35th European Rotorcraft Forum 2009
DocumentID: 101168

INERTIAL NAVIGATION OF AN AUTONOMOUS ROTORCRAFT VEHICLE USING STEREO-VISION

C.L. Bottasso and D. Leonello

Dipartimento di Ingegneria Aerospaziale,
Politecnico di Milano, Milano, Italy
carlo.bottasso@polimi.it, leonello@aero.polimi.it

Abstract

We describe a novel inertial navigation system based on measurement fusion which includes stereo-vision among its sensors. The vision-augmented system provides enhanced accuracy in the estimation of the vehicle states when flying in proximity of obstacles, and can operate without GPS signal, for example when flying under vegetation cover, indoors or in complex urban environments. Scene feature points are tracked between the left an right images and across time steps, yielding vision-based information on the state of motion of the vehicle which is fused together with other non-vision-based sensors. The proposed approach is demonstrated using simulation for an autonomous helicopter flying in an urban environment.

1 INTRODUCTION

Rotorcraft unmanned aerial vehicles (RUAVs) offer great potential for a wide range of challenging applications because of their ability to hover and their maneuverability, a particularly valuable asset in complex and confined environments. On-board flight control systems provide the vehicle with the required level of autonomy, which varies depending on the application needs, by providing from basic capabilities such as stabilization and trajectory following, to more sophisticated behavioral skills such as exploration of unknown and dynamic environments.

Irrespectively of the level of vehicle autonomy, onboard avionics must provide the flight control systems with real-time estimates of the vehicle states with a sufficient level of precision. Several lightweight low-cost hardware solutions based on off-the-shelf available hardware components have already been proposed and successfully demonstrated for autonomous rotorcraft [4, 3, 9]. The information provided by the various sensors are usually fused together using some variant of the Kalman filtering approach, to deliver estimates of the vehicle position, attitude, linear and angular velocities, while accounting for the stochastic nature of the problem by including the effects of measurement and process noise sources.

In this work we propose a sensor fusion approach to inertial navigation which incorporates stereo-vision in the classical framework described above. In our approach, the vision information provided by a Kanade-Lucas-Tomasi (KLT) feature tracker [10] is directly fused with the other sensors of the navigation system, which include in the current implementation a tri-axial accelerometer, a tri-axial gyro, a Ground Positioning System (GPS), a sonar altimeter and a tri-axial magnetometer. Each feature point tracked by the vision system is treated as an independent motion sensor, and it is fused in parallel with all other sensors in an extended Kalman filter (EKF). The filter uses quaternions for the parameterization of finite rotations and accounts for the different operating frequencies of the various sensors. Outliers within the vision sensor set (due to feature points on moving objects, excessive noise, tracking errors, etc.) are identified by verifying the compatibility of their optical flow with the vehicle motion. The algorithm accommodates in a straightforward way the dynamic insertion and deletion of feature points.

The overall system architecture is sketched in Fig. 1. Notice that the stereo-cameras used by the state estimator are shared with the map-building systems which, by enabling the recognition of obstacles and targets, support the higher levels of autonomy.

The use of vision sensors enhances the quality of the state estimates with respect to the classical non-vision-based inertial navigation systems. In fact, the number of independent vision sensors available at each instant of time can be quite high (of the order of one hundred with our current hardware), being only limited by the available computational resources and quality and resolution of the cameras. The quality of the information provided by the vision sensors increases when operating in close proximity of obstacles, exactly when higher precision flight is necessary, and degrades when flying in open environments with fewer and farther features to track, but where

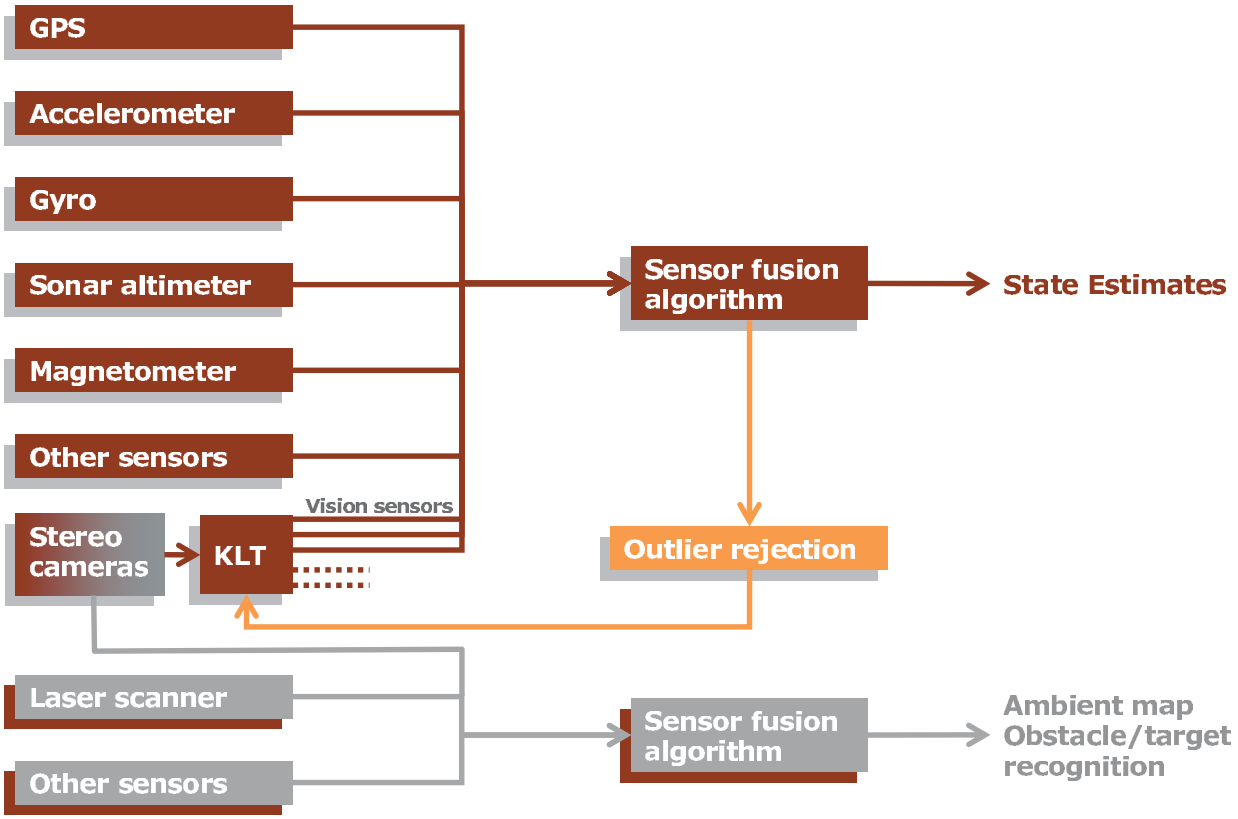

Figure 1: Vision-aided sensor fusion for the estimation of vehicle states.

the classical non-vision based inertial system already provides estimates of sufficient accuracy.

Another interesting characteristic of the proposed vision-aided approach is that it enables accurate estimation of the vehicle states even when the GPS looses satellite contact, for example when flying under vegetation cover or indoors. In fact, the GPS provides both absolute position and linear velocity information to the sensor fusion algorithm. While the former can not be replaced by a vision system, which can only provide relative position data, the latter is embedded in the spatio-temporal tracking of the point features. Hence good quality state estimates are available for feedback control even after the loss of the GPS signal enabling, for example, a seamless transition from outdoor to indoor flight, or viceversa.

The proposed approach combines in a synergistic way the classical inertial navigation sensors together with the vision-based ones, in a general probabilistic framework. Notice that the vision data provided by small on-board cameras are affected by high noise levels, so that their direct use for extracting real-time motion information (Structure from Motion, SfM, see e.g. [5] and references therein) is not reliable enough for providing high-quality state estimates to high performance agile flying vehicles such as helicopters. Furthermore, SfM suffers from drift of the estimates over time, which also calls for corrective actions.

The sensor fusion approach and underlying hardware described herein is capable of reconstructing the vehicle states with suitable levels of precision and bandwidth for loop closure for a small rotorcraft vehicle. The architecture is modular, so that other sensors can be easily accommodated (e.g. laser-scanner, fish-eyes, barometric altimeter, etc.), to further enhance the accuracy and robustness of the estimates.

2 INERTIAL NAVIGATION BY MEASUREMENT FUSION

In this section we describe a measurement fusion [12] approach for the estimation of the linear and angular velocity, position and attitude of a rigid body, based on measurements provided by a tri-axial accelerometer, a tri-axial gyroscope, a GPS, a sonar altimeter, and a tri-axial magnetometer.

2.1 Kinematics

We consider an inertial frame of reference centered at point O and denoted by a triad of unit vectors $\mathcal{E} \doteq (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$, $\boldsymbol{e}_1$ pointing North, $\boldsymbol{e}_2$ pointing East and $\boldsymbol{e}_3$ pointing down (NED navigational system). A

body-attached local frame of reference has origin in the generic material point B of the vehicle and has a triad of unit vectors $\mathcal{B} \doteq (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$.

The kinematic equations describing the motion of the body-attached reference frame with respect to the inertial one write

$$\dot{\boldsymbol{v}}_B = \boldsymbol{a}_B, \tag{1a}$$
$$\dot{\boldsymbol{\omega}} = \boldsymbol{\alpha}, \tag{1b}$$
$$\dot{\boldsymbol{r}}_{OB} = \boldsymbol{v}_B, \tag{1c}$$
$$\dot{\boldsymbol{q}} = \boldsymbol{T}(\boldsymbol{\omega})\boldsymbol{q}, \tag{1d}$$

where $\boldsymbol{v}_B$ is the velocity of point B and $\boldsymbol{a}_B$ its acceleration, $\boldsymbol{\omega}$ the angular velocity of triad $\mathcal{B}$ with respect to triad $\mathcal{E}$ and $\boldsymbol{\alpha}$ its angular acceleration, $\boldsymbol{r}_{OB}$ the position vector from point O to point B, and finally $\boldsymbol{q}$ are rotation parameters, which are chosen as quaternions in the present work.

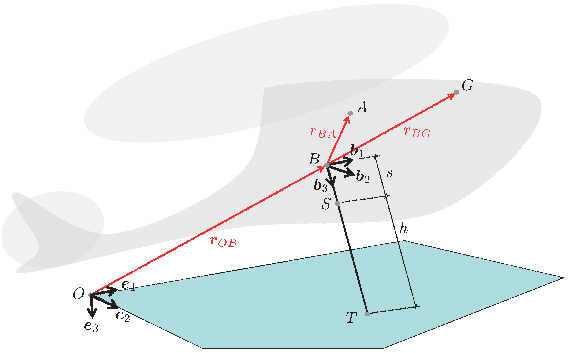

Figure 2: Reference frames and location of the sensors on the vehicle.

2.2 Sensors

The components of the acceleration in the body-attached frame are sensed by an accelerometer located at point A on the vehicle (see Fig. 2). The accelerometer yields a reading $\boldsymbol{a}_{\text{acc}}$ affected by noise $\boldsymbol{n}_{\text{acc}}$:

$$\boldsymbol{a}_{\text{acc}} = \boldsymbol{g}^{\mathcal{B}} - \boldsymbol{a}_A^{\mathcal{B}} + \boldsymbol{n}_{\text{acc}}. \tag{2}$$

$\boldsymbol{g}^{\mathcal{B}}$ indicates the body-attached components of the acceleration of gravity, where

$$\boldsymbol{g}^{\mathcal{B}} = \boldsymbol{R}\boldsymbol{g}^{\mathcal{E}}, \tag{3}$$

with $\boldsymbol{g}^{\mathcal{E}} = (0, 0, g)^T$ and $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{q})$ are the components of the rotation tensor which brings triad $\mathcal{E}$ into triad $\mathcal{B}$. Expressing the acceleration at point A, $\boldsymbol{a}_A$, in terms of the acceleration at point B, we have

$$\boldsymbol{a}_A = \boldsymbol{a}_B + \boldsymbol{\omega} \times \boldsymbol{\omega} \times \boldsymbol{r}_{BA} + \boldsymbol{\alpha} \times \boldsymbol{r}_{BA}, \tag{4}$$

and using Eq. (1a) and (2), we get

$$\dot{\boldsymbol{v}}_B^{\mathcal{E}} = \boldsymbol{g}^{\mathcal{E}} - \boldsymbol{R}(\boldsymbol{a}_{\text{acc}} + \boldsymbol{\omega}^{\mathcal{B}} \times \boldsymbol{\omega}^{\mathcal{B}} \times \boldsymbol{r}_{BA}^{\mathcal{B}} + \boldsymbol{\alpha}^{\mathcal{B}} \times \boldsymbol{r}_{BA}^{\mathcal{B}}) + \boldsymbol{R}\boldsymbol{n}_{\text{acc}}. \tag{5}$$

Gyroscopes measure the body-attached components of the angular velocity vector, yielding a reading $\boldsymbol{\omega}_{\text{gyro}}$ affected by a bias $\boldsymbol{b}_{\text{gyro}}$ and noise disturbance $\boldsymbol{n}_{\text{gyro}}$:

$$\boldsymbol{\omega}_{\text{acc}} = \boldsymbol{\omega}^{\mathcal{B}} + \boldsymbol{b}_{\text{gyro}} + \boldsymbol{n}_{\text{gyro}}. \tag{6}$$

The unknown bias can be identified by promoting it to a state variable; however, since one can not identify simultaneously both bias and angular velocity as it appears clearly from Eq. (6), this means that $\boldsymbol{\omega}$ can not be treated as a state variable together with its bias and that consequently Eq. (1b) should be dropped from system (1). An alternative approach is used in this work, whereby the gyro measures are used for computing an estimate of the angular acceleration. Since this implies a differentiation of the measures of the gyros, assuming a constant (or slowly varying) bias over the differentiation interval, knowledge of the bias becomes unnecessary. Hence we estimate the angular acceleration as

$$\boldsymbol{\alpha}^{\mathcal{B}} \simeq \boldsymbol{\alpha}_h(\boldsymbol{\omega}_{\text{gyro}}), \tag{7}$$

where $\boldsymbol{\alpha}_h$ is a discrete differentiation operator. In this work, the angular acceleration at time t_k is computed according to the following three-point stencil formula based on a parabolic interpolation:

$$\boldsymbol{\alpha}_h(t_k) = \frac{3\boldsymbol{\omega}_{\text{gyro}}(t_k) - 4\boldsymbol{\omega}_{\text{gyro}}(t_{k-1}) + \boldsymbol{\omega}_{\text{gyro}}(t_{k-2})}{2h}, \tag{8}$$

where $h \equiv t_k - t_{k-1} = t_{k-1} - t_{k-2}$.

At the light of the previous considerations, the kinematic equations (1) become

$$\dot{\boldsymbol{v}}_B^{\mathcal{E}} = \boldsymbol{g}^{\mathcal{E}} - \boldsymbol{R}\big(\boldsymbol{a}_{\text{acc}} + \boldsymbol{\omega}^{\mathcal{B}} \times \boldsymbol{\omega}^{\mathcal{B}} \times \boldsymbol{r}_{BA}^{\mathcal{B}} + \boldsymbol{\alpha}_h(\boldsymbol{\omega}_{\text{gyro}} + \boldsymbol{n}_{\text{gyro}}) \times \boldsymbol{r}_{BA}^{\mathcal{B}}\big) + \boldsymbol{R}\boldsymbol{n}_{\text{acc}}, \tag{9a}$$
$$\dot{\boldsymbol{\omega}}^{\mathcal{B}} = \boldsymbol{\alpha}_h(\boldsymbol{\omega}_{\text{gyro}} + \boldsymbol{n}_{\text{gyro}}), \tag{9b}$$
$$\dot{\boldsymbol{r}}_{OB}^{\mathcal{E}} = \boldsymbol{v}_B^{\mathcal{E}}, \tag{9c}$$
$$\dot{\boldsymbol{q}} = \boldsymbol{T}(\boldsymbol{\omega}^{\mathcal{B}})\boldsymbol{q}, \tag{9d}$$

where matrix $\boldsymbol{T}$ writes

$$\boldsymbol{T}(\boldsymbol{\omega}^{\mathcal{B}}) = \frac{1}{2}\begin{bmatrix} 0 & -\boldsymbol{\omega}^{\mathcal{B}^T} \\ \boldsymbol{\omega}^{\mathcal{B}} & -\boldsymbol{\omega}_{\times}^{\mathcal{B}} \end{bmatrix}. \tag{10}$$

The set of state-space equations (9) can be written in the following compact form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\nu}), \tag{11}$$

where the state vector is $\boldsymbol{x} \doteq (\boldsymbol{v}_B^{\mathcal{E}^T}, \boldsymbol{\omega}^{\mathcal{B}^T}, \boldsymbol{r}_{OB}^{\mathcal{E}^T}, \boldsymbol{q})^T$, the input vector is $\boldsymbol{u} \doteq (\boldsymbol{a}_{\text{acc}}^T, \boldsymbol{\omega}_{\text{gyro}}^T)^T$, and the process[1] noise vector $\boldsymbol{\nu} \doteq (\boldsymbol{n}_{\text{acc}}^T, \boldsymbol{n}_{\text{gyro}}^T)^T$.

A GPS is located at point G on the vehicle (see Fig. 2). The inertial components of the velocity and position vectors of point G, noted respectively $\boldsymbol{v}_G^{\mathcal{E}}$ and $\boldsymbol{r}_{OG}^{\mathcal{E}}$, can be expressed in terms of the states $\boldsymbol{x}$ as

$$\boldsymbol{v}_G^{\mathcal{E}} = \boldsymbol{v}_B^{\mathcal{E}} + \boldsymbol{R}\boldsymbol{\omega}^{\mathcal{B}} \times \boldsymbol{r}_{BG}^{\mathcal{B}}, \tag{12a}$$

$$\boldsymbol{r}_{OG}^{\mathcal{E}} = \boldsymbol{r}_{OB}^{\mathcal{E}} + \boldsymbol{R}\boldsymbol{r}_{BG}^{\mathcal{B}}. \tag{12b}$$

The GPS yields measurements of the position and velocity of point G affected by noise, i.e.

$$\boldsymbol{v}_{\text{gps}} = \boldsymbol{v}_G^{\mathcal{E}} + \boldsymbol{n}_{v_{\text{gps}}}, \tag{13a}$$

$$\boldsymbol{r}_{\text{gps}} = \boldsymbol{r}_{OG}^{\mathcal{E}} + \boldsymbol{n}_{r_{\text{gps}}}. \tag{13b}$$

Furthermore, a sonar altimeter measures the distance h along the body-attached vector $\boldsymbol{b}_3$ between its location at point S on the vehicle, being $\boldsymbol{r}_{OS}^{\mathcal{B}} = (0, 0, s)^T$, and point T on the terrain, as shown in Fig. 2. In the body-attached frame $\mathcal{B}$, the distance vector between S and T has components $\boldsymbol{r}_{ST}^{\mathcal{B}} = (0, 0, h)^T$, which are readily transformed into inertial components as $\boldsymbol{r}_{ST}^{\mathcal{E}} = \boldsymbol{R}\boldsymbol{r}_{ST}^{\mathcal{B}}$. Hence, we get

$$h = r_{OB_3}^{\mathcal{E}} / R_{33} - s, \tag{14}$$

where $r_{OB_3}^{\mathcal{E}} = \boldsymbol{r}_{OB} \cdot \boldsymbol{e}_3$ and $\boldsymbol{R} = [R_{ij}]$, $i, j = 1, 2, 3$. The sonar altimeter yields a reading h_{sonar} affected by noise n_{sonar}, i.e.

$$h_{\text{sonar}} = h + n_{\text{sonar}}. \tag{15}$$

Finally, we consider a magnetometer which senses the components of the magnetic field $\boldsymbol{m}$ of the Earth in the body-attached system $\mathcal{B}$, which can be expressed as

$$\boldsymbol{m}^{\mathcal{B}} = \boldsymbol{R}^T \boldsymbol{m}^{\mathcal{E}}, \tag{16}$$

where the inertial components $\boldsymbol{m}^{\mathcal{E}}$ are assumed to be known and constant in the area of operation of the vehicle. The magnetometer yields a measurement $\boldsymbol{m}_{\text{magn}}$ affected by noise $\boldsymbol{n}_{\text{magn}}$, i.e.

$$\boldsymbol{m}_{\text{magn}} = \boldsymbol{m}^{\mathcal{B}} + \boldsymbol{n}_{\text{magn}}. \tag{17}$$

Equations (12), (14) and (16) can be gathered together and written in compact form as

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}), \tag{18}$$

where $\boldsymbol{y} \doteq (\boldsymbol{v}_G^{\mathcal{E}^T}, \boldsymbol{r}_{OG}^{\mathcal{E}^T}, h, \boldsymbol{m}^{\mathcal{B}^T})^T$ is the output vector.

Similarly, Eqs. (13), (15) and (17) can be gathered together and written in compact form as

$$\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{\mu}, \tag{19}$$

where $\boldsymbol{z} \doteq (\boldsymbol{v}_{\text{gps}}^T, \boldsymbol{r}_{\text{gps}}^T, h_{\text{sonar}}, \boldsymbol{m}_{\text{magn}}^T)^T$ is the measurement vector, and $\boldsymbol{\mu} \doteq (\boldsymbol{n}_{v_{\text{gps}}}^T, \boldsymbol{n}_{r_{\text{gps}}}^T, n_{\text{sonar}}, \boldsymbol{n}_{\text{magn}}^T)^T$ the measurement noise vector.

2.3 State Estimation by Filtering

Equations (11), (18) and (19) are gathered here for convenience in a complete set of state-space equations

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\nu}(t)\big), \tag{20a}$$

$$\boldsymbol{y}(t_k) = \boldsymbol{h}\big(\boldsymbol{x}(t_k)\big), \tag{20b}$$

$$\boldsymbol{z}(t_k) = \boldsymbol{y}(t_k) + \boldsymbol{\mu}(t_k). \tag{20c}$$

The state estimation problem (20) can be solved with a number of filtering approaches. The Kalman filter is an optimal estimator for unconstrained linear systems with normally distributed process and measurement noise, while for non-linear problems various methods have been proposed, including the extended Kalman filter, the unscented Kalman filter, the sigma point and particle filters [11]. In this work we use the extended Kalman filter, which amounts to an approximate generalization of the Kalman filter to non-linear systems obtained by linearizing the dynamics at each time step. Theoretical results on the stability and convergence of this approach are discussed in Ref. [8].

The equations of motion (20a) are integrated on each sampling interval $[t_k, t_{k+1}]$ to yield a state prediction $\bar{\boldsymbol{x}}(t_{k+1})$ together with its associated output vector $\bar{\boldsymbol{y}}(t_{k+1})$ given by (20b). Next, at each sampling instant the state predictions are improved based on the innovations, i.e. the difference between the measurements $\boldsymbol{z}(t_{k+1})$ and the predicted outputs $\bar{\boldsymbol{y}}(t_{k+1})$, as

$$\hat{\boldsymbol{x}}(t_{k+1}) = \bar{\boldsymbol{x}}(t_{k+1}) + \boldsymbol{K}(t_{k+1})\big(\boldsymbol{z}(t_{k+1}) - \bar{\boldsymbol{y}}(t_{k+1})\big), \tag{21}$$

where $\boldsymbol{K}(t_{k+1})$ is a time-varying gain matrix, which is propagated forward in time together with the state estimates based on the covariances of the estimation error, and of the process and measurement noise. The current implementation of the software accounts for the fact that not all sensor measurements are available at each sampling instant; for example, the GPS yields new readings with a frequency of 1 Hz, while the other sensors operate at 50 Hz.

[1] Although noise terms appearing in dynamic equations are usually termed "process noise", in the present problem they are in reality due to measurement noise in the accelerometer and gyro readings.

3 VISION-AUGMENTED INERTIAL NAVIGATION

3.1 Stereo Projection and Vision-Based Position Sensors

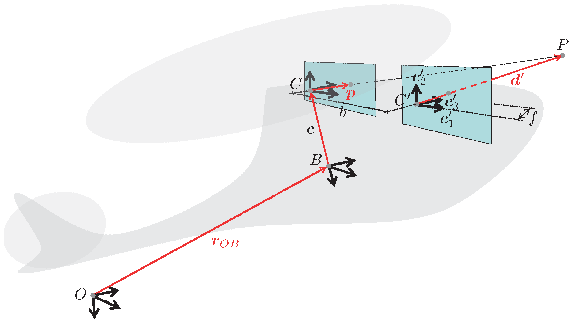

Figure 3: Stereo cameras and feature point projection.

Consider a pair of stereo cameras located on the vehicle, as shown in Fig. 3. A triad of unit vectors $\mathcal{C} \doteq (\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3)$ has origin at the optical center C of the left camera, where $\boldsymbol{c}_1$ is directed along the horizontal scanlines of the image plane while $\boldsymbol{c}_3$ is parallel to the optical axis, pointing towards the scene. We indicate with $\boldsymbol{C}$ the rotation tensor which brings triad $\mathcal{B}$ into triad $\mathcal{C}$, which is constant in time since the two are rigidly attached to the vehicle and move with it.

A feature point P has a position vector $\boldsymbol{d}$ with respect to point C, while its projection on the image plane has a position vector $\boldsymbol{p} = \boldsymbol{\pi}(\boldsymbol{d})$, where $\boldsymbol{\pi}(\cdot)$ is the projection operator. Assuming an ideal pinhole camera model, the components in $\mathcal{C}$ of the position vectors, $\boldsymbol{d}^{\mathcal{C}} = (d_1, d_2, d_3)^T$ and $\boldsymbol{p}^{\mathcal{C}} = (p_1, p_2, f)^T$, are related as

$$\boldsymbol{p}^{\mathcal{C}} = \frac{f}{d_3}\boldsymbol{d}^{\mathcal{C}}, \tag{22}$$

where f is the camera focal length, assumed known (calibrated camera).

The right camera has its optical axis and scanlines parallel to those of the left camera, i.e. $\mathcal{C}' \equiv \mathcal{C}$, where we use the symbol $(\cdot)'$ to indicate quantities of the right camera. The origin C' of triad $\mathcal{C}'$ has a distance $\boldsymbol{b} = b\boldsymbol{c}_1$ from C, where b is the stereo baseline.

A *vision-based position sensor* can be obtained by first writing the triangle vector equality

$$\boldsymbol{d} = \boldsymbol{b} + \boldsymbol{d}'. \tag{23}$$

Next, using Eq. (22), one gets an expression for the components of the distance vector of point P with respect to each camera in terms of the components measured on the image plane of the same camera, i.e.

$$\boldsymbol{d}^{\mathcal{C}} = \frac{b}{d}\boldsymbol{p}^{\mathcal{C}}, \tag{24}$$

where $d = p_1 - p_1'$ is the disparity; an identical relationship is obtained for the right camera.

Computing the disparity d requires solving the correspondence problem, i.e. finding points in each image which are projections of the same scene point. In this work, feature points are detected and tracked using the KLT algorithm [6, 10]. Feature selection accounts for good texture properties, so as to increase accuracy. The tracker uses an affine motion model based on rigid translation and linear warping, to exclude features subjected to visual obstruction and for the identification of outliers; possibly remaining outliers missed by the KLT algorithm are identified by the detection procedure described later on.

An example of feature point detection and tracking using the KLT algorithm is shown in Fig. 4. The two top images of the figure represent the left and right frames grabbed by the cameras at a certain instant of time, with a feature point that has been recognized between the two images. The two lower images represent the frames grabbed at a successive time instant. At first, the feature point in the right image at the earlier time step is tracked and identified in the right image at the later time step. This phase is essential to maintain feature consistency through time and to detect outliers. Next, the same scene point is matched to the corresponding image point in the left camera. Matching across time instants alternates between the left and right images; therefore, in the example of the figure, at the next time step points in the left image will be matched, where possible, to points in the left image at the later time.

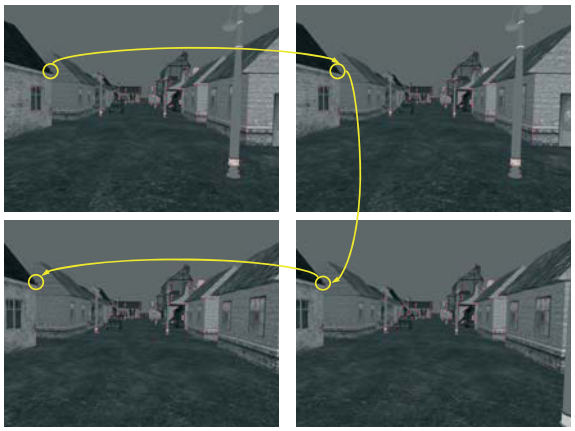

Figure 4: Tracking a scene point through two successive time steps using the KLT algorithm.

Cameras used on-board small UAVs are typically noisy and of low resolution. Hence, the expected ac-

curacy of the reconstructed position components in Eq. (24) is fairly low. An estimate of the accuracy can be readily derived by considering the influence of the disparity measure error on the computed feature position. For example, differentiating the third component of Eq. (24) we get

$$\delta d_3 = \pm \frac{fb}{w\tilde{d}^2}\,\delta\tilde{d}, \tag{25}$$

where we have set $d = \tilde{d}w$, being $\tilde{d}$ the disparity in pixel units and w the pixel width. Figure 5 reports the error in the estimation of d_3 for an error in the disparity of one pixel for the BumbleBee X3 camera [1], which has been used in the present work.

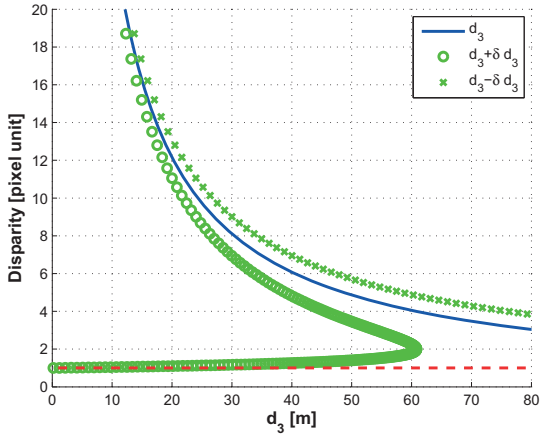

Figure 5: Stereo projection sensitivity. Solid line: reconstructed distance between feature and stereo camera. Symbols ∘ and ×: distance reconstruction errors for one pixel error in the disparity.

From this plot, it is clear that the use for navigation purposes of stereo information obtained with low resolution vision equipment should be performed with great care. To address this issue, in this paper we propose to fuse the information embedded in measures of the kind of Eq. (24) together with the information provided by the already described sensors (accelerometers, gyros, GPS, sonar altimeter, magnetometer and possibly others) in an augmented version of the filtering problem (20).

3.2 Vision-Based Motion Sensors

A *vision-based motion sensor* can be derived by first writing the vector closure expression

$$\boldsymbol{r}_{OP} = \boldsymbol{r} + \boldsymbol{c} + \boldsymbol{d}, \tag{26}$$

where $\boldsymbol{r} \doteq \boldsymbol{r}_{OB}$ is the position vector of the reference point B on the vehicle with respect to the origin O of the inertial frame, $\boldsymbol{c} \doteq \boldsymbol{r}_{BC}$ is the position vector of the camera optical center C with respect to the origin of the vehicle body-attached frame, and $\boldsymbol{d} \doteq \boldsymbol{r}_{CP}$ the position vector of point P with respect to the origin of the camera reference frame (cfr. Fig. 3). Next, by differentiating Eq. (26), using the fact that $\boldsymbol{v}_P \doteq \dot{\boldsymbol{r}}_{OP} \equiv 0$ if P is a fixed point, and expressing each of the terms in convenient reference frames, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}^{\mathcal{E}} + \boldsymbol{R}\boldsymbol{c}^{\mathcal{B}} + \boldsymbol{R}\boldsymbol{C}\boldsymbol{d}^{\mathcal{C}}) = 0, \tag{27}$$

which yields

$$\dot{\boldsymbol{d}}^{\mathcal{C}} = -\boldsymbol{C}^T\big(\boldsymbol{R}^T\boldsymbol{v}_B^{\mathcal{E}} + \boldsymbol{\omega}^{\mathcal{B}} \times (\boldsymbol{c}^{\mathcal{B}} + \boldsymbol{C}\boldsymbol{d}^{\mathcal{C}})\big). \tag{28}$$

The apparent (i.e. due to the motion of the vehicle) velocity of a feature point on the camera image plane is computed by differentiating Eq. (22) with respect to time

$$\dot{\boldsymbol{p}}^{\mathcal{C}} = \frac{f}{d_3}\dot{\boldsymbol{d}}^{\mathcal{C}} - f\frac{\dot{d}_3}{d_3^2}\boldsymbol{d}^{\mathcal{C}}, \tag{29a}$$

$$= \boldsymbol{M}\dot{\boldsymbol{d}}^{\mathcal{C}}, \tag{29b}$$

where

$$\boldsymbol{M} \doteq \frac{f}{d_3}\begin{bmatrix} 1 & 0 & -d_1/d_3 \\ 0 & 1 & -d_2/d_3 \\ 0 & 0 & 0 \end{bmatrix}. \tag{30}$$

By inserting the expression for $\dot{\boldsymbol{d}}^{\mathcal{C}}$ given by Eq. (28) into Eq. (29b), one gets

$$\dot{\boldsymbol{p}}^{\mathcal{C}} = -\boldsymbol{M}\boldsymbol{C}^T\big(\boldsymbol{R}^T\boldsymbol{v}_B^{\mathcal{E}} + \boldsymbol{\omega}^{\mathcal{B}} \times (\boldsymbol{c}^{\mathcal{B}} + \boldsymbol{C}\boldsymbol{d}^{\mathcal{C}})\big), \tag{31}$$

which is an expression in terms of the states $\boldsymbol{x}$ of Eqs. (11) for the apparent velocity on the image plane, a quantity which can be measured by tracking feature points across consecutive frames; in this sense, Eq. (31) can be interpreted as the output equation of a vision-based motion sensor.

Since the use of Eq. (31) requires computing the apparent velocity $\dot{\boldsymbol{p}}^{\mathcal{C}}$ of tracked feature points, which is a noisy operation, we prefer to develop a discrete version of Eq. (28) to be used as motion sensor, and reserve the use of Eq. (31) for the detection of outliers, i.e. those feature points which are not fixed with respect to the inertial frame.

To this end, considering two consecutive time instants t_k and t_{k+1}, one can write for both the left and right cameras the following vector closure relationship (cfr. Fig. (6))

$$\boldsymbol{r}(t_k) + \boldsymbol{c}(t_k) + \boldsymbol{d}(t_k) = \boldsymbol{r}(t_{k+1}) + \boldsymbol{c}(t_{k+1}) + \boldsymbol{d}(t_{k+1}). \tag{32}$$

By expressing each of the terms of Eq. (32) in convenient reference frames, we obtain

$$\begin{aligned}\boldsymbol{d}(t_{k+1})^{\mathcal{C}_{k+1}} = -\boldsymbol{C}^T\big(\boldsymbol{R}(t_{k+1})^T(\boldsymbol{r}(t_{k+1})^{\mathcal{E}} - \boldsymbol{r}(t_k)^{\mathcal{E}}) \\ + (\boldsymbol{I} - \boldsymbol{R}(t_{k+1})^T\boldsymbol{R}(t_k))(\boldsymbol{c}^{\mathcal{B}} + \boldsymbol{C}\boldsymbol{d}(t_k)^{\mathcal{C}_k})\big) + \boldsymbol{d}(t_k)^{\mathcal{C}_k}.\end{aligned} \tag{33}$$

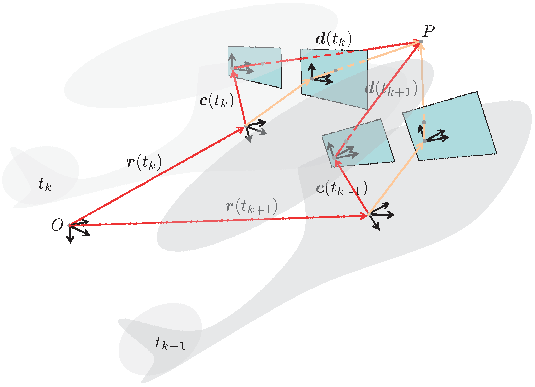

Figure 6: Geometry for the derivation of the discrete vision-based motion sensor.

This expression depends on the absolute position vector $\boldsymbol{r}$, a quantity which however can not be observed by a vision system, which only senses relative distances. Hence, since in the absence of GPS measurements the observed absolute position will drift away from the true one, it is advisable to rewrite the above equation in terms of velocities. By setting

$$\boldsymbol{I} - \boldsymbol{R}(t_{k+1})^T \boldsymbol{R}(t_k) = h\,\boldsymbol{\omega}(t_{k+1})^{\mathcal{B}}_{\times}, \tag{34a}$$

$$\boldsymbol{r}(t_{k+1})^{\mathcal{E}} - \boldsymbol{r}(t_k)^{\mathcal{E}}) = h\,\boldsymbol{v}^{\mathcal{E}}(t_{k+1}), \tag{34b}$$

we have

$$\begin{aligned}\boldsymbol{d}(t_{k+1})^{\mathcal{C}_{k+1}} = -h\,\boldsymbol{C}^T\big(&\boldsymbol{R}(t_{k+1})^T\boldsymbol{v}^{\mathcal{E}}(t_{k+1}) \\ &+ \boldsymbol{\omega}^{\mathcal{B}}(t_{k+1}) \times (\boldsymbol{c}^{\mathcal{B}} + \boldsymbol{C}\boldsymbol{d}(t_k)^{\mathcal{C}_k})\big) + \boldsymbol{d}(t_k)^{\mathcal{C}_k}.\end{aligned} \tag{35}$$

Clearly, the same result could have been obtained by temporal discretization of Eq. (28) using Eqs. (34). The right hand side of the previous equation depends on the current linear velocity $\boldsymbol{v}^{\mathcal{E}}(t_{k+1})$, angular velocity $\boldsymbol{\omega}^{\mathcal{B}}(t_{k+1})$ and orientation $\boldsymbol{q}(t_{k+1})$ (through $\boldsymbol{R}(t_{k+1})$), on past and hence known quantities $\boldsymbol{d}(t_k)$, and on known constant terms $\boldsymbol{c}^{\mathcal{B}}$ and $\boldsymbol{C}$. Hence, Eq. (35) represents an output equation which can be appended to the output system (18), defining a vision-augmented output vector

$$\begin{aligned}\boldsymbol{y} = (&\boldsymbol{v}_G^{\mathcal{E}^T}, \boldsymbol{r}_{OG}^{\mathcal{E}^T}, h, \boldsymbol{m}^{\mathcal{B}^T}, \ldots, \\ &\boldsymbol{d}(t_{k+1})^{\mathcal{C}_{k+1}^T}, \boldsymbol{d}(t_{k+1})^{\mathcal{C}_{k+1}'^T}, \ldots)^T.\end{aligned} \tag{36}$$

For each tracked feature point, we include in the augmented vector a new output for both the left and right cameras. When a feature point is lost or discarded because identified as an outlier, it is simply removed from the output vector (36).

The left hand side of Eq. (35), $\boldsymbol{d}(t_{k+1})^{\mathcal{C}_{k+1}}$, is computed at time step t_{k+1} by stereo reconstruction using Eq. (24), which yields an estimate $\boldsymbol{d}_{\text{vsn}}$ affected by noise $\boldsymbol{n}_{\text{vsn}}$:

$$\boldsymbol{d}_{\text{vsn}} = \boldsymbol{d}(t_{k+1})^{\mathcal{C}_{k+1}} + \boldsymbol{n}_{\text{vsn}}. \tag{37}$$

Accordingly, we append Eq. (37) to system (19), and we define vision-augmented measurement and noise vectors

$$\boldsymbol{z} \doteq (\boldsymbol{v}_{\text{gps}}^T, \boldsymbol{r}_{\text{gps}}^T, h_{\text{sonar}}, \boldsymbol{m}_{\text{magn}}^T, \ldots, \boldsymbol{d}_{\text{vsn}}^T, \boldsymbol{d}_{\text{vsn}}^{T\prime}, \ldots)^T, \tag{38a}$$

$$\boldsymbol{\mu} \doteq (\boldsymbol{n}_{v_{\text{gps}}}^T, \boldsymbol{n}_{r_{\text{gps}}}^T, n_{\text{sonar}}, \boldsymbol{n}_{\text{magn}}^T, \ldots, \boldsymbol{n}_{\text{vsn}}^T, \boldsymbol{n}_{\text{vsn}}^T, \ldots,)^T. \tag{38b}$$

3.3 Outlier Rejection

In this work, the identification of outliers is performed by using Eq. (31). The idea is to first estimate the expected apparent velocity of each candidate feature point on the basis of the currently estimated motion state of the vehicle. If $\hat{\boldsymbol{x}}(t_k)$ and $\hat{\boldsymbol{x}}(t_{k+1})$ are the available estimates at the previous and current time steps, respectively, the expected apparent velocity is

$$\begin{aligned}\dot{\hat{\boldsymbol{p}}}^{\mathcal{C}} = -\hat{\boldsymbol{M}}(t_{1/2})\boldsymbol{C}^T\big(&\hat{\boldsymbol{R}}(t_{1/2})^T\hat{\boldsymbol{v}}_B^{\mathcal{E}}(t_{1/2}) \\ &+ \hat{\boldsymbol{\omega}}^{\mathcal{B}}(t_{1/2}) \times (\boldsymbol{c}^{\mathcal{B}} + \boldsymbol{C}\boldsymbol{d}^{\mathcal{C}})\big),\end{aligned} \tag{39}$$

where $a(t_{1/2}) \doteq (a(t_k) + a(t_{k+1}))/2$ is a mid-step value.

Next, the apparent velocity is computed by measuring the optical flow of the feature point

$$\dot{\boldsymbol{p}}_h^{\mathcal{C}} = \frac{\boldsymbol{p}^{\mathcal{C}}(t_{k+1}) - \boldsymbol{p}^{\mathcal{C}}(t_k)}{h}, \tag{40}$$

where $h \doteq t_{k+1} - t_k$. Clearly, neglecting the reconstruction error on the current state of motion of the vehicle and on the measurement of the optical flow, the two quantities $\dot{\hat{\boldsymbol{p}}}^{\mathcal{C}}$ and $\dot{\boldsymbol{p}}_h^{\mathcal{C}}$ should match if the tracked feature point is fixed with respect to the inertial frame of reference; on the other hand, a mismatch between the two measures can be taken to indicate an outlier.

The coherence check between the two measures is performed as follows. First, we discard from the check all points that have optical flow vectors which are two small with respect to the pixel width, i.e. $\|\dot{\hat{\boldsymbol{p}}}^{\mathcal{C}}\|, \|\dot{\boldsymbol{p}}_h^{\mathcal{C}}\| < \varepsilon_L$, a typical value being $\varepsilon_L = 10w$. Next, a candidate point is considerate an outlier if either one or the other of the following criteria on magnitude and direction is met:

$$\left| \|\dot{\hat{\boldsymbol{p}}}^{\mathcal{C}}\| - \|\dot{\boldsymbol{p}}_h^{\mathcal{C}}\| \right| > \varepsilon_M, \tag{41a}$$

$$\left| \frac{\dot{\hat{\boldsymbol{p}}}^{\mathcal{C}}}{\|\dot{\hat{\boldsymbol{p}}}^{\mathcal{C}}\|} \cdot \frac{\dot{\boldsymbol{p}}_h^{\mathcal{C}}}{\|\dot{\boldsymbol{p}}_h^{\mathcal{C}}\|} \right| > \varepsilon_\theta, \tag{41b}$$

where ε_M and ε_θ are threshold values.

4 RESULTS AND APPLICATIONS

To illustrate the performance of the proposed vision-augmented navigation system, we consider two different simulated experiments, which use a flight mechanics model of a small RUAV, a model of the scene, a model of the camera, and models of the noise for all sensors. Tests in a simulation environment allow for the determination of the accuracy of the system, since the motion of the helicopter is known and reconstruction errors can be exactly measured. Testing in the field is in progress, and results will be reported soon in a forthcoming publication.

In the first experiment the helicopter is conducting a pirouette maneuver, i.e. it travels on a circular trajectory by maintaining a heading such that the vehicle nose points towards the center of the circle. The lateral speed is 2 m/s and the circle radius is 20 m. About one hundred small spherical objects are located within the circle, and the KLT algorithm operates at a frequency of 1 Hz, values chosen to show conservative results. The scene is very simple: since the vehicle travels around the spheres pointing towards them, all of them (except for occlusions) remain visible to the cameras and there are of the order of about 20 tracked features at all times. To show the capability of the system in ensuring an accurate estimation of the vehicle states even without GPS, we simulated a temporary signal loss between $t = 100$ s and $t = 200$ s.

Figure 7 shows the time history of the observed and true linear velocity components of the helicopter for five complete revolutions of the vehicle around the circle. The two vertical lines at $t = 100$ s and $t = 200$ s indicate the time instants when the GPS signal is lost and reacquired, respectively. The true reference components are shown using a dashed line, while the ones observed with the vision-augmented system using a solid line; the figure also shows using a dotted line quantities observed by the inertial navigation unit without vision-augmentation (Eqs. 20). Figure 8 shows the helicopter observed and real attitude in terms of the roll, pitch and yaw angles. Similar results were obtained for the component of the vehicle angular velocity, which are however not shown here for brevity.

It appears that, after a transient of about 50 s, the estimates match very well the true reference values for all states. Furthermore, the loss of GPS signal does not cause any appreciable degradation in the quality of the vision-augmented estimates nor the presence of any transient following these discrete events. On the other hand, the loss of GPS signal causes a rapid divergence of the estimates of the standard non-vision-augmented observer.

In the second experiment, we increase the fidelity of the simulation by using a more realistic model of

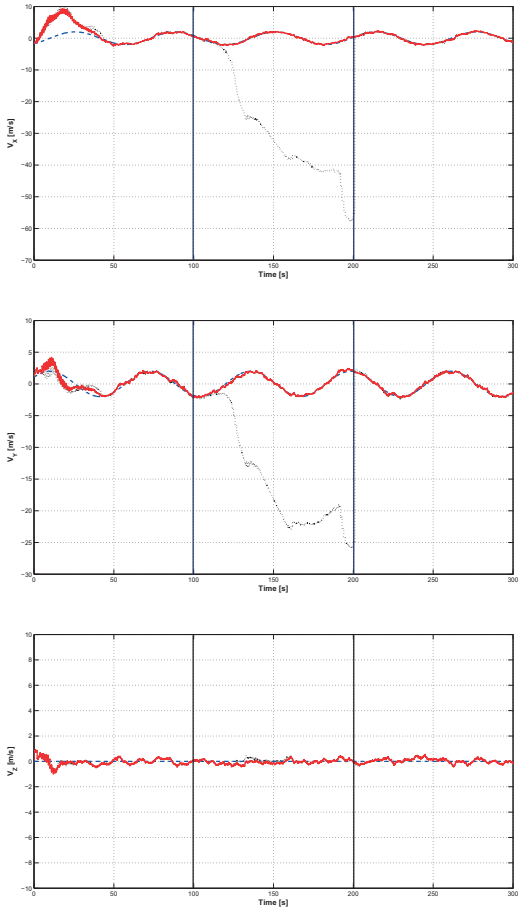

Figure 7: Pirouette maneuver. Linear velocity inertial components $\boldsymbol{v}^{\mathcal{E}} = (V_x, V_y, V_z)^T$ (from top to bottom). Dashed line: true states; solid line: vision-augmented observer; dotted line: standard non-augmented observer.

the scene. In this case, the helicopter performs a low level flight in an urban canyon within a small village, composed of houses and several other objects with realistic textures (see Fig. 9, top). The scene environment and image acquisition was simulated with the software `Gazebo` [2], a complete multi-robot simulator for outdoor environments of which we only used the graphical rendering and camera image grabbing features. The helicopter flies at an altitude of 2 m over the terrain among the houses of the village, following a rectangular path at a constant speed of 2 m/s. A view from above of the village and of the helicopter trajectory is shown in the bottom part of Fig. 9. The vehicle preforms three complete loops of the village path, and during the second loop it operates without GPS.

Figure 10 shows the true and estimated linear velocity inertial components; Fig. 11 shows the true and estimated roll angle and the roll and yaw angular velocity body-attached components, respectively from

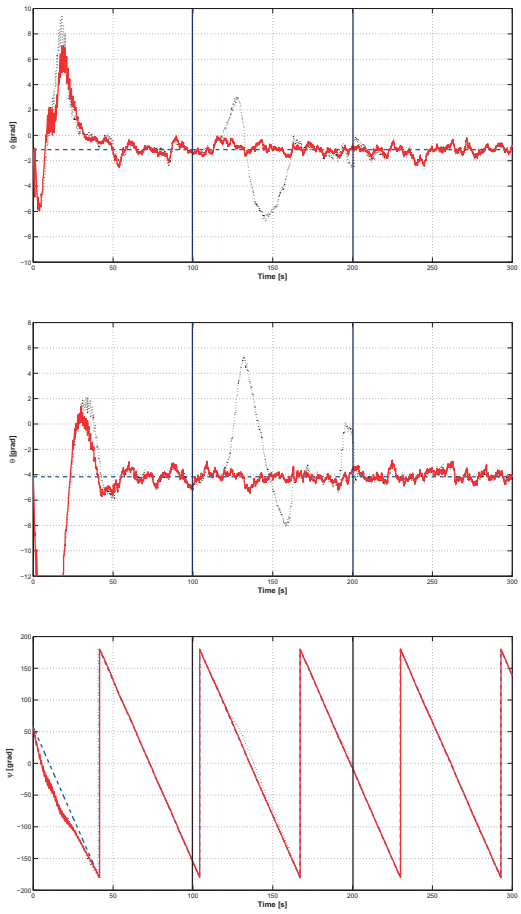

Figure 8: Pirouette maneuver. Vehicle attitude expressed in terms of the roll (ϕ), pitch (θ) and yaw (ψ) angles (from top to bottom). Dashed line: true states; solid line: vision-augmented observer; dotted line: standard non-augmented observer.

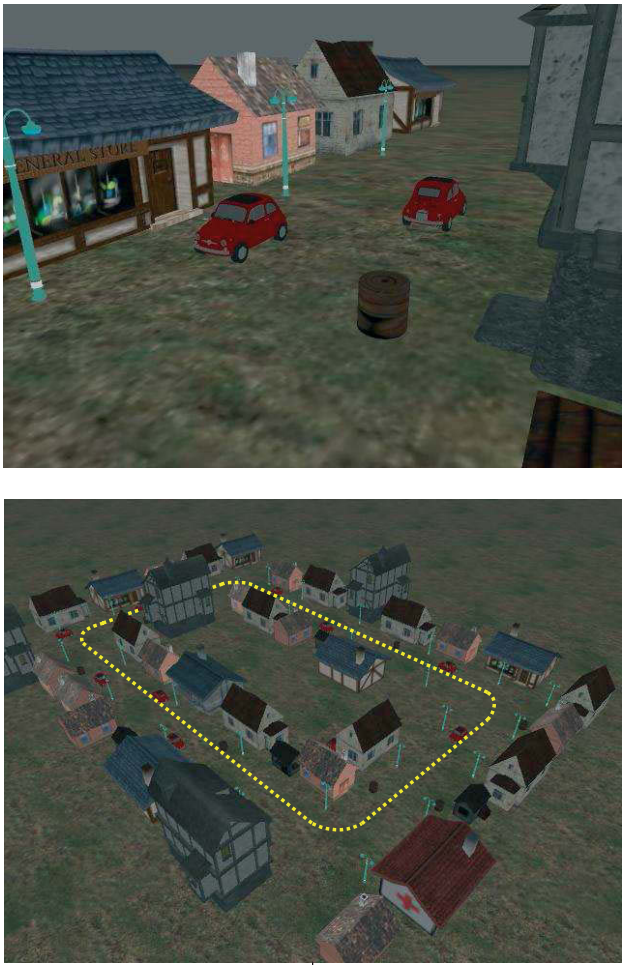

Figure 9: Flight in an urban environment. At top, view of detail of houses and other objects with texture; at bottom, view from above of the village and of the helicopter flight path.

top to bottom. For the angular velocity components, we show a zoom of the time histories between 100 s and 200 s, i.e. when the system is operating without GPS signal, to better appreciate the match between real and observed quantities. After an initial phase where the Kalman filter warms up, again of about 50 s in length, the vehicle states are reconstructed with good accuracy with all sensors on for the first path loop. Then, during the second loop, the GPS signal is lost, and the inertial unit continues its operation without it. At the end of the second loop the GPS signal is reacquired. The two events are indicated in the figures by think vertical solid lines.

Here again it appears that the vision-augmented observer is capable of producing good quality estimates of the vehicle states throughout the simulation. The loss of the GPS signal seems to a have a small effect on the quality of the estimates, which is especially noticeable in the observed linear velocity components, while the attitude and angular velocity estimates seem to be unaffected. In this example, the number of tracked features averaged about 20 at all times, a relatively small number which was chosen to show conservative results, and better accuracy can be obtained by increasing the number of tracked features.

5 CONCLUSIONS

We have presented a new approach to inertial navigation, which incorporates vision sensors in the system and fuses them together with other non-vision-based sensors. In fact, we have argued that the information provided by stereo cameras on-board the vehicle can and should not only be used for map-building (obstacle and target recognition), but also to enhance the quality of the state estimates which must be provided to the on-board flight control system. In our approach, scene points are tracked between simultaneous images and across time steps, yielding information on

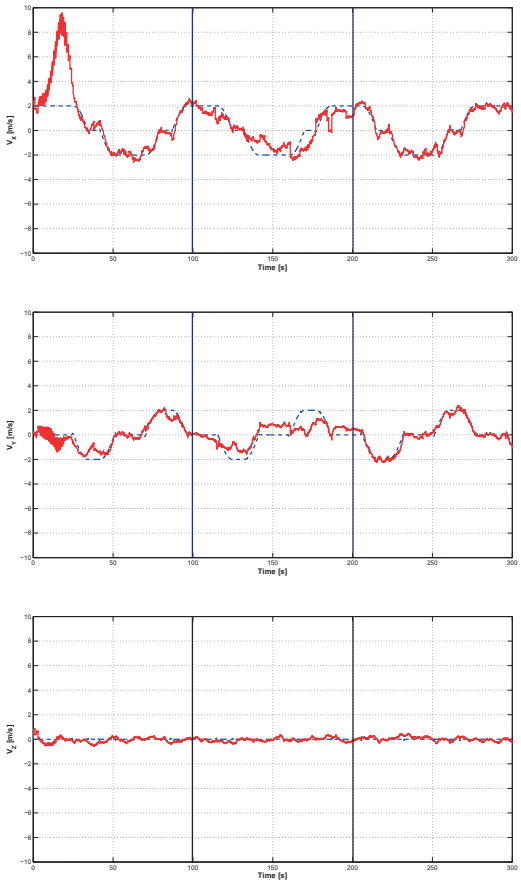

Figure 10: Flight in an urban environment. Linear velocity inertial components $v^{\mathcal{E}} = (V_x, V_y, V_z)^T$ (from top to bottom). Dashed line: true states; solid line: vision-augmented observer.

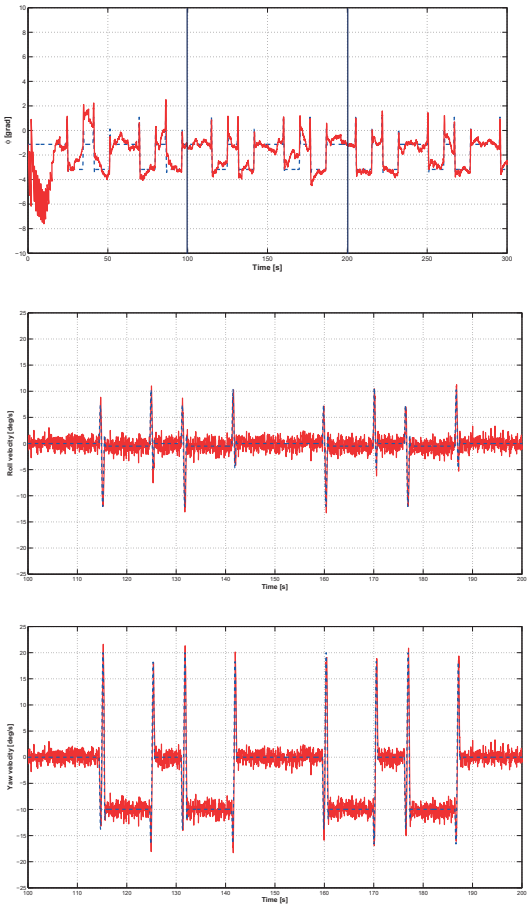

Figure 11: Flight in an urban environment. Vehicle roll angle (ϕ), and roll (p) and yaw (r) angular velocity body components (from top to bottom). Dashed line: true states; solid line: vision-augmented observer.

the state of motion of the vehicle which augments the information provided by other sensors. The implementation of the new system is relatively straightforward, since it is based on a standard measurement fusion algorithm and therefore it is easily integrated with an existing non-vision-based navigation system.

Testing performed in a simulation environment comprising models of the vehicle, of the scene, of all sensors including the cameras and of their noise characteristics, have shown the following facts:

- The incorporation of vision sensors in the inertial navigation system improves the observability of the vehicle linear and angular velocity vectors and of the vehicle attitude. When enough feature points can be tracked, no transients in the quality of the computed estimates can be noticed even after the discrete events of GPS loss or reacquisition. Clearly, without GPS signal the absolute position of the vehicle will drift away since it can not be observed by a vision system, but this not crucial since relative position information is probably more valuable in confined environments.

- The performance of the system degrades when not enough points can be tracked between stereo images and across time steps, for example because of high levels of noise in the images, too many outliers, not enough visually reach information, etc. Hence, the feature point tracking algorithm used by the inertial system plays a crucial role for the effectiveness of the proposed procedure.

- The accuracy of the vision sensors depends on the distance between tracked features and vehicle. Since relatively low precision cameras were used here, the positive effects brought to the state estimation problem by vision sensors are felt only when operating in the close proximity (distances of the order of meters-few tens of meters) of obstacles. These are however the typical distances expected during the operation of rotorcraft vehicles in complex urban environ-

ments; hence, it is felt that the proposed method has good potential for practical applicability in the field.

- The computational cost of the sensor fusion approach is compatible with the requirements of a real-time implementation with the necessary bandwidth for loop closure on a small helicopter using standard computing hardware, e.g. a mid level PC/104 small-form-factor computer as the one used here.

There are several aspects of the problem which require further investigation, and which will be the object of our attention in the very next future.

First, testing is in progress to evaluate the performance of the vision-augmented inertial navigation unit in the field, and the results of the experimental campaign will be documented soon in a forthcoming publication.

Furthermore, we have noticed that the manual tuning of the filter process and measurement noise covariances, crucial parameters which govern the convergence behavior of the observer, is not always straightforward to accomplish. To alleviate the need for careful tuning of such parameters, we are implementing an adaptive filtering method, which keeps a buffer of past values to automatically extract noise samples on-line during filtering [7].

The proposed procedure is modular and expandable: other sensors could be incorporated in the inertial navigation unit in a relatively straightforward manner, an interesting candidate among several others being a stereo laser-scanner.

References

[1] Point Grey Research, 12051 Riverside Way, Richmond, BC, Canada V6W 1K7, `http://www.ptgrey.com`.

[2] The Player/Stage/Gazebo Project, `http://playerstage.sourceforge.net`.

[3] Dittrich, J.S. and Johnson, E.N., “Multi-Sensor Navigation System for an Autonomous Helicopter,” Proceedings of the 21st Digital Avionics Systems Conference, Irvine, CA, USA, October 27–31, 2002.

[4] Gavrilets, V., Shterenberg, A., Dahleh, M.A. and Feron, E., “Avionics System for a Small Unmanned Helicopter Performing Aggressive Maneuvers,” Proceedings of the 19th Digital Avionics Systems Conferences, Philadelphia, PA, USA, October 7–13, 2000.

[5] Jin, H., Favaro, P. and Soatto, S., “A Semi-Direct Approach to Structure from Motion,” *The Visual Computer*, Vol. 19, No. 6, 2003, pp. 377–394.

[6] Lucas, B.D. and Kanade, T., “An Iterative Image Registration Technique with an Application to Stereo Vision,” Proceedings of IJCAI81, 1981, pp. 674–679.

[7] Myers, K.A. and Tapley, B.D., “Adaptive Sequential Estimation with Unknown Noise Statistics,” IEEE Transactions on Automatic Control, Vol. 21, 1976, pp. 520–523.

[8] Reif, K. and Unbehauen, R., “The Extended Kalman Filter as an Exponential Observer for Nonlinear Systems,” IEEE Transactions on Signal Processing, Vol. 47, 1999, pp. 2324–2328.

[9] Savini, B., *Development of a Test-Bed for Simulation and Control of an Unmanned Rotorcraft*, Ph.D. Thesis, Politecnico di Milano, Dipartimento di Ingegneria Aerospaziale, 2007.

[10] Shi, J. and Tomasi, C., “Good Features to Track,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600.

[11] Simon, D., *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, John Wiley & Sons, Inc., 2006.

[12] Willner, D., Chang, C.B. and Dunn, K.P., “Kalman Filter Algorithms for a Multi-Sensor System,” Proceedings of the IEEE Conference on Decision and Control, 1976, pp. 570–574.