Abstract—The aim of this paper is to present a method for integration of measurements provided by inertial sensors (gyroscopes and accelerometers), GPS and a video system in order to estimate position and attitude of an UAV (Unmanned Aerial Vehicle). Inertial sensors are widely used for aircraft navigation because they represent a low cost and compact solution, but their measurements suffer of several errors which cause a rapid divergence of position and attitude estimates. To avoid divergence inertial sensors are usually coupled with other systems as for example GNSS (Global Navigation Satellite System). In this paper it is examined the possibility to couple the inertial sensors also with a camera. A camera is generally installed on-board UAVs for surveillance purposes, it presents several advantages with respect to GNSS as for example great accuracy and higher data rate. Moreover, it can be used in urban area or, more in general, where multipath effects can forbid the application of GNSS. A camera, coupled with a video processing system, can provide attitude and position (up to a scale factor), but it has lower data rate than inertial sensors and its measurements have latencies which can prejudice the performances and the effectiveness of the flight control system. The integration of inertial sensors with a camera allows exploiting the better features of both the systems, providing better performances in position and attitude estimation.

I. INTRODUCTION

The integration of measurements provided by inertial sensors (gyroscopes and accelerometers) and video systems (VS) has gained an increasing popularity in the last decade [1]. These kind of sensors bring complementary characteristics. Indeed inertial sensors cannot discern a change in inclination from the body acceleration, due to Einstein’s equivalence principle. Moreover, inertial sensors present high uncertainty at low velocities and low relative errors at high velocities while camera motion estimation becomes less reliable at high velocity because of the low camera acquisition framerate. Furthermore, the camera estimates the motion up to a scale factor, i.e., a near object with low relative speed appears the same as a far object with high relative speed.

Based on these observations, the motivation for integration of vision and inertial sensing is clear. There are two broad approaches usually called loosely and tightly coupled. The loosely coupled approach uses separate INS and VS blocks, running at different rates and exchanging information. The tightly-coupled systems combine the disparate raw data of vision and inertial sensors in a single, optimum filter, rather than cascading two filters, one for each sensor. Typically, in order to fusion the data a Kalman Filter is used. The integration of visual and inertial sensing mode opens up new directions for the application in the field of autonomous navigation, robotics and many other fields.

The use of inertial sensors in machine vision applications has been proposed by now more than twenty years ago and further studies have investigated the cooperation between inertial and visual systems in autonomous navigation of mobile robots, in image correction and to improve motion estimation for 3D reconstruction of structures. More recently, a framework for cooperation between vision sensors and inertial sensors has been proposed. The use of gravity as a vertical reference allows the calibration of focal length camera with a single vanishing point, the segmentation of the vertical and the horizontal plane. In [2] is presented with a function for detecting the vertical (gravity) and 3D mapping, and in [3] such vertical inertial reference is used to improve the alignment and registration of the depth map.

Virtual reality applications have always required the use of motion sensors worn by the user. The augmented reality, where virtual reality is superimposed on a real time view, is particularly sensitive to any discrepancy between the estimated and the actual motion of the user. An accurate estimate of the user position can be obtained through the use of many sensors, using the external vision and specific markers such as radio transponder, ultrasound beams or lasers, etc. Having available MEMS inertial sensors at low cost, these are used in combination with computer vision techniques [4]. The objective is to have a visual-inertial tracker, able to operate in arbitrary environments relying only on natural observable characteristics, suitable for augmented reality applications. In [5] an extended Kalman filter is used to combine the low-frequency stability of vision sensors with the high frequency of the gyroscope, thereby reaching a stable tracking of the pose (6 degrees of freedom) both static and dynamic. The augmented reality systems are based on hybrid tracker to fuse the images successfully in real time with 3D models dynamic [4], [6], [7], [8].

Applications in robotics are increasing. Initial works on vision systems for automated passenger vehicles have also incorporated inertial sensors and have explored the benefits of visual-inertial tracking [9], [10]. Other applications include agricultural vehicles [11], wheelchairs [12] and robots in indoor environments [13], [14]. Other recent work related to Unmanned Aerial Vehicles (UAV) include fixed-wing air-
Using this mapping, we can express cross operation of two matrix. So vector operation such as cross product ($\times$) vector to a 3 by 3 matrix: \[ \sqrt{\sum_{i,j} a_{ij}^2} \] operations. Three-dimensional vectors. Matrix operations precede vector matrix operations such as matrix multiplication are applied to the field of view. In [26] the pitch and roll angles, and three body angular rates are determined by detection of the horizon in a sequence of images and by the analysis of the optical flow. The use of the horizon detection technique does not allow determining the value of the yaw angle, whose knowledge is required, for example, in trajectory tracking algorithms. Moreover, for some configurations of the video system or particular maneuvers, the horizon could not be present in the images (for example cameras could frame only the sky if they are also used for Collision Avoidance). In [27] it is proposed another method, based on the detection of the horizon in an images sequence, to estimate roll and pitch angles. In [28] the aircraft complete attitude is determined using a single camera observations. Specifically, the proposed method required the knowledge of the positions of the own aircraft and of the points identified in the camera images. (a georeferenced map of the surrounding space is required). In [29] it is proposed another method to estimate both aircraft position and attitude using a single camera. In this application is required the presence of some reference points or some known geometries in the surrounding space too.

This paper is organized as follow. In Section II some notations are introduced. Section III provides a description of the camera system. Section IV and Section V deal with the estimation of camera egomotion and the Kalman Filter theory respectively. Experimental setup and numerical analysis are described in Section VI. Finally, conclusions are drawn in Section VII.

II. NOTATIONS

We first introduce some notation. Matrices are denoted by capital italics. Vectors are denoted by bold fonts either capital or small. A three-dimensional column vector is specified by ($s_1, s_2, s_3)^T$. A vector is sometimes regarded as a column matrix. So vector operation such as cross product ($\times$) and matrix operations such as matrix multiplication are applied to three-dimensional vectors. Matrix operations precede vector operations. 0 denotes a zero vector. For a matrix $A = [a_{ij}]$, $\|A\|$ denotes the Euclidean norm of the matrix, i.e., $\|a_{ij}\| = \sqrt{\sum_{ij} a_{ij}^2}$. We define a mapping $[\cdot]_\times$ from a three-dimensional vector to a 3 by 3 matrix:

$$
[(x_1, x_2, x_3)^T]_\times = 
\begin{bmatrix}
0 & -x_3 & 0 \\
-x_1 & 0 & -x_2 \\
x_2 & x_1 & 0 
\end{bmatrix}.
$$

(1)

Using this mapping, we can express cross operation of two vectors by the matrix multiplication of a 3 by 3 matrix and a column matrix:

$$
X \times Y = [X]_\times Y.
$$

(2)

The reference system $s$, in which the coordinates of the vector $x$ are expressed, is reported as superscript on the upper-left corner of $x$ with the notation $x^s$.

The reference systems considered in the paper are the following:

- $i$ inertial system;
- $e$ the ECEF (Earth Cenetered Earth Fixed) system;
- $n$ the NED (North East Down) system, tangent to the earth ellipsoid, at a reference Latitude and Longitude $(\text{Lat}_0, \text{Lon}_0)$;
- $b$ the "body" system as seen by the IMU (Inertial Measurement Unit);
- $c$ the camera system, with fixed orientation with respect the IMU.

III. CAMERA EGO-MOTION

A. Coordinate System

The default coordinate system in which the UAV position, i.e. $(\text{Lat}, \text{Lon}, \text{Alt})$ of the vehicle center of mass, is given is the ECEF system. $C_n^e$ is the reference change matrix from NED to ECEF. The attitude and heading of the vehicle is given as $C_n^b(\phi, \theta, \psi)$ where $C_n^b$ is the reference change matrix associated with the roll ($\phi$), pitch ($\theta$), and yaw ($\psi$) angles. In general, the camera standard reference system might be rotated with respect to the body reference system. Thus, the transformation from the camera reference to ECEF is given by

$$
C_c^e = C_n^e(\text{Lat}, \text{Lon}) C_n^b(\phi, \theta, \psi) C_c^b
$$

(3)

where $C_n^b$ represents the constant reference change matrix from camera system $c$ to body unit system $b$.

B. Essential Matrix

Suppose we have a calibrated camera, i.e. with known intrinsic parameters, moving in a static environment following an unknown trajectory. Consider two images taken by the camera at different time. The equation linking two image points $p'$ and $p$ which are the projection of the same 3D point is:

$$
p' E p = 0,
$$

(4)

where $p' = [u', v', 1]^T$ and $p = [u, v, 1]^T$ are the (normalized) homogeneous coordinates in the second and the first camera reference frame respectively. The matrix $E = [r_{21}^c]^T C_c^e$ is called the essential matrix and contains the information about the camera movement. The essential matrix depends on 3 parameters for the rotation and 2 for the translation (since eq. (4) does not depend on the modulus of $r_{21}^c$).

Here $[r_{c1}^e, r_{c2}^e]$ represents the rigid transformation (rotation and translation) which brings points (i.e. transforms coordinates) from the camera 1 to the camera 2 standard reference system. The coordinates change equation is given by

$$
x'^c = C_{c1}^e x^c + r_{21}^c
$$

(5)
where \( x'_{\text{st}} \) represent the coordinates in the camera \( k \) standard reference system.

The translation vector between the two camera center \( r_{21}^t = O_2^t - O_1^t \) expressed in a generic reference system \( s \) is equal to \(-C_2^r r_{21}^t\).

The matrix \( R \) contains in it different contributions:

1) the rotation of the camera by the pan-tilt unit;

2) changes in the vehicle attitude and heading;

3) rotation between the two NED systems (in \( t_2 \) and \( t_1 \)).

Indeed it can be also expressed as

\[
C_{c1}^{c2} = C_{c}^{c_2} C_{c_1}^{c} = C_{c_1}^{c_2} \quad (6)
\]

By means of eq. (3) and with a little algebra we obtain

\[
C_{c1}^{c2} = C_{c_2}^{c_b} C_{c_1}^{c_n} C_{c}^{c_{n_2}} C_{c_1}^{c_{n_1}} \quad (7)
\]

In the above equation we supposed (Lat, Lon) did not change a lot between two consecutive frames, hence the identity approximation. This also means that the two NED reference systems (i.e., corresponding to \( n_1 \) and \( n_2 \)) can be thought to be coincident up to a translation. The eq.(7) then reads as

\[
C_{c1}^{c2} \approx C_{c_2}^{c_b} C_{c_1}^{c_n} C_{c_n}^{c_{n_2}} C_{c_{n_1}}^{c_{n_1}} \quad (8)
\]

Therefore, supposing a fixed camera orientation with respect to the body, the functional dependency is only in the rotation of the body, i.e. in the change of the vehicle attitude and heading,

\[
\begin{split}
C_{b_1}^{b_2} &\approx C_{n_2}^{b_2} C_{n_1}^{b_1} \approx R(d\phi, d\theta, d\psi) = R_x(d\phi) R_y(d\theta) R_z(d\psi)
\end{split} \quad (9)
\]

where \( d\phi = \phi_2 - \phi_1 \), \( d\theta = \theta_2 - \theta_1 \) and \( d\psi = \psi_2 - \psi_1 \).

The expression of \( R \) is given in (10) on the top of the next page. It can be approximated for small angle variations by

\[
R(d\phi, d\theta, d\psi) \approx \begin{pmatrix}
1 & d\psi & -d\theta \\
-d\psi & 1 & d\phi \\
d\theta & -d\phi & 1
\end{pmatrix} = I - W dt \quad (11)
\]

where \( \omega = [d\phi, d\theta, d\psi]^T \) / \( dt \) is the angular velocity (in the body reference system) and \( W = [\omega]^x \).

Once estimated \( R_{21}^t \) and \( C_{c1}^{c2} \) from the essential matrix \( E \), they give information on \( v(t) \) and \( \omega(t) \) respectively. Such an information can be used either as a prediction or as a measurement for the Kalman Filter. Similarly, the IMU output can be used as \textit{a priori} information for the vision algorithms.

IV. COMPUTATION OF \( E \) AND ERROR ANALYSIS

A. Fast computation of image point correspondences

To calculate the essential matrix by video analysis, the point correspondences between two successive frames can be calculated using a computer vision system based on the Lukas-Kanade method [30]. This method assumes that the displacements of the points belonging to the image plane is small and approximately constant. Thus the velocity of the points in the image plane must satisfy the Optical Flow equation

\[
I_u(p) \dot{u} + I_v(p) \dot{v} = -\dot{I}(p) \quad (12)
\]

where \( p = (u, v) \) is a point in the image plane, \( I_u, I_v \) are the components of the spatial gradient of the image intensity in the image plane and \( \dot{x} \) is the notation for the temporal derivative of \( x \).

The Lucas-Kanade method [30] can be used only when the image flow vector between the two frames is small enough for the differential equation of the optical flow to hold. When this condition does not hold, a rough preliminary prediction of the optical flow must be performed with other methods and the Lukas-Kanade method can be used for refinement only. When external sensors, like IMU or GNSS (Global Navigation Satellite System), are available, the camera motion is predictable, so that a rough estimation of the optical flow can be easily initialized starting from the sensor measurements [31]. To aim this task, a basic knowledge of the scene 3D geometry should be known or at least approximatively known. This is the case of the Earth surface that can be modeled as ellipsoid, geoid or digital elevation model. Using this prior information, the computational burden of the optical flow estimation can be strongly reduced, so that the delay of the video processing can be made compatible with the application of the proposed technology of sensor fusion to the navigation scenario.

To further reduce the computational burden of the computer vision algorithms an improvement of the Lukas-Kanade method, the Kanade-Lucas-Tomasi (KLT) feature matching algorithm [32], has been adopted. The KLT has been improved using the ellipsoid model of the Earth surface, increasing the speed of calculus of the Optical Flow.

B. Essential matrix calculation

Obtained a set of correspondences of image points for a couple of successive frames, our objective is to find the essential matrix \( E \). Each point correspondence gives one equation (4) which is linear and homogeneous in the elements of \( E \). \( n \) point correspondences give \( N \) such equations. Let

\[
E = \begin{bmatrix}
e_1 & e_4 & e_7 \\
e_2 & e_5 & e_8 \\
e_3 & e_6 & e_9
\end{bmatrix}, \quad c = (c_1, c_2, \ldots, c_9)^T. \quad (13)
\]

Rewriting equation (4) in the elements of \( E \) using \( k \) point correspondences, we have

\[
A e = 0. \quad (14)
\]

where \( A \) is

\[
A = \begin{bmatrix}
u_1 u_1' & u_1 v_1' & u_2 v_1' & v_1 & u_1' & v_1' & 1 \\
u_2 u_2' & u_2 v_2' & v_2 & v_2' & u_2' & v_2' & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
u_k u_k' & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}
\]

In presence of noise, \( A \) is a full rank matrix. Therefore we solve for unit vector \( h \) such that

\[
||Ah|| = \min. \quad (16)
\]
The solution of \( h \) is the unit eigenvector of \( A^T A \) associated with the smallest eigenvalue. Then \( E \) is determined by

\[
E = [E_1 E_2 E_3] = \sqrt{2} \begin{bmatrix}
    h_1 & h_4 & h_7 \\
    h_2 & h_5 & h_8 \\
    h_3 & h_6 & h_9
\end{bmatrix}.
\]

(17)

**C. Getting motion parameters from \( E \)**

Once got the essential matrix, the Huang and Faugeras theorem [33] allows to factorize \( E \) in rotation \( (C_{ci}^2) \) and translation \( (r_{21}^i) \). Let \( E = UDV^T \) be the singular value decomposition (SVD) of \( E \) and let

\[
S' = \begin{bmatrix}
    0 & 1 & 0 \\
    0 & 0 & 1
\end{bmatrix}
\quad \text{and} \quad
R' = \begin{bmatrix}
    0 & 1 & 0 \\
    -1 & 0 & 0 \\
    0 & 0 & 1
\end{bmatrix}
\]

Then, there are four possible factorizations for \( E = [r_{21}^i] \times C_{ci}^2 \) given by

\[
[r_{21}^i] = U(\pm S')U^T
\]

\[
C_{ci}^2 = \beta UR'V^T \quad \text{or} \quad C_{ci}^2 = \beta UR^TV^T
\]

(19) \quad (20)

where

\[
\beta = \det(UV^T).
\]

(21)

As observed in [34], the right factorization can be determined by imposing that 3D points (which can be found by triangulation), must lie always in front of the camera, *i.e.*, their \( z \)-coordinate must be positive.

**D. Error analysis**

In order to do an error analysis, we want to relate the errors in the correspondence matrix \( A \) to the solution of the minimization problem (16). The sources of errors in the image coordinates include spatial quantization, feature detector errors, point mis-matching and camera distortion. In a system that is well calibrated so that systematic errors are negligible, errors in the image coordinates of a feature can be modeled by random variables. These errors result in the errors of the estimates of the motion parameters.

The covariance matrix for \( e \) is

\[
\Sigma_e = 2\Sigma_h \simeq 2D_h \Sigma_{A^T} D_h^T,
\]

(22)

where \( \Sigma_{A^T} \) is the covariance matrix of \( A^T \) and

\[
D_h = G_h G_{A^T A}
\]

(23)

\[
G_{A^T A} = [F_{ij}] + [G_{ij}],
\]

(24)

with the \( i \)th column being the column vector \( A_j \) and all other columns being zeros. The matrix \( G_h \) is given by

\[
G_h = S\Delta S^T[h_1 I_3 h_2 I_3 \ldots h_9 I_3],
\]

(25)

where the columns of \( S \) represent the eigenvectors (ordered in non-decreasing order) of \( S^T S \), \( \Delta = \text{diag} \{0, (\lambda_1 - \lambda_2)^{-1}, \ldots, (\lambda_1 - \lambda_n)^{-1}\} \), \( \lambda \) being the eigenvalues of \( A^T A \), and \([h_1, \ldots, h_9]\) is the solution of the minimization problem (16).

Assume the errors between different points and different components in the image coordinates are uncorrelated, and they have the same variance \( \sigma^2 \). With this assumption we get

\[
\Sigma_{A^T} = \sigma^2 \text{diag} \{\Sigma_1, \Sigma_2, \ldots, \Sigma_n\},
\]

(26)

where \( \Sigma_i, 1 \leq i \leq n \), is a 9 by 9 submatrix:

\[
\Sigma_i = \begin{bmatrix}
    p_i^T \quad 0 & p_i^T \\
    0 & 0 & p_i^T \\
    0 & 0 & p_i^T
\end{bmatrix} + \begin{bmatrix}
    u_i u_i J & u_i v_i J & u_i J \\
    u_i v_i J & v_i v_i J & v_i J \\
    u_i J & v_i J & J
\end{bmatrix},
\]

(27)

where

\[
J = \begin{bmatrix}
    1 & 0 & 0 \\
    0 & 1 & 0 \\
    0 & 0 & 0
\end{bmatrix}.
\]

(28)

**V. KALMAN FILTER**

The final purpose of the proposed approach is the estimation of the position and attitude of the UAV. The data fusion algorithm is based on the Extended Kalman Filter (EKF), because dynamic and observation equations are non-linear in their original form.

The state vector includes the position and speed of the aircraft UAV in the tangent reference frame, and the rotation matrix from body to tangent reference frame. The EKF is applied to estimate the state variables. The input measurements for prediction of state variables are inertial accelerations and angular speeds processed by an IMU.

The fusion algorithm is composed by three EKF as sketched out in Figure 1. With reference to Figure 1:

- the EKF block GPS-IMU uses linear accelerations provided by the IMU and position and speed provided by GPS in order to compute position and speed of the aircraft center of gravity;
- the VIDEO-IMU(1) block is based on angular speed measurements provided by the IMU and attitude differences measured by the video-system in order to estimate aircraft attitude. Attitude estimates are provided to the other two EKF blocks;
the VIDEO-IMU (2) block takes the linear accelerations provided by the IMU as in GPS-IMU, but it uses as observables the position differences provided by the video system, in order to estimate position and speed of the aircraft center of gravity. This filter is used during GPS outages only in order to propagate position and speed. When it is not activated, its output is set equal to zero.

For the sake of simplicity, we will consider the IMU installed at the center of gravity of the aircraft (in order to eliminate lever arm effects) with its axes aligned with the body axes. Euler angles are used to represent aircraft attitude. Therefore, we have

\[
\begin{align*}
\dot{\hat{p}}_{cg} &= \hat{C}_b \hat{f}_b + g^n \quad (29) \\
\dot{\hat{r}}_{cg} &= \hat{v}_{cg} \quad (30)
\end{align*}
\]

and

\[
\begin{bmatrix}
\dot{\phi} \\
\dot{\theta} \\
\dot{\psi}
\end{bmatrix}_v = J \begin{bmatrix}
\dot{\phi} \\
\dot{\theta} \\
\dot{\psi}
\end{bmatrix}_v \cdot \hat{w}_{nb} \quad (31)
\]

Starting from these non-linear equations, EKF perturbated form of the prediction equations are derived. Output equations are based on the following GPS predicted measurements

\[
\begin{align*}
\hat{p}_{GPS}^s &= C_n \dot{r}_{eg}^n + C_b \cdot r_{eg-GPS}^b \\
\hat{v}_{GPS}^n &= \hat{v}_{vg}^n + C_b \cdot [\hat{w}_{nb}^b]_x \cdot r_{eg-GPS}^b.
\end{align*} \quad (32) \quad (33)
\]

For the video system considering that it provides delta-attitudes (derived from the essential matrix), a pseudo-attitude video based is built starting from an initial known value.

\[
\begin{bmatrix}
\dot{\phi} \\
\dot{\theta} \\
\dot{\psi}
\end{bmatrix}_{video(k)} = \begin{bmatrix}
\dot{\phi} \\
\dot{\theta} \\
\dot{\psi}
\end{bmatrix}_{video(k-1)} + \begin{bmatrix}
\frac{d\phi}{dt} \\
\frac{d\theta}{dt} \\
\frac{d\psi}{dt}
\end{bmatrix}_{video(k)} \quad (34)
\]

The output equation for the estimation of speed based on video measurements is:

\[
\Delta \dot{p} = \frac{\Delta V}{\Delta t} = \hat{v}_{cg}^n, \quad (35)
\]

where \(\Delta p\) is provided by the video system and represent the three unit vector in the direction of the movement, \(V\) is the Euclidean norm of the speed considered constant (is the last valid value obtained from the GPS-IMU sensor fusion algorithm) and \(\Delta t\) is the time difference between two measurements of the video system.

VI. NUMERICAL ANALYSIS

Two types of sensors configurations have been considered. The first one refers to a big lever arm between GPS and aircraft center of gravity (10 meters on each axis), while the second one is a more typical configuration because it consider a lever arm only for the y-body axis with a magnitude of two meters.

Figure 2 shows the UAV attitude estimation results relative to the first configuration. Attitude is reported in terms of roll, yaw and pitch angles. A performance improvement of attitude estimation accuracy is clearly observable since the accuracy of video based estimation is better than the GPS one. Indeed the static accuracy of Video System is of the order of only 1 degree, while the GPS accuracy is of the order of 5 degrees.

For the second configuration, results are reported in Figure 3. It can be observed that the \(\phi\) angle estimation shows the same behavior of the first configuration for both GPS and video assisted estimation algorithms. This happens because observability is guaranteed by the lever arm. On the contrary, for \(\theta\) (not reported here for the sake of brevity) and \(\psi\) angles, the estimation error with only the GPS is much greater with respect to the previous configuration. This is principally due to the absence of a significant lever arm, in which case the estimation error is that of simple INS measurements integration. However, the IMU+VIDEO algorithm is more robust to the lever arm effect since it still performs as well as in the first configuration.

The distance between the GPS antenna and the aircraft center of gravity is generally minimized because of the aircraft configuration and the need to reduce cables and wiring. Moreover, a big lever arm introduces errors in the estimation of aircraft center of gravity position and speed. Figures 4 and 5 show the results of position and speed estimations with video measurements. At time instant 120 sec a GPS outage has been simulated with a duration of 120 sec. This situation could happen, for example, while traveling through a tunnel or in wooded area, in indoor navigation and/or in hostile environments where the GPS signal is not available. The use of video measurements allows continuing estimation of position and speed with a good accuracy throughout the sequence. Without the fusion of these measurements, solution would diverge after short time, due to integration of inertial measurement errors.

VII. CONCLUSION

In this paper an innovative technique for estimation of the position and attitude of an UAV has been proposed. The main
innovative aspect of this technique concerns the introduction of a vision-based system composed of a low-cost camera device. The information carried by the camera is then integrated with classical data coming from the IMU and GPS units in a sensor fusion algorithm. A linearized Kalman Filter has been implemented, because dynamic and observation equations are non-linear in their original form.

The simulation setup is composed of accurate models for realistic UAV flight behavior, sensors simulation and 3D photorealistic rendering. Preliminary results are very encouraging, since they show the improvement up to one order of magnitude on the attitude accuracy, using an HD camera. Future work is the validation of simulation results on a real flight test using a mini-UAV platform.

REFERENCES

Fig. 4. Velocity estimation in North-East-Down (NED) reference system. Groundtruth (Black), IMU+Video estimation (Blue).

Fig. 5. Down position estimation. Top: entire simulation. Bottom: zoom during the GPS outage. Groundtruth (Black), IMU+Video estimation (Blue).


