Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination

Agostino Martinelli

Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination

Agostino Martinelli

***Abstract*—This paper investigates the problem of vision and inertial data fusion. A sensor assembling constituted by one monocular camera, three orthogonal accelerometers and three orthogonal gyroscopes is considered. The first paper contribution is the analytical derivation of all the observable modes, i.e. all the physical quantities that can be determined by only using the information in the sensor data acquired during a short time interval. Specifically, the observable modes are the speed and attitude (roll and pitch angles), the absolute scale and the biases affecting the inertial measurements. This holds even in the case when the camera only observes a single point feature. The analytical derivation of the aforementioned 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 aforementioned observable modes in terms of the visual and inertial 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. In addition, in all these contexts, the minimum number of camera images necessary for the observability is derived. The performance of the proposed approach is evaluated via extensive Monte Carlo simulations and real experiments.**

***Index Terms*—Sensor Fusion, Vision-aided Inertial Navigation, Computer Vision, Non linear Observability, Aerial Robotics**

I. 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.

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 these physical quantities starting from the raw sensor data.

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

It is very reasonable to expect that, when fusing vision and inertial measurements, the absolute scale is an observable mode and can be obtained by a closed-form solution. Let us consider the trivial case where a vehicle, 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 vehicle 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 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 vehicle 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 vehicle 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.

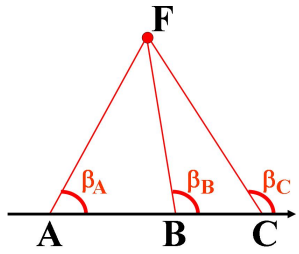

Fig. 1. A vehicle equipped with an accelerometer and a camera moves on a line. The camera performs three observations of the feature in F, repsectively from the points A, B and C.

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-

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 Autonomous System Lab at ETHZ in Zurich for providing us a $3D$ data set which includes a very reliable ground truth.

A. Martinelli is with INRIA Rhone Alpes, Montbonnot, France e-mail: agostino.martinelli@ieee.org

free and unbiased measurements. In this paper we will relax these restrictions by considering the case of a vehicle equipped with IMU[1] and bearing sensors. We want to know which are the observable modes, namely the physical quantities that can be determined without any a priori knowledge (i.e. by only collecting the data from the sensors during a short time interval). For instance, are the absolute scale, the vehicle speed and the vehicle orientation observable modes? Are they observable modes even in the case of biased IMU measurements? Are the biases (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)? And, if yes, what is the minimum number of camera images necessary for this determination?

An answer to the first three questions can be found by applying the method introduced in [21], 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 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 [21], this new concept of continuous symmetry has been adopted to deal with a calibration problem in the framework of wheeled robotics. In [22], this concept has been adopted to deal with the problem of vision and inertial data fusion. Specifically, 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 [21]. 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 III illustrates and summarizes the basic steps of the method introduced in [21], by dealing with a simple $2D$ localization problem. Section IV provides a mathematical description of the system. Starting from this description, in sections V and VI the observability analysis is performed. Then, in section VII, 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 VIII). Finally, conclusions are provided in section IX.

[1]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 gyroscopes.

II. Related Works

The problem of fusing 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 [4]. In [3], 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 [25] the inertial measurements are used in order to reduce the ambiguities in the structure from motion problem. Recent works investigate the observability properties of the vision-aided inertial navigation system [10], [12] and [24]. These works show that the absolute roll and pitch angles of the vehicle are observable modes while the yaw angle is unobservable. This result is consistent with the experimental results obtained in [2] which clearly show how the roll and pitch angles remain more consistent than the heading. In [11], the authors provide a theoretical investigation to analytically derive the motion conditions under which the vehicle state is observable. This analysis also includes the conditions under which the parameters describing the transformation camera-IMU are identifiable. On the other hand, a general theoretical investigation able to also derive the minimum number of camera observations[2] necessary for the state determination still lacks. The results presented in section VI address precisely these limitations. In addition, in section V, the observability analysis is performed in several contexts by also including the case of biased inertial measurements.

The majority of the approaches so far introduced, perform the fusion of vision and inertial sensors by filter-based algorithms. In [1], these sensors are used to perform egomotion estimation. The sensor fusion is obtained by an Extended Kalman Filter (EKF) and by an Unscented Kalman Filter (UKF). The approach proposed in [6] 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 [30] 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 [13] to perform $3D-$SLAM by fusing inertial and vision measurements. It was observed that any inconsistent attitude update severely affects any SLAM solution. The authors proposed to separate attitude update from position and velocity update. Alternatively, they proposed to use additional velocity observations, such as air velocity observation.

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 consid-

[2]Throughout this paper, we will adopt the term *camera observation* to mean the bearing measurements provided by the camera from a single pose, i.e. obtained by a single image.

ered in [18]. In particular, the proposed method is able to estimate the absolute scale 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 [19].

In [24] 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 [31]. 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 [29]. 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 [5] the graphical SLAM approach has been suggested to fuse the data from many different sensors: encoder, inertial, vision and GPS.

To the best of our knowledge, no prior work has addressed the problem of determining the trajectory of a platform in closed form, by only using visual and inertial measurements. Section VII addresses precisely this important problem by providing closed-form expressions of all the observable modes in several different contexts. These solutions have the advantage of not requiring any prior information about the state.

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 [23] the extrinsic calibration has been performed by using an EKF. Non-iterative solutions have been proposed in [9] and [16].

III. 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 vehicle 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 vehicle orientation and the distance of the vehicle 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 [21], 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 [21]. For the sake of clarity, these concepts will be illustrated by referring to a simple localization problem, which is introduced in section III-A.

A. 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. It is also possible to characterize the robot configuration by using the polar coordinates, i.e. $D \equiv \sqrt{x_R^2 + y_R^2}$ and $\phi_R \equiv \arctan 2(y_R, x_R)$. The dynamics are described by the following non-linear differential equations:

$$\begin{bmatrix} \dot{x}_R = v\cos\theta_R \\ \dot{y}_R = v\sin\theta_R \\ \dot{\theta}_R = \omega \end{bmatrix} \quad or \quad \begin{bmatrix} \dot{D} = v\cos(\theta_R - \phi_R) \\ \dot{\phi}_R = \dfrac{v}{D}\sin(\theta_R - \phi_R) \\ \dot{\theta}_R = \omega \end{bmatrix} \tag{1}$$

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. $2a$):

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

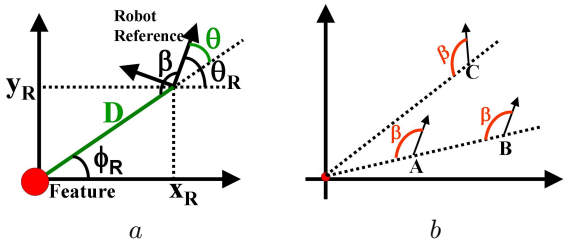

Fig. 2. A simple localization problem. The robot is equipped with odometry and bearing sensors able to evaluate the angle β. In b, the three initial robot configurations are compatible with the same initial observation (β).

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. 2*b* all the three positions A, B and C are compatible with the observation β, provided that the robot 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. 3*a* 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. 3*b*, the two indicated trajectories provide the same bearing observations at any time. Therefore, the dimension of the undistinguishable region is 1 and the dimension of the largest observable subsystem is $3 - 1 = 2$.

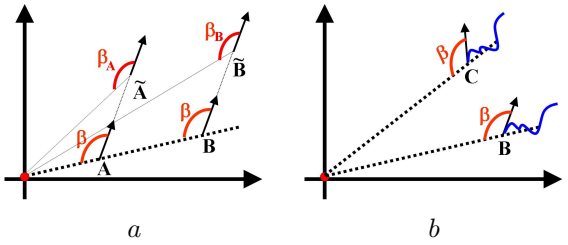

Fig. 3. In a the two initial positions (A and B) do not reproduce the same observations ($\beta_A \neq \beta_B$). In b the two indicated trajectories provide the same bearing observations at any time.

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 aforementioned 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:

$$\begin{cases} \dot{D} = v\cos\theta \\ \dot{\theta} = \omega - \dfrac{v}{D}\sin\theta \end{cases} \qquad y = \beta = \pi - \theta \tag{3}$$

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 two modes (D and θ) and the derivation of the equations in (3) 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 [8] and some basic tools introduced in [21] 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.

B. 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:

$$\begin{cases} \dot{\boldsymbol{S}} = \boldsymbol{f}(\boldsymbol{S}, \boldsymbol{u}) = \boldsymbol{f_0}(\boldsymbol{S}) + \sum_{i=1}^{L} \boldsymbol{f_i}(\boldsymbol{S}) u_i \\ y = h(\boldsymbol{S}) \end{cases} \tag{4}$$

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). The system defined by (1-2) (both in cartesian and in polar coordinates) and the one defined by (3) can be characterized by (4). For instance, for the system in (1) in polar coordinates, 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:

$$L^0\Lambda = \Lambda,\ L^{k+1}_{v_{i_1},...,v_{i_{k+1}}}\Lambda = \nabla_S \left(L^k_{v_{i_1},...,v_{i_k}}\Lambda \right) \ . \ v_{i_{k+1}} \tag{5}$$

where the symbol "." 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. $\boldsymbol{dL^k}_{\boldsymbol{f_{i_1}}, ..., \boldsymbol{f_{i_k}}} h \equiv \nabla_{\boldsymbol{S}} L^k_{\boldsymbol{f_{i_1}}, ..., \boldsymbol{f_{i_k}}} h$), 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 III-A, 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 (1-2) is straightforward. Let us use the polar coordinates. From (2), 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

$$\Gamma \equiv \left\{ w_1^T,\ w_2^T,\ ...,\ w_i^T,\ ... \right\} \tag{6}$$

is equal to two. We conclude that the largest observable subsystem has dimension two as derived in section III-A.

C. Continuous Symmetries

We refer to the input output system given in (4). In [21], 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 (4) 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 III-A. We show that the previous definition corresponds to a global rotation.

For the system defined in (1-2) only one continuous symmetry exists given, in polar coordinates, by the vector $\boldsymbol{w_s} = [0,\ 1,\ 1]^T$ (i.e. belonging to the null space of the matrix Γ in (6)). 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 [7]:

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

In [21] 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:

$$\sum_{i=1}^{n} w_{si}(\boldsymbol{S}) \frac{\partial g}{\partial S_i} = 0 \tag{7}$$

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 (7) to derive the two observable modes for the system discussed in section III-A. As previously mentioned, this system only has the symmetry $[0,\ 1,\ 1]^T$. Hence, the associated equation (7) 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 III-A.

We conclude this section by summarizing the main steps illustrated in this section to detect the observability properties of a given input-output system. The first step consists in the derivation of all the continuous symmetries. This is obtained by computing the analytical expression of the Lie derivatives[3]. Then, according to property 1, a system of partial differential equations is obtained and the observability properties are obtained by solving this system of partial differential equations. Indeed, all the independent observable modes are all the independent solutions of this system.

IV. The Considered System

Let us consider a sensor assembling constituted by a monocular camera and IMU sensors. The IMU consists of three orthogonal accelerometers and three orthogonal gyroscopes. We assume that the transformations among the camera frame and the IMU frames are known (we can assume that the local frame coincides with the camera frame). In the following, we will use the word *vehicle* to refer to this sensor assembling. 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 gravitational acceleration ($\boldsymbol{A_g}$). In particular, we have $\boldsymbol{A} = \boldsymbol{A_v} - \boldsymbol{A_g}$ 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.

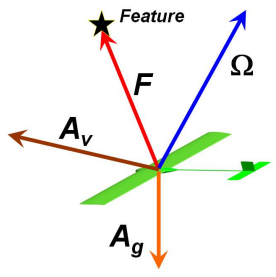

Fig. 4. The feature position ($\boldsymbol{F}$), the vehicle acceleration ($\boldsymbol{A_v}$) the vehicle angular speed ($\boldsymbol{\Omega}$) and the gravitational acceleration ($\boldsymbol{A_g}$).

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 [14].

[3]In section V we will see that sometimes the symmetries can easily be derived from physical considerations, i.e. by remarking the system invariance under several transformations. This allows us to avoid the computation of high order Lie derivative

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:

$$\begin{cases} \dot{\hat{r}} = \hat{v} \\ \dot{\hat{v}} = q\hat{A}_v q^* = q\hat{A}q^* + \hat{a}_g \\ \dot{q} = \frac{1}{2} q\hat{\Omega} \end{cases} \tag{8}$$

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:

$$h_{cam}(\hat{r},\ \hat{v},\ q) = [y_1,\ y_2]^T = \left[\frac{(q^*\hat{r}q)_x}{(q^*\hat{r}q)_z},\ \frac{(q^*\hat{r}q)_y}{(q^*\hat{r}q)_z} \right]^T \tag{9}$$

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):

$$h_{const}(\hat{r},\ \hat{v},\ q) = q^*q \tag{10}$$

A. 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 (8) together with the equations:

$$\dot{\boldsymbol{d}}_{\boldsymbol{i}} = [0\ 0\ 0]^T\ i = 1,\ ...,N_f - 1 \tag{11}$$

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:

$$h^i_{cam} = \left[\frac{(q^*(\hat{d}_i - \hat{r})q)_x}{(q^*(\hat{d}_i - \hat{r})q)_z},\ \frac{(q^*(\hat{d}_i - \hat{r})q)_y}{(q^*(\hat{d}_i - \hat{r})q)_z} \right]^T \quad i = 0, 1, ..., N_f - 1 \tag{12}$$

which coincides with the observation in (9) 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 (8) and (11) and the observations are given in (12) and (10).

B. The Case with Bias

We consider the case when the data provided by the IMU are biased. In other words, we assume that the measurements provided by the three accelerometers and the three gyroscopes are affected by an error which is not zero-mean. Let us denote with $\boldsymbol{A_{bias}}$ and with $\boldsymbol{\Omega_{bias}}$ the two 3D-vectors whose components are the mean values of the measurement errors from the accelerometers and the gyroscopes, respectively. The two vectors $\boldsymbol{A_{bias}}$ and $\boldsymbol{\Omega_{bias}}$ are time-dependent. However, during a short time interval, it is reasonable to consider them to be constant. Under these hypotheses, the dynamics in (8) become:

$$\begin{cases} \dot{\hat{r}} & = \hat{v} \\ \dot{\hat{v}} & = q\hat{A}_v q^* = q\hat{A}q^* + q\hat{A}_{bias}q^* + \hat{a}_g \\ \dot{q} & = \frac{1}{2} q\hat{\Omega} + \frac{1}{2} q\hat{\Omega}_{bias} \\ \dot{\boldsymbol{A}}_{\boldsymbol{bias}} & = \dot{\boldsymbol{\Omega}}_{\boldsymbol{bias}} = [0\ 0\ 0]^T \end{cases} \tag{13}$$

Note that these 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).

V. Observability Properties

We investigate the observability properties of the system whose dynamics are given in (8) and whose observations are given in (9) and (10). For the sake of clarity, we discuss both the case without gravity (V-A) and with gravity (V-B). Moreover, in V-C 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 (8), (10), (11) and (12). Then, the case when the IMU sensors are affected by a bias is investigated (V-D).

The observability analysis performed in this section takes into account all the degrees of freedom allowed by the dynamics in (8). In other words, the observability of the modes here derived, could require the vehicle to move along all these degrees of freedom. The modes derived in this section could become unobservable when the vehicle performs special motions. In section VI we discuss the observability properties for special vehicle motions.

A. The Case without Gravity

Let us set $g = 0$ in (8). By directly computing the Lie derivatives and their gradients, it is possible to detect three independent symmetries for the resulting system. They are:

$$\boldsymbol{w_s^{Rot_x}} = \begin{bmatrix} 0 & -r_z & r_y & 0 & -v_z & v_y & -\frac{q_x}{2} & \frac{q_t}{2} & -\frac{q_z}{2} & \frac{q_y}{2} \end{bmatrix}^T \tag{14}$$

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

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

According to definition 1, these vectors are orthogonal to all the gradients of all the Lie derivatives. 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 [7]:

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

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

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

that is:

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

On the other hand, without computing the Lie derivatives, we could not conclude that the rotational symmetries are *all* the symmetries for the considered system. In order to be sure that they 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.

Summarizing, we detected all the symmetries by proceeding in two separate steps. In the first, we 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 (7)). 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 $10 - 3 = 7$ [15]. 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 (14) but can also be proved by remarking that the absolute scale 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 (10); 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 (8), (9) and (10) 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).*

B. The Case with Gravity

We investigate the observability properties when $g \neq 0$. The presence of the gravity breaks two of the three rotational symmetries. In other words, the system remains invariant only with respect to rotations about the vertical axis. This means that $\boldsymbol{w_s^{Rot_x}}$ and $\boldsymbol{w_s^{Rot_y}}$ 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 $\boldsymbol{w_s^{Rot_z}}$.

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

$$-2r_y \frac{\partial \Lambda}{\partial r_x} + 2r_x \frac{\partial \Lambda}{\partial r_y} - 2v_y \frac{\partial \Lambda}{\partial v_x} + 2v_x \frac{\partial \Lambda}{\partial v_y} + \tag{15}$$
$$-q_z \frac{\partial \Lambda}{\partial q_t} - q_y \frac{\partial \Lambda}{\partial q_x} + q_x \frac{\partial \Lambda}{\partial q_y} + q_t \frac{\partial \Lambda}{\partial q_z} = 0$$

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) [15]. Hence, in this case we have two additional observable modes. They are:

$$Q_r \equiv \frac{q_t q_x + q_y q_z}{1 - 2(q_x^2 + q_y^2)}; \qquad Q_p \equiv q_t q_y - q_z q_x \tag{16}$$

Also for these two solutions it is possible to find a physical meaning. They are related to the roll and pitch angles [14]. 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 (15).

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 (8), (9) and (10). 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.*

C. 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 (8) and (11) and the observations are given in (12) and (10).

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 (8), (10), (11) and (12). 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.*

D. The Case with Bias

In this subsection we will prove that, even when the camera only observes a single feature, the biases affecting the accelerometers and the gyroscopes are observable. The system we are considering is defined by the state: $[\boldsymbol{r}\ \boldsymbol{v}\ q\ \boldsymbol{A_{bias}}\ \boldsymbol{\Omega_{bias}}]^T$, whose dimension is 16. This state satisfies the dynamics in (13). Finally, this system is characterized by the observations given in (9) and (10).

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 V-B). 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 (15), where, now, Λ also depends on the components of $\boldsymbol{A_{bias}}$ and $\boldsymbol{\Omega_{bias}}$. 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: $L^0 y_1,\ L^0 y_2,\ L^0 h_{const},\ L^1_{\boldsymbol{f_0}} y_1,\ L^1_{\boldsymbol{f_0}} y_2,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_0}} y_1,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_1}} y_1,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_4}} y_1,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_0}} y_2,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_4}} y_2,\ L^2_{\boldsymbol{f_0},\boldsymbol{f_5}} y_2,\ L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_5}} y_1,\ L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_6}} y_1,\ L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_2}} y_2,\ L^3_{\boldsymbol{f_0},\boldsymbol{f_0},\boldsymbol{f_6}} y_2$. 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 (13). 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 (15). They are: the 9 solutions provided in V-B and the six components of the two vectors $\boldsymbol{A_{bias}}$ and $\boldsymbol{\Omega_{bias}}$ (note that these components are trivial solutions of (15)).

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 (13), (9) and (10). All the independent observable modes are the same as in the case without bias and the six components of the two bias vectors* $\boldsymbol{A_{bias}}$ *and* $\boldsymbol{\Omega_{bias}}$*.*

E. Unknown Gravity

The results provided in the previous sections are obtained by assuming that the magnitude of the gravitational acceleration (g) is a priori known. In [20] we prove that g is among the observable modes even in the worst case when the inertial sensors are affected by a bias and when only a single feature is available. In other words, the following property holds:

Property 6 (Observability of gravity) *The gravity vector is observable even in the case of biased inertial measurements and when a single feature is available.*

VI. Observability for special trajectories and few camera images

The goal of this section is to discuss the following two issues:

1) Derivation of the observability properties for special vehicle trajectories;
2) Derivation of the minimum number of camera images necessary for the observability of the modes derived in section V.

As we will see, the second issue can be dealt starting from the results obtained by dealing with the first issue.

A. Special Trajectories

We are interested in deriving the observability properties for special trajectories. Mathematically, this can be done by introducing in (8) the constraints characterizing the trajectory we want to consider. Then, it suffices to apply the method described in section III to the system characterized by the new dynamics and the same observations (9) and (10). We only consider two special cases since they allow us to derive important necessary conditions on the minimum number of camera images (see theorem 1). However, there are many other special motions/feature configurations, for which the observability properties degenerate. Some of them will be discussed for the case of two point features in three camera images (section VII-B).

The following property holds:

Property 7 (Observability with constant acceleration) *When the vehicle moves with constant acceleration all the modes derived in section V are observable except the magnitude of the gravitational acceleration.*

Proof: The proof is provided in [20] ■

A special case of constant acceleration is the case of constant speed. In this case we have a nice property when the magnitude of the gravity is a priori known:

Property 8 (Observability with constant speed) *When the magnitude of the gravity is known and the vehicle moves with constant speed all the modes derived in section V are observable up to a scale factor.*

Proof: Our system is characterized by the dynamics given in (8), where the second equation is replaced by $\dot{\hat{v}} = 0$ and with the parameter g a priori known. The system outputs are given in (9) and (10) together with the observations provided by the accelerometers (in this case $\hat{A} = -q^* \hat{a}_g q$). We want to derive the observable modes of this system. According to the method illustrated in section III, we need, first of all, to detect the system symmetries. Instead of computing the Lie derivatives, we remark that, with respect to the case of a general motion (investigated in section V-B), the system is characterized by a further symmetry. Indeed, 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 observations are invariant with respect to the same change[4]. We conclude that, when the vehicle does not accelerate, the system does not contain the information to determine the absolute scale[5]. This result also holds in the case of multiple features. Indeed, the same invariance also characterizes the equations in (11) and (12) by also considering $\boldsymbol{d_i} \rightarrow \lambda \boldsymbol{d_i},\ i = 0,\ 1,\ ..., N_f - 1$ ■

[4]Note that this invariance corresponds to the continuous symmetry: $\boldsymbol{w_s^{scale}} = [r_x,\ r_y,\ r_z, v_x,\ v_y,\ v_z,\ 0,\ 0,\ 0,\ 0]^T$, which would have been obtained by the Lie derivatives and definition 1.

B. Minimum number of camera observations

The observability analysis performed so far, assumes that the observation is provided continuously during a given time interval. However, the following property, allows us to obtain necessary conditions on the number of camera observations.

Property 9 *Let us consider the systems defined in section IV. When the observability of a mode requires the vehicle to move with a non-constant speed, this mode cannot be determined by two camera images. Similarly, when the observability of a mode requires the vehicle to move with a non-constant acceleration, this mode cannot be determined by three camera images.*

Proof: The proof is provided in [20] ■

A consequence of properties 7, 8 and 9 is:

Theorem 1 (Minimum number of camera images) *In order to estimate the observable modes the camera must perform at least three observations (i.e. the observability requires to have at least three images taken from three distinct camera poses). When the magnitude of the gravitational acceleration (g) is unknown, the minimum number of camera images becomes four.*

Proof: The first part of this theorem is a simple consequence of properties 8 and 9. The second part of this theorem is a simple consequence of properties 7 and 9. ■

In most of cases, the magnitude of the gravitational acceleration (g) is known with good accuracy. Hence, considering the case of unknown gravity, could seem useless. On the other hand, considering this case has a very practical importance (see property 12 at the end of the next section).

VII. Closed-Form Solutions to Determine 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 (VII-A). Then, we provide the solution in presence of gravity (VII-B) and bias (VII-C). We also discuss the case of multiple features.

[5]Mathematically, this can be seen by proving that the expression of the scale factor (i.e. $\sqrt{r_x^2 + r_y^2 + r_z^2}$) is not a solution of the partial differential equation associated to $\boldsymbol{w_s^{scale}}$.

A. The case without Gravity

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:

$$\begin{bmatrix} \dot{\boldsymbol{F}} = M\boldsymbol{F} - \boldsymbol{V} \\ \dot{\boldsymbol{V}} = M\boldsymbol{V} + \boldsymbol{A} \end{bmatrix} \tag{17}$$

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 (17) 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 by means of (8).

In the local frame, the observation in (9) is:

$$h_{cam} = [y_1,\ y_2]^T = \left[\frac{F_x}{F_z},\ \frac{F_y}{F_z}\right]^T \tag{18}$$

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 (17). This seems to be useless 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 (17) 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 10 *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:*

$$\boldsymbol{F}(t) = C_F(t)\boldsymbol{F_0} + C_V(t)\boldsymbol{V_0} + \boldsymbol{C_B}(t) \tag{19}$$

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: See appendix B where C_F, C_V and $\boldsymbol{C_B}$ 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 (18) we obtain:

$$\begin{aligned} F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) &= y_1(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) \\ F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) &= y_2(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0}) \end{aligned} \tag{20}$$

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 10, the components of $\boldsymbol{F}(t)$ are linear on the unknowns. Hence, the equations in (20) 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 [20] 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.

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 (17) with the first equation repeated for all the features. The camera observation model is the one in (18), 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 (20). 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 theorem 1, 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.

B. The case with Gravity

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:

$$\begin{bmatrix} \dot{\boldsymbol{F}} = M\boldsymbol{F} - \boldsymbol{V} \\ \dot{\boldsymbol{V}} = M\boldsymbol{V} + \boldsymbol{A} + \boldsymbol{A_g} \\ \dot{\boldsymbol{q}} = m\boldsymbol{q} \end{bmatrix} \tag{21}$$

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 VII-A and the matrix m is:

$$m \equiv \frac{1}{2}\begin{bmatrix} 0 & -\Omega_x & -\Omega_y & -\Omega_z \\ \Omega_x & 0 & \Omega_z & -\Omega_y \\ \Omega_y & -\Omega_z & 0 & \Omega_x \\ \Omega_z & \Omega_y & -\Omega_x & 0 \end{bmatrix}$$

$\boldsymbol{A_g}$ is the gravitational 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 (21) 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 (21) 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 property 10 to the case with gravity:

Property 11 *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:*

$$\boldsymbol{F}(t) = C_F(t)\boldsymbol{F_0} + C_V(t)\boldsymbol{V_0} + C_\chi(t)\boldsymbol{\chi_g} + \boldsymbol{C_B}(t) \quad (22)$$

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: See appendix C where C_F, C_V, C_χ and $\boldsymbol{C_B}$ are computed ■

By proceeding as in the case without gravity we obtain the analogous of equations (20). The new equations also depend on the vector $\boldsymbol{\chi_g}$:

$$\begin{aligned} F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) &= y_1(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) \quad (23)\\ F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) &= y_2(t)\ F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}) \end{aligned}$$

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 11, the components of $\boldsymbol{F}(t)$ are linear on the unknowns. Hence, the equations in (23) 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.

2) Single feature; exploiting additional information: On the basis of property 3, we know that, regarding the vehicle 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 D 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 (37) 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 (23) are eight. 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. However, it is possible to prove that these equations are in general independent [20]. Let us denote by $A\boldsymbol{x} = \boldsymbol{b}$ the linear system in (23) (i.e., the entries of the nine-dimensional column vector $\boldsymbol{x}$ are the components of the vectors $\boldsymbol{F_0}, \boldsymbol{V_0}$ and $\boldsymbol{\chi_g}$). The rank of the matrix A is 8. Let us denote by $\boldsymbol{n}$ the unit vector spanning the null space of A (whose dimension is 1). The linear system $A\boldsymbol{x} = \boldsymbol{b}$ has infinite solution. Each solution satisfies the following equation: $\boldsymbol{x} = A^*\boldsymbol{b} + \gamma\boldsymbol{n}$, being A^* the pseudoinverse of A and γ a scalar number. The determination of γ is obtained by enforcing the constraint that the norm of the vector formed by the last three elements of $\boldsymbol{x}$ is equal to g.

$$|\P\,(A^*\boldsymbol{b} + \gamma\boldsymbol{n})| = g, \quad \P \equiv [0_{3\times 6},\ I_3] \quad (24)$$

where $0_{n\times m}$ is the $n \times m$ matrix whose entries are all zero and I_3 is the identity 3×3 matrix. The equation in (24) is a quadratic polynomial in γ and has two real roots. Hence, we obtain two discrete solutions for $\boldsymbol{x}$.

In the case we have $n_{obs} \geq 5$, the value of $\boldsymbol{x}$ is obtained by using the $2n_{obs}(\geq 10)$ equations in (23) (it suffices to compute the pseudoinverse of A, whose dimension is $(2n_{obs}\times 9)$). Then, the equations in (37) are used to obtain the roll and pitch angles. We have:

$$P = \arcsin\left(\frac{\chi_\alpha}{g}\right), \quad R = -\arcsin\left(\frac{\chi_\beta}{\sqrt{g^2 - \chi_\alpha^2}}\right) \quad (25)$$

The procedure described in this case of $n_{obs} \geq 5$ does not exploit a possible knowledge of the magnitude of the gravitational acceleration. This can be done by minimizing the cost function:

$$c(x) = |A\boldsymbol{x} - \boldsymbol{b}|^2 \quad (26)$$

under the constraint $|\boldsymbol{\chi_g}| = g$. This minimization problem can be solved by using the method of Lagrange multipliers.

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 (23). The number of unknowns are now $3N_f + 6$. We have the following property:

Property 12 *When the number of camera images is less or equal to three ($n_{obs} \leq 3$) the rank of the matrix characterizing the linear system in (23) is always smaller than the number of unknowns, independently of the number of features.*

Proof: According to theorem 1, when $n_{obs} = 3$ the value of g, i.e. the magnitude of the vector $\boldsymbol{\chi_g}$, cannot be determined. Hence, $\boldsymbol{\chi_g}$ cannot be determined by simply solving the linear system in (23). This means that the rank of the matrix characterizing that linear system is always smaller than the number of unknowns ■

Let us consider the case of two points features in three camera images. The unknowns are 12: the position of the two features in the local frame (6 unknowns), the vehicle speed in the local frame (3 unknowns) and the gravity vector in the local frame (3 unknowns). The number of equations in (23) is also 12. On the other hand, because of property 12, the rank of the matrix characterizing the linear system in (23) is less than 12. In [20] we prove that this rank is in general equal to 11 with the exception of the following special cases (when it is less than 11):

1) at least one of the camera pose is aligned with the two other features;
2) all the camera poses and the two features belong to the same plane.

In general, i.e. when the rank is 11, the estimation can be performed by using the value of g which must be a priori known. Enforcing $|\boldsymbol{\chi_g}| = g$ is obtained by solving equation (24), with $\P = [0_{3\times(3N_f+3)},\ I_3] = [0_{3\times 6},\ I_3]$. Hence, as in the case of a single feature in four images, two distinct solutions are obtained.

Property 12 states that when $n_{obs} = 3$, the determination of the observable modes cannot be obtained by computing a pseudoinverse also when the number of features is larger than two. On the other hand, it is possible to show that, with the exception of special cases, the observable modes can be determined by enforcing $|\boldsymbol{\chi_g}| = g$. Hence, when $n_{obs} = 3$ and $N_f \geq 2$, two distinct solutions are in general obtained. When $n_{obs} \geq 4$, the determination of the observable modes can be performed by the computation of a pseudoinverse, provided that the number of equations is at least as the number of unknowns and that the vehicle poses and the positions of the features do not satisfy special conditions, whose probability is zero (for instance when all the features and all the camera poses lie on the same plane).

4) Multiple features; exploiting additional information: As discussed in the second part of VII-B2, it is possible to exploit an a priori knowledge of the magnitude of the gravity to improve the precision. The procedure consists of the minimization of the cost function in (26), as for the case of one single feature.

C. 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 $\boldsymbol{A_{bias}} \neq [0\ 0\ 0]^T$ and $\boldsymbol{\Omega_{bias}} = [0\ 0\ 0]^T$. Indeed, all the matrices appearing in (22) depend on $\boldsymbol{\Omega}(\tau)$ and therefore on $\boldsymbol{\Omega_{bias}}$. Hence, when $\boldsymbol{\Omega_{bias}}$ is unknown, the dependence of $\boldsymbol{F(t)}$ on all the unknowns ($\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g}$ and $\boldsymbol{\Omega_{bias}}$) becomes non linear making more complex their derivation. In contrast, when the bias on the accelerometers is unknown, we obtain the following property, which extends property 11:

Property 13 *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 $\boldsymbol{A_{bias}}$. In other words:*

$$\boldsymbol{F}(t) = \tag{27}$$
$$= C_F(t)\boldsymbol{F_0} + C_V(t)\boldsymbol{V_0} + C_\chi(t)\boldsymbol{\chi_g} + C_{A_{bias}}(t)\boldsymbol{A_{bias}} + \boldsymbol{C_B}(t)$$

where $\boldsymbol{\chi_g} \equiv [\chi_\alpha,\ \chi_\beta,\ \chi_\gamma]^T$ and $C_F(t)$, $C_V(t)$, $C_\chi(t)$, $C_{A_{bias}}(t)$ 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 $C_{A_{bias}}(t)$ only depend on $\Omega(\tau)$, $\tau \in [T_0,\ t]$.

Proof: See the last paragraph of appendix C ■

By proceeding as in the case without bias we obtain the analogous of equations (23). The new equations also depend on the vector $\boldsymbol{A_{bias}}$:

$$\begin{aligned} F_x(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{A_{bias}}) &= y_1(t)\, F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{A_{bias}}) \\ F_y(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{A_{bias}}) &= y_2(t)\, F_z(t;\ \boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{A_{bias}}) \end{aligned} \tag{28}$$

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 $\boldsymbol{A_{bias}}$). On the basis of property 13, the components of $\boldsymbol{F}(t)$ are linear on the unknowns. Hence, the equations in (28) are linear and they allow us to determine the state $[\boldsymbol{F_0},\ \boldsymbol{V_0},\ \boldsymbol{\chi_g},\ \boldsymbol{A_{bias}}]^T$.

VIII. 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 VIII-A3. In all the other simulations and in the experiments, we assume unbiased inertial measurements.

A. 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.

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 observed 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 8, when the camera moves at constant speed, the absolute scale 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 parameters in order to be close to a real case (as in the experiment discussed in VIII-B; see also figure 7 b): $\sigma_a = 1\ ms^{-2}$, $\mu_\omega = 0\ deg\ s^{-1}$ and $\sigma_\omega = 1\ deg\ s^{-1}$. Regarding μ_a we considered the following two values $\mu_a = 0\ ms^{-2}$ and $\mu_a = 0.3\ ms^{-2}$. 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: $[0.3,\ 0.3,\ 0.3]ms^{-1}$.

2) Simulated Sensors: Starting from the performed 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 $\boldsymbol{A_i} = N\left(\boldsymbol{A_{v\ i}^{true}} - \boldsymbol{A_{gi}} - \boldsymbol{A_{bias\ i}}, P_{A_i}\right)$ 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 gravitational acceleration in the local frame and $\boldsymbol{A_{bias\ i}}$ 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: $\sigma_{gyro} \in [0.3,\ 10]\ deg\ s^{-1}$ and $\sigma_{acc} \in [0.01,\ 0.3]ms^{-2}$.

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.

3) Simulation Results: We start by showing the results related to an illustrative case, where the vehicle performs a $3D$ trajectory. In particular, the simulated vehicle moves during $100s$. Figure 5 a displays the vehicle trajectory together with the position of the point features.

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: $\sigma_{gyro} = 1\ deg\ s^{-1}$ and $\sigma_{acc} = 0.03ms^{-2}$. The number of observations for every estimation is $n_{obs} = 8$.

Figure 5b shows the norm of the vehicle speed. The blue dots are the true values while the red disks are the estimated ones. Figures 6 (left and right) display the roll and pitch angles and figure 7a 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.

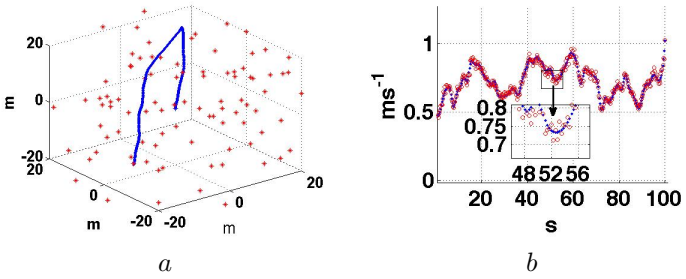

Fig. 5. In a: typical $3D$ motion generated in our simulations; the red stars indicate the point features. In b: the true (blue dots) and the estimated (red disks) vehicle speed.

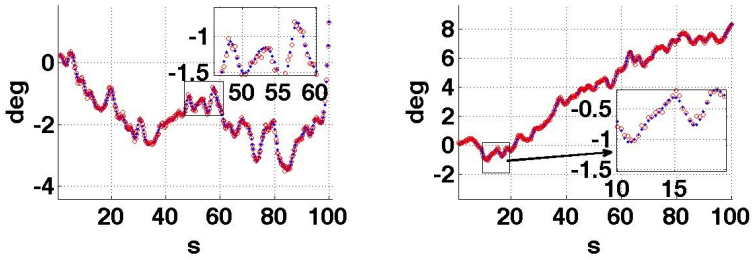

Fig. 6. Roll (left) and pitch (right) angles during the simulated experiment. The blue dots are the ground truth and the red disks the estimated values.

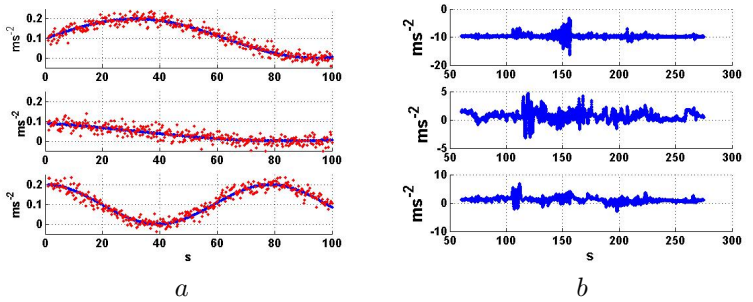

Fig. 7. In a: the three bias components of the accelerometers; from the bottom to the top the x, y and z components. In b: the three components of the acceleration provided by the tri-axial accelerometer in the real experiments (see section VIII-B); from the bottom to the top the x, y and z components.

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 performed simulations with $1 \leq N_f \leq 10$. We found that there is a significant precision improvement by passing from $N_f = 1$ to $N_f = 2$ while, for larger N_f, the precision improvement is negligible. For this reason, in this section only the results for $N_f = 1$ and $N_f = 2$ are provided. The position of the features was randomly generated with a uniform distribution on the box centered on the origin and with size $5m$. Figure 8 summarizes

the results of this investigation by displaying the estimation error vs the number of camera observations (n_{obs}). 16 subplots 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 subplot displays 4 distinct curves, which correspond to 4 different settings of the sensor noise (i.e. the values of σ_{gyro} and σ_{acc}). From the bottom to the top, the sensor noise increase. In particular, from the bottom to the top of every subplot the values are: $\sigma_{gyro} = 0.3\ deg\ s^{-1}\ \sigma_{acc} = 0.01\ ms^{-2}$, $\sigma_{gyro} = 1\ deg\ s^{-1}\ \sigma_{acc} = 0.03\ ms^{-2}$, $\sigma_{gyro} = 3\ deg\ s^{-1}\ \sigma_{acc} = 0.1\ ms^{-2}$ and $\sigma_{gyro} = 10\ deg\ s^{-1}\ \sigma_{acc} = 0.3\ ms^{-2}$. 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 VII-B3, when $N_f = 2$, three observations allow performing the estimation. This is the reason why the smallest n_{obs} is 3 when $N_f = 2$ (the subplots in the last two columns). Regarding the case of a single feature, as explained in section VII-B2, the smallest n_{obs} is 4.

Figure 8 clearly shows that, when the vehicle motion is characterized by a low acceleration ($\mu_a = 0ms^{-2}$, first and third column) the precision on the vehicle speed and on the absolute scale is worse than for the case of higher acceleration ($\mu_a = 0.3ms^{-2}$, second and fourth column). On the other hand, for the roll and the pitch angles, the precision increases by decreasing the acceleration.

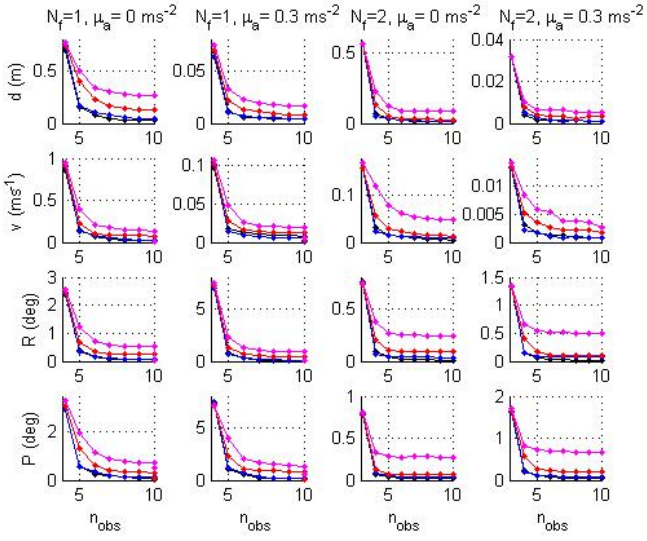

Fig. 8. Error on the observable modes averaged on 100 simulations.

Finally, we performed Monte Carlo simulations in order to investigate the statistical properties of the noise resulting from the estimation procedure. These simulations clearly show that the proposed procedure is not bias-affected and that the noise is well approximated by a Gaussian distribution. For the sake of brevity, only the case of the roll angle is shown. Figure 9 displays the error distribution together with the best Gaussian fit (solid line). This plot is obtained by counting for each bin of $0.2deg$ the number of simulations which provide an error on the roll angle falling in the considered bin. Then, the plotted points are normalized by enforcing the area to be 1. The number of simulations is 10^4. In every simulation, the procedure uses four consecutive camera images and two point features. The variances characterizing the sensors are $\sigma_{gyro} = 10\ deg\ s^{-1}\ \sigma_{acc} = 0.3\ ms^{-2}$, i.e. they are set as in the worst case considered in the simulations shown in fig. 8. Similar results have been obtained for the other estimated quantities and by using other noise sensor settings.

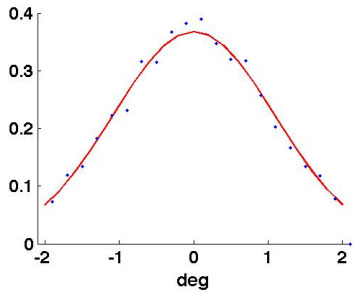

Fig. 9. Distribution of the noise on the roll angle (blue line) and its best Gaussian fit (red line).

B. 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 $3D$. For the sake of brevity, we show the results obtained with the $3D$ data set. The results obtained with the $2D$ data set can be found in [20].

The 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 [17], 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 [26]. Finally, the extrinsic calibration between the camera and the IMU has been obtained by using the strategy introduced in [16]. The experiment here analyzed lasted for about $250s$.

Figure 10 a shows the trajectory (ground truth) during the time interval $[200,\ 240]s$.

Figures 10 b and 11 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 disks are the estimated values.

IX. Conclusion

In this paper we investigated the problem of vision and inertial data fusion. Specifically, we considered a sensor assembling constituted by one monocular camera, three orthogonal

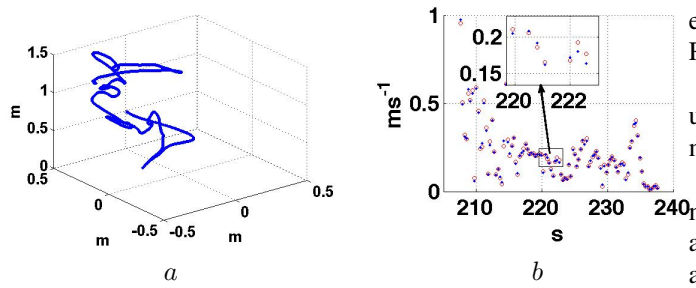

Fig. 10. In a: the trajectory (ground truth) in the $3D$ real data set during the time interval $[200,\ 240]s$. In b: the vehicle speed in the real $3D$ experiment. Blue dots are the ground truth and red disks the estimated values.

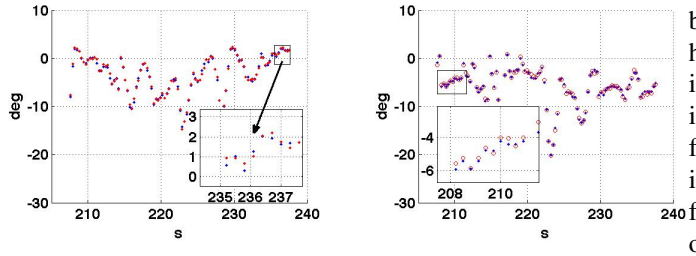

Fig. 11. Roll (left) and pitch (right) angles in the real $3D$ experiment. Blue dots are the ground truth and red disks the estimated values.

accelerometers and three orthogonal gyroscopes. We provided two main contributions:

1) The analytical derivation of all the observable modes, i.e. all the physical quantities that can be determined by only using the information in the sensor data acquired during a short time interval;
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 has been discussed in section V and VI. These sections 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, there is all the necessary information to determine the speed in the local frame, the position of the feature in the same frame, the absolute roll and pitch angles and the biases affecting the inertial measurements. This is a non intuitive result. In addition, the minimum number of camera images necessary to perform the state determination has been provided.

The second contribution provides closed form expressions 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. Another important aspect of these closed form expressions is that they can even work by only using a single feature. This allows us to design very efficient and robust computation methods, such as 1-point RANSAC [27], [28], to prune false matches and outliers.

The performance of the proposed approach has been evaluated via extensive Monte Carlo simulations and real experiments.

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. Additionally, the approach could be extended to incorporate the benefit of a possible previous knowledge on the state. To this regard, we remark that he proposed procedure is not optimal. The mentioned key advantage that it is able to determine the observable modes by only using the sensor data provided during a short interval, has the drawback that it does not exploit a possible previous information on these modes. 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 in the case without bias (section VII-B2).

Appendix A
Number of Independent Lie Derivatives for the System analyzed in V-A

The system is characterized by the state: $[\boldsymbol{r}\ \boldsymbol{v}\ \boldsymbol{q}]^T$, whose dimension is 10. The dynamics are given in (8) (without the term $\hat{a}_g$, since we are considering the case $g = 0$) and the observations are given in (9) and (10). In order to compute the Lie derivatives, we need to express the dynamics as in (4). 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:

$$\boldsymbol{f_0} = [v_x,\ v_y,\ v_z,\ \mathbf{0_7}]^T$$

$$\boldsymbol{f_1} = [\mathbf{0_3},\ q_t^2+q_x^2-q_y^2-q_z^2,\ 2q_tq_z+2q_yq_x,\ -2q_tq_y+2q_zq_x,\ \mathbf{0_4}]^T$$

$$\boldsymbol{f_2} = [\mathbf{0_3},\ -2q_tq_z+2q_yq_x,\ q_t^2+q_y^2-q_z^2-q_x^2,\ 2q_tq_x+2q_zq_y,\ \mathbf{0_4}]^T$$

$$\boldsymbol{f_3} = [\mathbf{0_3},\ 2q_tq_y+2q_zq_x,\ -2q_tq_x+2q_zq_y,\ q_t^2+q_z^2-q_x^2-q_y^2,\ \mathbf{0_4}]^T$$

$$\boldsymbol{f_4} = [\mathbf{0_6},\ -1/2q_x,\ 1/2q_t,\ 1/2q_z,\ -1/2q_y]^T$$

$$\boldsymbol{f_5} = [\mathbf{0_6},\ -1/2q_y,\ -1/2q_z,\ 1/2q_t,\ 1/2q_x]^T$$

$$\boldsymbol{f_6} = [\mathbf{0_6},\ -1/2q_z,\ 1/2q_y,\ -1/2q_x,\ 1/2q_t]^T$$

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 (9) and (10) with respect to all the 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.

APPENDIX 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 acceleration $\boldsymbol{A}(\tau)$ and angular speed $\boldsymbol{\Omega}(\tau)$, $\tau \in [T_0,\ t]$.

By discretizing the second equation in (17) 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 VII 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:

$$\boldsymbol{V_j} = \Xi_j \left(\boldsymbol{V_0} + \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{A_k} dt_k \right) \tag{29}$$

where:

$$\Xi_j \equiv \prod_{k=1}^{j} (I_3 + M_k dt_k) \tag{30}$$

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:

$$\boldsymbol{F_j} = \Xi_j \left(\boldsymbol{F_0} - \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{V_k} dt_k \right) = \tag{31}$$

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

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: $C_F(t_j) \equiv \Xi_j$, $C_V(t_j) \equiv (T_0 - t_j)\Xi_j$, $\boldsymbol{C_B}(t_j) \equiv -\Xi_j \sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} \boldsymbol{A_{k'}} dt_k dt_{k'}$.

APPENDIX C
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 acceleration $\boldsymbol{A}(\tau)$ 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 (21), 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:

$$\boldsymbol{\chi_g} = 2g \begin{bmatrix} q_{t0}q_{y0} - q_{x0}q_{z0} \\ -q_{t0}q_{x0} - q_{y0}q_{z0} \\ q_{x0}^2 + q_{y0}^2 - \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \chi_\alpha \\ \chi_\beta \\ \chi_\gamma \end{bmatrix} \tag{32}$$

and $\boldsymbol{A_g}(t) = \Xi(t)\boldsymbol{\chi_g}$, where $\Xi(t)$ is given in (30). We integrate the second equation in (21), obtaining:

$$\boldsymbol{V_j} = (I_3 + M_j dt_j)\boldsymbol{V_{j-1}} + \boldsymbol{B_j} dt_j \tag{33}$$

where $\boldsymbol{B_j} = \boldsymbol{A_j} + \boldsymbol{A_g}_j = \boldsymbol{A_j} + \Xi_j \boldsymbol{\chi_g}$.

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

$$\boldsymbol{V_j} = \Xi_j \left[\boldsymbol{V_0} + (t_j - T_0)\boldsymbol{\chi_g} + \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{A_k} dt_k \right] \tag{34}$$

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

$$\boldsymbol{F_j} = \Xi_j \left(\boldsymbol{F_0} - \sum_{k=1}^{j} \Xi_k^{-1} \boldsymbol{V_k} dt_k \right) = \Xi_j \left[\boldsymbol{F_0} + \right. \tag{35}$$

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

Hence, we have:

$$\boldsymbol{F_j} = C_F(t_j)\boldsymbol{F_0} + C_V(t_j)\boldsymbol{V_0} + C_\chi(t_j)\boldsymbol{\chi_g} + \boldsymbol{C_B}(t_j) \tag{36}$$

with: $C_F(t_j) \equiv \Xi_j$, $C_V(t_j) \equiv (T_0 - t_j)\Xi_j$, $C_\chi(t_j) \equiv -\Xi_j \frac{(t_j - T_0)^2}{2}$, $\boldsymbol{C_B}(t_j) \equiv -\Xi_j \sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} \boldsymbol{A_{k'}} dt_k dt_{k'}$ and the matrix Ξ_j, given in (30), is computed by only using the gyro's measurements in the time-interval $[T_0,\ t_j]$. 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. The only difference is that in (33), the term $\boldsymbol{B_j}$ also includes the bias $\boldsymbol{A_{bias}}$. In particular we have $\boldsymbol{B_j} = \boldsymbol{A_j} + \Xi_j \boldsymbol{\chi_g} + \boldsymbol{A_{bias}}$. The expression of $\boldsymbol{F_j}$ differs from the one in (36) since it includes a new term: $\boldsymbol{F_j} = C_F(t_j)\boldsymbol{F_0} + C_V(t_j)\boldsymbol{V_0} + C_\chi(t_j)\boldsymbol{\chi_g} + \boldsymbol{C_B}(t_j) + C_{A_{bias}}(t_j)\boldsymbol{A_{bias}}$, where $C_{A_{bias}}(t_j) \equiv -\Xi_j \left(\sum_{k=1}^{j} \sum_{k'=1}^{k} \Xi_{k'}^{-1} dt_{k'} dt_k \right)$.

APPENDIX D
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 [14]:

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

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

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

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

We use these 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:

$$\chi_\alpha = g\sin P,\ \ \chi_\beta = -g\sin R\cos P,\ \ \chi_\gamma = -g\cos R\cos P \tag{37}$$

REFERENCES

[1] 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

[2] 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

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

[4] 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.

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

[6] 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

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

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

[9] 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.

[10] E. Jones, A. Vedaldi, and S. Soatto, "Inertial Structure From Motion with Autocalibration," in Proc. IEEE Int'l Conf. Computer Vision Workshop on Dynamical Vision, Rio de Janeiro, Brazil, Oct. 2007.

[11] E. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach", The International Journal of Robotics Research, published on-line: January 17, 2011

[12] J. Kelly and G. Sukhatme, Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration, Int. Journal of Robotics Research, Oct. 2010

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

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

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

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

[17] 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

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

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

[20] A. Martinelli, Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination by Fusing Vision and Inertial Measurements, Internal Research Report, 2011, INRIA, http://hal.inria.fr/inria-00569083/en/

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

[22] 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

[23] 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, 2008.

[24] 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

[25] 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

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

[27] D. Scaramuzza, F. Fraundorfer F. and R. Siegwart, Real-Time Monocular Visual Odometry for On-Road Vehicles with 1-Point RANSAC, IEEE International Conference on Robotics and Automation (ICRA'09), Kobe, Japan

[28] D. Scaramuzza, 1-Point-RANSAC Structure from Motion for Vehicle-Mounted Cameras by Exploiting Non-Holonomic Constraints, International Journal of Computer Vision, 2011

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

[30] 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

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

Agostino Martinelli (1971) received the M.Sc. degree in theoretical physics from the University of Rome *Tor Vergata*, Rome, Italy, in 1994 and the Ph.D. degree in astrophysics from the University of Rome *La Sapienza*, Rome, in 1999. While working toward the Ph.D. degree, he spent one year at the University of Wales, Cardiff, U.K., and one year with the *Scuola Internazionale Superiore di Studi Avanzati* (SISSA), Trieste, Italy. His research focused on chemical and dynamical evolution in elliptical galaxies, in quasars, and in the intergalactic medium. He also introduced models based on general relativity to explain the anisotropies of cosmic background radiation. After receiving the Ph.D. degree, his interests moved to the problem of autonomous navigation. He spent two years at the University of Rome *Tor Vergata*, and in 2002, he moved to the *Autonomous Systems Laboratory, Ecole Polytechnique Federale de Lausanne*, Lausanne, Switzerland, as a Senior Researcher, where he lead several projects on multisensor fusion for robot localization, simultaneous localization and odometry error learning, and simultaneous localization and mapping. Since September 2006, he has been a Researcher with the *Institut National de Recherche en Informatique et en Automatique* (INRIA) Rhone Alpes, Grenoble, France. He is the author of more than 50 journal and conference papers.