



# Vision and IMU Data Fusion: Closed-Form Determination of the Absolute Scale, Speed and Attitude

Agostino Martinelli, Roland Siegwart

► **To cite this version:**

Agostino Martinelli, Roland Siegwart. Vision and IMU Data Fusion: Closed-Form Determination of the Absolute Scale, Speed and Attitude. Azim Eskandarian. Handbook of Intelligent Vehicles, Springer, 2012. <hal-00779460>

**HAL Id: hal-00779460**

**<https://hal.inria.fr/hal-00779460>**

Submitted on 22 Jan 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Vision and IMU Data Fusion: Closed-Form Determination of the Absolute Scale, Speed and Attitude

Agostino Martinelli\* and Roland Siegwart\*\*

\*INRIA Rhone Alpes, Grenoble, France

\*\*ETHZ, Zurich, Switzerland

## Abstract

This chapter describes an algorithm for determining the speed and the attitude of a sensor assembling constituted by a monocular camera and inertial sensors (three orthogonal accelerometers and three orthogonal gyroscopes). The system moves in a 3D unknown environment. The algorithm inputs are the visual and inertial measurements during a very short time interval. The outputs are: the speed and attitude, the absolute scale and the bias affecting the inertial measurements. The determination of these outputs is obtained by a simple closed form solution which analytically expresses the previous physical quantities in terms of the sensor measurements. This closed form determination allows performing the overall estimation in a very short time interval and without the need of any initialization or prior knowledge. This is a key advantage since allows eliminating the drift on the absolute scale and on the orientation. The performance of the proposed algorithm is evaluated with real experiments.

## 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. In addition, these sensors have very interesting complementarities and together provide rich information to build a system capable of vision-aided inertial navigation and mapping and a great effort has been done very recently in this direction (e.g. [1], [3]). A special issue of the *International Journal of Robotics Research* has recently been devoted to the integration of vision and inertial sensors

[6]. In [5], 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. The majority of the approaches so far introduced, perform the fusion of vision and inertial sensors by filter-based algorithms. In [2], these sensors are used to perform egomotion estimation. The sensor fusion is obtained with an Extended Kalman Filter (*EKF*) and with an Unscented Kalman Filter (*UKF*). The approach proposed in [7] 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 [18] 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 airborne SLAM, an *EKF* has been adopted in [9] to perform 3D-SLAM by fusing inertial and vision measurements. It was remarked 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. Regarding the robot attitude, in [4] it has been noted that roll and pitch angles remain more consistent than the heading.

A fundamental issue to address when fusing vision and inertial measurements, is to understand which are the *observable modes*, i.e. the physical quantities that the information contained in the sensor data allows uniquely determining. The next issue to address is to find a reliable and efficient method to determine all the previous physical quantities.

The following simple 1 –  $D$  example clearly shows that it is reasonable to expect that the absolute scale is an observable mode and can be obtained by a closed-form solution. 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 robot speed during the subsequent time steps and then the distances  $A - B$  and  $B - C$  by integrating the speed. The lengths  $A - F$  and  $B - F$  are obtained by a simple triangulation by using the two angles  $\beta_A$  and  $\beta_B$  from the bearing sensor. When the initial speed  $v_A$  is unknown, all the previous segment lengths can be obtained in terms of  $v_A$ . In other words, it is possible to obtain the analytical expression of  $A - F$  and  $B - F$  in terms of the

unknown  $v_A$  and all the sensor measurements performed while the robot navigates from  $A$  to  $B$ . By repeating the same computation with the bearing measurements in  $A$  and  $C$ , it is obtained 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 the value of  $v_A$  is obtained. Hence, the value of all the segment lengths in fig 1 is obtained in terms of the measurements performed by the accelerometer and the bearing sensor.

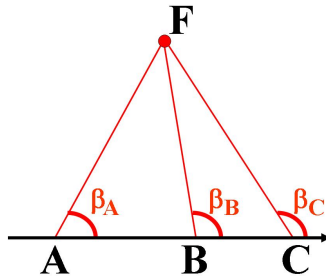


Fig. 1. A vehicle equipped with an accelerometer and a camera moves on a line. The camera performs three observations of the feature in  $F$ , respectively from the points  $A$ ,  $B$  and  $C$ .

The previous example is very simple because of several unrealistic restrictions. First of all, the motion is constrained on a line. Additionally, the accelerometer provides gravity-free and unbiased measurements.

In [15] these restrictions were relaxed. A vehicle equipped with IMU and bearing sensors was considered. The motion of the vehicle was not constrained. However, only the case of one single feature was considered. In addition, the inertial measurements were unbiased.

This chapter extends the results obtained in [15] by also considering the case of multiple features. Additionally, also the case when the accelerometers provide biased measurements will be considered.

The chapter is organized as follows. Section II provides a mathematical description of the system. Sections III and IV provide conditions for the state observability. Then, section V provides the analytical derivation of the closed-form solution to determine the speed and attitude. Section VI highlights the main steps of the proposed algorithm. The performance of the algorithm is evaluated in section VII. Conclusions are provided in section VIII.

## II. THE CONSIDERED SYSTEM

The system is a sensor assembling constituted by a monocular camera and *IMU* sensors. The IMU consists of three orthogonal accelerometers and three orthogonal gyroscopes. The transformations among the camera frame and the IMU frames are known (the local frame is the camera frame). In the following, the word *vehicle* will be used 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 ( $\mathbf{A}$ ) is not simply the vehicle acceleration ( $\mathbf{A}_v$ ). It also contains the gravitational acceleration ( $\mathbf{A}_g$ ). In particular,  $\mathbf{A} = \mathbf{A}_v - \mathbf{A}_g$  since, when the camera does not accelerate (i.e.  $\mathbf{A}_v$  is zero) the accelerometer perceives an acceleration which is the same of an object accelerated upward in the absence of gravity.

In the following, upper-case letters will indicate the vectors when expressed in the local frame and lower-case letters when they are expressed in the global frame. Hence, regarding the gravity:  $\mathbf{a}_g = [0, 0, -g]^T$ , being  $g \simeq 9.8 \text{ ms}^{-2}$ .

The camera is observing a point feature during a given time interval. The global frame will be attached to this feature. The vehicle and the feature are displayed in fig 2.

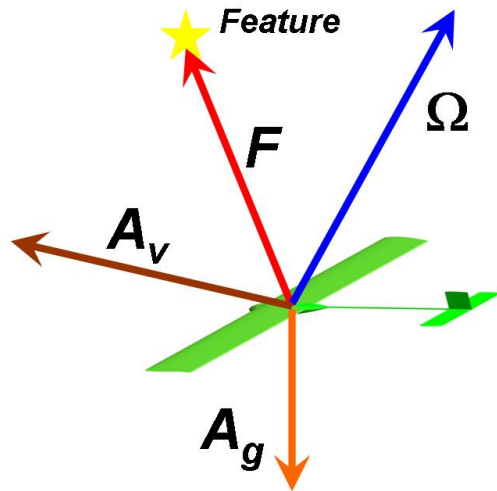


Fig. 2. The feature position ( $\mathbf{F}$ ), the vehicle acceleration ( $\mathbf{A}_v$ ) the vehicle angular speed ( $\mathbf{\Omega}$ ) and the gravitational acceleration ( $\mathbf{A}_g$ ).

Finally, a quaternion will be adopted 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 [10].

The system is characterized by the state  $[\mathbf{r}, \mathbf{v}, q]^T$  where  $\mathbf{r} = [r_x, r_y, r_z]^T$  is the 3D vehicle position,  $\mathbf{v}$  is its time derivative, i.e. the vehicle speed in the global frame ( $\mathbf{v} \equiv \frac{d\mathbf{r}}{dt}$ ),  $q = q_t + iq_x + jq_y + kq_z$  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  $\mathbf{w} = [w_x, w_y, w_z]^T$ , the imaginary quaternion  $\hat{w} \equiv 0 + iw_x + jw_y + kw_z$  will be associated with it. The dynamics of the state  $[\hat{r}, \hat{v}, q]^T$  are:

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

being  $q^*$  the conjugate of  $q$ ,  $q^* = q_t - iq_x - jq_y - kq_z$ . The camera observations can be expressed in terms of the same state  $([\hat{r}, \hat{v}, q]^T)$ . The camera provides the direction of the feature in the local frame. In other words, it provides the unit vector  $\frac{\mathbf{F}}{|\mathbf{F}|}$  (see fig. 2). Hence, the camera provides the two ratios  $y_1 = \frac{F_x}{F_z}$  and  $y_2 = \frac{F_y}{F_z}$ , being  $\mathbf{F} = [F_x, F_y, F_z]^T$ . 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  $-\mathbf{r}$ . Therefore,  $\mathbf{F}$  is obtained by the quaternion product  $\hat{F} = -q^*\hat{r}q$ . The observation function provided by the camera is:

$$h_{cam}(\hat{r}, \hat{v}, q) = [y_1, y_2]^T = \left[ \frac{(q^*\hat{r}q)_x}{(q^*\hat{r}q)_z}, \frac{(q^*\hat{r}q)_y}{(q^*\hat{r}q)_z} \right]^T \quad (2)$$

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

$$h_{const}(\hat{r}, \hat{v}, q) = q^*q \quad (3)$$

### A. The Case with Multiple Features

In the case when the camera observes  $N_f$  features, simultaneously, the global frame will be attached to one of the features.  $\mathbf{d}_i$  denotes the 3D vector which contains the cartesian coordinates

of the  $i^{th}$  feature ( $i = 0, 1, \dots, N_f - 1$ ). The global frame is attached to the  $0^{th}$  feature, i.e.  $\mathbf{d}_0 = [0 \ 0 \ 0]^T$ . The new system is characterized by the state  $[\hat{r}, \hat{v}, q, \hat{d}_1, \dots, \hat{d}_{N_f-1}]^T$ , whose dimension is  $7 + 3N_f$ . The dynamics of this state are given by (1) together with the equations:

$$\dot{\mathbf{d}}_i = [0 \ 0 \ 0]^T \quad i = 1, \dots, N_f - 1 \quad (4)$$

The position  $\mathbf{F}^i$  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:

$$h_{cam}^i = \left[ \begin{array}{c} (q^*(\hat{d}_i - \hat{r})q)_x \\ (q^*(\hat{d}_i - \hat{r})q)_y \\ (q^*(\hat{d}_i - \hat{r})q)_z \end{array} \right]^T, \quad i = 0, 1, \dots, N_f - 1 \quad (5)$$

which coincides with the observation in (2) when  $i = 0$ . Summarizing, the case of  $N_f$  features is described by the state  $[\hat{r}, \hat{v}, q, \hat{d}_1, \dots, \hat{d}_{N_f-1}]^T$ , whose dynamics are given in (1) and (4) and the observations are given in (5) and (3).

### B. The Case with Bias

$\mathbf{A}_{bias}$  denotes and  $\mathbf{\Omega}_{bias}$  denote the two 3D-vectors whose components are the mean values of the measurement errors from the accelerometers and the gyroscopes, respectively. The two vectors  $\mathbf{A}_{bias}$  and  $\mathbf{\Omega}_{bias}$  are time-dependent. However, during a short time interval, it is reasonable to consider them to be constant. Under these hypotheses, the dynamics in (1) become:

$$\left\{ \begin{array}{l} \dot{\hat{r}} = \hat{v} \\ \dot{\hat{v}} = q\hat{A}_vq^* = q\hat{A}q^* + q\hat{A}_{bias}q^* + \hat{a}_g \\ \dot{q} = \frac{1}{2}q\hat{\Omega} + \frac{1}{2}q\hat{\Omega}_{bias} \\ \dot{\mathbf{A}}_{bias} = \dot{\mathbf{\Omega}}_{bias} = [0 \ 0 \ 0]^T \end{array} \right. \quad (6)$$

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

### III. OBSERVABILITY PROPERTIES

We investigate the observability properties of the system whose dynamics are given in (1) and whose observations are given in (2) and (3). For the sake of clarity, we discuss both the case without gravity (III-A) and with gravity (III-B).

#### A. The Case without Gravity

Let us set  $g = 0$  in (1). By directly computing the Lie derivatives and their gradients, it is possible to detect three independent symmetries for the resulting system (see [14]). They are:

$$\mathbf{w}_s^1 = \left[ 0 \quad -r_z \quad r_y \quad 0 \quad -v_z \quad v_y \quad -\frac{q_x}{2} \quad \frac{q_t}{2} \quad -\frac{q_z}{2} \quad \frac{q_y}{2} \right]^T \quad (7)$$

$$\mathbf{w}_s^2 = \left[ r_z \quad 0 \quad -r_x \quad v_z \quad 0 \quad -v_x \quad -\frac{q_y}{2} \quad \frac{q_z}{2} \quad \frac{q_t}{2} \quad -\frac{q_x}{2} \right]^T$$

$$\mathbf{w}_s^3 = \left[ -r_y \quad r_x \quad 0 \quad -v_y \quad v_x \quad 0 \quad -\frac{q_z}{2} \quad -\frac{q_y}{2} \quad \frac{q_x}{2} \quad \frac{q_t}{2} \right]^T$$

According to definition of continuous symmetry introduced in [14], 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  $\epsilon$  about the vertical axis changes the state as follows [8]:

$$\begin{aligned} \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} &\rightarrow \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} + \epsilon \begin{bmatrix} -r_y \\ r_x \\ 0 \end{bmatrix} \\ \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} &\rightarrow \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} + \epsilon \begin{bmatrix} -v_y \\ v_x \\ 0 \end{bmatrix} \\ \begin{bmatrix} q_t \\ q_x \\ q_y \\ q_z \end{bmatrix} &\rightarrow \begin{bmatrix} q_t \\ q_x \\ q_y \\ q_z \end{bmatrix} + \frac{\epsilon}{2} \begin{bmatrix} -q_z \\ -q_y \\ q_x \\ q_t \end{bmatrix} \end{aligned}$$



that is:

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{v} \\ \mathbf{q} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{r} \\ \mathbf{v} \\ \mathbf{q} \end{bmatrix} + \epsilon \mathbf{w}_s^3$$

On the other hand, without computing the Lie derivatives, we could not conclude that the previous ones are *all* the symmetries for the considered system.

In [14] we proved that for every symmetry there is an associated partial differential equation and 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  $10 - 3 = 7$  [11]. 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.  $|\mathbf{r}|$ , is a solution of the three equations (this can be checked by substitution for the partial differential equations associated with the symmetries in (7) but can also be proved by remarking that the scale factor is invariant under rotations). This means that the distance of the feature is observable and it is one among the 7 independent solutions. On the other hand, since the camera provides the position of the feature in the local frame up to a scale factor, having the distance means that the feature position in the local frame is also observable. Therefore, the three components of the feature position in the local frame are three independent solutions. By using quaternions, we can say that three independent solutions are provided by the components of the imaginary quaternion  $q^* \hat{r} q$ . Additionally, since the three partial differential equations are invariant under the transformation  $\mathbf{r} \leftrightarrow \mathbf{v}$ , 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 (3); it can be in any case verified that it satisfies the three partial differential equations).

### B. The Case with Gravity

We investigate the observability properties when  $g \neq 0$ . The presence of the gravity breaks two of the previous three symmetries. In other words, the system remains invariant only with respect to rotations about the vertical axis. This means that  $\mathbf{w}_s^1$  and  $\mathbf{w}_s^2$  are no longer symmetries for the

new system. By directly computing the Lie derivatives, we were able to find nine independent Lie derivatives. Hence, the system has  $10 - 9 = 1$  symmetry which is  $\mathbf{w}_s^3$ .

The partial differential equation associated with  $\mathbf{w}_s^3$  is:

$$\begin{aligned} -2r_y \frac{\partial \Lambda}{\partial r_x} + 2r_x \frac{\partial \Lambda}{\partial r_y} - 2v_y \frac{\partial \Lambda}{\partial v_x} + 2v_x \frac{\partial \Lambda}{\partial v_y} + \\ -q_z \frac{\partial \Lambda}{\partial q_t} - q_y \frac{\partial \Lambda}{\partial q_x} + q_x \frac{\partial \Lambda}{\partial q_y} + q_t \frac{\partial \Lambda}{\partial q_z} = 0 \end{aligned} \quad (8)$$

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

$$Q_r \equiv \frac{q_t q_x + q_y q_z}{1 - 2(q_x^2 + q_y^2)}; \quad Q_p \equiv q_t q_y - q_z q_x \quad (9)$$

Also for these two solutions it is possible to find a physical meaning. They are related to the roll and pitch angles [10]. 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 (8).

### 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  $[\hat{r}, \hat{v}, q, \hat{d}_1, \dots, \hat{d}_{N_f-1}]^T$ , whose dynamics are given in (1) and (4) and the observations are given in (5) and (3).

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  $[\hat{r}, \hat{v}, q, \hat{d}_1, \dots, \hat{d}_{N_f-1}]^T$  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 the previous subsections can be summarized with the following property:

**Property 1** *Let us consider the system defined by (1), (3), (4) and (5). All the independent observable modes are the components of the imaginary quaternion  $q^*(\hat{d}_i - \hat{r})q$ ,  $i = 0, 1, \dots, N_f - 1$  (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 bias affecting the accelerometers and the gyros are observable. The system we are considering is defined by the state:  $[\mathbf{r} \ \mathbf{v} \ q \ \mathbf{b}_A \ \mathbf{b}_\Omega]^T$ , whose dimension is 16. This state satisfies the dynamics in (6). Finally, this system is characterized by the observations given in (2) and (3).

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 III-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 (8), where, now,  $\Lambda$  also depends on the components of  $\mathbf{b}_A$  and  $\mathbf{b}_\Omega$ . On the other hand, we do not know if the system has additional symmetries in which case the observable modes must satisfy additional partial differential equations, simultaneously. In order to prove that the system has a single symmetry, we must provide 15 independent Lie derivatives. By a direct computation, performed by using the symbolic Matlab computational tool, we were able to find the following

15 independent Lie derivatives:  $L^0 y_1, L^0 y_2, L^0 h_{const}, L^1_{\mathbf{f}_0} y_1, L^1_{\mathbf{f}_0} y_2, L^2_{\mathbf{f}_0, \mathbf{f}_0} y_1, L^2_{\mathbf{f}_0, \mathbf{f}_1} y_1, L^2_{\mathbf{f}_0, \mathbf{f}_4} y_1, L^2_{\mathbf{f}_0, \mathbf{f}_0} y_2, L^2_{\mathbf{f}_0, \mathbf{f}_4} y_2, L^2_{\mathbf{f}_0, \mathbf{f}_5} y_2, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_5} y_1, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_6} y_1, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_2} y_2, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_6} y_2$ .

As previously mentioned, we know that we cannot have more than 15 independent Lie derivatives

(otherwise, the yaw angle would be observable). 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 (8). They are: the 9 solutions provided in III-B and the six components of the two vectors  $\mathbf{b}_A$  and  $\mathbf{b}_\Omega$  (note that these components are trivial solutions of (8)).

### E. Unknown Gravity

The results provided in the previous sections are obtained by assuming that the magnitude of the gravity ( $g$ ) is a priori known. In this section we want to investigate if the information contained in the sensor data allows us to also estimate  $g$ . This investigation could seem useless since in most of cases the value  $g$  is known with good accuracy. On the other hand, this investigation allows us to derive several properties of practical importance.

We will show that  $g$  is among the observable modes even in the worst case when the inertial sensors are affected by bias and when only a single feature is available. We will proceed as in section III-D.

The system we are considering is defined by the state:  $[\mathbf{r} \ \mathbf{v} \ q \ \mathbf{b}_A \ \mathbf{b}_\Omega \ g]^T$ , whose dimension is 17. This state satisfies the dynamics in (6) with the additional equation  $\dot{g} = 0$ . Finally, this system is characterized by the observations given in (2) and (3).

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 III-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 (8), where  $\Lambda$  also depends on the components of  $\mathbf{b}_A$ ,  $\mathbf{b}_\Omega$  and on  $g$ . 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 16 independent Lie derivatives. By a direct computation, performed by using the symbolic Matlab computational tool, we were able to find the following 16 independent Lie derivatives:  $L^0 y_1, L^0 y_2, L^0 h_{const}, L^1_{\mathbf{f}_0} y_1, L^1_{\mathbf{f}_0} y_2, L^2_{\mathbf{f}_0, \mathbf{f}_0} y_1, L^2_{\mathbf{f}_0, \mathbf{f}_1} y_1, L^2_{\mathbf{f}_0, \mathbf{f}_4} y_1, L^2_{\mathbf{f}_0, \mathbf{f}_0} y_2, L^2_{\mathbf{f}_0, \mathbf{f}_4} y_2, L^2_{\mathbf{f}_0, \mathbf{f}_5} y_2, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_0} y_1, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_5} y_1, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_6} y_1, L^3_{\mathbf{f}_0, \mathbf{f}_3, \mathbf{f}_0} y_1, L^3_{\mathbf{f}_0, \mathbf{f}_0, \mathbf{f}_6} y_2$ . As previously mentioned, we know that we cannot have more than 16 independent Lie derivatives (otherwise, the yaw angle would be observable). The fact that we have 16 independent Lie derivatives means that there are no additional symmetries and, the independent

observable modes, are the independent solutions of (8). They are: the 15 solutions provided in III-D and  $g$ .

#### IV. NECESSARY CONDITIONS FOR OBSERVABILITY

The observability analysis performed so far takes into account all the degrees of freedom allowed by the dynamics in (1). In other words, the observability of the modes previously derived, could require the vehicle to move along all these degrees of freedom. It is interesting to understand what happens when only special trajectories are considered. Mathematically, this can be done by introducing in (1) the constraints characterizing the trajectory we want to consider. Then, it suffices to apply the method described in [14] to the system characterized by the new dynamics and the same observations (2) and (3).

By applying this technique we proved in [16] the two following properties:

**Property 2** *The absolute scale factor is not observable when the vehicle moves at constant speed.*

**Property 3** *When the vehicle moves at constant acceleration all the modes derived in section III are observable, with the exception of the magnitude of the gravitational acceleration ( $g$ ).*

A fundamental consequence of the previous two properties is:

**Theorem 1** *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 is unknown, the minimum number of camera images becomes four.*

*Proof:* The proof of this theorem is provided in [16]. In particular, it is shown that, if the observability of a given physical quantity requires to have a not constant speed, this observability also requires at least three camera observations. Similarly, it is shown that, if the observability of a given physical quantity requires to have a not constant acceleration, this observability also requires at least four camera observations. ■

In most of cases, the value  $g$  is known with good accuracy. Hence, considering the case of unknown magnitude of gravitational acceleration, could seem useless. On the other hand, considering this case has a very practical importance (see sections V and VI).

## V. CLOSED-FORM SOLUTIONS

This section provides closed form solutions which directly express the observable modes in terms of the sensor measurements collected during a short time interval. It starts by dealing with the case without bias.

### A. The case without Bias

In the local frame, the dynamics are:

$$\begin{cases} \dot{\mathbf{F}}^i = M\mathbf{F}^i - \mathbf{V} \\ \dot{\mathbf{V}} = M\mathbf{V} + \mathbf{A} + \mathbf{A}_g & i = 0, 1, \dots, N_f - 1 \\ \dot{\mathbf{q}} = m\mathbf{q} \end{cases} \quad (10)$$

where  $\mathbf{F}^i$  is the position of the  $i^{th}$  feature in the local frame ( $i = 0, 1, \dots, N_f - 1$ ),  $\mathbf{V}$  is the vehicle speed in the same frame,  $\mathbf{A}_g$  is the gravitational acceleration in the local frame, i.e.  $\hat{\mathbf{A}}_g = q^* \hat{a}_g q$ , and  $\mathbf{q}$  is the four vector whose components are the components of the quaternion  $q$ , i.e.  $\mathbf{q} = [q_t, q_x, q_y, q_z]^T$ . Finally:

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

The validity of (10) can be checked by using  $\hat{F} = -q^* \hat{r} q$ ,  $\hat{V} = q^* \hat{v} q$  and by computing their time derivatives with (1). In the local frame, the observation in (2) for the  $i^{th}$  feature is:

$$h_{cam} = [y_1, y_2]^T = \left[ \frac{F_x^i}{F_z^i}, \frac{F_y^i}{F_z^i} \right]^T \quad (11)$$

Because of the gravity, the first two equations in (10) cannot be separated from the equations describing the dynamics of the quaternion.

$\chi_g$  will denote the gravity vector in the local frame at a given time  $T_0$ . In other words,  $\chi_g \equiv \mathbf{A}_g(T_0)$ . Note that, determining  $\chi_g$  allows determining the roll and pitch angles ( $R_0$  and  $P_0$ ). Indeed, from the definition of the roll and pitch angles it is possible to obtain:

$$\boldsymbol{\chi}_g = g[\sin P_0, -\sin R_0 \cos P_0, -\cos R_0 \cos P_0]^T \quad (12)$$

$\mathbf{F}_0^i \equiv \mathbf{F}^i(T_0)$  will denote the position of the  $i^{\text{th}}$  feature ( $i = 0, 1, \dots, N_f - 1$ ) at  $T_0$ . Similarly,  $\mathbf{V}_0 \equiv \mathbf{V}(T_0)$  will denote the vehicle speed at  $T_0$ .

In the following, a closed form expression of the vectors  $\mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0$  and  $\boldsymbol{\chi}_g$  in terms of the sensor measurements in the time-interval  $[T_0, T_0 + T]$  will be provided.

To derive this closed form expression it is useful to first consider the special case where the vehicle does not rotate during the interval  $[T_0, T_0 + T]$ . In this case, the first two equations in (10) become:

$$\begin{cases} \dot{\mathbf{F}}^i = -\mathbf{V} \\ \dot{\mathbf{V}} = \mathbf{A} + \boldsymbol{\chi}_g \end{cases} \quad i = 0, 1, \dots, N_f - 1 \quad (13)$$

It is immediate to integrate the previous equations and obtain the position of the  $i^{\text{th}}$  feature in the local frame:

$$\mathbf{F}^i(t) = \mathbf{F}_0^i - \Delta t \mathbf{V}_0 - \frac{\Delta t^2}{2} \boldsymbol{\chi}_g - \int_{T_0}^t \int_{T_0}^{t'} \mathbf{A}(\tau) d\tau dt' \quad (14)$$

where  $\mathbf{A}(\tau)$  are provided by the accelerometers and  $\Delta t \equiv t - T_0$ .

$\Xi(t)$  will denote the matrix which characterizes the rotation occurred during the interval  $[T_0, t]$ . The equations in (14) correspond to the case when  $\Xi(t)$  is the identity  $3 \times 3$  matrix for any time  $t \in [T_0, T_0 + T]$ . In the general case, i.e. when the vehicle is not constrained to move with a fixed orientation,  $\Xi(t)$  can be evaluated by using the data from the gyroscopes during this time interval (see appendix A). Hence, it is possible to obtain the extension of (14) to a generic motion:

$$\mathbf{F}^i(t) = \Xi(t) \left( \mathbf{F}_0^i - \Delta t \mathbf{V}_0 - \frac{\Delta t^2}{2} \boldsymbol{\chi}_g - \int_{T_0}^t \int_{T_0}^{t'} \Xi^{-1}(\tau) \mathbf{A}(\tau) d\tau dt' \right), \quad i = 0, 1, \dots, N_f - 1 \quad (15)$$

In [16] the same result has been obtained by directly integrating the equations in (10).

The components of  $\mathbf{F}^i(t)$ , i.e.  $F_x^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \boldsymbol{\chi}_g)$ ,  $F_y^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \boldsymbol{\chi}_g)$  and  $F_z^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \boldsymbol{\chi}_g)$  are linear in the unknowns  $\mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \boldsymbol{\chi}_g$ .

By using (11) the following linear equations are obtained:

$$F_x^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \chi_g) = y_1(t) F_z^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \chi_g) \quad (16)$$

$$F_y^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \chi_g) = y_2(t) F_z^i(t; \mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \chi_g)$$

$$i = 0, 1, \dots, N_f - 1$$

In particular, each camera observation occurred at the time  $t \in [T_0, T_0 + T]$  provides  $2N_f$  linear equations in the  $3N_f + 6$  unknowns (which are the components of  $\mathbf{F}_0^i$  ( $i = 0, 1, \dots, N_f - 1$ ),  $\mathbf{V}_0$  and  $\chi_g$ ).

When the camera performs observations from  $n_{obs}$  distinct poses the number of equations provided by (16) is  $2n_{obs}N_f$  while the number of unknowns is  $3N_f + 6$ . In order to determine the unknowns, it is fundamental to know whether these equations are independent or not. To this regard, according to theorem 1, we know that the number of independent equations is always smaller than the number of unknowns for  $n_{obs} \leq 3$ . On the other hand, when  $n_{obs} = 3$  the knowledge of the magnitude of the gravity makes possible the determination of the modes.

1)  $n_{obs} \geq 4$ : In this case the equations in (16) are in general independent. On the other hand, when  $n_{obs} = 4$  and  $N_f = 1$ , the number of equations is 8, which is less than the number of unknowns 9. In section VI it is shown that, by using the knowledge of the gravity (i.e. the magnitude of the vector  $\chi_g$ ), it is possible to determine the unknowns by solving a second order polynomial equation. Hence, in this case, two solutions are determined. When  $n_{obs} \geq 5$  and/or  $N_f \geq 2$  the determination can be done by the computation of a pseudoinverse. Hence, a single solution can be obtained. Then, the knowledge of the magnitude of the gravitational acceleration, can be used to improve the precision (see section VI).

2)  $n_{obs} = 3$ : When  $N_f = 1$  the number of equations is 6 and the number of unknowns is 9. Hence the estimation cannot be performed. When  $N_f \geq 2$  the number of equations is larger or equal to the number of unknowns. On the other hand, according to theorem 1, the vector  $\chi_g$  cannot be determined, since its norm is not observable. In other words, the equations in (16) are not independent. As in the case  $n_{obs} = 4$ ,  $N_f = 1$ , it is possible to determine the unknowns by solving a second order polynomial equation. Hence, also in this case, two solutions are determined (see section VI and [16] for further details).

Note that the previous remarks hold in general. There are special situations, whose probability of occurrence is zero, where the determination cannot be carry out. For instance, in the case



$n_{obs} = 3$ ,  $N_f = 2$ , if one of the three camera poses is aligned along with the two features, the determination cannot be performed. Another special case is when the three camera poses and the two features belong to the same plane.

### B. The case with Bias

The closed-form solution will be derived only when the accelerometers are affected by a bias, i.e. in the case  $\mathbf{A}_{bias} \neq [0 \ 0 \ 0]^T$  and  $\mathbf{\Omega}_{bias} = [0 \ 0 \ 0]^T$ .

The expression in (15) can be easily extended to deal with this case by the substitution:  $\mathbf{A}(\tau) \rightarrow \mathbf{A}(\tau) + \mathbf{A}_{bias}$ :

$$\mathbf{F}^i(t) = \Xi(t) \left( \mathbf{F}_0^i - \Delta t \mathbf{V}_0 - \frac{\Delta t^2}{2} \boldsymbol{\chi}_g - \int_{T_0}^t \int_{T_0}^{t'} \Xi^{-1}(\tau) d\tau dt' \mathbf{A}_{bias} - \int_{T_0}^t \int_{T_0}^{t'} \Xi^{-1}(\tau) \mathbf{A}(\tau) d\tau dt' \right) \quad (17)$$

$$i = 0, 1, \dots, N_f - 1$$

By proceeding as in the case without bias the analogous of equations (16) is obtained. The new equations also depend on the vector  $\mathbf{A}_{bias}$ .

## VI. THE ALGORITHM

This section describes the algorithm which allows determining the speed, the attitude and the absolute scale starting from the inertial and visual data collected in a very short time interval. The extension to also determine the bias is straightforward.

As stated at the beginning of the previous section, the local frame is the camera frame. Hence, it is necessary to determine the expression of the acceleration and the angular speed of this frame starting from the acceleration and the angular speed provided by the IMU and from the knowledge of the transformation between the IMU frame and the camera frame (which is assumed a priori known).

The following notation will be adopted:

- $C$  is the matrix which transforms vectors in the IMU frame in vectors expressed in the camera frame;
- $D$  is the vector describing the position of the origin of the camera frame in the IMU frame;

- $\boldsymbol{\Omega}_{IMU}$  and  $\mathbf{A}_{IMU}$  are the angular speed and the acceleration in the IMU frame (i.e. provided by the inertial sensors).

The expressions of the angular speed and the acceleration in the camera frame are:

$$\boldsymbol{\Omega} = C\boldsymbol{\Omega}_{IMU}, \quad \mathbf{A} = C \left[ \mathbf{A}_{IMU} + \dot{\boldsymbol{\Omega}}_{IMU} \wedge \mathbf{D} + \boldsymbol{\Omega}_{IMU} \wedge (\boldsymbol{\Omega}_{IMU} \wedge \mathbf{D}) \right] \quad (18)$$

The previous expressions must be used to obtain all the inertial measurements in the camera frame.

The second step consists in computing the matrix  $\Xi$  at each time step when the inertial data are delivered (see appendix A). This allows computing the term  $-\int_{T_0}^t \int_{T_0}^{t'} \Xi^{-1}(\tau) \mathbf{A}(\tau) d\tau dt'$  for all the time  $t$  when a camera image is available.

The linear system in (16) will be denoted with:

$$\Gamma \mathbf{x} = \boldsymbol{\beta} \quad (19)$$

where the vector  $\mathbf{x}$  contains all the unknowns, i.e.  $\mathbf{x} \equiv [\mathbf{F}_0^0, \mathbf{F}_0^1, \dots, \mathbf{F}_0^{N_f-1}, \mathbf{V}_0, \boldsymbol{\chi}_g]^T$ .  $\Gamma$  and  $\boldsymbol{\beta}$  are respectively a  $(2n_{obs}N_f \times 3N_f + 6)$  matrix and a  $(2n_{obs}N_f \times 1)$  vector and are obtained as follows. For a camera observation occurred at time  $t$ , each feature contributes with two rows to the matrix  $\Gamma$  and with two entries to the vector  $\boldsymbol{\beta}$ .

For the  $j^{th}$  feature observed at time  $t$  the three rows of the matrix  $\Xi(t)$  will be denoted with  $\boldsymbol{\xi}_1(t)$ ,  $\boldsymbol{\xi}_2(t)$  and  $\boldsymbol{\xi}_3(t)$ . The two rows of the matrix  $\Gamma$  are, respectively:

$$\left[ \mathbf{0}_{3j-3}, \boldsymbol{\xi}_1(t) - y_1(t)\boldsymbol{\xi}_3(t), \mathbf{0}_{3N_f-3j}, -\Delta t(\boldsymbol{\xi}_1(t) - y_1(t)\boldsymbol{\xi}_3(t)), -\frac{\Delta t^2}{2}(\boldsymbol{\xi}_1(t) - y_1(t)\boldsymbol{\xi}_3(t)) \right]$$

$$\left[ \mathbf{0}_{3j-3}, \boldsymbol{\xi}_2(t) - y_2(t)\boldsymbol{\xi}_3(t), \mathbf{0}_{3N_f-3j}, -\Delta t(\boldsymbol{\xi}_2(t) - y_2(t)\boldsymbol{\xi}_3(t)), -\frac{\Delta t^2}{2}(\boldsymbol{\xi}_2(t) - y_2(t)\boldsymbol{\xi}_3(t)) \right]$$

where  $\mathbf{0}_n$  denotes the row-vector whose dimension is  $n$  and whose entries are all zeros. The corresponding two entries in the vector  $\boldsymbol{\beta}$  are, respectively:

$$[\boldsymbol{\xi}_1(t) - y_1(t)\boldsymbol{\xi}_3(t)] \int_{T_0}^t \int_{T_0}^{t'} \Xi^{-1}(\tau) \mathbf{A}(\tau) d\tau dt', \quad [\boldsymbol{\xi}_2(t) - y_2(t)\boldsymbol{\xi}_3(t)] \int_{T_0}^t \int_{T_0}^{t'} \Xi^{-1}(\tau) \mathbf{A}(\tau) d\tau dt'$$

As stated in the previous section, it is possible in general to compute the pseudoinverse (or the inverse) of the matrix  $\Gamma$  in the following cases:

- 1) when  $n_{obs} \geq 4$  and  $N_f \geq 2$ ;
- 2) when  $n_{obs} \geq 5$  and  $N_f = 1$ .

When the rank of  $\Gamma$  is one less than the number of its columns, the nullspace of  $\Gamma$  has dimension one. As discussed at the end of the previous section, this is in general the case when  $n_{obs} = 3$ ,  $N_f \geq 2$  (because of the theorem 1) or when  $n_{obs} = 4$ ,  $N_f = 1$ . In this case, the system in (19) has an infinite number of solutions. By denoting with  $\boldsymbol{\nu}$  the unit vector belonging to the nullspace of  $\Gamma$ , with  $\boldsymbol{x}_p$  one among the solutions of (19), any solution of (19) is

$$\boldsymbol{x} = \boldsymbol{x}_p + \lambda \boldsymbol{\nu}$$

where  $\lambda$  is a real number. On the other hand, by knowing the magnitude of the gravitational acceleration, it is possible to determine two values of  $\lambda$ . This is obtained by enforcing the constraint that the vector  $\boldsymbol{s}_\lambda$  constituted by the last three entries of the solution  $\boldsymbol{x}_p + \lambda \boldsymbol{\nu}$  is a vector with norm equal to  $g$ . In other words:

$$|\boldsymbol{s}_\lambda|^2 = g^2 \tag{20}$$

which is a second order polynomial equation in  $\lambda$ . Hence, in this case two solutions are determined.

Finally, when  $\Gamma$  is full rank, the knowledge of the magnitude of the gravitational acceleration can be exploited by minimizing the cost function:

$$c(\boldsymbol{x}) = |\Gamma \boldsymbol{x} - \boldsymbol{\beta}|^2 \tag{21}$$

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

The main steps of the algorithm are displayed in the algorithm 1.

**Algorithm 1 (Returns features's positions, speed and attitude)**

*Inputs:*  $\boldsymbol{A}_{IMU}(t)$ ,  $\boldsymbol{\Omega}_{IMU}(t)$ ,  $y_1^i(t)$ ,  $y_2^i(t)$ , ( $i = 0, 1, \dots, N_f - 1$ ),  $t \in [T_0, T_0 + T]$

*Outputs:*  $\boldsymbol{F}_0^i$ ,  $\boldsymbol{V}_0$ ,  $\boldsymbol{\chi}_g$ , ( $i = 0, 1, \dots, N_f - 1$ )

*Compute  $\mathbf{A}$  and  $\mathbf{\Omega}$  by using (18)*

*Build the matrix  $\Gamma$  and the vector  $\beta$  in (19)*

*Compute the rank ( $r$ ) of  $\Gamma$*

***if  $r = 3N_f + 6$  then***

*$\mathbf{x}_{in} = \text{pinv}(\Gamma)\beta$*

*minimize  $c(\mathbf{x})$  in (21) with initialization  $\mathbf{x}_{in}$*

***else***

***if  $r = 3N_f + 5$  then***

*Determine two solutions by (20)*

***else***

*Determination non possible*

***end if***

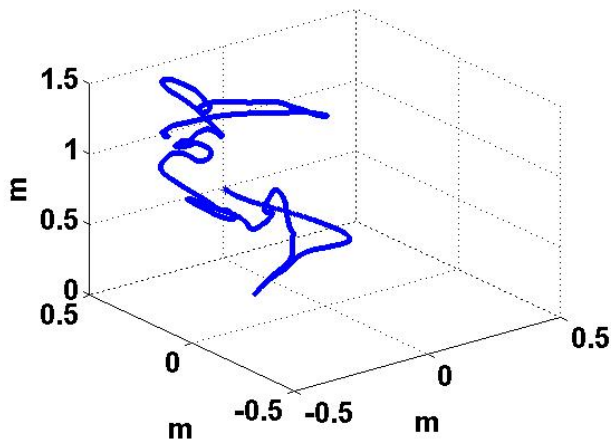
***end if***

## VII. PERFORMANCE EVALUATION

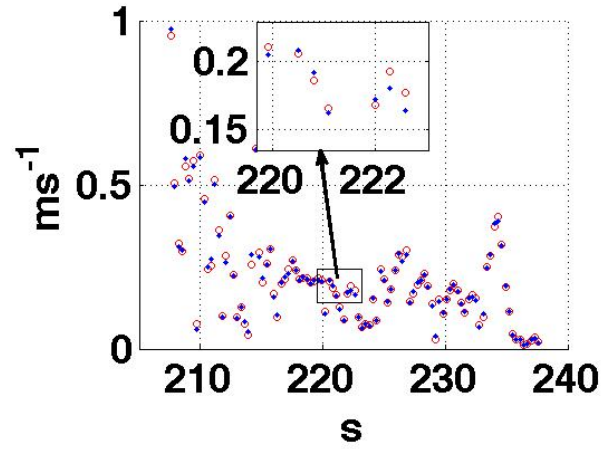
This section shows the results obtained by using the algorithm with a real data set. The data have been provided by the autonomous system laboratory at ETHZ in Zurich. The data are provided together with a reliable ground-truth, which has been obtained by performing the experiments at the ETH Zurich Flying Machine Arena [13], which is equipped with a Vicon motion capture system. The visual and inertial data are obtained with a monochrome USB-camera gathering  $752 \times 480$  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 [17]. Finally, the extrinsic calibration between the camera and the IMU has been obtained by using the strategy introduced in [12]. The experiment here analyzed lasted for about  $250s$ .

Figure 3 *a* shows the trajectory (ground truth) during the time interval  $[200, 240]s$ .

Figures 3 *b* and 4 show the results regarding the estimated speed, roll and pitch angles, respectively. In all these figures, the blue dots are the ground truth while the red disks are the estimated values.



*a*



*b*

Fig. 3. 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.

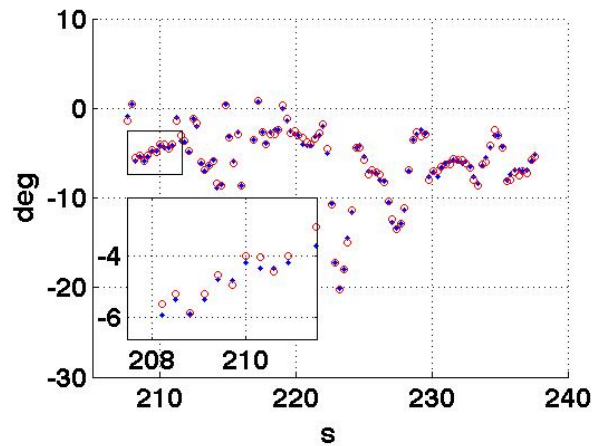
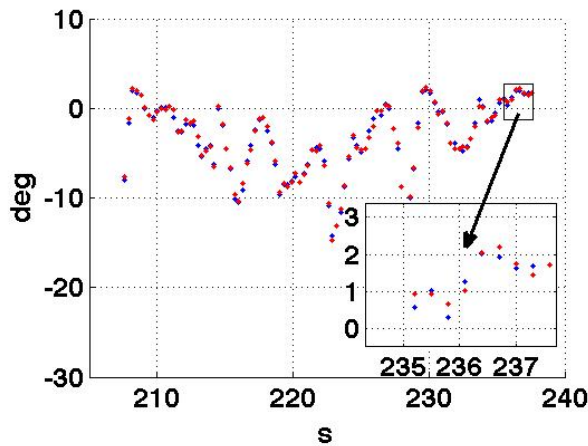


Fig. 4. Roll (left) and pitch (right) angles in the real 3D experiment. Blue dots are the ground truth and red disks the estimated values.

## VIII. CONCLUSIONS

This chapter described a method for determining the speed and the attitude of a vehicle equipped with a monocular camera and inertial sensors (i.e. one tri-axial accelerometer and one tri-axial gyrometer). The vehicle moves in a 3D unknown environment. It has been shown that, by collecting the visual and inertial measurements during a very short time interval, it is possible to determine the following physical quantities: the vehicle speed and attitude, the absolute distance

of the point features observed by the camera during the considered time interval and the bias affecting the inertial measurements. In particular, this determination, is based on a closed form solution which analytically expresses the previous physical quantities in terms of the sensor measurements. This closed form determination allows performing the overall estimation in a very short time interval and without the need of any initialization or a priori knowledge. This is a key advantage since allows eliminating the drift on the scale factor and on the vehicle orientation.

## APPENDIX A

### EXPRESSION OF THE ROTATION MATRIX $\Xi$ BY INTEGRATING THE ANGULAR SPEED

Let consider a vehicle and a frame attached to this vehicle. When the vehicle moves during the infinitesimal interval  $[t_j, t_j + \delta t]$ , the rotation matrix which transforms vectors in the reference before this motion and the reference after this motion is:  $I_3 + M_j \delta t$ , where  $I_3$  is the  $3 \times 3$  identity matrix and  $M_j$  is the skew-symmetric defined in section V at the time  $t_j$ .

Now, let us suppose that the vehicle moves during the interval of time  $[t_i, t_f]$ . In order to compute the rotation matrix which transforms vectors in the reference before this motion and the reference after this motion, the path is divided in many ( $N$ ) steps. For each step, the expression of the rotation matrix is the one previously provided. Then, it suffices to compute the product of all these matrices, namely:

$$\Xi = \prod_{k=1}^N (I_3 + M_k \delta t_k) \quad (22)$$

where  $t_1 = t_i$  and  $t_N = t_f$ .

## REFERENCES

- [1] Ahrens, S.; Levine, D.; Andrews, G.; How, J.P., Vision-based guidance and control of a hovering vehicle in unknown, gps-denied environments, IEEE International Conference on Robotics and Automation (ICRA 2009), Kobe, Japan, May, 2009.
- [2] 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
- [3] Blosch, M., Weiss, S., Scaramuzza, D., and Siegwart, R. (2010), Vision Based MAV Navigation in Unknown and Unstructured Environments, IEEE International Conference on Robotics and Automation (ICRA 2010), Anchorage, Alaska, May, 2010.

- [4] Bryson, M. and Sukkarieh, S., Building a Robust Implementation of Bearing-only Inertial SLAM for a UAV, *JFR*, 2007, 24, 113-143
- [5] P. Corke, J. Lobo, and J. Dias, An Introduction to Inertial and Visual Sensing, *International Journal of Robotics Research* 2007 26: 519-535
- [6] J. Dias, M. Vincze, 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.
- [7] 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
- [8] Goldstein, H. *Classical Mechanics*, 2nd ed. Reading, MA: Addison-Wesley, 1980
- [9] Kim, J. and Sukkarieh, S. Real-time implementation of airborne inertial-SLAM, *Robotics and Autonomous Systems*, 2007, 55, 62–71
- [10] *Quaternions and rotation Sequences: a Primer with Applications to Orbits, Aerospace, and Virtual Reality*. Kuipers, Jack B., Princeton University Press copyright 1999.
- [11] F. John, *Partial Differential Equations*, Springer-Verlag, 1982.
- [12] J. Lobo and J. Dias, Relative pose calibration between visual and inertial sensors, *International Journal of*
- [13] S. Lupashin, A. Schollig, M. Sherback and R. D’Andrea, A simple learning strategy for high-speed quadcopter multi-flips, *IEEE International Conference on Robotics and Automation*, Anchorage, 2010
- [14] Agostino Martinelli, State Estimation Based on the Concept of Continuous Symmetry and Observability Analysis: the Case of Calibration, *IEEE Transactions on Robotics*, vol. 27, no. 2, pp. 239-255, 2011
- [15] A. Martinelli, Closed-Form Solution for Attitude and Speed Determination by Fusing Monocular Vision and Inertial Sensor Measurements, to be presented at the *International Conference on Robotics and Automation, ICRA 2011*, Shanghai, China
- [16] A. Martinelli, Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination, Accepted for publication on *Transaction on Robotics*.
- [17] D. Scaramuzza, A. Martinelli and R. Siegwart, A toolbox for easy calibrating omnidirectional cameras, *IEEE International Conference on Intelligent Robots and Systems*, 2006
- [18] M. Veth, and J. Raquet, Fusing low-cost image and inertial sensors for passive navigation, *Journal of the Institute of Navigation*, 54(1), 2007