# Vision-Aided Inertial Navigation Using Virtual Features

Chiara Troiani, Agostino Martinelli

# Vision-Aided Inertial Navigation Using Virtual Features

Chiara Troiani and Agostino Martinelli

*Abstract*— In this paper we consider an aerial vehicle equipped with a monocular camera and inertial sensors. Additionally, a laser pointer is mounted on the vehicle and it produces a laser spot. The laser spot is observed by the monocular camera and it is the unique point feature used in the proposed approach. We focus our attention to the case when the vehicle moves in proximity of a planar surface and in particular when the laser spot belongs to this surface. 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 synthetic data. Additionally, preliminary tests with real data are provided and more complete experiments are in progress. The presented approach can be integrated in the framework of autonomous take-off and landing, safe touch-down and low altitude manoeuvres even in dark or featureless environment.

## 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 [21] an $EKF$ has been adopted. In this case, the proposed algorithm estimates a state containing the robot speed, position and attitude, together with the inertial sensor biases and the location of the features of interest. In the framework of airbone SLAM, an $EKF$ has been adopted in [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]. Natraj et al. [23] proposed a vision based approach, close to structured light, for roll, pitch and altitude estimation of UAV. They use a fisheye camera and a laser circle projector, assuming that the projected circle belongs to a planar surface. The latter must be orthogonal to the gravity vector in order to allow the estimation of the aforementioned quantities. The attitude estimation of the planar surface becomes crucial in order to extend the operational environment of UAVs. Shipboard operations, search and rescue cooperation between ground and aerial robots, low altitude manoeuvres, require to attenuate the position error and to track the platform attitude.

Recent works investigate the observability properties of the vision-aided inertial navigation system [9], [10], [11], [18], [19] and [20]. In particular, in [19], 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. The results provided in section III address precisely this issue (more details about the analytical computation can be found in [24]). 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 and 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 synthetic data (section V). Preliminary tests with real data are also provided and more complete experiments are in progress.

## 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 the surrounding 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 ($A$) is not simply the vehicle acceleration ($A_v$). It also contains the gravity acceleration ($G$). In particular, we have $A = A_v - G$ since, when the camera does not accelerate (i.e. $A_v$ 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: $g = [0, \ 0, \ -g]^T$, being $g \simeq 9.8 \ ms^{-2}$.

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 $[r, \ v, \ q]^T$ where $r = [r_x, \ r_y, \ r_z]^T$ is the $3D$ vehicle position, $v$ is its
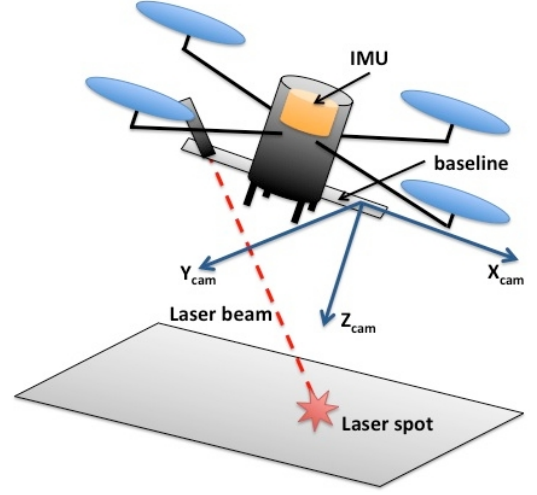


Fig. 1. Quadrotor equipped with a monocular camera, $IMU$ and a laser pointer. The laser spot is on a planar surface and its position in the camera frame is obtained by the camera up to a scale factor.

time derivative, i.e. the vehicle speed in the global frame ($v \equiv \frac{dr}{dt}$), $q = q_t + iq_x + jq_y + kq_z$ 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 $w = [w_x, \ w_y, \ w_z]^T$ we associate with it the imaginary quaternion $w_q \equiv 0 + iw_x + jw_y + kw_z$. The dynamics of the state $[r_q, \ v_q, \ q]^T$ are:

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

being $q^*$ the conjugate of $q$, $q^* = q_t - iq_x - jq_y - kq_z$ and $\Omega$ 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 (see figure 2). In addition, the camera frame is such that the laser beam intercept the $xy$−plane in $[L, \ 0, \ 0]$. In [24] 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 $[X_s, \ Y_s, \ Z_s]$ of the laser spot in the camera reference frame. We have:
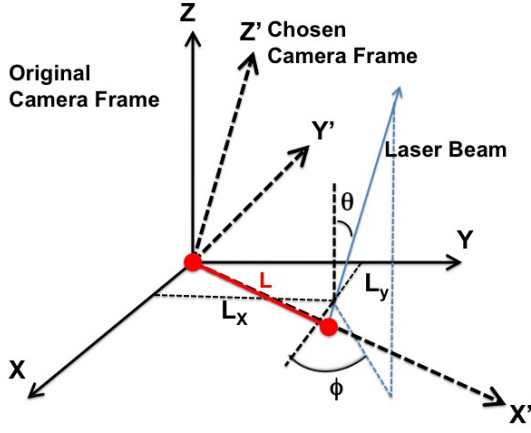
Fig. 2. The original camera frame $XYZ$, the chosen camera frame $X'Y'Z'$ and the laser module at the position $[L_x, \ L_y, \ 0]$ and the direction $(\theta, \ \phi)$ (in the original frame) and position $[L, \ 0, \ 0]$ and the direction $(0, \ 0)$ (in the chosen camera frame).

$$\begin{cases} X_s = L \\ Y_s = 0 \\ Z_s = \dfrac{r_z + 2q_z q_x L - 2q_y k q_x L - 2q_y q_t L - 2q_t L k q_z - k r_y}{2k q_z q_y - 2k q_t q_x - q_z^2 - q_t^2 + q_y^2 + q_x^2} \end{cases}$$
$$(2)$$

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

$$h_{cam} = \frac{L(2k q_z q_y - 2k q_t q_x - q_z^2 - q_t^2 + q_y^2 + q_x^2)}{r_z + 2q_z q_x L - 2q_y k q_x L - 2q_y q_t L - 2q_t L k q_z - k r_y}$$
$$(3)$$

### 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):

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

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 $\boldsymbol{S} = [r_q, \ v_q, \ q, k]^T$, whose dynamics are given in (1) and by the additional equation $\dot{k} = 0$.

We apply the method introduced in [17]. 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 [24] we perform this analytical computation and we show that the system has

six observable modes. Additionally, in [24], 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 $\tilde{x}, \ \tilde{y}, \ \tilde{z}$-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:

$$p = \cos\left(\frac{\arctan(k)}{2}\right) + i \sin\left(\frac{\arctan(k)}{2}\right) \qquad (5)$$

The state in this frame is $\tilde{\boldsymbol{S}} = [\tilde{r}_q \ \tilde{v}_q \ \tilde{q}, \ k]^T$, where:

$$\tilde{q} = p^* q \quad \tilde{r}_q = p^* r_q p \quad \tilde{v}_q = p^* v_q p \qquad (6)$$

The observable modes are (see the derivation in [24]):

$$\begin{aligned} m_1 &= \tilde{r}_z \\ m_2 &= \tilde{v}_z \\ m_3 &= 2(\tilde{q}_t \tilde{q}_x + \tilde{q}_y \tilde{q}_z) \\ m_4 &= 2(\tilde{q}_t \tilde{q}_y - \tilde{q}_x \tilde{q}_z) \\ m_5 &= k \\ m_6 &= \tilde{q}^* \tilde{q} \end{aligned} \qquad (7)$$

$m_1$ 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 [24]) 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 $m_1, \ ..., \ m_6$. 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 in (7). 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 $\tilde{S} = [\tilde{r}_q \ \tilde{v}_q \ \tilde{q}, \ k]^T$. We obtain [24]:

$$
\begin{cases}
\dot{\tilde{r}}_q = \tilde{v}_q \\
\dot{\tilde{v}}_q = \tilde{q} A_q \tilde{q}^* + p^* g_q p \\
\dot{\tilde{q}} = \dfrac{1}{2} \tilde{q} \Omega_q \\
\dot{k} = 0
\end{cases}
\tag{8}
$$

A local decomposition for the dynamics is:

$$
\begin{cases}
\dot{m}_1 = m_2 \\
\dot{m}_2 = -m_4 A_x + m_3 A_y + \xi(m_3, \ m_4) A_z + g_z \\
\dot{m}_3 = \Omega_x \ \xi(m_3, \ m_4) + \Omega_z m_4 \\
\dot{m}_4 = \Omega_y \ \xi(m_3, \ m_4) - \Omega_z m_3 \\
\dot{g}_z = 0
\end{cases}
\tag{9}
$$

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 $\tilde{q}_x^2 + \tilde{q}_y^2 - \frac{1}{2}$:

$$
\xi(m_3, \ m_4) \equiv
\begin{cases}
\sqrt{1 - m_3^2 - m_4^2} & if \quad \tilde{q}_x^2 + \tilde{q}_y^2 < \dfrac{1}{2} \\
-\sqrt{1 - m_3^2 - m_4^2} & if \quad \tilde{q}_x^2 + \tilde{q}_y^2 > \dfrac{1}{2}
\end{cases}
\tag{10}
$$

Hence, as previously said, we have two local decompositions for our original system. The validity of (9) can be checked by using (7) and (8). Note that deriving (9) is troublesome. In contrast, checking its validity is very simple since it 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:

$$
h_{cam} = \frac{L \xi(m_3, \ m_4)}{m_4 L - m_1}
\tag{11}
$$

The validity of (11) can be checked by using (3), (5), (6), (7) and (10).

The equations (9) and (11) 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 (9) provides the link between the observable modes and the $IMU$ data. Equation (11) 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 (9) and the Jacobian of the observation function in (11) (see [2]). We provide these Jacobians in [24].

## 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 ($\alpha$).

Specifically, in all the simulations the values of the estimated $d$, $v_o$, $R$, $P$ and $\alpha$ 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 $\mu_a$ and $\mu_\omega$ and whose variances will be denoted respectively with $\sigma_a^2$ and $\sigma_\omega^2$. We set the parameters in order to be close to a real case: $\mu_a = 0 \ ms^{-2}$, $\sigma_a = 1 \ ms^{-2}$, $\mu_\omega = 0 \ deg \ s^{-1}$ and $\sigma_\omega = 10 \ deg \ s^{-1}$. The initial vehicle position is at $\tilde{x} = 0$, $\tilde{y} = 0$, $\tilde{z} = 1m$. The initial vehicle speed is $[1, \ 0, \ 0]ms^{-1}$ 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 $\Omega_i^{true}$ and $A_{v \ i}^{true}$). Starting from them, the IMU sensors are simulated by randomly generating the angular speed and the linear acceleration at each step according to the following: $\Omega_i = N\left(\Omega_i^{true} - \Omega_{bias}, P_{\Omega_i}\right)$ and $A_i = N\left(A_{v \ i}^{true} - A_{gi} - A_{bias}, P_{A_i}\right)$ where:

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

- $\boldsymbol{\Omega_{bias}}$ is the bias affecting the data from the gyroscope.

In all the simulations we set both the matrices $P_{\Omega_i}$ 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 $3 \times 3$ matrix. We considered several values for $\sigma_{gyro}$ and $\sigma_{acc}$, in particular: $\sigma_{gyro} = 1 \ deg \ s^{-1}$ and $\sigma_{acc} = 0.01 \ ms^{-2}$.

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 computed[1]. They are computed each $0.1s$. The parameter $L$ is set equal to $0.3m$. Then, the camera readings are generated by adding to the true values zero-mean Gaussian errors whose variance is equal to $(1 \ deg)^2$ for all the readings.

*3) Simulation Results:* Figure 3 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 $\alpha = \frac{\pi}{8} \ rad = 22.5 \ deg$ with the gravity.



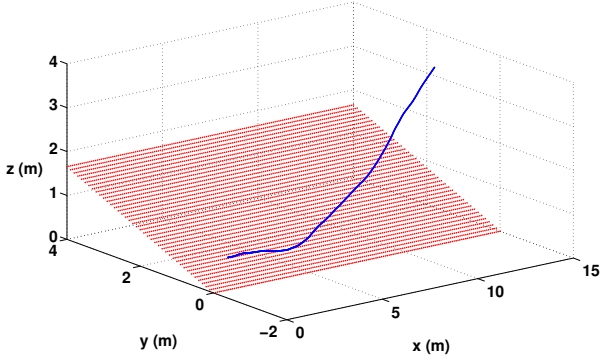Fig. 3.   A typical vehicle trajectory in our simulations.

Figures 4 $a$ and $b$ display the estimated $\alpha$ respectively in the case without and with bias. The values of the biases adopted in our simulations are: $\boldsymbol{\Omega_{bias}} = [0.03 \ 0.03 \ 0.03]^T (deg \ s^{-1})$ and $\boldsymbol{A_{bias}} = [0.03 \ 0.03 \ 0.03]^T (ms^{-2})$. As expected, the estimation in presence of bias becomes worse. However, the error on the estimated $\alpha$ in presence of bias is smaller than $1 \ deg$.
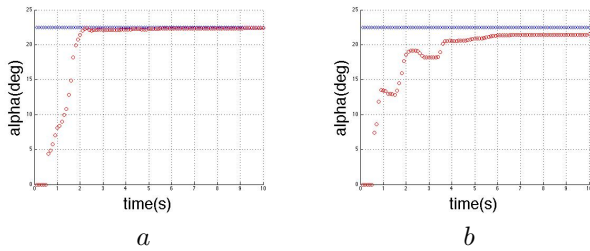


Fig. 4.   Estimated $\alpha$ in absence ($a$) and in presence of bias ($b$) on the inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.

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

Figures 5, 6, 7 and 8 $a$ and $b$ display respectively the estimated $P$, $R$, $v_o$ and $d$. In each figure, both the cases of unbiased and biased inertial measurements are displayed. We initialized the filter by using a value of the initial observable state which differs from the ground truth by a relative error in the range $[10, \ 20]\%$.
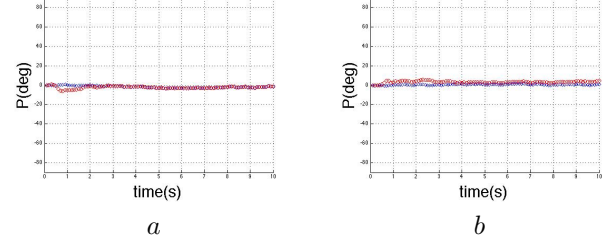


Fig. 5.   Estimated $P$ in absence ($a$) and in presence of bias ($b$) on the inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.
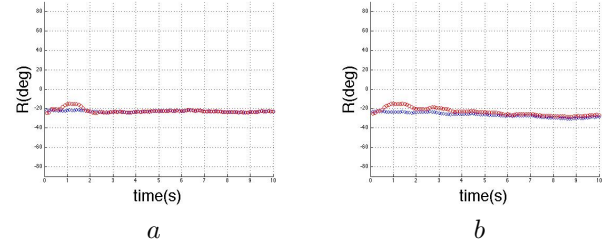


Fig. 6.   Estimated $R$ in absence ($a$) and in presence of bias ($b$) on the inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.
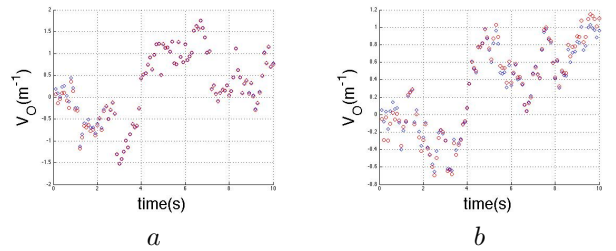


Fig. 7.   Estimated $v_0$ in absence ($a$) and in presence of bias ($b$) on the inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.

We also evaluated the robustness of the filter with respect to systematic errors on the imu-camera calibration and laser-camera calibration. Specifically, we performed simulations by introducing errors of one $cm$ and one $deg$ on the calibration parameters. These systematic errors affect the estimated $\alpha$ (the difference with respect to the ground truth is in the range $[4, \ 6]deg$) while for all the other observable modes the effect is negligible (less than one $deg$ for $R$ and $P$ and less than $1cms^{-1}$ and $1cm$ respectively for $v_o$ and $d$).

## B. Preliminary experiments

In this section we provide preliminary results obtained by using a data set provided by the autonomous system
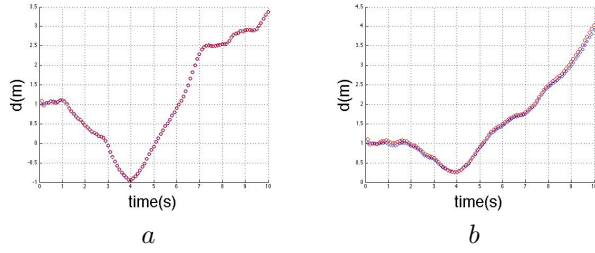
Fig. 8. Estimated $d$ in absence ($a$) and in presence of bias ($b$) on the inertial data. Blue dots indicate ground true values while red discs indicate the estimated values.



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.

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 [16], 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 $\alpha$. 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 that the convergence of the filter occurs in less than half second for all the observable modes.
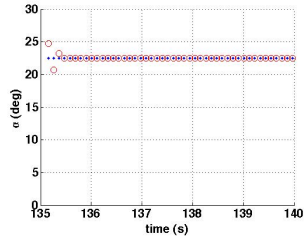


Fig. 9. Estimated $\alpha$ in the experiment. Blue dots indicate ground true values while red discs indicate the estimated values.
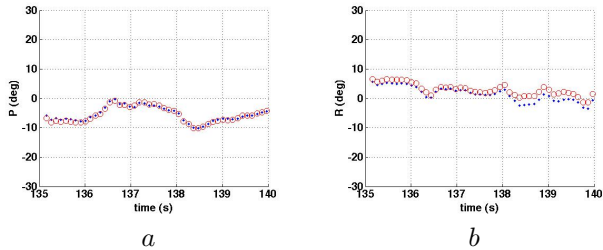


Fig. 10. Estimated $P$ ($a$) and $R$ ($b$) in the experiment. Blue dots indicate ground true values while red discs indicate the estimated values.
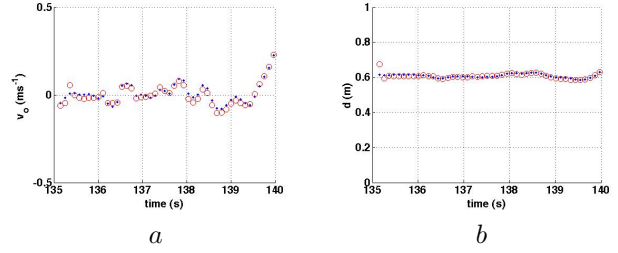
## C. Experiments in progress

In this last section we describe our vehicle and sensors. The robot platform is a *Pelican* from *Ascending Technologies* equipped with an Intel Atom processor board (*1.6 GHz, 1 GB RAM*). Our sensor suite is composed by an Inertial Measurement Unit (*3-Axis Gyro, 3-Axis Accelerometer*), a monocular camera (*Matrix Vision mvBlueFOX, FOV*: 130 $deg$) and a Laser Module (*SparkFun TTL, wavelength: 650nm, poweroutput: 0.45-0.8mW*). The Laser module and the monocular camera are mounted on a fixed baseline, and the latter is calibrated using the Camera Calibration Toolbox for Matlab [22]. The calibration between IMU and camera has been performed using the Inertial Measurement Unit and Camera Calibration Toolbox by Lobo [15], whereas the Camera-Laser Module calibration technique we used is described in [24]. By using this latter technique we obtained the following values: $\theta = 47.1\ deg$, $\phi = -3.1\ deg$ and $L_x = -0.146\ m$, $L_y = -0.005\ m$. The resulting $L$ is $0.100\ m$.

The IMU provides measurements update at a rate of $100Hz$, while the camera framerate is $13Hz$. We are currently equipping our lab with a motion capture system to perform experiments.
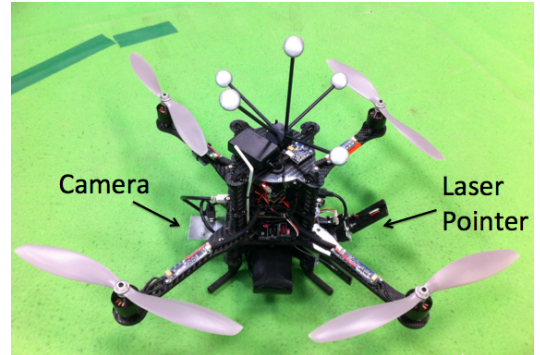


Fig. 12. The Pelican quadcopter equipped with a monocular camera, a laser module and passive markers which will be used to collect sensors data and ground truth.

## VI. CONCLUSION

In this paper we considered an aerial vehicle equipped with a monocular camera and inertial sensors. Additionally, a laser pointer is mounted on the vehicle and it produces a laser spot.

The laser spot is observed by the monocular camera and it is the unique point feature used in the proposed approach. We focused our attention to the case when the vehicle moves in proximity of a planar surface and in particular when the laser spot belongs to this surface. 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 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 and allowed us to introduce a simple recursive method to perform the estimation of all the observable modes (second contribution). The use of a virtual laser spot feature overcomes the limits of feature tracking algorithms and makes our approach suitable to work even in dark or featureless environment. The method is validated by using synthetic data. The validation with real data is in progress. We presented a low-cost low-weight sensor suite and a low computational complexity approach in the framework of aerial navigation. It can be integrated in the framework of autonomous takeoff and landing, safe touch-down and low altitude manoeuvres. 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).

A very important extension of the proposed strategy is to consider laser patterns in order to improve the precision. So far, the simplest case of a point-spot was considered. The performance of the considered system depends on the extrinsic parameters of the laser module with respect to the distance from the planar surface. Therefore we want to consider the option of a laser module with a servomotor in order to improve precision according to the trajectory. We finally want to implement this approach on our quadrotor platform to perform autonomous landing and safe touchdown tasks.

## 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] 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] 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. Vinzce, P. Corke, and J. Lobo, Editorial: Special Issue: 2nd Workshop on Integration of Vision and Inertial Sensors, The International Journal of Robotics Research, June 2007; vol. 26, 6.

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

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

[17] 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

[18] 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

[19] A. Martinelli, Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale and Bias Determination, Transactions on Robotics, Vol. 28, No. 1, pp 44–60, February 2012

[20] 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

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

[22] J.Y. Bouguet, Camera Calibration toolbox for Matlab. www.vision.caltech.edu/bouguetj.

[23] A. Natraj, C. Demonceaux, P. Vasseur, P. Sturm, Vision based attitude and altitude estimation for UAVs in dark environments, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2011

[24] A. Martinelli and Chiara Troiani, Vision-Aided Inertial Navigation Using Virtual Features, INRIA tech rep, http://hal.inria.fr/hal-00679529