature correspondences need to be more than four, 2) the sensitivity of the method to outliers, 3) the distance to the planar surface is unobservable when there is no translation in the system.

Although the observation of ground planar features has been implicitly used in visual inertial navigation systems [2, 11, 22, 25, 34, 35, 37, 40], this condition was first explicitly used in [30] for the VINS motion estimation. In [30], we proposed an IMU-camera ego-motion approach in which the ground planar features were directly used to construct the inertial model of the system. However, the developed system was for a downward looking camera. The assumption of the orthogonality of the camera optical axis to the ground is a nice property to overcome the use of additional sensors for recovering the lost scale factor, and the reported experimental results in [30] confirm the reliability and persistency of the developed method to the level of the noise, and the approach works well even with a few numbers of features. However, this assumption might be violated in a real scenario. In this paper, we relax this assumption and propose a general solution in which the camera is not restricted to be downward looking. That is, no specific restrictions are imposed on the camera motion other than that the camera is continuously observing features on the plane. The problem of motion estimation is formulated in a state space framework and is solved using a modified UKF. We study the performance of the proposed solution both in simulation and in an indoor environment, and show that it can be used as a promising positioning system, even at the presence of only few numbers of features (Fig. 2).

## 1.2 Notation

In the following sections scalars are denoted by lower-case letters ( $s$ ), vectors by bold letters ( $\mathbf{f}$ ), and matrices by bold capitals ( $\mathbf{K}$ ).  $\mathbf{I}_m$  denotes the  $m \times m$  square identity matrix.  $\mathbf{0}_{m \times n}$  denotes the  $m \times n$  zero matrix.



**Fig. 2** The IMU-camera sensor fusion system and its corresponding coordinate frames at current  $\{C\}$  time and  $\ell$  lags from the current one  $\{C_\ell\}$ . The relative rotation and translation between the time instants are denoted as  ${}^C\mathbf{C}_{C_\ell}$  and  ${}^C\mathbf{t}_{C_\ell}$ , respectively. The global frame of reference  $G$  is assumed to be located on a desired horizontal plane, where the normal of the plane with respect to the global frame  ${}^G\mathbf{n}_{dhp}$  is parallel with gravity vector. A sample feature point  $f$ , located on the desired horizontal plane, is considered to be in the cameras' field of view

$\mathbf{e}_i \in \mathbb{R}^3$  for  $i = 1, 2, 3$  and  $\mathbf{e}_1^\top = [1 \ 0 \ 0]$ ,  $\mathbf{e}_2^\top = [0 \ 1 \ 0]$ , and  $\mathbf{e}_3^\top = [0 \ 0 \ 1]$ . The global, IMU, and the current camera frames are denoted,  $\{G\}$ ,  $\{I\}$ , and  $\{C\}$ , respectively.  $\{C_\ell\}$  is the coordinate frame of the camera capturing an image  $\ell$  lags from the current one.  ${}^A\mathbf{p}_B$  represents the position of coordinate frame  $\{B\}$  in coordinate frame  $\{A\}$ , and  ${}^A\mathbf{v}_B$  denotes velocity of coordinate frame  $\{B\}$  in coordinate frame  $\{A\}$ . Based on the Euler rotation theorem, the principal rotation vector  $\theta$  is defined as  $\theta = \alpha \hat{\mathbf{k}}$  [36], where  $\hat{\mathbf{k}}$  is the unit vector along the axis and  $\alpha$  the angle of rotation. To represent the attitude, we use both the quaternion,  $\mathbf{q} \in \mathbb{R}^{4 \times 1}$ , and the Cayley-Gibbs-Rodrigues parameterization,  $\mathbf{s} \in \mathbb{R}^{3 \times 1}$ , where  $\mathbf{s} = \tan(\frac{\alpha}{2})\hat{\mathbf{k}}$ . Then,  ${}^A\mathbf{q}_B$  and  ${}^A\mathbf{s}_B$  are used to denote the orientation of the frame  $\{B\}$  in the frame of reference  $\{A\}$ ;  $\mathbf{C}(\mathbf{q})$  and  $\mathbf{C}(\mathbf{s})$  are the rotation matrices corresponding to  $\mathbf{q}$  and  $\mathbf{s}$ , respectively.

The IMU-camera extrinsic calibration parameters,  $\{{}^I\mathbf{p}_C, \mathbf{C}({}^I\mathbf{q}_C)\}$ , are estimated based on the approach of [29]. To preserve the clarity of the presentation, we assume that the IMU and camera frame of reference coincide.

Without loss of generality, we consider the global frame of reference on our desired horizontal plane

where its  $z$  axis is pointing in the opposite direction of the gravity field (see Fig. 3).

The skew-symmetric matrix of vector  $\mathbf{a}$  is represented by  $[\mathbf{a}]$  and the following properties of the cross product skew-symmetric matrix are used:  $\mathbf{a} \times \mathbf{b} = [\mathbf{b}]\mathbf{a} = -[\mathbf{a}]\mathbf{b}$ ,  $[\mathbf{a}]\mathbf{a} = \mathbf{0}_{3 \times 1}$ ,  $[\mathbf{a}][\mathbf{b}] = [\mathbf{b}\mathbf{a}^\top - (\mathbf{a}^\top\mathbf{b})\mathbf{I}_3]$ ,  $[\mathbf{A}\mathbf{b}] = \mathbf{A}[\mathbf{b}]\mathbf{A}^\top$ ,  $\mathbf{b}^\top[\mathbf{a}]\mathbf{b} = 0$ ,  $\forall \mathbf{A} \in \mathbb{R}^{3 \times 3}$  and  $\forall \{\mathbf{a}, \mathbf{b}\} \in \mathbb{R}^{3 \times 1}$ .

## 2 General System Model

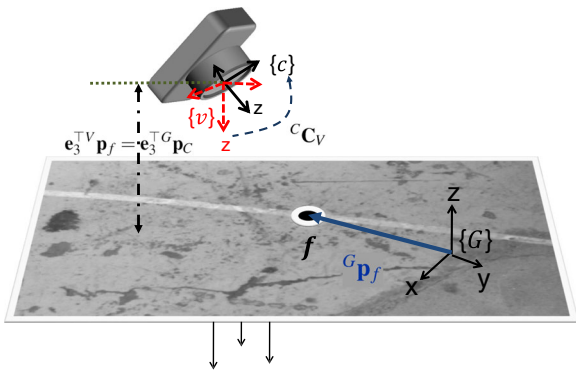
We hereafter describe the general INS propagation model in Section 2.1. The constructed propagation model is based on the IMU measurements and camera measurement model in Section 2.2, where the camera measurement is employed for state corrections. These two models are the basic propagation and measurement models which are used in later sections under different parameterizations for the purpose of observability analysis and motion estimation.

### 2.1 INS Propagation Model

We define the INS state variables in the form of the system state vector

$$\mathbf{x}^{\text{ins}} = \left[ {}^C\mathbf{q}_G^\top \quad {}^G\mathbf{v}_C^\top \quad {}^G\mathbf{p}_C^\top \quad \mathbf{b}_a^\top \quad \mathbf{b}_g^\top \right]^\top, \quad (1)$$

where  ${}^C\mathbf{q}_G$  is the quaternion that represents the orientation of the global frame  $\{G\}$  in the camera's frame of reference  $\{C\}$ .



**Fig. 3** Graphical representation of the virtual camera coordinate frame  $\{V\}$ , which coincides with real camera coordinate frame  $\{C\}$ . The optical axis of the virtual camera is selected to be always orthogonal to the desired plane

The velocity and the position of the camera in the global frame are denoted by  ${}^G\mathbf{v}_C$  and  ${}^G\mathbf{p}_C$ , respectively.  $\mathbf{b}_a$  and  $\mathbf{b}_g$  are the bias vectors affecting the accelerometer and gyroscope measurements, respectively.

The time evolution of the INS state is given by

$${}^C\dot{\mathbf{q}}_G(t) = \frac{1}{2} \begin{bmatrix} -[\boldsymbol{\omega}(t)] & \boldsymbol{\omega}(t) \\ \boldsymbol{\omega}(t)^\top & 0 \end{bmatrix} {}^C\mathbf{q}_G(t), \quad (2)$$

$${}^G\dot{\mathbf{v}}_C(t) = {}^G\mathbf{a}(t) = {}^G\mathbf{g} + \mathbf{C}({}^C\mathbf{q}_G(t))^\top (\mathbf{a}(t) - \mathbf{b}_a(t)),$$

$${}^G\dot{\mathbf{p}}_C(t) = {}^G\mathbf{v}_C(t), \quad \dot{\mathbf{b}}_a(t) = \mathbf{n}_{\delta a}, \quad \dot{\mathbf{b}}_g(t) = \mathbf{n}_{\delta g},$$

where  $\boldsymbol{\omega}(t) = [\omega_1 \ \omega_2 \ \omega_3]^\top$  and  $\mathbf{a}(t) = [a_1 \ a_2 \ a_3]^\top$  are the IMU rotational velocity and specific acceleration,  ${}^G\mathbf{g} = [0 \ 0 \ g]^\top$  is the gravitational acceleration,  $\mathbf{C}(\mathbf{q})$  is the rotation matrix corresponding to  $\mathbf{q}$ , and  $\mathbf{n}_{\delta a}$  and  $\mathbf{n}_{\delta g}$  are the accelerometer and gyroscope bias driving white Gaussian noises. The output measurement signals of the gyroscope,  $\boldsymbol{\omega}_m$ , and accelerometer,  $\mathbf{a}_m$ , are modeled as:

$$\boldsymbol{\omega}_m(t) = \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t), \quad (3)$$

$$\mathbf{a}_m(t) = \mathbf{C}({}^C\mathbf{q}_G(t)) ({}^G\mathbf{a}(t) - {}^G\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t), \quad (4)$$

where  $\mathbf{n}_g$  and  $\mathbf{n}_a$  are zero-mean white Gaussian noise processes.

### 2.2 Camera Measurement Model

Consider a feature point  $f$  observed by a camera, where its position relative to the camera center  $\{C\}$  is denoted by  ${}^C\mathbf{p}_f$ . Using the pinhole camera model [6], its homogenous and normalized pixel coordinates on the image plane are represented by  ${}^C\bar{\mathbf{z}}$  and  ${}^C\mathbf{z}$ , respectively, as:

$${}^C\bar{\mathbf{z}} = (\mathbf{e}_3^\top {}^C\mathbf{p}_f)^{-1} {}^C\mathbf{p}_f, \quad (5)$$

and

$${}^C\mathbf{z} = [\mathbf{I}_2 \ 0_{2 \times 1}] (\mathbf{e}_3^\top {}^C\mathbf{p}_f)^{-1} {}^C\mathbf{p}_f, \quad (6)$$

where

$${}^C\mathbf{p}_f = \mathbf{C}({}^C\mathbf{q}_G)({}^G\mathbf{p}_f - {}^G\mathbf{p}_C). \quad (7)$$

Figure 2 depicts the relative rotation and translation vectors between the camera, the feature point  $f$ , and the global frame of reference  $\{G\}$ .



### 3 Nonlinear Observability Analysis

In Section 3.1, we first provide an overview of the nonlinear observability rank condition of the method proposed in [7]. Then, in Section 3.2, we present our summary, [27], of the method of [5] for finding the observable and unobservable modes of a nonlinear system.

#### 3.1 Observability Analysis with Lie Derivatives

Consider a nonlinear system

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}_0(\mathbf{x}) + \sum_{i=1}^{\ell} \mathbf{f}_i(\mathbf{x})u_i \\ \mathbf{y} = \mathbf{h}(\mathbf{x}) \end{cases} \quad (8)$$

where  $\mathbf{x} \in \mathbb{R}^m$  is the state vector,  $\mathbf{u} = [u_1 \dots u_{\ell}]^T$  is the system input,  $\mathbf{y} \in \mathbb{R}^k$  is the system output, and  $\mathbf{f}_i$  for  $i \in \{0, \dots, \ell\}$  is the process function.

The zeroth order Lie derivative of a measurement function  $\mathbf{h}$  is the function itself, i.e.,  $\mathcal{L}^0 \mathbf{h} = \mathbf{h}(\mathbf{x})$ . For any  $n$ -th order Lie derivative,  $\mathcal{L}^n \mathbf{h}$ , the  $(n+1)$ -th order Lie derivative  $\mathcal{L}_{\mathbf{f}_i}^{n+1} \mathbf{h}$  with respect to a process function  $\mathbf{f}_i$  is computed as:

$$\mathcal{L}_{\mathbf{f}_i}^{n+1} \mathbf{h} = \nabla \mathcal{L}^n \mathbf{h} \cdot \mathbf{f}_i, \quad (9)$$

where  $\nabla$  denotes the gradient operator with respect to  $\mathbf{x}$  and ' $\cdot$ ' represents the vector inner product. Similarly, mixed higher order Lie derivatives is defined as:

$$\mathcal{L}_{\mathbf{f}_i \mathbf{f}_j \dots \mathbf{f}_k}^n \mathbf{h} = \mathcal{L}_{\mathbf{f}_i} (\mathcal{L}_{\mathbf{f}_j \dots \mathbf{f}_k}^{n-1} \mathbf{h}) = \nabla \mathcal{L}_{\mathbf{f}_j \dots \mathbf{f}_k}^{n-1} \mathbf{h} \cdot \mathbf{f}_i \quad (10)$$

where  $i, j, k \in \{0, \dots, \ell\}$ . The observability of a system is determined by calculating the dimension of the space spanned by the gradients of the Lie derivatives of its output functions [7]. Hence, the observability matrix  $\mathbf{O}$  of system (8) is defined as:

$$\mathbf{O} \triangleq \begin{bmatrix} \nabla \mathcal{L}^0 \mathbf{h} \\ \nabla \mathcal{L}_{\mathbf{f}_1}^1 \mathbf{h} \\ \vdots \\ \nabla \mathcal{L}_{\mathbf{f}_i \mathbf{f}_j \dots \mathbf{f}_k}^n \mathbf{h} \\ \vdots \end{bmatrix}. \quad (11)$$

To prove that a system is observable, it suffices to show that  $\mathbf{O}$  is of full column rank. However, to prove that a system is unobservable, we have to find the null space of matrix  $\mathbf{O}$ , which may have infinitely many rows. This can be challenging especially for high-dimensional systems, such as the VINS. To address this issue, in the following section we present the

method of [5, 9] for proving that a system is unobservable and finding its unobservable directions.

#### 3.2 Observability Analysis with Basis Functions

**Theorem 1** Assume that there exists a nonlinear transformation  $\boldsymbol{\beta}(\mathbf{x}) = [\boldsymbol{\beta}_1(\mathbf{x})^T, \dots, \boldsymbol{\beta}_n(\mathbf{x})^T]^T$  (i.e., a set of basis functions) of the variable  $\mathbf{x}$ , such that:

1. The system measurement equation can be written as a function of  $\boldsymbol{\beta}$ , i.e.,  $\mathbf{y} = \mathbf{h}(\mathbf{x}) = \bar{\mathbf{h}}(\boldsymbol{\beta})$
2.  $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \mathbf{f}_j$ , for  $j = 0, \dots, \ell$ , is a function of  $\boldsymbol{\beta}$

Then the observability matrix of system (8) can be factorized as:  $\mathbf{O} = \boldsymbol{\Xi} \boldsymbol{\Omega}$  where  $\boldsymbol{\Xi}$  is the observability matrix of system

$$\begin{cases} \dot{\boldsymbol{\beta}} = \mathbf{g}_0(\boldsymbol{\beta}) + \sum_{i=1}^{\ell} \mathbf{g}_i(\boldsymbol{\beta})u_i \\ \mathbf{y} = \bar{\mathbf{h}}(\boldsymbol{\beta}) \end{cases} \quad (12)$$

and  $\boldsymbol{\Omega} \triangleq \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}}$ .

*Proof* See [5].  $\square$

Note that system (12) results by pre-multiplying the process function in Eq. 8 with  $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}}$

$$\begin{cases} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial t} = \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \mathbf{f}_0(\mathbf{x}) + \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \sum_{i=1}^{\ell} \mathbf{f}_i(\mathbf{x})u_i \\ \mathbf{y} = \mathbf{h}(\mathbf{x}) \end{cases} \Rightarrow \begin{cases} \dot{\boldsymbol{\beta}} = \mathbf{g}_0(\boldsymbol{\beta}) + \sum_{i=1}^{\ell} \mathbf{g}_i(\boldsymbol{\beta})u_i \\ \mathbf{y} = \bar{\mathbf{h}}(\boldsymbol{\beta}) \end{cases}$$

where  $\mathbf{g}_i(\boldsymbol{\beta}) \triangleq \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{x}} \mathbf{f}_i(\mathbf{x})$  and  $\bar{\mathbf{h}}(\boldsymbol{\beta}) \triangleq \mathbf{h}(\mathbf{x})$ .

**Corollary 1** If  $\boldsymbol{\Xi}$  is of full column rank, i.e., system (12) is observable, then the unobservable directions of system (8) will be spanned by the null vectors of  $\boldsymbol{\Omega}$ .

*Proof* From  $\mathbf{O} = \boldsymbol{\Xi} \boldsymbol{\Omega}$ , we have  $\text{null}(\mathbf{O}) = \text{null}(\boldsymbol{\Omega}) \cup (\text{null}(\boldsymbol{\Xi}) \cap \text{range}(\boldsymbol{\Omega}))$ . Therefore, if  $\boldsymbol{\Xi}$  is of full column rank, i.e., system (12) is observable, then  $\text{null}(\mathbf{O}) = \text{null}(\boldsymbol{\Omega})$ .  $\square$

Based on Theorem 1 and Corollary 1, to find unobservable directions of a system, we first need to define the basis functions,  $\boldsymbol{\beta}$ , which fulfil the first and second conditions of Theorem 1. Then we should prove that the infinite-dimensional matrix  $\boldsymbol{\Xi}$  has full column rank, which satisfies the condition of Corollary 1.

To define the basis functions, we start with the system measurement equation and extract the initial bases as a function of the state variables (i.e., the first condition of the Theorem 1). The rest of the bases will be defined by projecting these initial bases into the process functions. Then any resulting term that cannot be expressed as a function of the previously defined bases is incorporated as a new basis (i.e., the second condition of the Theorem 1). Finally, we terminate the procedure of defining new basis functions when the second condition of Theorem 1 is satisfied.

#### 4 The System Observability Analysis

In what follows, we first introduce some modifications of the system model for the purpose of observability analysis in Section 4.1. Then, we define the basis functions for our VINS in Section 4.2. Finally, we derive its unobservable modes in Section 4.3.

##### 4.1 Modified System Model

To assist the observability analysis, the following modifications are considered in the system propagation and measurement model:

- In our analysis, we consider the position of the observed feature point with respect to the global coordinate frame,  ${}^G\mathbf{p}_f$ , as an unknown constant variable. Therefore, we append it to the state vector, and reflect its dynamic in the propagation model, see [8].
- Instead of using the quaternion parametrization, we re-parameterize the rotation matrix  $\mathbf{C}(\mathbf{q})$  in terms of the Cayley-Gibbs-Rodriguez parameter  $\mathbf{s}$ ,  $\mathbf{C}(\mathbf{s})$ . Then, the noise free system propagation model of Eq. 2, by considering time propagation of  ${}^G\mathbf{p}_f$ , is described by

$${}^C\dot{\mathbf{s}}_G(t) = \frac{1}{2}\mathbf{D}(\omega(t) - \mathbf{b}_g(t)), \quad (13)$$

$${}^G\dot{\mathbf{v}}_C(t) = {}^G\mathbf{g} + \mathbf{C}({}^C\mathbf{s}_G(t))^\top (\mathbf{a}(t) - \mathbf{b}_a(t)),$$

$${}^G\dot{\mathbf{p}}_C(t) = {}^G\mathbf{v}_C(t), \quad \dot{\mathbf{b}}_a(t) = \mathbf{0}_{3 \times 1},$$

$$\dot{\mathbf{b}}_g(t) = \mathbf{0}_{3 \times 1}, \quad {}^G\dot{\mathbf{p}}_f = \mathbf{0}_{3 \times 1}$$

where  $\frac{1}{2}\mathbf{D} \triangleq \frac{\partial \mathbf{s}}{\partial \theta} = \frac{1}{2}(\mathbf{I} + [\mathbf{s}] + \mathbf{s}\mathbf{s}^\top)$ , see [36].

- We state the following Lemma 1, to impose the geometric constraint of the visual observations in the system measurement (5). To analyze our VINS

observability properties, the results of Lemma 1 is used as our measurement equation for finding the system basis functions.

**Lemma 1** *For an arbitrary feature point  $f$  on a horizontal plane, the camera measurement model of the system based on the projective camera model can be represented as*

$$\mathbf{y} = \left( \mathbf{e}_3^\top {}^G\mathbf{p}_C \mathbf{I}_3 + \mathbf{C} \begin{pmatrix} {}^C\mathbf{s}_G \end{pmatrix} \right) \left( {}^G\mathbf{p}_f - {}^G\mathbf{p}_C \right) \mathbf{e}_3^\top \mathbf{C} \begin{pmatrix} {}^C\mathbf{s}_G \end{pmatrix}^\top \right) {}^C\bar{\mathbf{z}} = 0 \quad (14)$$

*Proof* Let us consider a virtual camera along with our real camera in which the center of its coordinate frame  $\{V\}$  coincides with the center of the real camera coordinate frame  $\{C\}$ . Moreover, the virtual camera optical axis is always restricted to be parallel with the normal of the reference plane, in this case the desired horizontal plane. Thus, the orientation of the virtual camera frame is fixed relative to the global frame  $\mathbf{C}({}^V\mathbf{s}_G) = \text{diag}(1, -1, -1)$ . Therefore, the position of  $\mathbf{p}_f$  in  $\{V\}$  along its optical axis,  $\mathbf{e}_3^\top {}^V\mathbf{p}_f$ , is equal to the height of the camera above the plane,

$$\mathbf{e}_3^\top {}^V\mathbf{p}_f = \mathbf{e}_3^\top {}^G\mathbf{p}_C. \quad (15)$$

As the transformation between the real camera frame and the virtual frame is related by a rotation,  ${}^C\mathbf{C}_V \triangleq \mathbf{C}({}^C\mathbf{s}_G)\mathbf{C}({}^V\mathbf{s}_G)^\top$ , the positions of  $\mathbf{p}_f$  in the two coordinate frames are related by

$${}^C\mathbf{p}_f = {}^C\mathbf{C}_V {}^V\mathbf{p}_f. \quad (16)$$

Substituting  ${}^V\mathbf{p}_f$  in Eq. 16 with its projection in the virtual image plane (5), we get

$${}^C\mathbf{p}_f = {}^C\mathbf{C}_V \mathbf{e}_3^\top {}^V\mathbf{p}_f {}^V\bar{\mathbf{z}}. \quad (17)$$

Moreover, the projection of feature point  $\mathbf{p}_f$  in the virtual image plane can be mapped to the real image plane as:

$${}^C\bar{\mathbf{z}} = \alpha {}^C\mathbf{C}_V {}^V\bar{\mathbf{z}} \quad (18)$$

in which the equality holds only up to a scale factor  $\alpha$ . Then, by substituting  ${}^V\bar{\mathbf{z}}$  from Eq. 18 to Eq. 17, we get

$$\begin{aligned} {}^C\mathbf{p}_f &= \alpha^{-1} \mathbf{e}_3^\top {}^V\mathbf{p}_f {}^C\mathbf{C}_V {}^C\mathbf{C}_V^\top {}^C\bar{\mathbf{z}} \\ &= \alpha^{-1} \mathbf{e}_3^\top {}^V\mathbf{p}_f {}^C\bar{\mathbf{z}} \\ &\Rightarrow \mathbf{e}_3^\top {}^V\mathbf{p}_f {}^C\bar{\mathbf{z}} - \alpha {}^C\mathbf{p}_f = 0 \end{aligned} \quad (19)$$



where

$$\begin{aligned}\alpha &= \mathbf{e}_3^T \mathbf{C} \mathbf{C}_V^T \mathbf{C} \bar{\mathbf{z}} = \mathbf{e}_3^T \mathbf{C} (\mathbf{C}_V^T \mathbf{C}) (\mathbf{C}_G^T \mathbf{C})^T \mathbf{C} \bar{\mathbf{z}} \\ &= -\mathbf{e}_3^T \mathbf{C} (\mathbf{C}_G^T \mathbf{C})^T \mathbf{C} \bar{\mathbf{z}}.\end{aligned}\quad (20)$$

Finally from Eqs. 7, 15, and 20, Eq. 19 is rewritten as

$$(\mathbf{e}_3^T \mathbf{p}_C \mathbf{I}_3 + \mathbf{C} (\mathbf{C}_G^T \mathbf{C})^T (\mathbf{p}_f - \mathbf{p}_C) \mathbf{e}_3^T \mathbf{C} (\mathbf{C}_G^T \mathbf{C})^T) \mathbf{C} \bar{\mathbf{z}} = 0 \quad (21)$$

which is Eq. 14.  $\square$

#### 4.2 Defining the basis functions

For simplicity, we retain only a few of the subscripts and superscripts in the state elements and write the augmented state vector as

$$\mathbf{x} = [\mathbf{s}^T \quad \mathbf{v}^T \quad \mathbf{p}^T \quad \mathbf{b}_a^T \quad \mathbf{b}_g^T \quad \mathbf{p}_f^T]^T. \quad (22)$$

Moreover, following the structure of system (8), we rewrite the state propagation equation in Eq. 13 as

$$\begin{bmatrix} \dot{\mathbf{s}} \\ \dot{\mathbf{v}} \\ \dot{\mathbf{p}} \\ \dot{\mathbf{b}}_a \\ \dot{\mathbf{b}}_g \\ \dot{\mathbf{p}}_f \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{1}{2} \mathbf{D} \mathbf{b}_g \\ \mathbf{g} - \mathbf{C}^T \mathbf{b}_a \\ \mathbf{v} \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \end{bmatrix}}_{\mathbf{f}_0} + \underbrace{\begin{bmatrix} \frac{1}{2} \mathbf{D} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \end{bmatrix}}_{\mathbf{F}_1} \omega + \underbrace{\begin{bmatrix} \mathbf{0}_{3 \times 3} \\ \mathbf{C}^T \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \end{bmatrix}}_{\mathbf{F}_2} \mathbf{a} \quad (23)$$

where  $\mathbf{C} \triangleq \mathbf{C}(\mathbf{s})$ . Note that  $\mathbf{f}_0$  is a  $24 \times 1$  vector, while  $\mathbf{F}_1$  and  $\mathbf{F}_2$  are both  $24 \times 3$  matrices representing three process functions as

$$\begin{aligned}\mathbf{F}_1 \omega &= \mathbf{f}_{11} \omega_1 + \mathbf{f}_{12} \omega_2 + \mathbf{f}_{13} \omega_3 \\ \mathbf{F}_2 \mathbf{a} &= \mathbf{f}_{21} a_1 + \mathbf{f}_{22} a_2 + \mathbf{f}_{23} a_3.\end{aligned}\quad (24)$$

Following the first condition of Theorem 1, we define the system's initial bases using the unknown terms appearing in the measurement function (14), i.e.,

$$\beta_1 \triangleq \mathbf{e}_3^T \mathbf{p}, \quad \beta_2 \triangleq \mathbf{C}(\mathbf{p}_f - \mathbf{p}), \quad \beta_3 \triangleq \mathbf{C} \mathbf{e}_3.$$

To check the second condition of Theorem 1, we compute their spans with respect to the state vector  $\mathbf{x}$

$$\begin{aligned}\frac{\partial \beta_1}{\partial \mathbf{x}} &= [\mathbf{0}_{1 \times 3} \quad \mathbf{0}_{1 \times 3} \quad \mathbf{e}_3^T \quad \mathbf{0}_{1 \times 3} \quad \mathbf{0}_{1 \times 3} \quad \mathbf{0}_{1 \times 3}] \\ \frac{\partial \beta_2}{\partial \mathbf{x}} &= [\mathbf{C}(\mathbf{p}_f - \mathbf{p})] \frac{\partial \theta}{\partial \mathbf{s}} \quad \mathbf{0}_{3 \times 3} \quad -\mathbf{C} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{C}] \\ \frac{\partial \beta_3}{\partial \mathbf{x}} &= [\mathbf{C} \mathbf{e}_3] \frac{\partial \theta}{\partial \mathbf{s}} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3}]\end{aligned}$$

and project them onto all the process functions. Specifically, for the span of  $\beta_1 \triangleq \mathbf{e}_3^T \mathbf{p}$ , we have:

$$\frac{\partial \beta_1}{\partial \mathbf{x}} \mathbf{f}_0 = \mathbf{e}_3^T \mathbf{v} \triangleq \beta_4, \quad \frac{\partial \beta_1}{\partial \mathbf{x}} \mathbf{f}_{1i} = 0, \quad \frac{\partial \beta_1}{\partial \mathbf{x}} \mathbf{f}_{2i} = 0,$$

for  $i = \{1, 2, 3\}$ , where we have defined a new basis element  $\beta_4 \triangleq \mathbf{e}_3^T \mathbf{v}$ . Similarly, for the span of  $\beta_2 \triangleq \mathbf{C}(\mathbf{p}_f - \mathbf{p})$ , we have:

$$\begin{aligned}\frac{\partial \beta_2}{\partial \mathbf{x}} \mathbf{f}_0 &= -[\mathbf{C}(\mathbf{p}_f - \mathbf{p})] \mathbf{b}_g - \mathbf{C} \mathbf{v} \\ &= -[\beta_2] \mathbf{b}_g - \mathbf{C} \mathbf{v} \triangleq -[\beta_2] \beta_5 - \beta_6\end{aligned}$$

$$\frac{\partial \beta_2}{\partial \mathbf{x}} \mathbf{f}_{1i} = [\mathbf{C}(\mathbf{p}_f - \mathbf{p})] \mathbf{e}_i = [\beta_2] \mathbf{e}_i$$

$$\frac{\partial \beta_2}{\partial \mathbf{x}} \mathbf{f}_{2i} = \mathbf{0}_{3 \times 1}$$

where  $\frac{\partial \theta}{\partial \mathbf{s}} \frac{1}{2} \mathbf{D} = \frac{\partial \theta}{\partial \mathbf{s}} \frac{\partial \mathbf{s}}{\partial \theta} = \mathbf{I}_3$ , and the newly defined bases are  $\beta_5 \triangleq \mathbf{b}_g$  and  $\beta_6 \triangleq \mathbf{C} \mathbf{v}$ .

Finally, for the span of  $\beta_3 \triangleq \mathbf{C} \mathbf{e}_3$ , we have:

$$\frac{\partial \beta_3}{\partial \mathbf{x}} \mathbf{f}_0 = -[\mathbf{C} \mathbf{e}_3] \mathbf{b}_g = -[\beta_3] \beta_5$$

$$\frac{\partial \beta_3}{\partial \mathbf{x}} \mathbf{f}_{1i} = [\mathbf{C} \mathbf{e}_3] \mathbf{e}_i = [\beta_3] \mathbf{e}_i$$

$$\frac{\partial \beta_3}{\partial \mathbf{x}} \mathbf{f}_{2i} = \mathbf{0}_{3 \times 1}.$$

In the next step, we repeat the same process of projecting the span of the newly defined basis functions  $\beta_4$ ,  $\beta_5$ , and  $\beta_6$  on the process functions. Specifically, we have:

$$\beta_4 \triangleq \mathbf{e}_3^T \mathbf{v}:$$

$$\frac{\partial \beta_4}{\partial \mathbf{x}} = [\mathbf{0}_{1 \times 3} \quad \mathbf{e}_3^T \quad \mathbf{0}_{1 \times 3} \quad \mathbf{0}_{1 \times 3} \quad \mathbf{0}_{1 \times 3} \quad \mathbf{0}_{1 \times 3}]$$

$$\begin{aligned}\frac{\partial \beta_4}{\partial \mathbf{x}} \mathbf{f}_0 &= \mathbf{e}_3^T (\mathbf{g} - \mathbf{C}^T \mathbf{b}_a) = \mathbf{e}_3^T \mathbf{C}^T (\mathbf{C} \mathbf{g} - \mathbf{b}_a) \\ &\triangleq \beta_3^T (g \beta_3 - \beta_7)\end{aligned}$$

$$\frac{\partial \beta_4}{\partial \mathbf{x}} \mathbf{f}_{1i} = 0, \quad \frac{\partial \beta_4}{\partial \mathbf{x}} \mathbf{f}_{2i} = \mathbf{e}_3^T \mathbf{C}^T \mathbf{e}_i = \beta_3^T \mathbf{e}_i,$$

where  $\beta_7 \triangleq \mathbf{b}_a$  and  $g = |\mathbf{g}|$  is the norm of the gravitational acceleration.

$$\beta_5 \triangleq \mathbf{b}_g:$$

$$\frac{\partial \beta_5}{\partial \mathbf{x}} = [\mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{I}_3 \quad \mathbf{0}_{3 \times 3}]$$

$$\frac{\partial \beta_5}{\partial \mathbf{x}} \mathbf{f}_0 = \mathbf{0}_{3 \times 1}, \quad \frac{\partial \beta_5}{\partial \mathbf{x}} \mathbf{f}_{1i} = \mathbf{0}_{3 \times 1}, \quad \frac{\partial \beta_5}{\partial \mathbf{x}} \mathbf{f}_{2i} = \mathbf{0}_{3 \times 1}$$

$$\beta_6 \triangleq \mathbf{C}\mathbf{v}:$$

$$\frac{\partial \beta_6}{\partial \mathbf{x}} = [\lfloor \mathbf{C}\mathbf{v} \rfloor \frac{\partial \theta}{\partial s} \quad \mathbf{C} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3}]$$

$$\begin{aligned} \frac{\partial \beta_6}{\partial \mathbf{x}} \mathbf{f}_0 &= -\lfloor \mathbf{C}\mathbf{v} \rfloor \mathbf{b}_g + \mathbf{C}\mathbf{g} - \mathbf{b}_a \\ &\triangleq -\lfloor \beta_6 \rfloor \beta_5 + g\beta_3 - \beta_7 \end{aligned}$$

$$\frac{\partial \beta_6}{\partial \mathbf{x}} \mathbf{f}_{1i} = \lfloor \mathbf{C}\mathbf{v} \rfloor \mathbf{e}_i = \lfloor \beta_6 \rfloor \mathbf{e}_i$$

$$\frac{\partial \beta_6}{\partial \mathbf{x}} \mathbf{f}_{2i} = \mathbf{C}\mathbf{C}^\top \mathbf{e}_i = \mathbf{e}_i$$

$$\beta_7 \triangleq \mathbf{b}_a:$$

$$\frac{\partial \beta_7}{\partial \mathbf{x}} = [\mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3} \quad \mathbf{I}_3 \quad \mathbf{0}_{3 \times 3} \quad \mathbf{0}_{3 \times 3}]$$

$$\frac{\partial \beta_7}{\partial \mathbf{x}} \mathbf{f}_0 = \mathbf{0}_{3 \times 1}, \quad \frac{\partial \beta_7}{\partial \mathbf{x}} \mathbf{f}_{1i} = \mathbf{0}_{3 \times 1}, \quad \frac{\partial \beta_7}{\partial \mathbf{x}} \mathbf{f}_{2i} = \mathbf{0}_{3 \times 1}.$$

Since all the terms in the preceding projections are defined based on the existing basis functions (i.e., second condition of Theorem 1 is satisfied), we have found a complete set of basis functions. The corresponding new process model for the defined bases can be expressed as:

$$\begin{bmatrix} \dot{\beta}_1 \\ \dot{\beta}_2 \\ \dot{\beta}_3 \\ \dot{\beta}_4 \\ \dot{\beta}_5 \\ \dot{\beta}_6 \\ \dot{\beta}_7 \end{bmatrix} = \underbrace{\begin{bmatrix} \beta_4 \\ -\lfloor \beta_2 \rfloor \beta_5 - \beta_6 \\ -\lfloor \beta_3 \rfloor \beta_5 \\ \beta_3^\top (g\beta_3 - \beta_7) \\ \mathbf{0}_{3 \times 1} \\ -\lfloor \beta_6 \rfloor \beta_5 + g\beta_3 - \beta_7 \\ \mathbf{0}_{3 \times 1} \end{bmatrix}}_{\mathbf{g}_0} + \underbrace{\begin{bmatrix} \mathbf{0}_{1 \times 3} \\ \lfloor \beta_2 \rfloor \\ \lfloor \beta_3 \rfloor \\ \mathbf{0}_{1 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \lfloor \beta_6 \rfloor \\ \mathbf{0}_{3 \times 3} \end{bmatrix}}_{\mathbf{G}_1} \omega + \underbrace{\begin{bmatrix} \mathbf{0}_{1 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \\ \beta_3^\top \mathbf{I} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{I} \\ \mathbf{0}_{3 \times 3} \end{bmatrix}}_{\mathbf{G}_2} \mathbf{a} \quad (25)$$

and the measurement equation in terms of the basis functions is

$$\mathbf{y} = \bar{\mathbf{h}} = (\beta_1 \mathbf{I}_3 + \beta_2 \beta_3^\top) \bar{\mathbf{z}}. \quad (26)$$

Equation 26 is used in Appendix A for constructing matrix  $\mathcal{E}$  and showing that it is full rank, Corollary 1. It should be mentioned that the given proof in Appendix A is valid under the condition that the inputs to the system, i.e., corresponding coefficients of  $G_1$  and  $G_2$ , are not zero.

#### 4.3 Unobservable Directions of $\Omega$

Based on Corollary 1 and given that  $\mathcal{E}$  is full rank, proved in Appendix A, the unobservable directions of

system (8) lie in the null space of matrix  $\Omega$ . By stacking the derivatives of the basis functions with respect to the variable  $\mathbf{x}$ , the matrix  $\Omega$  is

$$\Omega = \begin{bmatrix} \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{e}_3^\top & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} \\ \lfloor \mathbf{C}(\mathbf{p}_f - \mathbf{p}) \rfloor \frac{\partial \theta}{\partial s} & \mathbf{0}_{3 \times 3} & -\mathbf{C} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{C} \\ \lfloor \mathbf{C}\mathbf{e}_3 \rfloor \frac{\partial \theta}{\partial s} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{1 \times 3} & \mathbf{e}_3^\top & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I} & \mathbf{0}_{3 \times 3} \\ \lfloor \mathbf{C}\mathbf{v} \rfloor \frac{\partial \theta}{\partial s} & \mathbf{C} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix}.$$

To describe the null space of  $\Omega$ , we need to find a matrix

$$\mathbf{A} = [\mathbf{A}_1^\top \quad \mathbf{A}_2^\top \quad \mathbf{A}_3^\top \quad \mathbf{A}_4^\top \quad \mathbf{A}_5^\top \quad \mathbf{A}_6^\top]^\top \neq \mathbf{0}, \quad (27)$$

such that

$$\Omega \mathbf{A} = \mathbf{0}. \quad (28)$$

From Eq. 28, we have:<sup>1</sup>

- Multiplying the fifth and the seventh block rows of  $\Omega$  with  $\mathbf{A}$ , we get  $\mathbf{A}_4 = \mathbf{A}_5 = \mathbf{0}$ .
- Multiplying the third block row of  $\Omega$  with  $\mathbf{A}$ , we have  $\lfloor \mathbf{C}\mathbf{e}_3 \rfloor \frac{\partial \theta}{\partial s} \mathbf{A}_1 = \mathbf{0}$ , which implies that either  $\mathbf{A}_1 = \mathbf{0}$  or  $\mathbf{A}_1 = \frac{\partial s}{\partial \theta} \mathbf{C}\mathbf{e}_3$ .
  1. If  $\mathbf{A}_1 = \mathbf{0}$ , then from the sixth block row of  $\Omega \mathbf{A} = \mathbf{0}$  we have:  $\mathbf{C}\mathbf{A}_2 = \mathbf{0} \Rightarrow \mathbf{A}_2 = \mathbf{0}$ , since  $\mathbf{C}$  is a rotation matrix and full rank. For the second block row of  $\Omega \mathbf{A}$  to be zero, we have:  $\mathbf{A}_3 = \mathbf{A}_6$ . Finally, from the first block row of  $\Omega \mathbf{A} = \mathbf{0}$ , we get  $\mathbf{e}_3^\top \mathbf{A}_3 = 0$ . This implies that  $\mathbf{A}_3$  is spanned by  $\mathbf{e}_1$  and  $\mathbf{e}_2$ , i.e.,

$$\mathbf{A}_3 = \mathbf{A}_6 = \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \end{bmatrix}.$$

2. If  $\mathbf{A}_1 = \frac{\partial s}{\partial \theta} \mathbf{C}\mathbf{e}_3$ , from the second block row of  $\Omega \mathbf{A} = \mathbf{0}$ , we have  $\mathbf{A}_3 = -\lfloor \mathbf{p} \rfloor \mathbf{e}_3$  and  $\mathbf{A}_6 = -\lfloor \mathbf{p}_f \rfloor \mathbf{e}_3$ . Then, from its sixth block row, we get  $\mathbf{A}_2 = -\lfloor \mathbf{v} \rfloor \mathbf{e}_3$ .

<sup>1</sup>The choice of the null space bases is done so that meaningful physical interpretations can be made.

Hence, the system's unobservable directions are spanned by

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}_{3 \times 2} & \frac{\partial s}{\partial \theta} \mathbf{C} \mathbf{e}_3 \\ \mathbf{0}_{3 \times 2} & -[\mathbf{v}] \mathbf{e}_3 \\ \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \end{bmatrix} & -[\mathbf{p}] \mathbf{e}_3 \\ \mathbf{0}_{3 \times 2} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 2} & \mathbf{0}_{3 \times 1} \\ \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \end{bmatrix} & -[\mathbf{p}_f] \mathbf{e}_3 \end{bmatrix}. \quad (29)$$

Therefore, the unobservable directions correspond to the system's planar motion over the desired horizontal plane and the landmark's position (first and second column of  $\mathbf{A}$ ), and the rotation around the gravity vector (third column of  $\mathbf{A}$ ). On the other hand, the orthogonal translation along the normal of the plane is observable. It should be noted that adding more point features does not change the observable and unobservable states of the system (special case is studied in [28]).

It is worth mentioning that, using the results in [27], one can intuitively prove the observability of the orthogonal distance between camera and the plane. Alternatively, the geometry of the features on the image plane can be used for the observability analysis [15, 23]. In contrast, our rigorous observability analysis in this section is advantageous as we used the *exact system measurement equation* in Eq. 5 and the *full state variables* including position, velocity, rotation, and IMU biases. Another important benefit of our analysis is that we not only prove that the orthogonal distance of the camera with respect to the horizontal plane is observable, but also, we derived all the unobservable directions of the system.

## 5 Motion Estimation Based on Observing Planar Features

In this section, we first describe INS error propagation model in Section 5.1. Then, we present the geometric constraint of the planar features as a function of the state variables in Section 5.2. In agreement with the derived geometric constraint, we define a set of state augmentation in Section 5.3. The camera measurement model is described in Section 5.4. Finally, the motion estimation algorithm is presented in Section 5.5.

### 5.1 INS Error States

The system discrete-time error state space model is derived based on the standard additive error definition for the position, velocity, and biases ( $\delta x \simeq x - \hat{x}$ ), where  $\simeq$  denotes an equality where only the dominant terms have been retained, and quaternion error for the rotational euler angles  $\theta$  ( $\delta \mathbf{q} \simeq [1 \quad \frac{\delta \theta}{2}]^\top$ ).<sup>2</sup> Then the errors can be concatenated into an error state vector as:

$$\delta \mathbf{x}^{\text{ins}} = \begin{bmatrix} \delta^C \boldsymbol{\theta}_G^\top & \delta^G \mathbf{v}_C^\top & \delta^G \mathbf{p}_C^\top & \delta \mathbf{b}_a^\top & \delta \mathbf{b}_g^\top \end{bmatrix}^\top. \quad (30)$$

During a short period of time  $\delta t$ , we describe the nonlinear INS propagation model in the form of a discrete-time linear state-space model

$$\delta \mathbf{x}_{k+1}^{\text{ins}} = \mathbf{F}_k^{\text{ins}} \delta \mathbf{x}_k^{\text{ins}} + \mathbf{G}_k^{\text{ins}} \mathbf{n}_k \in \mathbb{R}^{15}, \quad (31)$$

where  $\mathbf{F}_k^{\text{ins}}$  and  $\mathbf{G}_k^{\text{ins}}$  are known as discrete time state and system noise propagation matrices, respectively:

$$\mathbf{F}_k^{\text{ins}} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & -dt \mathbf{C}(\hat{\mathbf{C}} \hat{\mathbf{q}}_G) \\ dt [\mathbf{C}(\hat{\mathbf{C}} \hat{\mathbf{q}}_G) \hat{\mathbf{a}}]^\top & \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & dt \mathbf{C}(\hat{\mathbf{C}} \hat{\mathbf{q}}_G) & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & dt \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_3 & \mathbf{0}_{3 \times 3} & \mathbf{I}_3 \end{bmatrix} \quad (32)$$

and

$$\mathbf{G}_k^{\text{ins}} = \begin{bmatrix} \mathbf{0}_{3 \times 3} & -dt \mathbf{C}(\hat{\mathbf{C}} \hat{\mathbf{q}}_G) & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ dt \hat{\mathbf{C}} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & dt \mathbf{I}_3 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & dt \mathbf{I}_3 \end{bmatrix}. \quad (33)$$

$\mathbf{n}_k = [\mathbf{n}_a^\top \quad \mathbf{n}_g^\top \quad \mathbf{n}_{\delta a}^\top \quad \mathbf{n}_{\delta g}^\top]^\top$  is the process noise (assumed to be wide-sense stationary) with the corresponding diagonal covariance matrix  $\mathbf{Q} \in \mathbb{R}^{12 \times 12}$ ,  $\mathbf{C}(\hat{\mathbf{C}} \hat{\mathbf{q}}_G)$  is the estimated rotation matrix in which  $\hat{\omega}(t) = \omega_m(t) - \hat{\mathbf{b}}_g(t)$ , and  $\hat{\mathbf{a}} = \mathbf{a}_m - \hat{\mathbf{b}}_a$ .

### 5.2 Geometric Constraint of the Planar Features

Let us consider an arbitrary feature point  $f$ , which is in both  $\{C_\ell\}$  and  $\{C\}$  field of view. See Fig. 2 for graphical illustration of the used notations. Then the

<sup>2</sup>The advantage of using the quaternion error definition for the rotational euler angles is the direct use of error angle vectors  $\delta \theta$  instead of  $\delta \mathbf{q}_b^q$  in the error models.

position of  $f$  with respect to  $\{C_\ell\}$  and  $\{C\}$  are related as:

$${}^C \mathbf{p}_f = {}^C \mathbf{C}_{C_\ell} {}^{C_\ell} \mathbf{p}_f + {}^C \mathbf{t}_{C_\ell} \quad (34)$$

where

$${}^C \mathbf{C}_{C_\ell} \triangleq \mathbf{C}({}^C \mathbf{q}_G) \mathbf{C}({}^{C_\ell} \mathbf{q}_G)^\top,$$

$${}^C \mathbf{t}_{C_\ell} \triangleq \mathbf{C}({}^C \mathbf{q}_G)({}^G \mathbf{p}_{C_\ell} - {}^G \mathbf{p}_C).$$

Equation 34 enables us to relate an observation of  $\mathbf{p}_f$  in two different views to the motion between the views, parameterized by the rotation matrix  ${}^C \mathbf{C}_{C_\ell}$  and the translation vector  ${}^C \mathbf{t}_{C_\ell}$ . Thus the projection of  ${}^C \mathbf{p}_f$  is correlated with the motion of the system and therefore imposes constraints on the accumulated INS errors.

The feature point observed in the past frame  ${}^{C_\ell} \mathbf{p}_f$  is, however, an unknown nuisance parameter. Estimating the distance of  ${}^{C_\ell} \mathbf{p}_f$  to the camera is a poorly conditioned problem when it is large relative to the translation between views. However, by exploiting the fact that the point belongs to the defined horizontal plane,  ${}^{C_\ell} \mathbf{p}_f$  can be expressed in terms of camera pose  ${}^G \mathbf{p}_{C_\ell}$ ,  $\mathbf{C}({}^G \mathbf{q}_{C_\ell})$ , and observation  ${}^{C_\ell} \bar{\mathbf{z}}$ . Such a geometric relation is obtained from (14), as:

$${}^{C_\ell} \mathbf{p}_f = \frac{-\mathbf{e}_3^\top {}^G \mathbf{p}_{C_\ell}}{\mathbf{e}_3^\top \mathbf{C}({}^G \mathbf{q}_{C_\ell})^\top {}^{C_\ell} \bar{\mathbf{z}}} {}^{C_\ell} \bar{\mathbf{z}}. \quad (35)$$

Using Eq. 35, the general transformation (34) can now be rewritten for  ${}^{C_\ell} \mathbf{p}_f$  on the defined horizontal plane as:

$${}^C \mathbf{p}_f = {}^C \mathbf{C}_{C_\ell} \frac{-\mathbf{e}_3^\top {}^G \mathbf{p}_{C_\ell}}{\mathbf{e}_3^\top \mathbf{C}({}^G \mathbf{q}_{C_\ell})^\top {}^{C_\ell} \bar{\mathbf{z}}} {}^{C_\ell} \bar{\mathbf{z}} + {}^C \mathbf{t}_{C_\ell}. \quad (36)$$

### 5.3 State Augmentation

In order to impose the geometric constraint (36) to the previous camera poses and the normalized pixel coordinates, we define the following sets of auxiliary state variables:

- 1) Given a set of the  $L$  most recent images taken at different locations  $\{{}^G \mathbf{p}_{C_\ell}\}_{\ell=1}^L$ , we define the camera pose state vector

$$\mathbf{x}^{\text{cam}} = \left[ {}^G \mathbf{p}_{C_1}^\top \quad {}^{C_1} \mathbf{q}_G^\top \quad \dots \quad {}^G \mathbf{p}_{C_L}^\top \quad {}^{C_L} \mathbf{q}_G^\top \right]^\top, \quad (37)$$

and its corresponding error state vector as

$$\delta \mathbf{x}^{\text{cam}} = \left[ \delta {}^G \mathbf{p}_{C_1}^\top \quad \delta {}^{C_1} \theta_G^\top \quad \dots \quad \delta {}^G \mathbf{p}_{C_\ell}^\top \quad \delta {}^{C_\ell} \theta_G^\top \right]^\top. \quad (38)$$

Since the recorded pose errors are static, the discrete-time error state space model of the camera position is

$$\delta \mathbf{x}_{k+1}^{\text{cam}} = \delta \mathbf{x}_k^{\text{cam}}. \quad (39)$$

- 2) The feature points on the defined horizontal plane,  ${}^{C_\ell} \mathbf{z}$ , matched and observed in a past frame  $C_\ell$  are added as states. Hence,  $P_\ell$  points are augmented into the state vector

$${}^{C_\ell} \mathbf{m} = \left[ {}^{C_\ell} \mathbf{z}_1^\top \quad \dots \quad {}^{C_\ell} \mathbf{z}_{P_\ell}^\top \right]^\top, \quad (40)$$

that is added and removed for each incoming image depending on the feature point matches. A total number of  $P$  points are matched from  $L$  images and represented by  $\mathbf{m}$ .<sup>3</sup>

### 5.4 Camera Measurement

The observations for the current frame  $\{C\}$  are the feature points on the defined plane, matched to those observed in the previous frame  $\{C_\ell\}$ . Suppose feature point  $f_i$  has been matched. Then its measured coordinates on the normalized image plane is

$${}^C \mathbf{z}_i = \left[ \mathbf{I}_2 \quad \mathbf{0}_{2 \times 1} \right] \left( \mathbf{e}_3^\top {}^C \mathbf{p}_{f_i} \right)^{-1} {}^C \mathbf{p}_{f_i} + \mathbf{n}_i, \quad (41)$$

where the pixel noise  $\mathbf{n}_i$  is assumed to be zero-mean with covariance matrix  $\mathbb{C}_i = \sigma_{\text{pix}}^2 \mathbf{I}_2$ . The importance of parametrization of the feature points position  ${}^C \mathbf{p}_{f_i}$  (36) is appearing here since it relates the observation to the previous frame  ${}^{C_\ell} \mathbf{z}_i$  and camera pose,  $\{{}^G \mathbf{p}_{C_\ell}, {}^{C_\ell} \mathbf{q}_G\}$ , both of which are states.

By stacking all  $P_\ell$  feature point matches corresponding to view  $\ell$ , one obtains the nonlinear measurement equation

$$\begin{aligned} {}^C \mathbf{y}_{\text{pts}, \ell} &= \left[ {}^{C_\ell} \mathbf{z}_1^\top \quad \dots \quad {}^{C_\ell} \mathbf{z}_{P_\ell}^\top \right]^\top \\ &= \mathbf{h}(\delta \mathbf{x}^{\text{ins}}, \delta \mathbf{x}^{\text{cam}}, {}^{C_\ell} \mathbf{m}) + \mathbf{n}_{\text{pts}, \ell}. \end{aligned} \quad (42)$$

The feature points not matched to  $\ell$  are subsequently matched to a previous older frame  $\ell + 1$ , where  $\ell = 1, \dots, L$ , until at most  $P$  observations are collected.

<sup>3</sup>The proposed state augmentations partially follows our algorithm in [33].

The aggregate vector of dimension  $2P$  is denoted by  ${}^C\mathbf{y}_{\text{pts}}$ . Consequently, the error covariance matrix is  $\mathbb{C}_{\text{pts}} = \sigma_{\text{pix}}^2 \mathbf{I}_{2P}$ .

### 5.5 Estimation Framework

The joint state-space model for the INS and camera error states, along with the measurement (42) is compactly written as

$$\begin{aligned}\delta\mathbf{x}_{k+1} &= \mathbf{F}_k \delta\mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k \\ {}^C\mathbf{y}_k &= \mathbf{h}(\delta\mathbf{x}_k, \mathbf{m}_k) + \mathbf{n}_k,\end{aligned}\quad (43)$$

where  $\delta\mathbf{x}_k = [\delta\mathbf{x}_k^{\text{ins}} \ \delta\mathbf{x}_k^{\text{cam}}]$ . We model the statistics of the feature point vector  $\mathbf{m}$  (Section 5.4) based on the following simplifying assumptions: Its expected value is taken as the observed coordinates on the image plane,  $\hat{\mathbf{m}}$ , and its covariance matrix is  $\mathbb{C}_{\mathbf{m}} = \sigma_{\text{pix}}^2 \mathbf{I}_{2P}$ . Further,  $\mathbf{m}$  and  $\delta\mathbf{x}$  are assumed to be uncorrelated.

Since the process model is linear, the predicted errors are zero with error covariance matrix propagated as:

$$\mathbf{P}_{k+1}^- = \mathbf{F}_k \mathbf{P}_k^- \mathbf{F}_k^\top + \mathbf{G}_k \mathbf{Q} \mathbf{G}_k^\top, \quad (44)$$

where

$$\mathbf{F}_k \triangleq \begin{bmatrix} \mathbf{F}_k^{\text{ins}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{3L} \end{bmatrix} \text{ and } \mathbf{G}_k \triangleq \begin{bmatrix} \mathbf{G}_k^{\text{ins}} \\ \mathbf{0}_{3L \times 12} \end{bmatrix}.$$

When a new image is recorded an error state  $\delta^G \mathbf{p}_{C_\ell}$  and  $\delta^{C_\ell} \theta_G$  are augmented to  $\delta\mathbf{x}^{\text{cam}}$ , (38), and the error covariance matrix is updated as:

$$\mathbf{P}' = \begin{bmatrix} \mathbf{P} & \mathbf{P}\bar{\mathbf{T}}^\top \\ \bar{\mathbf{T}}\mathbf{P} & \bar{\mathbf{T}}\mathbf{P}\bar{\mathbf{T}}^\top \end{bmatrix} \quad (45)$$

where  $\bar{\mathbf{T}} = [\mathbf{I}_3 \ \mathbf{0}_{3 \times 3} \ \mathbf{I}_3 \ \mathbf{0}_{3 \times 6}]$  and  $\bar{\mathbf{T}} = [\bar{\mathbf{T}} \ \mathbf{0}]$ . When the buffer of  $L$  views is full, the oldest state and its corresponding rows and columns of the covariance matrix are decimated.

The UKF is set to perform a measurement update when features on the defined plane have been detected and matched. Based on the state-space model in Eq. 43, the UKF estimates the errors linearly as:

$$\delta\hat{\mathbf{x}}_k = \mathbf{K}_k ({}^C\mathbf{y}_k - {}^C\hat{\mathbf{y}}_k), \quad (46)$$

where  ${}^C\hat{\mathbf{y}}_k$  is the measurement prediction and  $\mathbf{K}_k$  is the Kalman gain matrix. An overview of the algorithm is given in Algorithm 1.

The joint statistics of  $\delta\mathbf{x}$  and  $\mathbf{m}$  are propagated as  $2N + 1$  sigma points, where  $N$  is the total number of states  $N = 15 + 6L + 2P$ . Let the joint state vector be denoted by  $\mathbf{x} = [\delta\mathbf{x}^\top \ \mathbf{m}^\top]^\top$  and  $\hat{\mathbf{x}}^- =$

$[\mathbf{0}^\top \ \hat{\mathbf{m}}^\top]^\top$  be the prediction with error covariance matrix  $\mathbf{P}^- \oplus \mathbb{C}_{\mathbf{m}}$ , where  $\oplus$  denotes the direct sum of matrices. The sigma points  $\mathcal{X}_j = [\delta\mathcal{X}_j^\top \ \mathcal{M}_j^\top]^\top$  are then generated using the matrix square-root of the joint error covariance matrix. By propagating the sigma points through the measurement (42), the correlations between the error states  $\delta\mathbf{x}$  and observations  ${}^C\mathbf{y}$  can be approximated [14]. These correlations form the basis of the Kalman gain.

**Algorithm 1** Proposed algorithm: Motion estimation approach in UKF framework for  $N = 15 + 6L + 2P$

```

1: Initialize  $\mathbf{P}_0^-$ 
2: for  $k = 0, \dots$  do
3:   Update INS state estimates
4:   if new image  $\exists$  then
5:     Extract and match feature points  $\{{}^C\mathbf{z}_i\}$ 
6:     Detect feature points on defined plane
7:   end if
8:   if  ${}^C\mathbf{y}_{\text{pts}}$  exists then
9:     %Generate sigma points and prediction:
10:     $\mathcal{X}_j = \hat{\mathbf{x}}_k^- \pm \eta \cdot \text{column}[(\mathbf{P}_k^- \oplus \mathbb{C}_{\mathbf{m}})^{1/2}]_j$ 
11:     $\mathcal{Y}_j = \mathbf{h}(\delta\mathcal{X}_j, \mathcal{M}_j)$ 
12:     ${}^C\hat{\mathbf{y}}_k = \sum_{j=0}^{2N} w_j^m \mathcal{Y}_j$ 
13:    %Measurement update:
14:     $\mathbf{C}_{\mathbf{e},k} = \sum_{j=0}^{2N} w_j^c (\mathcal{Y}_j - {}^C\hat{\mathbf{y}}_k) (\mathcal{Y}_j - {}^C\hat{\mathbf{y}}_k)^\top +$ 
        $\mathbb{C}_{\mathbf{n},k}$ 
15:     $\mathbf{D}_k = \sum_{j=0}^{2N} w_j^c \delta\mathcal{X}_j (\mathcal{Y}_j - {}^C\hat{\mathbf{y}}_k)^\top$ 
16:     $\mathbf{K}_k = \mathbf{D}_k \mathbf{C}_{\mathbf{e},k}^{-1}$ 
17:     $\delta\hat{\mathbf{x}}_k = \mathbf{K}_k ({}^C\mathbf{y}_k - {}^C\hat{\mathbf{y}}_k)$ 
18:     $\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{C}_{\mathbf{e},k} \mathbf{K}_k^\top$ 
19:    Use  $\delta\hat{\mathbf{x}}_k$  to correct state estimates
20:     $\mathbf{P} := \mathbf{P}_k$ 
21:   else
22:      $\mathbf{P} := \mathbf{P}_k^-$ 
23:   end if
24:   if new image  $\exists$  then
25:     Record view pose  ${}^G\hat{\mathbf{p}}_{C_\ell}$  and  ${}^{C_\ell}\hat{\mathbf{q}}_G$ 
26:     Decimate and augment error states  $\delta^G \mathbf{p}_{C_\ell}$  and
        $\delta^{C_\ell} \theta_G$ 
27:     Update  $\mathbf{P}$  correspondingly
28:   end if
29:   %Prediction:
30:    $\mathbf{P}_{k+1}^- = \mathbf{F}_k \mathbf{P}_k^- \mathbf{F}_k^\top + \mathbf{G}_k \mathbf{Q} \mathbf{G}_k^\top$ 
31: end for
```

The weights in the UKF are set as  $w_l^c = w_l^m = \frac{1}{2(N+\lambda)}$  for  $l = 1, \dots, 2N$ . For  $l = 0$ ,  $w_0^c = \frac{\lambda}{N+\lambda} + (1-\alpha^2+\beta)$  and  $w_0^m = \frac{\lambda}{N+\lambda}$ . Here  $\lambda = \alpha^2(N+\kappa)-N$ , with parameters set to  $\alpha = 0.1$ ,  $\beta = 2$  and  $\kappa = 0$  that also determine the spread of the sigma points through the weight  $\eta \triangleq \sqrt{N+\lambda}$ , [14].

## 6 Results

The key findings of our observability analysis and the performance of our proposed estimator are validated both with simulation and experimental data. The results indicate the accuracy and reliability of our proposed method where only a few features are used for motion estimation. In this section, we present some of the results from the simulation studies (Section 6.1) and experiments (Section 6.2).

### 6.1 Simulation results

For the simulation study, we consider the following setting: The sampling rate of the accelerometer and gyroscope output signals of the IMU,  $\mathbf{a}_m$  and  $\boldsymbol{\omega}_m$ , is set to 100 Hz. The IMU accelerometers and gyros biases are  $[2, 2, 2] \cdot 10^{-3} \text{ m/s}^2$  and  $[-4, 4, 2] \cdot 10^{-4} \text{ rad/s}$  and the standard deviations of the corresponding incremental noise processes,  $\mathbf{n}_{\delta a}$  and  $\mathbf{n}_{\delta g}$ , are  $6 \cdot 10^{-3}$  and  $3 \cdot 10^{-3}$ , respectively. For the camera, we consider the pinhole model, with the focal length of 833 mm, and sampling rate 10 Hz. Zero-mean white Gaussian noise with standard deviation of  $\sigma_{\text{pixel}} = 2 \text{ pixel}$  is added to the projected image. We set maximum number of tracked feature points between successive images to be 10 over the whole simulation process between two consecutive images.

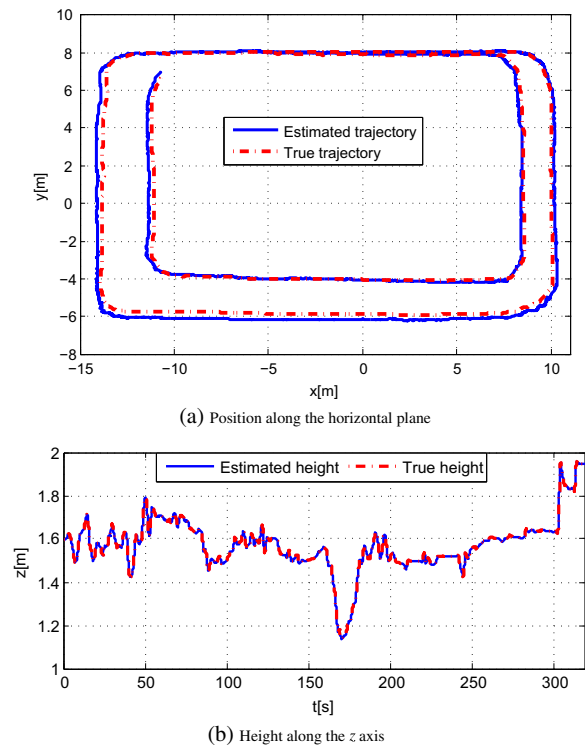
The estimated trajectory along with the ground truth are plotted in Fig 4. The simulated trajectory is for about 360 s and final estimated error in the position is about  $[0.28, -0.31, 0.019]$  meters.

Furthermore to validate the accuracy of the estimator, the estimated errors of the position, attitude, and velocity along with the  $\pm\sigma$ -bounds are plotted in Fig 5. The  $\sigma$ -bounds are computed from the corresponding diagonal elements of the filter's error covariance matrix that provides a representation of its estimate uncertainty. The results indicates the capacity of the filter to constrain the rate of the error growth.

Moreover, the estimated value of the accelerometer and gyro biases are plotted in Fig 6. The results show that the estimated value of the biases both for the accelerometers and gyroscopes are approximately stable over time.

### 6.2 Experimental Validation

Figure 7a shows the hardware used for this experiment. The camera is an AVT Guppy monochrome



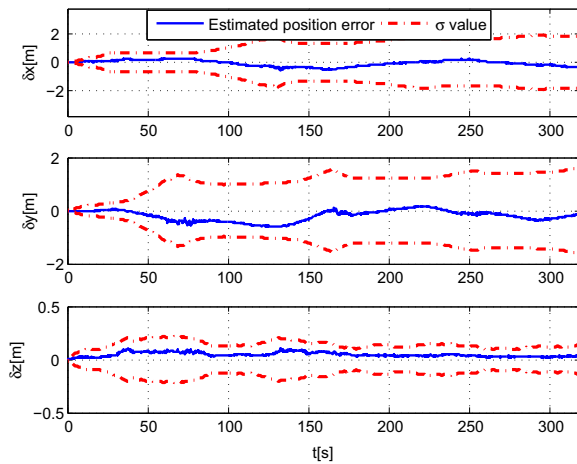
**Fig. 4** Estimated trajectory over the ground truth along  $x$ ,  $y$  and  $z$ -axes over 320 s

camera that is rigidly mounted on top of a MicroStrain 3DMGX2 IMU with sampling rate 250 Hz. Images are captured with resolution  $752 \times 480$  pixels, the camera sampling rate is 10 Hz. The camera internal calibration parameters are estimated using [1], and the IMU-camera calibration parameters are estimated from [29]. For the feature extraction and matching, the MATLAB Computer Vision Toolbox implementation of SURF is used. The ground plane features are selected using a predefined box in the lower part of the image while the outliers are rejected using the filters estimates over time. Detail studies about the ground plane feature detection and outlier removal are given in [31, 32]. Time synchronization and temporal ordering between the IMU and the camera measurements are based on methods in [26].

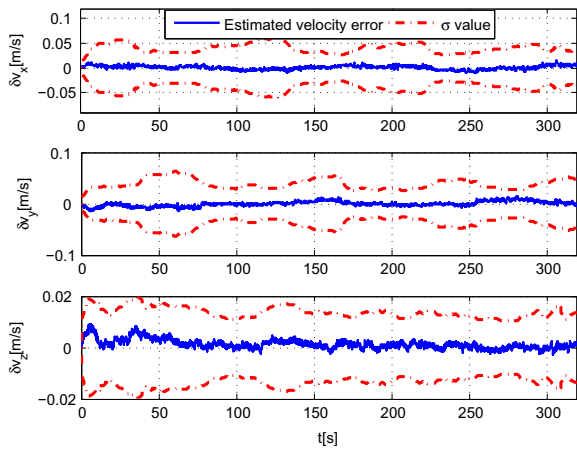
The initialization of the IMU gyro and accelerometer biases and their corresponding covariance values in the filter was done by placing the IMU in a static position for a few seconds prior to each experiment. The mean and variance of biases during the stationary phase are then used as initial values in the filter.

We consider ground plane as our desired horizontal plane. Such that we can run the experiment





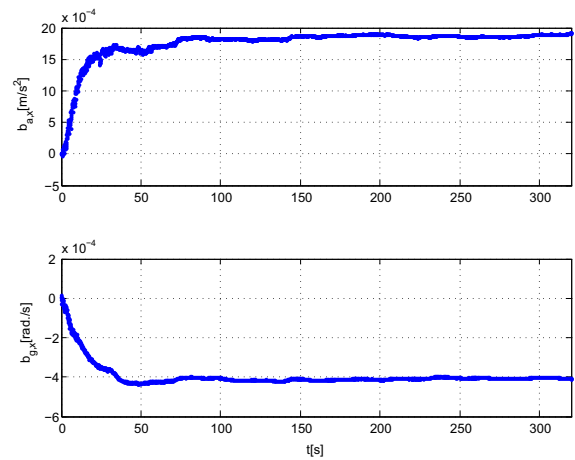
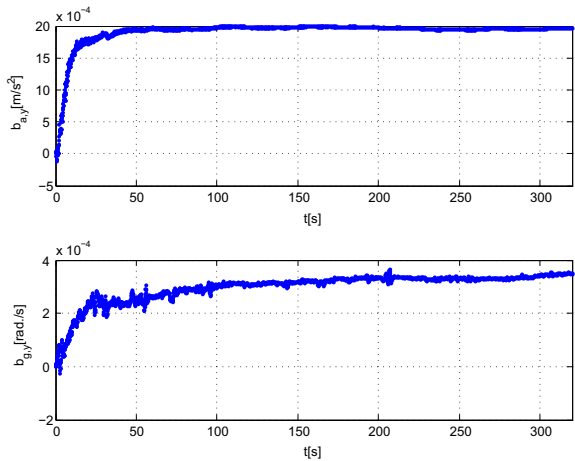
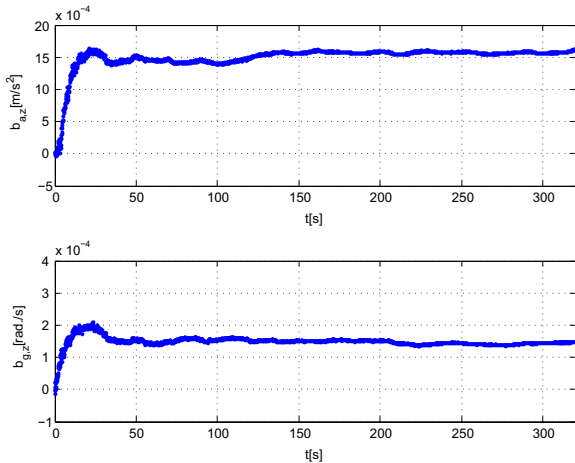
(a) Position



(b) Velocity

**Fig. 5** Estimated errors for the position and velocity in the body frame with corresponding  $\sigma$ -bounds

over larger area and for longer period. To evaluate the performance of the proposed method in real scenario, a controlled test environment was set up in an underground hall (R1) at the university of KTH [3]. Figure 7b depicts some sample images from the test environment. The underground location of the test environment provides a controlled level of radio interference for a reference commercial ultra-wideband (UWB) system [17] used as a ground truth for evaluating the results, more details on the accuracy of the UWB system are given in [3]. When line-of-sight is lost, however, the UWB position estimates become unreliable. During the experiments all data were stored on a computer and processing was done off-line.

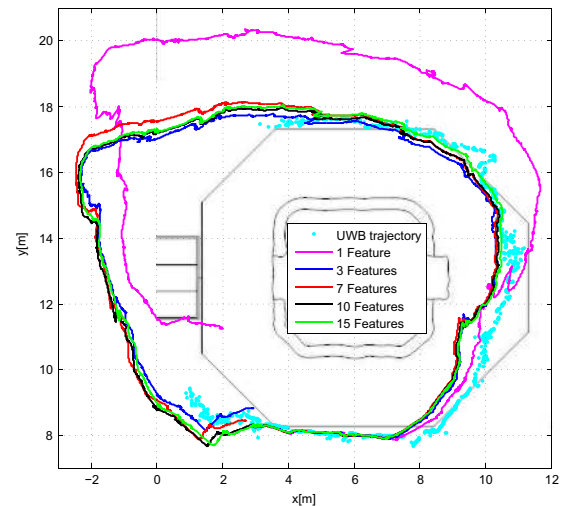
(a) Biases along the  $x$  axis(b) Biases along the  $y$  axis(c) Biases along the  $z$  axis

**Fig. 6** Estimated biases in accelerometers and gyroscopes of the IMU over 320s

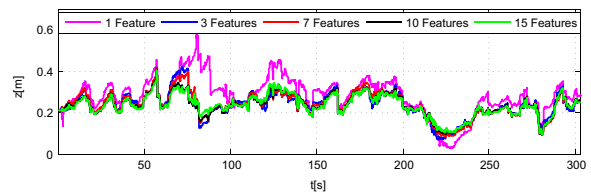
Figure 8a depicts an example of estimated trajectories for an experiment along with the UWB data used as the ground truth over the map of the test environment. In this experiment 3081 image frames were recorded within 304 seconds. The IMU and camera were placed at the height of about 21 cm on a trolley, moving in a closed loop trajectory. The system was initialized at the measured height of 21 cm above the ground. The roll and pitch were estimated using the gravitational force during first samples when the system was stationary [4]. In the implementation, the number of views in memory was set to  $L = 5$ .

To evaluate the performance of the estimator, we have plotted the estimated trajectory under different settings for the maximum number of matched features,  $\{1, 3, 7, 10, 15\}$ . The results show that the UKF is capable of drastically reducing the growth rate of the accumulated INS errors over the experiment. The trajectories follow the UWB position estimates reasonably well, although the UWB sensors were not in line-of-sight during a substantial part of the trajectory. Moreover, similar patterns are observed for different maximum number of matched features, especially for  $\{3, 7, 10, 15\}$ , which in turn show that the proposed system is able to achieve a good performance even for low number of features.

Furthermore, the estimated height of the mobile system is shown in Fig. 8b, for the corresponding estimated trajectories in Fig. 8a. As can be seen in the figure, the system is capable of good error correction that leads to a quite accurate estimate for the height. Also, note that the estimated height has briefly deviated from the reference height (21 cm) at around 210–230 seconds, which can be explained by considering the potential problems in accurately detecting



(a) Estimated position along the  $x$  and  $y$  axes



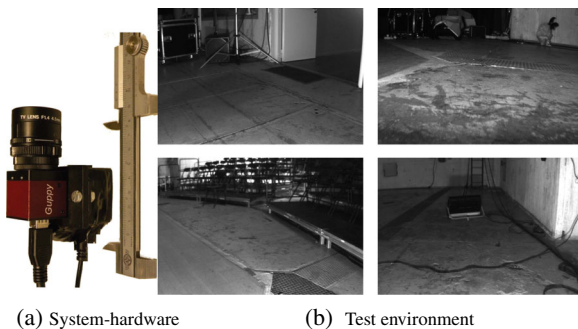
(b) Estimated height along the  $z$  axis

**Fig. 8** The estimated trajectory of the IMU along with the UWB reference data, overlaid on the map of the test environment. The estimated trajectories are plotted for different maximum number of tracked features,  $\{1, 3, 7, 10, 15\}$

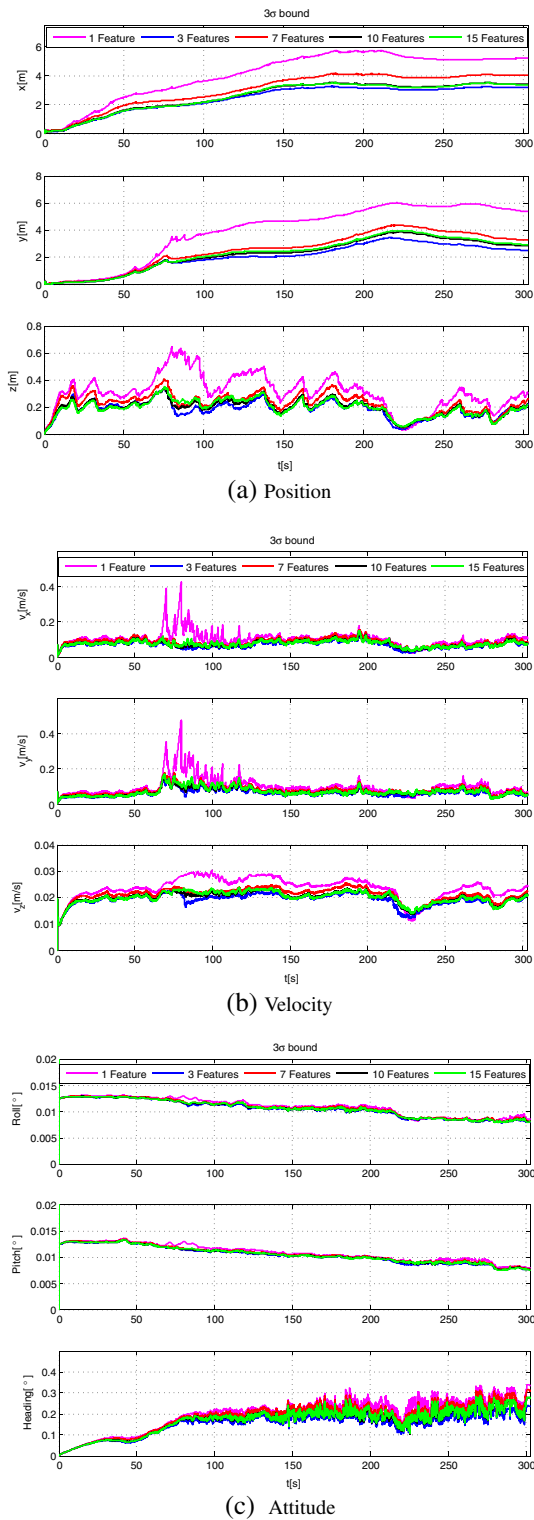
and matching the feature points. However, this deviation does not last long and the algorithm corrects the height quickly.

To provide a representation of the estimation uncertainty the  $3\sigma$ -bounds for the error in the camera position, orientation, and velocity along the three axes are shown in Fig 9. As it is depicted in Fig. 9a, the uncertainty of the position along the  $x$  and  $y$  axes grows slowly since the system has no access to the absolute position update. The behavior of the  $3\sigma$ -bounds for the position uncertainty along the  $z$  axis confirms the findings of our observability analysis, i.e., the height is observable. These results indicate that the proposed method is capable of achieving a rough estimate of the states even when maximum number of features is set to one. However, in this case the errors have higher uncertainties, especially for the unobservable states.

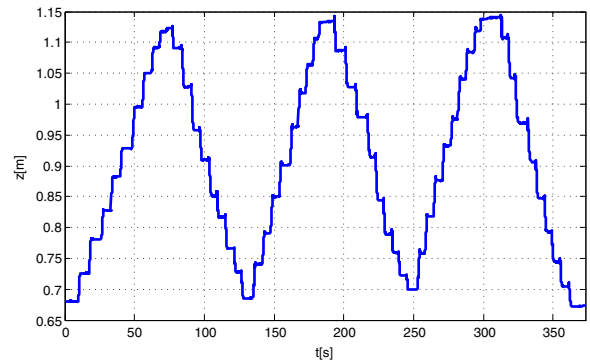
Based on our theoretical results for the system unobservable modes, we expect increasing uncertainties of the estimators along the unobservable modes.



**Fig. 7** **a** An AVT Guppy monochrome camera, with resolution  $752 \times 480$  that is rigidly mounted on top of a MicroStrain 3DMGX2 IMU. **b** Sample images from the test environment



**Fig. 9** The  $3\sigma$ -bounds for the error in **a** position, **b** velocity, and **c** attitude. The  $\sigma$  values are the square root of the corresponding diagonal elements of the states error covariance matrix



**Fig. 10** Estimated height along the z axis, where the x and y positions were fixed

However, as it is illustrated in Fig. 9a for the positions along the x and y axis and heading, the estimation uncertainties are decreasing at some regions. This behavior, known as estimator inconsistency, has been recently studied in [8, 10, 16] for a VINS. This estimator inconsistency might be explained by considering that the filter employs a linearized state-space model, where the unobservable subspace of the obtained system has a lower dimension than the unobservable subspace of the underlying nonlinear system [16]. Hence, the estimator gains spurious information from the measurements and therefore incorrectly reduces the estimated covariance.

As a future work, a further study with more focus on estimator inconsistency is therefore suggested for our system.

Moreover, to evaluate the performance of our proposed method for the height estimation, we performed the following experiment: the IMU-camera was placed on a table, and without having any motion along the x – y plane the table was lifted up and down using a step motor. Figure 10 illustrates the estimated height of the system for about 373 seconds. The initial and the final measured height was approximately 68 cm and height of the peaks was in the range of  $113 \pm 2$  cm. It is worth mentioning that the estimated error along the x and y directions was in the order of millimeter. For this estimate, the maximum number of features was set to 7 between two successive images.

For the reported experiments in this section, we have tried to capture different types of motions, such as sharp and smooth turns, constant speed and sudden acceleration. Moreover, in all the experiments, for some period of time the sensors were left stationary, without any movement. The stationary periods are

well depicted in Fig. 10. Additionally, in a separate experiment, to investigate the stability of the system under stationary conditions over a long period of time, we evaluated the performance of the system for a case where the system was left stationary on a table without any movement. The results show that after about 30 min the error along the  $x$  and  $y$  axes were about 50 cm and along the  $z$  axis it was about 150 cm. The estimated biases of the accelerometer gyro for this experiment are plotted in Fig. 11, where it can be seen that the biases are quite stable for a long period of time.

It is worth mentioning that, for this case where there is no excitation in the system (linear acceleration and rotational velocity), the observability analysis is not strictly valid since the required conditions regarding existence of excitation are not satisfied.

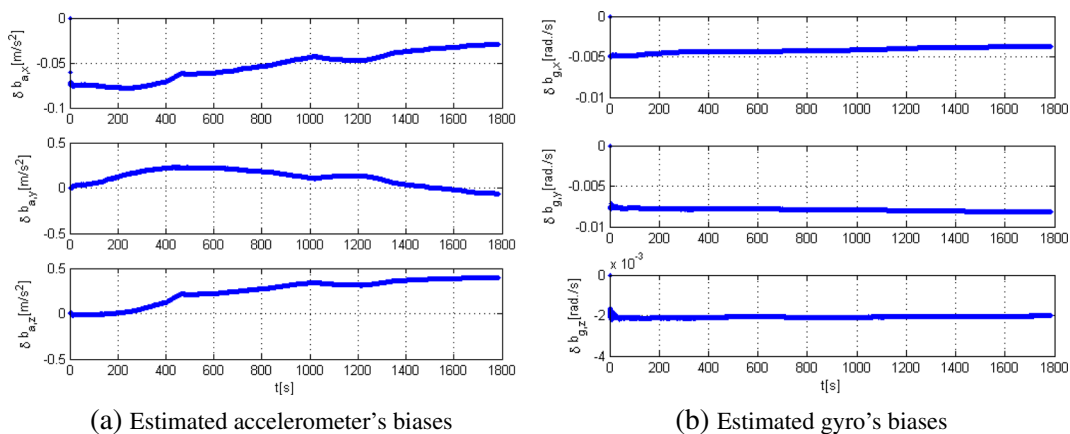
As was mentioned earlier, the focus of this paper is mainly theoretical in nature and the considered simulations and experiments target the validation of the main results, rather than the full behavior of an UAV close to landing. Such system performance evaluation is of importance, but the outcome will also depend on the physical platform and its configuration. Accordingly, such full-scale experiments are beyond the scope of the paper.

## 7 Discussion

Imposing the IMU and camera data to the geometrical derivations in Section 5, we develop a 6-DoF motion estimation approach. Moreover, we have carefully implemented and analyzed the performance of

our proposed system. The achieved results validate both the findings of our observability analysis in Section 4 and the accuracy of the developed nonlinear filter in Section 5. Compared to the state-of-the-art VINS ego-motion estimation [25] or SLAM-type of approaches, our developed closed-form measurement model significantly reduces the filter computational complexity. Since the length of the state vector in our method is  $15 + 6L$  (ignoring the feature points positions in the image plane which are just *nuisance parameters*). Additionally, our measurement equation is independent of movements in the system. That is, the system provides reliable information even when the camera is static, and there is no motion in the system. Because the validity of Eq. 36 is preserved even when the camera does not move or is subject to very slow motions.

In this work, we focus on the problem of 6-DoF IMU-camera motion estimation when the camera observations are features on a horizontal plane. It should be noted that, the choice of selecting features on any planar surface whose normal is known does not influence the results. That is, for any arbitrary planar surface, we first need to align the virtual camera optical axis with the normal of the plane. Then, we have to calculate the corresponding rotation between the virtual camera coordinate frame and the global reference frame,  $C(Vq_G)$ . It should be noted that, changing the direction of the plane normal affects the observability properties of the system. That is, the observable mode of the translation,  $Cp_G$ , is always the orthogonal translation of the camera along the normal of the plane. For instance, for the entire horizontal planes only the vertical translation is observable while



**Fig. 11** Estimated accelerometer and gyro biases, where the system was left stationary for about 1784 seconds

for vertical planes depending on their normal directions planar translation along the  $x$  or  $y$  axis becomes observable.

For the observability analysis, we consider the full INS parameter vector (including position, velocity, rotation, and inertial sensor biases) as well as the 3D position of the observed feature point. It is straightforward to show that by concatenating the IMU-camera calibration parameters and the value of the gravity acceleration into our state vector, while our current results for the system is preserved, these additional parameters are observable, as it is shown in [15, 23].

## 8 Conclusion

In this paper, we have introduced a visual inertial navigation system, which is based on tracking features on a horizontal plane. For this nonlinear system, we first study its observability properties. We show that compared to the general VINS, where features are not constrained to be all planar, the orthogonal translation of the camera to the plane is observable. Furthermore, we have derived all the observable and unobservable directions of the system using only one feature point as the camera measurement. Secondly, we seek to the problem of ego-motion estimation for the proposed system. We solve the 6-DoF VINS motion estimation using the state space formulation in the unscented Kalman filter framework. The key advantages of our solution is the proposed new measurement equation, which reflects both the geometric of the planar features and the state of the system. Finally, we have experimentally verified the key features of our observability analysis and motion estimation. The experimental results are provided under different conditions where accurate results with low computational complexity are achieved. Accordingly, the method presented in this paper forms a basis for many applications, e.g., accurate UAV landing systems.

**Acknowledgment** The authors would like to thank Prof. Stergios Roumeliotis for the discussions on the observability techniques that are used in this paper.

## Appendix A

We study the observability of system (12) (or specifically for our system (25) and (26)) by the algebraic

test. In the algebraic test, the gradients of the Lie derivatives of the measurement functions are first derived.

To prove that matrix  $\mathcal{E}$  is full column rank, it suffices to show that a subset of its rows, whose dimension is the same or larger than the number of columns, is linearly independent. To show this, it is sufficient to construct a sub-matrix of  $\mathcal{E}$  by selecting the minimum number of Lie derivative gradients of the measurement functions that leads to a full column rank matrix.

In the following, we compute only those Lie derivatives of  $\bar{\mathbf{h}}$  whose spans are used to prove that  $\mathcal{E}$  is of full column rank.

- The zeroth order Lie derivatives of the measurement function is

$$\mathcal{L}^0 \bar{\mathbf{h}}_1 = (\beta_1 \mathbf{I}_3 + \beta_2 \beta_3^\top) \bar{\mathbf{z}} = \mathbf{0}_{3 \times 1}$$

Then, its gradient (the span) is

$$\begin{aligned} \nabla \mathcal{L}^0 \bar{\mathbf{h}}_1 &= \frac{\partial \bar{\mathbf{h}}_1}{\partial \beta} \\ &= [\bar{\mathbf{z}} \ \beta_3^\top \bar{\mathbf{z}} \mathbf{I}_3 \ \beta_2 \bar{\mathbf{z}}^\top \ \mathbf{0}_{3 \times 1} \ \mathbf{0}_{3 \times 3} \ \mathbf{0}_{3 \times 3} \ \mathbf{0}_{3 \times 3}] \end{aligned}$$

- The first order Lie derivatives of  $\bar{\mathbf{h}}_1$  with respect to  $\mathbf{g}_0$ , and  $\mathbf{g}_{1i}$  are computed, respectively, as:

$$\begin{aligned} \mathcal{L}_{\mathbf{g}_0}^1 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}^0 \bar{\mathbf{h}}_1 \mathbf{g}_0 = \bar{\mathbf{z}} \beta_4 + \beta_3^\top \bar{\mathbf{z}} (-[\beta_2] \beta_5 - \beta_6) \\ &\quad - \beta_2 \bar{\mathbf{z}}^\top [\beta_3] \beta_5 \\ \mathcal{L}_{\mathbf{g}_{1i}}^1 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}^0 \bar{\mathbf{h}}_1 \mathbf{g}_{1i} = \beta_3^\top \bar{\mathbf{z}} [\beta_2] \mathbf{e}_i + \beta_2 \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_i \end{aligned}$$

and their corresponding gradients are given by

$$\begin{aligned} \nabla \mathcal{L}_{\mathbf{g}_0}^1 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_0}^1 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= [\mathbf{0}_{3 \times 1} \ \Pi_2 \ \Pi_3 \ \bar{\mathbf{z}} \ \Pi_5 \ -\beta_3^\top \bar{\mathbf{z}} \mathbf{I}_3 \ \mathbf{0}_{3 \times 3}] \end{aligned}$$

where

$$\begin{aligned}\Pi_2 &= \beta_3^\top \bar{z} \lfloor \beta_5 \rfloor - \bar{z}^\top \lfloor \beta_3 \rfloor \beta_5 \mathbf{I}_3, \\ \Pi_3 &= (-\lfloor \beta_2 \rfloor \beta_5 - \beta_6) \bar{z}^\top + \beta_2 \bar{z}^\top \lfloor \beta_5 \rfloor, \\ \Pi_5 &= -\beta_3^\top \bar{z} \lfloor \beta_2 \rfloor - \beta_2 \bar{z}^\top \lfloor \beta_3 \rfloor.\end{aligned}$$

$$\begin{aligned}\nabla \mathcal{L}_{\mathbf{g}_{li}}^1 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_{li}}^1 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} - \beta_3^\top \bar{z} \lfloor \mathbf{e}_i \rfloor \\ + \bar{z}^\top \lfloor \beta_3 \rfloor \mathbf{e}_i \mathbf{I}_3 \lfloor \beta_2 \rfloor \mathbf{e}_i \bar{z}^\top \\ - \beta_2 \bar{z}^\top \lfloor \mathbf{e}_i \rfloor \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \end{bmatrix}\end{aligned}$$

- Second order Lie derivatives, and their corresponding gradients are:<sup>4</sup>

$$\begin{aligned}\mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0}^2 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}_{\mathbf{g}_{li}}^1 \bar{\mathbf{h}}_1 \mathbf{g}_0 \\ &= (-\beta_3^\top \bar{z} \lfloor \mathbf{e}_i \rfloor + \bar{z}^\top \lfloor \beta_3 \rfloor \mathbf{e}_i \mathbf{I}_3) \\ &\quad \times (-\lfloor \beta_2 \rfloor \beta_5 - \beta_6) \\ &\quad - (\lfloor \beta_2 \rfloor \mathbf{e}_i \bar{z}^\top - \beta_2 \bar{z}^\top \lfloor \mathbf{e}_i \rfloor) \lfloor \beta_3 \rfloor \beta_5 \\ \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}}^2 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}_{\mathbf{g}_0}^1 \bar{\mathbf{h}}_1 \mathbf{g}_{2i} = \bar{z} \beta_3^\top \mathbf{e}_i - \beta_3^\top \bar{z} \mathbf{e}_i \\ \nabla \mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0}^2 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0}^2 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} \tilde{\Pi}_{2i} \tilde{\Pi}_{3i} \mathbf{0}_{3 \times 1} \tilde{\Pi}_{5i} \beta_3^\top \bar{z} \lfloor \mathbf{e}_i \rfloor \\ - \bar{z}^\top \lfloor \beta_3 \rfloor \mathbf{e}_i \mathbf{I}_3 \mathbf{0}_{3 \times 3} \end{bmatrix} \\ \nabla \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}}^2 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}}^2 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} \bar{z} \mathbf{e}_i^\top \\ - \mathbf{e}_i \bar{z}^\top \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \end{bmatrix}\end{aligned}$$

<sup>4</sup> $\tilde{\Pi}_{ij}$  and  $\tilde{\Pi}_{ji}$  will be removed later on through Gaussian elimination, thus we do not show their explicit expressions here.

- Third order Lie derivatives, and their corresponding gradients:

$$\begin{aligned}\mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0\mathbf{g}_0}^3 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0}^2 \bar{\mathbf{h}}_1 \mathbf{g}_0 \\ &= \tilde{\Pi}_{2i} (-\lfloor \beta_2 \rfloor \beta_5 - \beta_6) \\ &\quad + \tilde{\Pi}_{3i} (-\lfloor \beta_3 \rfloor \beta_5) \\ &\quad + (\beta_3^\top \bar{z} \lfloor \mathbf{e}_i \rfloor - \bar{z}^\top \lfloor \beta_3 \rfloor \mathbf{e}_i \mathbf{I}_3) \\ &\quad \times (-\lfloor \beta_6 \rfloor \beta_5 + \beta_3 - \beta_7)) \\ \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1j}}^3 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}}^2 \bar{\mathbf{h}}_1 \mathbf{g}_{1j} \\ &= (\bar{z} \mathbf{e}_i^\top - \mathbf{e}_i \bar{z}^\top) \lfloor \beta_3 \rfloor \mathbf{e}_i \\ \nabla \mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0\mathbf{g}_0}^3 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_{li}\mathbf{g}_0\mathbf{g}_0}^3 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} \tilde{\Pi}_{2i} \tilde{\Pi}_{3i} \tilde{\Pi}_{4i} \tilde{\Pi}_{5i} \tilde{\Pi}_{6i} \\ - \beta_3^\top \bar{z} \lfloor \mathbf{e}_i \rfloor + \bar{z}^\top \lfloor \beta_3 \rfloor \mathbf{e}_i \mathbf{I}_3 \end{bmatrix} \\ \nabla \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1j}}^3 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1j}}^3 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} - (\bar{z} \mathbf{e}_i^\top - \mathbf{e}_i \bar{z}^\top) \lfloor \mathbf{e}_j \rfloor \\ \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \end{bmatrix}\end{aligned}$$

for  $i = j$

$$\begin{aligned}\nabla \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1i}}^3 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1i}}^3 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} \mathbf{e}_i \bar{z}^\top \lfloor \mathbf{e}_i \rfloor \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \end{bmatrix}\end{aligned}$$

- Fourth order Lie derivatives, and their corresponding gradients:

$$\begin{aligned}\mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1i}\mathbf{g}_0}^4 \bar{\mathbf{h}}_1 &= \nabla \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1i}}^3 \bar{\mathbf{h}}_1 \mathbf{g}_0 \\ &= \mathbf{e}_i \bar{z}^\top \lfloor \mathbf{e}_i \rfloor \lfloor \beta_3 \rfloor \beta_5 \\ \nabla \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1i}\mathbf{g}_0}^4 \bar{\mathbf{h}}_1 &= \frac{\partial \mathcal{L}_{\mathbf{g}_0\mathbf{g}_{2i}\mathbf{g}_{1i}\mathbf{g}_0}^4 \bar{\mathbf{h}}_1}{\partial \beta} \\ &= \begin{bmatrix} \mathbf{0}_{3 \times 1} \mathbf{0}_{3 \times 3} - \mathbf{e}_i \bar{z}^\top \lfloor \mathbf{e}_i \rfloor \lfloor \beta_5 \rfloor \\ \mathbf{0}_{3 \times 1} \mathbf{e}_i \bar{z}^\top \lfloor \mathbf{e}_i \rfloor \lfloor \beta_3 \rfloor \mathbf{0}_{3 \times 3} \mathbf{0}_{3 \times 3} \end{bmatrix}\end{aligned}$$



Stacking together the computed span of the Lie derivatives, a subset of the observability matrix  $\mathcal{E}$  is constructed as:

$$\begin{bmatrix} \nabla \mathcal{L}_{g_0}^0 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0}^1 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{11}}^1 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{12}}^1 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{13}}^1 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{11}g_0}^2 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{12}g_0}^2 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{13}g_0}^2 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0g_{21}}^2 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0g_{22}}^2 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0g_{23}}^2 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{11}g_0g_0}^3 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{12}g_0g_0}^3 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_{13}g_0g_0}^3 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0g_{21}g_{11}g_0}^4 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0g_{22}g_{12}g_0}^4 \bar{\mathbf{h}}_1 \\ \nabla \mathcal{L}_{g_0g_{23}g_{13}g_0}^4 \bar{\mathbf{h}}_1 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{z}} & \beta_3^\top \bar{\mathbf{z}} \mathbf{I}_3 & \beta_2^\top \bar{\mathbf{z}}^\top & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \Pi_2 & \Pi_3 & \bar{\mathbf{z}} & \Pi_5 & -\beta_3^\top \bar{\mathbf{z}} \mathbf{I}_3 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_1] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_1 \mathbf{I}_3 & [\beta_2] \mathbf{e}_1 \bar{\mathbf{z}}^\top - \beta_2^\top \bar{\mathbf{z}}^\top [\mathbf{e}_1] & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_2] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_2 \mathbf{I}_3 & [\beta_2] \mathbf{e}_2 \bar{\mathbf{z}}^\top - \beta_2^\top \bar{\mathbf{z}}^\top [\mathbf{e}_2] & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_3] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_3 \mathbf{I}_3 & [\beta_2] \mathbf{e}_3 \bar{\mathbf{z}}^\top - \beta_2^\top \bar{\mathbf{z}}^\top [\mathbf{e}_3] & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{21} & \tilde{\Pi}_{31} & \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{51} & \beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_1] - \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_1 \mathbf{I}_3 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{22} & \tilde{\Pi}_{32} & \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{52} & \beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_2] - \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_2 \mathbf{I}_3 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{23} & \tilde{\Pi}_{33} & \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{53} & \beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_3] - \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_3 \mathbf{I}_3 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \bar{\mathbf{z}} \mathbf{e}_1^\top - \mathbf{e}_1 \bar{\mathbf{z}}^\top & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \bar{\mathbf{z}} \mathbf{e}_2^\top - \mathbf{e}_2 \bar{\mathbf{z}}^\top & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \bar{\mathbf{z}} \mathbf{e}_3^\top - \mathbf{e}_3 \bar{\mathbf{z}}^\top & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{21} & \tilde{\Pi}_{31} & \tilde{\Pi}_{41} & \tilde{\Pi}_{51} & \tilde{\Pi}_{61} & -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_1] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_1 \mathbf{I}_3 \\ \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{22} & \tilde{\Pi}_{32} & \tilde{\Pi}_{42} & \tilde{\Pi}_{52} & \tilde{\Pi}_{62} & -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_2] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_2 \mathbf{I}_3 \\ \mathbf{0}_{3 \times 1} & \tilde{\Pi}_{23} & \tilde{\Pi}_{33} & \tilde{\Pi}_{43} & \tilde{\Pi}_{53} & \tilde{\Pi}_{63} & -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_3] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_3 \mathbf{I}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & -\mathbf{e}_1 \bar{\mathbf{z}}^\top [\mathbf{e}_1] [\beta_3] & \mathbf{0}_{3 \times 1} & \mathbf{e}_1 \bar{\mathbf{z}}^\top [\mathbf{e}_1] [\beta_3] & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & -\mathbf{e}_2 \bar{\mathbf{z}}^\top [\mathbf{e}_2] [\beta_3] & \mathbf{0}_{3 \times 1} & \mathbf{e}_2 \bar{\mathbf{z}}^\top [\mathbf{e}_2] [\beta_3] & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & -\mathbf{e}_3 \bar{\mathbf{z}}^\top [\mathbf{e}_3] [\beta_3] & \mathbf{0}_{3 \times 1} & \mathbf{e}_3 \bar{\mathbf{z}}^\top [\mathbf{e}_3] [\beta_3] & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix}.$$

One way to show that matrix (43) is full rank is through performing Gauss elimination. The elimination is performed based on the knowledge that the following matrices are of full column rank:

$$\mathbf{A} = \begin{bmatrix} -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_1] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_1 \mathbf{I}_3 \\ -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_2] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_2 \mathbf{I}_3 \\ -\beta_3^\top \bar{\mathbf{z}} [\mathbf{e}_3] + \bar{\mathbf{z}}^\top [\beta_3] \mathbf{e}_3 \mathbf{I}_3 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \bar{\mathbf{z}} \mathbf{e}_1^\top - \mathbf{e}_1 \bar{\mathbf{z}}^\top \\ \bar{\mathbf{z}} \mathbf{e}_2^\top - \mathbf{e}_2 \bar{\mathbf{z}}^\top \\ \bar{\mathbf{z}} \mathbf{e}_3^\top - \mathbf{e}_3 \bar{\mathbf{z}}^\top \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{e}_1 \bar{\mathbf{z}}^\top [\mathbf{e}_1] [\beta_3] \\ \mathbf{e}_2 \bar{\mathbf{z}}^\top [\mathbf{e}_2] [\beta_3] \\ \mathbf{e}_3 \bar{\mathbf{z}}^\top [\mathbf{e}_3] [\beta_3] \end{bmatrix}.$$

Performing elementary column and row operations on (43), we finally get:

$$\begin{bmatrix} 1 & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{1 \times 1} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & 1 & \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{I} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{I}_3 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{I}_3 & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{I}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 1} & \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \end{bmatrix},$$

which clearly is a full column rank matrix.

## References

1. Bouguet, J.Y.: Camera calibration toolbox. Computaion vision at Caltech, [www.vision.caltech.edu/bouguetj/](http://www.vision.caltech.edu/bouguetj/) (last accessed Mar. 2012)
2. Conrad, D., DeSouza, G.N.: Homography-based ground plane detection for mobile robot navigation using a modified em algorithm. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Anchorage, Alaska, pp. 910–915 (2010)
3. De Angelis, A., Händel, P., Rantakokko, J.: Measurement report. laser total station campaign in KTH R1 for ubisense system accuracy evaluation : Laser total station campaign in KTH R1 for ubisense system accuracy evaluation. In: Technical Report (2012)
4. Farrell, J.A., Barth, M.: Global Positioning System, Inertial Navigation and Integration. McGraw-Hill Companies (1999)
5. Guo, C.X., Roumeliotis, S.I.: IMU-RGBD camera extrinsic calibration: Observability analysis and consistency improvement. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany (2013)
6. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000). ISBN: 0521623049
7. Hermann, R., Krener, A.: Nonlinear controllability and observability. IEEE Trans. Autom. Control **22**(4), 728–740 (1977)

8. Hesch, J.A., Kottas, D.G., Bowman, S.L., Roumeliotis, S.I.: Towards consistent vision-aided inertial navigation. In: 10th International Workshop on the Algorithmic Foundations of Robotics, Cambridge, Massachusetts (2012)
9. Hesch, J.A., Kottas, D.G., Bowman, S.L., Roumeliotis, S.I.: Camera-IMU-based localization: Observability analysis and consistency improvement. *Int. J. Robot. Res.* **33**(1), 182–201 (2014)
10. Hesch, J.A., Kottas, D.G., Bowman, S.L., Roumeliotis, S.I.: Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Trans. Robot.* **30**(1), 158–176 (2014)
11. Hide, C., Botterill, T., Andreotti, M.: Low cost vision-aided IMU for pedestrian navigation. In: IEEE Conference on Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), pp. 1–7 (2010)
12. Jones, E., Vedaldi, A., Soatto, S.: Inertial structure from motion and autocalibration. In: Workshop on Dynamical Vision (2007)
13. Jones, E.S., Soatto, S.: Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *Int. J. Robot. Res.* **30**(4), 407–430 (2011)
14. Julier, S.J., Uhlmann, J.K.: A new extension of the Kalman filter to nonlinear systems. In: Proceedings of Signal Processing, Sensor fusion, and Target Recognition, vol. 4, pp. 182–193 (1997)
15. Kelly, J., Sukhatme, G.S.: Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *Int. J. Robot. Res.* **30**(1), 56–79 (2011)
16. Li, M., Mourikis, A.I.: Improving the accuracy of EKF-based visual-inertial odometry. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, Minnesota, pp. 828–835 (2012)
17. Ltd., U.: The ubisense precise real-time location system - series 7000 sensor (<http://www.ubisense.net/>), last accessed Mar.2012)
18. Martinelli, A.: State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration. *IEEE Trans. Robot.* **27**(2), 239–255 (2011)
19. Martinelli, A.: Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination. *IEEE Trans. Robot.* **28**(1), 44–60 (2012)
20. Martinelli, A.: Observability properties and deterministic algorithms in visual-inertial structure from motion. *Foundations and Trends in Robotics* **3**(3), 139–209 (2014)
21. Maybeck, P.S.: Stochastic models, estimation, and control, vol. I. Academic press, New York (1979)
22. Mirisola, L.G.B., Dias, J., De Almeida, A.T.: Trajectory recovery and 3d mapping from rotation-compensated imagery for an airship. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1908–1913 (2008)
23. Mirzaei, F.M., Roumeliotis, S.I.: A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Trans. Robot.* **24**(5), 1143–1156 (2008). doi:[10.1109/TRO.2008.2004486](https://doi.org/10.1109/TRO.2008.2004486)
24. Mourikis, A.I., Roumeliotis, S.: A multi-state constraint Kalman filter for vision-aided inertial navigation. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Roma, Italy, pp. 3565–3572 (2007)
25. Mourikis, A.I., Trawny, N., Roumeliotis, S.I., Johnson, A.E., Ansar, A., Matthies, L.: Vision-aided inertial navigation for spacecraft entry, descent, and landing. *IEEE Trans. Robot.* **25**(2), 264–280 (2009)
26. Nilsson, J., Händel, P.: Time synchronization and temporal ordering of asynchronous sensor measurements of a multi-sensor navigation system. In: IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 897–902 (2010)
27. Panahandeh, G., Guo, C.X., Jansson, M., Roumeliotis, S.I.: Observability analysis of a vision-aided inertial navigation system using planar features on the ground. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, pp. 4187–4194 (2013)
28. Panahandeh, G., Guo, C.X., Jansson, M., Roumeliotis, S.I.: Observability analysis of a vision-aided inertial navigation system using planar features on the ground: Supplemental material. In: Technical Report (2013)
29. Panahandeh, G., Jansson, M.: IMU-camera self-calibration using planar mirror reflection. In: Proceedings of IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), Guimares, Portugal, pp. 1–7 (2011)
30. Panahandeh, G., Jansson, M.: Vision-aided inertial navigation based on ground plane feature detection. *IEEE/ASME Trans. Mechatron.* **19**(4), 1206–1215 (2014)
31. Panahandeh, G., Jansson, M., Hutchinson, S.: IMU-camera data fusion: Horizontal plane observation with explicit outlier rejection. In: IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), Montbéliard, France (2013)
32. Panahandeh, G., Mohammadiha, N., Jansson, M.: Ground plane feature detection in mobile vision-aided inertial navigation. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Algarve, Portugal, pp. 3607–3611 (2012)
33. Panahandeh, G., Zachariah, D., Jansson, M.: Exploiting ground plane constraints for visual-inertial navigation. In: Proceedings of IEEE-ION Position Location and Navigation Symposium, Myrtle Beach, South Carolina, pp. 527–534 (2012)
34. Pears, N., Liang, B.: Ground plane segmentation for mobile robot visual navigation. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Outrigger Wailea Resort, Maui, Hawaii, pp. 1513–1518 (2001)
35. Roumeliotis, S.I., Johnson, A.E., Montgomery, J.F.: Augmenting inertial navigation with image-based motion estimation. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Washington D.C., pp. 4326–4333 (2002)
36. Shuster, M.D.: A survey of attitude representations. *Astronaut. Sci.* **41**(4), 439–517 (1993)
37. Song, X., Seneviratne, L.D., Althoefer, K.: A Kalman filter-integrated optical flow method for velocity sensing of mobile robots. *IEEE/ASME Trans. Mechatron.* **16**(3), 551–563 (2011)
38. Troiani, C., Martinelli, A.: Vision-aided inertial navigation using virtual features. In: Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Algarve, Portugal, pp. 4828–4834 (2012)

39. Wagner, D., Schmalstieg, D.: ARToolKitPlus for pose tracking on mobile devices. In: Proceedings of 12th Computer Vision Winter Workshop (CVWW), pp. 139–146 (2007)
40. Weiss, S., Brockers, R., Matthies, L.: 4 DoF drift free navigation using inertial cues and optical flow. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, pp. 4180–4186 (2013)

**Ghazaleh Panahandeh** received the M.Sc. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2008. She completed her Ph.D. studies at the Signal Processing Department, KTH Royal Institute of Technology, Stockholm, Sweden, in 2014. Her research interests include inertial navigation and positioning, vision-aided inertial navigation, estimation theory, and applied machine learning. Dr. Panahandeh spent one year, 2014–2015, as a part-time postdoctoral fellow at the Signal Processing Department, KTH Royal Institute of Technology, Stockholm. Since 2014, she is working on active safety at Volvo cars, Gothenburg, Sweden.

**Seth Hutchinson** received the Ph.D. degree from Purdue University, West Lafayette, IN, in 1988. In 1990, he joined the faculty at the University of Illinois at Urbana-Champaign, where he is currently a Professor in the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute for Advanced Science and Technology. He served as Associate Department Head of ECE from 2001 to 2007. He has published more than 200 papers on the topics of robotics and computer vision, and is coauthor of the books *Principles of Robot Motion: Theory, Algorithms, and Implementations* (Cambridge, MA: MIT Press, 2005) and *Robot Modeling and Control* (New York: Wiley, 2005). Dr. Hutchinson currently serves on the editorial boards of the *International Journal of Robotics Research* and the *Journal of Intelligent Service Robotics*. He was Founding Editor-in-Chief of the *IEEE Robotics and Automation Societys Conference Editorial Board* (2006–2008) and Editor-in-Chief of the *IEEE TRANSACTIONS ON ROBOTICS* (2008–2013).

**Peter Händel** received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 1993. He was with Uppsala University from 1987 to 1993, and Ericsson AB, Kista, Sweden, from 1993 to 1997. From 1996 to 1997, he was a Visiting Scholar with the Tampere University of Technology, Tampere, Finland. Since 1997, he has been with the Royal Institute of Technology (KTH), Stockholm, Sweden, where he is currently a Professor of Signal Processing and the Head of the Department of Signal Processing. From 2000 to 2006, he held an adjunct position with the Swedish Defence Research Agency, Stockholm. He was a Guest Professor with the Indian Institute of Science, Bangalore, India, and the University of Gävle, Gävle, Sweden. He has served as an Associate Editor of the *IEEE TRANSACTIONS ON SIGNAL PROCESSING*.

**Magnus Jansson** received the Master of Science, Technical Licentiate, and Ph.D. degrees in electrical engineering from KTH Royal Institute of Technology, Stockholm, Sweden, in 1992, 1995 and 1997, respectively. During 1997–98 he held a lecturer position in the Control group at KTH. Since 1998 he has held various positions in the Signal Processing lab at KTH: 1998–2003 Assistant Professor, 2003–2012 Associate Professor, and since 2013 Professor. In January 2002 he was appointed Docent in Signal Processing at KTH. Dr Jansson spent one year during 1998–99 at the Department of Electrical and Computer Engineering, University of Minnesota.

Dr Jansson served as Associate Editor of *IEEE Signal Processing Letters* 2008–2012. Currently he is Associate Editor of *EURASIP Journal on Advances in Signal Processing* (2007–), and Senior Area Editor of *IEEE Signal Processing Letters* (2012–).

His research interests include statistical signal processing; navigation and positioning, sensor array processing, time series analysis, and system identification.