XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ - BRAZIL

Visual-Inertial Navigation Applied to a Motorboat

Lucas Braga Proença[1], Davi Antônio dos Santos[2], Stiven Schwanz Dias[1]

[1]*Embraer S.A. lucas.proenca@embraer.com.br, stiven.dias@embraer.com.br*
[2]*Aeronautics Institute of Technology (ITA) davists@ita.br*

Abstract. This work is concerned with the problem of state estimation for a small, nine-meter-long motorboat navigating in a GPS-denied environment. It proposes a visual-inertial navigation method based on a continuous-discrete formulation of the extended Kalman filter (EKF) for estimating the state vector of the motorboat, which includes its position, velocity and attitude, as well as the sensors' biases. The filter relies on inertial measurements, provided by an accelerometer and a rate gyro, and on position measurements obtained by a panoramic camera system. By employing Monte Carlo simulation, this work concludes that the proposed method has an estimation performance comparable to GPS/INS approaches.

Keywords: sensor fusion, state-estimation, visual-inertial navigation

1 Introduction

Sensor fusion is a computational process where multiple physical sensors' measurements are combined to produce more accurate information of the environment, as defined by Elmenreich [1]. This field of research has several applications, such as navigation, image processing, data science, artificial intelligence, and robotics. This work, in particular, employs sensor fusion techniques for navigation.

Inertial sensors constitute the basis of a navigation system. The so-called Inertial Measurement Unit (IMU) is a combination of sensors such as accelerometers, rate gyro and magnetometers in one single device, which is often integrated to a computer that calculates body's position, velocity and orientation. This integrated system is commonly referred as Inertial Navigation System (INS), such as in Farrell and Barth [2]. Woodman [3] demonstrates that these systems are usually not sufficient for navigation because they are associated to drifting errors. For that, navigation systems often combine INS with Global Positioning System (GPS), as it is explained by Farrell [4].

Although the blended GPS/INS is suitable for most navigation applications, KNIGHT [5] and MUELLA et al. [6] show that GPS signal is considerably affected by atmospheric conditions, electromagnetic phenomena and by obstruction in the so-called GPS-denied environments. For that reason, some applications require a navigation approach independent from GPS, that is, robust navigation approaches such as the ones detailed by BACHRACH et al. [7], which are able to deal with GPS signal unavailability and to external interference such as jamming and spoofing. In this context, this work proposes a novel navigation approach: the vehicle state (position, velocity and attitude) is estimated by using Bayesian filtering, which combines inertial measurements and vehicle's pose-related information coming from a camera system. This kind of approach is commonly referred as *visual-inertial navigation* and is often employed in Robotics.

In the context of the *Aerial Robotics Laboratory*, at the Aeronautics Institute of Technology (ITA - Brazil), several works have been developed in the field of robot navigation, especially for micro aerial vehicles (MAVs), which have been increasingly used for their capability of operating virtually anywhere, even indoors. The visual-inertial navigation is presented as a promising approach in this cases, taking advantage of the rich information on the environment and allowing increased navigation autonomy, obstacle avoidance, target tracking and mapping.

A navigation algorithm based on the Extended Kalman Filter (EKF) is presented by BALLET et al. [8] for low-cost MAVs, which uses measurements from inertial sensors, a downward-facing camera and an ultrasonic range sensor, navigating through an environment with known visible landmarks on the ground. In the same field of research, BARBOSA et al. [9] presented a visual-inertial navigation method for estimating position, velocity and heading of a MAV based on the Unscented Kalman Filter (UKF) and using low-cost inertial sensors and a camera. Also, it shows a performance comparison between the proposed method and a similar EKF-based method, concluding that the UKF has a slightly better performance with a higher computational effort.

In a more recent work, ASFORA [10] performed an investigation of the Ensemble Kalman Filter (EnKF) for the visual-aided navigation of multirotor MAVs. In this work, the EnKF is chosen for its ability to deal with a very large number of variables and high volume of data. The filter is evaluated for attitude determination, navigation, and simultaneous navigation and tracking. The performance of the EnKF filter is compared to more traditional algorithms for nonlinear systems, leading to the conclusion that the EnKF presents similar performance with respect to the EKF and a computational burden equivalent to the UKF. Besides, the simulation results suggest that the EnKF should be preferred in more complex navigation problems because of its scalability and tolerance to uncertainty.

In line with the referred works, this paper proposes a visual-inertial navigation method based on discrete and continuous-time filtering techniques to estimate in a Bayesian sense the state vector of a small motorboat, that is, its position, velocity, attitude and sensors' biases. The filter relies on inertial measurements, provided by an accelerometer and a rate-gyro, and on landmark measurements provided by a panoramic camera system.

2 Problem Definition

This section presents the mathematical modeling for the problem of motorboat navigation within a known environment.

2.1 Coordinate Systems

Consider the cartesian coordinate systems (CCS's) illustrated in Fig. 1. Let $\mathbb{S}_G = \{x_G, y_G, z_G\}$ denote an inertial reference system in the 3D space, with its origin in a point G, on ground, and with z_G aligned with the local vertical. Also, let $\mathbb{S}_B = \{x_B, y_B, z_B\}$ denote a local coordinate system which is fixed to the motorboat, considered to be lying on the motorboat's inertial sensors suite, with z_B parallel to z_G. Lastly, let $\mathbb{S}_C = \{x_C, y_C, z_C\}$ to be lying on the camera's optical center, with z_C pointing to the camera forwards, and translated by $r^{C/B}$ with respect to $\mathbb{S}_B$. $\mathbb{S}_B$ and $\mathbb{S}_C$ are attached to the boat and are solidary to its movement. Besides the coordinate systems in 3D space, we define a bi-dimensional coordinate system $\mathbb{S}_I = \{x_I, y_I\}$ for representing the position of the observed objects in the camera's *image plane*.

In any equation within this work, the reader should consider that, unless a different CCS is explicit, vectors are in $\mathbb{S}_G$.

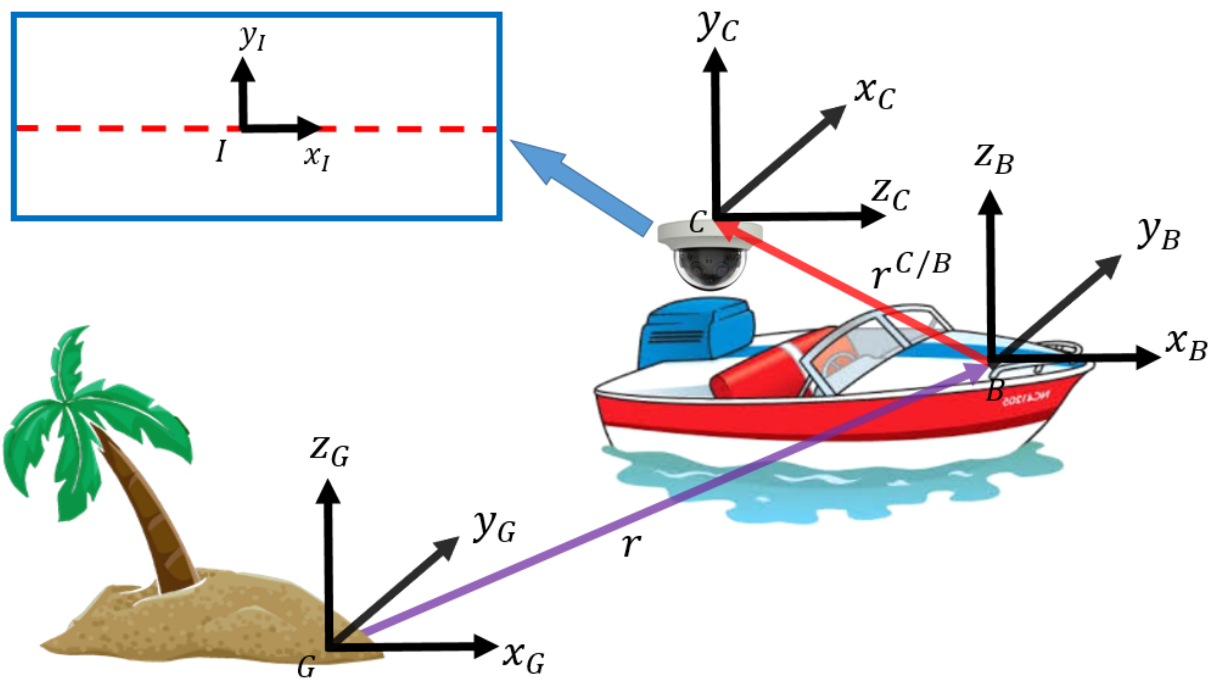

Figure 1. Coordinate systems defined for the navigation problem of a motorboat using landmark camera measurements.

2.2 Vehicle Kinematics

The body's accelerations, $\mathbf{a}_B \in \mathbb{R}^3$, and turn rates, $\boldsymbol{\omega_B} \in \mathbb{R}^3$, measured in $\mathbb{S}_B$ are obtained by the equations

$$\mathbf{a}_B = \mathbf{D}_G^B(\boldsymbol{\alpha})\, \dot{\mathbf{v}}, \tag{1}$$

$$\boldsymbol{\omega}_B = \mathbf{A}(\boldsymbol{\alpha})^{-1}\, \dot{\boldsymbol{\alpha}} \tag{2}$$

where $\mathbf{v}$ is the vehicle's velocity; $\boldsymbol{\alpha} \in \mathbb{R}^3$ is the vehicle's attitude (roll, pitch and yaw angles); $\mathbf{D}_G^B(\boldsymbol{\alpha})$ is the attitude matrix of $\mathbb{S}_G$ with respect to $\mathbb{S}_B$; and $\mathbf{A}(\boldsymbol{\alpha})$ is the parameterization of the attitude matrix $\mathbf{D}_G^B$ by Euler angles in rotation sequence 1-2-3.

2.3 Inertial Sensors

Assuming that the accelerometer and gyroscope are aligned with $\mathbb{S}_B$, the inertial measurements are simulated as

$$\tilde{\mathbf{a}}^B = \mathbf{a}^B + \boldsymbol{\beta}^a + \mathbf{w}^a, \tag{3}$$

$$\tilde{\boldsymbol{\omega}}^B = \boldsymbol{\omega}^B + \boldsymbol{\beta}^g + \mathbf{w}^g, \tag{4}$$

$$\dot{\boldsymbol{\beta}}^a = \mathbf{w}_{\boldsymbol{\beta}^a}, \tag{5}$$

$$\dot{\boldsymbol{\beta}}^g = \mathbf{w}_{\boldsymbol{\beta}^g}, \tag{6}$$

where $\mathbf{w}^a$ and $\mathbf{w}^g$ are the measurement noises, modeled as zero-mean white Gaussian noise with covariances $\mathbf{Q}_{\mathbf{w}^a}$ and $\mathbf{Q}_{\mathbf{w}^g}$, respectively; $\boldsymbol{\beta}^a \in \mathbb{R}^3$ and $\boldsymbol{\beta}^g \in \mathbb{R}^3$ are the sensors' biases, modeled as Wiener processes with initial conditions $\boldsymbol{\beta}^a(0) = \boldsymbol{\beta}_0^a$ and $\boldsymbol{\beta}^g(0) = \boldsymbol{\beta}_0^g$, where $\mathbf{w}_{\boldsymbol{\beta}^a}$ and $\mathbf{w}_{\boldsymbol{\beta}^g}$ are zero-mean white Gaussian noise with covariances $\mathbf{Q}_{\boldsymbol{\beta}^a}$ and $\mathbf{Q}_{\boldsymbol{\beta}^g}$, respectively.

2.4 Camera System

Let $\mathbf{p}^i$ be the position of the ith observed landmark in $\mathbb{S}_G$, and $\mathbf{s}^i$ be the position of the ith landmark, in S_G, with respect to $\mathbb{S}_C$. From the scheme presented in Fig. 2, we have

$$\mathbf{s}_C^i = \mathbf{D}_B^C\, \mathbf{D}_G^B\, \underbrace{(\mathbf{p}^i - \mathbf{r} - (\mathbf{D}_G^B)^T\, \mathbf{r}^{C/B})}_{\mathbf{s}^i}. \tag{7}$$

A camera model is necessary to translate the position of the landmarks into a camera measurement, $\mathbf{y}_k^i$, which is defined in the camera's image plane. This work uses the *pinhole* camera model proposed by FORSYTH and PONCE [11], which uses a first-order approximation of the mapping from a 3D scene to a 2D image. The measured landmark position provided by the *pinhole* camera model is the one proposed by BALLET et al. [8]

$$\mathbf{y}_k^i = \frac{f}{\mathbf{e}_3^T \mathbf{s}_C^i} \begin{bmatrix} \mathbf{e}_1^T \mathbf{s}_C^i \\ \mathbf{e}_2^T \mathbf{s}_C^i \end{bmatrix}, \tag{8}$$

where k is a discrete time instant $t = kT$, with T being the camera's sampling time; $\mathbf{e}_i$'s are the canonic unit vectors[1].

The *pinhole* camera model brings, however, some limitations. The model works well only for narrow field-of-views (FOVs), since it has a singularity for $\mathbf{e}_3^T \mathbf{s}_C$, which represents the projection of the object's position into the camera's optical axis. In physical terms, this singularity happens because, for an object located in an angular position perpendicular to the camera's optical axis, its projection does not exist.

This problem can be worked around by defining eight different camera models with a 45-degree FOV, such that the singularity is avoided, and referred as $j = \{1, 2, ..., 8\}$. Let j be the j-th camera model and consider that the second camera ($j = 2$) is rotated by 45 degrees about the y_C axis, with respect to the first camera, and that the

[1] $\mathbf{e}_1^T = [1, 0, 0]$, $\mathbf{e}_2^T = [0, 1, 0]$, $\mathbf{e}_3^T = [0, 0, 1]$

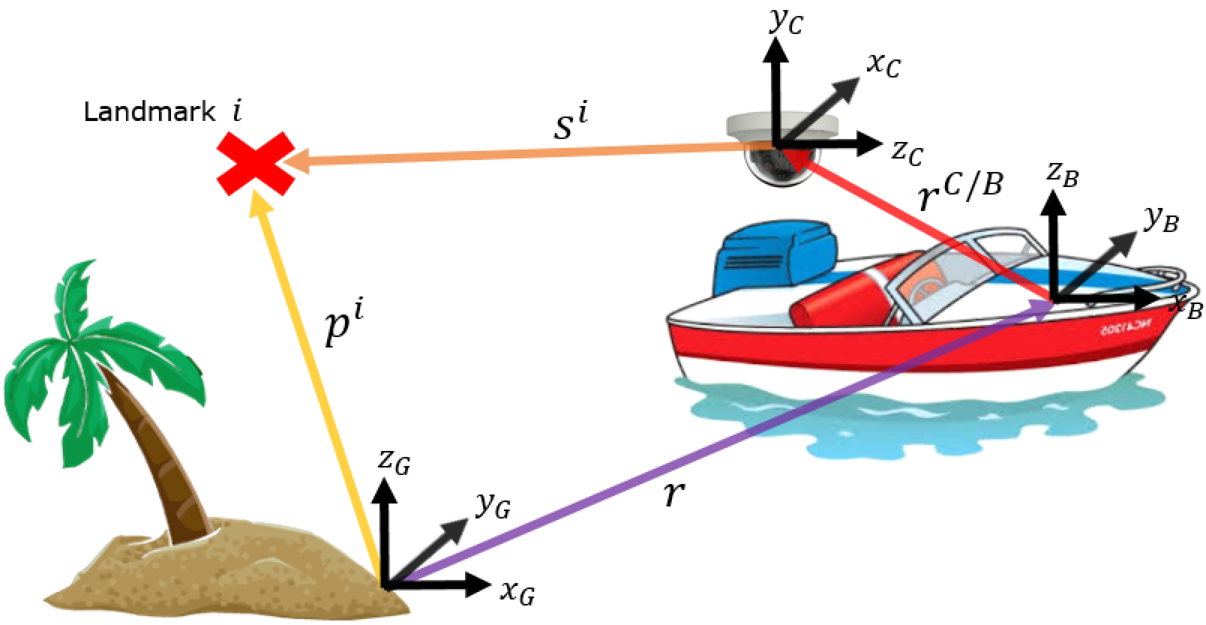

Figure 2. Position of a landmark with respect to the camera system.

third camera is rotated by 45 degrees with respect to the second one, and so on. Then, for the j-th camera model we have

$$\mathbf{s}_C^i = \mathbf{D}_B^{C^j} \left(\mathbf{D}_G^B \left(\mathbf{p}^i - \mathbf{r} \right) - \mathbf{r}^{C/B} \right). \tag{9}$$

with $\mathbf{D}_B^{C^j} = R_{y_C}\left((j-1)\, \frac{\pi}{4} \right) \mathbf{D}_B^C$, where $R_{y_C}(\gamma)$ represents a rotation of γ radians about the y_C axis.

3 Problem Solution

This section presents the proposed solution for the navigation problem.

3.1 Kinematics

For obtaining an observable model, the filter proposed in this work estimates the position of the motorboat in the xy plane and neglects the component z of its position, which represents the height of the vehicle with respect to sea level. For the context of navigation of a motorboat, neglecting the z component is acceptable, once this varies slightly while the vessel travels. Then, let $\mathbf{r}'$ and $\mathbf{v}'$ to be the motorboat's two-dimensional position and velocity, respectively, and $\mathbf{a}'_B$ to be the motorboat's linear accelerations in $\mathbb{S}_B$. Then, we have

$$\dot{\mathbf{r}}' = \mathbf{v}', \tag{10}$$

$$\dot{\mathbf{v}}' = \mathbf{M}_{2D}\, \mathbf{D}_G^B(\boldsymbol{\alpha})^T \begin{bmatrix} \mathbf{a}'_B \\ 0 \end{bmatrix} \tag{11}$$

where $\mathbf{M}_{2D}$ is the transformation that projects a $3D$ coordinate into the xy plane.

3.2 State-Equation

By using the kinematic model described by eq. (10) and eq. (11), along with the inertial sensors' measures modeled by eq. (3), eq. (4) and eq. (5), we obtain the state equation given by

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{G}(\mathbf{x}(t))\mathbf{w}(t), \tag{12}$$

where

$$\mathbf{x}(t) \triangleq \left[(\mathbf{r}')^T \; (\mathbf{v}')^T \; \boldsymbol{\alpha}^T \; \boldsymbol{\beta}^{a\prime T} \; (\boldsymbol{\beta}^g)^T\right]^T \in \mathbb{R}^{12},$$

$$\mathbf{u}(t) \triangleq \left[(\tilde{\mathbf{a}}'_B)^T \; (\tilde{\boldsymbol{\omega}}_B)^T\right]^T \in \mathbb{R}^5,$$

$$\mathbf{w}(t) \triangleq \left[(\mathbf{w}^a)^T \; (\mathbf{w}^g)^T \; (\mathbf{w}^{\beta^a})^T \; (\mathbf{w}^{\beta^g})^T\right]^T \in \mathbb{R}^{10},$$

$$\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) \triangleq \begin{bmatrix} \mathbf{v} \\ \mathbf{M}_{2D}\, \mathbf{D}_G^B(\boldsymbol{\alpha})^T [(\tilde{\mathbf{a}}_B - \boldsymbol{\beta}^a)\; 0]^T \\ \mathbf{A}(\boldsymbol{\alpha})(\tilde{\boldsymbol{\omega}}^B - \boldsymbol{\beta}^g) \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix},$$

$$\mathbf{G}(t) \triangleq \begin{bmatrix} \mathbf{0}_{2\times2} & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times3} \\ -\mathbf{M}_{2D}\, \mathbf{D}_G^B(\boldsymbol{\alpha})^T & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times3} \\ \mathbf{0}_{3\times2} & -\mathbf{A}(\boldsymbol{\alpha}) & \mathbf{0}_{3\times2} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times3} & \mathbf{I}_{2\times2} & \mathbf{0}_{2\times3} \\ \mathbf{0}_{3\times2} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times2} & \mathbf{I}_{3\times3} \end{bmatrix}.$$

3.3 Measurement Equation

The coordinates of the ith landmark in the camera's coordinate system, $\mathbf{s}_C^i$, are taken from

$$\mathbf{s}_C^i = \mathbf{D}_B^C \left[\mathbf{D}_G^B \left(\mathbf{p}^i - \mathbf{r}\right) - \mathbf{r}^{C/B}\right]. \tag{13}$$

However, as the navigation filter proposed in this work does not estimate the z component of the motorboat's position, it works out an adaptation to eq. (13) by making $\mathbf{r}_k \triangleq \begin{bmatrix} \mathbf{r}' \\ 0 \end{bmatrix}$. The coordinates of the ith landmark in $\mathbb{S}_I$ can be obtained by using the *pinhole* camera model, as in eq. (8). Let $\mathbf{s}_I^i$ be the position of the ith landmark in $\mathbb{S}_I$. Then, we define a discrete-time measurement model given by

$$\mathbf{h}_{k+1}^i(\mathbf{x}) \triangleq \mathbf{s}_{I\,k+1}^i, \tag{14}$$

and obtain a measurement equation for the ith landmark, given by

$$\mathbf{y}_{k+1}^i = \mathbf{h}_{k+1}^i(\mathbf{x}) + \mathbf{v}_{k+1}^i, \tag{15}$$

where $\mathbf{v}_{k+1}^i$ is the measurement noise, assumed to be a zero-mean Gaussian noise with covariance matrix $\mathbf{R}_{k+1}$.

In order to simulate practical limitations of the landmark identification, the simulation considers that only landmarks within five kilometers from the camera are detected.

3.4 Visual-Inertial Navigation Filter

The filter proposed in this work is a Continuous-Discrete Extended Kalman Filter (CDEKF) with sequential update detailed by Bar-Shalom et al. [12]. The sequential update works around the problem of variable number of available measurements during the estimation, since not all landmarks are supposed to be identified at a specific time instant either because they are out of camera's range or not identified by image processing algorithms. Algorithm 1 presents the implemented filter.

Algorithm 1 Visual-Inertial Navigation Filter

$\hat{\mathbf{x}}_{1|1} \leftarrow \bar{\mathbf{x}}$
$\mathbf{P}_{1|1} \leftarrow \bar{\mathbf{P}}$
$\mathbf{Q}(t) \leftarrow diag(\mathbf{Q}_a, \mathbf{Q}_g, \mathbf{Q}_{\beta^a}, \mathbf{Q}_{\beta^g})$
for $k \geq 1$ **do**
    $\mathbf{F}(t) \triangleq \left.\frac{\partial \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t))}{\partial \mathbf{x}}\right|_{\mathbf{x}(t)=\hat{\mathbf{x}}(t)}$
    Integrate from t_k to t_{k+1}:
        $\dot{\hat{\mathbf{x}}} = \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}_k)$
        $\dot{P}(t) = \mathbf{F}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}(t)^T + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}(t)^T$
    with initial conditions $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$
    $\hat{\mathbf{x}}_{k+1|k+1} \leftarrow \hat{\mathbf{x}}_{k+1|k}$
    $\mathbf{P}_{k+1|k} \leftarrow \mathbf{P}(t_{k+1})$
    $\mathbf{q}_k \leftarrow$ labels of identified landmarks, at instant *k*
    for $j := 1, j <= length(\mathbf{q}_k)$ **do**
        $i \leftarrow \mathbf{q}_k(j)$
        $\mathbf{H}_{k+1} = \left.\frac{d\mathbf{h}_{k+1}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k+1|k}}$
        $\hat{\mathbf{y}}^i_{k+1|k} = \mathbf{h}^i_{k+1}(\hat{\mathbf{x}}_{k+1|k})$
        $\mathbf{P}^Y = \mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}^T_{k+1} + \mathbf{R}_{k+1}$
        $\mathbf{P}^{XY} = \mathbf{P}_{k+1|k}\mathbf{H}^T_{k+1}$
        $\mathbf{K}_{k+1} \leftarrow \mathbf{P}^{XY}_{k+1|k}(\mathbf{P}^Y_{k+1|k})^{-1}$
        $\hat{\mathbf{x}}_{k+1|k+1} \leftarrow \hat{\mathbf{x}}_{k+1|k+1} + \mathbf{K}_{k+1}\,(\mathbf{y}^i_{k+1} - \hat{\mathbf{y}}^i_{k+1|k})$
        $\mathbf{P}_{k+1|k+1} \leftarrow \mathbf{P}_{k+1|k} - \mathbf{P}^{XY}(\mathbf{P}^Y)^{-1}(\mathbf{P}^{XY})^T$
    end for
end for

4 Conclusions

This work presented a visual-inertial navigation method for a small motorboat traveling in a GPS-denied environment. The method is a continuous-discrete formulation of the extended Kalman filter (EKF) which allows the estimation of the motoboat's position, velocity and attitude, as well as the sensors' biases.

The filter relies on inertial measurements provided by an accelerometer and a rate gyro, and on position measurements from a panoramic camera system. In order to evaluate the estimation performance of the filter, Monte Carlo simulation is performed over a real scenario where inertial sensors and camera system are simulated. The filter estimation is compared to ground-truth data recorded from the real sea trial, which includes the motorboat's position, velocity and attitude.

The obtained results suggest that the proposed filter performs equivalently to GPS/INS navigation approaches.

Acknowledgements. This work was supported by *Casimiro Montenegro Filho* Foundation in a partnership with EMBRAER and ITA.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Elmenreich, W., 2002. *An introduction to sensor fusion*. Institut fur Technische Informatik, Vienna University of Technology, Austria.

[2] Farrell, J. A. & Barth, M., 2007. *The Global Positioning System & Inertial Navigation*. Wiley-Interscience.

[3] Woodman, O. J., 2007. *An introduction to inertial navigation*. University of Cambridge - Computer Laboratory.

[4] Farrell, J. A., 2008. *Aided Navigation: GPS with High Rate Sensors*. McGraw-Hill, New York.

[5] KNIGHT, M. F., 2000. *Ionospheric Scintillation Effects on Global Positioning System Receivers*. PhD thesis, The University of Adelaide.

[6] MUELLA, M., DE PAULA, E., KANTOR, I., BATISTA, I., SOBRAL, J., ABDU, M., KINTNER, P., GROVES, K., & SMORIGO, P., 2008. Gps l-band scintillations and ionospheric irregularity zonal drifts inferred at equatorial and low-latitude regions. *Journal of Atmospheric and Solar-Terrestrial Physics, Elsevier*.

[7] BACHRACH, A., DE WINTER, A., He, R., HEMANN, G., PRENTICE, S., & ROY, N., 2010. Robust autonomous navigation in gps-denied environments. *IEEE Proceedings - International Conference on Robotics and Automation*.

[8] BALLET, R., SANTOS, D. A., & ZORAN, S., 2016. Position, velocity and heading estimation for unmanned aerial vehicles using camera and inertial sensors. *Congresso Brasileiro de Engenharia Mecânica (CONEM)*.

[9] BARBOSA, J. P. d. A., BALLET, R., SANTOS, D. A., & DIAS, S. S., 2016. Visual-inertial navigation algorithm for micro aerial vehicles using the unscented kalman filter. *Congresso Brasileiro de Automática (CBA)*.

[10] ASFORA, B. A., 2018. Inertial-visual navigation and tracking for unmanned aerial vehicles using the ensemble kalman filter.

[11] FORSYTH, D. & PONCE, J., 2011. *Computer Vision: A Modern Approach*. Prentice Hall.

[12] Bar-Shalom, Y., Kirubarajan, T., & Li, R. X., 2002. *Estimation with Applications to Tracking and Navigation*. Jonh Wiley & Sons.