Observabilty Properties and Deterministic Algorithms in Visual-Inertial Structure from Motion

Agostino Martinelli

Foundations and Trends® in
sample
Vol. xx, No xx (xxxx) 1–68

DOI: xxxxxx

now
the essence of knowledge

Observabilty Properties and Deterministic Algorithms in Visual-Inertial Structure from Motion

Agostino Martinelli[1]

[1] *655 Avenue de l'Europe, MontbonnotGrenobleF-38334, France, agostino.martinelli@inria.fr*

Abstract

This paper discusses the visual-inertial structure from motion problem (VI-SfM problem) with special focus on the following three fundamental issues: observability properties, resolvability in closed form and data association. Regarding the first issue, after a discussion about the current state of the art, the paper first investigates more complex scenarios. Specifically, with respect to the common formulation, which assumes three orthogonal accelerometers and three orthogonal gyroscopes, the analysis is extended to cope with the cases of a reduced number of inertial sensors and any number of point features observed by monocular vision. In particular, the minimal case of a single accelerometer, no gyroscope and a single point feature is addressed. Additionally, the analysis accounts for biased measurements and unknown extrinsic camera calibration. The results derived for these new and very challenging scenarios have interesting consequences both from a technological and neuroscientific perspective. Regarding the second issue, a simple closed form solution to the VI-SfM is presented. This solution expresses the structure of the scene and the motion only in terms of the visual and in-

ertial measurements collected during a short time interval. This allows introducing deterministic algorithms able to simultaneously determine the structure of the scene together with the motion without the need for any initialization or prior knowledge. Additionally, the closed-form solution allows us to identify the conditions under which the VI-SfM has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on their arrangement in the 3D space and on the number of camera images. Finally, the paper discusses the third issue, i.e., the data association problem. Starting from basic results in computer vision, it is shown that, by exploiting the information provided by the inertial measurements, a single point correspondence (in the case of a planar motion) and two point correspondences (for a general 3D motion) are sufficient to characterize the motion between two camera poses. This allows us to use a 1-point RANSAC algorithm (in the planar case) or a 2-point RANSAC algorithm (in the general 3D case) to detect outliers. The paper concludes with some discussion about connections to related research fields both in the framework of computer science and neuroscience.

Contents

1

Introduction

The term *Structure from Motion* (SfM) was coined by the computer vision community to define the problem of estimating the three-dimensional structure of the scene and the motion from two-dimensional image sequences. In this paper we consider the same estimation problem but the sensor suit is now composed by a monocular camera and inertial sensors (accelerometers and gyroscopes). We will refer to this problem as to the *Visual-Inertial Structure from Motion* problem (from now on the VI-SfM problem). The VI-SfM problem has particular interest and has been investigated in many disciplines, both in the framework of computer science (e.g., [5, 22, 23, 36, 48]) and in the framework of neuroscience (e.g., [4, 11, 14]). These sensors require no external infrastructure and this is a key advantage for robots operating in unknown environments where GPS signals are shadowed. For this reason, vision and inertial sensing have received great attention by the mobile robotics community in the last years and many approaches have been introduced.

According to [8], we distinguish between loosely coupled and tightly coupled approaches. In the former, the sensor data processing takes place in separate modules, which exchange information each other. The

information delivered by the inertial can be used to speed up the tracking task of the features by predicting their locations within the next frame; in turn, data from the visual sensor allows updating the calibration parameters of inertial sensors. In the latter (tightly coupled approaches), all measurements, both visual and inertial, are combined and processed using a common filter-based approach.

A special issue of the *International Journal of Robotics Research* has recently been devoted to the visual and inertial sensor fusion [10]. In [8], a tutorial introduction to the vision and inertial sensing was presented. This work provides a biological point of view and it illustrates how vision and inertial sensors have useful complementarities allowing them to cover the respective limitations and deficiencies. In [46] the inertial measurements are used to reduce the ambiguities in the structure from motion problem.

The majority of the approaches so far introduced, perform the fusion of vision and inertial sensors by filter-based algorithms. In [2], these sensors are used to perform egomotion estimation. The sensor fusion is obtained by an Extended Kalman Filter (EKF) and by an Unscented Kalman Filter (UKF). The approach proposed in [16] extends the previous one by also estimating the structure of the environment where the motion occurs. In particular, new landmarks are inserted on line into the estimated map. This approach has been validated by conducting experiments in a known environment where a ground truth was available. Also, in [51] an EKF has been adopted. In this case, the proposed algorithm estimates a state containing the robot speed, position and attitude, together with the inertial sensor biases and the location of the features of interest. In the framework of airbone SfM, an EKF has been adopted in [24] to perform VI-SfM . It was observed that any inconsistent attitude update severely affects any SfM solution. The authors proposed to separate attitude update from position and velocity update. Alternatively, they proposed to use additional velocity observations, such as air velocity observation. Very recently, in the frame work of micro aerial robotics, flight stabilization and fully autonomous navigation have been achieved by using monocular vision and inertial sensors as the only exteroceptive sensors. Also in this case, the sensor fusion was carried out by a filter based algorithm [52, 53].

Since most of the previous algorithms to fuse visual and inertial measurements are based on linear estimators, in the last years, the effect that the observability properties can have on the consistency of a linearized estimator, has been investigated [21, 25, 27]. More in general, recent works have investigated the observability properties of the VI-SfM problem for various scenarios (see section 2.1 for a detailed state of the art).

The first goal of this paper (chapter 2) is to provide the main results achieved in the last years about the observability properties of the VI-SfM problem and to provide an important step forward by analyzing new and very challenging scenarios. Specifically, in the second part of chapter 2, the observability analysis is extended to cope with the cases of a reduced number of inertial sensors and any number of point features observed by the monocular vision. In particular, the minimal case of a single accelerometer, no gyroscope and a single point feature is addressed. Additionally, the analysis accounts for biased measurements and unknown extrinsic camera calibration. The results derived for these new and very challenging scenarios have interesting consequences both from a technological and neuroscientific perspective.

There are very few methods able to perform the fusion of image and inertial measurements without a filter-based approach. One algorithm of this type has been suggested in [48]. This algorithm is a batch method that performs SfM from image and inertial measurements. Specifically, it minimizes a cost function by using the Levenberg Marquardt algorithm [45]. This minimization process starts by initializing the velocities, the gravity and the biases to zero.

When using a recursive estimator (e.g. an EKF), or an off-line optimization method to minimize a suitable cost function, an important issue that arises is the initialization problem. Indeed, because of the system non-linearities, an erroneous initialization can irreparably compromise the entire estimation process. This problem has firstly been considered in [30]. In particular, the proposed method is able to estimate the scale factor by using a square root information filter. Additionally, the same authors proposed an EKF that does not suffer from the initialization of the speed and of the orientation [31, 32]. An efficient solution to the initialization problem is obtained by introducing a

deterministic algorithm able to determine the initial state by using the visual and inertial measurements acquired during a short time interval. This issue has been addressed only very recently [12, 35, 36, 37, 38]. Specifically, in [35] a first closed-form solution to VI-SfM has been obtained. Then, new extensions of this solution have been derived to cope with the cases of biased accelerometer's measurements [36, 37, 38] and an unknown camera-IMU extrinsic calibration [12]. In chapter 3, we provide a simple closed-form solution to the VI-SfM and, starting from this, we identify the conditions under which the VI-SfM has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on their arrangement in the 3D space and on the number of camera images.

Finally, a fundamental issue that arises in any visual motion estimation problem is data association. Any matching algorithm suffers from outliers, which must be detected and removed. To achieve this objective, the *RANdom SAmple Consensus* (RANSAC) introduced in [15] has been extensively used in visual motion estimation. In the past, concerning the case of only visual measurements (i.e., in SfM), the 5-point RANSAC [44] has been adopted [26, 49]. Indeed, five correspondences are in general necessary to identify the five parameters that characterize both the rotation (3 parameters) and the translation up to a scale (2 parameters), occurred between two camera frames. In the special case of a planar motion, the rotation can be characterized by a single parameter and the translation up to a scale by a further parameter. Hence, a 2-point RANSAC can be adopted to detect outliers for any planar motion. In the context of wheeled and indoor robotics, the motion not only is planar but also satisfies the non-holonomic constraint. This further information has been exploited in [47] to use a 1-point RANSAC for outliers detection.

The information provided by the inertial measurements, dramatically simplifies the data association task. Indeed, the rotation occurred between two camera frames can be efficiently obtained by integrating the inertial measurements. The translation up to a scale only depends on two parameters (for a general $3D$ motion) and one parameter (in the planar case). This makes possible the use of a 2-point RANSAC in

the general $3D$ case and a 1-point RANSAC in the planar case [50]. We discuss this issue very briefly in chapter 4.

2

Observability properties in VI-SfM

This chapter discusses the observability properties in visual inertial structure from motion. The discussion starts by considering the standard formulation of the problem, where the system is a sensor assembling constituted by a monocular camera and an Inertial Measurement Unit (IMU), which consists of three orthogonal accelerometers and three orthogonal gyroscopes. In this case the considered state contains the vehicle speed and attitude, the biases on the inertial measurements, the position of the features observed by the camera, the parameters characterizing the extrinsic camera-inertial sensors calibration and the magnitude of the gravity. In the case of a single feature, this state consists of 24 components. The observability analysis is carried out by using the method of *continuous symmetries*, which has recently been introduced [34]. Due to the system complexity, a direct derivation of the system continuous symmetries would require an excessive computational cost. For this reason, new theoretical results, which extend the theory developed in [34], are here derived to significantly reduce the load of symbolic computation. Specifically, we introduce the *quasi-projection* operation, which allows us to reduce the space dimensionality when deriving the observability properties of any estimation

problem. The first result provided in this chapter is the proof that, the information contained in the data provided by a monocular camera observing a single point-feature and by an IMU, allows estimating the absolute scale, the speed in the local frame, the absolute roll and pitch angles, the biases that affect the accelerometer's and the gyroscope's measurements, the magnitude of the gravitational acceleration and the extrinsic camera-IMU calibration.

Then, the chapter analyzes the observability properties when the number of inertial sensors is reduced. This analysis has never been provided before. The main result achieved in this context is that the observability properties of visual inertial structure from motion do not change by removing all the three gyroscopes and one of the accelerometers. In other words, exactly the same properties hold when the sensor assembling only consists of a monocular camera and two accelerometers. By removing a further accelerometer (i.e., by considering the case of a monocular camera and a single accelerometer) the system loses part of its observability properties. In particular, a new symmetry arises that corresponds to an internal rotation around the accelerometer axis. This means that some of the internal parameters that define the extrinsic camera calibration, are no longer observable. Although this symmetry does not affect the observability of the absolute scale and the magnitude of the velocity, it reflects in an indistinguishability of all the initial speeds that differ for a rotation around the accelerometer axis. On the other hand, as the camera is extrinsically calibrated (i.e., as the relative transformation between the camera frame and the accelerometer frame is known) this symmetry disappears and the system gains again full observability, as in the case of three orthogonal accelerometers and gyroscopes. This contribution is very important since allows us to reduce the number of inertial sensors. Additionally, it provides a new perspective in the framework of neuroscience to the process of vestibular and visual integration for depth perception and self motion perception.

To analyze these systems with a reduced number of inertial sensors, the chapter introduces a new methodology to derive the observability properties of a system when part of its input controls are unknown. To the best of our knowledge, this methodology is also an original contribution in the framework of control theory. It is an extension of

the theory developed by Herman and Krener [20] to investigate the observability properties of a control system when all or part of its input controls are unknown. Indeed, the theory developed in [20] always assumes that all the inputs of the system are known.

The chapter is articulated as follows. Section 2.1 provides the state of the art. The system and its basic equations are provided in section 2.2. Section 2.3 reminds the reader some basic concepts in non linear observability. In particular, it provides the main results introduced in [20] and [34]. Then, section 2.4 provides an observability analysis for the standard formulation, i.e., when all the input controls are known. Section 2.5 contains the extension of the theory in [20] to investigate the observability properties of a control system when all or part of its input controls are unknown. Finally, section 2.6 contains the new results about the observability properties when the number of inertial sensors is reduced. This includes the case of a single accelerometer.

2.1 State of the art

In [21, 22, 23, 25, 27, 36, 41] and [53] the observability properties of the VI-SfM have been investigated. In [21, 25] and [27] the authors investigated the estimator inconsistency in the VI-SfM problem. In [21] the authors found that standard EKF-based estimators lead to spurious information gain along unobservable directions. They also suggested a modification on the basic estimator to enforce the unobservable directions and thus to reduce inconsistency. In [22, 23, 41] and [53] the observability properties have been derived by accounting an unknown transformation between the camera and the IMU frames and an unknown magnitude of the gravity. Additionally, in [23] and [53] also the case of biased inertial measurements has been considered. We remark that in [22, 23] and [53] the observability properties have been derived starting from basic results in computer vision. Specifically, in [22] and [23], starting from the results derived in [6], a global frame has been fixed by constraining three directions determined by three points on the image plane. In [53], the camera is considered as a localization sensor up to a scale. This is based on the assumption that the camera is observing a number of features (at least five [43]), which guarantees

that its motion can be reconstructed up to a scale. This significantly simplifies the observability analysis since, the expression of the observation provided by the camera consists of three components of the state which defines the system. In [36] the observability properties have been derived without using the previous mentioned results from computer vision and this allowed us to deal with the case when a single point feature is observed by the camera. The analysis was based on the concept of *continuous symmetry* introduced in [34]. Since under these conditions the camera observation has an expression much more complex, the analysis in [36] was limited to the case when the camera extrinsic calibration in the IMU frame is a priori known.

2.2 The System

We consider a system (from now on we call it the *vehicle*) consisting of a monocular camera and an Inertial Measurement Unit (*IMU*). The IMU consists of three orthogonal accelerometers and three orthogonal gyroscopes. We introduce a global frame to characterize the motion of the vehicle moving in a $3D$ environment. Its z-axis points vertically upwards. We will adopt lower-case letters to denote vectors in this frame (e.g., the gravity is $\boldsymbol{g} = [0,\ 0,\ -g]^T$, where g is the magnitude of the gravitational acceleration). We define the vehicle local frame as the IMU frame. We will adopt upper-case letters to denote vectors in the vehicle frame. The camera frame differs from the local frame. We characterize the transformation between these two frames through $\boldsymbol{R^c}$ and q^c, where $\boldsymbol{R^c} \equiv [X^c,\ Y^c,\ Z^c]^T$ is the position of the camera optical center in the local frame and $q^c \equiv q_t^c + q_x^c i + q_y^c j + q_z^c k$ is the unit quaternion that characterizes the orientation of the camera frame in the local frame. We assume that both $\boldsymbol{R^c}$ and q^c are independent of time and are unknown.

The IMU provides the vehicle angular speed and acceleration. We will denote the measured quantities by $\boldsymbol{\Omega}$ and $\boldsymbol{A}$, respectively. These quantities differ from the true values, $\boldsymbol{\Omega^{true}}$ and $\boldsymbol{A^{true}}$. Regarding the angular speed, the one measured by the gyroscopes includes a bias and a zero-mean error, i.e.: $\boldsymbol{\Omega} = \boldsymbol{\Omega^{true}} + \boldsymbol{\Omega^{bias}} + \boldsymbol{n_\Omega}$. Regarding the acceleration, the one measured by the accelerometers includes the inertial

acceleration ($\boldsymbol{A}^{inertial}$), the gravitational acceleration ($\boldsymbol{G}$), a bias and a zero-mean error. In other words: $\boldsymbol{A} = \boldsymbol{A}^{inertial} + \boldsymbol{A}^{bias} - \boldsymbol{G} + \boldsymbol{n_A}$. Note that the gravity comes with a minus since, when the vehicle does not accelerate (i.e. $\boldsymbol{A}^{inertial}$ is zero), the accelerometers perceive an acceleration that is the same of an object accelerated upward in absence of gravity.

Our system is characterized by the state $[\boldsymbol{r},\ \boldsymbol{v},\ q]^T$ where $\boldsymbol{r} = [r_x,\ r_y,\ r_z]^T$ is the $3D$ vehicle position in the global frame, $\boldsymbol{v}$ is its time derivative, i.e. the vehicle speed in the global frame ($\boldsymbol{v} \equiv \frac{d\boldsymbol{r}}{dt}$) and $q \equiv q_t + q_x i + q_y j + q_z k$ is the unit quaternion that characterizes the vehicle orientation in the global frame.

In the following, we want to derive the analytical expression of the dynamics and the camera observations. For the sake of simplicity, we consider the case of noiseless measurements. The case with noise can be easily obtained with the substitution $\boldsymbol{A} \to \boldsymbol{A} + \boldsymbol{n_A}$ and $\boldsymbol{\Omega} \to \boldsymbol{\Omega} + \boldsymbol{n_\Omega}$. The dynamics of the previous state can be easily provided by expressing all the $3D$ vectors as imaginary quaternions. In practice, given a $3D$ vector $\boldsymbol{w} = [w_x,\ w_y,\ w_z]^T$ we associate with it the imaginary quaternion $w_q \equiv 0 + w_x i + w_y j + w_z k$. The dynamics of the state $[r_q,\ v_q,\ q]^T$ are:

$$\begin{cases} \dot{r}_q &= v_q \\ \dot{v}_q &= q A_q^{inertial} q^* = q A_q q^* - q A_q^{bias} q^* + g_q \\ \dot{q} &= \frac{1}{2} q \Omega_q - \frac{1}{2} q \Omega_q^{bias} \end{cases} \tag{2.1}$$

being q^* the conjugate of q, $q^* = q_t - i q_x - j q_y - k q_z$. By considering the case of unknown biases, unknown magnitude of the gravity and unknown transformation between the IMU and the camera frames, the state that defines our system becomes the following $24-$dimensional vector:

$$\boldsymbol{X} \equiv \left[\boldsymbol{r},\ \boldsymbol{v},\ q,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{R^c}, q^c,\ g\right]^T \tag{2.2}$$

whose dynamics are given in (2.1) with the following trivial additional equations:

$$\begin{cases} \dot{\boldsymbol{A}}^{\boldsymbol{bias}} = \dot{\boldsymbol{\Omega}}^{\boldsymbol{bias}} = \dot{\boldsymbol{R}}^{\boldsymbol{c}} = [0\ 0\ 0]^T \\ \dot{g} = \dot{q}^c = 0 \end{cases} \tag{2.3}$$

Note that this is the state that defines our system when a single point feature is observed by the camera. In this case the origin of the global frame can be chosen as coincident with the observed feature. In the case of multiple features, the state dimension becomes $24 + 3(N_f - 1)$ (N_f being the number of observed features) and the coordinates of the further $N_f - 1$ features are included in the state (see [36] for more details). On the other hand, the state defined in (2.2) is not a suitable choice to characterize our system. Indeed, the expression of the camera observations in terms of it involves the product of five quaternions: $(q^c)^* q^* r_q q q^c$. This makes impossible to efficiently derive the observability properties. To have a simple expression of the camera observations it is much more convenient to adopt a new state. Let us refer to the case of a single feature. The new state is:

$$\boldsymbol{X_n} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F},\ \boldsymbol{V},\ q,\ \boldsymbol{A}^{\boldsymbol{bias}},\ \boldsymbol{\Omega}^{\boldsymbol{bias}},\ \boldsymbol{R}^{\boldsymbol{c}}, q^c,\ g\right]^T \tag{2.4}$$

where ${}^{\boldsymbol{c}}\boldsymbol{F} \equiv [{}^cF_x,\ {}^cF_y,\ {}^cF_z]^T$ is the position of the feature in the camera frame and $\boldsymbol{V}$ is the vehicle speed in the local frame (i.e., in the IMU frame). By using the equations in (2.1) we obtain the following dynamics for the new state:

$$\begin{cases} {}^{\boldsymbol{c}}\dot{\boldsymbol{F}} &= M({}^{\boldsymbol{c}}\boldsymbol{\Omega})\,{}^{\boldsymbol{c}}\boldsymbol{F} - R_{q^c}[\boldsymbol{V} + (\boldsymbol{\Omega} - \boldsymbol{\Omega}^{\boldsymbol{bias}}) \wedge \boldsymbol{R}^{\boldsymbol{c}}] \\ \dot{\boldsymbol{V}} &= M(\boldsymbol{\Omega} - \boldsymbol{\Omega}^{\boldsymbol{bias}})\boldsymbol{V} + \boldsymbol{A} - \boldsymbol{A}^{\boldsymbol{bias}} + \boldsymbol{G} \\ \dot{q} &= \frac{1}{2} q \Omega_q - \frac{1}{2} q \Omega_q^{bias} \\ \dot{\boldsymbol{A}}^{\boldsymbol{bias}} &= \dot{\boldsymbol{\Omega}}^{\boldsymbol{bias}} = \dot{\boldsymbol{R}}^{\boldsymbol{c}} = [0\ 0\ 0]^T \\ \dot{g} &= \dot{q}^c = 0 \end{cases} \tag{2.5}$$

where:

- $M(\boldsymbol{\Omega}) \equiv \begin{bmatrix} 0 & \Omega_z & -\Omega_y \\ -\Omega_z & 0 & \Omega_x \\ \Omega_y & -\Omega_x & 0 \end{bmatrix}$;

- $^{c}\boldsymbol{\Omega}$ is the angular speed in the camera frame, i.e., $^{c}\Omega_q = (q^c)^*(\Omega_q - \Omega_q^{bias})q^c$;
- R_{q^c} is the rotation matrix associated with the quaternion q^c (i.e., for a $3D$ vector $\boldsymbol{w} = [w_x,\ w_y,\ w_z]^T$, $(R_{q^c}\boldsymbol{w})_q = (q^c)^* w_q q^c$);
- the symbol "$\wedge$" denotes the vectorial product.

Figure 2.1 displayes the three reference frames together with some of the previous vectors.

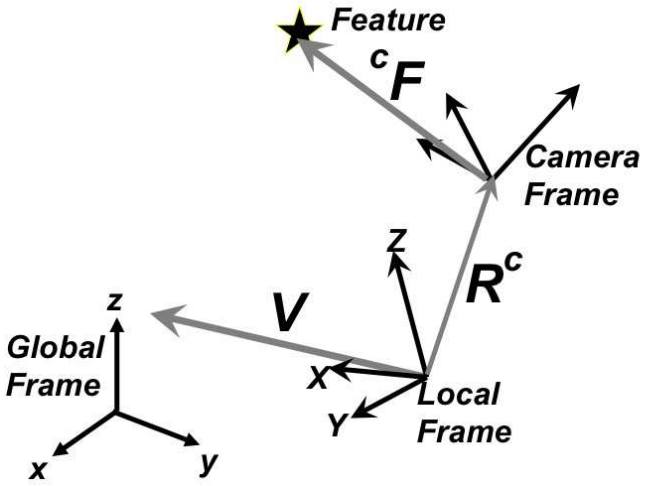


Fig. 2.1 Global frame, local (IMU) frame and camera frame with the feature position ($^{c}\boldsymbol{F}$) in the camera frame and the vehicle speed ($\boldsymbol{V}$) and the camera position ($\boldsymbol{R^c}$) in the local frame.

The expression of the camera observations in terms of the new state is trivial. Indeed, the camera provides the direction of the observed feature in its own frame. Hence, it provides the vector $^{c}\boldsymbol{F}$ up to a scale, or, equivalently, the two following ratios:

$$\boldsymbol{h_{cam}}(\boldsymbol{X_n}) \equiv [h_u,\ h_v]^T = \left[\frac{^{c}F_x}{^{c}F_z},\ \frac{^{c}F_y}{^{c}F_z}\right]^T \tag{2.6}$$

We have also to consider the two constraints $q^*q = 1$ and $(q^c)^*q^c = 1$. These can be dealt as further observations:

$$\boldsymbol{h_{const}}(\boldsymbol{X_n}) \equiv [h_q,\ h_{q^c}]^T = [q^* q,\ (q^c)^* q^c]^T \tag{2.7}$$

Finally, the case of multiple features can be characterized by including in the state the position of each feature in the camera frame, i.e., ${}^{\boldsymbol{c}}\boldsymbol{F} \rightarrow {}^{\boldsymbol{c}}\boldsymbol{F^i}$, $i = 1, 2, \cdots, N_f$. The resulting state has dimension $24+3(N_f-1)$.

2.3 Observability analysis

In this section we remind the reader some basic concepts in the theory of observability for non linear systems. For the sake of clarity, we will refer to a simple example. This will allow us to better illustrate these concepts and, in section 2.5.1, to introduce new concepts to deal with non linear systems when part (or even all) of the input controls are unknown.

2.3.1 A simple 2D localization problem

We consider a vehicle moving in a $2D$-environment. The configuration of the vehicle in a global reference frame, can be characterized through the vector $[x_v,\ y_v,\ \theta_v]^T$ where x_v and y_v are the cartesian vehicle coordinates, and θ_v is the vehicle orientation. We assume that the dynamics of this vector satisfy the unicycle differential equations:

$$\begin{cases} \dot{x}_v = v\cos\theta_v \\ \dot{y}_v = v\sin\theta_v \\ \dot{\theta}_v = \omega \end{cases} \tag{2.8}$$

where v and ω are the linear and the rotational vehicle speed, respectively, and they are the system input controls. In [34] we considered the same system. In that case, we assumed that the vehicle was equipped with a range sensor able to evaluate the distance from the origin (for instance, this can be obtained by fixing at the origin a marker that can be detected by the range sensor). Here, we assume that the vehicle is equipped with a GPS able to provide its position, i.e.:

$$\boldsymbol{z} = [x_v,\ y_v]^T \tag{2.9}$$

Our system is characterized by the previous two equations: the former describes its dynamics, the latter its observations. As the majority of real control systems, the dynamics given in (2.8) are affine in the controls, i.e. they can be written as follows:

$$\begin{cases} \dot{\boldsymbol{S}} = \boldsymbol{f}(\boldsymbol{S}, \boldsymbol{u}) = \boldsymbol{f_0}(\boldsymbol{S}) + \sum_{i=1}^{n_c} \boldsymbol{f_i}(\boldsymbol{S}) u_i \end{cases} \tag{2.10}$$

with $\boldsymbol{S} = [x_v,\ y_v,\ \theta_v]^T$, $n_c = 2$, $\boldsymbol{u} = [u_1,\ u_2]^T = [v,\ \omega]^T$, $\boldsymbol{f_0}(\boldsymbol{S}) = [0, 0, 0]^T$, $\boldsymbol{f_1}(\boldsymbol{S}) = [\cos\theta_v,\ \sin\theta_v,\ 0]^T$, $\boldsymbol{f_2}(\boldsymbol{S}) = [0,\ 0,\ 1]^T$. Additionally, our system is characterized by two observation functions (system outputs), which are $h_x(\boldsymbol{S}) = x_v$ and $h_y(\boldsymbol{S}) = y_v$.

2.3.2 Observability rank criterion

This criterion was introduced in 1977 by Hermann and Krener [20], to investigate the observability properties of a non linear system which satisfies (2.10) and with one or more outputs (observations). It requires to compute the Lie derivatives of all the observation functions with respect to all the vector fields $\boldsymbol{f_0}(\boldsymbol{S}),\ \boldsymbol{f_1}(\boldsymbol{S}),\ \cdots,\ \boldsymbol{f_{n_c}}(\boldsymbol{S})$. The Lie derivatives are defined recursively. The zero order Lie derivative of a given observation function $h(\boldsymbol{S})$ is the function itself, i.e., $\mathcal{L}^0 h \equiv h$. Then, the $(k+1)$-order Lie derivative of the observation function $h(\boldsymbol{S})$ with respect to $\boldsymbol{f_{i_1}}(\boldsymbol{S}),\ \cdots,\ \boldsymbol{f_{i_k}}(\boldsymbol{S}),\ \boldsymbol{f_{i_{k+1}}}(\boldsymbol{S})$ (with $i_1,\ i_2,\ \cdots,\ i_{k+1} = 0, 1, \cdots, n_c$) is $\mathcal{L}^{k+1}_{i_1,\cdots,i_k,i_{k+1}} h \equiv \boldsymbol{\nabla_S} \mathcal{L}^k_{i_1,\cdots,i_k} h \cdot \boldsymbol{f_{i_{k+1}}}(\boldsymbol{S})$. Note that this operation is not commutative with respect to the indexes's order. Hence, for a given observation function, we have $(n_c + 1)^k$ $k-$order Lie derivatives.

Let us denote with Ω, the space of all the Lie derivatives up to the $k-$order and with $\nabla\Omega$ the vector space spanned by the gradients of these functions. In this notation, the observability rank criterion can be expressed in the following way: *The dimension of the largest observable sub-system at a given* $\boldsymbol{S}$ *is equal to the dimension of* $\nabla\Omega$[1]. As a consequence, if for a given $k-$order the dimension of $\nabla\Omega$ is equal to the dimension of $\boldsymbol{S}$, the state $\boldsymbol{S}$ is observable and it is not necessary to compute higher order Lie derivatives.

[1] Actually, this condition guarantees that the system is *locally weakly* observable. The reader is referred to [20] for a detailed description of *weak* and *local* observability.

Let us apply the observability rank criterion to our example. The system has two outputs: $h_x \equiv x_v$ and $h_y \equiv y_v$. By definition, they coincide with their zero-order Lie derivatives. Their gradients with respect to the state $\boldsymbol{S}$ are, respectively: $[1,\ 0,\ 0]^T$ and $[0,\ 1,\ 0]^T$. Hence, the space spanned by the zero-order Lie derivatives has dimension two. In particular, by considering only the zero-order Lie derivatives, we can only conclude that the first 2 state components are observable. We do not know whether the third component, i.e., the vehicle orientation, is observable or not. Let us compute the first order Lie derivatives. We obtain: $\mathcal{L}_1^1 h_x = \cos\theta$, $\mathcal{L}_1^1 h_y = \sin\theta$, $\mathcal{L}_2^1 h_x = \mathcal{L}_2^1 h_y = 0$. Hence, the space spanned by the Lie derivatives up to the first order span the entire configuration space and we conclude that also the vehicle orientation is observable.

2.4 Observability Properties for the Standard Problem

In [36] we investigated the observability properties of the VI-SfM problem in several contexts which include the case of a single and multiple features, the case of known and unknown magnitude of the gravity, the case of biased and unbiased inertial measurements. In all the considered cases, the extrinsic camera-IMU transformation was assumed known. In the case of a single feature, the state adopted to characterize the case of biased inertial measurements and unknown magnitude of the gravity was $\left[\boldsymbol{r},\ \boldsymbol{v},\ q,\ \boldsymbol{A}^{\boldsymbol{bias}},\ \boldsymbol{\Omega}^{\boldsymbol{bias}},\ g\right]^T$, whose dimension is 17. The results of the observability analysis carried out in [36] can be summarized as follows:

Theorem 1 (Known extrinsic calibration) *Let us consider the VI-SfM problem with biased inertial measurements. Additionally, the magnitude of the gravity is unknown while the camera-IMU transformation is known. All the independent observable modes are: the positions in the local frame of all the observed features, the* 3 *components of the speed of the vehicle in the local frame, the two biases affecting the accelerometer and gyroscope measurements, the roll and the pitch angle and the magnitude of the gravity.*

The derivation of this result required to analytically compute the Lie derivatives up to the third order. In general, the complexity of the computation of the Lie derivatives and the determination of their dependence or independence dramatically depends on the state dimension.

Let us consider now the case when the extrinsic camera-IMU transformation is unknown. To solve the structure from motion we also need to estimate the parameters that characterize this transformation. In other words, the state that defines our system, is the one given in (2.2) or in (2.4), for the case of a single feature. Let us use the state in (2.4). Our system is characterized by the equations (2.5), (2.6) and (2.7). Also this system is characterized by the general formulation given in (2.10) with $\boldsymbol{S} = \boldsymbol{X_n}$, $n_c = 6$ and $\boldsymbol{u} = [u_1,\ u_2,\ \cdots,\ u_{n_c}]^T = [A_x,\ A_y,\ A_z,\ \Omega_x,\ \Omega_y,\ \Omega_z]^T$. The expression of the vector fields $\boldsymbol{f_i}$, $i = 0, 1, \cdots, n_c$, which is necessary for the computation of the Lie derivatives, can be obtained by a direct computation. For instance, the analytical expression of $\boldsymbol{f_0}$ is obtained by setting $A_x = A_y = A_z = \Omega_x = \Omega_y = \Omega_z = 0$ in (2.5).

In this case we found prohibitive to analytically deal with second-order Lie derivatives. Specifically, by using the symbolic computation tool of Matlab running on a 2.7GHz dual-core Intel Core i7 processor with 4MB shared L3 cache, the time demanded to compute the rank of the matrix whose lines are the gradients of all the Lie derivatives up to the second order, is equal to $101734s$ and the analytical determination of its null space required $127683s$. One of the goals of this chapter is to show how the theory developed in [34] can be used to deal with such complex cases, namely, how the observability properties can be derived by avoiding the computation of high-order Lie derivatives and how the space dimensionality where the rank must be computed can be reduced.

2.4.1 Quasi-Projection Operation and Matrix

In this section we introduce new theoretical results that allow significantly reducing the load of symbolic computation requested to derive the observability properties of a given system starting from the analysis of its continuous symmetries. Throughout this section, we denote

by n the dimension of the state. We start by reminding the reader the definition of *continuous symmetry* and the basic result derived in [34], which characterizes the observable modes.

Definition 1 (Continuous Symmetry) *The vector field $\boldsymbol{w_s(S)}$ is a continuous symmetry in $\boldsymbol{S}$ for the system defined in (2.10) if and only if it is a non null vector belonging to the null space of the matrix whose lines are the gradients of all the Lie derivatives computed in $\boldsymbol{S}$.*

Property 1 *$g(\boldsymbol{S})$ is an observable mode if and only if its gradient is orthogonal to all the symmetries.*

According to this result, the derivation of all the observable modes requires, first of all, to derive all the symmetries. The following remark provides an upper bound on the number of Lie derivatives that must be computed to detect the system symmetries:

Remark 1 *To detect the symmetries of the system in (2.10), only the first $(n-1)-$order Lie derivatives must be computed.*

The reader is referred to [1] to verify the validity of this remark. On the other hand, this remark is not useful to deal with our case since it is prohibitive even to deal with the second-order Lie derivatives.

Let us denote with $\boldsymbol{\phi^1}, \cdots, \boldsymbol{\phi^{n'}}$ $(1 \leq n' < n)$ a set of independent column vectors of dimension n that are orthogonal to the gradients of $n - n'$ independent Lie derivatives. We have the following result:

Property 2 *A sufficient condition for $g(\boldsymbol{S})$ to be an observable mode is that its gradient is orthogonal to all the vectors $\boldsymbol{\phi^1}, \cdots, \boldsymbol{\phi^{n'}}$.*

Proof. According to definition 1, if $n - n'$ is the number of all the independent Lie derivatives, $\boldsymbol{\phi^1}, \cdots, \boldsymbol{\phi^{n'}}$ are all the symmetries of the system. In this case, as stated in property 1, the previous condition is also necessary. When $n - n'$ is smaller than the number of all the independent Lie derivatives, we only know that the vector space generated by the symmetries is a subspace of the space generated by $\boldsymbol{\phi^1}, \cdots, \boldsymbol{\phi^{n'}}$. Hence, if the gradient of $g(\boldsymbol{S})$ is orthogonal to this latter

vector space it is also orthogonal to any of its subspace. From property 1 we conclude that $g(\boldsymbol{S})$ is an observable mode ■

This result is useful since it can be used by computing any number of independent Lie derivatives and not necessarily all of them. Additionally, this simple result allows us to reduce the system dimensionality in the special case when a component of the state verifies the sufficient condition stated in the previous property. In this case, since we know that this component is observable, to detect the other observable modes we can consider the reduced system that is obtained by removing this component from the state.

The second new result allows us to determine the dependence or independence of a given set of Lie derivatives by computing the rank of matrices whose dimension is smaller than the dimension of the original state. Let us suppose that we detected $n-n'$ independent Lie derivatives $(\mathcal{Q}^1,\ \cdots,\ \mathcal{Q}^{n-n'})$ and, as before, let us denote with $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$ $(1 \leq n' < n)$ a set of independent column vectors orthogonal to their gradients. Let us suppose that we compute m additional Lie derivatives $(\mathcal{P}^1,\ \cdots,\ \mathcal{P}^m)$. We need to determine if they are independent from the previous ones and also if they are independent one each other. In other words, we want to know the number of independent Lie derivatives in the set $\mathcal{Q}^1,\ \cdots,\ \mathcal{Q}^{n-n'},\ \mathcal{P}^1,\ \cdots,\ \mathcal{P}^m$. The easiest way to determine this is to compute the gradient of all these functions and then the rank of the matrix that contains these gradients. The dimension of this matrix is $(n-n'+m)\times n$. We know that this rank is larger or equal to $n-n'$ since we assumed that $\mathcal{Q}^1,\ \cdots,\ \mathcal{Q}^{n-n'}$ are independent. Let us denote this rank by $n-n'+p$. We remark that determining p is very easy if the vectors $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$ are orthonormal. In this special case, it is immediate to project the gradients of the Lie derivatives $(\mathcal{P}^1,\ \cdots,\ \mathcal{P}^m)$ onto the vectors $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$. This allows us to reduce the space dimensionality: instead of working in the original $n-$dimensional space we can work in a $n'-$dimensional space. In other words, p is the rank of the $m\times n'$ matrix, whose lines are the projections of the gradients of the Lie derivatives $(\mathcal{P}^1,\ \cdots,\ \mathcal{P}^m)$ onto the vectors $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$. On the other hand, even in the case when the vectors $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$ are not orthonormal, p can be computed by computing a rank of a $m\times n'$ matrix. We have the following result:

Property 3 *p is the rank of the $m \times n'$ matrix, whose i, j entry is the scalar product $\boldsymbol{\nabla_S} \mathcal{P}^i \cdot \boldsymbol{\phi^j}$.*

Proof. In this proof we will use the following two results from linear algebra about the matrix rank [40]:

(1) $rank(AB) = rank(A)$ when $rank(B) = k$, where k is the number of lines of B;
(2) If A is a matrix over the real numbers, $rank(AA^T) = rank(A^TA) = rank(A) = rank(A^T)$.

Let us introduce the following notation:

- $\mathcal{D}$ is the $(n-n') \times n$ matrix whose lines are the gradients of the functions $\mathcal{Q}^1, \cdots, \mathcal{Q}^{n-n'}$, i.e.:
$$\mathcal{D} \equiv \begin{bmatrix} \boldsymbol{\nabla_S} \mathcal{Q}^1 \\ \cdots \\ \boldsymbol{\nabla_S} \mathcal{Q}^{n-n'} \end{bmatrix}$$
- $\mathcal{E}$ is the $m \times n$ matrix whose lines are the gradients of the functions $\mathcal{P}^1, \cdots, \mathcal{P}^m$, i.e.:
$$\mathcal{E} \equiv \begin{bmatrix} \boldsymbol{\nabla_S} \mathcal{P}^1 \\ \cdots \\ \boldsymbol{\nabla_S} \mathcal{P}^m \end{bmatrix}$$
- $\mathcal{M}$ is the $(n-n'+m) \times n$ matrix whose lines are all the previous gradients, i.e.: $\mathcal{M} \equiv \begin{bmatrix} \mathcal{D} \\ \mathcal{E} \end{bmatrix}$
- $\mathcal{F}$ is the $n \times n'$ matrix whose columns are the vectors $\boldsymbol{\phi^1}, \cdots, \boldsymbol{\phi^{n'}}$;
- $\mathcal{N}$ is the following $n \times n$ matrix: $\mathcal{N} \equiv \begin{bmatrix} \mathcal{D}^T & \mathcal{F} \end{bmatrix}$

We have: $n - n' + p = rank(\mathcal{M})$. Since $rank(\mathcal{N}) = n$, by using the first property mentioned at the beginning of this proof, we also have $rank(\mathcal{M}) = rank(\mathcal{M}\mathcal{N})$. We have:

$$\mathcal{M}\mathcal{N} = \begin{bmatrix} \mathcal{D}\mathcal{D}^T & 0_{(n-n') \times n'} \\ \mathcal{E}\mathcal{D}^T & \mathcal{E}\mathcal{F} \end{bmatrix} \tag{2.11}$$

where $0_{(n-n')\times n'}$ is the zero $(n-n')\times n'$ matrix. Because of the second property mentioned at the beginning, $rank(\mathcal{D}\mathcal{D}^T) = rank(\mathcal{D})$, which is $n-n'$, because of the independence of $\mathcal{Q}^1,\ \cdots,\ \mathcal{Q}^{n-n'}$. To compute the rank of the matrix in (2.11), we detect the largest number of independent columns. Let us consider any linear combination of its columns that includes all the first $n-n'$ columns. A necessary condition to be the null vector is that the first $n-n'$ coefficients are equal to zero. Hence, a necessary and sufficient condition such that this linear combination vanishes if and only if all the coefficients are equal to zero, is that the columns selected among the last m columns of the matrix $\mathcal{M}\mathcal{N}$ are independent among them. This proves that $rank(\mathcal{M}\mathcal{N}) = n-n'+rank(\mathcal{E}\mathcal{F})$. But the $i,\ j$ entry of $\mathcal{E}\mathcal{F}$ is precisely the scalar product $\boldsymbol{\nabla_S}\mathcal{P}^i\ \cdot\boldsymbol{\phi^j}$ ■

This result is very important since allows us to verify the independence of a set of Lie derivatives by computing the rank of matrices with low dimension. Specifically, the gradients of the Lie derivatives are $n-$dimensional vectors. Once we have detected $n-n'$ independent Lie derivatives and a basis for its null space (i.e., $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$), to detect further independent Lie derivatives, we can work with $n'-$dimensional vectors. Indeed, for a given Lie derivative $\mathcal{P}$, instead of working with the $n-$dimensional vector $\boldsymbol{\nabla_S}\mathcal{P}$, we can work with the $n'-$dimensional vector, which is defined as the vector whose j^{th} entry is $\boldsymbol{\nabla_S}\mathcal{P}\ \cdot\boldsymbol{\phi^j}$ $(j=1,\cdots,n')$. This operation would be a projection if the basis $\boldsymbol{\phi^1},\ \cdots,\ \boldsymbol{\phi^{n'}}$ were orthonormal. For this reason we will call this operation a *quasi-projection*. This result is very powerful since it can be applied more consecutive times. Specifically, let us suppose that, by quasi-projecting the gradients of additional Lie derivatives in the $n'-$dimensional space we detect additional $n'-n''$ independent Lie derivatives $(1\leq n''<n')$. Let us denote with $\boldsymbol{\phi'^1},\ \cdots,\ \boldsymbol{\phi'^{n''}}$ a basis of the null space of the matrix made by these gradients quasi-projected in the $n'-$dimensional space. Now, to work in the new $n''-$dimensional space, we must quasi-project the gradient of a given Lie derivative, first on the $n'-$dimensional space and then on the $n''-$dimensional space. By computing the matrix whose $i,\ k$ entry is $\sum_j \boldsymbol{\phi_i^j}\boldsymbol{\phi_j'^k}$ we can simply multiply the gradient of the Lie derivative by this matrix. We we

will call this matrix the *quasi-projection* matrix. Therefore, each time we apply the result in property 3, we need to compute the new quasi-projection matrix.

2.4.2 VI-SfM Observability

We illustrate the power of the previous two properties by referring to our case, where the original state has dimension $n = 24$. We remind the reader that the time demanded to compute the rank of the matrix whose lines are the gradients of all the Lie derivatives up to the second order, is equal to $101734s$. We start by considering the matrix whose lines are the gradients of the following Lie derivatives: $\mathcal{L}^0 h_u$, $\mathcal{L}^0 h_v$, $\mathcal{L}^0 h_q$, $\mathcal{L}^0 h_p$, $\mathcal{L}^1_5 h_u$, $\mathcal{L}^1_6 h_u$, $\mathcal{L}^1_6 h_v$. The computation of its rank and its null space demands $0.37s$ and $0.24s$, respectively. In particular, the rank is 7 proving that these functions are independent. In a second step we quasi-project the gradient of the following Lie derivatives: $\mathcal{L}^1_0 h_u$ and $\mathcal{L}^1_4 h_u$. We build with them a 2×17 matrix. The computation of its rank and null space requires $0.027s$ and $3.11s$, respectively. The rank is 2. The computation of the quasi-projection matrix requires $6.48s$. In a third step we quasi-project the gradient of the following Lie derivatives: $\mathcal{L}^1_4 h_v$ and $\mathcal{L}^1_5 h_v$. We build with them a 2×15 matrix. The computation of its rank and null space requires $64s$ and $81s$, respectively. The rank is 2. The computation of the quasi-projection matrix requires $0.34s$. Finally, we quasi-project the gradient of the Lie derivative $\mathcal{L}^1_0 h_v$. We build with it a 1×13 matrix. The computation of its rank and null space requires $0.041s$ and $0.089s$, respectively. The computation of the quasi-projection matrix requires $0.44s$. We can proceed further to detect all the independent Lie derivatives. However, it is more convenient to use property 2. Specifically, we can check if some of the components of the state are observable (property 2 is a sufficient condition). We compute the gradient of the components of the original state and we quasi-project them on the final $12-$dimensional space. Regarding the four components of the quaternion q^c we obtain the null vector, meaning that these components are four observable modes. Hence, we remove these coponents from the original state and we start again the observability analysis of a new system whose state has now dimension equal

to 20. As previously mentioned, the complexity of the computation dramatically depends on the dimension of the state. Specifically, with the new state, it is possible to work with third-order Lie derivatives. In particular, we detect the following 19 independent Lie derivatives: $\mathcal{L}^0 h_u$, $\mathcal{L}^0 h_v$, $\mathcal{L}^0 h_q$, $\mathcal{L}^1_0 h_u$, $\mathcal{L}^1_5 h_u$, $\mathcal{L}^1_6 h_u$, $\mathcal{L}^1_0 h_v$, $\mathcal{L}^1_6 h_v$, $\mathcal{L}^2_{00} h_u$, $\mathcal{L}^2_{01} h_u$, $\mathcal{L}^2_{04} h_u$, $\mathcal{L}^2_{05} h_u$, $\mathcal{L}^2_{06} h_u$, $\mathcal{L}^2_{00} h_v$, $\mathcal{L}^2_{04} h_v$, $\mathcal{L}^3_{400} h_u$, $\mathcal{L}^3_{004} h_u$, $\mathcal{L}^3_{005} h_u$, $\mathcal{L}^3_{006} h_v$. The computation of the null space of the matrix whose lines are their gradients requires $0.46s$ and provides the vector:

$$\boldsymbol{w_s^{Rot_z}} = [0_{1\times 6}, -q_z, -q_y, q_x, q_t, 0_{1\times 10}]^T$$

where $0_{i\times j}$ denotes the $i \times j$ zero matrix. This vector expresses the system invariance under rotations around the vertical axis. Indeed, an infinitesimal rotation of magnitude ϵ about the vertical axis leaves the vectors ${}^{\boldsymbol{c}}\boldsymbol{F}$, $\boldsymbol{V}$, $\boldsymbol{A^{bias}}$, $\boldsymbol{\Omega^{bias}}$ and $\boldsymbol{R^c}$ unchanged and the quaternion q changes as follows [17]:

$$\begin{bmatrix} q_t \\ q_x \\ q_y \\ q_z \end{bmatrix} \rightarrow \begin{bmatrix} q_t \\ q_x \\ q_y \\ q_z \end{bmatrix} + \frac{\epsilon}{2} \begin{bmatrix} -q_z \\ -q_y \\ q_x \\ q_t \end{bmatrix} \tag{2.12}$$

i.e., the reduced $20-$dimensional state $\boldsymbol{X_r} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F},\ \boldsymbol{V},\ q,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{R^c},\ g\right]^T$ changes as follows: $\boldsymbol{X_r} \rightarrow \boldsymbol{X_r} + \frac{\epsilon}{2}\boldsymbol{w_s^{Rot_z}}$. Hence, $\boldsymbol{w_s^{Rot_z}}$ is a continuous symmetry for our system and the only non-observable mode is the yaw angle. In other words, we extended the result in theorem 1 to the case of unknown camera-IMU calibration. We have the following new result:

Theorem 2 (Unknown extrinsic calibration) *Let us consider the VI-SfM problem with biased inertial measurements. Additionally, both the magnitude of the gravity and the camera-IMU transformation are unknown. All the independent observable modes are: the positions in the local frame of all the observed features, the three components of the speed of the vehicle in the local frame, the two biases affecting the accelerometer and gyroscope measurements, the roll and the pitch angle, the magnitude of the gravity and the transformation between the camera and IMU frames.*

Note that this result holds even in the case when the camera observes a single point-feature.

2.5 Extension of the Theory of Herman and Krener

This section introduces a new methodology to derive the observability properties of a non linear system when all or part of its input controls are unknown. As it is common, we will refer to non linear systems whose dynamics are affine in the controls, i.e., they can be written as in the equation (2.10). Additionally, we will refer to the case when the observation (system output) is a scalar function of the state, i.e., $z = h(\boldsymbol{S})$. The theory of the observability is based on the assumption that both the system inputs (i.e., u_i in (2.10), $i = 1, \cdots, n_c$) and the system output (i.e., z) are known during a given time interval. This is a basic assumption. Specifically, the *observability rank criterion* introduced in [20] and used in [34] to define the concept of continuous symmetry, is based on this assumption. To extend the observability rank criterion, let us focus on the main steps in the theory introduced in [20]. Let us denote with $[T_{in},\ T_{fin}]$ the interval of time where the functions $u_i(t)$ $(i = 1, \cdots, n_c)$ and $z(t) = h(\boldsymbol{S}(t))$ are known.

The observability rank criterion is obtained by proceeding with the following three steps:

(1) The Taylor's theorem allows us to state that the knowledge of the functions $u_i(t)$ $(i = 1, \cdots, n_c)$ and $z(t)$ for $t \in [T_{in},\ T_{fin}]$ is equivalent to the knowledge of any order time derivative for $t = T_{in}$ of the functions u_i $(i = 1, \cdots, n_c)$ and z;
(2) the values of all the Lie derivatives of the function $h(\boldsymbol{S})$ in $\boldsymbol{S}(T_{in})$ along all the directions $\boldsymbol{f_i}(\boldsymbol{S})$, $i = 0, 1, \cdots, n_c$ can be obtained by inverting a linear system whose coefficients and constant terms are given by the previous time derivatives;
(3) the inverse function theorem allows us to identify whether the vector $\boldsymbol{S}(T_{in})$ can be obtained starting from the knowledge of the Lie derivatives (which are non linear scalar functions of this vector).

The goal of this section is to extend the observability rank criterion

to deal with the case when all or part of the n_c input controls of our system are unknown. In this case, we do not know some of the functions $u_i(t)$ $(i = 1, \cdots, n_c)$ $t \in [T_{in},\ T_{fin}]$. Let us denote with n_k and n_u the number of known and unknown input controls, respectively. We have $n_k + n_u = n_c$. Additionally, we order the inputs such that the known input are the first n_k. In other words, $u_i(t)$ $(i = 1, \cdots, n_k)$ are known for any $t \in [T_{in},\ T_{fin}]$, while $u_i(t)$ $(i = n_k + 1, \cdots, n_c)$ are unknown. Hence, the time derivatives of $u_i(t)$ $(i = n_k + 1, \cdots, n_c)$ are not available and the second step mentioned above cannot be used to obtain the Lie derivatives. Our basic idea consists in modifying the original state to be able to select some Lie derivatives, which can be obtained even without knowing all the time derivatives of $u_i(t)$ $(i = n_k + 1, \cdots, n_c)$ at $t = T_{in}$. This allows us to obtain a sufficient condition for the state observability and, by using the theoretical results derived in section 2.4, we obtain sufficient conditions to check whether a given mode is observable or not. We will call the new criterion the *extended observability rank criterion*. It will be introduced in 2.5.1. Then, to better illustrate the proposed methodology, we consider again the simple localization problem discussed in 2.3.1.

2.5.1 Extended observability rank criterion

Let us refer to the non linear system described by equation (2.10) and a given observation function $z = h(\boldsymbol{S})$. It is possible to analytically derive the expression of the n^{th} time derivative of the observation function in $t = T_{in}$ in terms of all the Lie derivatives of the function h along all the directions $\boldsymbol{f_i}(\boldsymbol{S})$, $i = 0, 1, \cdots, n_c$ computed in $\boldsymbol{S_0} \equiv \boldsymbol{S}(T_{in})$ up to the n^{th} order and all the time derivatives of the functions u_i $(i = 1, \cdots, n_c)$ computed in $t = T_{in}$. We have:

$$\left.\frac{d^n h(\boldsymbol{S}(t))}{dt^n}\right|_{t=T_{in}} = \tag{2.13}$$

$$= \sum_{p=1}^{n} \sum_{i_1 i_2 \cdots i_p = 0}^{n_c} \mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) \sum_{k_1, k_2, \cdots, k_p = 0,\ |\sum_{j=1}^{p} k_j = n-p}^{n-p} C^{n,\ p}_{k_1, k_2, \cdots, k_p} u_{i_1}^{(k_1)} \cdots u_{i_p}^{(k_p)}$$

where:

- $u_i^{(k)} \equiv \frac{d^k u_i}{dt^k}$, $k = 0, 1, \cdots, n$; $i = 1, \cdots, n_c$;
- $u_0 \equiv 1$ and $u_0^{(k)} = 0$, $k > 1$;
- $C^{n,\ p}_{k_1,k_2,\cdots,k_p}$, are real numbers satisfying the following recursive equation: $C^{n+1,\ p}_{k_1,k_2,\cdots,k_p} = \begin{bmatrix} C^{n,\ p-1}_{k_1,k_2,\cdots,k_{p-1}} + C^{n,\ p}_{k_1-1,k_2,\cdots,k_p} + C^{n,\ p}_{k_1,k_2-1,\cdots,k_p} + \cdots + C^{n,\ p}_{k_1,k_2,\cdots,k_{p-1}-1,k_p} & k_p = 0 \\ C^{n,\ p}_{k_1-1,k_2,\cdots,k_p} + C^{n,\ p}_{k_1,k_2-1,\cdots,k_p} + \cdots + C^{n,\ p}_{k_1,k_2,\cdots,k_{p-1}-1,k_p} & k_p \neq 0 \end{bmatrix}$, which is obtained by directly differentiating the expression in (2.13) with respect to time.

The expression in (2.13) allows us to perform the second step mentioned above, i.e., it allows us to obtain the Lie derivatives of h starting from the knowledge of the time derivatives of the system inputs and output by inverting a linear system. When n_u inputs are unknown, this step cannot be performed starting directly from (2.13). We split the expression in (2.13) as follows:

$$\left.\frac{d^n h(\boldsymbol{S}(t))}{dt^n}\right|_{t=T_{in}} = \sum_{p=1}^{n} \Bigg\{ \tag{2.14}$$

$$\Bigg\{ \sum_{i_1 i_2 \cdots i_p = 0}^{n_k} \mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) \sum_{k_1,k_2,\cdots,k_p=0,\ |\sum_{j=1}^{p} k_j = n-p}^{n-p} C^{n,\ p}_{k_1,k_2,\cdots,k_p} u_{i_1}^{(k_1)} \cdots u_{i_p}^{(k_p)} +$$

$$+ \sum_{i_1 i_2 \cdots i_p = remaining} \mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) \sum_{k_1,k_2,\cdots,k_p=0,\ |\sum_{j=1}^{p} k_j = n-p}^{n-p} C^{n,\ p}_{k_1,k_2,\cdots,k_p} u_{i_1}^{(k_1)} \cdots u_{i_p}^{(k_p)} \Bigg\}$$

The first sum only contains the know controls (i.e., $i_1, i_2, \cdots, i_p = 0, 1, \cdots n_k$) while, for each addend in the second sum, i.e., the sum where the indexes $i_1, i_2, \cdots, i_p$ take the remaining values, at least one control is unknown. In the special case when all the Lie derivatives $\mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0})$ vanish when at least one index $i_1, i_2, \cdots, i_p$ is larger than n_k, the second sum, which also contains unknown controls, vanishes as well. Hence, the expression in (2.14) can still be used to obtain all the Lie derivatives and the observability rank criterion can still be adopted. Obviously, this is a very special case. Our idea is to extend the original

state to artificially reproduce such a situation. In particular, we include the unknown inputs in the state together with their time derivatives. By including the time derivatives up to the n^{th} order, we will obtain $\mathcal{L}^p_{i_1 i_2 \cdots i_p} h(\boldsymbol{S_0}) = 0$ when at least one index $i_1, i_2, \cdots, i_p$ is larger than n_k and for all $p = 0, 1, \cdots, n$. Let us illustrate this by referring to the case when $n_k = n_c - 1$ (and, consequently, $n_u = 1$). Obviously, the zero order Lie derivative can be obtained without the necessity to know the inputs (it is trivially the output at $t = T_{in}$, $\mathcal{L}^0 h(\boldsymbol{S_0}) = h(\boldsymbol{S}(T_{in}))$). Let us consider the first order time derivative ($n = 1$). The expression in (2.14) becomes:

$$\left.\frac{dh(\boldsymbol{S}(t))}{dt}\right|_{t=T_{in}} = \sum_{i=0}^{n_c-1} \mathcal{L}^1_i h(\boldsymbol{S_0}) u_i + \mathcal{L}^1_{n_c} h(\boldsymbol{S_0}) u_{n_c} \tag{2.15}$$

Let us include the unknown u_{n_c} in the state, i.e., $\boldsymbol{S} \rightarrow \boldsymbol{S^e} \equiv [\boldsymbol{S},\ u_{n_c}]^T$. We have:

$$\begin{cases} \dot{\boldsymbol{S^e}} = \boldsymbol{f^e_0}(\boldsymbol{S^e}) + \sum_{i=1}^{n_c} \boldsymbol{f^e_i}(\boldsymbol{S^e}) u^e_i \end{cases} \tag{2.16}$$

where

- $\boldsymbol{f^e_0}(\boldsymbol{S^e}) \equiv [\boldsymbol{f_0}(\boldsymbol{S})^T + \boldsymbol{f_{n_c}}(\boldsymbol{S})^T u_{n_c},\ 0]^T$;
- $\boldsymbol{f^e_i}(\boldsymbol{S^e}) \equiv [\boldsymbol{f_i}(\boldsymbol{S})^T,\ 0]^T$, $i = 1, \cdots, n_c - 1$;
- $\boldsymbol{f^e_{n_c}}(\boldsymbol{S^e}) \equiv [\boldsymbol{0},\ 1]^T$, with $\boldsymbol{0}$ the line vector whose entries are all zero and whose dimension is equal to the one of $\boldsymbol{S}^T$;
- $u^e_i \equiv u_i$, $i = 1, \cdots, n_c - 1$;
- $u^e_{n_c} \equiv u^{(1)}_{n_c} = \dot{u}_{n_c}$.

It is immediate to realize that the first order Lie derivative along the direction $\boldsymbol{f^e_{n_c}}(\boldsymbol{S^e})$ is identically zero. This allows us to obtain all the other first order Lie derivatives. By including in the state also the first time derivative of u_{n_c} (namely, $u^{(1)}_{n_c}$) we can obtain all the Lie derivatives, up to the second order, along the first $n_c - 1$ directions. By including higher order time derivatives of the unknown input control u_{n_c} we can obtain higher order Lie derivatives along the first $n_c - 1$ directions. At this point, the third step in the Herman & Krener theory

previously mentioned can be performed by using the Lie derivatives that are available. By analyzing these Lie derivatives it is possible to detect potential symmetries according to the theory developed in [34] and property 3 provides a sufficient condition for a given mode to be observable.

2.5.2 Observability of the system in 2.3.1 with unknown input controls

We illustrate the methodology introduced in the previous section to investigate the observability properties of the simple system introduced in 2.3.1 when part, or all, of the input controls are unknown. We already know that the state $\boldsymbol{S} = [x_v,\ y_v,\ \theta_v]^T$ is observable when all the input controls are known (i.e., when the functions $v(t)$ and $\omega(t)$ are known for any $t \in [T_{in},\ T_{fin}]$). Intuitively, we know that the knowledge of both the inputs is unnecessary to have the full observability of the entire state. Indeed, the first two state components can be directly obtained from the GPS. By knowing these two components in a given time interval, we also know their time derivatives. In particular, we known $\dot{x}_v(T_{in})$ and $\dot{y}_v(T_{in})$. From (2.8) we easily obtain: $\theta_v(T_{in}) = atan\left(\frac{\dot{y}_v(T_{in})}{\dot{x}_v(T_{in})}\right)$. Hence, also the initial orientation is observable, by only using the GPS measurements.

We want to show that, by applying the methodology introduced in the previous section, we obtain exactly the same result. We start by including in the original state the unknown v, i.e., $\boldsymbol{S^e} = [x_v,\ y_v,\ \theta_v,\ v]^T$. We obtain:

$$\begin{cases} \dot{x}_v = v\cos\theta_v \\ \dot{y}_v = v\sin\theta_v \\ \dot{\theta}_v = \omega \\ \dot{v} = v^{(1)} \end{cases} \tag{2.17}$$

We have: $n_c = 2$, $\boldsymbol{u} = [u_1^e,\ u_2^e]^T = [\omega,\ v^{(1)}]^T$, $\boldsymbol{f_0^e}(\boldsymbol{S^e}) = [v\cos\theta_v,\ v\sin\theta_v,\ 0,\ 0]^T$, $\boldsymbol{f_1^e}(\boldsymbol{S^e}) = [0,\ 0,\ 1,\ 0]^T$, $\boldsymbol{f_2^e}(\boldsymbol{S^e}) = [0,\ 0,\ 0,\ 1]^T$. The first order Lie derivatives are: $\mathcal{L}_0^1 h_x(\boldsymbol{S^e}) = v\cos\theta_v$, $\mathcal{L}_1^1 h_x(\boldsymbol{S^e}) = \mathcal{L}_2^1 h_x(\boldsymbol{S^e}) = 0$. By chance, also $\mathcal{L}_1^1 h_x(\boldsymbol{S^e}) = 0$ and we do

not need to include also ω in the state. By using (2.14) up to the first order (i.e., $n \leq 1$), we can determine $\mathcal{L}_0^1 h_x(\boldsymbol{S^e})$. In other words, we can determine $v \cos\theta_v$. By considering the second observation function (i.e., h_y) we find that we can also determine $\mathcal{L}_0^1 h_y(\boldsymbol{S^e}) = v \sin\theta_v$. The gradients of the functions $\mathcal{L}^0 h_x(\boldsymbol{S^e})$, $\mathcal{L}^0 h_y(\boldsymbol{S^e})$, $\mathcal{L}_0^1 h_x(\boldsymbol{S^e})$ and $\mathcal{L}_0^1 h_y(\boldsymbol{S^e})$ span the entire configuration space of the state $\boldsymbol{S^e}$ meaning that this extended state is observable.

2.6 Observability Properties with unknown inputs

We use the methodology described in section 2.5 to analyze the VI-SfM problem when the number of accelerometers and gyroscopes is reduced. In other words, we consider the system defined in section 2.2 when the input vector $\boldsymbol{u} = [u_1,\ u_2,\ \cdots,\ u_6]^T = [A_x,\ A_y,\ A_z,\ \Omega_x,\ \Omega_y,\ \Omega_z]^T$ is partially known, i.e., some of its components are known during a given time interval and the others are totally unknown. Specifically, we analyze three different cases:

(1) camera extrinsically calibrated and $n_k = 1$ (i.e., the inertial sensors only consist of a single accelerometer);
(2) camera extrinsically uncalibrated and $n_k = 2$ (i.e., the inertial sensors only consist of two accelerometers);
(3) camera extrinsically uncalibrated and $n_k = 1$.

For the sake of simplicity, in the last two cases we do not consider the extreme case of a single feature. In particular, we assume that the camera is able to provide its position, orientation and speed up to a scale. This significantly reduces the computational load.

2.6.1 Single Accelerometer and Camera extrinsically calibrated

We directly analyze the extreme case of a single point feature. In the case of three orthogonal accelerometers and gyroscopes and unknown camera-IMU extrinsic calibration, the state that characterizes the vehicle was given in (2.4). Here, we assume that the camera is extrinsically calibrated. Without loss of generality, we assume that the camera and

the IMU frame coincide (i.e., $\boldsymbol{R^c} = [0\ 0\ 0]^T$ and $q^c = 1$) and that the single accelerometer points towards the $z-$direction of the camera frame. The state that characterizes this system is the following 12−dimensional vector:

$$\boldsymbol{X_1^c} \equiv \left[\boldsymbol{F},\ \boldsymbol{V},\ q,\ A_z^{bias},\ g\right]^T \tag{2.18}$$

where we adopt the subscript 1 to denote the case of a single accelerometer and the apex c to denote the case of a calibrated camera. $\boldsymbol{F}$ is the position of the feature in the local frame, $\boldsymbol{V}$ is the vehicle speed in the local frame, q the unit quaternion that describes the orientation of the local frame in the global frame, A_z^{bias} the bias of the accelerometer and g the magnitude of the gravity. The dynamics of this state are obtained by using (2.5):

$$\begin{cases} \dot{\boldsymbol{F}} & = M(\boldsymbol{\Omega})\boldsymbol{F} - \boldsymbol{V} \\ \dot{\boldsymbol{V}} & = M(\boldsymbol{\Omega})\boldsymbol{V} + \boldsymbol{A} - \boldsymbol{A^{bias}} + \boldsymbol{G} \\ \dot{q} & = \dfrac{1}{2} q \Omega_q \\ \dot{g} & = \dot{A}_z^{bias} = 0 \end{cases} \tag{2.19}$$

where the matrix M is defined in section 2.2, $\boldsymbol{\Omega}$ is the unknown angular speed of the camera, $\boldsymbol{A^{bias}} = [0\ 0\ A_z^{bias}]^T$ and $\boldsymbol{A - A^{bias}}$ is the camera acceleration in the local frame, whose first two components are unknown and the third component is known up to the bias tanks to the accelerometer.

From (2.6) we easily obtain the expression of the camera observations:

$$\boldsymbol{h_{cam}}(\boldsymbol{X_1^c}) \equiv [h_u,\ h_v]^T = \left[\frac{F_x}{F_z},\ \frac{F_y}{F_z}\right]^T \tag{2.20}$$

We have also to consider the constraint $q^* q = 1$. This provides the further observation:

$$\boldsymbol{h_{const}}(\boldsymbol{X_1^c}) \equiv h_q = q^* q \tag{2.21}$$

The dynamics in (2.19) provide seven independent directions along which the Lie derivatives can be computed. On the other hand, only two

directions are available. They are the vector $\boldsymbol{f_0}(\boldsymbol{X_1^c})$, which is obtained by setting $\boldsymbol{\Omega} = \boldsymbol{A} = [0\ 0\ 0]^T$ in (2.19) and the vector $\boldsymbol{f_3}(\boldsymbol{X_1^c})$, which is obtained by setting $\boldsymbol{\Omega} = [0\ 0\ 0]^T$ and $\boldsymbol{A} = [0\ 0\ 1]^T$ in the dynamics in (2.19), once $\boldsymbol{f_0}(\boldsymbol{X_1^c})$ has been removed. Since the Lie derivatives along the other five directions are not null, we have to proceed as in section 2.5.2. We do this in several subsequent steps. In each step we check, first of all, which highest order of Lie derivatives of the observations can be used. This is obtained by checking that, for a given order, all the Lie derivatives up to this order, computed along at least one of the directions that are not available (i.e., $\boldsymbol{f_1}$, $\boldsymbol{f_2}$, $\boldsymbol{f_4}$, $\boldsymbol{f_5}$ and $\boldsymbol{f_6}$) are identically zero. Once this highest order is identified, we find the largest number of independent Lie derivatives up to this order. Then, we compute the set of all of vectors that are orthogonal to the gradients of these Lie derivatives. Finally, we apply the property 3 to detect which components of the vector in (2.18) are observable. We include new time derivatives of the unknown inputs (i.e., $A_x,\ A_y,\ \Omega_x,\ \Omega_y,\ \Omega_z$) to make usable higher order Lie derivatives, as explained in section 2.5.

2.6.1.1 First step

We start with the 12−dimensional state given by the vector in (2.18). Since the first order Lie derivatives along $\boldsymbol{f_4}$, $\boldsymbol{f_5}$ and $\boldsymbol{f_6}$ are different from zero both for h_u and h_v, we can only use zero-order Lie derivatives. On the other hand, the first order Lie derivatives along $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ (and also $\boldsymbol{f_3}$) vanish. Hence, it suffices to include $\Omega_x,\ \Omega_y,\ \Omega_z$ in the state to use the Lie derivatives up to the first order. The zero-order Lie derivatives are the three functions in (2.20, 2.21). They are independent. Hence the system has three observable modes. The set of vectors orthogonal to the gradients of these three functions can be determined. Property 3 does not allow us to prove the observability for any component of the state in (2.18).

2.6.1.2 Second step

We include $\Omega_x,\ \Omega_y,\ \Omega_z$ in the state. The dimension of the new state is 15. We can use all the Lie derivatives up to the first order. We detect the additional independent functions from them: $\mathcal{L}_0^1 h_u$ and $\mathcal{L}_0^1 h_v$. Hence

the system has five independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 10 vectors. Again, property 3 does not allow us to prove the observability for any component of the state in (2.18).

2.6.1.3 Third step

To use the second order Lie derivatives we need to include $A_x,\ A_y,\ \Omega_x^{(1)} \equiv \dot{\Omega}_x,\ \Omega_y^{(1)} \equiv \dot{\Omega}_y,\ \Omega_z^{(1)} \equiv \dot{\Omega}_z$ in the state. The dimension of the new state is 20. We can use all the Lie derivatives up to the second order. We detect the following additional independent functions: $\mathcal{L}^2_{00}h_u$, $\mathcal{L}^2_{03}h_u$ and $\mathcal{L}^2_{00}h_v$. Hence the system has eight independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 12 vectors. This time, property 3 allows us to conclude that the first three components of the vector in (2.18) (i.e., $\boldsymbol{F}$) are observable. On the other hand, it does not allow us to prove the observability for any other component of the state.

2.6.1.4 Forth step

To use the third order Lie derivatives we need to include also $A_x^{(1)} \equiv \dot{A}_x,\ A_y^{(1)} \equiv \dot{A}_y,\ \Omega_x^{(2)} \equiv \dot{\Omega}_x^{(1)},\ \Omega_y^{(2)} \equiv \dot{\Omega}_y^{(1)},\ \Omega_z^{(2)} \equiv \dot{\Omega}_z^{(1)}$ in the state. The dimension of the new state is 25. We detect the following additional independent functions: $\mathcal{L}^3_{000}h_u$, $\mathcal{L}^3_{003}h_u$, $\mathcal{L}^3_{030}h_u$, $\mathcal{L}^3_{000}h_v$ and $\mathcal{L}^3_{003}h_v$. Hence the system has 13 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 12 vectors. As in the previous step, property 3 allows us to conclude that the first three components of the vector in (2.18) (i.e., $\boldsymbol{F}$) are observable. On the other hand, it does not allow us to prove the observability for any other component of the state.

2.6.1.5 Fifth step

To use the fourth order Lie derivatives we need to include also $A_x^{(2)} \equiv \dot{A}_x^{(1)},\ A_y^{(2)} \equiv \dot{A}_y^{(1)},\ \Omega_x^{(3)} \equiv \dot{\Omega}_x^{(2)},\ \Omega_y^{(3)} \equiv \dot{\Omega}_y^{(2)},\ \Omega_z^{(3)} \equiv \dot{\Omega}_z^{(2)}$ in the state. The dimension of the new state is 30. We detect the following additional independent functions: $\mathcal{L}^4_{0000}h_u$, $\mathcal{L}^4_{0003}h_u$, $\mathcal{L}^4_{0030}h_u$, $\mathcal{L}^4_{0300}h_u$, $\mathcal{L}^4_{0000}h_v$ and

$\mathcal{L}^4_{0003}h_v$. Hence the system has 19 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 11 vectors. Property 3 allows us to conclude that the first six components of the vector in (2.18) (i.e., both the vector $\boldsymbol{F}$ and $\boldsymbol{V}$) are observable. On the other hand, it does not allow us to prove the observability for the other components of the state.

2.6.1.6 Sixth step

To use the fifth order Lie derivatives we need to include also $A_x^{(3)} \equiv \dot{A}_x^{(2)}$, $A_y^{(3)} \equiv \dot{A}_y^{(2)}$, $\Omega_x^{(4)} \equiv \dot{\Omega}_x^{(3)}$, $\Omega_y^{(4)} \equiv \dot{\Omega}_y^{(3)}$, $\Omega_z^{(4)} \equiv \dot{\Omega}_z^{(3)}$ in the state. The dimension of the new state is 35. We detect the following additional independent functions: $\mathcal{L}^5_{00000}h_u$, $\mathcal{L}^5_{00003}h_u$, $\mathcal{L}^5_{00030}h_u$, $\mathcal{L}^5_{03000}h_u$, $\mathcal{L}^5_{00000}h_v$ and $\mathcal{L}^5_{00003}h_v$. Hence the system has 25 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 10 vectors. By using property 3 we find the same properties obtained in the previous step.

2.6.1.7 Seventh step

To use the sixth order Lie derivatives we need to include also $A_x^{(4)} \equiv \dot{A}_x^{(3)}$, $A_y^{(4)} \equiv \dot{A}_y^{(3)}$, $\Omega_x^{(5)} \equiv \dot{\Omega}_x^{(4)}$, $\Omega_y^{(5)} \equiv \dot{\Omega}_y^{(4)}$, $\Omega_z^{(5)} \equiv \dot{\Omega}_z^{(4)}$ in the state. The dimension of the new state is 40. We detect the following additional independent functions: $\mathcal{L}^6_{000000}h_u$, $\mathcal{L}^6_{000003}h_u$, $\mathcal{L}^6_{000030}h_u$, $\mathcal{L}^6_{030000}h_u$, $\mathcal{L}^6_{000000}h_v$ and $\mathcal{L}^6_{000003}h_v$. Hence the system has 31 independent Lie derivatives. The set of vectors orthogonal to the gradients of these functions consists of 9 vectors. Again, by using property 3, we find the same properties obtained in the previous step.

2.6.1.8 Eighth step

To use the seventh order Lie derivatives we need to include also $A_x^{(5)} \equiv \dot{A}_x^{(4)}$, $A_y^{(5)} \equiv \dot{A}_y^{(4)}$, $\Omega_x^{(6)} \equiv \dot{\Omega}_x^{(5)}$, $\Omega_y^{(6)} \equiv \dot{\Omega}_y^{(5)}$, $\Omega_z^{(6)} \equiv \dot{\Omega}_z^{(5)}$ in the state. The dimension of the new state is 45. We detect the following additional independent functions: $\mathcal{L}^7_{0000000}h_u$, $\mathcal{L}^7_{0000003}h_u$, $\mathcal{L}^7_{0000030}h_u$, $\mathcal{L}^7_{0300000}h_u$, $\mathcal{L}^7_{0000000}h_v$ and $\mathcal{L}^7_{0000003}h_v$. Hence the system has 37 independent Lie derivatives. The set of vectors orthogonal to the gradients

of these functions consists of 8 vectors. By using property 3, we find that the first six components of the vector in (2.18) (i.e., both the vector $\boldsymbol{F}$ and $\boldsymbol{V}$) and the last two components of this vector (i.e., the accelerometer bias A_z^{bias} and the magnitude of the gravity g) are observable. Regarding the other four components, namely the four components of the quaternion q that characterizes the camera orientation, we detect the same rotation symmetry described by equation (2.12). This means that the roll and the pitch angles are also observable and that the only unobservable mode is the yaw. Since this unobservable mode is a consequence of the system invariance with respect to rotations about the vertical axis, it is useless to include higher order Lie derivatives: the observability properties of the state in (2.18) do not change.

We conclude this section with the following important novel result:

Theorem 3 (Single accelerometer, calibrated camera) *Let us consider the VI-SfM problem when the inertial sensors only consist of a single accelerometer and no gyroscope. Additionally, the magnitude of the gravity is unknown and the camera-IMU transformation is known. All the independent observable modes are: the positions in the local frame of all the observed features, the three components of the speed of the vehicle in the local frame, the bias affecting the accelerometer measurements, the roll and the pitch angle, the magnitude of the gravity. This holds even in the extreme case of a single point feature.*

This result is very important and basically means that, in the case when the camera is extrinsically calibrated, the observability properties of the VI-SfM problem do not change by removing two of the three accelerometers and all the gyroscopes. The information provided by these additional five inertial sensors is redundant.

2.6.2 Two Accelerometers and Uncalibrated Camera

In this section and in the next one we account an unknown camera-IMU calibration. For simplicity sake, we assume that the camera is observing at least five features. This allows us to consider the camera as a sensor able to provide its orientation, its angular speed and its position

and speed up to a scale in a global reference frame attached to these features. Obviously, the gravity in this global frame is not necessarily along the vertical axis. We denote this vector by $\boldsymbol{g} \equiv [g_x,\ g_y,\ g_z]^T$, which is unknown (both in magnitude and in direction). Additionally, we denote by $\frac{1}{\mu}$ the unknown absolute scale.

Without loss of generality, we assume that the two available accelerometers are along the y and the $z-$axis in the IMU frame. The state that describes our system is the one given in (2.4), with only two bias components instead of six, since the inertial sensors only consist of two accelerometers. Additionally, we include in the state the position of all the observed N_f features in the camera frame ($N_f \geq 5$). The state is:

$$\boldsymbol{X_2^u} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F^1},\ \cdots,\ {}^{\boldsymbol{c}}\boldsymbol{F^{N_f}},\ \boldsymbol{V},\ q,\ A_y^{bias},\ A_z^{bias},\ \boldsymbol{R^c}, q^c,\ \boldsymbol{g},\ \mu\right]^T \quad (2.22)$$

where the subscript 2 denotes two accelerometers and the apex u the fact that we are considering the case of a camera extrinsically uncalibrated. The dynamics of this state can be easily obtained from (2.5). We have ($i = 1, \cdots, N_f$):

$$\begin{cases} {}^{\boldsymbol{c}}\dot{\boldsymbol{F}}^{\boldsymbol{i}} & = M({}^{\boldsymbol{c}}\boldsymbol{\Omega}){}^{\boldsymbol{c}}\boldsymbol{F^i} - R_{q^c}[\boldsymbol{V} + \boldsymbol{\Omega} \wedge \boldsymbol{R^c}] \\ \dot{\boldsymbol{V}} & = M(\boldsymbol{\Omega})\boldsymbol{V} + \boldsymbol{A} - \boldsymbol{A^{bias}} + \boldsymbol{G} \\ \dot{q} & = \frac{1}{2} q \Omega_q \\ \dot{\boldsymbol{g}} & = \dot{\boldsymbol{R}}^{\boldsymbol{c}} = [0\ 0\ 0]^T \\ \dot{A}_y^{bias} & = \dot{A}_z^{bias} = \dot{q}^c = \dot{\mu} = 0 \end{cases} \quad (2.23)$$

where $\boldsymbol{\Omega}$ is the unknown angular speed in the IMU frame, $\boldsymbol{A}$ is the acceleration in the IMU frame, whose first component is unknown and $\boldsymbol{A^{bias}} = [0,\ A_y^{bias},\ A_z^{bias}]^T$. Since at least five point features are available, the angular speed is known in the camera frame, i.e., the vector ${}^{\boldsymbol{c}}\boldsymbol{\Omega}$ ($= R_{q^c}\ \boldsymbol{\Omega}$) can be obtained from the visual measurements. Additionally, the visual measurements provide $3N_f + 9$ scalar functions (system outputs). The first $3N_f$ are the components of the vectors $\mu\ {}^{\boldsymbol{c}}\boldsymbol{F^i}$, $i = 1, \cdots, N_f$, which are the positions of the features in the camera frame up to the scale. Hence, we have:

$$h_{F_x^i} \equiv \mu\ {}^cF_x^i; \quad h_{F_y^i} \equiv \mu\ {}^cF_y^i; \quad h_{F_z^i} \equiv \mu\ {}^cF_z^i; \quad i = 1, \cdots, N_f \tag{2.24}$$

Regarding the remaining nine outputs, three of them are the components of the speed in the camera frame up to a scale, i.e.:

$$h_{V_x} \equiv \mu\ {}^cV_x; \quad h_{V_y} \equiv \mu\ {}^cV_y; \quad h_{V_z} \equiv \mu\ {}^cV_z \tag{2.25}$$

where $\boldsymbol{{}^cV} \equiv [{}^cV_x,\ {}^cV_y,\ {}^cV_z]^T = R_{q^c}[\boldsymbol{V} + \boldsymbol{\Omega} \wedge \boldsymbol{R^c}]$. Since the camera provides its orientation in the global frame, also the component of the quaternion qq^c can be considered system outputs. We have:

$$h_t \equiv (qq^c)_t; \quad h_x \equiv (qq^c)_x; \quad h_y \equiv (qq^c)_y; \quad h_z \equiv (qq^c)_z \tag{2.26}$$

Finally, both the quaternions q and q^c must be unit quaternions. Hence, we have the two outputs:

$$h_q \equiv q_t^2 + q_x^2 + q_y^2 + q_z^2; \quad h_{q^c} \equiv (q_t^c)^2 + (q_x^c)^2 + (q_y^c)^2 + (q_z^c)^2; \tag{2.27}$$

We start by investigating the observability properties of a simplified system that is obtained by referring to the state in (2.22) with $N_f = 1$ and with the $3 + 9 = 12$ outputs given by (2.24) for a single feature and (2.25-2.27). Note that we are using the four observations in (2.26): this implicitly assumes that we are actually exploiting the camera observations related to at least five features, simultaneously.

The dynamics in (2.23) provide seven independent directions along which the Lie derivatives can be computed. On the other hand, one of these directions is not available. This is the vector $\boldsymbol{f_1(X_2^u)}$ that is obtained by setting $\boldsymbol{{}^c\Omega} = [0\ 0\ 0]^T$ and $\boldsymbol{A} = [1\ 0\ 0]^T$ in (2.23), once $\boldsymbol{f_0(X_2^u)}$ has been removed[2]. Since the Lie derivatives along this direction are not null, we have to proceed as in section 2.6.1. We must proceed in several subsequent steps. In each step we check, first of all, which highest order of Lie derivatives of the observations can be used.

[2] Note that in this case $\boldsymbol{u} = [u_1,\ u_2,\ \cdots,\ u_{n_c}]^T = [A_x,\ A_y,\ A_z,\ {}^c\Omega_x,\ {}^c\Omega_y,\ {}^c\Omega_z]^T$, i.e., the last three inputs are the components of the angular speed in the camera frame, which is known.

This is obtained by checking that, for a given order, all the Lie derivatives up to this order, computed at least once along $\boldsymbol{f_1}$, are identically zero. Once this highest order is identified, we find the largest number of independent Lie derivatives up to this order. We include new time derivatives of the unknown inputs (i.e., A_x) to make usable higher order Lie derivatives, as explained in section 2.5. We will show that, by including A_x and its first time derivative, we can prove the observability of the entire state.

2.6.2.1 First step

We start with the 23−dimensional state given by the vector in (2.22), with a single feature. By chance, the first order Lie derivatives of the functions h_{F_x}, h_{F_y} and h_{F_z} along $\boldsymbol{f_1}$ are null. Regarding the other nine outputs, the first order Lie derivatives along this direction are different from zero. Among the usable Lie derivatives, we detect 14 independent functions that are: $\mathcal{L}^0 h_{F_x}$, $\mathcal{L}^0 h_{F_y}$, $\mathcal{L}^0 h_{F_z}$, $\mathcal{L}^0 h_{V_x}$, $\mathcal{L}^0 h_{V_y}$, $\mathcal{L}^0 h_{V_z}$, $\mathcal{L}^0 h_t$, $\mathcal{L}^0 h_x$, $\mathcal{L}^0 h_y$, $\mathcal{L}^0 h_q$, $\mathcal{L}^0 h_{q^c}$, $\mathcal{L}^1_5 h_{F_x}$, $\mathcal{L}^1_6 h_{F_x}$ and $\mathcal{L}^1_6 h_{F_y}$.

2.6.2.2 Second step

We include A_x in the state. The new state has dimension 24. Now we can use the first order Lie derivatives of all the outputs and the second order Lie derivatives of the first three outputs. We detect seven additional independent Lie derivatives that are: $\mathcal{L}^2_{00} h_{F_x}$, $\mathcal{L}^2_{02} h_{F_x}$, $\mathcal{L}^2_{03} h_{F_x}$, $\mathcal{L}^2_{00} h_{F_y}$, $\mathcal{L}^2_{02} h_{F_y}$, $\mathcal{L}^2_{03} h_{F_y}$ and $\mathcal{L}^2_{00} h_{F_z}$. Hence, the total number of independent Lie derivatives that are usable is 21.

2.6.2.3 Third step

We include $A_x^{(1)} \equiv \dot{A}_x$ in the state. The new state has dimension 25. Now we can use the second order Lie derivatives of all the outputs. We detect four additional independent Lie derivatives that are: $\mathcal{L}^2_{00} h_{V_x}$, $\mathcal{L}^2_{05} h_{V_x}$, $\mathcal{L}^2_{06} h_{V_x}$ and $\mathcal{L}^2_{06} h_{V_y}$. Hence, the total number of independent Lie derivatives that are usable is 25, which coincides with the dimension of the state. This means that the entire state is observable.

If we include in the state all the observed features, we obtain that all of them are observable. Indeed, the camera provides for each feature its position up to a scale. On the other hand, we have shown that μ is observable, meaning that it is possible to get the scale. We conclude this section with the following important novel result:

Theorem 4 (Two accelerometers, uncalibrated camera) *Let us consider the VI-SfM problem when the inertial sensors only consist of two accelerometers and no gyroscope. Additionally, the magnitude of the gravity and the camera-IMU transformation are unknown and at least five features are available. All the independent observable modes are: the positions in the local frame of all the observed features, the three components of the speed of the vehicle in the local frame, the bias affecting the accelerometer measurements, the vehicle orientation, the magnitude of the gravity and the camera-IMU transformation.*

Note that, the fact that the entire orientation (i.e., also the yaw angle) is observable, is a consequence of the fact that we fixed the global frame on the point features that are available from the beginning. Theorem 4 is very important and basically states that, even in the case when the camera is not extrinsically calibrated, the observability properties of the VI-SfM problem do not change by removing one of the three accelerometers and all the gyroscopes. The information provided by these additional four inertial sensors is redundant.

2.6.3 Single Accelerometer and Uncalibrated Camera

We want to know how the observability properties change by removing a further accelerometer, namely, we want to investigate the case when only a single accelerometer is available. We make the same assumptions as in the previous section, i.e., we consider the case when at least five features are available. Our system is now described by the following state:

$$\boldsymbol{X_1^u} \equiv \left[{}^{\boldsymbol{c}}\boldsymbol{F^1},\ \cdots,\ {}^{\boldsymbol{c}}\boldsymbol{F^{N_f}},\ \boldsymbol{V},\ q,\ A_z^{bias},\ \boldsymbol{R^c}, q^c,\ \boldsymbol{g},\ \mu \right]^T \tag{2.28}$$

where N_f is the number of observed features and we assumed, without loss of generality, that the available accelerometer is along the $z-$axis in the IMU frame.

Before computing the Lie derivatives to apply the extended observability rank criterion introduced in section 2.5, we derive a continuous symmetry by using an intuitive procedure. Let us suppose to collect the data from the camera and the accelerometer during a given time interval for a generic vehicle motion, starting from a given initial state. We remark that, independently of the motion, by rotating the initial state around the accelerometer axis (i.e., around the $z-$axis of the IMU frame) we obtain exactly the same measurements. Let us derive how this rotation changes the initial state by referring to an infinitesimal rotation of an angle ϵ. We rotate all the features, the camera frame (namely its position and orientation in the IMU frame) the initial vehicle speed and orientation, simultaneously, around the $z-$axis of the IMU frame, by the angle ϵ. The camera configuration in the IMU frame changes as follows [17]:

$$\boldsymbol{R^c} \rightarrow \boldsymbol{R'^c} = \boldsymbol{R^c} + \epsilon \begin{bmatrix} Y^c \\ -X^c \\ 0 \end{bmatrix}; \quad q^c \rightarrow q'^c = q^c + \frac{\epsilon}{2} \begin{bmatrix} q_z^c \\ q_y^c \\ -q_x^c \\ -q_t^c \end{bmatrix}$$

The initial speed in the IMU frame ($\boldsymbol{V} \equiv [V_x, V_y, V_z]^T$) changes as follows:

$$\boldsymbol{V} \rightarrow \boldsymbol{V'} = \boldsymbol{V} + \epsilon \begin{bmatrix} V_y \\ -V_x \\ 0 \end{bmatrix}$$

Let us derive how the initial orientation changes. The state in (2.28) contains the quaternion q, which describes the IMU orientation in the global frame and not the orientation of the global frame in the IMU frame. This last orientation is described by the quaternion $p \equiv p_t + ip_x + jp_y + kp_z \equiv q^* = qt - iq_x - jq_y - kq_z$. The quaternion p changes as q^c, namely:

$$p \to p' = p + \frac{\epsilon}{2} \begin{bmatrix} p_z \\ p_y \\ -p_x \\ -p_t \end{bmatrix}$$

Hence, we have: $q_t^* \to q_t^* + \frac{\epsilon}{2} q_z^*$, $q_x^* \to q_x^* + \frac{\epsilon}{2} q_y^*$, $q_y^* \to q_y^* - \frac{\epsilon}{2} q_x^*$ and $q_z^* \to q_z^* - \frac{\epsilon}{2} q_t^*$. By using $q_t^* = q_t$, $q_x^* = -q_x$, $q_y^* = -q_y$ and $q_z^* = -q_z$ we obtain:

$$q \to q' = q + \frac{\epsilon}{2} \begin{bmatrix} -q_z \\ q_y \\ -q_x \\ q_t \end{bmatrix}$$

The rotation does not affect all the remaining quantities in the state in (2.28). Indeed, μ and A_z^{bias} are scalar quantities. The vectors ${}^{\boldsymbol{c}}\boldsymbol{F}^{\mathbf{1}}, \cdots, {}^{\boldsymbol{c}}\boldsymbol{F}^{\boldsymbol{N_f}}$ are the relative positions of the features in the camera frame. Since, by definition, we are both rotating the features and the camera frame, these relative positions are unvaried. Finally, the vector $\boldsymbol{g}$ remains unvaried since we are rotating the global frame and the gravity, simultaneously.

The rotation described above, is characterized by the following symmetry:

$$\boldsymbol{w^{int}} \equiv \tag{2.29}$$

$$\equiv \left[\mathbf{0}_{3N_f},\ V_y, -V_x, 0,\ -\frac{q_z}{2}, \frac{q_y}{2}, -\frac{q_x}{2}, \frac{q_t}{2},\ 0,\ Y^c, -X^c, 0,\ \frac{q_z^c}{2}, \frac{q_y^c}{2}, -\frac{q_x^c}{2}, -\frac{q_t^c}{2},\ 0, 0, 0,\ 0 \right]^T$$

where $\mathbf{0}_{3N_f}$ is the zero $1 \times 3N_f$ vector. Namely, the transformation:

$$\boldsymbol{X_1^u} \to \boldsymbol{X_1^u} + \epsilon \boldsymbol{w^{int}}$$

on the initial state, cannot be detected by analyzing the measurements delivered by the camera and the accelerometer independently of the trajectory.

Now we proceed as in the previous section, in several subsequent steps. As in the previous section, we refer to the case of a single feature

and we show that $\boldsymbol{w^{int}}$ is the only continuous symmetry. The dynamics of (2.28) provide seven independent directions along which the Lie derivatives can be computed. On the other hand, two of these directions are not available. These are the vectors $\boldsymbol{f_1(X_1^u)}$ and $\boldsymbol{f_2(X_1^u)}$, which are obtained by setting ${}^{\boldsymbol{c}}\boldsymbol{\Omega} = [0\ 0\ 0]^T$, $\boldsymbol{A} = [1\ 0\ 0]^T$ and ${}^{\boldsymbol{c}}\boldsymbol{\Omega} = [0\ 0\ 0]^T$, $\boldsymbol{A} = [0\ 1\ 0]^T$ in the dynamics, once $\boldsymbol{f_0(X_1^u)}$ has been removed. Since the Lie derivatives along these directions are not null, we have to proceed as in section 2.6.2. We must proceed in several subsequent steps. In each step we check, first of all, which highest order of Lie derivatives of the observations can be used. This is obtained by checking that, for a given order, all the Lie derivatives up to this order, computed at least once along $\boldsymbol{f_1}$ or $\boldsymbol{f_2}$, are identically zero. Once this highest order is identified, we find the largest number of independent Lie derivatives up to this order. We include new time derivatives of the unknown inputs (i.e., A_x and A_y) to make usable higher order Lie derivatives, as explained in section 2.5. We will show that, by including A_x and A_y and their first time derivative, we can collect a number of independent Lie derivatives that is equal to the dimension of the state minus one. Hence, the system has a unique symmetry, which is the vector $\boldsymbol{w^{int}}$ in (2.29).

2.6.3.1 First step

We start with the 22−dimensional state given by the vector in (2.28), with a single feature. By chance, the first order Lie derivatives of the functions h_{F_x}, h_{F_y} and h_{F_z} along $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ are null. Regarding the other nine outputs, the first order Lie derivatives along these directions are different from zero. Among the usable Lie derivatives, we detect 14 independent functions, which are: $\mathcal{L}^0 h_{F_x}$, $\mathcal{L}^0 h_{F_y}$, $\mathcal{L}^0 h_{F_z}$, $\mathcal{L}^0 h_{V_x}$, $\mathcal{L}^0 h_{V_y}$, $\mathcal{L}^0 h_{V_z}$, $\mathcal{L}^0 h_t$, $\mathcal{L}^0 h_x$, $\mathcal{L}^0 h_y$, $\mathcal{L}^0 h_q$, $\mathcal{L}^0 h_{q^c}$, $\mathcal{L}^1_5 h_{F_x}$, $\mathcal{L}^1_6 h_{F_x}$ and $\mathcal{L}^1_6 h_{F_y}$.

2.6.3.2 Second step

We include A_x and A_y in the state. The new state has dimension 24. Now we can use the first order Lie derivatives of all the outputs and the second order Lie derivatives of the first three outputs. We detect six ad-

ditional independent Lie derivatives that are: $\mathcal{L}^2_{00}h_{F_x}$, $\mathcal{L}^2_{03}h_{F_x}$, $\mathcal{L}^2_{00}h_{F_y}$, $\mathcal{L}^2_{03}h_{F_y}$, $\mathcal{L}^2_{00}h_{F_z}$ and $\mathcal{L}^2_{03}h_{F_z}$. Hence, the total number of independent Lie derivatives that are usable is 20.

2.6.3.3 Third step

We include $A_x^{(1)} \equiv \dot{A}_x$ and $A_y^{(1)} \equiv \dot{A}_y$ in the state. The new state has dimension 26. Now we can use the second order Lie derivatives of all the outputs. We detect five additional independent Lie derivatives which are: $\mathcal{L}^2_{00}h_{V_x}$, $\mathcal{L}^2_{05}h_{V_x}$, $\mathcal{L}^2_{06}h_{V_x}$, $\mathcal{L}^2_{00}h_{V_y}$ and $\mathcal{L}^2_{06}h_{V_y}$. Hence, the total number of independent Lie derivatives that are usable is 25, which is the dimension of the state minus one. This means that the system has a unique symmetry, which is the vector $\boldsymbol{w^{int}}$ in (2.29).

If we include in the state all the observed features, we obtain that all of them are observable. Indeed, the camera provides for each feature its position up to a scale. On the other hand, it is immediate to check that the gradient of μ computed with respect to the state is orthogonal to the symmetry $\boldsymbol{w^{int}}$. From property 1 we obtain that μ is observable, meaning that it is possible to get the scale. We conclude this section with the following important novel result:

Theorem 5 (Single accelerometer, uncalibrated camera) *Let us consider the VI-SfM problem when the inertial sensors only consist of a single accelerometer and no gyroscope. Additionally, the magnitude of the gravity and the camera-IMU transformation are unknown and at least five features are available. The system loses some of the observability properties since it arises a continuous internal symmetry. As a consequence, the initial speed and orientation and the camera configuration in the IMU frame are not fully observable: all these quantities cannot be distinguished from the same quantities rotated around the accelerometer axis. All the remaining states are observable as in the case of two or more accelerometers.*

3

Resolvability in closed form

This chapter investigates the possibility of solving the visual-inertial structure from motion problem in closed form. This makes possible to introduce deterministic algorithms that are very important to efficiently and robustly initialize any filter-based or optimization approach to the VI-SfM problem. The chapter starts by deriving a simple closed-form solution to this problem (section 3.1). This solution could be extended to cope with the case of biased accelerometer measurements [36, 38] and the case of unknown camera-IMU extrinsic calibration [12]. For the sake of simplicity and brevity, in this chapter we do not consider these extensions. Special attention is devoted to identify the conditions under which the problem has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on their arrangement in the 3D space and on the number of camera images (section 3.2). In the case of biased measurements, more images and additional restrictive conditions arise [37, 38]. Finally, section 3.3 provides some results obtained by performing Monte Carlo simulations. Specifically, the closed-form solution is used in conjunction with a filtering approach to show its

benefit in the initialization task.

3.1 Solution of VI-SfM in closed form

In the following we assume that the camera and the IMU frames coincide and that the IMU biases and the magnitude of the gravity are known. These assumptions will be relaxed in section 3.3. Since the local frame is time dependent, we will adopt the following notation: $\boldsymbol{W}_t(\tau)$ will be the vector with global coordinates $\boldsymbol{w}(\tau)$ in the local frame at time t. Additionally, we will denote with $C_{t_1}^{t_2}$ the matrix that characterizes the rotation occurred during the time interval $(t_1,\ t_2)$ and with $C_{t_2}^{t_1}$ its inverse (i.e., $(C_{t_1}^{t_2})^{-1} = C_{t_2}^{t_1}$). Let us refer to vectors that are independent of the origin of the reference frame (e.g., speed, acceleration, etc.). For these vectors we have: $\boldsymbol{W}_{t_1}(\tau) = C_{t_1}^{t_2}\boldsymbol{W}_{t_2}(\tau)$. Finally, C^t will denote the rotation matrix between the global frame and the local frame at time t, i.e., $\boldsymbol{w}(\tau) = C^t\boldsymbol{W}_t(\tau)$. We assume that the camera is observing one or more point-features during the time interval $[T_{in},\ T_{fin}]$. Our goal is to express in closed-form the observable modes at a given time T_{in} only in terms of the visual and inertial measurements obtained in the time interval $[T_{in},\ T_{fin}]$.

The position of the vehicle $\boldsymbol{r}$ at any time $t \in [T_{in},\ T_{fin}]$ satisfies the equation $\boldsymbol{r}(t) = \boldsymbol{r}(T_{in}) + \boldsymbol{v}(T_{in})\Delta t + \int_{T_{in}}^{t}\int_{T_{in}}^{\tau}\boldsymbol{a}(\xi)d\xi d\tau$. The last term contains a double integral over time that can be simplified in a single integral by integrating by parts. We obtain:

$$\boldsymbol{r}(t) = \boldsymbol{r}(T_{in}) + \boldsymbol{v}(T_{in})\Delta t + \int_{T_{in}}^{t}(t-\tau)\boldsymbol{a}(\tau)d\tau \tag{3.1}$$

where $\boldsymbol{a} \equiv \frac{d\boldsymbol{v}}{dt}$ and $\Delta t \equiv t - T_{in}$. We write equation (3.1) by highlighting the vector $\boldsymbol{A}_\tau(\tau)$ provided by the accelerometer:

$$\boldsymbol{r}(t) = \boldsymbol{r}(T_{in}) + \boldsymbol{v}(T_{in})\Delta t + \boldsymbol{g}\frac{\Delta t^2}{2} + C^{T_{in}}\boldsymbol{S}_{T_{in}}(t) \tag{3.2}$$

where:

$$\boldsymbol{S}_{T_{in}}(t) \equiv \int_{T_{in}}^{t}(t-\tau)C_{T_{in}}^{\tau}\boldsymbol{A}_\tau(\tau)d\tau$$

The matrix $C^{\tau}_{T_{in}}$ can be obtained from the angular speed during the interval $[T_{in},\ \tau]$ provided by the gyroscopes [13]. Hence, the vector $\boldsymbol{S}_{T_{in}}(t)$ can be obtained by integrating the data provided by the gyroscopes and the accelerometers delivered during the interval $[T_{in},\ t]$.

Let us suppose that N_f point-features are observed, simultaneously. Their position in the local frame are $\boldsymbol{F}^i_t(t)$ and, in the global frame $\boldsymbol{f}^i$.

$$\boldsymbol{f}^i = \boldsymbol{r}(t) + C^{T_{in}} C^t_{T_{in}} \boldsymbol{F}^i_t(t) \tag{3.3}$$

We write this equation at time $t = T_{in}$ obtaining:

$$\boldsymbol{f}^i - \boldsymbol{r}(T_{in}) = C^{T_{in}} \boldsymbol{F}^i_{T_{in}}(T_{in}) \tag{3.4}$$

By inserting the expression of $\boldsymbol{r}(t)$ provided in (3.2) into equation (3.3), by using (3.4) and by pre multiplying by the rotation matrix $(C^{T_{in}})^{-1}$ (we remind the reader that, according to our notation, $\boldsymbol{v}(T_{in}) = C^{T_{in}} \boldsymbol{V}_{T_{in}}(T_{in})$ and $\boldsymbol{g} = C^{T_{in}} \boldsymbol{G}_{T_{in}}$) we finally obtain the following equation:

$$\begin{aligned} C^t_{T_{in}} \boldsymbol{F}^i_t(t) = \boldsymbol{F}^i_{T_{in}}(T_{in}) - \boldsymbol{V}_{T_{in}}(T_{in})\Delta t - \boldsymbol{G}_{T_{in}} \frac{\Delta t^2}{2} + \\ - \boldsymbol{S}_{T_{in}}(t); \qquad i = 1,\ 2,\ \cdots,\ N_f \end{aligned} \tag{3.5}$$

A single image provides the bearing angles of all the point-features in the local frame. In other words, an image taken at time t provides all the vectors $\boldsymbol{F}^i_t(t)$ up to a scale. Since the data provided by the gyroscopes during the interval $(T_{in},\ T_{fin})$ allow us to build the matrix $C^t_{T_{in}}$, having the vectors $\boldsymbol{F}^i_t(t)$ up to a scale, allows us to also know the vectors $C^t_{T_{in}} \boldsymbol{F}^i_t(t)$ up to a scale.

We assume that the camera provides n_i images of the same N_f point-features at the consecutive times: $t_1 = T_{in} < t_2 < \cdots < t_{n_i} = T_{fin}$. From now on, for the sake of simplicity, we adopt the following notation:

- $\boldsymbol{F}^i_j \equiv C^{t_j}_{T_{in}} \boldsymbol{F}^i_{t_j}(t_j)$, $i = 1,\ 2,\ \cdots,\ N_f$; $j = 1,\ 2,\ \cdots,\ n_i$
- $\boldsymbol{F}^i \equiv \boldsymbol{F}^i_{T_{in}}(T_{in})$, $i = 1,\ 2,\ \cdots,\ N_f$
- $\boldsymbol{V} \equiv \boldsymbol{V}_{T_{in}}(T_{in})$
- $\boldsymbol{G} \equiv \boldsymbol{G}_{T_{in}}$

- $\boldsymbol{S}_j \equiv \boldsymbol{S}_{T_{in}}(t_j)$, $j = 1,\ 2,\ \cdots,\ n_i$

We remark that the difference $\boldsymbol{F}^i_j - \boldsymbol{F}^i_1,\ i = 1, 2, \cdots, N,\ j = 2, \cdots, n_i$, is independent of i (see equation (3.5), where, by definition, $C^{t_j}_{T_{in}} \boldsymbol{F}^i_{t_j}(t_j) = \boldsymbol{F}^i_j$). Hence, we will set $\boldsymbol{\chi}_j \equiv \boldsymbol{F}^i_j - \boldsymbol{F}^i_1$. This quantity characterizes the motion of the vehicle. We will denote with $\boldsymbol{\mu}^i_j$ the unit vector with the same direction of $\boldsymbol{F}^i_j$ and we introduce the unknowns λ^i_j such that $\boldsymbol{F}^i_j = \lambda^i_j \boldsymbol{\mu}^i_j$. Finally, without loss of generality, we can set $T_{in} = 0$, i.e., $\Delta t = t$. Our sensors provide $\boldsymbol{\mu}^i_j$ and $\boldsymbol{S}_j$ for $i = 1,\ 2,\ \cdots,\ N_f$; $j = 1,\ 2,\ \cdots,\ n_i$. Equation (3.5) can be written as follows:

$$\boldsymbol{F}^i - \boldsymbol{V} t_j - \boldsymbol{G}\frac{t_j^2}{2} - \lambda^i_j \boldsymbol{\mu}^i_j = \boldsymbol{S}_j \tag{3.6}$$

The VI-SfM problem is the determination of the vectors: $\boldsymbol{F}^i$, ($i = 1,\ 2,\ \cdots,\ N_f$), $\boldsymbol{V}$, $\boldsymbol{G}$. We can use the equations in (3.6) to determine these vectors. On the other hand, the use of (3.6) requires to also determine the quantities λ^i_j. By considering $j = 1$ in (3.6), i.e. $t_j = t_1 = T_{in} = 0$, we easily obtain: $\boldsymbol{F}^i = \lambda^i_1 \boldsymbol{\mu}^i_1$. Then, we write the linear system in (3.6) as follows:

$$\left[\begin{array}{ll} -\boldsymbol{G}\frac{t_j^2}{2} - \boldsymbol{V} t_j + \lambda^1_1 \boldsymbol{\mu}^1_1 - \lambda^1_j \boldsymbol{\mu}^1_j & = \boldsymbol{S}_j \\ \lambda^1_1 \boldsymbol{\mu}^1_1 - \lambda^1_j \boldsymbol{\mu}^1_j - \lambda^i_1 \boldsymbol{\mu}^i_1 + \lambda^i_j \boldsymbol{\mu}^i_j & = 0_3 \end{array}\right. \tag{3.7}$$

where $j = 2, \cdots, n_i$, $i = 2, \cdots, N_f$ and 0_3 is the 3×1 zero vector. This linear system consists of $3(n_i - 1)N_f$ equations in $N_f n_i + 6$ unknowns. Let us define the two column vectors $\boldsymbol{X}$ and $\boldsymbol{S}$:

$$\boldsymbol{X} \equiv [\boldsymbol{G}^T,\ \boldsymbol{V}^T,\ \lambda^1_1,\ \cdots,\ \lambda^{N_f}_1,\ \cdots,\ \lambda^1_{n_i},\ \cdots,\ \lambda^{N_f}_{n_i}]^T$$

and

$$\boldsymbol{S} \equiv [\boldsymbol{S}_2^T,\ 0_3,\ \cdots,\ 0_3,\ \boldsymbol{S}_3^T,\ 0_3,\ \cdots,\ 0_3,\ \cdots,\ \boldsymbol{S}_{n_i}^T,\ 0_3,\ \cdots,\ 0_3]^T$$

and the matrix Ξ (see equation (3.8) at the next page), where $T_j \equiv -\frac{t_j^2}{2} I_3$, $S_j \equiv -t_j I_3$ and I_3 is the identity 3×3 matrix; 0_{33} is the 3×3 zero matrix (note that the third set of columns disappear in absence

$$\Xi \equiv \left[\begin{array}{c|c|c|c|c|c|c|c|c|c|c} T_2 & S_2 & \boldsymbol{\mu}_1^1 & 0_3 & 0_3 & -\boldsymbol{\mu}_2^1 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & -\boldsymbol{\mu}_1^2 & 0_3 & -\boldsymbol{\mu}_2^1 & \boldsymbol{\mu}_2^2 & 0_3 & 0_3 & 0_3 & 0_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & 0_3 & -\boldsymbol{\mu}_1^{N_f} & -\boldsymbol{\mu}_2^1 & 0_3 & \boldsymbol{\mu}_2^{N_f} & 0_3 & 0_3 & 0_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ T_{n_i} & S_{n_i} & \boldsymbol{\mu}_1^1 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & -\boldsymbol{\mu}_{n_i}^1 & 0_3 & 0_3 \\ 0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & -\boldsymbol{\mu}_1^2 & 0_3 & 0_3 & 0_3 & 0_3 & -\boldsymbol{\mu}_{n_i}^1 & \boldsymbol{\mu}_{n_i}^2 & 0_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0_{33} & 0_{33} & \boldsymbol{\mu}_1^1 & 0_3 & -\boldsymbol{\mu}_1^{N_f} & 0_3 & 0_3 & 0_3 & -\boldsymbol{\mu}_{n_i}^1 & 0_3 & \boldsymbol{\mu}_{n_i}^{N_f} \end{array}\right] \tag{3.8}$$

of bias). The linear system in (3.7) can be written in the following compact format:

$$\Xi \boldsymbol{X} = \boldsymbol{S} \tag{3.9}$$

The sensor information is completely contained in the above linear system. Additionally, we assume that the magnitude of the gravitational acceleration is a priori known. This extra information is obtained by adding to our linear system the following quadratic equation: $|\boldsymbol{G}| = g$. By introducing the following $3 \times (N_f n_i + 6)$ matrix, $\Pi \equiv [I_3,\ 0_3\ \cdots\ 0_3]$, this quadratic constraint can be written in terms of $\boldsymbol{X}$ as follows:

$$|\Pi \boldsymbol{X}|^2 = g^2 \tag{3.10}$$

The VI-SfM problem can be solved by finding the vector $\boldsymbol{X}$, which satisfies (3.9) and (3.10).

3.2 Existence and number of distinct solutions

We are interested in understanding how the existence and the number of solutions of the VI-SfM problem depend on the motion, on the number of observed point-features, on the point-features arrangement in the 3D space and on the number of camera images. The resolvabil-

ity of the VI-SfM problem can be investigated by computing the null space of the matrix Ξ in (3.8). Let us denote with $\mathcal{N}(\Xi)$ this space. In [37, 38] we prove the following theorem that allows us to obtain all the properties of the VI-SfM problem by investigating the null space of Ξ:

Theorem 6 (Number of Solutions) *The VI-SfM problem has a unique solution if and only if $\mathcal{N}(\Xi)$ is empty. It has two solutions, if and only if $\mathcal{N}(\Xi)$ has dimension 1 and, for any $\boldsymbol{n} \in \mathcal{N}(\Xi)$, $|\Pi \boldsymbol{n}| \neq 0$. It has infinite solutions in all the other cases.*

In [37, 38] we also prove the following important property:

Property 4 *When $n_i \leq 2$ the dimension of $\mathcal{N}(\Xi)$ is at least 3. When $n_i = 3$ the dimension of $\mathcal{N}(\Xi)$ is at least 1. Finally, when $n_i \geq 4$ and the vehicle moves with constant acceleration the dimension of $\mathcal{N}(\Xi)$ is at least 1.*

Regarding the cases $n_i = 3$ and $n_i \geq 4$ with constant acceleration, the proof is obtained by showing that there is at least one $3D-$vector $\boldsymbol{\alpha}_0$ and one $3D-$vector $\boldsymbol{\nu}_0$ such that the following vector always belongs to $\mathcal{N}(\Xi)$:

$$\boldsymbol{n}_0 = [\boldsymbol{\alpha}_0,\ \boldsymbol{\nu}_0,\ \bar{n}_1^1,\ ..\bar{n}_1^i..,\ ..\bar{n}_j^1..,\ ..\bar{n}_j^i..]^T \tag{3.11}$$

where $\bar{n}_1^1 = -1$, $\bar{n}_j^1 = 1$, $\bar{n}_1^i = 1$, $\bar{n}_j^i = -1$ $(j = 2, 3;\ i = 2, \cdots, N_f)$. In the case of constant acceleration, $\boldsymbol{\alpha}_0$ is precisely its value. In the following, we discuss the number of solutions of the VI-SfM problem depending on the number of camera images (n_i).

3.2.1 $n_i \leq 2$

From property 4 we know that the dimension of $\mathcal{N}(\Xi)$ is at least 3 and, consequently, by using theorem 6, we conclude that the VI-SfM has always infinite solutions.

3.2.2 $n_i = 3$

From property 4 we know that the dimension of $\mathcal{N}(\Xi)$ is at least 1, independently of the number of point-features. When $N_f = 1$, Ξ is a

6×9 matrix. Hence, the dimension of $\mathcal{N}(\Xi)$ is at least 3. Let us consider the case when $N_f = 2$. In this case Ξ is a 12×12 matrix. In [37, 38] we prove the following property:

Property 5 ($n_i = 3,\ N_f = 2$) *The dimension of $\mathcal{N}(\Xi)$ is 1 if and only if the following two conditions are met:*

(1) for a given j (e.g., for $j = 2$), the three vectors $\boldsymbol{F}_1^1$, $\boldsymbol{F}_1^2$ and $\boldsymbol{\chi}_j$ span the entire $3D-$space;
(2) for the other value of j (e.g., for $j = 3$) $\boldsymbol{F}_j^i$ is not proportional to $\boldsymbol{F}_j^k$, $\forall i, k = 1, 2, \cdots, N_f$.

From now on, we will say that a condition is satisfied *in general* when the probability that it is not satisfied is zero. We remark that both conditions (i) and (ii) are met in general. For $N_f \geq 2$ we have:

Property 6 ($n_i = 3,\ N_f \geq 2$) *When $n_i = 3$ and $N_f \geq 2$ the VI-SfM problem has in general two distinct solutions. In some special cases it has infinite solutions.*

3.2.3 $n_i \geq 4$

When $n_i \geq 4$ the number of equations is larger than the number of unknowns, except when $n_i = 4$ and $N_f = 1$. In this case the matrix Ξ is 9×10 and the dimension of its null space is at least 1. We have the following property (see [37, 38] for its proof):

Property 7 ($n_i = 4,\ N_f = 1$) *The dimension of $\mathcal{N}(\Xi)$ is 1 if and only if the four vectors $\boldsymbol{F}_1^1$, $\boldsymbol{\chi}_2$, $\boldsymbol{\chi}_3$ and $\boldsymbol{\chi}_4$ span the entire $3D-$space.*

We do not provide here necessary and sufficient conditions for any value of n_i and N_f. The following property holds (see [37, 38] for its proof):

Property 8 ($n_i \geq 4$) *When $n_i = 4$ and $N_f = 1$ the VI-SfM problem has in general two distinct solutions. If $n_i = 4,\ N_f \geq 2$ or if $n_i \geq 5,\ \forall N_f$ it has in general a unique solution.*

3.2.4 Constant acceleration and constant speed

From property 4 we know that when the vehicle moves with constant acceleration, the dimension of $\mathcal{N}(\Xi)$ is at least 1. In [37, 38] we also provide a sufficient condition that guarantees that the dimension of $\mathcal{N}(\Xi)$ is exactly 1. Hence, the VI-SfM has two distinct solutions. A special case of constant acceleration occurs when the vector $\boldsymbol{\alpha}_0$ vanishes, i.e., when the vehicle moves with constant speed. Since $|\Pi \boldsymbol{n}_0| = |\boldsymbol{\alpha}_0| = 0$, according to theorem 6, the VI-SfM has infinite solutions. In [37, 38] we also prove that in this case the roll and pitch angles can be uniquely determined. Hence, we have:

Property 9 (Constant speed) *Let us suppose that the vehicle moves with constant speed. The VI-SfM has infinite solutions. Additionally, the orientation of the vehicle with respect to the horizontal plane can be uniquely determined.*

Table 3.1 summarizes the results of this section by providing the number of solutions case by case. Note that this table does not account the point-features arrangement in the 3D space. Specifically, the motion and the point-features are not supposed to be either coplanar or collinear. Regarding these cases, necessary conditions are provided in [37, 38] where the previous analysis is also extended to cope with the case of biased accelerometer's measurements.

3.3 Performance Evaluation

In this section we evaluate the performance of the closed-form solution in solving the VI-SfM starting from noisy and biased synthetic data generated through Monte Carlo simulations. Additionally, we consider the case when the transformation between the visual and inertial sensors is not perfectly known.

3.3.1 Simulated Trajectories

All the trajectories are randomly generated starting from the following initial true state:

$$\boldsymbol{r}(T_{in}) = [0.5,\ 0.5,\ 0.5]m;\ \boldsymbol{v}(T_{in}) = [0.1,\ 0.1,\ 0.1]ms^{-1};$$

Cases	Number of Solutions
Varying Acceleration $n_i = 4,\ N_f \geq 2\ ;\ n_i \geq 5,\ \forall N_f$	Unique Solution
Varying Acceleration $n_i = 3,\ N_f \geq 2;\ n_i = 4,\ N_f = 1$	Two Solutions
Constant and non null Acceleration $n_i = 3,\ N_f \geq 2;\ n_i \geq 4,\ \forall N_f$	Two Solutions
Constant Speed $\forall n_i,\ \forall N_f$	Infinite Solutions
Any Motion $n_i \leq 2,\ \forall N_f;\ n_i = 3,\ N_f = 1$	Infinite Solutions

Table 3.1 Number of distinct solutions for the VI-SfM problem.

$q(T_{in}) = 1$, which corresponds to the vehicle attitude $roll = pitch = yaw = 0\ deg$. Then, the trajectories are generated by randomly generating the linear and angular acceleration of the vehicle at 100 Hz. In particular, at each time step, the three components of the linear acceleration and the angular speed are generated as zero-mean Gaussian independent variables whose covariance matrices are equal to $(1ms^{-2})^2 I_3$ and $(10\ deg\ s^{-1})^2 I_3$, respectively.

3.3.2 Simulated Sensors

We simulate an Xsens Mtx device, according to the signal noise analysis provided in [55], which is based on the Allan Variance (see [55] for more details). The Mtx is a MEMS IMU characterized by low weight (about 30 g) and low cost. Hence it is suitable for many applications.

Starting from the accomplished trajectory, the true angular speed and the linear acceleration are computed at each time step of $0.01s$ (respectively, at the j^{th} time step, we denote them with $\boldsymbol{\Omega_j^{true}}$ and $\boldsymbol{A_j^{true}}$). Starting from them, the IMU sensors are simulated by randomly generating the angular speed and the linear acceleration at each step according to the following:

$$\boldsymbol{\Omega_j} = \mathcal{N}\left(\boldsymbol{\Omega_j^{true}} - \boldsymbol{\Omega^{bias}},\ P_\Omega\right)$$

$$\boldsymbol{A_j} = \mathcal{N}\left(\boldsymbol{A_j^{true}} - \boldsymbol{G}(t_j) - \boldsymbol{A^{bias}},\ P_A\right)$$

where:

- $\mathcal{N}(.,\ .)$ indicates the Normal distribution whose first entry is the mean value and the second its covariance matrix;
- P_Ω and P_A are the covariance matrices characterizing the thermo-mechanical white noise that affects both the gyroscopes and the accelerometers;
- $\boldsymbol{\Omega^{bias}}$ and $\boldsymbol{A^{bias}}$ are the gyroscope and accelerometer's biases, respectively.

In all the simulations we set both the matrices P_Ω and P_A diagonal and in particular: $P_\Omega = (1\ deg\ s^{-1})^2\ I_3$ and $P_A = (1\ cm\ s^{-2})^2\ I_3$. The value $(1\ deg\ s^{-1})^2$ is obtained from the experimental results provided in [55] on the Angular Random Walk (ARW). Specifically, this value slightly exceeds the one that can be inferred from the ARW provided for the examined Xsens Mtx device. Similarly, the value $(1\ cm\ s^{-2})^2$ is obtained from the experimental results provided in [55] on the Velocity Random Walk (VRW). Also this value slightly exceeds the one that can be inferred from the VRW provided for the examined Xsens Mtx device.

Finally, the two bias vectors are set as follows: $\boldsymbol{A^{bias}} = 0.001\ \hat{\boldsymbol{\mu}}\ m\ s^{-2}$, $\boldsymbol{\Omega^{bias}} = 0.01\ \hat{\boldsymbol{\mu}}\ deg\ s^{-1}$ where $\hat{\boldsymbol{\mu}}$ is the unit vector pointing in the direction $[1,\ 1,\ 1]$; Also in this case these values are obtained starting from the experimental results provided in [55] about the bias instability. Specifically, we are assuming that the bias are the ones that arise after about 100 s from the last calibration.

Regarding the camera, first of all we assume that its frame coincides with the IMU frame (i.e., the two frames have the same origin and the same orientation). We characterize an error in the extrinsic calibration by setting the actual position of the origin of the camera frame in the IMU frame to $[0.002,\ -0.003,\ 0.004]m$ and the actual orientation $q_{cam} = 1 - 2.3\ 10^{-5} + (3.5i - 5.2j + 2.6k)\ 10^{-3}$, which corresponds to the attitude $roll = 0.4\ deg$, $pitch = -0.6\ deg$ and $yaw = 0.3\ deg$. The camera readings are then generated in the following way. By knowing

the true trajectory and the true camera-IMU transformation, the true bearing angles of the point-features in the camera frame are computed. They are computed each $0.1s$. Then, the camera readings are generated by adding to the true values zero-mean Gaussian errors whose variance is equal to $(1\ deg)^2$ for all the readings.

3.3.3 Results

We considered three distinct scenarios that correspond to a different number of point features in the environment. Specifically, we considered the following three values: $N_f = 1, 2, 5$. For each scenario we performed 1000 trials. In each trial the position of the features is generated randomly pursuant to a uniform distribution on a box of $1m$ side centred on the starting position $\boldsymbol{r}(T_{in})$. For each trial, we computed the following errors that characterize the performance of the closed-form solution in solving the VI-SfM problem:

- E_v, defined as the norm of the vector that is the difference between the true initial speed and the one determined by the closed-form solution;
- E_μ, defined as the error on the scale in %;
- E_{roll}, defined as the difference between the true initial roll and the one determined by the closed-form solution;
- E_{pitch}, defined as E_{roll} but by considering the initial pitch instead of the initial roll.

These errors are plotted in fig. 3.1. Table 3.2 displays the mean values of these errors, averaged on the 1000 trials that have been performed. Regarding E_{roll} and E_{pitch} the mean value is computed starting from their absolute values. We remark that there is a huge improvement (nearly one order of magnitude) by passing from $N_f = 1$ to $N_f = 2$ while the improvement obtained by further increasing N_f is almost negligible.

N_f	$\langle E_v \rangle$ (ms^{-1})	$\langle E_\mu \rangle$ (%)	$\langle E_{roll} \rangle$ (deg)	$\langle E_{pitch} \rangle$ (deg)
1	0.0809	2.5482	0.4870	0.4745
2	0.0083	0.6063	0.0948	0.0950
5	0.0071	0.5418	0.0874	0.0848

Table 3.2 The mean values of the four errors for the three scenarios.

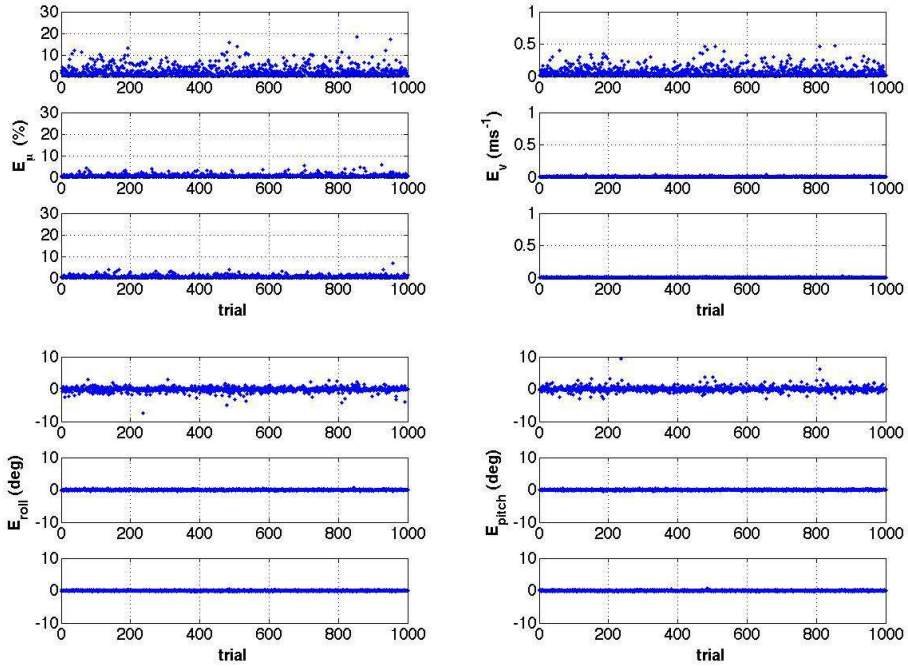


Fig. 3.1 The value of the four errors computed to evaluate the performance of the closed-form solution in solving the VI-SfM problem. Each figure contains three plots: they correspond to the three scenarios defined by $N_f = 1, 2, 3$ respectively from the top to the bottom.

3.4 Deterministic initialization of an EKF-based VI-SfM

In this section we show the benefit of using the closed-form solution for initializing a filter based approach to solve the VI-SfM problem. In section 3.1 we formulated the VI-SfM problem as the problem of determining the vectors: $\boldsymbol{F}^i$, $(i = 1,\ 2,\ \cdots,\ N_f)$ and $\boldsymbol{V}$, $\boldsymbol{G}$. For the

sake of clarity, in this section we choose to display the results in a global frame. For this reason, we need to consider at least two point-features. Indeed, two is the minimum number of point-features to uniquely define a global frame, provided that they do not lie on the same vertical axis (defined by the gravity). We define the global frame as follows: first, we define one of the point-feature as the origin of the frame. The z-axis coincides with the gravity axis but with opposite direction. Finally, the x-axis is defined by requiring that the second point-feature belongs to the xz-plane. In other words, the second point-feature has zero $y-$coordinate. In these settings, the VI-SfM can be defined as the estimation of the vehicle configuration and the estimation of the x and the z coordinate of the second point-feature (from now on, p_x and p_z, respectively). By adding more point-features, the state to be estimated also includes all the three coordinates of each point-feature. We adopt an Extended Kalman Filter (EKF) to perform this estimation. The state to be estimated is:

$$\boldsymbol{x^e} \equiv [\boldsymbol{r},\ \boldsymbol{v},\ q,\ p_x,\ p_z,\ \boldsymbol{A^{bias}},\ \boldsymbol{\Omega^{bias}},\ \boldsymbol{f}^3,\ \cdots,\ \boldsymbol{f}^{N_f}]^T$$

By collecting the sensor measurements during the time-interval $[T_{in},\ T_{fin}]$, the closed-form solution discussed in the previous sections allows us to determine the vectors $\boldsymbol{F}^i$, $(i = 1,\ 2,\ \cdots,\ N_f)$, $\boldsymbol{V}$ and $\boldsymbol{G}$ at the time T_{in}. Note that, when $N_f \geq 2$, having the vectors $\boldsymbol{F}^i$, $\boldsymbol{V}$ and $\boldsymbol{G}$ at the time T_{in}, allows us to build the state $\boldsymbol{x^e}$ at time T_{in} (with the exception of $\boldsymbol{A^{bias}}$ and $\boldsymbol{\Omega^{bias}}$). Since our simulated measurements are corrupted by noise and also include a bias on the IMU and an error on the extrinsic camera-IMU calibration, the values obtained with the closed-form solution will differ from the true values.

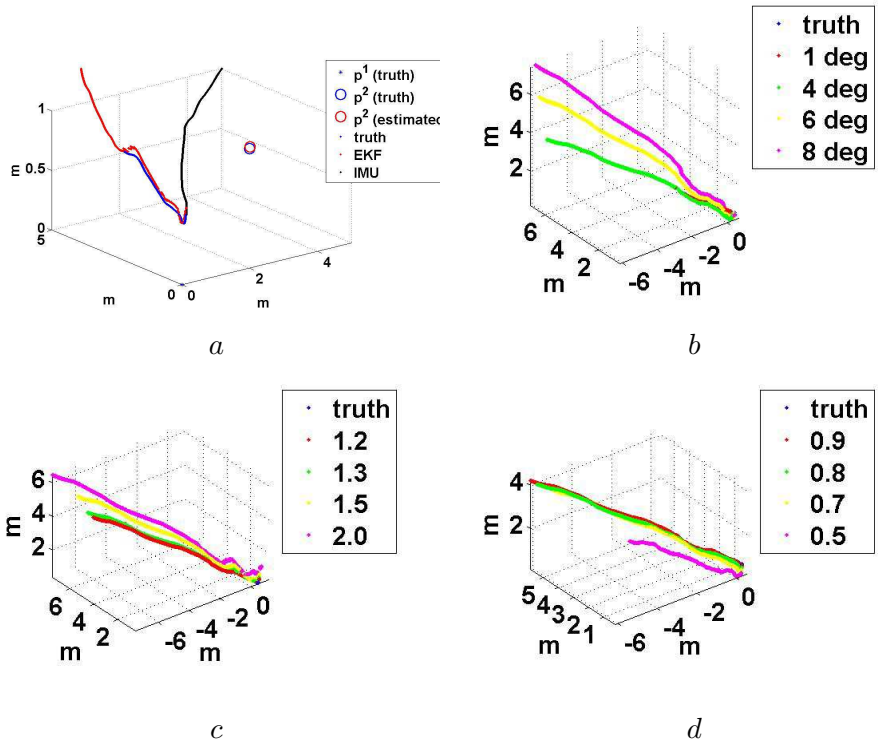
In this section, we investigate how the performance of the EKF depends on its initialization and how this performance can be improved by using the closed-form solution to initialize the state. Since the closed-form solution does not provide the initial $\boldsymbol{A^{bias}}$ and $\boldsymbol{\Omega^{bias}}$, their initial values will be set to zero.

3.4.1 Results

We adopted the same parameters that characterize the simulations described in the previous section. As previously explained, we need at least two point features. The first one is the origin and the second one is characterized by $p_x = 2m$ and $p_z = 1m$. We also considered the case of more than two point-features ($N_f \geq 3$), obtaining similar results in terms of performance and, for the sake of brevity, in the following we only refer to the case of $N_f = 2$.

We first investigate the convergence of the EKF vs the initialization of the state. In all the considered initializations we set the initial accelerometer and gyroscope biases to zero. In general, the EKF diverges when: (a) the initial scale factor error exceeds 20%; (b) the initial attitude error exceeds 4 deg. These conclusions on the EKF convergence have been obtained by running many simulations with the settings specified in 3.3.1 and 3.3.2. As an illustration, we display here the results obtained with a particular trial. Figures $3.2a-d$ display the trajectories estimated by the EKF when the initial state differs from the true state because of an error on the absolute scale and on the attitude (as said, the initial state is also affected by an error on the inertial sensors' biases since they are always initialized to zero). Figure $3.2a$ displays the true trajectory (blue) together with the one estimated by only using inertial measurements (black) and the one estimated by the EKF with an initial absolute scale set to 1.1 times the true value and an error of 1 deg on the roll, pitch and yaw angles. Figure $3.2b$ displays the trajectories estimated by the EKF with an initial state affected by an error on the attitude (same error on the roll, pitch and yaw) and correct absolute scale. Figures $3.2c$ and d display the trajectories estimated by the EKF with an initial state affected by an error on the absolute scale and correct attitude.

By using the first 6 camera observations (i.e. by considering the time interval $[T_{in} = 0,\ T_{fin} = 0.6]s$) we obtain the initial position $[0.4961,\ 0.4975,\ 0.5017]m$, the initial speed $[0.1024,\ 0.1028,\ 0.1222]m\ s^{-1}$ and the initial attitude $q = 1 - 4.3\ 10^{-6} + (1.0i - 2.3j + 1.6k)\ 10^{-3}$, which corresponds to the attitude $roll = 0.11\ deg$, $pitch = -0.26\ deg$ and $yaw = 0.18\ deg$. By



a *b* *c* *d*

Fig. 3.2 Fig *a* displays the true trajectory (blue), the trajectory estimated by the EKF (red) with initial errors (10% on the scale and 1 *deg* on the attitude) and the trajectory estimated by only using inertial measurements (black). Figs *b*, *c* and *d* display the trajectories estimated by the EKF when the initial state is affected by an error on the attitude (*b*) and on the absolute scale (larger and smaller than the true one, in *c* and *d*, respectively).

running many simulations, we found that the initial state determined through the closed-form solution is never affected by an error larger than 8% regarding the absolute scale and than 0.7 *deg* regarding the attitude.

4

Data Association

All the results illustrated so far assume a perfect data association. In autonomous robotics, any feature-based approach to SfM, must be able, first of all, to autonomously remove all the outliers. In this chapter we briefly address this important problem. In section 4.1 we present a method able to detect outliers by using the closed-form solution described in the previous chapter with a single point feature. In sections 4.2 and 4.3 we describe a simpler approach obtained by combining some concepts from epipolar geometry with basic concepts in inertial navigation.

4.1 Outliers detection by using the closed-form solution

In this section we show the use of the closed-form solution to detect outliers in the features matching process. We start by remarking that the closed-form solution can work even in the extreme case of a single feature correspondence in consecutive images. Specifically, with a single point feature, it is possible to determine the following physical quantities that regard the vehicle motion: the absolute roll, the absolute pitch and the three components of the speed in the local frame. This

suggests to find inliers by a simple histogram voting. In particular, we can build an independent histogram for each of the previous physical quantities. According to table 3.1, we need at least 5 consecutive images to have a unique solution for these quantities. Let us suppose to have $n_i \geq 5$ images and that the same N_f features have been detected and correctly matched in the first $n_i - 1$ images. Now, we consider the last image, i.e. the n_i^{th}. We start from a given matching of the previous N_f features with N_f image points detected on this n_i^{th} image. We will use the closed-form solution to detect the inliers and outliers in this last matching by proceeding as follows.

We consider each feature separately. For each one we compute the three components of the speed and the roll and pitch angles. Hence, we obtain for each of these 5 quantities N_f values that correspond to the N_f features. Starting from these values we build a histogram voting for each of these 5 physical quantities. We use each histogram to remove outliers. By combining two or more histograms we can remove more outliers.

We tested this procedure on synthetic data. We adopted the same parameters used in the simulations described in the previous chapter, with the exception of the number of features that are now $N_f = 50$. We considered $n_i = 8$ consecutive images and we correctly matched these 50 features in the first 7 images, while in the matching with the last image we introduced a given percentage (P_o) of outliers. Specifically, we considered the following two cases, $P_o = 10\%, 40\%$. We ran 1000 trials for each case. Table 4.1 displays the results. We provide the percentage of the inliers that have been missed by our procedure (the Missed Inliers (M.I.) in the second, third and fourth column) and the percentage of outliers classified as inliers (the False Inliers (F.I.) in the fifth, sixth and seventh column). In particular, the results reported in the second and fifth column are obtained by only using the histograms obtained with the three components of the speed ($V_x V_y V_z$); the results in the third and in the sixth column are obtained by only using the histograms obtained with the roll and pitch angles (r p); finally, the results in the fourth and in the seventh column are obtained by using all the previous histograms.

P_o	M.I. (Speed) $V_xV_yV_z$	M.I. $r\ p$	M.I. $V_xV_yV_z\ r\ p$	F.I. $V_xV_yV_z$	F.I. $r\ p$	F.I. $V_xV_yV_z\ r\ p$
10%	22%	13%	29%	11%	0.7%	0.07 %
40%	33%	18%	42%	14%	6.9%	0.09 %

Table 4.1 Simulation results on outliers and inliers detection.

4.2 Essential matrix and epipolar constraint

The previous procedure requires to already have a correct data association in the first $n_i - 1$ images. We now want to refer to two consecutive images, without assuming any prior knowledge. We will refer to the RANdom SAmple Consensus (RANSAC) [15], which is the most popular method to estimate a given model from a set of observed data that contains outliers. In the case of two camera images, the model consists of the rotation and the translation up to a scale occurred between the two camera poses, i.e., the transformation up to a scale between the two reference frames attached to the camera at the two times when the two images are taken. This model can be obtained from matched points starting from the epipolar constraint.

Let us consider two camera images and let us denote with $\boldsymbol{m} \equiv [u,\ v,\ 1]^T$ and $\boldsymbol{m'} \equiv [u',\ v',\ 1]^T$ the coordinates (up to a scale) of the same scene point from the two camera configurations, respectively. Additionally, we denote with $\boldsymbol{s} \equiv [s_x,\ s_y,\ s_z]^T$ the translation between the two camera frames and with R the rotation between them. The epipolar constraint [19] is the following homogeneous equation:

$$\boldsymbol{m'}^T\ E\boldsymbol{m} = 0 \tag{4.1}$$

where $E = [\boldsymbol{s}]_\times R$ is the essential matrix and $[\boldsymbol{s}]_\times \equiv \begin{bmatrix} 0 & -s_z & s_y \\ s_z & 0 & -s_x \\ -s_y & s_x & 0 \end{bmatrix}$.

The epipolar constraint can be considered an equation in the unknown rotation matrix R and the unknown translation $\boldsymbol{s}$. Since it is a homogeneous equation, it is independent of the magnitude of the translation. For this reason, it is convenient to adopt the following

parametrizations:

(1) $\boldsymbol{s} = \rho[\cos\phi,\ \sin\phi,\ 0]^T$ for a general planar motion;
(2) $\boldsymbol{s} = \rho[\cos\theta\cos\phi,\ \cos\theta\sin\phi,\ \sin\theta]^T$ for a general $3D$ motion.

In both cases, the magnitude of the translation can be neglected. For a general $3D$ motion, the model that must be inferred from matched points consists of five parameters. Three of them characterize the rotation, i.e., define the matrix R. The remaining two parameters are the previous θ and ϕ. Hence, at least five correspondences must be used to estimate the model [43]. In the planar case, the model consists of two parameters: the former characterizes the rotation, the latter (the previous ϕ) the translation up to a scale. Therefore, to detect and remove outliers, a five-point and a two-point RANSAC can be adopted in the $3D$ and planar case, respectively.

4.3 Simplified model

When monocular vision is coupled with an IMU, the rotation occurred between two camera configurations can be easily inferred from the inertial measurements. Hence, we can assume that the matrix R is known. This fact, significantly simplifies the epipolar constraint, by reducing the number of unknowns in the essential matrix. In the following, we denote with $\boldsymbol{m''} \equiv [u'',\ v'',\ 1]^T$, the vector obtained by pre multiplying $\boldsymbol{m}$ by R and then by dividing all the components by the third component. The epipolar constraint (i.e., the equation (4.1)) reads:

$$\boldsymbol{m'}^T \begin{bmatrix} 0 & 0 & \sin\phi \\ 0 & 0 & -\cos\phi \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \boldsymbol{m''} = 0 \tag{4.2}$$

in the planar case and

$$\boldsymbol{m'}^T \begin{bmatrix} 0 & -\sin\theta & \cos\theta\sin\phi \\ \sin\theta & 0 & -\cos\theta\cos\phi \\ -\cos\theta\sin\phi & \cos\theta\cos\phi & 0 \end{bmatrix} \boldsymbol{m''} = 0 \tag{4.3}$$

in the $3D$ case. We discuss these two cases separately.

4.3.1 Planar motion

A single correspondence is sufficient to get the model (i.e., the parameter ϕ). From equation (4.2) we have: $u' \sin\phi - v' \cos\phi - u'' \sin\phi + v'' \cos\phi = 0$. Hence:

$$\phi = atan\left(\frac{v' - v''}{u' - u''}\right) \tag{4.4}$$

Hence, 1-point RANSAC can be easily adopted to detect and remove outliers [50].

4.3.2 3D motion

In this case the model consists of two parameters, ϕ and θ. We need to use the epipolar equation in (4.3), for two distinct points. Let us denote them by the subscript 1 and 2. By a direct computation it is possible to obtain the following expression for ϕ:

$$\phi = \tag{4.5}$$
$$= atan\left(\frac{u''_2 v'_2 v'_1 - u''_2 v'_2 v''_1 + v''_2 u'_2 v''_1 - v''_2 u'_2 v'_1 - v''_1 u'_1 v''_2 + v''_1 u'_1 v'_2 + u''_1 v'_1 v''_2 - u''_1 v'_1 v'_2}{u''_2 v'_2 u'_1 - u''_2 v'_2 u''_1 - v''_2 u'_2 u'_1 + v''_2 u'_2 u''_1 + v''_1 u'_1 u'_2 - v''_1 u'_1 u''_2 - u''_1 v'_1 u'_2 + u''_1 v'_1 u''_2}\right)$$

Then, one of the two epipolar equations (e.g., the first one) can be used to obtain θ:

$$\theta = atan\left(\frac{(u''_1 - u'_1)\sin\phi + (v'_1 - v''_1)\cos\phi}{u''_1 v'_1 - v''_1 u'_1}\right) \tag{4.6}$$

Hence, 2-point RANSAC can be easily adopted to detect and remove outliers.

5

Conclusion and discussion

In this paper we provided a series of new results about the visual-inertial structure from motion problem. In particular, these new results significantly improve the current state of the art by providing new properties related to three fundamental issues: observability properties, resolvability in closed-form and data association. These results are important from a technological point of view. Additionally, they can provide a new insight for the comprehension of the process of vestibular and visual integration, which has been investigated in the framework of neuroscience.

5.1 Observability properties

The analysis started by considering the common formulation, which assumes three orthogonal accelerometers and three orthogonal gyroscopes. Theorem 2 provides the most general result in this sensor settings, by accounting biased inertial measurements, unknown magnitude of the gravity, unknown camera-IMU extrinsic calibration and the case of any number of point-features. This theorem states that all the independent observable modes are: the positions in the local frame of all

the observed features, the three components of the speed in the local frame, the biases affecting the inertial measurements, the roll and the pitch angle, the magnitude of the gravity and the transformation between the camera and IMU frames. The fact that the yaw angle is not observable is an obvious consequence of the system invariance under rotation about the gravity vector.

Then, the analysis is extended to derive the observability properties when the number of inertial sensors is reduced. The case of a single accelerometer and no gyroscope is investigated. Theorem 3 basically states that, if the camera is extrinsically calibrated, the observability properties do not change. On the other hand, as the camera is not extrinsically calibrated, an internal symmetry arises (see theorem 5). Specifically, from the sensor measurements collected during any time interval and by accomplishing any motion, it is not possible to distinguish all the physical quantities rotated around the accelerometer axis. This means that, in this setting, it is not possible to fully perceive self-motion. As an additional accelerometer is introduced, the system gains again the full observability (theorem 4). These results clearly show that the information provided by an IMU is redundant. Additionally, these results are consistent with our knowledge about the vestibular system, which provides balance in most mammals. Indeed, the otoliths, which indicate linear accelerations, consists of two organs (the *utricle* and the *saccule*) able to sense the acceleration only along two independent axes.

5.2 Resolvability in closed-form and data association

In the paper, we discussed the resolvability of the VI-SfM problem, in closed form. In the case of visual measurements only, the SfM has been solved up to a scale [6, 9, 18, 29, 43] and a closed-form solution has also been derived [18, 29, 43], allowing the determination of the three-dimensional structure of the scene, without the need for any prior knowledge. In chapter 3 we extend this deterministic solution to the case of the VI-SfM problem. This makes possible to introduce deterministic algorithms, which are very important to efficiently and robustly initialize any filter-based or optimization approach.

The most useful applications of this closed-form solution will be in

all the applicative domains that need to solve the SfM problem with low-cost sensors and that do not demand any infrastructure (e.g., in GPS denied environment). In these contexts, there is often the need to perform the estimation without any prior knowledge. Typical examples of applicative domains are the emergent fields of space robotics [42], humanoid robotics and unmanned aerial navigation in urban-like environments [53], where the use of the GPS is often forbidden.

The closed-form solution could also play an important role in the framework of neuroscience by providing a new insight on the process of vestibular and visual integration for depth perception and self-motion perception. The influence of extra retinal cues in depth perception has extensively been investigated in the last decades [7, 28, 33, 54]. In particular, a very recent study investigates this problem by performing trials with passive head movements [11]. The conclusion of this study is that the combination of retinal image with vestibular signals can provide rudimentary ability to depth perception. Our findings could provide a new insight to this integration mechanism for depth perception since, according to the closed-solution here provided, by combining retinal image with vestibular signals it is possible to determine the scale factor even without any knowledge about the initial speed. Our findings also show that it is possible to easily distinguish linear acceleration from gravity. Specifically, the closed-form solution performs this determination by a very simple matrix inversion. This problem has also been investigated in neuroscience [3, 39]. Our results could provide a new insight to this mechanism since they clearly characterize the conditions (type of motion, features arrangement in the 3D space) under which this determination can be performed.

Finally, in the last chapter of the paper, we discussed very briefly the data association problem. We showed that, since the rotation occurred between two camera frames can be efficiently obtained by integrating the inertial measurements, the parameters necessary to characterize the model reduce to two for a general $3D$ motion and to one in the planar case. This makes possible the use of a 2-point RANSAC in the general $3D$ case and a 1-point RANSAC in the planar case.

References

[1] Anguelova M., Non linear Observability and Identifiability: General Theory and a Case Study of a Kinetic Model, PhD thesis, Chapter 4, Goteborg, April 2004

[2] Armesto L., Tornero J., and Vincze M., Fast Ego-motion Estimation with Multi-rate Fusion of Inertial and Vision, The International Journal of Robotics Research, 26, pp 577–589, (2007)

[3] Angelaki D. E., Shaikh A. G., Green A. M. and Dickman J. D., Neurons compute internal models of the physical laws of motion, Nature, 430, pp 560–564, (2004)

[4] Berthoz A., Pavard B. and Young L.R., Perception of Linear Horizontal Self-Motion Induced by Peripheral Vision (Linearvection) Basic Characteristics and Visual-Vestibular Interactions, Exp. Brain Res. 23, pp 471–489, (1975).

[5] Bryson M. and Sukkarieh S., Observability Analysis and Active Control for Airbone SLAM, IEEE Transaction on Aerospace and Electronic Systems, 44, no. 1, pp 261–280, (2008)

[6] ChiusoA., Favaro P., Jin H. and Soatto S., "Structure from Motion Causally Integrated Over Time", IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), pp 523–535, (2002)

[7] Cornilleau-Peres V. and Droulez J., The visual perception of three-dimensional shape from self-motion and object-motion, Vision research, 34, pp 2331–2336, (1994)

[8] Corke P., Lobo J. and Dias J., An Introduction to Inertial and Visual Sensing, International Journal of Robotics Research 26, pp 519-535, (2007)

[9] Davison A. J., Reid I.D., Molton N. D. and Stasse O., "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), pp 1052–1067, (2007)

[10] Dias J., Vinzce M., Corke P. and Lobo J., Editorial: Special Issue: 2nd Workshop on Integration of Vision and Inertial Sensors, The International Journal of Robotics Research, 26, 6: pp. 515-517 (2007).

[11] Dokka K., MacNeilage P. R., De Angelis G. C. and Angelaki D. E., Estimating distance during self-motion: a role for visual-vestibular interactions, Journal of Vision 11(13):2, pp. 1-16, (2011)

[12] Dong-Si T.C. and Mourikis A. I., Estimator Initialization in Vision-aided Inertial Navigation with Unknown Camera-IMU Calibration, International Conference on Intelligent Robot and System, Iros 2012, Vilamoura, Portugal, October 2012.

[13] J. Farrell A., Aided Navigation: GPS and High Rate Sensors. McGraw- Hill, 2008.

[14] Fetsch C. R., De Angelis G. C. and Angelaki D. E., Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory, Eur J Neurosci. 31(10): pp. 1721-1729, (2010)

[15] Fischler M. A. and Bolles R. C., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, vol. 24, no. 6, pp. 381395, (1981)

[16] Gemeiner P., Einramhof P. and Vincze M., Simultaneous Motion and Structure Estimation by Fusion of Inertial and Vision Data, The International Journal of Robotics Research 26: pp. 591-605, (2007)

[17] Goldstein, H. Classical Mechanics, 2nd ed. Reading, MA: Addison-Wesley, 1980

[18] Hartley R. I., In Defense of the Eight-Point Algorithm". IEEE Transaction on Pattern Recognition and Machine Intelligence 19 (6): pp. 580–593, (1997)

[19] Hartley R. I. and Zisserman A., Multiple view Geometry in Computer Vision, Cambridge University Press, ISBN:0521540518

[20] Hermann R. and Krener A. J., Nonlinear Controllability and Observability, Transaction On Automatic Control, AC-22(5): pp. 728-740, (1977)

[21] Hesch J. A., Kottas D. G., Bowman S. L. and Roumeliotis S. I., Towards consistent vision-aided inertial navigation, Int. Workshop on the Algorithmic Foundations of Robotics, Cambridge, MA, Jun. 1315, 2012.

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

[23] Kelly J. and Sukhatme G., Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration, International Journal of Robotics Research, vol. 30, no. 1, pp. 56–79, (2011)

[24] Kim J. and Sukkarieh S., Real-time implementation of airborne inertial-SLAM, Robotics and Autonomous Systems, 55, pp. 62-71, (2007)

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

[26] Lhuillier M., Automatic structure and motion using a catadioptric camera, IEEE workshop on omnidirectional vision, 2005.

[27] Li M. and Mourikis A. I., Improving the accuracy of EKF-based visual inertial odometry, IEEE International Conference on Robotics and Automation, Icra 2012, St. Paul, MN, May 1418, 2012

[28] Longuet-Higgins H. C. and Prazdny K., The interpretation of a moving retinal image, Royal Society of London B: Biological Sciences, 208, pp 385–397, (1980)

[29] Longuet-Higgins H. C. , A computer algorithm for reconstructing a scene from two projections, Nature 293: 133–135, (1981)

[30] Lupton T. and Sukkarieh S., Removing scale biases and ambiguity from 6DoF monocular SLAM using inertial, International Conference on Robotics and Automation, Icra 2008

[31] Lupton T. and Sukkarieh S., Efficient Integration of Inertial Observations into Visual SLAM without Initialization, International Conference on Inteligent Robot and System, Iros 2009

[32] Lupton T. and Sukkarieh S., Visual-Inertial-Aided Navigation for High-Dynamic Motion in Built Environments Without Initial Conditions, Transaction on Robotics, Volume 28, Issue 1 pp 61–76, (2012)

[33] MacNeilage P. R., Banks M. S. De Angelis G. C. and Angelaki D. E, Vestibular heading discrimination and sensitivity to linear acceleration in head and world coordinates, J. of Neuroscience, 30, pp 9084–9094, (2010)

[34] Martinelli A., State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration, Transactions on Robotics, Vol. 27, No. 2, pp 239–255, (2011)

[35] Martinelli A., Closed-Form Solution for Attitude and Speed Determination by Fusing Monocular Vision and Inertial Sensor Measurements, International Conference on Robotics and Automation, Icra 2011, Shangai

[36] Martinelli A., Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination, Transaction on Robotics, Volume 28, Issue 1 (February), pp 44–60, (2012)

[37] Martinelli A., Resolvability of Visual-Inertial Structure from Motion in Closed-form, INRIA technical Report (2012)

[38] Martinelli A., Closed-form solution of visual-inertial structure from motion, International Journal of Computer Vision, DOI 10.1007/s11263-013-0647-7, (2013)

[39] Merfeld D. M., Zupan L. and Peterka R. J., Humans use internal models to estimate gravity and linear acceleration, Nature, 398, pp 615–618, (1999)

[40] Mirsky L., An Introduction to Linear Algebra, 1990, ISBN 0-486-66434-1

[41] Mirzaei F. M. and Roumeliotis S. I., A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation, IEEE Transactions on Robotics, 2008, Vol. 24, No. 5, pp. 1143–1156, (2008)

[42] Mourikis A. I., Trawny N., Roumeliotis S. I., Johnson A., Ansar A. and Matthies L., Vision-Aided Inertial Navigation for Spacecraft Entry, Descent, and Landing, Transactions on Robotics, 25(2), pp. 264–280, (2009)

[43] Nistér D., An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 26(6):756-770, (2004)
[44] Nistér D., Preemptive ransac for live structure and motion estimation, Machine Vision and Applications, 16, pp. 321–329, (2005)
[45] Numerical Recipes in C, Chapter 15.5: Nonlinear models
[46] Quian G., Zheng Q. and Chellappa R., Reduction of inherent ambiguities in structure from motion problem using inertial data, IEEE International Conference on Image Processing, (2000)
[47] Scaramuzza, D., 1-Point-RANSAC Structure from Motion for Vehicle-Mounted Cameras by Exploiting Non-Holonomic Constraints, International journal of computer vision, 95:74–85, (2011)
[48] Strelow D. and Singh S., Motion estimation from image and inertial measurements, International Journal of Robotics Research, 23(12), (2004)
[49] Tardif J., Pavlidis Y. and Daniilidis K., Monocular visual odometry in urban environments using an omnidirectional camera, International Conference on Intelligent Robot and System, Iros 2008.
[50] Troiani C., Martinelli A., Laugier C., Scaramuzza D., 1-Point-based Monocular Motion Estimation for Computationally- limited Micro Aerial Vehicles, European Conference on Mobile Robotics, Barcelona, Spain, ECMR 2013.
[51] Veth M. and Raquet J., Fusing low-cost image and inertial sensors for passive navigation, Journal of the Institute of Navigation, vol. 54(1), (2007)
[52] Weiss S., Scaramuzza D. and Siegwart R., Monocular-SLAM-Based Navigation for Autonomous Micro Helicopters in GPS-Denied Environments, Journal of Field Robotics, Volume 28, issue 6, (2011)
[53] Weiss. S., Vision Based Navigation for Micro Helicopters, PhD thesis, Diss. ETH No. 20305
[54] Wexler M., Panerai F., Lamouret I. and Droulez J., Self-motion and perception of stationary objects, Nature, 409, pp 85–88, (2001)
[55] Woodman O. J., An introduction to inertial navigation, Technical Report, University of Cambridge, Computer Laboratory, 2007, UCAM-CL-TR-696