Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination by Fusing Vision and Inertial Measurements

Agostino Martinelli

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination by Fusing Vision and Inertial Measurements

Agostino Martinelli

N° 7530

Février 2011

Thèmes COG et NUM

ISSN 0249-6399 ISRN INRIA/RR--7530--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
GRENOBLE - RHÔNE-ALPES

Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination by Fusing Vision and Inertial Measurements

Agostino Martinelli

Thèmes COG et NUM — Systèmes cognitifs et Systèmes numériques
Équipe-Projet Emotion

Rapport de recherche n° 7530 — Février 2011 — 43 pages

Abstract: This paper analyzes the problem of data fusion when the adopted sensors are a monocular camera and inertial sensors (i.e. one tri-axial accelerometer and one tri-axial gyrometer). The paper provides two main contributions. The first is the analytical derivation of all the observable modes, i.e. all the physical quantities that can be estimated by only using the information in the sensor data acquired during a short time interval. Specifically, the observable modes are the vehicle speed and attitude, the absolute scale and the bias affecting the inertial measurements. This holds even in the case when the camera only observes a single point feature. The analytical derivation of the previous observable modes is based on a non standard observability analysis, which fully accounts the system non linearities. The second contribution is the analytical derivation of closed-form solutions which analytically express all the previous observable modes in terms of the sensor measurements collected during a very short time interval. This allows introducing a very simple and powerful new method able to simultaneously estimate all the observable modes without the need of any initialization or a priori knowledge. Both the observability analysis and the derivation of the closed-form solutions are carried out in several different contexts, including the case of biased and unbiased inertial measurements, the case of a single and multiple features, and in presence and absence of gravity. The performance of the proposed approach is evaluated via extensive Monte Carlo simulations and by using two distinct real data sets.

Key-words: Sensor Fusion, Inertial Sensors, Vision, Non linear Observability, Aerial Robotics

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Résumé : Ce document considère le problème de la fusion sensorielle dans le cas très importante des capteurs inertielles et vison. Ce document contient deux contributions. La première est la dérivation analytique de tous les modes observables, c'est á dire les quantités physiques qu'on peut estimer a partir des données sensorielles. Plus précisément, les modes observables sont: la vitesse et l'orientation du véhicule, le facteur d'échelle et le bias qui est dans les données inertielles. Ces quantités physiques sont observables même dans le cas ou seulement une feature est disponible. La deuxième contribution est la dérivation d'une formule analytique qui permet d'exprimer les modes observables précédents en fonction des données sensorielles (de la camera et des capteurs inertielles) prises dans un intervalle de temps très court. Cela permet de déterminer les modes observables sans la nécessite d'aucune connaissance a priori et sans aucune initialisation. Le document contient simulations et données réels pour tester la méthode d'estimation proposée.

Mots-clés : Fusion Sensoriel, Capteurs inertiels, Vision, Observabilité, Robotique Aérienne

1 Introduction

In recent years, vision and inertial sensing have received great attention by the mobile robotics community. These sensors require no external infrastructure and this is a key advantage for robots operating in unknown environments where GPS signals are shadowed. Additionally, these sensors have very interesting complementarities and together provide rich information to build a system capable of vision-aided inertial navigation and mapping and a great effort has been done very recently in this direction (e.g. [28], [4], [1] and references therein).

When fusing vision and inertial measurements, the following two issues must be addressed:

1. find all the physical quantities that the information contained in the sensor data allows us to estimate;

2. find a reliable and efficient method to estimate the previous physical quantities starting from the raw sensor data.

Throughout this paper, we will call these physical quantities the *Observable Modes*.

The previous two issues have been addressed in the case of vision measurements alone. To this regard, one of the main results achieved by the computer vision community is the capacity of performing structure from motion by only using the data provided by a single camera in an unknown environment [6]. The eight-point algorithm is one of the most famous solutions to this problem [6, 13]. The algorithm's name derives from the fact that it estimates the essential matrix from eight or more corresponding image points. However, it is also possible modify the original algorithm in order to achieve the same result with fewer than eight points (at least five, [29]). More recently, the same structure from motion problem has been considered by the mobile robotics community under the name of visual *SLAM* (Visual Simultaneous Localization and Mapping). In [8], a new algorithm, called MonoSLAM, has been introduced. This algorithm is based on the Extended Kalman Filter and it is able to perform real-time localization and mapping with a single freely moving camera as the only data source. On the other hand, when a single camera is adopted to perform the structure from motion, the estimation can be done up to a scale factor. This is due to the lack of metric information. Usually, the initial camera observations[1] are used in order to define a common scale factor which must be the same during the entire camera motion. On the other hand, the camera observations are affected by measurement errors. Hence, the error on the scale factor accumulates over the motion when the camera moves in an unknown environment. As a result, the error on this scale factor is affected by an unbounded drift. This fact becomes a serious inconvenience for practical applications as soon as the environment is large. In order to overcome this inconvenience, it is necessary to fuse the camera data with the data provided by at least one supplementary sensor. It is very reasonable to expect that, when fusing vision and inertial measurements, the scale factor is an observable mode and can be obtained by

[1] Throughout this paper, we will adopt the term *camera observation* to mean the bearing measurements provided by the camera from a single pose

a closed-form solution. Let us consider the trivial case where a robot, equipped with a bearing sensor (e.g. a camera) and an accelerometer, moves on a line (see fig 1). If the initial speed in A is known, by integrating the data from the accelerometer, it is possible to determine the robot speed during the subsequent time steps and then the distances $A-B$ and $B-C$ by integrating the speed. The lengths $A-F$ and $B-F$ are obtained by a simple triangulation by using the two angles β_A and β_B from the bearing sensor. Let us now assume that the initial speed v_A is unknown. In this case, all the previous segment lengths can be obtained in terms of v_A. In other words, we obtain the analytical expression of $A-F$ and $B-F$ in terms of the unknown v_A and all the sensor measurements performed while the robot navigates from A to B. By repeating the same computation with the bearing measurements in A and C, we have a further analytical expression for the segment $A-F$, in terms of the unknown v_A and the sensor measurements performed while the robot navigates from A to C. The two expressions for $A-F$ provide an equation in the unknown v_A. By solving this equation we finally obtain all the lengths in terms of the measurements performed by the accelerometer and the bearing sensor.

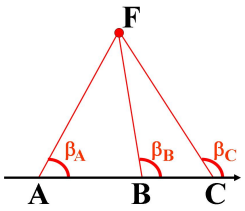


Figure 1: A robot equipped with an accelerometer and a camera moves on a line. The camera performs three observations of the feature in F, repsectively from the points A, B and C.

The previous example is very simple because of several unrealistic restrictions. First of all, the motion is constrained on a line. Additionally, the accelerometer provides gravity-free and unbiased measurements. In this paper we will relax these restrictions by considering the case of a robot equipped with IMU[2] and bearing sensors. We want to know which are the observable modes, namely the physical quantities that we can estimate without any a priori knowledge (i.e. by only collecting the data from the previous sensors during a short time interval). For instance, are the scale factor, the robot speed and the robot orientation observable modes? Are they observable modes even in the case of biased IMU measurements? Are the bias (affecting the IMU measurements) observable modes? And more importantly: is it possible to determine all these quantities by a closed form solution (as in the simple unrealistic example previously provided)?

An answer to the first three questions can be found by applying the method introduced in [23] and [24], where a non standard observability analysis, based on the new concept of continuous symmetry, has been introduced. The advantages of this non standard observability analysis is that, in contrast to previous

[2]Throughout this paper, we will adopt the term IMU (Inertial Measurement Unit) to indicate the sensor assembling constituted by three orthogonal accelerometers and three orthogonal gyrometers.

approaches, it is able not only to check whether a given state is observable or not, but, in the negative case, it is also able to detect the quantities which are observable. In particular, by analyzing the continuous symmetries of a given system, it is possible to obtain a system of partial differential equations. The observable modes are all the independent solutions of this system of partial differential equations. In [25] the observable modes have been provided in the case of one feature and in the case of unbiased IMU measurements. Additionally, a closed-form solution has been derived in this special case and the performance of an estimator based on an Extended Kalman Filter has also been discussed. In this paper, we also provide the analytical derivation of the observable modes starting from the theory developed in [23] and [24]. Additionally, also new realistic contexts are considered, by including the case of biased and unbiased inertial measurements, the case of single and multiple features, and in presence and absence of gravity.

The paper is organized as follows. Section 3 illustrates and summarizes the basic steps of the method introduced in [23] and [24], by dealing with a simple $2D$ localization problem. Section 4 provides a mathematical description of the system, constituted by a mobile robot moving in $3D$ and equipped with vision and IMU sensors. Starting from this description, in section 5 the observability analysis is performed. This analysis allows us to derive all the observable modes. Then, in section 6, we provide closed-form expressions of the observable modes in terms of the sensor measurements. The performance of the method in estimating the observable modes is evaluated by using synthetic and real data (section 7). Finally, conclusions are provided in section 8.

2 Related Works

The problem of fusing the vision and inertial data has been extensively investigated in the past. A special issue of the *International Journal of Robotics Research* has recently been devoted to this important topic [9]. In [7], a tutorial introduction to the vision and inertial sensing is 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 [30] the inertial measurements are used in order 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 [3], these sensors are used to perform egomotion estimation. The sensor fusion is obtained with an Extended Kalman Filter (EKF) and with an Unscented Kalman Filter (UKF). The approach proposed in [11] 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 [34] 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 SLAM, an EKF has been adopted in [16] to perform $3D-$SLAM by fusing inertial and vision measurements. It was remarked that any inconsistent attitude update severely affects any SLAM so-

lution. 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. Regarding the robot attitude, in [5] it has been noted that roll and pitch angles remain more consistent than the heading.

When using an EKF, an important issue which arises is the initialization problem. Indeed, because of the system non-linearities, an erroneous initialization can irreparably damage the entire estimation process. This problem has been considered in [21]. 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 which does not suffer from the initialization of the speed and of the orientation [22].

In [27] it is introduced a measurement model that is able to express the geometric constraints that arise when the same feature is observed from multiple camera poses. This measurement model does not require to include the feature position in the state which is estimated by an EKF. A similar idea is adopted in [35]. Also in this case, the problem of estimating the location of each feature is avoided, by using epipolar points on the image plane.

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 [33]. This algorithm is a batch method which performs SLAM from image and inertial measurements. Specifically, it minimizes a cost function by using the Leven-Marquardt algorithm. This minimization process starts by initializing the velocities, the gravity and the biases to zero. In [10] the graphical SLAM approach has been suggested to fuse the data from many different sensors: encoder, inertial, vision and GPS.

Finally, an important issue which arises when inertial and vision sensors are simultaneously used, is the problem of the extrinsic calibration, i.e. the estimation of the relative pose of these sensors. This problem has been approached in the past and several iterative and non-iterative solutions have been proposed. In [26] the extrinsic calibration has been performed by using an EKF. Non-iterative solutions have been proposed in [15] and [19].

3 Observable Modes and Continuous Symmetries

When a state is not observable, there are in general infinite initial states reproducing exactly the same inputs and outputs. Let us consider for instance, the $2D$ localization problem when the robot moves along a corridor, equipped with odometry sensors and sensors able to perform relative observations (e.g. bearing and range sensors). In this situation, all the initial states differing for a shift along the corridor, reproduce exactly the same inputs and outputs. Intuitively, we remark that the entire system has one continuous symmetry that is the invariance of the corridor with respect to a shift. It is obvious that the only quantities that we can estimate (i.e. the observable modes) are invariant with respect to this continuous symmetry (i.e. the robot orientation and the distance of the robot from the corridor walls). The previous consideration regarding this simple localization problem is quite trivial and it's not required to

introduce special mathematical tools. However, there are cases where deriving the observable modes is a very challenging task. The key to deal with these cases is to first provide a mathematical definition of continuous symmetry able to generalize the intuitive idea of symmetry. In [23] and [24], a procedure which allows us to analytically derive the observable modes for a generic system, has been introduced. This procedure is based on the concept of continuous symmetry, whose mathematical definition has also been provided. In this section we remind the reader the basic concepts characterizing the theory developed in [23] and [24]. For the sake of clarity, these concepts will be illustrated by referring to a simple localization problem, which is introduced in section 3.1.

3.1 A Simple Localization Problem

We consider a mobile robot moving in a $2D$-environment. The configuration of the robot in a global reference frame, can be characterized through the vector $[x_R,\ y_R,\ \theta_R]^T$ where x_R and y_R are the cartesian robot coordinates and θ_R is the robot orientation. The dynamics of this vector are described by the following non-linear differential equations:

$$\begin{cases} \dot{x}_R = v\cos\theta_R \\ \dot{y}_R = v\sin\theta_R \\ \dot{\theta}_R = \omega \end{cases} \tag{1}$$

where v and ω are the linear and the rotational robot speed, respectively. The robot is equipped with proprioceptive sensors which are able to evaluate these two speeds. We assume that a point feature exists in our environment and, without loss of generality, we fix the global reference frame onto it (see figure $2a$). The robot is also equipped with a bearing sensor (e.g. a camera), able to evaluate the bearing angle of the point feature in its own frame. Therefore, our system has the following output (see fig. $2a$):

$$y = \beta \equiv \pi - \theta_R + atan2(y_R, x_R) \tag{2}$$

We also provide the equations for the same system in polar coordinates, i.e. when the robot configuration is described by the coordinates D, $\phi_R \equiv atan2(y_R, x_R)$ and θ_R.

$$\begin{cases} \dot{D} &= v\cos(\theta_R - \phi_R) \\ \dot{\phi}_R &= \dfrac{v}{D}\sin(\theta_R - \phi_R) \\ \dot{\theta}_R &= \omega \\ y &= \beta = \pi - \theta_R + \phi_R \end{cases} \tag{3}$$

To check whether the robot configuration $[x_R,\ y_R,\ \theta_R]^T$ is observable or not, we have to prove that it is possible to uniquely reconstruct the initial robot configuration by knowing the input controls and the outputs (observations) in a given time interval. When at the initial time, the bearing angle β of the origin is available, the robot can be everywhere in the plane but, for each position, only one orientation provides the right bearing β. In fig. $2b$ all the three positions A, B and C are compatible with the observation β, provided that the robot

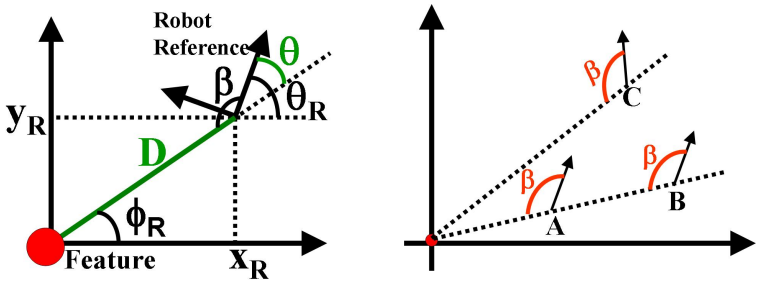


Figure 2: A simple localization problem. The robot is equipped with odometry and bearing sensors able to evaluate the angle β. In b, the three initial robot configurations are compatible with the same initial observation (β).

orientation satisfies (2). In particular, the orientation is the same for A and B but not for C.

Let us suppose that the robot moves according to the inputs $v(t)$ and $\omega(t)$. With the exception of the special motion consisting of a line passing by the origin, by only performing a further bearing observation it is possible to distinguish all the points belonging to the same line passing by the origin. In fig. 3a the two initial positions in A and B do not reproduce the same observations ($\beta_A \neq \beta_B$). On the other hand, all the initial positions whose distance from the origin is the same, cannot be distinguished independently of the chosen trajectory. In fig. 3b, the two indicated trajectories provide the same bearing observations at every time. Therefore, the dimension of the undistinguishable region is 1 and the dimension of the largest observable subsystem is $3 - 1 = 2$.

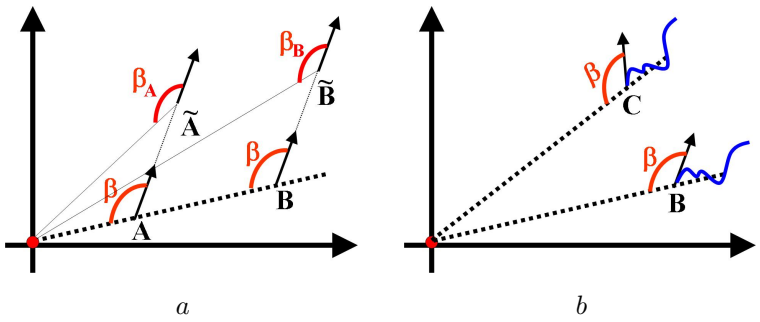


Figure 3: In a the two initial positions (A and B) do not reproduce the same observations ($\beta_A \neq \beta_B$). In b the two indicated trajectories provide the same bearing observations at every time.

We remark that the system has a continuous symmetry: the system inputs ($v(t)$ and $\omega(t)$), and outputs ($y(t)$), are invariant with respect to a rotation of the global frame about the vertical axis (in the next section we will provide a mathematical definition for a general continuous symmetry). Based on the fact that the dimension of the largest observable subsystem is two, we know that we can only estimate two independent modes. In addition, these two modes must satisfy the previous system invariance, i.e. they must be rotation

invariant. A possible choice is provided by the two quantities D and θ in figure 2*a* ($\theta \equiv \theta_R - atan2(y_R, x_R)$).

The new system is characterized by the following equations:

$$\begin{cases} \dot{D} = v\cos\theta \\ \dot{\theta} = \omega - \dfrac{v}{D}\sin\theta \\ y = \beta = \pi - \theta \end{cases} \tag{4}$$

which express the link between the new state $[D,\ \theta]^T$ and the proprioceptive data $(v,\ \omega)$ and the exteroceptive data (β).

The detection of the previous two modes and the derivation of the equations in (4) is fundamental. Indeed, estimating the original state brings inconsistencies with catastrophic consequences.

In the next subsections we remind the reader some concepts in the theory by Hermann and Krener in [14] and some basic tools introduced in [23] and [24] in order to perform the same analysis in the case of more complex systems. This will allow us to derive the observable modes when fusing monocular vision and IMU sensor measurements.

3.2 Observability Rank Criterion

A general characterization for systems in the framework of autonomous navigation, is provided by the following two equations, which describe the dynamics and the observation respectively:

$$\begin{cases} \dot{\boldsymbol{S}} = \boldsymbol{f}(\boldsymbol{S}, \boldsymbol{u}) = \boldsymbol{f_0}(\boldsymbol{S}) + \sum_{i=1}^{L} \boldsymbol{f_i}(\boldsymbol{S}) u_i \\ y = h(\boldsymbol{S}) \end{cases} \tag{5}$$

where $\boldsymbol{S} \in \Sigma \subseteq \Re^n$ is the state, $\boldsymbol{u} = [u_1,\ u_2,\ ...,\ u_L]^T$ are the system inputs, $y \in \Re$ is the output (we are considering a scalar output for the sake of clarity; the extension to a multi dimensional output is straightforward). Both the systems defined by (1-2) or (3) and the one defined by (4) can be characterized by (5). For instance, for the system in (3), we have: $\boldsymbol{S} = [D,\ \phi_R,\ \theta_R]^T$, $\boldsymbol{f_0} = [0, 0, 0]^T$, $L = 2$, $u_1 = v$, $u_2 = \omega$, $\boldsymbol{f_1}(\boldsymbol{S}) = [\cos(\theta_R - \phi_R),\ \frac{\sin(\theta_R - \phi_R)}{D},\ 0]^T$, $\boldsymbol{f_2}(\boldsymbol{S}) = [0,\ 0,\ 1]^T$, $h(\boldsymbol{S}) = \pi - \theta_R + \phi_R$.

We indicate the k^{th} order Lie derivative of a field Λ along the vector fields $\boldsymbol{v_{i_1}}, \boldsymbol{v_{i_2}}, ..., \boldsymbol{v_{i_k}}$ with $L^k_{\boldsymbol{v_{i_1}}, \boldsymbol{v_{i_2}}, ..., \boldsymbol{v_{i_K}}} \Lambda$. The definition of the Lie derivative is provided by the following two equations:

$$\begin{aligned} L^0 \Lambda &= \Lambda, \\ L^{k+1}_{\boldsymbol{v_{i_1}}, ..., \boldsymbol{v_{i_k}}, \boldsymbol{v_{i_{k+1}}}} \Lambda &= \nabla_{\boldsymbol{S}} \left(L^k_{\boldsymbol{v_{i_1}}, ..., \boldsymbol{v_{i_k}}} \Lambda \right) \cdot \boldsymbol{v_{i_{k+1}}} \end{aligned} \tag{6}$$

where the symbol "$\cdot$" denotes the scalar product and $\nabla_{\boldsymbol{S}}$ the gradient operation with respect to the state $\boldsymbol{S}$. We remark that the Lie derivatives quantify the impact of changes in the control input (u_i) on the output function (h). Additionally, we denote with $\boldsymbol{dL^k}_{\boldsymbol{f_{i_1}}, ..., \boldsymbol{f_{i_k}}} h$, the gradient of the corresponding Lie

derivative (i.e. $\boldsymbol{dL^k}_{\boldsymbol{f_{i_1}, ..., f_{i_k}}} h \equiv \nabla \boldsymbol{S} L^k_{\boldsymbol{f_{i_1}, ..., f_{i_k}}} h$), and, we denote with $d\Omega$, the space spanned by all these gradients.

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* $\boldsymbol{S}_0$ *is equal to the dimension of* $d\Omega$.

We consider again the simple example introduced in 3.1, and we show that by using the observability rank criterion, we find the same result obtained by following intuitive reasoning (i.e. that the dimension of the largest observable subsystem is 2).

The computation of the rank for the system in (3) is straightforward. From the last equation in (3), we obtain: $L^0 h = \pi - \theta_R + \phi_R$ whose gradient is $\boldsymbol{dL^0} h \equiv \boldsymbol{w_1} = [0, -1, 1]$. The first order Lie derivatives are: $L^1_{\boldsymbol{f_1}} h = -\frac{\sin(\theta_R - \phi_R)}{D}$ and $L^1_{\boldsymbol{f_2}} h = 1$. We have: $\boldsymbol{dL^1}_{\boldsymbol{f_1}} h \equiv \boldsymbol{w_2} = [\frac{\sin(\theta_R - \phi_R)}{D^2}, -\frac{\cos(\theta_R - \phi_R)}{D}, \frac{\cos(\theta_R - \phi_R)}{D}]$. It is easy to realize that each vector $\boldsymbol{w_i}$ obtained by extending the previous computation to every Lie derivative order, has the structure: $\boldsymbol{w_i} = [\varrho_i, \varsigma_i, -\varsigma_i]$. Indeed, every Lie derivative will depend on θ_R and ϕ_R only through the quantity $\theta_R - \phi_R$, whose sign changes with respect to the change $\theta_R \leftrightarrow \phi_R$. Therefore, the rank of the matrix

$$\Gamma \equiv \left\{ \begin{array}{c} \boldsymbol{w_1} \\ ... \\ \boldsymbol{w_i} \\ ... \end{array} \right\} \tag{7}$$

is equal to two. We conclude that the largest observable sub-system has dimension two as derived in section 3.1.

3.3 Continuous Symmetries

We refer to the input output system given in (5). In [23] and [24], we introduced the following definition of continuous symmetry:

Definition 1 (Continuous Symmetry) *The vector field* $\boldsymbol{w_s}(\boldsymbol{S})$ *(*$\boldsymbol{S} \in \Sigma$*) is a continuous symmetry in* $\boldsymbol{S}$ *for the system defined in (5) 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* $\boldsymbol{S}$*.*

We discuss again the simple example provided in section 3.1. We show that the previous definition corresponds with a global rotation.

For the system defined in (3) it exists only one continuous symmetry given by the vector $\boldsymbol{w_s} = [0,\ 1,\ 1]^T$ (i.e. belonging to the null space of the matrix Γ in (7)). Let us provide an intuitive interpretation of this continuous symmetry. It is possible to see that this symmetry corresponds to an infinitesimal rotation. Indeed, an infinitesimal rotation of magnitude ϵ about the vertical axis changes the state as follows [12]:

$$\begin{bmatrix} D \\ \phi_R \\ \theta_R \end{bmatrix} \rightarrow \begin{bmatrix} D \\ \phi_R \\ \theta_R \end{bmatrix} + \epsilon \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} D \\ \phi_R \\ \theta_R \end{bmatrix} + \epsilon\ \boldsymbol{w_s}$$

In [23] and [24] we proved the following fundamental property:

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

This property can be expressed by a system of partial differential equations, one for each symmetry:

$$\sum_{i=1}^{n} w_{si}(\boldsymbol{S}) \frac{\partial g}{\partial S_i} = 0 \tag{8}$$

where $w_{si}(\boldsymbol{S})$ is the i^{th} component of the symmetry $\boldsymbol{w_s}$. In other words, for every symmetry there is an associated partial differential equation which must be satisfied by all the observable modes.

We use (8) to derive the two observable modes for the system discussed in section 3.1. As previously mentioned, this system has only the symmetry $[0,\ 1,\ 1]^T$. Hence, the associated equation (8) becomes:

$$\frac{\partial g}{\partial \phi_R} + \frac{\partial g}{\partial \theta_R} = 0$$

and two independent solutions are $g = D$ and $g = \theta_R - \phi_R$. This is the same result we obtained in section 3.1.

We conclude this section with an important remark which allows us to restrict the computation when detecting the symmetries of a given system.

Remark 1 *In order to detect the symmetries of the system in (5), only the first $(n-1)-$order Lie derivatives must be computed.*

The reader is addressed to [2] where it is proven that the first $n-1$ Lie derivatives determine the observability properties of a control system.

4 The Considered System

Let us consider an aerial vehicle equipped with a monocular camera and IMU sensors. The IMU consists of three orthogonal accelerometers and three orthogonal gyrometers. We assume that the transformations among the camera frame and the IMU frames are known (we can assume that the vehicle frame coincides with the camera frame). The IMU provides the vehicle angular speed and acceleration. Actually, regarding the acceleration, the one perceived by the accelerometer ($\boldsymbol{A}$) is not simply the vehicle acceleration ($\boldsymbol{A_v}$). It also contains the gravity acceleration ($\boldsymbol{A_g}$). In particular, we have $\boldsymbol{A = A_v - A_g}$ since, when the camera does not accelerate (i.e. $\boldsymbol{A_v}$ is zero) the accelerometer perceives an acceleration which is the same of an object accelerated upward in the absence of gravity.

We will use uppercase letters when the vectors are expressed in the local frame and lowercase letters when they are expressed in the global frame. Hence, regarding the gravity we have: $\boldsymbol{a_g} = [0,\ 0,\ -g]^T$, being $g \simeq 9.8\ ms^{-2}$.

We assume that the camera is observing a point feature during a given time interval. We fix a global frame attached to this feature. The vehicle and the feature are displayed in fig 4.

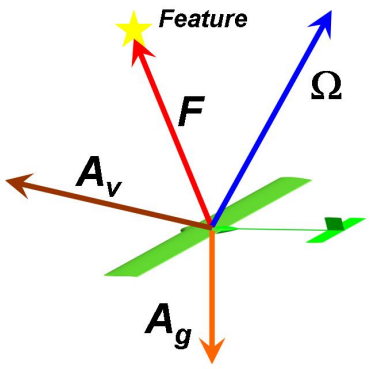


Figure 4: The feature position ($\boldsymbol{F}$), the vehicle acceleration ($\boldsymbol{A_v}$) the vehicle angular speed ($\boldsymbol{\Omega}$) and the gravity acceleration ($\boldsymbol{A_g}$).

Finally, we will adopt a quaternion to represent the vehicle orientation. Indeed, even if this representation is redundant, it is very powerful since the dynamics can be expressed in a very easy and compact notation [17].

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, $\boldsymbol{v}$ is its time derivative, i.e. the vehicle speed in the global frame ($\boldsymbol{v} \equiv \frac{d\boldsymbol{r}}{dt}$), $q = q_t + iq_x + jq_y + kq_z$ is a unitary quaternion (i.e. satisfying $q_t^2 + q_x^2 + q_y^2 + q_z^2 = 1$) and characterizes the vehicle orientation. The analytical expression of the dynamics and the camera observations 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 $\hat{w} \equiv 0 + iw_x + jw_y + kw_z$. The dynamics of the state $[\hat{r},\ \hat{v},\ q]^T$ are:

$$\begin{cases} \dot{\hat{r}} = \hat{v} \\ \dot{\hat{v}} = q\hat{A}_v q^* = q\hat{A}q^* + \hat{a}_g \\ \dot{q} = \frac{1}{2} q\hat{\Omega} \end{cases} \tag{9}$$

being q^* the conjugate of q, $q^* = q_t - iq_x - jq_y - kq_z$. We now want to express the camera observations in terms of the same state ($[\hat{r},\ \hat{v},\ q]^T$). We remark that the camera provides the direction of the feature in the local frame. In other words, it provides the unit vector $\frac{\boldsymbol{F}}{|\boldsymbol{F}|}$ (see fig. 4). Hence, we can assume that the camera provides the two ratios $y_1 = \frac{F_x}{F_z}$ and $y_2 = \frac{F_y}{F_z}$, being $\boldsymbol{F} = [F_x,\ F_y,\ F_z]^T$. We need to express $\boldsymbol{F}$ in terms of $[\hat{r},\ \hat{v},\ q]^T$. We note that the position of the feature in the frame with the same orientation of the global frame but shifted in such a way that its origin coincides with the one of the local frame is $-\boldsymbol{r}$. Therefore, $\boldsymbol{F}$ is obtained by the quaternion product $\hat{F} = -q^*\hat{r}q$. The observation function provided by the camera is:

$$h_{cam}(\hat{r},\ \hat{v},\ q) = [y_1,\ y_2]^T = \left[\frac{(q^*\hat{r}q)_x}{(q^*\hat{r}q)_z},\ \frac{(q^*\hat{r}q)_y}{(q^*\hat{r}q)_z}\right]^T \tag{10}$$

where the pedices x, y and z indicate respectively the i, j and k component of the corresponding quaternion. We have also to consider the constraint $q^*q = 1$. This can be dealt as a further observation (system output):

$$h_{const}(\hat{r},\ \hat{v},\ q) = q^*q \tag{11}$$

4.1 The Case with Multiple Features

We consider the case when the camera observes N_f features, simultaneously. We fix the global frame on one of the features. Let us denote with $\boldsymbol{d_i}$ the $3D$ vector which contains the cartesian coordinates of the i^{th} feature ($i = 0,\ 1,\ ..., N_f - 1$). We assume that the global frame is attached to the 0^{th} feature, i.e. $\boldsymbol{d_0} = [0\ 0\ 0]^T$. The new system is characterized by the state $[\hat{r},\ \hat{v},\ q,\ \hat{d}_1,\ ...,\ \hat{d}_{N_f-1}]^T$, whose dimension is $7 + 3N_f$. The dynamics of this state are given by (9) together with the equations:

$$\dot{\boldsymbol{d}}_{\boldsymbol{i}} = [0\ 0\ 0]^T\ \ i = 1,\ ..., N_f - 1 \tag{12}$$

The position $\boldsymbol{F_i}$ of the i^{th} feature in the local frame is obtained by the quaternion product $\hat{F}_i = q^*(\hat{d}_i - \hat{r})q$. The corresponding observation function is:

$$h^i_{cam} = \left[\frac{(q^*(\hat{d}_i - \hat{r})q)_x}{(q^*(\hat{d}_i - \hat{r})q)_z},\ \frac{(q^*(\hat{d}_i - \hat{r})q)_y}{(q^*(\hat{d}_i - \hat{r})q)_z}\right]^T\ \ i = 0, 1, ..., N_f - 1 \tag{13}$$

which coincides with the observation in (10) when $i = 0$. Summarizing, the case of N_f features is described by the state $[\hat{r},\ \hat{v},\ q,\ \hat{d}_1,\ ...,\ \hat{d}_{N_f-1}]^T$, whose dynamics are given in (9) and (12) and the observations are given in (13) and (11).

4.2 The Case with Bias

We consider the case when the data provided by the IMU are affected by a bias. In other words, we assume that the measurements provided by the three accelerometers and the three gyrometers are affected by an error which is not zero-mean. Let us denote with $\boldsymbol{b_A}$ and with $\boldsymbol{b_\Omega}$ the two 3D-vectors whose components are the mean values of the measurement errors from the accelerometers and the gyros, respectively. The two vectors $\boldsymbol{b_A}$ and $\boldsymbol{b_\Omega}$ are time-dependent. However, during a short time interval, it is reasonable to consider them to be constant. Under these hypotheses, the dynamics in (9) become:

$$\begin{cases} \dot{\hat{r}} &= \hat{v} \\ \dot{\hat{v}} &= q\hat{A}_v q^* = q\hat{A}q^* + q\hat{b}_A q^* + \hat{a}_g \\ \dot{q} &= \frac{1}{2}q\hat{\Omega} + \frac{1}{2}q\hat{b}_\Omega \\ \dot{\boldsymbol{b}}_{\boldsymbol{A}} = \dot{\boldsymbol{b}}_{\boldsymbol{\Omega}} &= [0\ 0\ 0]^T \end{cases} \tag{14}$$

Note that the previous equations only hold for short time intervals. In the following, we will use these equations only when this hypothesis is satisfied (in particular, during time intervals allowing the camera to perform at most ten consecutive observations).

5 Observability Properties

We investigate the observability properties of the system whose dynamics are given in (9) and whose observations are given in (10) and (11). For the sake of clarity, we discuss both the case without gravity (5.1) and with gravity (5.2). Moreover, in 5.3 we discuss the case when the camera is observing simultaneously more than one feature, namely we investigate the observability properties of the system defined by (9), (11), (12) and (13). Then, the case when the IMU sensors are affected by a bias is investigated (5.4). Finally, in 5.6, we derive two conditions which must be fulfilled in order to guarantee the observability of the modes derived in the previous subsections.

5.1 The Case without Gravity

Let us set $g = 0$ in (9). By directly computing the Lie derivatives and their gradients, it is possible to detect three independent symmetries for the resulting system. They are:

$$\boldsymbol{w_s^1} = \begin{bmatrix} 0 & -r_z & r_y & 0 & -v_z & v_y & -\frac{q_x}{2} & \frac{q_t}{2} & -\frac{q_z}{2} & \frac{q_y}{2} \end{bmatrix}^T \tag{15}$$

$$\boldsymbol{w_s^2} = \begin{bmatrix} r_z & 0 & -r_x & v_z & 0 & -v_x & -\frac{q_y}{2} & \frac{q_z}{2} & \frac{q_t}{2} & -\frac{q_x}{2} \end{bmatrix}^T$$

$$\boldsymbol{w_s^3} = \begin{bmatrix} -r_y & r_x & 0 & -v_y & v_x & 0 & -\frac{q_z}{2} & -\frac{q_y}{2} & \frac{q_x}{2} & \frac{q_t}{2} \end{bmatrix}^T$$

According to definition 1, these vectors are orthogonal to all the gradients of all the Lie derivatives. Because of the remark 1 (where $n = 10$), the detection of the previous symmetries only requires the computation of a limited number of Lie derivatives (up to the nine order). These symmetries could also be derived by remarking the system invariance with respect to rotations about all the three axes. For instance, an infinitesimal rotation of magnitude ϵ about the vertical axis changes the state as follows [12]:

$$\begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \rightarrow \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} + \epsilon \begin{bmatrix} -r_y \\ r_x \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \rightarrow \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} + \epsilon \begin{bmatrix} -v_y \\ v_x \\ 0 \end{bmatrix}$$

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

that is:

$$\begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{v} \\ \boldsymbol{q} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{v} \\ \boldsymbol{q} \end{bmatrix} + \epsilon \boldsymbol{w_s^3}$$

On the other hand, without computing the Lie derivatives, we could not conclude that the previous ones are *all* the symmetries for the considered system. In order to be sure that the previous are all the symmetries, we must detect $10 - 3 = 7$ independent Lie derivatives. In appendix A, we provide a possible choice of 7 independent Lie derivatives.

We have shown two methods to detect the symmetries of our system. The former consisted in the computation of all the Lie derivatives up to the nine order. Then, the symmetries are the vectors belonging to the null space spanned by the gradients of the previous Lie derivatives. The latter used the system invariance under rotations which allowed us to immediately detect three symmetries. Then, by providing 7 independent Lie derivatives, we concluded that these are all the symmetries.

According to property 1, for every symmetry there is an associated partial differential equation (the one provided in (8)). Hence, every observable mode must satisfy simultaneously all the three partial differential equations. Since our system is defined by 10 variables, the number of independent solutions satisfying all the three partial differential equations is $10-3 = 7$ [18]. On the other hand, their derivation, once the three symmetries are detected, is easy. Indeed, it is immediate to prove that the distance of the feature from the camera, i.e. $|\boldsymbol{r}|$, is a solution of the three equations (this can be checked by substitution for the partial differential equations associated with the symmetries in (15) but can also be proved by remarking that the scale factor is invariant under rotations). This means that the distance of the feature is observable and it is one among the 7 independent solutions. On the other hand, since the camera provides the position of the feature in the local frame up to a scale factor, having the distance means that the feature position in the local frame is also observable. Therefore, the three components of the feature position in the local frame are three independent solutions. By using quaternions, we can say that three independent solutions are provided by the components of the imaginary quaternion $q^*\hat{r}q$. Additionally, since the three partial differential equations are invariant under the transformation $\boldsymbol{r} \leftrightarrow \boldsymbol{v}$, three other independent solutions are the components of the imaginary quaternion $q^*\hat{v}q$. Physically, this means that the vehicle speed in the local frame is also observable. Finally, the last solution is q^*q since it is directly observed (see equation (11); it can be in any case verified that it satisfies the three partial differential equations).

The analytical results derived in this subsection can be summarized with the following property:

Property 2 (Observable Modes without Gravity) *Let us consider the system defined by (9), (10) and (11) in absence of gravity (i.e. $g = 0$). All the independent observable modes are 7 and they are the three components of the imaginary quaternion $q^*\hat{r}q$ (i.e. the position of the observed feature in the local frame), the three components of the imaginary quaternion $q^*\hat{v}q$ (i.e. the vehicle speed in the local frame) and the product q^*q (i.e. the norm of the the quaternion).*

5.2 The Case with Gravity

We investigate the observability properties when $g \neq 0$. The presence of the gravity breaks two of the previous three symmetries. In other words, the system

remains invariant only with respect to rotations about the vertical axis. This means that $\boldsymbol{w_s^1}$ and $\boldsymbol{w_s^2}$ are no longer symmetries for the new system. By directly computing the Lie derivatives, we were able to find nine independent Lie derivatives (the computation is similar to the one illustrated in appendix A). Hence, the system has $10 - 9 = 1$ symmetry which is ${\boldsymbol{w_s}}^3$.

The partial differential equation associated with ${\boldsymbol{w_s}}^3$ is:

$$\begin{aligned} -2r_y\frac{\partial\Lambda}{\partial r_x} + 2r_x\frac{\partial\Lambda}{\partial r_y} - 2v_y\frac{\partial\Lambda}{\partial v_x} + 2v_x\frac{\partial\Lambda}{\partial v_y} + \\ -q_z\frac{\partial\Lambda}{\partial q_t} - q_y\frac{\partial\Lambda}{\partial q_x} + q_x\frac{\partial\Lambda}{\partial q_y} + q_t\frac{\partial\Lambda}{\partial q_z} = 0 \end{aligned} \tag{16}$$

The number of independent solutions $\Lambda = \Lambda(r_x,\ r_y,\ r_z,\ v_x,\ v_y,\ v_z,\ q_t,\ q_x,\ q_y,\ q_z)$ is equal to the number of variables (i.e. 10) minus the number of equations (i.e. 1) [18]. Hence, in this case we have two additional observable modes. They are:

$$Q_r \equiv \frac{q_t q_x + q_y q_z}{1 - 2(q_x^2 + q_y^2)}; \qquad Q_p \equiv q_t q_y - q_z q_x \tag{17}$$

Also for these two solutions it is possible to find a physical meaning. They are related to the roll and pitch angles [17]. In particular, the first solution provides the roll angle which is $R = \arctan(2Q_r)$. The latter provides the pitch angle which is $P = \arcsin(2Q_p)$. Finally, we remark that the expression of the yaw, $Y = \arctan\left(2\frac{q_t q_z + q_x q_y}{1-2(q_y^2+q_z^2)}\right)$, does not satisfy (16).

The analytical results derived in this subsection can be summarized with the following property:

Property 3 (Observable Modes with Gravity) *Let us consider the system defined by (9), (10) and (11). All the independent observable modes are* 9 *and they are the* 7 *observable modes for the case without gravity together with the roll and pitch angles.*

5.3 The Case with Multiple Features

Let us suppose that the vehicle is observing $N_f > 1$ features, simultaneously. The new system is characterized by the $(7+3N_f)-$ dimensional state $[\hat{r},\ \hat{v},\ q,\ \hat{d}_1,\ ...,\ \hat{d}_{N_f-1}]^T$, whose dynamics are given in (9) and (12) and the observations are given in (13) and (11).

It is immediate to realize that all the camera observations are invariant with respect to the same symmetries found in the case of one single feature (for instance, the camera observations do not change when the initial state $[\hat{r},\ \hat{v},\ q,\ \hat{d}_1,\ ...,\ \hat{d}_{N_f-1}]^T$ is rotated about the vertical axis). Hence, in presence of gravity, the yaw angle is still unobservable. In absence of gravity, also the roll and pitch angles are unobservable. Hence, in presence of gravity, the number of independent modes cannot exceed $7+3N_f-1 = 6+3N_f$. In absence of gravity, this number cannot exceed $7 + 3N_f - 3 = 4 + 3N_f$.

On the basis of the results obtained in the previous subsections, we know that the position of each feature in the local frame provides 3 observable modes. Also, the vehicle speed in the local frame provides 3 observable modes. In addition, an observable mode is the norm of the quaternion. Therefore, in both

the cases with and without gravity, we have $3N_f + 4$ observable modes. In absence of gravity, these are all the observable modes. In presence of gravity, also the roll and pitch angles are observable modes, since they are observable modes with a single feature.

The analytical results derived in this subsection can be summarized with the following property:

Property 4 (Observable Modes with Multiple Features) *Let us consider the system defined by (9), (11), (12) and (13). All the independent observable modes are the components of the imaginary quaternion* $q^*(\hat{d}_i - \hat{r})q,\ i = 0,\ 1,\ ..., N_f - 1$ *(i.e. the position of the observed features in the local frame), the three components of the imaginary quaternion* $q^*\hat{v}q$ *(i.e. the vehicle speed in the local frame) and the product* q^*q *(i.e. the norm of the quaternion). In addition, in presence of gravity, also the roll and pitch angles are observable modes.*

5.4 The Case with Bias

In this subsection we will prove that, even when the camera only observes a single feature, the bias affecting the accelerometers and the gyros are observable. The system we are considering is defined by the state: $[\boldsymbol{r}\ \boldsymbol{v}\ q\ \boldsymbol{b_A}\ \boldsymbol{b_\Omega}]^T$, whose dimension is 16. This state satisfies the dynamics in (14). Finally, this system is characterized by the observations given in (10) and (11).

We know that the state is not observable. Indeed, even without bias, we know that it is not possible to estimate the yaw angle (section 5.2). In other words, also this system is invariant with respect to rotations about the vertical axis. Hence, its observable modes must satisfy the equation in (16), where, now, Λ also depends on the components of $\boldsymbol{b_A}$ and $\boldsymbol{b_\Omega}$. On the other hand, we do not know if the system has additional symmetries in which case the observable modes must satisfy additional partial differential equations, simultaneously. In order to prove that the system has a single symmetry, we must provide 15 independent Lie derivatives. By a direct computation, performed by using the symbolic Matlab computational tool, we were able to find the following 15 independent Lie derivatives: $L^0y_1,\ L^0y_2,\ L^0h_{const},\ L^1_{\boldsymbol{f_0}}y_1,\ L^1_{\boldsymbol{f_0}}y_2,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_0}}y_1,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_1}}y_1,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_4}}y_1, L^2_{\boldsymbol{f_0},\boldsymbol{f_0}}y_2, L^2_{\boldsymbol{f_0},\boldsymbol{f_4}}y_2, L^2_{\boldsymbol{f_0},\boldsymbol{f_5}}y_2, L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_5}}y_1, L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_6}}y_1, L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_2}}y_2, L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_6}}y_2$. As previously mentioned, we know that we cannot have more than 15 independent Lie derivatives (otherwise, the yaw angle would be observable). Note that in the previous computation the expression of the vector fields $\boldsymbol{f_0},\ \boldsymbol{f_1},\ ...,\boldsymbol{f_6}$ is not the one given in appendix A. The right one must be computed starting from the dynamics in (14). The fact that we have 15 independent Lie derivatives means that there are no additional symmetries and, the independent observable modes, are the independent solutions of (16). They are: the 9 solutions provided in 5.2 and the six components of the two vectors $\boldsymbol{b_A}$ and $\boldsymbol{b_\Omega}$ (note that these components are trivial solutions of (16)).

The analytical results derived in this subsection can be summarized with the following property:

Property 5 (Observable Modes in Presence of Bias) *Let us consider the system defined by (14), (10) and (11). All the independent observable modes*

are the same as in the case without bias and the six components of the two bias vectors $\boldsymbol{b_A}$ *and* $\boldsymbol{b_\Omega}$.

5.5 The Case with Bias and Unknown Gravity

The results provided in the previous section are based on the dynamics given in (14). In these equations, the parameter g in $\hat{a}_g$ is assumed to be known. When g is unknown, the state must be extended. Hence, in the case of one single feature, it contains 17 components and its dynamics must be completed with $\dot{g} = 0$. By applying the method described in section 3 it is possible to detect the symmetry:

$$\boldsymbol{w_s^4} = [0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 2(q_t q_y - q_x q_z),$$
$$,\ -2(q_t q_x + q_y q_z),\ -q_t^2 + q_x^2 + q_y^2 - q_z^2,\ 0,\ 0,\ 0,\ 1]^T$$

in addition to $\boldsymbol{w_s^3}$. This new symmetry corresponds to the invariance with respect to the following changes:

$$g \rightarrow g + \epsilon; \qquad \hat{b}_A \rightarrow \hat{b}_A - \epsilon q^* k q$$

being k the quaternion $0 + 0i + 0j + 1k$. In other words, this symmetry expresses the intuitive result that it is not possible to distinguish the component of the bias along the vertical axis from the gravity acceleration.

5.6 Necessary Conditions for Observability

The observability analysis performed so far takes into account all the degrees of freedom allowed by the dynamics in (9). In other words, the observability of the modes previously derived, could require the vehicle to move along all these degrees of freedom. It is interesting to understand what happens when only special trajectories are considered. Mathematically, this can be done by introducing in (9) the constraints characterizing the trajectory we want to consider. Then, it suffices to apply the method described in section 3 to the system characterized by the new dynamics and the same observations (10) and (11).

Let us consider the case when the vehicle moves at constant speed (i.e. when the three components of $\boldsymbol{A_v}$ are zero) . It is easy to understand that the scale factor is not observable . Indeed, the second equation in (9) becomes $\dot{\hat{v}} = 0$. Hence, the new dynamics are invariant with respect to the change $\boldsymbol{r} \rightarrow \lambda \boldsymbol{r}$, $\boldsymbol{v} \rightarrow \lambda \boldsymbol{v}$, being λ a real number. In addition, also the two observations in (10) and (11) are invariant with respect to the same change. Therefore, when the vehicle does not accelerate, the system does not contain the information to perform the estimation of the scale factor. This result also holds in the case of multiple features. Indeed, the same invariance also characterizes the equations in (12) and (13) by also considering $\boldsymbol{d_i} \rightarrow \lambda \boldsymbol{d_i},\ i = 0,\ 1,\ ..., N_f - 1$. Hence we have:

Property 6 (First Necessary Condition for Observability) *In order to estimate the observable modes the vehicle cannot move at constant linear speed.*

A fundamental consequence of the previous property is the following property:

Property 7 (Second Necessary Condition for Observability) *In order to estimate the observable modes the camera must perform at least three observations.*

Proof: Let us suppose to have only two camera observations, one taken at the time 0 and one at the time T. Let us suppose that the vehicle moves during the time interval $[0,\ T]$ with the linear acceleration $\boldsymbol{A_v}(t)$ and the angular speed $\boldsymbol{\Omega}(t)$, $t \in [0,\ T]$. Let us denote with $\boldsymbol{s}(t)$ the position of the vehicle at the time t in the reference frame attached to the vehicle at the time 0. Let us denote with $p(t)$ the quaternion characterizing the vehicle orientation at time t in the same frame. Note that this reference frame is a global frame since does not move. Hence, we can use the equations in (9), obtaining:

$$\begin{cases} \dot{\hat{s}} = \hat{\nu} \\ \dot{\hat{\nu}} = p\hat{A}_v p^* = p\hat{A}p^* + \hat{a}_g \\ \dot{p} = \dfrac{1}{2} p\hat{\Omega} \end{cases} \tag{18}$$

where $\boldsymbol{\nu}(t)$ is the vehicle speed in the same frame. Obviously, the configuration of the vehicle at the time T is completely determined by the values of $\boldsymbol{A_v}(t)$, $\boldsymbol{\Omega}(t)$, $t \in [0,\ T]$ and the initial vehicle speed $\boldsymbol{\nu_0}$. In particular, for the position $\boldsymbol{s}(T)$ we have:

$$\hat{s}(T) = \hat{\nu}_0 T + \int_0^T \int_0^t p(\tau)\hat{A}_v(\tau)p(\tau)^* d\tau dt$$

where $p(\tau)$ is obtained by integrating the third equation in (18) with $p(0) = 1$. From the previous equation we obtain that the motion obtained with the new acceleration ($\boldsymbol{A'_v}(t)$), the new angular speed ($\boldsymbol{\Omega'}(t)$) and the new initial speed ($\boldsymbol{\nu'_0}$) defined as follows:

$$\begin{aligned} \boldsymbol{A'_v}(t) &= [0,\ 0,\ 0]^T \\ \boldsymbol{\Omega'}(t) &= \boldsymbol{\Omega}(t) \\ \hat{\nu}'_0 \quad &= \hat{\nu}_0 + \frac{1}{T}\int_0^T \int_0^t p(\tau)\hat{A}_v(\tau)p(\tau)^* d\tau dt \end{aligned}$$

reproduces exactly the same final vehicle position $\boldsymbol{s}(T)$ and the final orientation $p(T)$. Now let us consider the case where the vehicle moves with linear acceleration $\boldsymbol{A'_v}(t)$ and angular speed $\boldsymbol{\Omega'}(t)$ and the camera continuously performs observations during the interval $[0,\ T]$. Since the linear acceleration is identically zero, on the basis of property 6 we obtain that the modes cannot be observed. Let us consider the original case, i.e. the case where the vehicle moves with linear acceleration $\boldsymbol{A_v}(t)$ and angular speed $\boldsymbol{\Omega}(t)$ and performs only the observations at the time 0 and T. Both these observations are also performed in the previous case since the initial and the final vehicle poses coincide. Therefore, the information acquired in the original case is at most the same as the one acquired in the other case. Hence, the modes cannot be estimated also in the original case ■

6 Closed-Form Solutions to Perform the Estimation of All the Observable Modes

We provide closed form solutions which directly express the observable modes in terms of the sensor measurements collected during a short time interval. For the sake of clarity, we start by providing the closed-form solution in the case without gravity (6.1). Then, we provide the solution in presence of gravity (6.2) and bias (6.3). We also discuss the case of multiple features.

6.1 The case without Gravity

6.1.1 Single Feature

We start by discussing the case of one feature. Property 2 states that the sensor data collected during a given time interval contain the information to estimate the vehicle speed and the position of the feature in the local frame. Hence, we start by expressing the dynamics and the observation in this frame. We have:

$$\begin{bmatrix} \dot{\boldsymbol{F}} = M\boldsymbol{F} - \boldsymbol{V} \\ \dot{\boldsymbol{V}} = M\boldsymbol{V} + \boldsymbol{A} \end{bmatrix} \tag{19}$$

where $\boldsymbol{F}$ is the position of the feature in the local frame and $\boldsymbol{V}$ is the vehicle speed in the same frame. The matrix M depends on the angular speed:

$$M \equiv \begin{bmatrix} 0 & \Omega_z & -\Omega_y \\ -\Omega_z & 0 & \Omega_x \\ \Omega_y & -\Omega_x & 0 \end{bmatrix}$$

The validity of (19) can be checked by a direct substitution, i.e. by using $\hat{F} = -q^* \hat{r} q$, $\hat{V} = q^* \hat{v} q$ and by computing their time derivatives with (9).

In the local frame, the observation in (10) is:

$$h_{cam} = [y_1,\ y_2]^T = \left[\frac{F_x}{F_z},\ \frac{F_y}{F_z}\right]^T \tag{20}$$

Let us consider a given time interval, $[T_0,\ T_0 + T]$. Our goal is to estimate the position of the feature and the vehicle speed in the local frame at T_0, i.e. $\boldsymbol{F_0} \equiv \boldsymbol{F}(T_0)$ and $\boldsymbol{V_0} \equiv \boldsymbol{V}(T_0)$, by only using the data from the camera and the IMU during the interval $[T_0,\ T_0 + T]$. The measurements provided by the IMU are usually delivered at a very high frequency ($\sim 100\ Hz$). This allows us to integrate the equations in (19). This seems to be not useful since we do not know the initial state $[\boldsymbol{F_0},\ \boldsymbol{V_0}]^T$. In fact, our goal is to estimate $[\boldsymbol{F_0},\ \boldsymbol{V_0}]^T$. The basic idea is the following. We numerically integrate the equations in (19) by leaving symbolic the unknown components of the initial state. In other words, we obtain for every time $t > T_0$ the analytical expression of the state $[\boldsymbol{F}(t),\ \boldsymbol{V}(t)]^T$ in terms of its initial value $[\boldsymbol{F_0},\ \boldsymbol{V_0}]^T$.

The following fundamental property holds:

Property 8 *The position of the feature at any time, $\boldsymbol{F}(t)$, linearly depends on the initial feature position, $\boldsymbol{F_0}$, and on the initial vehicle speed, $\boldsymbol{V_0}$. In other words:*

$$\boldsymbol{F}(t) = C_F(t)\boldsymbol{F_0} + C_V(t)\boldsymbol{V_0} + \boldsymbol{C_B}(t) \tag{21}$$

where $C_F(t)$, $C_V(t)$ are 3×3 matrices and $\boldsymbol{C_B}(t)$ is a $3D-$vector. In addition, $C_F(t)$ and $C_V(t)$ only depend on $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$.

Proof: In appendix B we provide the expression of $\boldsymbol{F}(t)$ and $\boldsymbol{V}(t)$ in terms of the the initial values $\boldsymbol{F_0}$ and $\boldsymbol{V_0}$ and the vehicle acceleration $\boldsymbol{A_v}(\tau)$ and angular speed $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$. In particular, the matrices $C_F(t)$ and $C_V(t)$ and the vector $\boldsymbol{C_B}(t)$ are computed ■

We consider the components of $\boldsymbol{F}(t)$, i.e. $F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0})$ $F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0})$ and $F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0})$. By using (20) we obtain:

$$\begin{aligned} F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) &= y_1(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) \\ F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) &= y_2(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) \end{aligned} \tag{22}$$

These are two independent equations in our six unknowns, (which are the components of $\boldsymbol{F_0}$ and $\boldsymbol{V_0}$). On the basis of property 8, the components of $\boldsymbol{F}(t)$ are linear on the unknowns. Hence, the equations in (22) are linear and, by having at least $n_{obs} = 3$ camera observations, we can easily obtain the initial state $[\boldsymbol{F_0},\ \boldsymbol{V_0}]^T$. In appendix C we analyze the case $n_{obs} = 3$ and we prove that the 6 equations are independent (with the exception of special cases whose probability is zero). Hence, in this case, the components of $\boldsymbol{F_0}$ and $\boldsymbol{V_0}$ are obtained by inverting a (6×6) matrix. For larger n_{obs}, it suffices to compute the pseudoinverse of a $(2n_{obs} \times 6)$ matrix.

6.1.2 Multiple Features

Let us consider the case when the camera observes N_f features. Let us denote their position in the local frame with $\boldsymbol{F^i}$, $i = 0,\ 1,\ ..., N_f - 1$. On the basis of property 4, we know that we can estimate the state $[\boldsymbol{F^0},\ \boldsymbol{F^1},\ ..., \boldsymbol{F^{N_f-1}},\ \boldsymbol{V}]$ whose dynamics are given by (19) with the first equation repeated for all the features. The camera observation model is the one in (20), repeated for all the features. Each camera observation consists of $2N_f$ measurements, $y_1^i,\ y_2^i,\ i = 0,\ 1,\ ..., N_f - 1$. By proceeding as in the case of one feature, we obtain a system of linear equations similar to the one in (22). The number of unknowns are now $3N_f + 3$. By considering n_{obs} camera observations, the number of equations are $2n_{obs}N_f$. When $n_{obs} = 2$, we have $4N_f$ equations. For $N_f \geq 3$ the number of equations is larger than the number of unknowns, i.e. $4N_f \geq 3N_f + 3$ when $N_f \geq 3$. On the other hand, on the basis of property 7, we know that these equations are not independent. Hence, the minimum number of observations is 3 for any value of N_f. However, a higher value of N_f will increase the precision of the estimation.

6.2 The case with Gravity

6.2.1 Single Feature

As in the previous subsection, we start by considering the case of a single feature. On the basis of property 3, we know that the sensor data collected during a given time interval, contain the information to estimate the vehicle speed and the position of the feature in the local frame, and, the absolute roll and pitch angles. We express the dynamics and the observation in the local frame. We have:

$$\left[\begin{array}{l} \dot{\boldsymbol{F}} = M\boldsymbol{F} - \boldsymbol{V} \\ \dot{\boldsymbol{V}} = M\boldsymbol{V} + \boldsymbol{A} + \boldsymbol{A_g} \\ \dot{\boldsymbol{q}} = m\boldsymbol{q} \end{array}\right. \tag{23}$$

where $\boldsymbol{q}$ is the four vector whose components are the components of the quaternion q, i.e. $\boldsymbol{q} = [q_t,\ q_x,\ q_y,\ q_z]^T$. The matrix M is provided in 6.1 and the matrix m is:

$$m \equiv \frac{1}{2}\left[\begin{array}{cccc} 0 & -\Omega_x & -\Omega_y & -\Omega_z \\ \Omega_x & 0 & \Omega_z & -\Omega_y \\ \Omega_y & -\Omega_z & 0 & \Omega_x \\ \Omega_z & \Omega_y & -\Omega_x & 0 \end{array}\right]$$

$\boldsymbol{A_g}$ is the gravity acceleration in the local frame, i.e. $\hat{A}_g = q^* \hat{a}_g q$. We remark that, because of the gravity, the first two equations in (23) cannot be separated from the equations describing the dynamics of the quaternion, in contrast to the case without gravity.

Let us consider a given time interval, $[T_0,\ T_0 + T]$. In contrast to the previous case, our goal is now to also estimate the absolute roll and pitch angles at the time T_0. In other words, the goal is the estimation of the state $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ R_0,\ P_0]^T$, by only using the data from the camera and the IMU during the interval $[T_0,\ T_0 + T]$. We proceed as in the previous case. We numerically integrate the equations in (23) by leaving symbolic the unknown components of the initial state. On the other hand, the components of $\boldsymbol{q}(T_0)$ are not observable since the yaw angle is not observable. In order to proceed as in the previous subsection, we need to know how the position of the feature at the time t, i.e. $\boldsymbol{F}(t)$, depends on $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ R_0,\ P_0]^T$. We have the following fundamental property, which extends the property 8 to the case with gravity:

Property 9 *The position of the feature at any time, $\boldsymbol{F}(t)$, linearly depends on the initial feature position, $\boldsymbol{F_0}$, on the initial vehicle speed, $\boldsymbol{V_0}$, and on the three quantities: $\chi_\alpha \equiv 2g(q_{t0}q_{y0} - q_{x0}q_{z0})$, $\chi_\beta \equiv -2g(q_{t0}q_{x0} + q_{y0}q_{z0})$ and $\chi_\gamma \equiv 2g(q_{x0}^2 + q_{y0}^2) - g$. In other words:*

$$\boldsymbol{F}(t) = C_F(t)\boldsymbol{F_0} + C_V(t)\boldsymbol{V_0} + C_\chi(t)\boldsymbol{\chi_g} + \boldsymbol{C_B}(t) \tag{24}$$

where $\boldsymbol{\chi_g} \equiv [\chi_\alpha,\ \chi_\beta,\ \chi_\gamma]^T$ is the gravity vector in the local frame at time T_0, $C_F(t)$ $C_V(t)$, $C_\chi(t)$ are 3×3 matrices and $\boldsymbol{C_B}(t)$ is a $3D-$vector. In addition, $C_F(t)$, $C_V(t)$ and $C_\chi(t)$ only depend on $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$.

Proof: In appendix D we provide the expression of $\boldsymbol{F}(t)$ and $\boldsymbol{V}(t)$ in terms of the the initial values $\boldsymbol{F_0}$, $\boldsymbol{V_0}$, q_0 and the vehicle acceleration $\boldsymbol{A_v}(\tau)$ and angular speed $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$. The dependence on the initial quaternion q_0 is only through the three components of the vector $\boldsymbol{\chi_g}$ and it is linear, as it is the dependence on $\boldsymbol{F_0}$ and $\boldsymbol{V_0}$. In the appendix, the matrices $C_F(t)$, $C_V(t)$, $C_\chi(t)$ and the vector $\boldsymbol{C_B}(t)$ are computed ■

By proceeding as in the case without gravity we obtain the analogous of equations (22). The new equations also depend on the vector $\boldsymbol{\chi_g}$:

$$\begin{aligned} F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) &= y_1(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) \\ F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) &= y_2(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) \end{aligned} \tag{25}$$

i.e., each camera observation occurred at the time $t \in [T_0,\ T_0+T]$ provides two equations in the nine unknowns (which are the components of $\boldsymbol{F_0}$, $\boldsymbol{V_0}$ and $\boldsymbol{\chi_g}$). On the basis of property 9, the components of $\boldsymbol{F}(t)$ are linear on the unknowns. Hence, the equations in (25) are linear and, by having at least $n_{obs} = 5$ camera observations, we can easily obtain the initial state $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}]^T$. In particular, when $n_{obs} \geq 5$, the components of $\boldsymbol{F_0}$, $\boldsymbol{V_0}$ and $\boldsymbol{\chi_g}$ are obtained by computing the pseudoinverse of a $(2n_{obs} \times 9)$ matrix.

6.2.2 Single feature; exploiting additional information

On the basis of property 3, we know that, regarding the robot orientation, only the roll and pitch angles are observable modes. Hence, it must be possible to express the components of the vector $\boldsymbol{\chi_g}$ only in terms of these two angles. In appendix E we provide these expressions. These expressions contain additional information to estimate $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}]^T$. Indeed, the components of $\boldsymbol{\chi_g}$ are three but they only depend on two quantities. An important consequence due to this additional information is that it is possible to estimate $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}]^T$ even when the camera only performs $n_{obs} = 4$ observations. On the other hand, when more than four observations are available ($n_{obs} \geq 5$), the expressions in (40) can be adopted to improve the precision. We discuss the case of $n_{obs} = 4$ observations and we provide a procedure to perform the estimation. When $n_{obs} = 4$, the equations in (25) are eight[3]. Hence, it is not possible to determine the components of $\boldsymbol{F_0}, \boldsymbol{V_0}$ and $\boldsymbol{\chi_g}$ by a simple matrix inversion. On the other hand, the equations in (25) allow us to express eight among the nine unknowns in terms of one of them. Let us suppose to express the components of $\boldsymbol{F_0}$ and $\boldsymbol{V_0}$, and the first two components of $\boldsymbol{\chi_g}$ in terms of χ_γ. We have:

$$\boldsymbol{w} = \boldsymbol{c}\ \chi_\gamma + \boldsymbol{d} \tag{26}$$

where $\boldsymbol{w} = [F_{x0},\ F_{y0},\ F_{z0}, V_{x0},\ V_{y0},\ V_{z0}, \chi_\alpha,\ \chi_\beta]^T$ and $\boldsymbol{c}$ and $\boldsymbol{d}$ are two vectors whose components are obtained by using the eight linear equations provided by (25) where the expression of $\boldsymbol{F}(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g})$ is provided in the appendix D. $\boldsymbol{c}$ and $\boldsymbol{d}$ only depend on the vehicle angular speed and linear acceleration during the interval $[T_0,\ T_0+T]$. We start by considering the last two equations in (26). They are:

[3]By proceeding as in the appendix C, it is possible to prove that these equations are independent, with the exception of special cases whose probability is zero (e.g. when the four vehicle poses and the observed feature lie on the same plane)

$$\chi_\alpha = c_7\chi_\gamma + d_7$$
$$\chi_\beta = c_8\chi_\gamma + d_8$$

where $c_7,\ c_8$ are the 7^{th} and 8^{th} component of $\boldsymbol{c}$ and $d_7,\ d_8$ are the 7^{th} and 8^{th} component of $\boldsymbol{d}$. By using (40) we know that the norm of the vector $\boldsymbol{\chi_g}$ is g. Hence, by using the previous two expressions we have:

$$|\boldsymbol{\chi_g}|^2 = (c_7\chi_\gamma + d_7)^2 + (c_8\chi_\gamma + d_8)^2 + \chi_\gamma^2 = g^2 \tag{27}$$

which is a second degree equation in χ_γ. Therefore, by solving this equation and by using (26), we immediately obtain $\boldsymbol{w}$, and so $\boldsymbol{F_0}$, $\boldsymbol{V_0}$ and $\boldsymbol{\chi_g}$ (actually, two solutions are obtained since (27) is a second degree equation). Fig. 5 displays the steps of the procedure previously described.

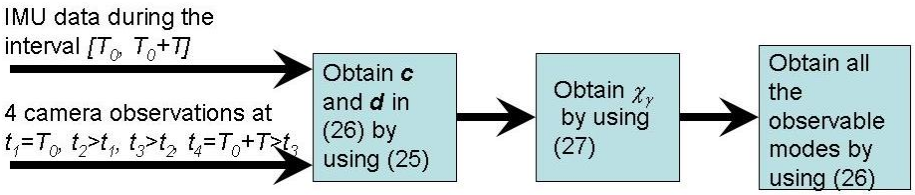


Figure 5: The steps performed to estimate the observable modes when $n_{obs} = 4$ and $N_f = 1$.

In the case we have $n_{obs} \geq 5$, the values of $\boldsymbol{F_0}, \boldsymbol{V_0}$ and $\boldsymbol{\chi_g}$ are obtained by using the $2n_{obs}(\geq 10)$ equations in (25) (it suffices to compute the pseudoinverse of a $(2n_{obs} \times 9)$ matrix). Then, the equations in (40) are used to obtain the roll and pitch angles. We have:

$$P = \arcsin\left(\frac{\chi_\alpha}{g}\right), \quad R = -\arcsin\left(\frac{\chi_\beta}{\sqrt{g^2 - \chi_\alpha^2}}\right) \tag{28}$$

The previous expressions only depend on χ_α and χ_β. In other words, by using them to estimate the roll and pitch angles, the information contained in χ_γ is not exploited. A possible way to exploit this information is to minimize the cost function:

$$c(R,\ P) = (g\sin P - \chi_\alpha)^2 + \tag{29}$$
$$+ (-g\sin R\cos P - \chi_\beta)^2 + (-g\cos R\cos P - \chi_\gamma)^2$$

where the initial values are set by using (28).

6.2.3 Multiple Features

Let us consider the case where the camera observes N_f features. As in the previous section, we denote their position in the local frame with $\boldsymbol{F^i}$, $i = 0,\ 1,\ ..., N_f - 1$. On the basis of property 4 we know that we can estimate the state $[\boldsymbol{F^0},\ \boldsymbol{F^1},\ ...,\boldsymbol{F^{N_f-1}},\ \boldsymbol{V},\ \boldsymbol{\chi_g}]$. Each camera observation consists of $2N_f$ measurements, $y_1^i,\ y_2^i,\ i = 0,\ 1,\ ..., N_f - 1$. By proceeding as in the case of one

feature, we obtain a system of linear equations similar to the one in (25). The number of unknowns are now $3N_f + 6$. On the basis of property 7, we know that at least $n_{obs} = 3$ observations are necessary to perform the estimation. When $n_{obs} = 3$ and $N_f = 2$, the number of equations (twelve) is larger than the number of unknowns. On the other hand, as illustrated in appendix F, in this case the estimation cannot start by the computation of a pseudoinverse but it requires to perform a procedure similar to the one described in section 6.2.2. For $n_{obs} \geq 4$, the estimation can start by the computation of a pseudoinverse.

6.2.4 Multiple features; exploiting additional information

As discussed in the second part of 6.2.2, it is possible to exploit the dependence among the components of $\boldsymbol{\chi_g}$ to improve the precision on the estimation of the roll and pitch angles. The procedure consists of the minimization of the cost function in (29), as for the case of one single feature.

6.3 The Case with Bias

We derive a closed-form solution only when the accelerometers are affected by a bias, i.e. we will consider the case $\boldsymbol{b_A} \neq [0\ 0\ 0]^T$ and $\boldsymbol{b_\Omega} = [0\ 0\ 0]^T$. Indeed, only in this case a simple closed-form solution has been derived. The case of having a bias also on the gyros will be considered in a future work.

By proceeding as in the case without bias we obtain the following property, which extends property 9:

Property 10 *The position of the feature at any time, $\boldsymbol{F}(t)$, linearly depends on the initial feature position, $\boldsymbol{F_0}$, on the initial vehicle speed, $\boldsymbol{V_0}$, on $\boldsymbol{\chi_g}$ and on the bias on the accelerometers $\boldsymbol{b_A}$. In other words:*

$$\boldsymbol{F}(t) = C_F(t)\boldsymbol{F_0} + C_V(t)\boldsymbol{V_0} + C_\chi(t)\boldsymbol{\chi_g} + C_{b_A}(t)\boldsymbol{b_A} + \boldsymbol{C_B}(t) \tag{30}$$

where $\boldsymbol{\chi_g} \equiv [\chi_\alpha,\ \chi_\beta,\ \chi_\gamma]^T$ and $C_F(t)$, $C_V(t)$, $C_\chi(t)$, $C_{b_A}(t)$ are 3×3 matrices and $\boldsymbol{C_B}(t)$ is a $3D-$vector. In addition, $C_F(t)$, $C_V(t)$, $C_\chi(t)$ and $C_{b_A}(t)$ only depend on $\Omega(\tau)$, $\tau \in [T_0,\ t]$.

Proof: At the end of appendix D we provide the expression of $\boldsymbol{F}(t)$ in terms of $\boldsymbol{F_0}, \boldsymbol{V_0}, \boldsymbol{\chi_g}, \boldsymbol{b_A}$ and the vehicle acceleration $\boldsymbol{A_v}(\tau)$ and angular speed $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$. In particular, the matrices $C_F(t)$, $C_V(t)$, $C_\chi(t)$, $C_{b_A}(t)$ and the vector $\boldsymbol{C_B}(t)$ are computed ■

By proceeding as in the case without bias we obtain the analogous of equations (25). The new equations also depend on the vector $\boldsymbol{b_A}$:

$$\begin{aligned} F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{b_A}) &= y_1(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{b_A}) \\ F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{b_A}) &= y_2(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{b_A}) \end{aligned} \tag{31}$$

i.e., each camera observation occurred at the time $t \in [T_0,\ T_0 + T]$ provides two equations in the 12 unknowns (which are the components of $\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}$ and $\boldsymbol{b_A}$). On the basis of property 10, the components of $\boldsymbol{F}(t)$ are linear on the unknowns. Hence, the equations in (31) are linear. On the other hand, they are not independent because of the symmetry discussed in section 5.5. Hence, the determination of the state $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{b_A}]^T$ must be performed by using

a procedure similar to the one discussed in section 6.2.2. This means that the value of the gravity must be a priori known and, as shown in section 6.2.2, two distinct solutions are determined.

7 Performance Evaluation

We evaluate the performance of the proposed strategy by using both synthetic and real data. The advantage of simulations is that the ground truth is perfectly known and this allows us a quantitative evaluation of the proposed strategy. We also investigate the accuracy of the proposed approach in the case where the data from the accelerometers are affected by a bias. This will be considered in a single simulation discussed in 7.1.3. In all the other simulations and in the experiments, we assume unbiased inertial measurements.

7.1 Accuracy of the Algorithm via Monte Carlo Simulations

We simulate many different trajectories in $3D$. For all the simulations we use the proposed strategy to estimate the distance of the N_f observed features ($d_i \equiv |\boldsymbol{d_i} - \boldsymbol{r}| = \left|\boldsymbol{F^i}\right|$, $i = 0,\ 1,\ ...N_f - 1$), the speed of the camera ($v \equiv |\boldsymbol{v}| = \sqrt{v_x^2 + v_y^2 + v_z^2} = \sqrt{V_x^2 + V_y^2 + V_z^2}$) and the roll and the pitch angles ($R \equiv \arctan(2Q_r)$ and $P \equiv \arcsin(2Q_p)$). Specifically, in all the simulations the values of the estimated d_i, v, R, P will be compared with the ground truth values and the difference is provided.

7.1.1 Simulated Trajectories

The trajectories are generated by randomly generating the linear and angular acceleration of the camera at 100 Hz. In particular, at each time step, the three components of the linear acceleration and the angular speed are generated as Gaussian independent variables whose mean values will be denoted respectively with μ_a and μ_ω and whose variances will be denoted respectively with σ_a^2 and σ_ω^2. By performing many simulations we remarked that the precision of the proposed strategy in estimating the roll and pitch angles is almost independent of μ_ω, σ_ω^2 and σ_a^2. On the other hand, the precision on the estimated d_i and v significantly depends on μ_a and also depends on σ_a^2. This is not surprising. Indeed, according to property 6, when the camera moves at constant speed, the scale factor cannot be estimated. Hence, we expect that when μ_a becomes smaller the precision on the estimation of d_i and v becomes worse. We set the previous parameters in order to be close to a real case (as in the experiment discussed in 7.2.2; see also figure 8 *b*): $\sigma_a^2 = \left(1\ \frac{m}{s^2}\right)^2$, $\mu_\omega = 0\frac{deg}{s}$ and $\sigma_\omega^2 = \left(1\ \frac{deg}{s}\right)^2$. Regarding μ_a we considered the following two values $\mu_a = 0\frac{m}{s^2}$ and $\mu_a = 0.3\frac{m}{s^2}$.

The initial vehicle position is at the origin. We adopt many different values for the initial speed. In the simulations here provided it is set equal to: $[0.3,\ 0.3,\ 0.3]\frac{m}{s}$.

7.1.2 Simulated Sensors

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 time step i, we denote them with $\boldsymbol{\Omega_i^{true}}$ and $\boldsymbol{A_{v\ i}^{true}}$). 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: $\boldsymbol{\Omega_i} = N\left(\boldsymbol{\Omega_i^{true}}, P_{\Omega_i}\right)$ and $\boldsymbol{A_i} = N\left(\boldsymbol{A_{v\ i}^{true}} - \boldsymbol{A_{gi}} - \boldsymbol{b_{Ai}}, P_{A_i}\right)$ where:

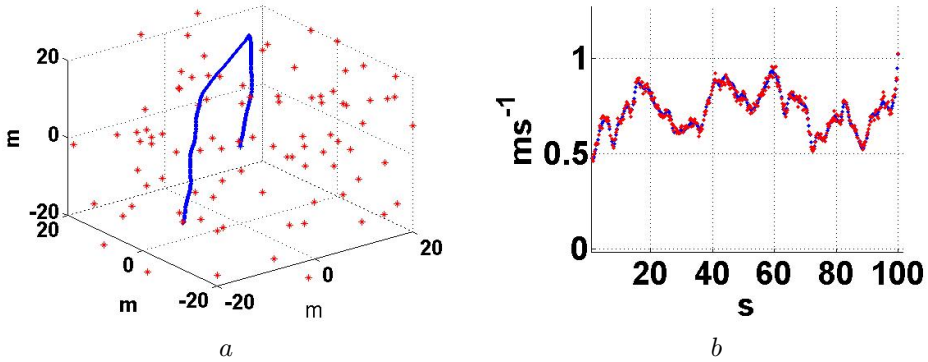
- N indicates the Normal distribution whose first entry is the mean value and the second its covariance matrix;
- P_{Ω_i} and P_{A_i} are the covariance matrices characterizing the accuracy of the IMU;
- $\boldsymbol{A_{gi}}$ is the gravity acceleration in the local frame and $\boldsymbol{b_{Ai}}$ is the bias affecting the data from the accelerometer.

In all the simulations we set both the matrices P_{Ω_i} and P_{A_i} diagonal and in particular: $P_{\Omega_i} = \sigma_{gyro}^2 I_3$ and $P_{A_i} = \sigma_{acc}^2 I_3$, where I_3 is the identity 3×3 matrix. We considered several values for σ_{gyro} and σ_{acc}, in particular: $\sigma_{gyro} \in [0.3,\ 10]\ deg$ and $\sigma_{acc} \in [1,\ 30]\frac{cm}{s^2}$.

Regarding the camera, the provided readings are generated in the following way. By knowing the true trajectory, the true bearing angles of the feature in the camera frame are computed. They are computed each $0.3s$. 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.

7.1.3 Simulation Results

We start by showing the results related to an illustrative case, where the vehicle accomplishes a $3D$ trajectory. In particular, the simulated vehicle moves during $100s$. Figure 6 a displays the vehicle trajectory together with the position of the point features.



a *b*

Figure 6: In a: typical $3D$ motion generated in our simulations; the red stars indicate the point features. In b: the true (blue dots) and the estimated (red dots) vehicle speed.

The camera observes all the features whose distance is smaller than $5m$. In this simulation, the parameters characterizing the error on the IMU are set as follows: $\sigma_{gyro} = 1\ deg$ and $\sigma_{acc} = 3\frac{cm}{s^2}$. The number of observations for every estimation is $n_{obs} = 8$.

Figure 6*b* shows the norm of the vehicle speed. The blue dots are the true values while the red dots are the estimated ones. Figures 7 (left and right) display the roll and pitch angles and figure 8*a* shows the three components of the bias affecting the tri-axial accelerometer. The camera performs a new observation every $0.3s$. Since $n_{obs} = 8$, the length of the time interval necessary to perform a single estimation is $2.4s$. Note that the value of the bias is changing very slowly with time and it can be assumed constant during every estimation process.

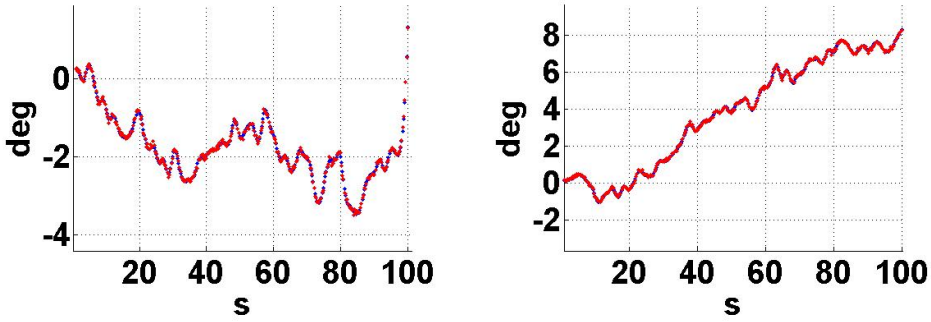
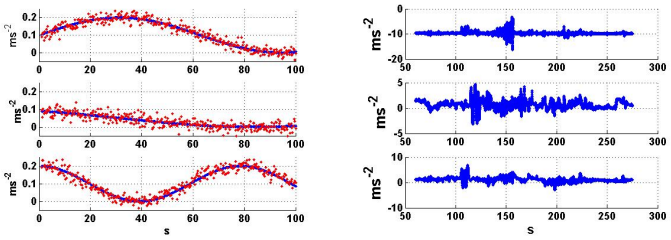


Figure 7: Roll (left) and pitch (right) angles during the simulated experiment. The blue dots are the ground truth and the red dots the estimated values.



a *b*

Figure 8: In *a*: the three bias components of the accelerometers; from the bottom to the top the x, y and z components. In *b*: the three components of the acceleration provided by the tri-axial accelerometer in the real experiments (see section 7.2); from the bottom to the top the x, y and z components.

In order to have more quantitative results we performed many simulations. We considered different scenarios by varying the number of observed features (N_f), the values of σ_{gyro} and σ_{acc}, the number of observations n_{obs} and the parameter μ_a which characterizes the motion. Regarding N_f, we remarked that there is a significant precision improvement by passing from $N_f = 1$ to $N_f = 2$. On the other hand, for larger N_f, the precision improvement is negligible. The

position of the features was randomly generated with a uniform distribution on the box centered on the origin and with size $5m$. Figure 9 summarizes the results of this investigation by displaying the estimation error vs the number of camera observations (n_{obs}). 16 pictures are provided. From the bottom to the top they display the error on the pitch angle, the roll angle, the vehicle speed and the distance of the observed features, respectively. From the left to the right they regard the case of $N_f = 1,\ \mu_a = 0\ ms^{-2}$, $N_f = 1,\ \mu_a = 0.3\ ms^{-2}$, $N_f = 2,\ \mu_a = 0\ ms^{-2}$ and $N_f = 2,\ \mu_a = 0.3\ ms^{-2}$. Every picture displays 4 distinct curves, which correspond to 4 different settings for the values of σ_{gyro} and σ_{acc}. From the bottom to the top, the previous variances increase. In particular, from the bottom to the top of every picture the values are: $\sigma_{gyro} = 0.3\ deg\ \sigma_{acc} = 1\ \frac{cm}{s^2}$, $\sigma_{gyro} = 1\ deg\ \sigma_{acc} = 3\ \frac{cm}{s^2}$, $\sigma_{gyro} = 3\ deg\ \sigma_{acc} = 10\ \frac{cm}{s^2}$ and $\sigma_{gyro} = 10\ deg\ \sigma_{acc} = 30\ \frac{cm}{s^2}$. Each value is computed by running 100 Monte Carlo simulations. Regarding the distance d, the provided error (the four pictures at the top) is averaged on the two features when $N_f = 2$. As stated in section 6.2.3, when $N_f = 2$, three observations allows performing the estimation. This is the reason because the smallest n_{obs} is 3 when $N_f = 2$ (the pictures in the last two columns). Regarding the case of a single feature, as explained in section 6.2.2, the smallest n_{obs} is 4.

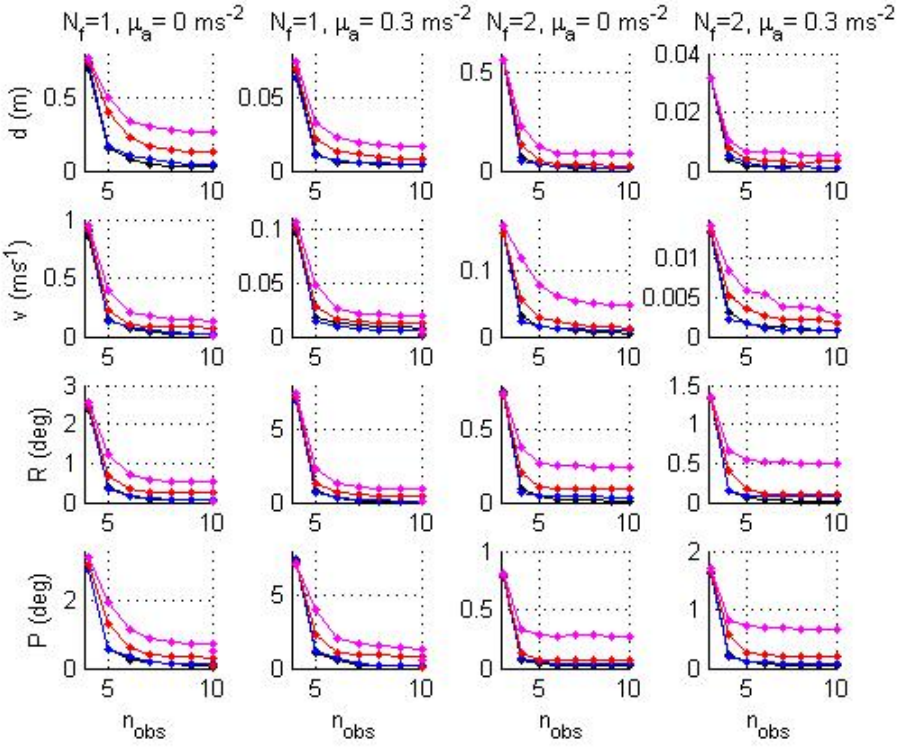


Figure 9: Error on the observable modes obtained by running 100 Monte Carlo simulations. From the bottom to the top the pictures display the error on the pitch angle, the roll angle, the robot speed and the distance of the observed features, respectively. Every picture displays 4 distinct curves which correspond to 4 different settings for the values of σ_{gyro} and σ_{acc}.

7.2 Performance Evaluation with Real Data

We evaluate the performance of the proposed algorithm by using two distinct data sets, the first is in $2D$ and the second in $3D$.

7.2.1 $2D$ Data Set

We adopted the data set provided in [31]. This is an excellent test bed since it also provides a thorough ground truth. The only drawback for our purposes is that our strategy works also in $3D$ while this data set regards experiments carried out in $2D$.

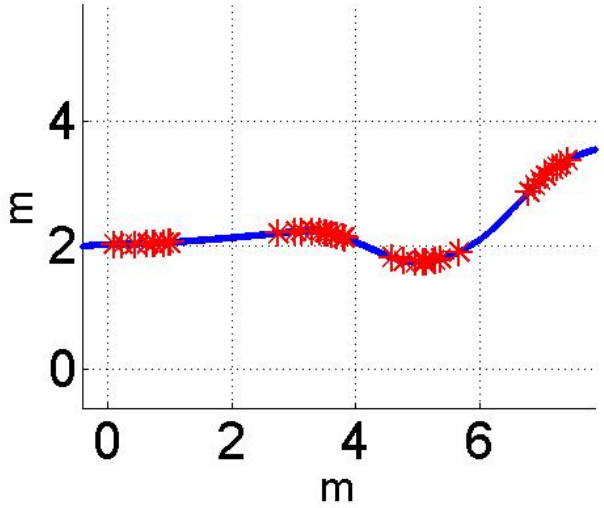


Figure 10: A piece of the $2D$ robot trajectory as recovered by the ground truth data file in the session Bicocca-2009-02-26a of [31]. The four groups of stars represent the points where the same point feature has been extracted.

We adopted the data provided in the session *Bicocca-2009-02-26a*. The robot trajectory in the ground truth data file is provided at around 50 Hz. This allowed us to get a reliable ground truth for the vehicle speed. The data provided by the IMU are also available. These data are delivered at around 130 Hz. Finally, by using the provided vision data files, we were able to extract several point features. In fig 10 we display a piece of the robot trajectory (as provided by the ground truth data file). In particular, all the points in blue represent the robot positions. In the figure four groups of points are also displayed by using red stars. Each group of star marks represent the true robot positions where the same point feature has been extracted from the vision data file. Unfortunately, through the provided ground truth data set, we do not have the actual position of our extracted four point features. For this reason, we cannot evaluate the performance of our strategy in evaluating the distance of these features. On the other hand, by having the true robot speed as previously mentioned, we evaluate the accuracy of the proposed strategy in estimating the

	v_1	v_2	v_3	v_4
True Speed	0.466	0.638	0.585	0.661
Estimated Speed	0.49	0.65	0.57	0.63

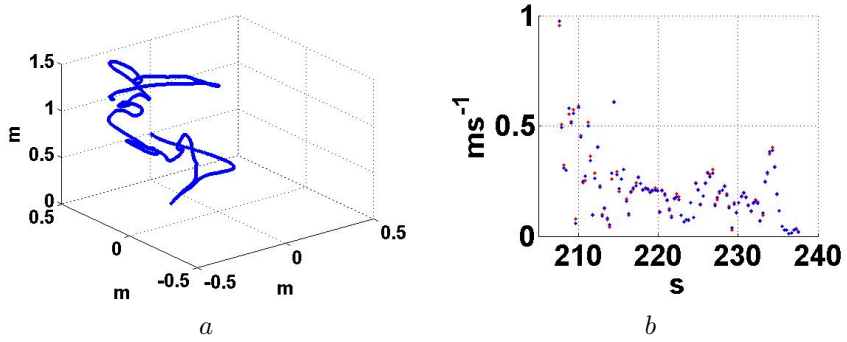
Table 1: True and estimated robot speeds (ms^{-1})

robot speed. In table 1 we report the true and the estimated speeds of the robot for the four considered group of points. The speed for each group is the initial one, i.e. the one that the robot has when it is at the point on the left.

7.2.2 $3D$ Data Set

We evaluated the performance of the proposed algorithm by using a $3D$ data set. These data have been provided by the autonomous system laboratory at ETHZ in Zurich. The data are provided together with a reliable ground-truth, which has been obtained by performing the experiments at the ETH Zurich Flying Machine Arena [20], which is equipped with a Vicon motion capture system. The visual and inertial data are obtained with a monochrome USB-camera gathering 752×480 images at $15Hz$ and a Crossbow VG400CC-200 IMU providing the data at 75 Hz. The camera field of view is 150 deg. The calibration of the camera was obtained by using the omnidirectional camera toolkit by Scaramuzza [32]. Finally, the extrinsic calibration between the camera and the IMU has been obtained by using the strategy introduced in [19]. The experiment here analyzed lasted for about $250s$.

Figure 11 a shows the trajectory (ground truth) during the time interval $[200,\ 240]s$.



a b

Figure 11: In a: the trajectory (ground truth) in the $3D$ real data set during the time interval $[200,\ 240]s$. In b: the vehicle speed in the real $3D$ experiment. Blue dots are the ground truth and red dots the estimated values.

Figures 11 b and 12 show the results regarding the estimated speed, roll and pitch angles, respectively. In all those figures, the blue dots are the ground truth while the red dots are the estimated values.

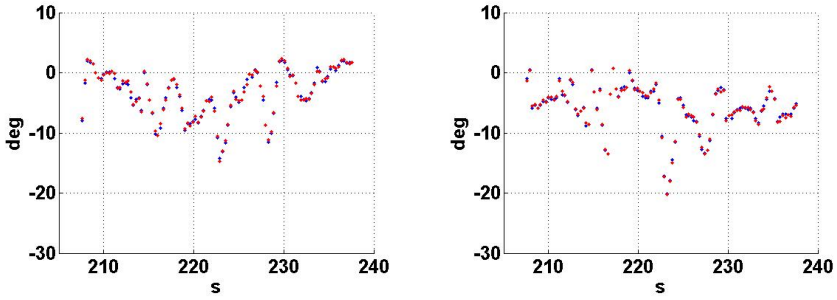


Figure 12: Roll (left) and pitch (right) angles in the real $3D$ experiment. Blue dots are the ground truth and red dots the estimated values.

8 Conclusions

In this paper we investigated the problem of data fusion when the adopted sensors are a monocular camera and inertial sensors (i.e. one tri-axial accelerometer and one tri-axial gyrometer). We provided two main contributions:

1. The analytical derivation of all the observable modes, based on a non standard observability analysis, which fully accounts the system non linearities.

2. The analytical derivation of closed-form solutions which analytically express the observable modes in terms of the sensor measurements collected during a very short time interval.

The first contribution was marginally approached in the past. To the best of our knowledge, no quantitative results, based on analytical computation, have been provided so far. The computation carried out in section 5 allowed us to provide quantitative results in many different contexts, including the case of biased and unbiased inertial measurements, the case of a single and multiple features, and in presence and absence of gravity. In our opinion, there are cases where the provided results are not intuitive. Property 5 states that, by only observing one single feature and by collecting the data from one tri-axial gyrometer and one tri-axial accelerometer, there is all the necessary information to estimate the robot speed in the local frame, the position of the feature in the same frame, the absolute roll and pitch angles and the bias affecting the inertial measurements. This is a non intuitive result.

The second contribution provides closed form solutions which allow us to simultaneously determine all the observable modes without the need of any initialization or a priori knowledge. In particular, only few camera observations are necessary. This is a key advantage since it allows us to quickly recover the observable modes even after a kidnapping. In mobile robotics, and in particular in aerial navigation, this becomes a fundamental advantage.

The performance of the proposed approach has been evaluated via extensive Monte Carlo simulations and by using two distinct real data sets.

Future works will be devoted to extend the proposed estimation approach by also taking into account varying sensor accuracies in order to give preferential

weighting to the more accurate sensor in the results. We also want to analytically investigate the independence of the equations in the closed form solutions in presence of bias. In particular, we want to investigate the cases when the number of observations and features are the minimum required to perform the estimation on the basis of the observability analysis. Currently, this analysis has been done only in the case without bias (in section 6.2.2 and in the appendices C and F).

Acknowledgment

This work was supported by the European Project FP7-ICT-2007-3.2.2 Cognitive Systems, Interaction, and Robotics under the contract #231855 (sFLY). We also acknowledge the Rawseeds project for the data sets provided on line [31]. Finally, we acknowledge the Autonomous System Lab at ETHZ in Zurich for providing us a $3D$ data set which includes a very reliable ground truth.

A Number of Independent Lie Derivatives for the System analyzed in 5.1

The system is characterized by the state: $[\boldsymbol{r}\ \boldsymbol{v}\ \boldsymbol{q}]^T$, whose dimension is 10. The dynamics are given in (9) (without the term $\hat{a}_g$, since we are considering the case $g = 0$) and the observations are given in (10) and (11). In order to compute the Lie derivatives, we need to express the dynamics as in (5). We have $L = 6$ and the six inputs are the three components of the acceleration, $\boldsymbol{A}$, and the three components of the angular speed, $\boldsymbol{\Omega}$. Hence: $u_1 = A_x$, $u_2 = A_y$, $u_3 = A_z$, $u_4 = \Omega_x$, $u_5 = \Omega_y$, $u_6 = \Omega_z$. The seven vector functions $\boldsymbol{f_0},\ \boldsymbol{f_1},\ \ldots,\boldsymbol{f_6}$ are:

$$
\begin{aligned}
\boldsymbol{f_0} &= [v_x,\ v_y,\ v_z,\ \mathbf{0_7}]^T \\
\boldsymbol{f_1} &= [\mathbf{0_3},\ q_t^2 + q_x^2 - q_y^2 - q_z^2,\ 2q_tq_z + 2q_yq_x,\ -2q_tq_y + 2q_zq_x,\ \mathbf{0_4}]^T \\
\boldsymbol{f_2} &= [\mathbf{0_3},\ -2q_tq_z + 2q_yq_x,\ q_t^2 + q_y^2 - q_z^2 - q_x^2,\ 2q_tq_x + 2q_zq_y,\ \mathbf{0_4}]^T \\
\boldsymbol{f_3} &= [\mathbf{0_3},\ 2q_tq_y + 2q_zq_x,\ -2q_tq_x + 2q_zq_y,\ q_t^2 + q_z^2 - q_x^2 - q_y^2,\ \mathbf{0_4}]^T \\
\boldsymbol{f_4} &= [\mathbf{0_6},\ -1/2q_x,\ 1/2q_t,\ 1/2q_z,\ -1/2q_y]^T \\
\boldsymbol{f_5} &= [\mathbf{0_6},\ -1/2q_y,\ -1/2q_z,\ 1/2q_t,\ 1/2q_x]^T \\
\boldsymbol{f_6} &= [\mathbf{0_6},\ -1/2q_z,\ 1/2q_y,\ -1/2q_x,\ 1/2q_t]^T
\end{aligned}
$$

where we denoted with $\mathbf{0_n}$ the vector line whose dimension is n and whose entries are all zeros.

We must compute the Lie derivatives of all the three observations function given in (10) and (11) with respect to all the previous vector fields. By a direct computation, performed by using the symbolic Matlab computational tool, we were able to find the following 7 independent Lie derivatives: L^0y_1, L^0y_2, L^0h_{const}, $L^1_{\boldsymbol{f_0}}y_1$, $L^1_{\boldsymbol{f_0}}y_2$, $L^2_{\boldsymbol{f_0},\boldsymbol{f_0}}y_1$, $L^2_{\boldsymbol{f_0},\boldsymbol{f_1}}y_1$. We know that we cannot have more than 7 independent Lie derivatives (otherwise, we would have less than three symmetries). Hence, the number of independent Lie derivatives is 7.

B Computation of $\boldsymbol{F}(t)$ and $\boldsymbol{V}(t)$ in the case without gravity

We provide the expression of $\boldsymbol{F}(t)$ and $\boldsymbol{V}(t)$ in terms of the the initial values $\boldsymbol{F}(T_0) = \boldsymbol{F_0}$ and $\boldsymbol{V}(T_0) = \boldsymbol{V_0}$ and the vehicle acceleration $\boldsymbol{A_v}(\tau)$ and angular speed $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$.

By discretizing the second equation in (19) and by denoting with j the j^{th} time step (corresponding with t_j), we obtain:

$$\boldsymbol{V_j} = (I_3 + M_j dt_j)\boldsymbol{V_{j-1}} + \boldsymbol{A_j} dt_j$$

where M_j is the matrix M provided in section 6 at the time step j, I_3 is the identity matrix 3×3 and $dt_j = t_j - t_{j-1}$.

The previous expression for $\boldsymbol{V_j}$ provides the following expression in terms of the initial conditions:

$$\boldsymbol{V_j} = \Xi_j \left(\boldsymbol{V_0} + \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{A_k} dt_k \right) \tag{32}$$

where:

$$\Xi_j \equiv \prod_{k=1}^{j} (I_3 + M_k dt_k) \tag{33}$$

which is the rotation matrix between the local frame at time T_0 and the local frame at time t_j. In the same way we finally obtain the expression of $\boldsymbol{F_j}$ in terms of the initial conditions:

$$\boldsymbol{F_j} = \Xi_j \left(\boldsymbol{F_0} - \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{V_k} dt_k \right) = \tag{34}$$

$$= \Xi_j \left(\boldsymbol{F_0} - (t_j - T_0)\boldsymbol{V_0} - \sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} \boldsymbol{A_{k'}} dt_k dt_{k'} \right)$$

Hence, we have:

$$\boldsymbol{F_j} = C_F(t_j)\boldsymbol{F_0} + C_V(t_j)\boldsymbol{V_0} + \boldsymbol{C_B}(t_j)$$

with:

$$C_F(t_j) \equiv \Xi_j, \quad C_V(t_j) \equiv (T_0 - t_j)\Xi_j,$$

$$\boldsymbol{C_B}(t_j) \equiv -\Xi_j \sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} \boldsymbol{A_{k'}} dt_k dt_{k'}$$

Note that both $C_F(t_j)$ and $C_V(t_j)$ only depend on $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t_j]$.

C The closed-form solution in the case $n_{obs} = 3$, $N_f = 1$ and in absence of gravity

We will prove that the equations in (22) are independent in the case of three distinct camera observations ($n_{obs} = 3$) and one feature ($N_f = 1$). As discussed in section 6.1, in this case the number of equations in (22) is six as the number of unknowns (which are the components of the vehicle speed in the local frame and the components of the feature position in the same frame).

In order to prove the independence of the previous equations, we proceed by a direct computation. We have three consecutive camera observations. Let us denote with t_0, t_1 and t_2 the three consecutive times when the three observations occur. Let us refer to the local frame at time t_0. Let us denote by $[a,\ b,\ c]^T$ the position of the feature in this local frame. In addition, we denote by $[x_1,\ y_1,\ z_1, R_1, P_1, Y_1]$ and $[x_2,\ y_2,\ z_2, R_2, P_2, Y_2]$ the two camera poses at the times t_1 and t_2 in this local frame, where, for each pose, the first three components describe the position and the last three the orientation (given by the three angles of roll, pitch and yaw). We derive the analytical expression of the observation in (10) in terms of the previous variables. We obtain: $h_{cam}(t_0) = \left[\frac{a}{c}, \frac{b}{c}\right]^T$,

$$h_{cam}(t_1) = \left[\frac{cY_1cP_1(a - x_1) + (cY_1sP_1sR_1 - sY_1cR_1)(b - y_1) + (cY_1sP_1cR_1 + sY_1sR_1)(c - z_1)}{-sP_1(a - x_1) + cP_1sR_1(b - y_1) + cP_1cR_1(c - z_1)} \right.$$

$$\frac{sY_1cP_1(a-x_1)+(sY_1sP_1sR_1+cY_1cR_1)(b-y_1)+(sY_1sP_1cR_1-cY_1sR_1)(c-z_1)}{-sP_1(a-x_1)+cP_1sR_1(b-y_1)+cP_1cR_1(c-z_1)}\Bigg]^T$$

and the same expression as above for $h_{cam}(t_2)$ (with 2 instead of 1). The c and s symbols are adopted for the cosinus and sinus function (e.g. $cP_1 \equiv \cos(P_1)$). Starting from these expressions, it is possible to obtain the matrix characterizing the system of equations in (22), i.e. the matrix which has to be inverted to obtain the six unknowns. By using the MATLAB symbolic tool it is possible to compute its determinant and investigate the conditions under which it vanishes. We found that it is zero only when all the three poses and the position of the feature lie on the same plane. A part this special situation (which occurs with zero probability) the determinant is different from zero meaning that the equations in (22) are in general independent.

D Computation of $\boldsymbol{F}(t)$ and $\boldsymbol{V}(t)$ in the case with gravity

We provide the expression of $\boldsymbol{F}(t)$ and $\boldsymbol{V}(t)$ in terms of the the initial values $\boldsymbol{F}(T_0) = \boldsymbol{F_0}$, $\boldsymbol{V}(T_0) = \boldsymbol{V_0}$, $q(T_0) = q_0$ and the vehicle acceleration $\boldsymbol{A_v}(\tau)$ and angular speed $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$. As we will see, the dependence on the initial quaternion q_0 is only through the three quantities: $\chi_\alpha \equiv 2g(q_{t0}q_{y0} - q_{x0}q_{z0})$, $\chi_\beta \equiv -2g(q_{t0}q_{x0} + q_{y0}q_{z0})$ and $\chi_\gamma \equiv 2g(q_{x0}^2 + q_{y0}^2) - g$, which are the component of the gravity vector in the local frame at time T_0. In addition, this dependence is linear as it is linear the dependence on $\boldsymbol{F_0}$ and $\boldsymbol{V_0}$.

Before integrating the second equation in (23), as in the appendix B, we consider the new term $\boldsymbol{A_g}$, which depends on the quaternion. In particular, we separate in this term the time-dependent part from the part which is time-independent. Specifically, we introduce the quaternion $p(t)$ such that $q(t) = q_0 p(t)$.

$$\hat{A}_g(t) = q(t)^* \hat{a}_g q(t) = p(t)^* q_0^* \hat{a}_g q_0 p(t)$$

$p(t)$ satisfies the same time differential equation as $q(t)$, i.e. $\dot{p} = \frac{1}{2} p \hat{\Omega}$, but, $p(0) = 1$. Let us denote with $\boldsymbol{\chi_g}$ the $3D$ vector associated with the quaternion $q_0^* \hat{a}_g q_0$, i.e. $\hat{\chi}_g \equiv q_0^* \hat{a}_g q_0$. By a direct computation we obtain:

$$\boldsymbol{\chi_g} = 2g \begin{bmatrix} q_{t0}q_{y0} - q_{x0}q_{z0} \\ -q_{t0}q_{x0} - q_{y0}q_{z0} \\ q_{x0}^2 + q_{y0}^2 - \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \chi_\alpha \\ \chi_\beta \\ \chi_\gamma \end{bmatrix} \tag{35}$$

and:

$$\boldsymbol{A_g}(t) = \Gamma(t)\boldsymbol{\chi_g}, \quad \Gamma(t) \equiv$$

$$\begin{bmatrix} p_t^2 + p_x^2 - p_y^2 - p_z^2 & 2p_tp_z + 2p_xp_y & -2p_tp_y + 2p_zp_x \\ -2p_tp_z + 2p_xp_y & p_t^2 + p_y^2 - p_x^2 - p_z^2 & 2p_tp_x + 2p_zp_y \\ 2p_tp_y + 2p_xp_z & -2p_tp_x + 2p_yp_z & p_t^2 + p_z^2 - p_x^2 - p_y^2 \end{bmatrix}$$

Note that $\Gamma(t)$ only depends on $p(t)$. $p(t)$ is obtained by integrating the equation $\dot{p} = \frac{1}{2} p \hat{\Omega}$, with $p(0) = 1$. Hence, $p(t)$ only depends on the values of the angular speed for $t > T_0$. As a result, the matrix $\Gamma(t)$ only depends on these values.

In particular, $\Gamma(t)$ is independent of the initial state. The matrix $\Gamma(t)$ is the rotation matrix transforming vectors from the local frame at time T_0 into local frame at the time t. We integrate the second equation in (23), obtaining:

$$\boldsymbol{V_j} = (I_3 + M_j dt_j)\boldsymbol{V_{j-1}} + \boldsymbol{B_j} dt_j \tag{36}$$

where $\boldsymbol{B_j} = \boldsymbol{A_j} + \boldsymbol{A_{g\ j}} = \boldsymbol{A_j} + \Gamma_j \boldsymbol{\chi_g}$.

The previous expression for $\boldsymbol{V_j}$ provides the following expression in terms of the initial conditions:

$$\boldsymbol{V_j} = \Xi_j \left[\boldsymbol{V_0} + (t_j - T_0)\boldsymbol{\chi_g} + \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{A_k} dt_k \right] \tag{37}$$

In the same way we finally obtain the expression of $\boldsymbol{F_j}$ in terms of the initial conditions:

$$\boldsymbol{F_j} = \Xi_j \left(\boldsymbol{F_0} - \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{V_k} dt_k \right) = \Xi_j \left[\boldsymbol{F_0} + \right. \tag{38}$$

$$\left. -(t_j - T_0)\boldsymbol{V_0} - \frac{(t_j - T_0)^2}{2}\boldsymbol{\chi_g} - \sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} \boldsymbol{A_{k'}} dt_k dt_{k'} \right]$$

Hence, we have:

$$\boldsymbol{F_j} = C_F(t_j)\boldsymbol{F_0} + C_V(t_j)\boldsymbol{V_0} + C_\chi(t_j)\boldsymbol{\chi_g} + \boldsymbol{C_B}(t_j) \tag{39}$$

with:

$$C_F(t_j) \equiv \Xi_j, \quad C_V(t_j) \equiv (T_0 - t_j)\Xi_j,$$

$$C_\chi(t_j) \equiv -\Xi_j \frac{(t_j - T_0)^2}{2},$$

$$\boldsymbol{C_B}(t_j) \equiv -\Xi_j \sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} \boldsymbol{A_{k'}} dt_k dt_{k'}$$

Note that $C_F(t_j)$, $C_V(t_j)$ and $C_\chi(t_j)$ only depend on $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t_j]$.

In the case where the tri-axial accelerometer is affected by a bias, the derivation of the expression of $\boldsymbol{F_j}$ is very similar to the previous one. The only difference is that in (36), the term $\boldsymbol{B_j}$ also includes the bias $\boldsymbol{b_A}$. In particular we have $\boldsymbol{B_j} = \boldsymbol{A_j} + \Gamma_j \boldsymbol{\chi_g} + \boldsymbol{b_A}$. The expression of $\boldsymbol{F_j}$ differs from the one in (39) since it includes a new term: $\boldsymbol{F_j} = C_F(t_j)\boldsymbol{F_0} + C_V(t_j)\boldsymbol{V_0} + C_\chi(t_j)\boldsymbol{\chi_g} + \boldsymbol{C_B}(t_j) + C_{b_A}(t_j)\boldsymbol{b_A}$, where $C_{b_A}(t_j) \equiv -\Xi_j \left(\sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} dt_{k'} dt_k \right)$.

E Analytical expression of χ_α, χ_β and χ_γ in terms of the roll and pitch angles

Let us consider the unit quaternion: $q_t + q_x i + q_y j + q_z k$. By denoting with R, P and Y respectively the roll, pitch and yaw angles, we have [17]:

$$q_t = \cos\frac{R}{2}\cos\frac{P}{2}\cos\frac{Y}{2} + \sin\frac{R}{2}\sin\frac{P}{2}\sin\frac{Y}{2}$$

$$q_x = \sin\frac{R}{2}\cos\frac{P}{2}\cos\frac{Y}{2} - \cos\frac{R}{2}\sin\frac{P}{2}\sin\frac{Y}{2}$$

$$q_y = \cos\frac{R}{2}\sin\frac{P}{2}\cos\frac{Y}{2} + \sin\frac{R}{2}\cos\frac{P}{2}\sin\frac{Y}{2}$$

$$q_z = \cos\frac{R}{2}\cos\frac{P}{2}\sin\frac{Y}{2} - \sin\frac{R}{2}\sin\frac{P}{2}\cos\frac{Y}{2}$$

We use the previous expressions to obtain $\chi_\alpha = 2g(q_t q_y - q_x q_z)$, $\chi_\beta = -2g(q_t q_x + q_y q_z)$ and $\chi_\gamma = 2g(q_x^2 + q_y^2) - g$ in terms of the roll, pitch and yaw angles. As expected on the basis of property 3, they only depend on the roll and pitch angles. By a direct substitution we obtain:

$$\chi_\alpha = g\sin P,\ \chi_\beta = -g\sin R\cos P,\ \chi_\gamma = -g\cos R\cos P \tag{40}$$

F The closed-form solution in the case $n_{obs} = 3$, $N_f = 2$ and in presence of gravity

As stated in section 6.2.3, in this case the number of equations in (25) is 12 and the number of unknowns is 9. Note that the independent unknowns are actually 8 since three of them are the components of the gravity in the local frame, whose magnitude can be assumed a priori known.

We adopt the same notation as in appendix C. Since we have one additional feature, we denote the position of the features in the local frame at time t_0 by $[a_1,\ b_1,\ c_1]^T$ and $[a_2,\ b_2,\ c_2]^T$, respectively.

By proceeding as in appendix C, we first derive the analytical expressions of the observations in (10) in terms of the previous variables. These expressions are similar to the expressions provided in appendix C. The only difference is that, at each time step, we have an additional observation since we have one additional feature. In other words, we have the same expressions repeated twice once with $[a_1,\ b_1,\ c_1]^T$ instead of $[a,\ b,\ c]^T$ and once with $[a_2,\ b_2,\ c_2]^T$ instead of $[a,\ b,\ c]^T$. Starting from these expressions, it is possible to obtain the matrix characterizing the system of equations in (25). By using the MATLAB symbolic tool, it is possible to compute its rank. We found that it is equal to 11 with the exception of special cases, which occur with zero probability (e.g. when one pose is aligned with both the features or when all the three poses and the two features lie on the same plane). In these special cases the rank is 10.

We conclude that the rank is in general equal to 11. Hence, the unknowns cannot be obtained by a simple computation of a pseudoinverse. On the other hand, since in general the matrix only miss 1 to be full rank, it is possible to obtain the unknowns as for the case $n_{obs} = 4$ and $N_f = 1$, which has been discussed in section 6.2.2. Obviously, also in this case, two distinct solutions can be obtained in general.

References

[1] Ahrens, S.; Levine, D.; Andrews, G.; How, J.P., Vision-based guidance and control of a hovering vehicle in unknown, gps-denied environments, IEEE International Conference on Robotics and Automation (ICRA 2009), Kobe, Japan, May, 2009.

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

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

[4] Blosch, M., Weiss, S., Scaramuzza, D., and Siegwart, R. (2010), Vision Based MAV Navigation in Unknown and Unstructured Environments, IEEE International Conference on Robotics and Automation (ICRA 2010), Anchorage, Alaska, May, 2010.

[5] Bryson, M. and Sukkarieh, S., Building a Robust Implementation of Bearing-only Inertial SLAM for a UAV, Journal of Field Robotics, 2007, 24, 113-143

[6] H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature 293: 133135.

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

[8] Andrew J. Davison, Ian D. Reid, Nicholas D.Molton and Olivier Stasse, "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), pp 1052–1067, 2007

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

[10] J. Folkesson and H. I. Christensen, SIFT Based Graphical SLAM on a Packbot, Field and Service Robotics, 2007, Chamonix, France

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

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

[13] Richard I. Hartley (June 1997). "In Defense of the Eight-Point Algorithm". IEEE Transaction on Pattern Recognition and Machine Intelligence 19 (6): 580593.

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

[15] J. D. Hol, T. B. Schn, and F. Gustafsson, Modeling and Calibration of Inertial and Vision Sensors, The International Journal of Robotics Research, February 2010, vol. 29, 2-3: pp. 231-244.

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

[17] Quaternions and rotation Sequences: a Primer with Applications to Orbits, Aerospace, and Virtual Reality. Kuipers, Jack B., Princeton University Press copyright 1999.

[18] F. John, Partial Differential Equations, Springer-Verlag, 1982.

[19] J. Lobo and J. Dias, Relative pose calibration between visual and inertial sensors, International Journal of Robotics Research, 26(6):2007, 561-575.

[20] S. Lupashin, A. Schollig, M. Sherback and R. D'Andrea, A simple learning strategy for high-speed quadrocopter multi-flips, IEEE International Conference on Robotics and Automation, Anchorage, 2010

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

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

[23] A. Martinelli, Continuous Symmetries and Observability Properties in Autonomous Navigation, Internal Research Report, INRIA, http://hal.inria.fr/inria-00421233_v1/

[24] A. Martinelli, State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration, Accepted for publication on Transaction on Robotics.

[25] A. Martinelli, Closed-Form Solution for Attitude and Speed Determination by Fusing Monocular Vision and Inertial Sensor Measurements, accepted for presentation at ICRA 2011, Shangai

[26] F.M. Mirzaei and S.I. Roumeliotis, "A Kalman Filter-based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation", Transactions on Robotics, 24(5), pp. 1143-1156

[27] A.I. Mourikis and S.I. Roumeliotis, "A Multi-State Constrained Kalman filter for Vision-aided Inertial Navigation", In Proc. 2007 IEEE International Conference on Robotics and Automation (ICRA'07), Rome, Italy, Apr. 10-14, pp. 3565-3572

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

[29] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 26(6):756-770, June 2004

[30] G. Quian and Q. Zheng and R. Chellappa, Reduction of inherent ambiguities in structure from motion problem using inertial data, IEEE International Conference on Image Processing, 2000

[31] The Rawseeds Project http://www.rawseeds.org/home/

[32] D. Scaramuzza, A. Martinelli and R. Siegwart, A toolbox for easy calibrating omnidirectional cameras, IEEE International Conference on Inteligent Robots and Systems, 2006

[33] D. Strelow and S. Singh, Motion estimation from image and inertial measurements, International Journal of Robotics Research, 23(12), 2004

[34] M. Veth, and J. Raquet, Fusing low-cost image and inertial sensors for passive navigation, Journal of the Institute of Navigation, vol. 54(1), 2007

[35] D. Zachariah and Magnus Jansson, Camera-aided inertial navigation using epipolar points, Proceedings of PLANS 2010

Contents

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399