



HAL
open science

Vision-Aided Inertial Navigation Using Virtual Features

Chiara Troiani, Agostino Martinelli

► **To cite this version:**

Chiara Troiani, Agostino Martinelli. Vision-Aided Inertial Navigation Using Virtual Features. Iros 2012, Oct 2012, Portugal. pp.4828–4834. hal-00743327v1

HAL Id: hal-00743327

<https://hal.science/hal-00743327v1>

Submitted on 18 Oct 2012 (v1), last revised 12 Nov 2012 (v2)

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-Aided Inertial Navigation Using Virtual Features

Chiara Troiani and Agostino Martinelli

Abstract—This paper considers the following scenario. A vehicle accomplishes a 3D-trajectory in the surrounding of a planar surface. The vehicle is equipped with a monocular camera and inertial sensors. Additionally, a laser pointer is mounted on the vehicle and it produces a laser spot on the planar surface. This laser spot is observed by the monocular camera and it is the unique point feature used in the proposed approach. The paper provides two main contributions. The former is the analytical derivation of all the observable modes, i.e. all the physical quantities that can be determined by only using the inertial data and the camera observations of the laser spot during a short time-interval. Specifically, it is shown that the observable modes are: the distance of the vehicle from the planar surface; the component of the vehicle speed, which is orthogonal to the planar surface; the relative orientation of the vehicle with respect to the planar surface; the orientation of the planar surface with respect to the gravity. The second contribution is the introduction of a simple recursive method to perform the estimation of all the aforementioned observable modes. This method is based on a local decomposition of the original system, which separates the observable modes from the rest of the system. The method is validated by using both synthetic and real data.

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.

A special issue of the *International Journal of Robotics Research* has recently been devoted to the problem of fusing vision and inertial data [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. 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 [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 [20] 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 [12] 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. More recently, a vision based navigation approach in unknown and unstructured environments has been suggested [3].

Recent works investigate the observability properties of the vision-aided inertial navigation system [9], [10], [11], [17], [18] and [19]. In particular, in [18], the observable modes are expressed in closed-form in terms of the sensor measurements acquired during a short time-interval.

In this paper we consider a vehicle which accomplishes a 3D-trajectory in the surrounding of a planar surface. The vehicle is equipped with a monocular camera and inertial sensors. This is the typical navigation problem in an indoor environment or in a city-like environment. All the approaches previously mentioned, require to extract natural features from the images provided by the camera and in particular to detect the same features in different images. The feature matching task becomes critical in outdoor environment because of possible illumination changes. In order to significantly reduce the computational burden required to perform these tasks and to make the feature matching more robust, we introduce a *virtual* feature by equipping our vehicle with a laser pointer. The laser beam produces a laser spot on the planar surface. This laser spot is observed by the monocular camera and it is the unique point feature used by the proposed approach.

To the best of our knowledge, this problem has never been considered so far. Compared to classical vision and IMU data fusion problems, the feature is moving in the environment but we exploit the hypothesis that it moves on a planar surface. The first question which arises is to understand which are the observable modes, i.e. the physical quantities that can be determined by only using the inertial data and the camera observation of the laser spot during a short time-interval. We will show (in section III and appendix II) that the observable modes are: the distance of the vehicle from the planar surface; the component of the vehicle speed, which is orthogonal to the planar surface; the relative orientation of

This work was supported by the European Project FP7-ICT-2007-3.2.2 Cognitive Systems, Interaction, and Robotics, 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 and C. Troiani are with INRIA Rhone Alpes, Montbonnot, France e-mail: agostino.martinelli@ieee.org, chiara.troiani@inria.fr

the vehicle with respect to the planar surface; the orientation of the planar surface with respect to the gravity. Then, the second step we consider is to analytically determine the link between the observable modes and the sensor data. This is obtained by performing a local decomposition of the original system (section IV). This decomposition separates the observable modes from the rest of the system. This decomposition will allow us to introduce a simple recursive method to perform the estimation of all the observable modes (second part of section IV). The method is validated by using both synthetic and real data (section V).

II. THE CONSIDERED SYSTEM

Let us consider an aerial vehicle equipped with a monocular camera and *IMU* sensors. The vehicle is also equipped with a laser pointer. The configuration of the laser pointer in the camera reference frame is known. The vehicle moves in front of a planar surface and we assume that the laser spot produced by the laser beam belongs to this planar surface (see fig. 1). The position and the orientation of this planar surface are unknown. The camera observations consist in the position of the laser spot in the camera frame up to a scale factor. The *IMU* consists of three orthogonal accelerometers and three orthogonal gyroscopes. We assume that the monocular camera is intrinsically calibrated and that the transformations among the camera frame and the *IMU* frames are known (we can assume that the vehicle frame coincides with the camera frame). The *IMU* provides the vehicle angular speed and acceleration. Actually, regarding the acceleration, the one perceived by the accelerometer (\mathbf{A}) is not simply the vehicle acceleration (\mathbf{A}_v). It also contains the gravity acceleration (\mathbf{G}). In particular, we have $\mathbf{A} = \mathbf{A}_v - \mathbf{G}$ 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.

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: $\mathbf{g} = [0, 0, -g]^T$, being $g \simeq 9.8 \text{ ms}^{-2}$.

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 [13].

Our system is characterized by the state $[\mathbf{r}, \mathbf{v}, \mathbf{q}]^T$ 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 $\mathbf{w} = [w_x, w_y, w_z]^T$ we associate with it the imaginary quaternion $w_q \equiv 0 + iw_x + jw_y + kw_z$. The dynamics of the state $[\mathbf{r}_q, \mathbf{v}_q, \mathbf{q}]^T$ are:

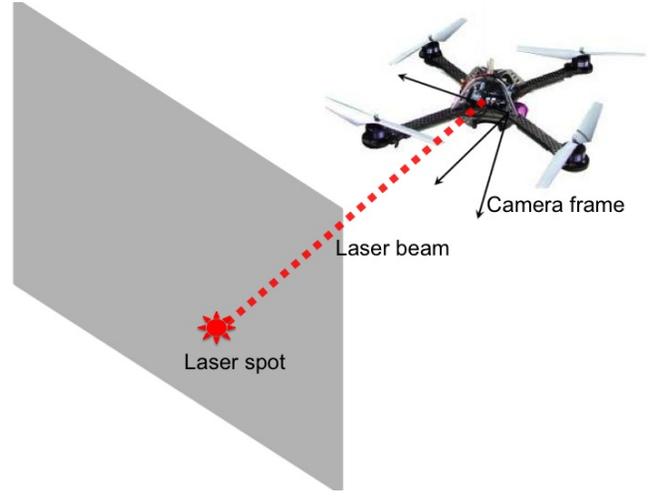


Fig. 1. Helicopter equipped with a monocular camera, *IMU* and a laser pointer. The laser spot is on a planar surface and its position in the camera frame is obtained by the camera up to a scale factor.

$$\begin{cases} \dot{r}_q = v_q \\ \dot{v}_q = qA_{vq}q^* = qA_qq^* + g_q \\ \dot{q} = \frac{1}{2}q\Omega_q \end{cases} \quad (1)$$

being q^* the conjugate of q , $q^* = q_t - iq_x - jq_y - kq_z$ and Ω the angular velocity.

We derive the expression of the camera observation consisting in the position of the laser spot in the camera frame up to a scale factor. The laser spot is on a planar surface whose configuration is unknown. Without loss of generality, we choose the camera frame with the z -axis parallel to the laser pointer. In addition, the camera frame is such that the laser beam intercept the xy -plane in $[L, 0, 0]$. In appendix I we introduce a simple and efficient method in order to determine the parameter L together with the rotation to transform vectors from the original camera frame into the chosen camera frame.

Finally, we characterize the planar surface in the global frame with the equation $z = ky$, where k is an unknown parameter.

In these settings, by carrying out analytical computation (which uses the basic quaternion rules) we obtain the analytical expression of the position $[X, Y, Z]$ of the laser spot in the camera reference frame. We have:

$$\begin{cases} X = L \\ Y = 0 \\ Z = \frac{r_z + 2q_zq_xL - 2q_ykq_xL - 2q_yq_tL - 2q_tLkq_z - kr_y}{2kq_zq_y - 2kq_tq_x - q_z^2 - q_t^2 + q_y^2 + q_x^2} \end{cases} \quad (2)$$

The camera provides the vector $[X, Y, Z]$ up to a scale factor. This is equivalent to the two ratios $\frac{X}{Z}$ and $\frac{Y}{Z}$. Hence, since the latter is identically zero, the camera observation is given by $h_{cam} = \frac{X}{Z}$, which is:

$$h_{cam} = \frac{L(2kq_zq_y - 2kq_tq_x - q_z^2 - q_t^2 + q_y^2 + q_x^2)}{r_z + 2q_zq_xL - 2q_ykq_xL - 2q_yq_tL - 2q_tLkq_z - kr_y} \quad (3)$$

III. OBSERVABILITY PROPERTIES

We investigate the observability properties of the system whose dynamics are given in (1) and whose observation function is given in (3). We have also to consider the constraint $q^*q = 1$. This can be dealt as a further observation (system output):

$$h_{const}(r_q, v_q, q) = q^*q \quad (4)$$

Finally, we want to investigate whether the parameter k is identifiable or not. This is done by performing an observability analysis on the extended state $\mathcal{S} = [r_q, v_q, q, k]^T$, whose dynamics are given in (1) and by the additional equation $\dot{k} = 0$.

We apply the method introduced in [16]. This will allow us to detect all the observable modes, i.e. all the physical quantities that can be determined by only using the information contained in the data provided by the IMU and the camera in a given time-interval. In appendix II we perform this analytical computation and we show that the system has six observable modes. Additionally, in the appendix II, we provide the physical meaning of these observable modes. Specifically, we found convenient to express the observable modes in a new global reference frame, denoted with \tilde{x} , \tilde{y} , \tilde{z} . In this frame the \tilde{z} -axis coincides with the axis normal to the planar surface. From now on, we will adopt this global frame to characterize the vehicle configuration. The state in this frame is $\tilde{\mathcal{S}} = [\tilde{r}_q, \tilde{v}_q, \tilde{q}, k]^T$. The observable modes are (see the derivation in appendix II):

$$\begin{aligned} m_1 &= \tilde{r}_z \\ m_2 &= \tilde{v}_z \\ m_3 &= 2(\tilde{q}_t\tilde{q}_x + \tilde{q}_y\tilde{q}_z) \\ m_4 &= 2(\tilde{q}_t\tilde{q}_y - \tilde{q}_x\tilde{q}_z) \\ m_5 &= k \\ m_6 &= \tilde{q}^*\tilde{q} \end{aligned} \quad (5)$$

m_1 is the \tilde{z} -coordinate of the vehicle, m_2 the component of the vehicle speed along the \tilde{z} -axis, m_3 and m_4 are related to the roll and pitch angles of the vehicle in this frame. In particular, the roll is $\arctan\left(\frac{m_3}{\sqrt{1-m_3^2-m_4^2}}\right)$ and the pitch is $\arcsin(m_4)$. m_5 is related to the orientation of the $\tilde{x}\tilde{y}$ -plane with respect to the gravity. In particular, the \tilde{z} -axis makes an angle $\arctan(k) = \arctan(m_5)$ with the gravity. m_6 is trivially the magnitude of the quaternion \tilde{q} , which is 1 since it describes a rotation.

By summarizing the results of the observability analysis performed in this section (and in the appendix II) we say that the information contained in the data provided by the IMU and the camera during a given time-interval, allows us to determine the six modes m_1, \dots, m_6 . For this reason, in the

rest of the paper, we will focus our attention only on these six quantities (actually, on the first five, since m_6 trivially expresses the constraint of having a unitary quaternion).

IV. LOCAL DECOMPOSITION AND RECURSIVE ESTIMATION

The goal of this section is to provide a method able to estimate the observable modes, i.e. the distance of the vehicle from the planar surface ($|m_1|$), the component of the vehicle speed orthogonal to the surface (m_2), the attitude of the vehicle relative to the surface (characterized by m_3 and m_4) and the orientation of the planar surface with respect to the gravity ($\arctan(m_5)$). To achieve this goal, the first step is to determine the link between the observable modes and the sensor data. By adopting the terminology introduced in [8], we have to perform a local decomposition of our system. We remind the reader that the local decomposition is the extension of the Kalman canonical decomposition [4] to the case of a non linear system. It consists in writing the equations characterizing the dynamics and the observation only in terms of the observable modes. We also remind the reader that in the non linear case it is often impossible to characterize the system with a unique decomposition. The decomposition only holds in a local region of the space of states. This is the reason why it is called *local* decomposition. To cover the entire space of states more than one decomposition is required (see [8]). In the following, we will show that for our system the number of decompositions is two.

We first provide the dynamics of the state $\tilde{\mathcal{S}} = [\tilde{r}_q, \tilde{v}_q, \tilde{q}, k]^T$. We use (11), (12), (13) and (1). We obtain:

$$\begin{cases} \dot{\tilde{r}}_q = \tilde{v}_q \\ \dot{\tilde{v}}_q = \tilde{q}A_q\tilde{q}^* + p^*g_qp \\ \dot{\tilde{q}} = \frac{1}{2}\tilde{q}\Omega_q \\ \dot{k} = 0 \end{cases} \quad (6)$$

A local decomposition for the dynamics is:

$$\begin{cases} \dot{m}_1 = m_2 \\ \dot{m}_2 = -m_4A_x + m_3A_y + \xi(m_3, m_4)A_z + g_z \\ \dot{m}_3 = \Omega_x \xi(m_3, m_4) + \Omega_z m_4 \\ \dot{m}_4 = \Omega_y \xi(m_3, m_4) - \Omega_z m_3 \\ \dot{g}_z = 0 \end{cases} \quad (7)$$

where g_z is the component of the gravity along the \tilde{z} -axis, i.e. $g_z = -g \cos(\arctan(k)) = \frac{-g}{\sqrt{1+k^2}} = \frac{-g}{\sqrt{1+m_5^2}}$; the function $\xi(m_3, m_4)$ depends on the original state and in particular changes its sign depending on the sign of $\tilde{q}_x^2 + \tilde{q}_y^2 - \frac{1}{2}$:

$$\xi(m_3, m_4) \equiv \begin{cases} \sqrt{1-m_3^2-m_4^2} & \text{if } \tilde{q}_x^2 + \tilde{q}_y^2 < \frac{1}{2} \\ -\sqrt{1-m_3^2-m_4^2} & \text{if } \tilde{q}_x^2 + \tilde{q}_y^2 > \frac{1}{2} \end{cases} \quad (8)$$

Hence, as previously said, we have two local decompositions for our original system. The validity of (7) can be checked by using (5) and (6). Note that deriving (7) is troublesome. In contrast, checking its validity is very simple since only demands to perform differentiation.

To complete the local decomposition we need to express the camera observation function in (3) in terms of the observable modes. We obtain:

$$h_{cam} = \frac{L\xi(m_3, m_4)}{m_4L - m_1} \quad (9)$$

The validity of (9) can be checked by using (3), (5), (8), (11), (12) and (13).

The equations (7) and (9) represent a local decomposition for our system. They provide the analytical expression of the link between the observable modes and the sensor data. Specifically, equation (7) provides the link between the observable modes and the *IMU* data. Equation (9) provides the link between the observable modes and the data delivered by the monocular camera. Having these equations allows us to perform the estimation of the state $[m_1, m_2, m_3, m_4, g_z]$. An efficient and simple approach is obtained by using an Extended Kalman Filter, *EKF*. To implement this filter it suffices to compute the Jacobian of the dynamics in (7) and the Jacobian of the observation function in (9) (see [2]). We provide these Jacobians in appendix III.

V. 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.

A. Simulations

We simulate many different trajectories in *3D* and many different scenarios corresponding to different orientation of the planar surface. For all the simulations we use the proposed strategy to estimate the observable modes, i.e.:

- 1) the distance of the vehicle from the planar surface ($d = |m_1|$);
- 2) the component of the vehicle speed orthogonal to the planar surface ($v_o = m_2$);
- 3) the roll (R) and the pitch (P) angles in the \tilde{x} , \tilde{y} , \tilde{z} -frame (i.e. the frame where the \tilde{x} , \tilde{y} -plane coincides with the planar surface);
- 4) the orientation of the plane with respect to the gravity (α).

Specifically, in all the simulations the values of the estimated d , v_o , R , P and α 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 vehicle 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 . We set the parameters in order to be close to a real case: $\mu_a = 0 \text{ ms}^{-2}$, $\sigma_a = 1 \text{ ms}^{-2}$, $\mu_\omega = 0 \text{ deg s}^{-1}$ and $\sigma_\omega = 1 \text{ deg s}^{-1}$. The initial vehicle position is at $\tilde{x} = 0$, $\tilde{y} = 0$, $\tilde{z} = 1m$. The initial vehicle speed is $[1, 0, 0] \text{ ms}^{-1}$ in the \tilde{x} , \tilde{y} , \tilde{z} -frame.

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 Ω_i^{true} and \mathbf{A}_v^{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: $\Omega_i = N(\Omega_i^{true} - \Omega_{bias}, P_{\Omega_i})$ and $\mathbf{A}_i = N(\mathbf{A}_v^{true} - \mathbf{A}_{gi} - \mathbf{A}_{bias}, P_{A_i})$ 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*;
- \mathbf{A}_{gi} is the gravitational acceleration in the local frame and \mathbf{A}_{bias} is the bias affecting the data from the accelerometer;
- Ω_{bias} is the bias affecting the data from the gyroscope.

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: $\sigma_{gyro} \in [0.3, 10] \text{ deg s}^{-1}$ and $\sigma_{acc} \in [0.01, 0.3] \text{ ms}^{-2}$.

Regarding the camera, the provided readings are generated in the following way. By knowing the true trajectory, and the position and the orientation of the planar surface, the true bearing angles of the laser spot in the camera frame are computed¹. They are computed each 0.1s. The parameter L is set equal to 0.1m. Then, the camera readings are generated by adding to the true values zero-mean Gaussian errors whose variance is equal to $(1 \text{ deg})^2$ for all the readings.

3) *Simulation Results*: Figure 2 displays a typical *3D* trajectory obtained in our simulations. The figure also displays the planar surface, consisting of a plane, which makes an angle of $\alpha = \frac{\pi}{8} \text{ rad} = 22.5 \text{ deg}$ with the gravity.

Figures 3 *a* and *b* display the estimated α respectively in the case without and with bias. The values of the biases adopted in our simulations are: $\Omega_{bias} = [1 \ 0 \ 0]^T (\text{deg s}^{-1})$ and $\mathbf{A}_{bias} = [0 \ 0.1 \ 0]^T (\text{ms}^{-2})$. As expected, the estimation in presence of bias becomes worse. However, the error on the estimated α in presence of bias is smaller than 1 *deg*.

Figures 4 *a* and *b* display the estimated P and R in the bias-free case while figures 5 *a* and *b* display the estimated P and R in the case of biased inertial data. Figures 6 *a* and *b* display the estimated v_o and d in the bias-free case while figures 7 *a* and *b* display the estimated v_o and d in the case of biased inertial data. We remark that the convergence of the filter occurs in less than 2 seconds for all the observable modes. To this regard, we initialized the filter by using a

¹This is obtained also by knowing that the laser pointer has the same orientation as the camera and that it is located at the position $[L, 0, 0]$

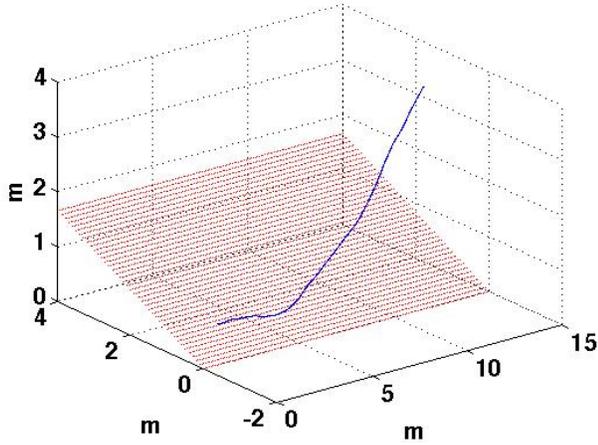


Fig. 2. A typical vehicle trajectory in our simulations.

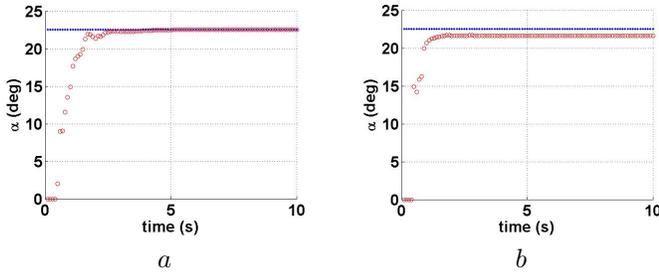


Fig. 3. Estimated α in absence (a) and in presence of bias (b) on the inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.

value of the initial observable state which differs from the ground truth by a relative error in the range $[10, 20]\%$. We also remark that the presence of bias corrupts the estimation of the pitch while it is negligible on the other observable modes.

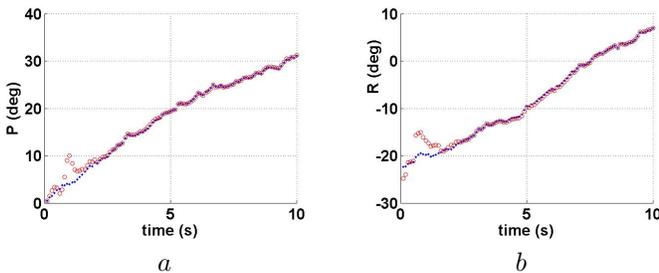


Fig. 4. Estimated P (a) and R (b) in the bias-free case. Blue dots indicate ground true values while red discs indicate the estimated values.

B. Preliminary experiments

In this last section we provide preliminary results obtained by using the Pelican helicopter, shown in figure 8. These are preliminary results since, as will be explained below, the observations of the laser spot by the monocular camera are simulated. The data have been provided by the autonomous

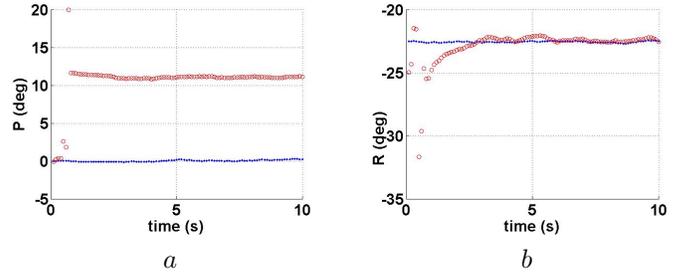


Fig. 5. Estimated P (a) and R (b) in the case of biased inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.

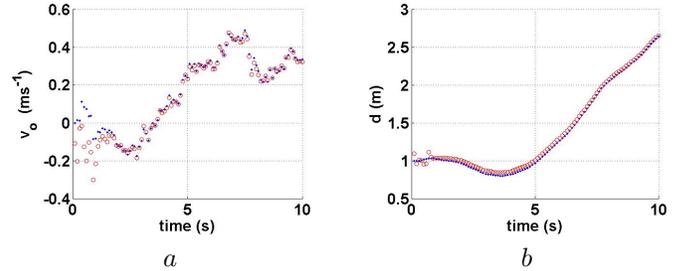


Fig. 6. Estimated v_o (a) and d (b) in the bias-free case. Blue dots indicate ground true values while red discs indicate the estimated values.

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 [15], which is equipped with a Vicon motion capture system. As previously said, the observations of the laser spot are simulated. This was possible thanks to the fact that a reliable ground truth was provided together with the inertial data. In particular, given the true trajectory, we simulated the same planar surface described in the previous section. By having the true vehicle configuration it was possible to create the observations performed by the camera on the laser spot produced by a laser pointer as in the simulations (see the last paragraph in V-A.2).

Figure 9 displays the estimated α . Figures 10 a and b display the estimated P and R and figures 11 a and b display the estimated v_o and d . All the observable modes are estimated with very good accuracy. Additionally, we remark

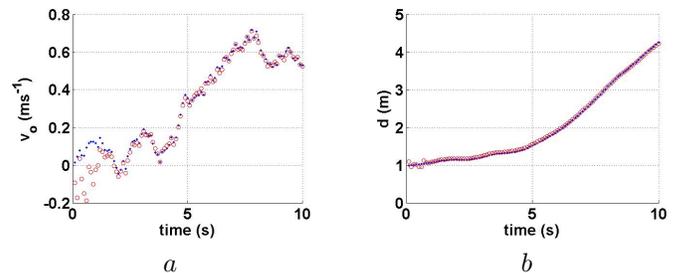


Fig. 7. Estimated v_o (a) and d (b) in the case of biased inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.



Fig. 8. The Pelican Helicopter used to take inertial data and ground truth data.

that the convergence of the filter occurs in less than half second for all the observable modes.

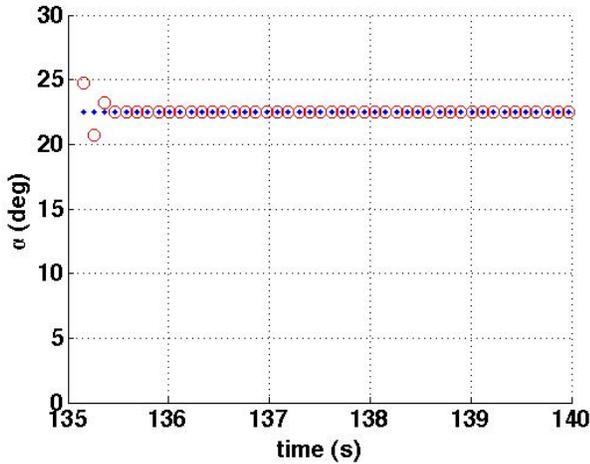


Fig. 9. Estimated α in the experiment. Blue dots indicate ground true values while red discs indicate the estimated values.

VI. CONCLUSION

In this paper we considered a vehicle which accomplishes a 3D-trajectory in the surrounding of a planar surface. The vehicle is equipped with a monocular camera and inertial sensors. Additionally, it is equipped with a laser pointer which produces a laser spot on the planar surface. This laser spot is observed by the monocular camera and it is the unique point feature used by the proposed approach. The paper provided two main contributions. The former is the analytical derivation of all the observable modes, i.e. all the physical quantities that can be determined by only using the inertial data and the camera observation of the laser spot during a short time-interval. Specifically, it is shown that the observable modes are: the distance of the vehicle from the planar surface; the component of the vehicle speed, which is

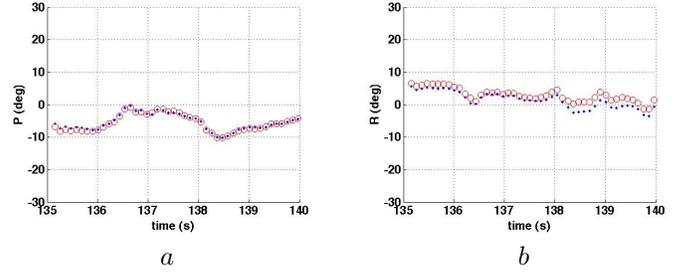


Fig. 10. Estimated P (a) and R (b) in the experiment. Blue dots indicate ground true values while red discs indicate the estimated values.

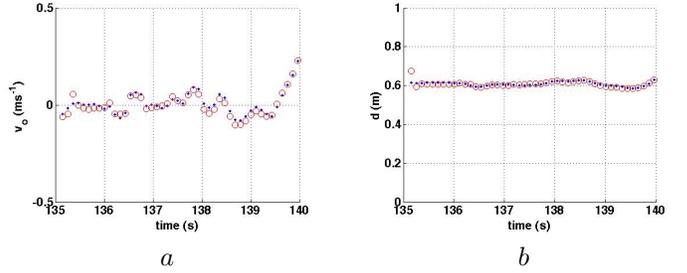


Fig. 11. Estimated v_o (a) and d (b) in the experiment. Blue dots indicate ground true values while red discs indicate the estimated values.

orthogonal to the planar surface; the relative orientation of the vehicle with respect to the planar surface; the orientation of the planar surface with respect to the gravity. Once the observed modes have been derived, a local decomposition of the original system has also been provided. This decomposition separates the observable modes from the rest of the system. This decomposition allowed us to introduce a simple recursive method to perform the estimation of all the observable modes (second contribution). The method was validated by using both synthetic and real data. In the paper the method has been implemented and used in the framework of aerial navigation. However, we want to emphasize that both the paper contributions are very general and can be applied in other frameworks. In particular, in all the environment where GPS is denied and where the most of objects have planar surfaces (e.g. in an indoor or city-like environment). For instance, these paper contributions could be used in the framework of humanoid robotics (where visual and inertial sensing are often adopted and the navigation usually occurs in an indoor environment).

An important aspect to be considered in future works is the extrinsic camera-laser pointer calibration. We assumed that the laser is aligned with the camera and that the position of the laser in the camera frame is $[L, 0, 0]$. An extrinsic calibration must estimate the angle between the camera frame and the orientation of the laser pointer together with the position of the laser pointer in the camera frame. This calibration requires to use both the two bearing angles of the laser spot in the camera frame. In the paper, one of them was always zero because of the assumed perfect alignment and because the y -coordinate of the laser pointer in the camera

frame was exactly zero.

Finally, a very important extension of the proposed strategy consists in considering other kind of laser spots. So far, the simplest case of a point-spot was considered. Other kind of shapes will be considered in order to improve the precision.

APPENDIX I CAMERA-LASER MODULE CALIBRATION

In figure 12, we display the position and the direction of the laser pointer in the original camera frame. The calibration consists in estimating the four parameters L_x , L_y , θ , ϕ . In other words, it consists in estimating the line made by the laser beam in the original camera frame. This line is determined starting from the position of the laser spot in the original camera frame for at least two spots. To have an accurate estimate the two spots must be as far as possible one each other. The precision can be further improved by considering more than two spots and by finding the best line fit. In order to have the Cartesian coordinates of a single spot in the original camera frame it suffices to produce the spot on a chess. By using the Matlab calibration camera toolbox, it is possible to get the equation of the plane for the chess in the original camera frame and, by having the direction of the spot from the camera measurement, the 3D position is finally obtained.

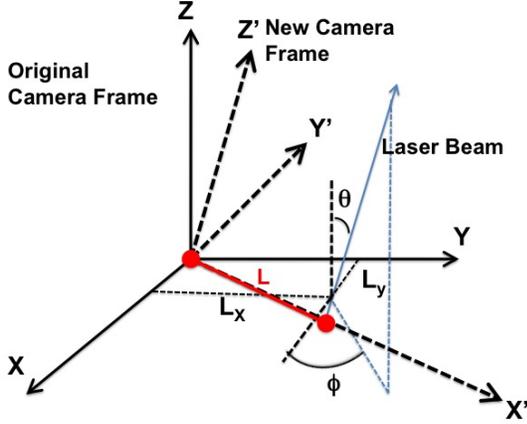


Fig. 12. The original camera frame XYZ , the new camera frame $X'Y'Z'$ and the laser module at the position $[L_x, L_y, 0]$ and the direction (θ, ϕ) (in the original frame) and position $[L, 0, 0]$ and the direction $(0, 0)$ (in the new frame).

The new camera frame is obtained by rotating the original frame such that in the new frame the z -axis has the same orientation of the laser beam. Additionally, we also require that the laser beam intersects the new x -axis. In other words, we require that the laser beam intersects the new xy -plane in the point $[L, 0, 0]^T$, for a given L . We want to obtain the quaternion q which characterizes this rotation. This will allow us to express the vectors provided by the camera in the new frame. Note that, since the two frames

only differ by a rotation (i.e., they share the same origin), we are allowed to express the vectors provided by the camera in the new frame, even if these vectors are defined up to a scale. Finally, in this section we want to determine the value of L . As we will see, both the quaternion q and the parameter L only depend on the calibration parameters: L_x , L_y , θ , ϕ .

We start by rotating the original frame of ϕ about its z -axis. The quaternion characterizing this rotation is $q_{z-axis}(\phi) = \cos\left(\frac{\phi}{2}\right) + k \sin\left(\frac{\phi}{2}\right)$. Then, we rotate the frame obtained with this rotation of θ about its y -axis. The quaternion characterizing this rotation is $q_{y-axis}(\theta) = \cos\left(\frac{\theta}{2}\right) + j \sin\left(\frac{\theta}{2}\right)$. Hence, the previous two rotations are characterized by the quaternion $q_{zy} \equiv q_{z-axis}(\phi)q_{y-axis}(\theta)$. The obtained frame has the z -axis aligned with the laser beam. On the other hand, the laser beam does not intersect necessarily the x -axis. In order to obtain this, we have to rotate again the frame about its current z -axis. Let us compute the intersection of the laser beam with the xy -plane. We compute this point in the original frame. By a direct computation we obtain: $r^{inters} = [L_x - v_x^2 L_x - v_x v_y L_y, L_y - v_x v_y L_x - v_y^2 L_y, -v_x v_z L_x - v_y v_z L_y]^T$. We then compute this vector in the rotated frame by doing the quaternion product: $R_q^{inters} = q_{zy}^* r^{inters} q_{zy}$. Let us denote R^{inters} with $[L', L', 0]^T$. We finally have: $L = \sqrt{L_x'^2 + L_y'^2}$ and $q = q_{zy} q_{z-axis}(\alpha)$, where $q_{z-axis}(\alpha) = \cos\left(\frac{\alpha}{2}\right) + j \sin\left(\frac{\alpha}{2}\right)$ and $\alpha = \arctan\left(\frac{L_y'}{L_x'}\right)$.

APPENDIX II CONTINUOUS SYMMETRIES AND OBSERVABLE MODES FOR THE SYSTEM DEFINED IN SECTION II

The system is characterized by the state: $\mathcal{S} = [r_q \ v_q \ q, k]^T$, whose dimension is 11 (r_q and v_q are imaginary quaternions, i.e. they are characterized by 3 real numbers; q contains 4 real numbers; k is a real number). The dynamics are given in (1) together with the equation $\dot{k} = 0$ and the observations are given in (3) and (4). In order to compute the Lie derivatives, we need to express the dynamics as follows:

$$\dot{\mathcal{S}} = \mathbf{f}(\mathcal{S}, \mathbf{u}) = \mathbf{f}_0(\mathcal{S}) + \sum_{i=1}^L \mathbf{f}_i(\mathcal{S}) u_i \quad (10)$$

We have $L = 6$ and the six inputs are the three components of the acceleration, \mathbf{A} , and the three components of the angular speed, $\mathbf{\Omega}$. Namely: $u_1 = A_x$, $u_2 = A_y$, $u_3 = A_z$, $u_4 = \Omega_x$, $u_5 = \Omega_y$, $u_6 = \Omega_z$. The seven vector functions $\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_6$ are:

$$\begin{aligned} \mathbf{f}_0 &= [v_x, v_y, v_z, 0, 0, -g, \mathbf{0}_5]^T \\ \mathbf{f}_1 &= [\mathbf{0}_3, q_t^2 + q_x^2 - q_y^2 - q_z^2, 2q_t q_z + 2q_y q_x, -2q_t q_y + 2q_z q_x, \mathbf{0}_5]^T \\ \mathbf{f}_2 &= [\mathbf{0}_3, -2q_t q_z + 2q_y q_x, q_t^2 + q_y^2 - q_z^2 - q_x^2, 2q_t q_x + 2q_z q_y, \mathbf{0}_5]^T \\ \mathbf{f}_3 &= [\mathbf{0}_3, 2q_t q_y + 2q_z q_x, -2q_t q_x + 2q_z q_y, q_t^2 + q_z^2 - q_x^2 - q_y^2, \mathbf{0}_5]^T \\ \mathbf{f}_4 &= [\mathbf{0}_6, -1/2q_x, 1/2q_t, 1/2q_z, -1/2q_y, 0]^T \end{aligned}$$

$$\mathbf{f}_5 = [\mathbf{0}_6, -1/2q_y, -1/2q_z, 1/2q_t, 1/2q_x, 0]^T$$

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

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 the two observation functions given in (3) and (4) with respect to all the vector fields. By a direct computation, performed by using the symbolic Matlab computational tool, we were able to find not more than 6 independent Lie derivatives².

In particular, according to the notation introduced in [16], the system has 5 continuous symmetries which are:

$$\mathbf{w}_s^1 = [0, 1, k, \mathbf{0}_8]^T$$

$$\mathbf{w}_s^2 = [1, \mathbf{0}_{10}]^T$$

$$\mathbf{w}_s^3 = [\mathbf{0}_4, 1, k, \mathbf{0}_5]^T$$

$$\mathbf{w}_s^4 = [\mathbf{0}_3, 1, \mathbf{0}_7]^T$$

$$\mathbf{w}_s^5 = [\mathbf{0}_6, q_z - kq_y, q_y + kq_z, -q_x + kq_t, -q_t - kq_x, 0]^T$$

The observable modes are all the solutions of the system of partial differential equations associated with the five symmetries. For instance, the equation associated with w_s^1 is $\frac{\partial}{\partial r_y} + k \frac{\partial}{\partial r_z} = 0$ (see [16] for more details). Since this system of partial differential equations consists of 5 equations on a manifold whose dimension is 11, the number of independent solutions is $6 = 11 - 5$ [14]. A possible choice of these solutions is:

$$r_z - kr_y$$

$$v_z - kv_y$$

$$2[c_k s_k (-q_t^2 + q_x^2 - q_y^2 + q_z^2) + (2c_k^2 - 1)(q_t q_x + q_y q_z)]$$

$$4c_k s_k (q_t q_z + q_x q_y) + 2(2c_k^2 - 1)(q_t q_y - q_x q_z)$$

$$k$$

$$q^* q$$

where $c_k \equiv \cos\left(\frac{\arctan(k)}{2}\right)$ and $s_k \equiv \sin\left(\frac{\arctan(k)}{2}\right)$. By knowing the value of the first solution, $r_z - kr_y$ and the value of the fifth solution, k , we can determine the quantity $\frac{|r_z - kr_y|}{\sqrt{1+k^2}}$, which is the distance of the vehicle from the planar surface³. Hence, to better visualize the physical meaning, it is convenient to select the following 6 observable modes:

²A possible choice of 6 independent Lie derivatives is: $L^0 h_{const}$, $L^0 h_{cam}$, $L^1_{f_0} h_{cam}$, $L^1_{f_4} h_{cam}$, $L^1_{f_5} h_{cam}$, $L^2_{f_0, f_0} h_{cam}$.

³In other words, also $\frac{r_z - kr_y}{\sqrt{1+k^2}}$ is a solution of the system of partial differential equations.

$$m_1 = \frac{r_z - kr_y}{\sqrt{1+k^2}}$$

$$m_2 = \frac{v_z - kv_y}{\sqrt{1+k^2}}$$

$$m_3 = 2[c_k s_k (-q_t^2 + q_x^2 - q_y^2 + q_z^2) + (2c_k^2 - 1)(q_t q_x + q_y q_z)]$$

$$m_4 = 4c_k s_k (q_t q_z + q_x q_y) + 2(2c_k^2 - 1)(q_t q_y - q_x q_z)$$

$$m_5 = k$$

$$m_6 = q^* q$$

Finally, the physical meaning of also m_3 and m_4 becomes clear by referring to a new global frame \tilde{x} , \tilde{y} , \tilde{z} . This frame has the $\tilde{x}\tilde{y}$ -plane coincident with the planar surface. In other words, this new global frame has the vertical axis coincident with the axis orthogonal to the planar surface. In this new frame, m_1 is the \tilde{z} -coordinate of the vehicle, m_2 is the component of the vehicle speed along the \tilde{z} -axis. m_3 and m_4 are related to the roll and pitch angles of the vehicle in the new frame. In particular, the roll angle is $\arctan\left(\frac{m_3}{\sqrt{1-m_3^2-m_4^2}}\right)$ and the pitch is $\arcsin(m_4)$. m_5 is related to the orientation of the $\tilde{x}\tilde{y}$ -plane with respect to the gravity. In particular, the \tilde{z} -axis makes an angle $\arctan(k) = \arctan(m_5)$ with the gravity. m_6 is trivially the magnitude of the quaternion, which is 1 since it describes a rotation.

From now on, we will adopt the new frame to characterize the vehicle configuration. The state in this frame is $\tilde{\mathbf{S}} = [\tilde{r}_q \tilde{v}_q \tilde{q}, k]^T$. The \tilde{x} , \tilde{y} , \tilde{z} -frame is obtained by rotating the x, y, z -frame about the x -axis of the angle $\arctan(k)$. Hence, it is characterized by the quaternion:

$$p = \cos\left(\frac{\arctan(k)}{2}\right) + i \sin\left(\frac{\arctan(k)}{2}\right) \quad (11)$$

Therefore, $q = p\tilde{q}$ or:

$$\tilde{q} = p^* q \quad (12)$$

By using the quaternion p it is also possible to obtain:

$$\tilde{r}_q = p^* r_q p \quad \tilde{v}_q = p^* v_q p \quad (13)$$

By using (12) and (13), we obtain the expressions of the observable modes in the new coordinates. We have:

$$m_1 = \tilde{r}_z$$

$$m_2 = \tilde{v}_z$$

$$m_3 = 2(\tilde{q}_t \tilde{q}_x + \tilde{q}_y \tilde{q}_z)$$

$$m_4 = 2(\tilde{q}_t \tilde{q}_y - \tilde{q}_x \tilde{q}_z)$$

$$m_5 = k$$

$$m_6 = \tilde{q}^* \tilde{q}$$

APPENDIX III EKF-EQUATIONS

Let us consider the state $\mathbf{m} = [m_1, m_2, m_3, m_4, g_z]$. The basic ingredients to implement an *EKF*, which estimates \mathbf{m} , are the analytical expression of the Jacobians of the dynamics, and the observation [2].

The Jacobian of the observation is obtained by differentiating the expression of h_{cam} given in (9) with respect to \mathbf{m} , i.e.

$$H \equiv \frac{\partial h_{cam}}{\partial \mathbf{m}} = \frac{L\xi(m_3, m_4)}{(m_4L - m_1)^2} \times \left[\begin{array}{ccccc} 1 & 0 & \frac{-m_3(m_4L - m_1)}{1 - m_3^2 - m_4^2} & \frac{Lm_3^2 - L + m_1m_4}{1 - m_3^2 - m_4^2} & 0 \end{array} \right]$$

where the function $\xi(m_3, m_4)$ is defined in (8). Regarding the Jacobian of the dynamics, we need first of all to discretize the equations in (7). Let us denote with δt the discretization time step. The Jacobian of the dynamics with respect to the state \mathbf{m} is:

$$F_m = \begin{bmatrix} 1 & \delta t & 0 & 0 & 0 \\ 0 & 0 & \delta t(A_y - A_z\bar{m}_3) & -\delta t(A_x + A_z\bar{m}_4) & 0 \\ 0 & 0 & -\delta t\Omega_x\bar{m}_3 & \delta t(\Omega_z - \Omega_x\bar{m}_4) & 0 \\ 0 & 0 & -\delta t(\Omega_z + \Omega_y\bar{m}_3) & -\delta t\Omega_y\bar{m}_4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where $\bar{m}_3 = \frac{m_3}{\xi(m_3, m_4)}$ and $\bar{m}_4 = \frac{m_4}{\xi(m_3, m_4)}$. The Jacobian of the dynamics with respect to the input $\mathbf{u} = [A_x, A_y, A_z, \Omega_x, \Omega_y, \Omega_z]^T$ is:

$$F_u = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -m_4 & m_3 & \xi & 0 & 0 & 0 \\ 0 & 0 & 0 & \xi & 0 & m_4 \\ 0 & 0 & 0 & 0 & \xi & -m_3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $\xi \equiv \xi(m_3, m_4)$.

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] Bar-Shalom Y., *Tracking and data association*, Academic Press Professional, Inc. San Diego, CA, USA, 1987
- [3] Bloesch, 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] Brogan W. L. *Modern Control Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1991
- [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] Isidori A., *Nonlinear Control Systems*, 3rd ed., Springer Verlag, 1995.
- [9] 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.
- [10] 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
- [11] J. Kelly and G. Sukhatme, Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration, *Int. Journal of Robotics Research*, Oct. 2010
- [12] Kim, J. and Sukkarieh, S. Real-time implementation of airborne inertial-SLAM, *Robotics and Autonomous Systems*, 2007, 55, 62-71
- [13] Quaternions and rotation Sequences: a Primer with Applications to Orbits, Aerospace, and Virtual Reality. Kuipers, Jack B., Princeton University Press copyright 1999.
- [14] F. John, *Partial Differential Equations*, Springer-Verlag, 1982.
- [15] 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
- [16] 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
- [17] A. Martinelli, Closed-Form Solution for Attitude and Speed Determination by Fusing Monocular Vision and Inertial Sensor Measurements, *International Conference on Robotics and Automation, ICRA 2011*, Shanghai, China
- [18] A. Martinelli, Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination, *Transaction on Robotics*, to appear on Volume 27 (2011), Issue 6 (December). INRIA tech rep, <http://hal.archives-ouvertes.fr/inria-00569083>.
- [19] 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
- [20] M. Veth, and J. Raquet, Fusing low-cost image and inertial sensors for passive navigation, *Journal of the Institute of Navigation*, 54(1), 2007