Journal of Intelligent & Robotic Systems (2019) 94:203–217
https://doi.org/10.1007/s10846-018-0775-y

Monocular SLAM System for MAVs Aided with Altitude and Range Measurements: a GPS-free Approach

Sarquis Urzua[1] · **Rodrigo Munguía**[1] · **Antoni Grau**[2]

Received: 25 January 2017 / Accepted: 15 January 2018 / Published online: 2 February 2018

Abstract

A typical navigation system for a Micro Aerial Vehicle (MAV) relies basically on GPS for position estimation. However, for several kinds of applications, the precision of the GPS is inappropriate or even its signal can be unavailable. In this context, and due to its flexibility, Monocular Simultaneous Localization and Mapping (SLAM) methods have become a good alternative for implementing visual-based navigation systems for MAVs that must operate in GPS-denied environments. On the other hand, one of the most important challenges that arises with the use of the monocular vision is the difficulty to recover the metric scale of the world. In this work, a monocular SLAM system for MAVs is presented. In order to overcome the problem of the metric scale, a novel technique for inferring the approximate depth of visual features from an ultrasonic range-finder is developed. Additionally, the altitude of the vehicle is updated using the pressure measurements of a barometer. The proposed approach is supported by the theoretical results obtained from a nonlinear observability test. Experiments performed with both computer simulations and real data are presented in order to validate the performance of the proposal. The results confirm the theoretical findings and show that the method is able to work with low-cost sensors.

Keywords Micro aerial vehicles · Monocular SLAM · Visual-based navigation · GPS-denied

1 Introduction

The ability of navigating is an important requirement for an autonomous Micro Aerial Vehicle (MAV). The Global Positioning System (GPS) and Inertial Measurements Units (IMU), or their fused variant, the Inertial Navigation Systems (INS), represent the most common approaches for addressing the problem of the MAVs navigation. Available Attitude and Heading Reference Systems (AHRS) can be used for providing a reliable estimation of the attitude of the vehicle. On the other hand, the position estimation can still impose several challenges for different scenarios.

For instance, in urban canyons, indoor or cluttered environments, the availability of the GPS signal can be compromised. Even, when enough satellites are available, several sources of error affect the accuracy of GPS position measurements. The cumulative effect of each of these error sources is called the user-equivalent range error (UERE). In [1] these errors are characterized as a combination of slowly varying biases and random noise. In [2] it is stated that the total UERE is approximately 4.0 m (σ), from which 0.4 m (σ) correspond to random noise.

Those errors can make GPS-based navigation unreliable for several scenarios when precision maneuvers are required [3]. Therefore, additional sensory information is sometimes integrated into the system in order to improve accuracy.

All those issues have propitiated the use of cameras in MAVs for performing visual-based navigation in periods or circumstances when the GPS sensor is unreliable. Cameras meet the requirements for embedded systems providing at the same time lots of information. In this context, thanks to visual SLAM (Simultaneous Localization and Mapping) methods, a MAV can operate in a priori unknown environment using only on-board sensors to simultaneously build a map of its surroundings and locate itself inside this map (for instance [4] and [5]).

✉ Antoni Grau
antoni.grau@upc.edu

Rodrigo Munguía
rodrigo.munguia@upc.edu

[1] Department of Computer Science, CUCEI, University of Guadalajara, Guadalajara, México

[2] Department of Automatic Control, Technical University of Catalonia, Barcelona, Spain

Compared with other visual configurations (e.g., stereo SLAM), monocular SLAM presents some advantages in terms of weight, space, power consumption and scalability. For example, in stereo rigs, the fixed base-line between cameras could limit the operation range instead.

On the other hand, the use of monocular vision introduces some technical challenges as the impossibility of directly recovering the metric scale of the world [6].

1.1 Related Work

In navigation systems for MAVs based on the monocular vision, different approaches have been followed for addressing the problem of the metric scale. One of the first approaches has been the use of a feature pattern with known dimensions. For example in [7], the initial map features are determined by the metrics of the feature pattern. Also, some human-supervised schemes have been used for the same purpose. For instance, in [5] and [8], the first map estimates are aligned by hand with a ground-truth at the beginning of the operation, in order to set the metric scale. Another approach consists on taking advantage of the topology and configuration of specific environments. For example, in [9] a method is proposed with several assumptions about the structure of the environment. Some of these assumptions are the flatness of the floor, and the known relative altitude of the MAV respect the floor by the use of an ultrasonic range sensor as well as the distance from the MAV to the wall of the corridor. Due to the several assumptions, this method is intended to be only applied in environments formed by corridors, like those commonly found in office buildings. Other methods like [10] or [11] fuse inertial measurements obtained from an IMU in order to recover the metric scale. In these approaches, the scale is explicitly considered as a variable in the system state, and it is estimated by means of an Extended Kalman Filter (EKF). In this case, the innovation error is defined as the difference between the measured acceleration in the vertical axis (obtained from the IMU) and the unscaled acceleration (obtained from the monocular vision). One potential problem with this scheme is related to the fact that the acceleration obtained by the IMU has a dynamic bias which is difficult to estimate. This bias introduces at the same time a bias in the estimated scale. Also, for these approaches, a careful calibration process is needed in order to align the camera and the IMU. In [12] an EKF-based method for visual odometry is presented (there is not a mapping process). In that work, the trajectory of the MAV is estimated fusing data from a monocular downward facing camera, inertial sensors, and a range sensor (sonar altimeter). In this case, the orography of the terrain is assumed to be completely planar (flat terrain assumption), using the range finder to directly measure the altitude of the vehicle.

1.2 Objectives and Contributions

This work presents a monocular SLAM system to be used in micro aerial vehicles. The proposed method is intended to address the problem of the visual-based navigation in fully GPS-denied environments or as a backup system in periods where GPS signal is unreliable. In order to overcome the problem of recovering the metric scale which is inherent to this kind of systems, a nonlinear observability analysis is developed. From this analysis, conditions of sufficiency for the observability of the metric scale are presented. The design of the proposed method is based on these theoretical findings. In particular, a novel technique for inferring the approximate depth of visual features from an ultrasonic range finder is developed. In this sense, new techniques for initializing the features into the map as well as for updating the filter by means of visual and range measurements are presented. Additionally, the altitude of the vehicle is proposed to be updated using the pressure measurements of a barometer.

Unlike other methods, as [9] or [12], in this work the use of accelerometers for recovering the metric scale is avoided. Therefore, the problem associated with the dynamic error bias of the accelerometer is thus avoided. Also, there is no need of an extensive pre-calibration routine for aligning the IMU and the camera, like in [10, 11] or [12]. Compared with [12], where the ultrasonic range finder is used for directly measuring the altitude of the vehicle, in this work, the range finder is used for computing the approximate depth of the visual features that are more likely to be detected by the sensor. Thus, the assumption of a completely flat terrain is relaxed by the assumption of a terrain with soft but continuous changes in altitude.

2 Preliminaries

A simplified 3DOF model will be used for introducing the problem to be addressed in a clear manner. Note that this simplified model retains the main aspects to be addressed in the full problem.

Let the following unconstrained model $\dot{x}_c = f(x, u)$ represent the basic dynamics of a camera on board a MAV (see Fig. 1):

$$\begin{aligned} \dot{x}_r &= v_x \dot{z}_r = v_z \dot{\theta}_r = \omega_c \dot{v}_x = V_x \\ \dot{v}_z &= V_z \dot{\omega}_c = \Omega \end{aligned} \tag{1}$$

where $x_c = [x_r, z_r, \theta_r, v_x, v_z, \omega_c]^T$ is the state vector of the camera, and $[x_r, z_r, \theta_r]$ and $[v_x, v_z, \omega_c]$ represent respectively the position and orientation of the camera, and their first derivatives. In model (1), it is assumed that the process is governed by an unknown input $u = [V_x, V_z, \Omega]^T$ of linear and angular accelerations with zero mean and

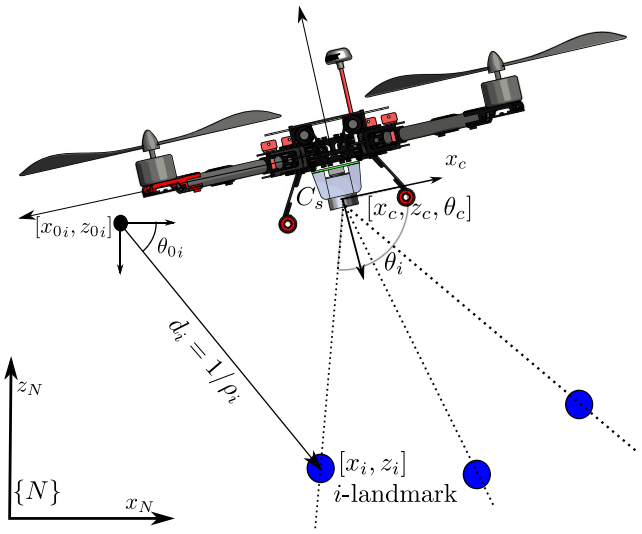

Fig. 1 System Parametrization

known covariance Gaussian processes. Note that a constant-acceleration camera is assumed. Also, it is assumed that the camera modeled by Eq. 1 is able to provide angular measurements from a set of 2D feature points composing the map. In this case, the measurement process is modeled by the following equation:

$$y_i = h_{\theta i}(x) = \arctan2\left(\frac{z_c - z_i}{x_c - x_i}\right) - \theta_c \tag{2}$$

where $[x_i, z_i]$ represents the position of a i-th feature which is coded in its inverse form:

$$\begin{aligned} x_i &= (1/\rho_i)cos(\theta_{0i}) + x_{0i} \\ z_i &= (1/\rho_i)sin(\theta_{0i}) + z_{0i} \end{aligned} \tag{3}$$

In this case, the state vector of the i-th feature y_i is defined by $y_i = [x_{0i}, z_{0i}, \theta_{0i}, \rho_i]^T$. Here $[x_{0i}, z_{0i}]$ represents the location of the camera when the feature was first detected, θ_{0i} is the initial angular measurement, and $\rho_i = 1/d_i$ represents the inverse of the feature depth d_i (see Fig. 1).

The full state vector x that is intended to be estimated is integrated by the camera state x_c, and the states y_i of the features composing the map.

$$x = [x_c, y_1, y_2, ..., y_n]^T \tag{4}$$

For the system state x it is assumed that the map features y_i remain static (rigid scene).

3 Observability of Metric Scale

In SLAM, the map and trajectory can only be estimated up to scale, if the system works only with monocular vision without additional information added to it, [6]. In this section, the problem of the observability of the metric scale in a monocular system is investigated. The theoretical findings obtained with this analysis provide support to the design of a monocular-based SLAM method able to recover the metric scale without the need of GPS data.

Firstly, the state vector defined in Eq. 4 will be split into a metric parameter s, unobservable when only angular measurements are available, and into a dimensionless map and camera part. For this purpose, the approach proposed in [13] is followed. The new system state x_s is defined by:

$$x_s = [s, \Pi_{x_c}, \Pi_{z_c}, \theta_c, \Pi_{v_x}, \Pi_{v_z}, \Pi_{\omega_c}, \Pi_{y_1}..., \Pi_{y_n}]^T \tag{5}$$

Camera measurements will reduce the uncertainty related to the scene geometry, but not the uncertainty related to the metric parameter s. The transformation of the state vector x_s and the original state vector x is defined by the following non-linear relationships:

$$\begin{cases} x_c = s\Pi_{x_c} \quad z_c = s\Pi_{z_c} \quad v_x = s\Pi_{v_x}\Delta t \\ v_z = s\Pi_{v_z}\Delta t \;\; \omega_c = s\Pi_{\omega_c}\Delta t \\ y_i = [s\Pi_{x_{0i}}, s\Pi_{z_{0i}}, \theta_i, \Pi_{\rho_i}/s] \end{cases} \tag{6}$$

Second, the dynamics of the 3DOF monocular SLAM system is defined, see Section 2, in terms of the metric parameter s and the dimensionless parameters. For this purpose, Eq. 6 is substituted into Eqs. 1–3, and the system state is augmented with s. Therefore, the new system dynamics becomes:

$$\begin{array}{llll} \dot{s} = 0 & \dot{x}_c = s\Pi_{v_x}\Delta t & \dot{z}_c = s\Pi_{v_z}\Delta t & \dot{\theta}_c = s\Pi_{\omega_c}\Delta t \\ \dot{v}_x = 0 & \dot{v}_z = 0 & \dot{\omega}_c = 0 & \dot{\rho}_i = 0 \end{array} \tag{7}$$

Note that in the system defined by Eq. 7 the metric parameter s is defined to be constant. The new system output equation is:

$$y_i = h_{\theta i}(x) = \arctan2\left(\frac{s\Pi_{z_c} - z_i}{s\Pi_{x_c} - x_i}\right) - \theta_c \tag{8}$$

where:

$$\begin{aligned} x_i &= (s/\Pi_{\rho_i})cos(\theta_i) + x_{0i} \\ z_i &= (s/\Pi_{\rho_i})sin(\theta_i) + z_{0i} \end{aligned} \tag{9}$$

If n landmarks are measured by the camera, the system output is defined as $y = [h_{\theta 1}, ..., h_{\theta n}]^T$.

Remark 1 When the problem is focused on the position estimation of the MAV camera, it is assumed that the orientation of the MAV can be determined from some independent device (e.g. an AHRS device). Hence, the system state can be simplified by removing the variables related to orientation.

Remark 2 According to Section 2, the state of the i-th feature y_i is defined by $y_i = [x_{0i}, z_{0i}, \theta_i, \rho_i]$ where $[x_{0i}, z_{0i}]$ is the position of the camera C_s when the feature was first detected, θ_i is the first bearing measurement, and $\rho_i = 1/d_i$ is the inverse of the feature depth d_i. Hence, note that $[x_{0i}, z_{0i}, \theta_i]$ is directly given when the i-th feature is

initialized. In this case, for the observability analysis only the state of the camera and the inverse depth ρ_i of features, will be considered.

According to Remark 1 and Remark 2 the system state x_s is defined as follows:

$$x_s = [s, \Pi_{x_c}, \Pi_{z_c}, \Pi_{v_x}, \Pi_{v_z}, \Pi_{\rho_1}, \Pi_{\rho_2}, ..., \Pi_{\rho_n}]^T \tag{10}$$

The property of the observability has important consequences for a SLAM system. A system is defined as observable if the initial state x_0, at any initial time t_0, can be determined given the state transition model of the system $\dot{x} = f(x, u)$, the observation model $y = h(x)$ and the measurements $y[t_0, t]$ from time t_0 to a finite time t. When a system is fully observable, the lower bound of the error in the estimations of its state will only depend on the noise parameters of the system and will not be reliant on initial information about the state.

A nonlinear system is determined to be *locally weakly observable* if the observability rank condition $\text{rank}(\mathcal{O}) = \dim(x)$ is verified [14]. For the investigated problem, the observability matrix $\mathcal{O}$ can be computed from:

$$\mathcal{O} = \left[\frac{\mathcal{L}_f^0(h_{\theta 1})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{\theta 1})}{\partial x}^T ... \frac{\mathcal{L}_f^0(h_{\theta n})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{\theta n})}{\partial x}^T \right]^T \tag{11}$$

where $\mathcal{L}_f^i(h)$ is the i-th-order Lie Derivative [15] of the scalar field of the measurement h with respect to the vector field f. Note that in Eq. 11 the observability matrix $\mathcal{O}$ is composed of the zero-order and the first-order Lie Derivatives for the n angular measurements $y_i = h_{\theta i}(x)$.

For the system state vector defined in Eq. 10, the results have been obtained using the MATLAB symbolic toolbox and:

– As it could be expected, the system is partially observable. The maximum degree of observability was obtained with 4 landmarks. In this case, $\dim(x_s) = 9$, $\text{rank}(\mathcal{O}) = 8$. Adding more landmarks does not improve the observability. Based on [13], the non-observable mode should correspond to the metric parameter s.

In order to improve the observability of the system, a different sensory source must be considered. In this case, it will be considered that measurements of the altitude of the MAV camera are available. The additional system output equation y_a is:

$$y_a = h_{z_c}(x) = z_c = s\Pi_{z_c} \tag{12}$$

For n landmarks being measured by the camera, the full system output is now defined by $y = [h_{z_c}, h_{\theta 1}, ..., h_{\theta n}]^T$. The new observability matrix $\mathcal{O}$ can be now computed from:

$$\mathcal{O} = \left[\frac{\mathcal{L}_f^0(h_{z_c})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{z_c})}{\partial x}^T \frac{\mathcal{L}_f^0(h_{\theta 1})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{\theta 1})}{\partial x}^T ... \frac{\mathcal{L}_f^0(h_{\theta n})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{\theta n})}{\partial x}^T \right]^T \tag{13}$$

The following results have been obtained with this observability matrix $\mathcal{O}$:

– The system becomes observable. The maximum degree of observability was obtained with 3 landmarks, that is, $\dim(x_s) = 8$, $\text{rank}(\mathcal{O}) = 8$.
– A sufficient condition for observability is defined by $(\Pi_{v_z} \neq 0)$. That means that the altitude of the vehicle must be varying in order to improve the observability of the system.

As a step forward, it will be considered that measurements of range (depth) are available for a subset of map features instead of altitude readings. A range measurement for an i-th feature can be modeled by:

$$y_d = h_{\rho_i}(x) = 1/\rho_i = s/\Pi_{\rho_i} \tag{14}$$

Therefore, if n landmarks are being measured by the camera C_s, and m range measurements ($m \leq n$) are available, the system output can be defined by $y = [h_{\theta 1}, ..., h_{\theta n}, h_{\rho 1}, ..., h_{\rho m}]^T$. The new observability matrix $\mathcal{O}$ can be computed as:

$$\mathcal{O} = \left[\frac{\mathcal{L}_f^0(h_{\theta 1})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{\theta 1})}{\partial x}^T ... \frac{\mathcal{L}_f^0(h_{\theta n})}{\partial x}^T \frac{\mathcal{L}_f^1(h_{\theta n})}{\partial x}^T , \times \frac{\mathcal{L}_f^0(h_{\rho 1})}{\partial x}^T ... \frac{\mathcal{L}_f^0(h_{\rho m})}{\partial x}^T \right]^T \tag{15}$$

Note that in Eq. 15, only the zero-order Lie Derivative is calculated for range measurements $[h_{\rho 1}, ..., h_{\rho m}]^T$. The following results have been obtained with this observability matrix $\mathcal{O}$:

– The system becomes observable. In this case, the observability does not depend on the number of features included into the system state. Full observability, $\dim(x_s) = \text{rank}(\mathcal{O})$, can be reached by including a single measurement of range h_{ρ_i}.

Remark 3 As of that moment on, after analyzing the observability of the system, the metric parameter s will be implicitly considered again into the system state.

4 Method Description

4.1 Sensor Fusion Approach

As previously seen, when the monocular vision is used as the unique sensory input to the SLAM system is not possible to recover the metric information of estimations. Nevertheless, according to the theoretical results obtained with the observability test, the metric scale becomes observable if the measurements of the MAV absolute altitude are considered. Therefore based on this theoretical result, in the proposed system the altitude of the vehicle is updated through the atmospheric pressure measurements obtained from a barometer. On the other hand, according to the same analysis, when altitude measurements are used, the observability of the scale depends on the movement of the vehicle along the vertical axis. For different scenarios, this dependency on the movement of the MAV could affect the effectiveness of the scale recovery if only altitude measurements are used for this purpose.

In order to improve the robustness of the proposed system for recovering the metric scale, a technique that makes use of an ultrasonic range finder for incorporating depth information of visual features is developed. The technique is based on a second theoretical result which states that the metric scale can become observable if the measurement of the depth of a single feature is available.

4.2 System Assumptions

In this work, a quadrotor MAV moving freely in $\mathbb{R}^3 \times SO(3)$ has been considered, anyway, it is not difficult to extend the research to another aerial configurations. In the presented contribution, it is assumed that the monocular camera, as well as the ultrasonic range finder, are mounted on a servo-controlled gimbal that counteracts the movements of the MAV in order to stabilize its orientation toward the ground (See Fig. 2). Also, it is assumed that the range sensor is mounted near the camera and aligned with its optical axis.

It is important to note that the use of the servo-controlled gimbal adds some extra weight to the MAV. In this sense, the inherent benefit related to the use of monocular vision in terms of weight could be affected. On the other hand, it is also well known that the stabilized video facilitates the operation of visual-based SLAM systems.

In order to obtain measurements of the altitude of the MAV, it is assumed that a barometer is available. Also, it is assumed that the location of the origin of the camera frame respect to other elements of the quadcopter (e.g. barometer) is known and fixed.

Additionally to the assumption regarding the orientation of the camera is maintained stably by the use of the gimbal, in this work, it is assumed that an Attitude and Heading Reference Systems (AHRS) is available for providing attitude and heading estimates with enough accuracy. Considering the above, for the sake of simplicity, the attitude variables are removed from the SLAM filter. And thus, this work is focused only on position estimation. Also, note that the observability results were obtained under the same assumption. Fortunately, in practical applications, the available Attitude and Heading Reference Systems (AHRS), can address the problem of the estimation of the attitude (roll and pitch) of the MAVs in a robust manner (e.g. [16] and [17]). However, it is important to note that a typical AHRS makes use of magnetometers for updating the Yaw (heading) of the vehicle. Magnetometers are often difficult to calibrate and are sensitive to external disturbances. In this work, it is assumed that the AHRS provides an acceptable heading estimation. However, this aspect should be studied more carefully in a future work.

The system proposed in this work is mostly intended to be applied in scenarios involving flight trajectories near to the origin of navigation reference frame. Therefore, the initial position of the MAV defines the origin of the navigation coordinates frame. The navigation system follows the NED (North, East, Down) convention. The magnitudes expressed in the navigation and camera frame are denoted respectively by the superscripts ^{N}, and ^{C}. All the coordinate systems are right-handed defined.

4.3 Sensor Measurement Models

4.3.1 Visual Measurements

A central-projection camera model is used for modeling the monocular camera on board the MAV. In this case, the image plane is assumed to be located in front of the camera's origin where a non-inverted image is formed. The camera frame ^{C} is right-handed with the z-axis pointing to the field of view. The $\mathbb{R}^3 \Rightarrow \mathbb{R}^2$ projection of a 3D point, located at $p^N = (x, y, z)^T$, to the image plane is defined by:

$$u = \frac{x'}{z'} \quad v = \frac{y'}{z'} \tag{16}$$

where u and v are the image coordinates (in pixels) of the projection of the 3D point, and:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} p^C \tag{17}$$

where p^C is the same 3D point p^N, but expressed in the camera frame ^{C} by $p^C = R^{NC} p^N$. R^{NC} is the rotation matrix transforming from the navigation frame ^{N} to the camera frame ^{C}. Note that R^{CN} is known by the use of the gimbal and it fulfills that $R^{NC} = (R^{CN})^T$.

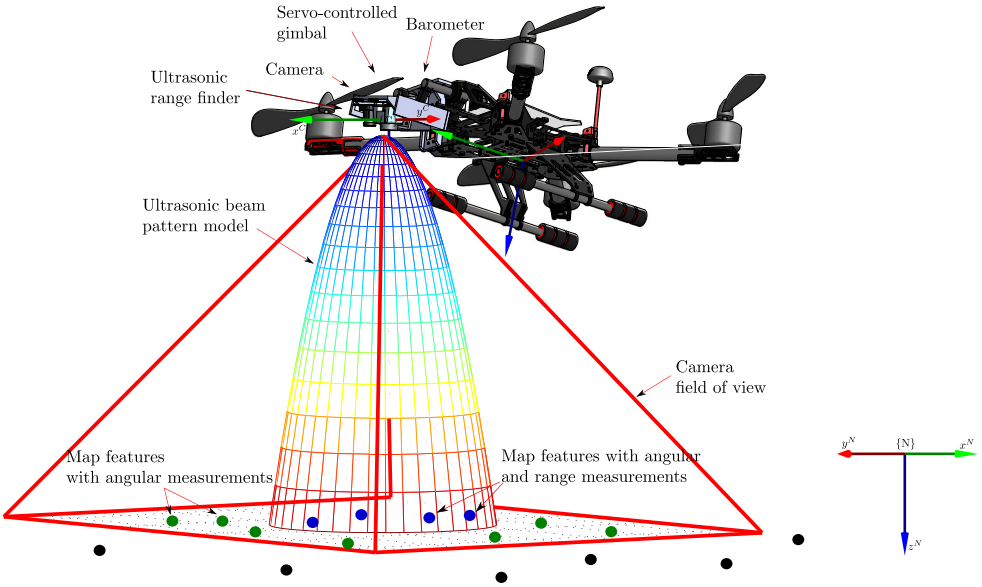

Fig. 2 Aerial platform and coordinate systems. The ultrasonic range finder is used for estimating the approximate depth of visual features lying in its beam pattern. An elliptic paraboloid is used as a simple model of the actual beam pattern of the ultrasonic range finder

Reciprocally, a vector $h^C = [h_x^C, h_y^C, h_z^C]^T$, pointing from the camera optical center position to the 3D point location, can be obtained from the image point coordinates u and v by means of:

$$h^C(u, v) = \left[\frac{u_0 - u}{f}, \frac{v_0 - v}{f}, 1\right]^T \tag{18}$$

The vector h^C can also be expressed in the navigation frame by $h^N = R^{CN}h^C$. Note that for the $\mathbb{R}^2 \Rightarrow \mathbb{R}^3$ transformation defined in Eq. 18, depth information is unavailable.

The camera lens distortion is considered through the model described in [18]. Using the former model (and its inverse form), undistorted pixel coordinates (u, v) can be obtained from (u_d, v_d) and conversely. In this work, it is assumed that the intrinsic parameters of the camera are already known: focal length f, principal point (u_0, v_0), and radial lens distortion $k_1, ..., k_n$.

4.3.2 Measurement of Visual Features Depth

According to the theoretical analysis, the metric scale can become observable when a single measurement of a feature depth is available. Therefore, a new technique is proposed for determining the approximate depth of visual features (whenever is possible). The idea is to define a subset of visual features whose depths are more likely to be related to the distance measured by the ultrasonic range finder. For this end, an image region will be computed in order to serve as criteria for choosing the visual features whose approximate depth will be inferred from the ultrasonic sensor. Note that the operation rate of ultrasonic range finder sensors is typically lower (3-4 Hz) than the frame rate of cameras. Therefore, the proposed technique can only be applied to a subset of frames. In this case, in order to establish some degree of synchrony between frames and ultrasonic readings, every time that a range measurement is available, the next available camera frame is associated with it.

An elliptic paraboloid with equation $z = ax^2 + ay^2$ is used for modeling the actual beam pattern of the ultrasonic sensor (see Fig. 3). For adjusting the model to the actual beam pattern (as best as possible), the parameter a is chosen using the sensor datasheet.

A circle $z_r = ax^2 + ay^2$, with radius $r_m = \sqrt{z_r/a}$, can be determined by the intersection of the paraboloid and the plane $z = z_r$, where z_r is the current reading obtained from the ultrasonic range finder. The interior of this circle defines a ground region where the landmarks that lie inside can be associated with the range measurements provided by the ultrasonic sensor (see Fig. 2).

In order to project the circular ground region to the image plane, a single 3D point $p^N = (r_m, 0, z_r)$ is projected using Eq. 16. The radius r_c (in pixels units) of the circular image region is determined by:

$$r_c = ||[u_c, v_c]^T - [u_0, v_0]^T|| \tag{19}$$

where $[u_c, v_c]$ are the image coordinates obtained from projecting the 3D point $p^N = (r_m, 0, z_r)$. Note that $r_c = f(z_r, a)$ is a function of the beam pattern shape of the ultrasonic sensor as well as the range it measures (see Fig. 3).

Any visual feature i satisfying $r_i < r_c$, where $r_i = ||[u_i, v_i]^T - [u_0, v_0]^T||$, lies inside the circular region. Hence, for such a feature it is assumed that it has an approximate depth d_i given by:

$$d_i = \frac{z_r||h^C||}{h_z^C} \tag{20}$$

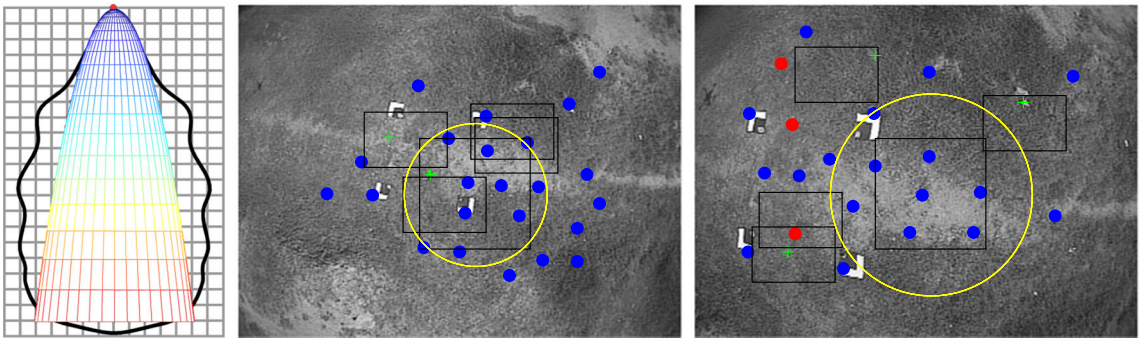

Fig. 3 An elliptic paraboloid is used as a simple model of the actual beam pattern of the ultrasonic range finder (left). The size of the terrain region detected by the ultrasonic range finder is a function of the beam pattern of the sensor and the flight altitude of the MAV. Frame captured at 7m of altitude (middle). Frame captured at 3.9m of altitude (right)

where h^C is computed with Eq. 18 and it represents the directional vector pointing from the camera optical center to the 3D location of the i visual feature.

4.3.3 Altitude Measurements

Under certain conditions, the observability of the metric scale can be improved by incorporating altitude measurements into the system. In this case, the altitude or height of the MAV above a local ground location can be computed from the change in the atmospheric pressure between the ground and the altitude of interest. The following equation can be used to obtain altitude readings z_a from a barometer:

$$z_a = \left(1 - \left(\frac{P}{P_g}\right)^{\frac{RL_0}{Mg}}\right)\frac{T}{L_0} \tag{21}$$

where P is the current barometric pressure measurement; P_g is the barometric pressure at the initial position (home position); $R = 8.31432$ N-m/(mol-K) is the universal gas constant for air; $L_0 = -.0065$ K/m is the rate of temperature decrease in the lower atmosphere; $M = 0.0289644$ kg/mol is the standard molar mass of atmospheric air; $g = 9.80665$ m/s^2 is the gravitational constant and T is the temperature at flight location in Kelvin degrees.

In this work, a measurement z_a of the actual MAV's altitude z^N is modeled by:

$$z_a = z^N + \mathrm{x}_z + v_z \tag{22}$$

where x_z is an additive error (bias) and v_z is a Gaussian white noise with a power spectral density (PSD) σ_z^2.

To model the transient behavior of the bias x_z, a Gauss-Markov process is used:

$$\dot{\mathrm{x}}_z = -\lambda_z \mathrm{x}_z + v_\lambda \tag{23}$$

where the constant parameter λ_z is a correlation time factor which models how fast the bias of altitude measurements are varying, and v_λ is a Gaussian white noise with a power spectral density (PSD) σ_λ^2.

4.4 EKF-SLAM

The proposed method makes use of the standard EKF-SLAM methodology for estimating the system state. Interested readers are referred to [19] and [20] for an extensive review on the EKF-SLAM methodology.

The system state to be estimated is:

$$x = [r^N, v^N, \mathrm{x}_z, y_1, y_2, ..., y_n]^T \tag{24}$$

Let $r^N = [x_r, y_r, z_r]$ represent the position of the MAV camera expressed in the navigation frame; let $v^N = [v_x, v_y, v_z]$ denote the linear velocity of the vehicle expressed in the navigation frame; let x_z represent the barometer bias (in meters), and let $y_1, y_2, ..., y_n$ represent the n features that compose the stochastic map.

To represent the features in the map two different models are used: i) Euclidean parametrization, and ii) inverse-depth parametrization. Map features whose depth is assumed to be approximately known are represented in their Euclidean form. On the other hand, the inverse-depth parametrization is more convenient for features whose depth has a high degree of uncertainty.

Euclidean features are defined by:

$$y_{e_i} = [x_i, y_i, z_i]^T \tag{25}$$

Let $[x_i, y_i, z_i]$ be the coordinates of the i-th feature, expressed in the navigation frame.

Inverse-depth (ID) features are defined by:

$$y_{id_i} = [r_i, \theta_i, \phi_i, \rho_i]^T \tag{26}$$

Let $r_i = [x_{0_i}, y_{0_i}, z_{0_i}]$ be the coordinates of the center of the camera when the feature was observed for the very

first time; let θ_i and ϕ_i represent azimuth and elevation respectively, and let $\rho_i = 1/d_i$ be the inverse of the depth d_i, and:

$$\theta_i = \text{atan2}(h_y^N, h_x^N) \qquad \phi_i = \text{acos}\left(\frac{h_z^N}{\sqrt{(h_x^N)^2+(h_y^N)^2+(h_z^N)^2}}\right) \tag{27}$$

Here, $h^N = [h_x^N, h_y^N, h_z^N]^T$ is computed from Eq. 18.

Note that the Euclidean form is more computationally efficient since it requires half of the parameters of the inverse-depth parametrization.

4.4.1 System Prediction

The following discrete model is used to take a step forward to the system state x:

$$\begin{cases} r_{[k+1]}^N = r_{[k]}^N + v_{[k]}^N \Delta t \\ v_{[k+1]}^N = v_{[k]}^N + V^N \\ \mathrm{x}_{z[k+1]} = (1 - \lambda_z \Delta t)\mathrm{x}_{z[k]} + V_\lambda \\ y_{1[k+1]} = y_{1[k]} \\ \vdots \\ y_{n[k+1]} = y_{n[k]} \end{cases} \tag{28}$$

At every step, it is assumed that there is an unknown linear velocity with acceleration zero-mean and known-covariance Gaussian processes σ_a, producing an impulse of linear velocity: $V^N = \sigma_a^2 \Delta t$. In Eq. 28, the equation used for modeling the behavior of the barometer bias x_z, was derived from the first order discretization of the Eq. 23, in this case $V_\lambda = \sigma_\lambda^2 \Delta t$. Let Δt be the system sample time.

4.4.2 Initialization of Map Features

New features are initialized in those frames where a range measurement z_r is available (see Section 4.3.2). To perform this process, a random search is conducted in the image in order to detect regions with salient visual points. In this work, the detector proposed in [21] is used for this purpose.

Every visual point i detected with image coordinates (u_i, v_i), is tested in order to know if an approximate depth d_i can be associated with it (see Section 4.3.2). Then, the visual points are initialized according two possible cases:

Case 1 Visual points whose depth is assumed to be available are initialized as Euclidean map features. The initialization function $y_{e_{new}} = f_e(x, u_i, v_i, d_i)$ used for computing new Euclidean features is:

$$y_{e_{new}} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = r^N + R^{CN}\left(\frac{h^C}{\| h^C \|} d_i\right) \tag{29}$$

Let R^{CN} be the camera to navigation rotation matrix, let r^N be the actual camera-vehicle position, h^C is computed from Eq. 18, and let d_i be the approximate depth computed from Eq. 20.

Case 2 Visual points whose depth is assumed to be uncertain are initialized in their inverse-depth form. The initialization function $y_{id_{new}} = f_{id}(x, u_i, v_i, d_i)$ used for computing new ID features is:

$$y_{id_{new}} = [r_0, \theta_0, \phi_0, \rho_0]^T \tag{30}$$

with $[r_0, \theta_0, \phi_0]$ calculated as in Eq. 26; and $\rho_0 = 1/d_i$. Note that d_i is taken as the best hypotheses of depth for ID features.

The system state x is augmented in a similar manner for both kinds of features: $x_{new} = [r^N, v^N, y_1, y_2, ..., y_n, y_{new}]^T$, either $y_{new} = y_{e_{new}}$ or $y_{new} = y_{id_{new}}$.

Also, in each case, the new covariance matrix P_{new} is computed by:

$$P_{new} = \nabla J \begin{bmatrix} P_{old} & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix} \nabla J^T \tag{31}$$

Let ∇J be the Jacobian computed from the proper initialization function (either f_e or f_{id}), let R be the measurement noise covariance matrix for (u_i, v_i), and σ_d is chosen according to the kind of map feature (see Fig. 4):

- For Euclidean features, σ_d is chosen with an small value. For example, in experiments good results were found with $\sigma_d = 1/(10 \times d_i)$.
- For ID features, σ_d is chosen as in [22] ($\sigma_d = \rho_0/2$), in order to cover a big uncertainty region.

4.4.3 Measurement of Map Features

In frames with an associated range z_r, when a feature is re-observed its image coordinates (u_i, v_i) are tested in order to know whether an approximate depth d_i can be associated to it (see Section 4.3.2). Features satisfying the above condition will be used for updating the filter with both range and bearing measurements. If visual features do not satisfy such a condition, and for all the visual features re-observed in frames without an associated range, only visual information will be available for updating the filter.

The following procedures are used for the two possible cases:

Case 1: For features with an associated depth d_i, both Euclidean and ID features, it is assumed that an indirect 3D

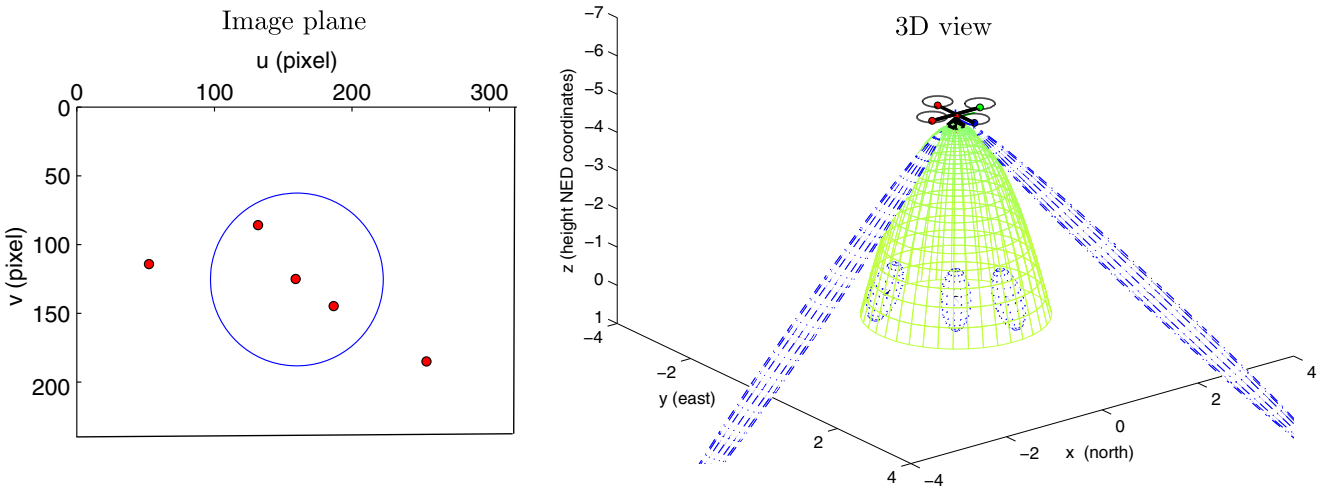

Fig. 4 Initialization of new map features: Camera image plane (left plot). 3D view of the estimates (right plot). The operational range of the ultrasonic sensor is indicated respectively by the blue circle in the image plane and by the green paraboloid in the 3D view. The blue ellipsoids indicate the initial uncertainty region of each new feature 3D location. Note that the visual features, that are assumed to be detected by the ultrasonic sensor, are initialized with a smaller initial uncertainty in depth

position measurement $z_{mi} = [x_i, y_i, z_i]$ is available. The measurement z_{mi} is computed as follows:

$$z_{mi} = \frac{h^C}{\| h^C \|} d_i \tag{32}$$

where h^C is computed by Eq. 18, using the undistorted pixel coordinates obtained from (u_i, v_i). The measurement model $[x_i, y_i, z_i] = h_i(x, y_i)$ is defined as follows:

$$p^C = h_i(x, y_i) = R^{NC}(p^N - r^N) \tag{33}$$

where R^{NC} is the camera to navigation rotation matrix, and p^C is the predicted 3D position of the feature y_i, expressed in the camera frame. In case of Euclidean features, $p^N = y_e$. In case of ID features:

$$p^N = r_i + \frac{1}{\rho_i} m(\theta_i, \phi_i) \tag{34}$$

where $m(\theta_i, \phi_i) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi)^T$ is the unit vector defined by the pair of azimuth-elevation angles in y_{id}.

Case 2: For features without an associated depth d_i, the visual measurement $z_{mi} = [u_i, v_i]$ is used together with the standard measurement model $[u_i, v_i] = h_i(x, y_e)$. In this case p^C is computed using Eq. 33 as well. Then p^C is projected to the image plane by Eqs. 16 and 17 (See Section 4.3.1).

4.4.4 Map Management

This work is mainly intended to address the local navigation problem, that is, the proposed system is intended to be applied in scenarios involving flight trajectories near to the origin of the navigation frame. Hence, large-scale SLAM and loop-closing are not considered in this work. On the other hand, a SLAM framework that works reliably in a local way can be applied to large-scale problems using different methods, such as sub-mapping or graph-based global optimization [23].

In order to improve the computational efficiency of the algorithm, ID features whose depth estimation has converged, are converted to Euclidean features using the method proposed in [24]. In a similar way, visual features with high percentage of mismatching are removed from the system state and covariance matrix.

4.4.5 Altitude Updates

When a new barometer reading is available, the system can be updated with an altitude measurement, using the standard update equations. In this case, altitude measurements z_a are computed using Eq. 21, from the pressure readings.

The measurement model $h_a(x)$ is defined by:

$$h_a = z^N + \mathrm{x}_z \tag{35}$$

where z^N and x_z are taken directly from the system state x.

5 Experimental Results

In order to validate the performance of the proposed method, several experiments have been carried out either in simulation and with real data. The methods used in the experiments were implemented in MATLAB®.

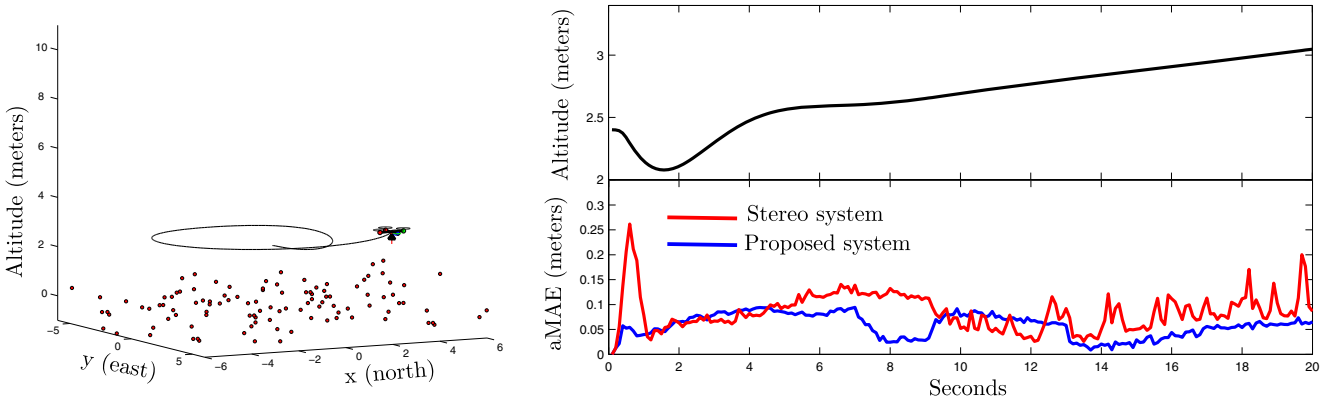

Fig. 5 Comparison between the proposed system and a stereo system, during a flight trajectory at low altitude. In this case, the mean absolute error (MAE) in position is similar

5.1 Experiments with Simulations

In simulations, the model used to implement the vehicle dynamics was taken from [25]. The monocular camera was simulated using the same parameter values of the camera used in the experiments with real data. The barometer was simulated using the data presented in Section 4.3.3. The ultrasonic sensor was simulated using parameters values taken from the datasheet of the real device used in experiments with real data. In each case, Gaussian noise was added in order to model system uncertainty. Also, it is assumed that the visual features can be detected and matched without errors. The simulation results were obtained averaging 20 Monte Carlo executions.

Figures 5 and 6 present the results obtained from comparing the proposed method with a stereo system. Both cameras of the stereo system (with a fixed baseline of 20 centimeters) are simulated using the same parameters of the monocular camera used by the proposed method. The same magnitude of noise is added to both the monocular system and the stereo system. It is well worth to observe that for flight trajectories, at low altitude above the ground (less than three meters), both systems perform reasonably well (See Fig. 5). Nevertheless, the performance of stereo system behaves considerably worse as the altitude of the vehicle increases (see Fig. 5). This fact is due to the fixed baseline between the cameras of the stereo system. Therefore, the operating range of the stereo system is highly restricted to low altitudes.

It is also important to note that an ultrasonic sensor has a limited operating range (typically under ten meters). Above that altitude, the ultrasonic sensor is unable to provide information about the depth of visual features. In these conditions, the recovering of the metric scale of the estimates rely only on the altitude information provided by the barometer. However, as it was previously mentioned, the efficacy of this latter approach heavily depends on the movements of the MAV.

Figures 7 and 8 present the results obtained from analyzing the effects of the inclusion of the range-finder model into the proposed method. It is important to recall that previous approaches assume that the orography of

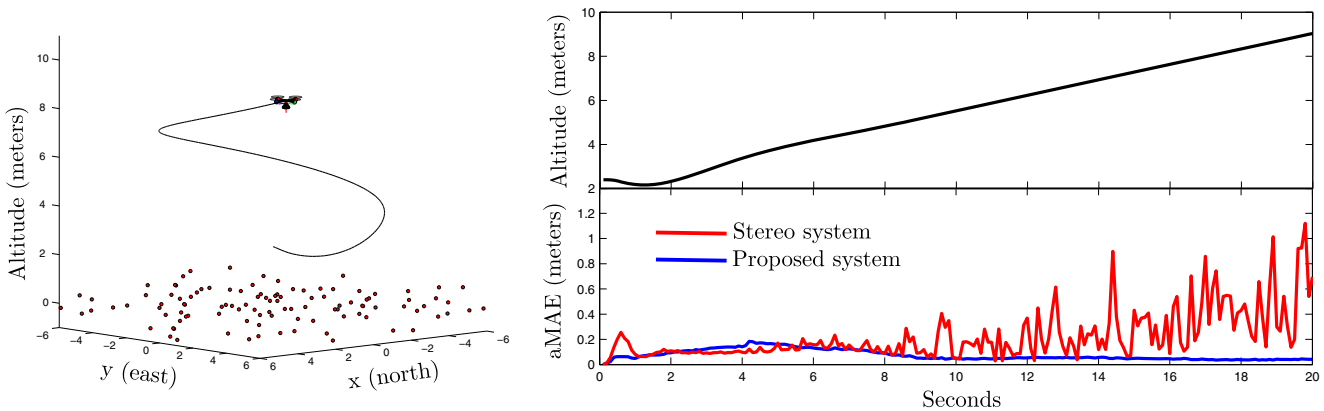

Fig. 6 Comparison between the proposed system and a stereo system, during an ascendant flight trajectory. Note that for the stereo system, the error in the position considerably increases as the vehicle gets more altitude

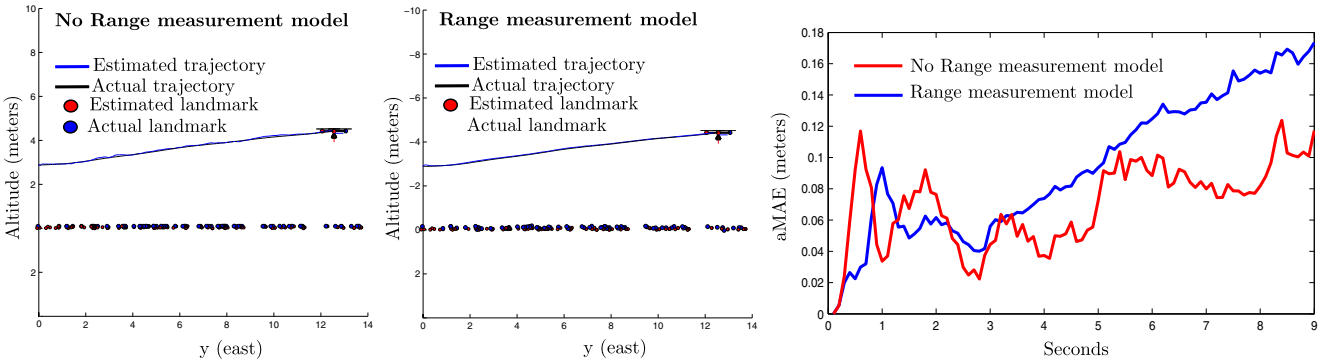

Fig. 7 Analysis of the use of the range-finder model for flights performed over terrain without orography changes (flat terrain)

the ground is completely planar (e.g.[12]). On the other hand, in the proposed method, it is assumed that only the portion of the ground detected by the ultrasonic sensor is approximately flat. For this experiment, a variant of the proposed method, that does not make use of the technique presented in Section 4.3.2, was used for comparison purposes. Such a technique defines what visual features are more likely to be related to the distance measured by the ultrasonic range finder. Without this technique, the depth of all the visual features are inferred from the readings of the ultrasonic sensor.

It is remarkable to observe that when a terrain without orography changes is considered (see Fig. 7), the performance when no range-finder model is used is even slightly better. This is because more depth information is incorporated into the system. However, when a terrain with a slope of 10 degrees is considered (see Fig. 8), the variant without range-finder model is unable of correctly mapping the landmarks. This fact has as a consequence a high drift in position error. On the other hand, with the use of the range-finder model, the proposed system is able to correctly estimate the map and trajectory. In this manner, the assumption of a completely flat terrain is relaxed by the assumption of a terrain with soft but continuous changes in altitude.

Figure 9 presents the results of an experiment carried out for validating the ability of the filter for estimating the barometer bias x_z. In this case, an initial bias of 0.5 meters was considered. For this experiment, the convergence time was about of 15 seconds.

For the proposed method, it is assumed that a servo-controlled Gimbal stabilizes the orientation of the camera and range-finder toward the ground. Figure 10 shows the results of an experiment that analyzes the effects on estimation, derived from the Gimbal angle control error. For this experiment, the same ascendant flight trajectory shown in Fig. 6 has been used. But in this case, sinusoidal signals have been used for perturbing the actual attitude of the Gimbal, while the estimation algorithm still assumes that the device points perfectly toward the ground. The experiment was repeated varying the magnitude of the angle control error. Table 1 summarizes the average mean absolute error (MAE) that was obtained for different angle control errors of the Gimbal. It is important to note that according to different manufacturers of low-cost servo-controlled Gimbal for MAVs, a typical value for the precision control angle is within the range of $\pm 0.1°$. Therefore, observing the results, it can be inferred that the use of a commercial low-cost Gimbal should not introduce a significant additional error to the estimates.

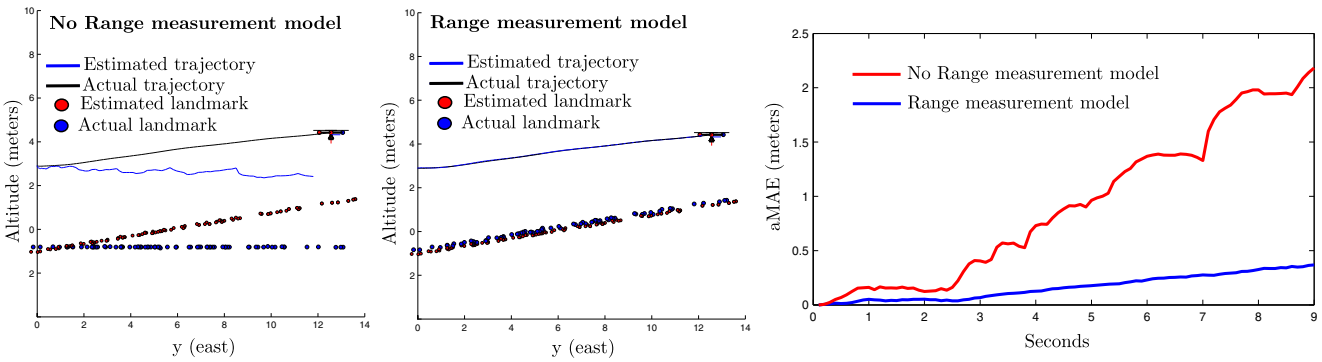

Fig. 8 Analysis of the use of the range-finder model for flights performed over terrain with a slope of 10 degrees

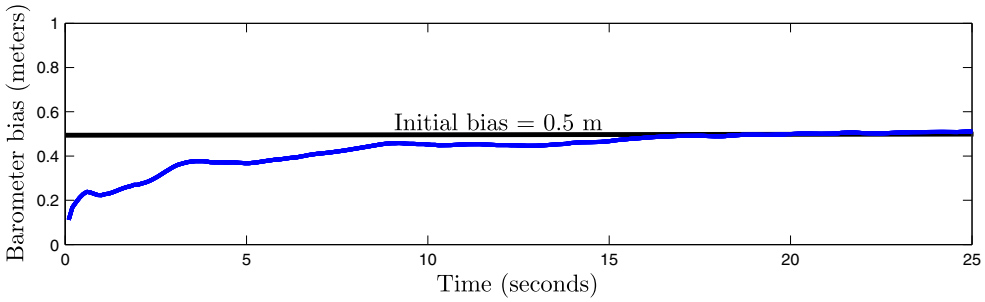

Fig. 9 Barometer bias estimation

5.2 Experiments with Real Data

A custom-built quadrotor equipped with low-cost sensors was used for capturing real data. The set of sensors is composed of: i) DX201 DPS camera with wide angle lens, ii) ultrasonic range-finder XL-MaxSonar-EZ0, iii) barometer integrated with the flight controller (Ardupilot, [26]).

The camera and the range-finder are mounted over a low-cost TURNIGY 2-Axis brushless gimbal. The experimental data sets were captured while the vehicle was manually radio-controlled. A ground-based application received the video from the camera, by mean of a 5.8-Ghz video link, as well as the data sensors, by means of a 915 MHz radio telemetry. The camera frames and data sensors are time stamped and stored. The camera gray scale frames with a resolution of 320×240 pixels were captured at 26 fps. The barometer signal was captured at 7 Hz. The range-finder signal was captured at 5 Hz.

It is important to recall that the proposed method is mainly intended to be applied to local small-scale navigation applications. In order to validate the performance of the proposal under the above conditions, an independent flight trajectory reference was computed using a perspective on 4-point (P4P) technique [27]. For this purpose four marks were placed in the floor, forming a square of known dimensions (see Fig. 3). The trajectory obtained by means of this technique should not be considered as a perfect reference of ground-truth. However this approach was very helpful to have a fully independent reference of flight for evaluation purposes.

Figure 11 shows the experimental results obtained for a small-scale flight. In this experiment the proposed method was compared with two additional approaches:

(i) The undelayed inverse depth method (UID) [22], with an initial inverse-depth value of $\rho_0 = 1/2m$. In this case, because no extra sensory information is used, the metric scale of estimations depends only on the initial inverse-depth value. This approach was considered in order to illustrate the difficulty of recovering the metric scale using only monocular vision. (ii) The UID method with altitude measurements obtained from the barometer (A+UID). This approach was considered in order to validate the observability result regarding the inclusion of altitude measurements.

Table 2 summarizes the results obtained in the foregoing experiment. In the table, the proposed method is indicated by (A+R), the UID+Altitude variant is indicated by (A+UID), and the undelayed inverse depth method by (UID). The following results have been computed for each method: i) number of the features initialized into the system state (NIF); ii) number of features deleted from the system state (NDF); iii) Number of features been tracked at each frame (FPF); iv) total time of execution (TTE) in miliseconds; and v) average mean absolute error (aMAE) of the vehicle position in meters. For computing the aMAE, the P4P trajectory has been used as an independent reference of the vehicle position.

As it would be expected, the UID method was able only to estimate the flight trajectory in a scaled manner. By means of the inclusion of altitude measurements, the

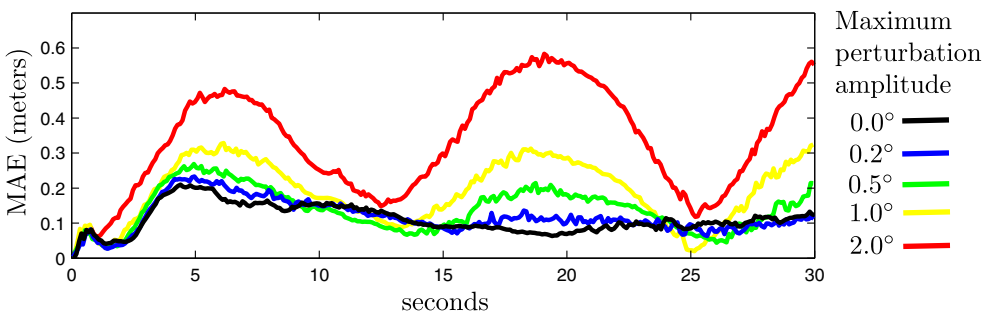

Fig. 10 Sensitivity analysis for the Gimbal control errors. The actual attitude (roll and pitch) of the Gimbal is perturbed with sinusoidal signals of different amplitudes. As it could be expected, the MAE in position increases as the angle control error of the Gimbal increases

Table 1 Average MAE obtained for different angle control errors of the Gimbal

Maximum control error (degrees)	average MAE (meters)
0.0°	0.111 ±0.040σ
±0.2°	0.114 ±0.046σ
±0.5°	0.133 ±0.060σ
±1.0°	0.185 ±0.087σ
±2.0°	0.327 ±0.147σ

Table 2 Results for flight trajectories (a)

Method	NIF	NDF	FPF	TTE (ms)	aMAE (m)
A+R	76.8±5.6σ	27.6±1.8σ	35.2±10.9σ	170±21σ	.25±.09σ
A+UID	89.6±4.2σ	32.4±3.5σ	43.2±11.3σ	197±11σ	.87±.54σ
UID	84.2±4.9σ	27.6±4.0σ	43.5±11.7σ	192±9σ	1.43±.73σ

A+UID method was able to partially recover the metric scale. However, it is important to recall that in this case the observability of the metric scale depends largely on the velocity of the vehicle in the vertical axis. Note that this particular flight trajectory presents considerable variations in altitude. On the other hand, with the inclusion of range measurements in the proposed method (A+R), a better concordance was found with the P4P visual reference. An additional experiment was carried out in order to validate the performance of the method by mean of a medium-scale flight trajectory. Figure 12 shows an aerial view of the estimated map and trajectory for this flight. The P4P technique can only be applied, for obtaining a trajectory reference, if the four marks placed on the floor are observed during all the flight. Therefore, for this experiment a GPS device equipped on the quadcopter was used to obtain a trajectory reference. For this experiment, the duration of the flight was about one minute and a half.

According to the experiments with simulations and with real data, it is interesting to note that the theoretical findings presented in Section 3 are well supported by the empirical results. In particular, the proposed method is able to recover the metric scale of the estimations in a reasonable manner. Also, it is shown that the proposed method is capable of working with the data obtained from low-cost sensors.

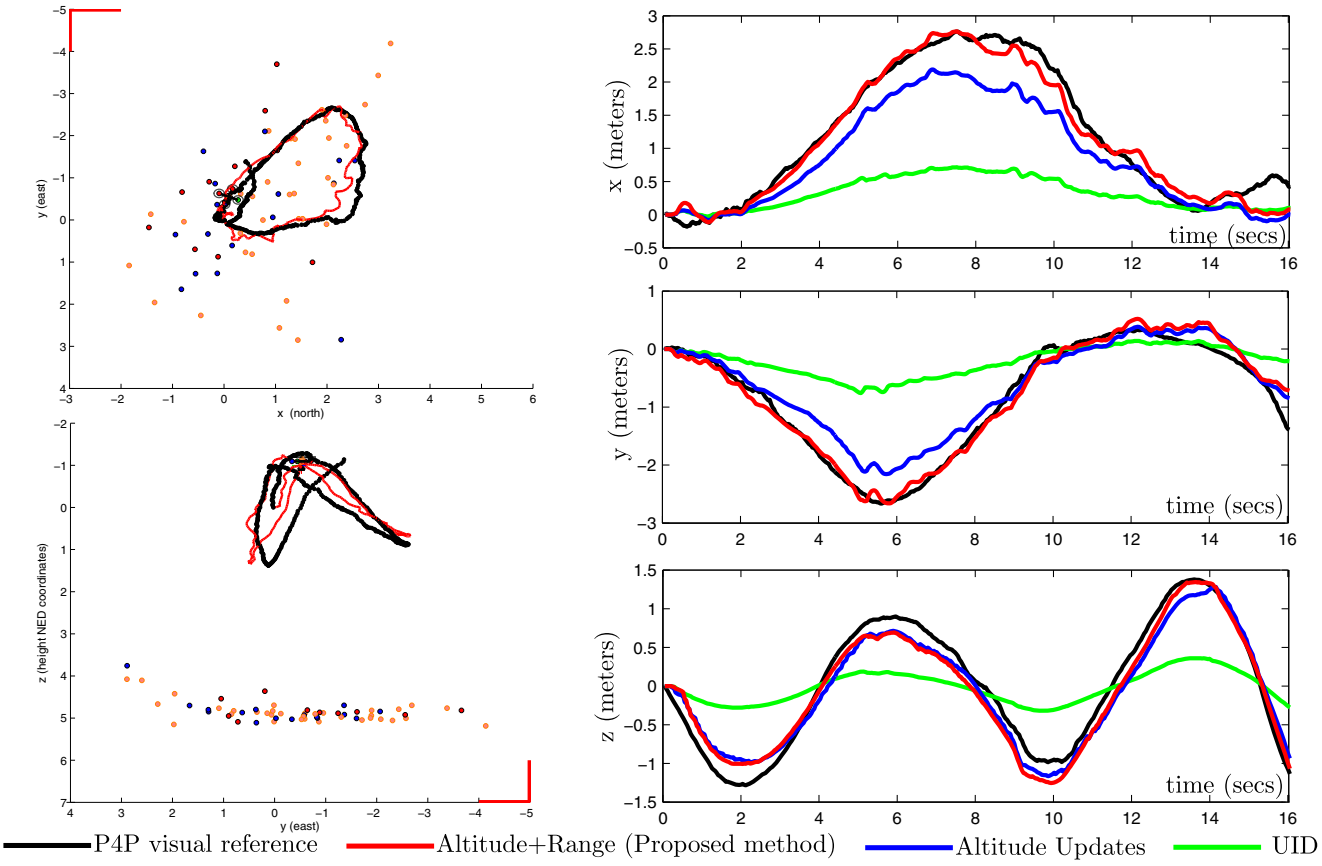

Fig. 11 Experimental results obtained with real data for a small-scale flight. Left plots show the sectional (z-x/y) view of maps and estimated trajectories for the proposed method. Right plots show the evolution over time for the estimated position expressed in each coordinate North, East, and Down (x, y, z), for the proposed method and the UID and A+UID methods

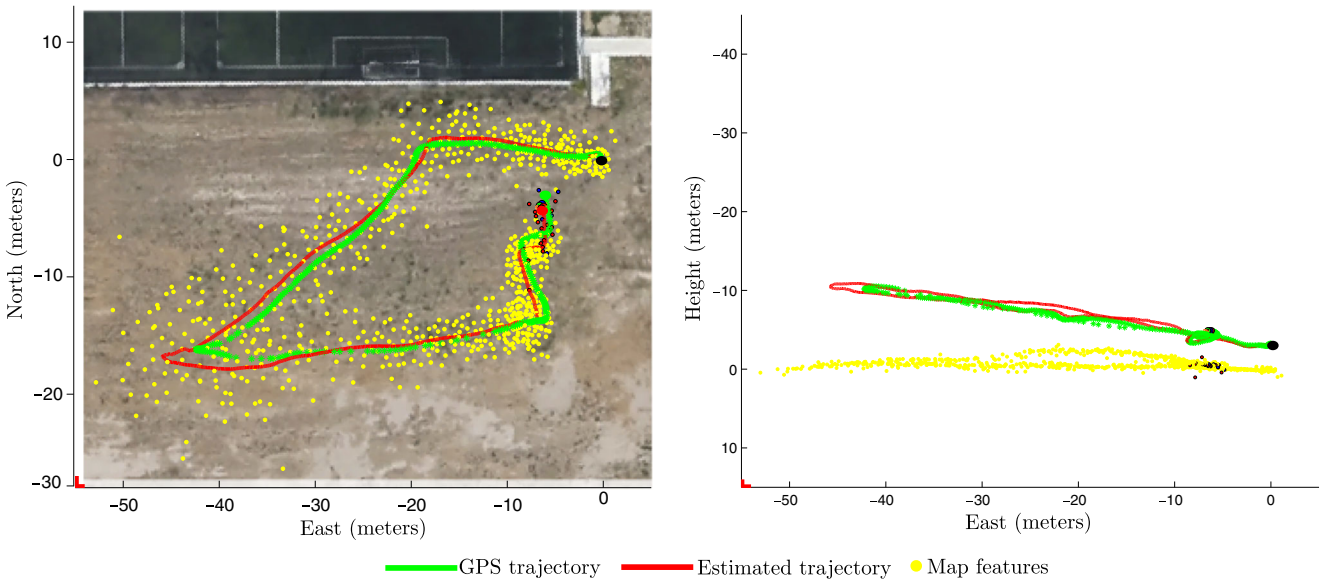

Fig. 12 Comparison between the trajectory estimated with the proposed method and the GPS trajectory for a medium-scale flight. Aerial top view (left plot). Lateral view (right plot)

6 Conclusions

This work presented a monocular SLAM system applied to micro aerial vehicles which is aided with altitude and range measurements. The proposed method is intended to be useful for performing visual-based local navigation in periods or circumstances where the GPS is not available or is unreliable.

One of the challenging aspects of working with monocular vision has to do with the impossibility of directly recovering the metric scale of the world.

In order to address this problem, the observability of the metric scale has been investigated by means of a nonlinear observability test. Two conditions of sufficiency were found from the observability test: i) The metric scale can become observable if measurements of altitude are included into the system and there is a movement of the vehicle along the vertical axis. ii) The metric scale can become observable if a measurement of the depth of a single map feature is available. Based on the theoretical findings, a novel technique for inferring the approximate depth of visual features from an ultrasonic range-finder is developed. Additionally, in the proposed method, the altitude of the vehicle is updated using the pressure measurements of a barometer.

Computer simulations as well as experiments with real data, obtained from a custom-built quadrotor equipped with low-cost sensors, have been carried out in order to validate the performance of the proposed method.

The experimental results confirm the theoretical findings and show that with the proposed system is possible to estimate the flight trajectory of the MAV, as well as a map of the environment, using only its onboard sensors.

Acknowledgements This research was funded by Spanish Science ministry project "ColRobTrans MINECO DPI2016-78957-R AEI/FEDER EU".

References

1. Parkinson, B.: Global positionig system: Theory and Applications. American Institute of Aeronautics and Astronautics, Reston (1996)
2. Zogg, J.: Essentials of satellite navigation, tech. rep., u-blox (2009)
3. Munguia, R., Urzua, S., Bolea, Y., Grau, A.: Vision-based slam system for unmanned aerial vehicles. Sensors **16**(3), 372 (2016)
4. Artieda, J., Sebastian, J.M., Campoy, P., Correa, J., Mondragon, I.F., Martinez, C., Olivares, M.: Visual 3-d slam from uavs. J. Intell. Robot. Syst. **55**, 299–321 (2009)
5. Weiss, S., Scaramuzza, D., Siegwart, R.: Monocular-slam based navigation for autonomous micro helicopters in gps-denied environments. J. Field Rob. **28**(6), 854–874 (2011)
6. Davison, A., Reid, I., Molton, N., Stasse, O.: Monoslam: Real-time single camera slam. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1052–1067 (2007)
7. Mirzaei, F., Roumeliotis, S.: A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation. IEEE Trans. Robot. **24**, 1143–1156 (2008)
8. Forster, C., Lynen, S., Kneip, L., Scaramuzza, D.: Collaborative monocular slam with multiple micro aerial vehicles. In: IROS, pp. 3962–3970. IEEE (2013)
9. Celik, K., Somani, A.K.: Monocular vision slam for indoor aerial vehicles. Journal of Electrical and Computer Engineering (2013)
10. Nutzi, G., Weiss, S., Scaramuzza, D., Siegwart, R.: Fusion of imu and vision for absolute scale estimation in monocular slam. J. Intell. Robot. Syst. **61**(1-4), 287–299 (2011)

11. Wang, C.-L., Wang, T.-M., Liang, J.-H., Zhang, Y.-C., Zhou, Y.: Bearing-only visual slam for small unmanned aerial vehicles in gps-denied environments. Int. J. Autom. Comput. **10**(5), 387–396 (2014)
12. Chowdhary, G., Johnson, E.N., Magree, D., Wu, A., Shein, A.: Gps-denied indoor and outdoor monocular vision aided navigation and control of unmanned aircraft. J. Field Rob. **30**(3), 415–438 (2013)
13. Civera, J., Davison, A.J., Montiel, J.M.M.: ch. Dimensionless Monocular SLAM, pp. 412–419. Lecture Notes in Computer Science (2007)
14. Hermann, R., Krener, A.J.: Nonlinear controllability and observability, IEEE Transactions on Automatic Control (1977)
15. Slotine, J.E., Li, W.: Applied nonlinear control. Prentice-Hall Englewood Cliffs, New Jersey (1991)
16. Munguia, R., Grau, A.: A practical method for implementing an attitude and heading reference system. International Journal of Advanced Robotic Systems, vol. 11 (2014)
17. Euston, M., Coote, P., Mahony, R., Kim, J., Hamel, T.: A complementary filter for attitude estimation of a fixed-wing uav, in Intelligent Robots and Systems. In: IEEE/RSJ International Conference on 2008. IROS 2008, pp. 340–345 (2008)
18. Bouguet, J.: Camera calibration toolbox for matlab, in online (2008)
19. Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: part i. IEEE Robot. Autom. Mag. **13**, 99–110 (2006)
20. Bailey, T., Durrant-Whyte, H.: Simultaneous localization and mapping (slam): part ii. IEEE Robot. Autom. Mag. **13**, 108–117 (2006)
21. Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: IEEE International Conference on Computer Vision, vol. 2, pp. 1508–1511 (2005)
22. Montiel, J.M.M., Civera, J., Davison, A.: Unified inverse depth parametrization for monocular slam. In: Proceedings of the Robotics Science and Systems Conference (2006)
23. Strasdat, H., Montiel, J., Davison, A.: Real-time monocular slam: Why filter? In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 2657–2664 (2010)
24. Civera, J., Davison, A., Montiel, J.: Inverse depth to depth conversion for monocular slam. In: 2007 IEEE International Conference on Robotics and Automation, pp. 2778–2783 (2007)
25. Corke, P.I.: Vision & robotics control: fundamental algorithms in matlab. Springer, Berlin (2011)
26. Community, O.S.: Ardupilot. http://ardupilot.com (2015)
27. Chatterjee, C., Roychowdhury, V.P.: Algorithms for coplanar camera calibration. Mach. Vis. Appl. **12**, 84–97 (2000)

Sarquis Urzua received an M.Sc. degree in mechanical engineering from the University of Guadalajara (UDG), Mexico in 2007. He is currently a Ph.D. candidate in the Department of Computer Science of the UDG. His research interest includes mobile robotics and unmanned aerial vehicles.

Rodrigo Munguía received the Ph.D. degree in automation and robotics from the Technical University of Catalonia (UPC), Spain in 2009. Currently, he is a titular professor with the Department of Computer Science of the University of Guadalajara, México. His research interests include mobile robotics, unmanned aerial vehicles, navigation systems for autonomous vehicles, automatic control, optimal state estimation and computer vision. He is a member of the researcher's national system of Mexico.

Antoni Grau received the M.Sc. and Ph.D. degrees in computer science from the Technical University of Catalonia (UPC), Barcelona, in 1990 and 1997, respectively. He is currently a Professor with the Department of Automatic Control, UPC, giving lectures on computer vision, digital signal processing, and robotics at the School of Informatics of Barcelona (FIB). His research areas are computer vision, pattern recognition, robotics, factory automation, and education on sustainable development. Dr. Grau is member of the International Association for Pattern Recognition and the International Federation of Automatic Control. He chaired several international conferences. He is also member of IEEE Industrial Electronics Society (IEEE-IES), co-chairing the Technical Subcommittee "FA11: Intelligent and Cooperative Robotics". He serves as Associate Editor of the Transactions on Industrial Informatics.