

Article

Cooperative Monocular-Based SLAM for Multi-UAV Systems in GPS-Denied Environments [†]

Juan-Carlos Trujillo ¹, Rodrigo Munguia ¹, Edmundo Guerra ² and Antoni Grau ^{2,*}

¹ Department of Computer Science, CUCEI, University of Guadalajara, Guadalajara 44430, Mexico; juancarlos_max@hotmail.com (J.-C.T.); rodrigo.munguia@upc.edu (R.M.)

² Department of Automatic Control, Technical University of Catalonia UPC, 08034 Barcelona, Spain; edmundo.guerra@upc.edu

* Correspondence: antoni.grau@upc.edu

[†] This paper is an extended version of our paper published in the 5th International Symposium on Sensor Science (I3S 2017), Barcelona, Spain, 27–29 September 2017.

Received: 21 January 2018; Accepted: 19 April 2018; Published: date



Abstract: This work presents a cooperative monocular-based SLAM approach for multi-UAV systems that can operate in GPS-denied environments. The main contribution of the work is to show that, using visual information obtained from monocular cameras mounted onboard aerial vehicles flying in formation, the observability properties of the whole system are improved. This fact is especially notorious when compared with other related visual SLAM configurations. In order to improve the observability properties, some measurements of the relative distance between the UAVs are included in the system. These relative distances are also obtained from visual information. The proposed approach is theoretically validated by means of a nonlinear observability analysis. Furthermore, an extensive set of computer simulations is presented in order to validate the proposed approach. The numerical simulation results show that the proposed system is able to provide a good position and orientation estimation of the aerial vehicles flying in formation.

Keywords: state estimation; unmanned aerial vehicle; monocular vision; localization; mapping; observability; cooperative

1. Introduction

Nowadays, unmanned aerial vehicles (UAVs) have received great attention from the robotics research community. In this case, one of the main objectives has been the improvement of the autonomy of these systems. In particular, the multi-rotor aerial systems allow great versatility of movements, making this kind of aerial platform very useful for a great variety of applications. Altogether with the recent advances in computational processing, computer vision has become an important tool in order to improve the autonomy of robotics systems. Cameras are well adapted for embedded systems because they are inexpensive, lightweight and power-saving. For instance, applications of surveillance [1], tracking and rescue [2], among others, seem to be feasible for aerial robots equipped with onboard cameras.

A fundamental requirement in order to improve the autonomy of an aerial robot has to do with the capacity of self-location and perception of the operational environment. In this case, for most applications, GPS (Global Positioning System) still represents the main alternative for addressing the localization problem. Nevertheless, the use of GPS presents some drawbacks, for instance, the precision error can be substantial, and it provides poor operability due to multipath propagation. However, several mission profiles require the UAVs to fly in GPS-challenging or GPS-denied environments, as in natural and urban canyons [3]. The use of range sensors like laser, sonar or radar (see [4–6])

allows obtaining knowledge about the environment of the robot. However, this kind of sensor can be expensive and sometimes heavy, and its use in outdoor environments can be somewhat limited. For instance, sonar systems have a limited range of operation. Active laser systems (e.g., LiDAR) represent a very interesting sensing technology; they can operate under any visibility condition (i.e., both day and night, unlike cameras) and can directly provide 3D measurements about the surrounding environment. On the other hand, LiDAR is generally expensive; it can overload the system for certain applications; and it has moving parts, which can generate error.

1.1. Related Work

Visual SLAM is a technique that makes use of visual features as landmarks. Visual SLAM is intended to address the navigation problem of a robot moving in a previously unknown environment, while it provides information about the environment, using mainly angular measurements obtained from cameras. Currently, there are two main approaches for implementing vision-based SLAM systems: (i) filtering-based methods [7–9] and (ii) the optimization-based methods [10,11]. While the latter approach is shown to give accurate results when the availability of computational power is enough, filtering-based SLAM methods might be still beneficial if limited processing power is available [12].

Some examples of visual SLAM approaches applied to unmanned aerial vehicles are [13,14]. In [15], a visual SLAM proposal that adds inertial measurements given by an Inertial Measurement Unit (IMU) is presented. The potential problem with this kind of approach is related to the fact that the acceleration obtained from the IMU has a dynamic bias, which is difficult to estimate. In [16], an EKF-based (Extended Kalman Filter) method is proposed in order to perform visual odometry with an unmanned aircraft. This method uses inertial sensors, a monocular downward facing camera and a range sensor (sonar altimeter). Unlike vision-based SLAM, in visual odometry approaches, there is not a mapping process. Furthermore, in those approaches, the operating altitude of the UAV is limited by the operating range of the sonar. More recently, new approaches appeared addressing the problem of visual-based navigation in GPS-denied environments, such as [17–19].

Multi-robot systems have also received great attention from the robotics research community. This attention is motivated by the inherent versatility that this kind of system has for performing tasks that could be difficult for a single robot. The use of several robots shows advantages like cost reductions, more robustness, better performance and efficiency [20,21]. In the case of the SLAM problem, in [22,23], a centralized architecture is used where all vehicles send their sensor data to a unique Kalman filter. In [16,24,25], the idea of combining monocular SLAM with cooperative, multi-UAV information to improve navigation capabilities in GPS-challenging environments is presented.

In works like [26–29], it has been shown that 6DOF-SLAM (six degrees of freedom), based only on angular measurements (i.e., monocular SLAM), is a partially observable system that can be applied to both the single-robot case and the multi-robot case. In [30], cooperative localization with visual information is addressed. According to the analysis presented in that work, the proposed system is completely observable. However, in this case, only distances and the relative orientations between robots are estimated. This fact can represent a clear drawback for applications where global measurements of the system are required (e.g., absolute position).

1.2. Objectives and Contributions

In this work, nonlinear observability properties of an aerial multi-robot system are analyzed. Based on this analysis, it is shown that the observability properties of this kind of system are improved by the inclusion of measurements of the relative distance between the aerial robots. Furthermore, based on the observability analysis, it is shown that the cooperative approach has theoretical advantages with respect to other configurations like the single-robot monocular SLAM approach. In addition, it is demonstrated that in a system composed of several UAVs, the observation of common landmarks is a sufficient condition in order to propagate through the whole system the information provided by

the measurement of the relative distance between two robots. This property allows flexibility in the system as opposed to the absolute need for multiple contacts between robots.

In order to take advantage of all the above theoretical results, in this work, a novel cooperative monocular-based SLAM approach for multi-UAV systems is proposed. The system model is composed of the dynamics of each aerial robot and the Euclidean position of each landmark. The measurements of the system are the projections of the landmarks in the images, provided by the monocular cameras carried individually in every aerial robot. Additionally, as was mentioned before, the availability of some measurements about the relative distance between two robots is assumed.

In order to accomplish the requirement of having measurements of the relative distance between two robots, a technique based on a homography is also presented in this research. The main idea is to exploit the physical structure of the aerial robots in order to obtain measurements of relative distances by means of visual information. In this case, the method is developed assuming a team of quadrotors. It is important to remark that this proposed approach could be also applied to many other aerial platforms. The only requirement for the presented approach is that at least one robot has to be maintained inside the field of view of another aerial robot, while sharing the observation of one common visual landmark (see Figure 1).

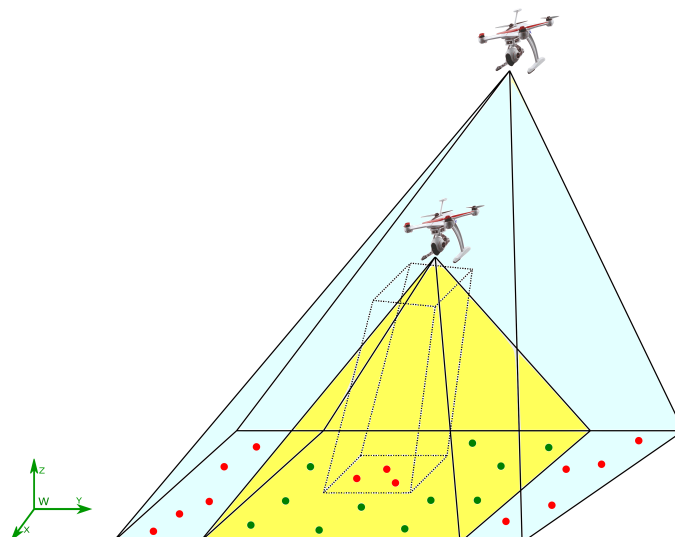


Figure 1. Cooperative monocular-based SLAM.

The geometric structure of a typical quadrotor is cross-shaped, and therefore, each rotor is mounted at the different ends of the cross. This kind of physical geometry can allow a standard computer vision algorithm to extract and track the centroids of the rotors. In this case, those centroids can be assumed to be coplanar. In order to compute the relative distance from one quadcopter in the field of view of another one, a homography is applied from the camera coordinate reference system of the observing robot to the plane formed by the four rotors of the robot being observed. The information obtained by the homography is fused with the orientation of the observing robot, provided by an IMU, which finally allows one to obtain measurements of relative distances. It is important to note that, based on the theoretical results presented in this work, it should be straightforward to replace the homographic technique used for estimating the relative distance between UAVs by another technique that would provide a similar measurement.

In addition to the benefit of improving the observability of the system, the relative distance obtained between any pair of robots provides metric information of the system, which is an important issue to be addressed in monocular-based systems. For example, in other configurations, the metric information is obtained purely from inertial systems (i.e., monocular/Inertial Navigation Systems

(INS) solutions), but inertial sensors present some drawbacks due to the large drift bias, which is inherent to this kind of sensor [31,32].

In the proposed system, in order to take advantage of the multi-UAV cooperative configuration, the initialization process of new map features is carried out through a pseudo-stereo system composed of two monocular cameras mounted on two UAVs respectively (one camera per UAV) that observe common landmarks. This approach allows initializing landmarks with less uncertainty than a pure monocular system since 3D information of the position of landmarks is gathered from the beginning of the observation. It is well known that, in visual SLAM, the initialization process can play an important role in the convergence of the filter. Having a flexible baseline in the pseudo-stereo system allows one to initialize landmarks at distances far away with less uncertainty, unlike stereo systems with a rigid baseline [32] or delayed monocular initialization methods. The above fact allows the proposed cooperative system to have better performance in environments where landmarks are far from the measurement system, contrary to SLAM approaches based on depth cameras, stereo systems, monocular cameras or sonars.

1.3. Paper Outline

The document is organized in the following manner: Section 2 presents the specifications of the system; Section 3 presents the nonlinear observability analysis that represents the theoretical basis of the proposed method; Section 4 presents the proposed cooperative approach for monocular-based SLAM; in Section 5, the results obtained from numerical simulations are presented in order to validate the proposal, and finally, in Section 6, some final remarks are given.

2. System Specification

In this section, the models used in this work are introduced. The model used for representing the dynamics of a camera carried by a quadcopter is presented. The representation of the landmarks as map features is also defined. The camera projection model used in this work is described. The technique based on homographies that is used for estimating the relative distance between two quadcopters is introduced, as well.

2.1. Dynamics of the System

Let us consider the following continuous-time model describing the dynamics of the j -th UAV-camera system (see Figure 2):

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{x}}_c^j \\ \dot{\mathbf{q}}_c^j \\ \dot{\mathbf{v}}_c^j \\ \dot{\boldsymbol{\omega}}_c^j \\ \dot{\mathbf{x}}_a^i \end{bmatrix} = \begin{bmatrix} \mathbf{v}_c^j \\ \frac{1}{2}\boldsymbol{\Omega}_c^j \mathbf{q}_c^j \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \end{bmatrix} \quad (1)$$

where the state vector \mathbf{x} is defined by:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_c^j & \mathbf{q}_c^j & \mathbf{v}_c^j & \boldsymbol{\omega}_c^j & \mathbf{x}_a^i \end{bmatrix}^T \quad (2)$$

With $i = 1, \dots, n_1$, let n_1 be the number of landmarks, and with $j = 1, \dots, n_2$, let n_2 be the number of UAV-camera systems.

Additionally, let $\mathbf{x}_c^j = \begin{bmatrix} x_c^j & y_c^j & z_c^j \end{bmatrix}^T$ represent the position of the reference system C of the j -th camera, with respect to the reference system W . Let $\mathbf{q}_c^j = \begin{bmatrix} q_0^j & q_x^j & q_y^j & q_z^j \end{bmatrix}^T$ be a unit quaternion representing the orientation of the reference system C of the j -th camera, with respect to

the reference system W . Let $\mathbf{v}_c^j = [\dot{x}_c^j \ \dot{y}_c^j \ \dot{z}_c^j]^T$ represent the linear velocity of the j -th camera. Let $\boldsymbol{\omega}_c^j = [\omega_x^j \ \omega_y^j \ \omega_z^j]^T$ represent the angular velocity of the j -th camera. Finally, let $\mathbf{x}_a^i = [x_a^i \ y_a^i \ z_a^i]^T$ be the position of the i -th landmark with respect to the reference system W , defined by its Euclidean parameterization. Furthermore the next definitions should be considered:

$$\boldsymbol{\Omega}_c^j = \begin{bmatrix} 0 & -\boldsymbol{\omega}_c^{jT} \\ \boldsymbol{\omega}_c^j & [\boldsymbol{\omega}_c^j]_{\times} \end{bmatrix}, \quad [\boldsymbol{\omega}_c^j]_{\times} = \begin{bmatrix} 0 & -\omega_z^j & \omega_y^j \\ \omega_z^j & 0 & -\omega_x^j \\ -\omega_y^j & \omega_x^j & 0 \end{bmatrix} \quad (3)$$

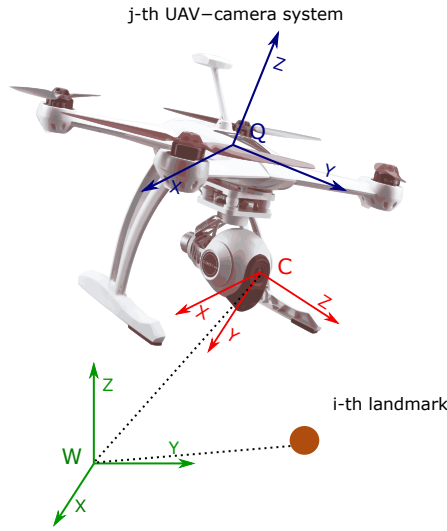


Figure 2. Coordinate reference systems.

In (1), it is assumed that every UAV-camera is moving freely in the space with six degrees of freedom: three for translation and three for rotation. Furthermore, note that a non-acceleration model is assumed for UAV-camera systems, and the landmarks are assumed to remain static.

2.2. Camera Measurement Model

Consider the projection of a specific landmark over the image plane of a camera. Using the pinhole model [33] (see Figure 3), the following expression can be defined:

$${}^i\mathbf{z}_c^j = {}^i\mathbf{h}_c^j = \begin{bmatrix} {}^i u_c^j \\ {}^i v_c^j \end{bmatrix} = \frac{1}{z_d^j} \begin{bmatrix} \frac{f_c^j}{d_u^j} & 0 \\ 0 & \frac{f_c^j}{d_v^j} \end{bmatrix} \begin{bmatrix} {}^i x_d^j \\ {}^i y_d^j \end{bmatrix} + \begin{bmatrix} c_u^j + d_{ur}^j + d_{ut}^j \\ c_v^j + d_{vr}^j + d_{vt}^j \end{bmatrix} \quad (4)$$

Let $[{}^i u_c^j, {}^i v_c^j]$ define the coordinates (in pixels) of the projection of the i -th landmark over the image of the j -th camera. Let f_c^j be the focal length (in meters) of the j -th camera. Let $[d_u^j, d_v^j]$ be the conversion parameters (in m/pixel) for the j -th camera. Let $[c_u^j, c_v^j]$ be the coordinates (in pixels) of the image central point of the j -th camera. Let $[d_{ur}^j, d_{vr}^j]$ be components (in pixels) accounting for the radial distortion of the j -th camera. Let $[d_{ut}^j, d_{vt}^j]$ be components (in pixels) accounting for the tangential distortion of the j -th camera. All the intrinsic parameters of the j -th camera are assumed to be known

by means of some calibration method. Let ${}^i\mathbf{p}_d^j = [{}^i x_d^j \ {}^i y_d^j \ {}^i z_d^j]^T$ represent the position (in meters) of the i -th landmark with respect to the coordinate reference system C of the j -th camera. Additionally,

$${}^i\mathbf{p}_d^j = {}^W\mathbf{R}_c^j(\mathbf{x}_a^i - \mathbf{x}_c^j) \quad (5)$$

Let ${}^W\mathbf{R}_c^j(\mathbf{q}_c^j) \in SO3$ be the rotation matrix, obtained from the quaternion \mathbf{q}_c^j , that transforms the world coordinate reference system W to the coordinate reference system C of the j -th camera.

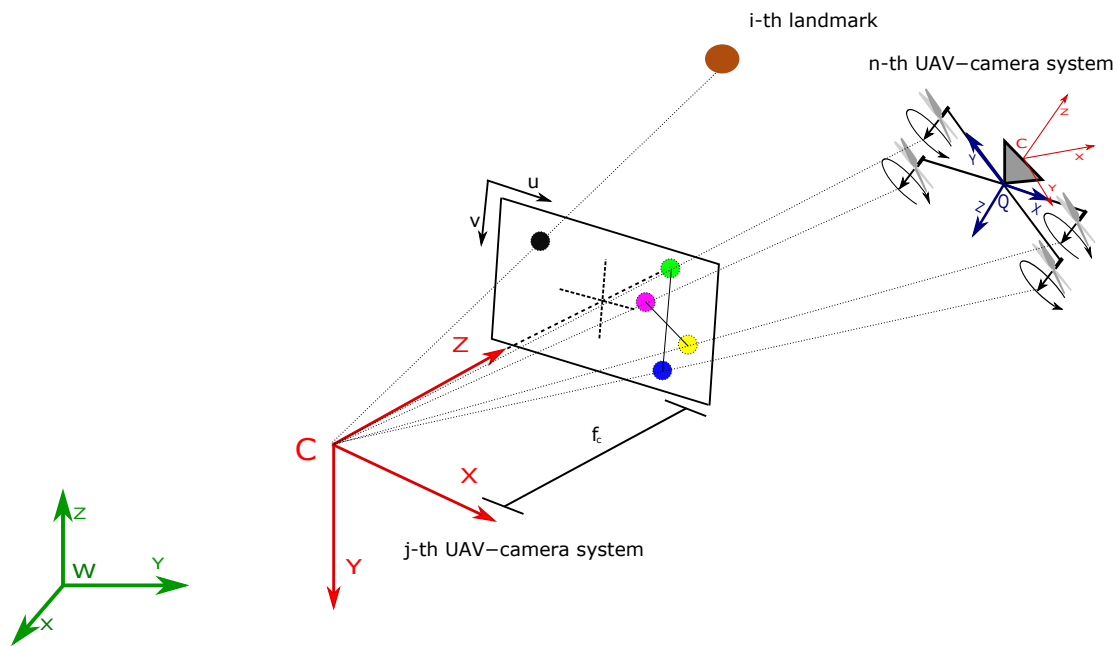


Figure 3. Pinhole camera projection model.

2.3. Relative Distance Measurement Model

To estimate the relative distance between UAV-camera systems, the physical structure of the aerial robots (quadcopters in this research) is exploited. In this case, the rotors of a quadcopter are considered as highlighted points in the images captured by another quadcopter (see Figures 1 and 3). These points can be considered coplanar in the reference system Q of the n -th quadcopter. Therefore, knowing their geometry, it is possible to formulate a homography [33]. In order to determine the exact correspondences between the motors' positions in the image plane and their real positions in reference Q , it is assumed that one rotor has a distinguishable color or geometry from the other ones. The other three correspondences can be determined given only the first one because it is also assumed that the quadrotor will not fly upside down. The homographic process will allow estimating the distance between the reference system of the camera to the plane to which the four points belong.

According to Equation (4), the following expression is obtained:

$${}^j\gamma_m^n \begin{bmatrix} {}^j u_m^n \\ {}^j v_m^n \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_c^j & \mathbf{0}_{3 \times 1} \end{bmatrix} {}^j\mathbf{E}_c^n \begin{bmatrix} x_m^n \\ y_m^n \\ z_m^n \\ 1 \end{bmatrix} \quad (6)$$

With $m = \{1, \dots, 4\}$, let $[{}^j u_m^n, {}^j v_m^n]$ define the coordinates (in pixels) of the projection of the m -th point of the n -th quadcopter over the image of the j -th camera. Let $[x_m^n, y_m^n, z_m^n]$ represent the position

of the m -th point with respect to the reference system Q of the n -th quadcopter, and let ${}^j\gamma_m^n$ be a scale factor. Additionally, it is defined:

$$\mathbf{T}_c^j = \begin{bmatrix} \frac{f_c^j}{d_u^j} & 0 & c_u^j + d_{ur}^j + d_{ut}^j \\ 0 & \frac{f_c^j}{d_v^j} & c_v^j + d_{vr}^j + d_{vt}^j \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

$${}^j\mathbf{E}_c^n = \begin{bmatrix} {}^j\mathbf{R}_q^n & {}^j\mathbf{d}_q^n \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (8)$$

Let ${}^j\mathbf{d}_q^n$ be the translation vector from the reference system Q of the n -th quadcopter to the reference system C of the j -th camera. Let ${}^j\mathbf{R}_q^n \in SO3$ be the rotation matrix that transforms the coordinate reference system Q of the n -th quadcopter to the coordinate reference system C of the j -th camera. The assumption that the four m points are coplanar implies that $z_m^n = 0$; therefore, Equation (6) can take the following form:

$$\begin{bmatrix} {}^ju_m^n \\ {}^jv_m^n \\ 1 \end{bmatrix} = \frac{1}{{}^j\gamma_m^n} \begin{bmatrix} \mathbf{T}_c^j & \mathbf{0}_{3 \times 1} \end{bmatrix} {}^j\mathbf{E}_c^n \begin{bmatrix} x_m^n \\ y_m^n \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} {}^js_{11}^n & {}^js_{12}^n & {}^js_{13}^n \\ {}^js_{21}^n & {}^js_{22}^n & {}^js_{23}^n \\ {}^js_{31}^n & {}^js_{32}^n & {}^js_{33}^n \end{bmatrix} \begin{bmatrix} x_m^n \\ y_m^n \\ 1 \end{bmatrix} = {}^j\mathbf{S}_c^n \begin{bmatrix} x_m^n \\ y_m^n \\ 1 \end{bmatrix} \quad (9)$$

where ${}^j\mathbf{S}_c^n$ is a non-singular homogeneous matrix. In this case, it is allowed to scale the matrix in such a way that ${}^js_{33}^n = 1$. This fact does not affect the projective transformation [33]. Therefore, the matrix ${}^j\mathbf{S}_c^n$ can be redefined as:

$${}^j\mathbf{S}_c^n = \begin{bmatrix} {}^js_{11}^n & {}^js_{12}^n & {}^js_{13}^n \\ {}^js_{21}^n & {}^js_{22}^n & {}^js_{23}^n \\ {}^js_{31}^n & {}^js_{32}^n & 1 \end{bmatrix} = \begin{bmatrix} {}^j\mathbf{s}_1^n & {}^j\mathbf{s}_2^n & {}^j\mathbf{s}_3^n \end{bmatrix} \quad (10)$$

In (10), the values of ${}^j\mathbf{S}_c^n$ are unknown; therefore, the following equation system can be formed from (9):

$${}^j\mathbf{G}_m^n {}^j\mathbf{t}_c^n = {}^j\mathbf{g}_m^n \quad (11)$$

where:

$${}^j\mathbf{G}_m^n = \begin{bmatrix} x_m^n & y_m^n & 1 & 0 & 0 & 0 & -x_m^n {}^ju_m^n & -y_m^n {}^ju_m^n \\ 0 & 0 & 0 & x_m^n & y_m^n & 1 & -x_m^n {}^jv_m^n & -y_m^n {}^jv_m^n \end{bmatrix} \quad (12)$$

$${}^j\mathbf{t}_c^n = \begin{bmatrix} {}^js_{11}^n & {}^js_{12}^n & {}^js_{13}^n & {}^js_{21}^n & {}^js_{22}^n & {}^js_{23}^n & {}^js_{31}^n & {}^js_{32}^n \end{bmatrix}^T \quad (13)$$

$${}^j\mathbf{g}_m^n = \begin{bmatrix} {}^ju_m^n & {}^jv_m^n \end{bmatrix}^T \quad (14)$$

Considering the projection of the four points, the solution to the system can be given as follows:

$${}^j\mathbf{t}_c^n = ({}^j\mathbf{G}_c^n)^{-1} {}^j\mathbf{g}_c^n \quad (15)$$

with:

$${}^j\mathbf{G}_c^n = \begin{bmatrix} {}^j\mathbf{G}_1^n & {}^j\mathbf{G}_2^n & {}^j\mathbf{G}_3^n & {}^j\mathbf{G}_4^n \end{bmatrix}^T \quad {}^j\mathbf{g}_c^n = \begin{bmatrix} {}^j\mathbf{g}_1^n & {}^j\mathbf{g}_2^n & {}^j\mathbf{g}_3^n & {}^j\mathbf{g}_4^n \end{bmatrix}^T \quad (16)$$

From the method proposed in [34], where the orthonormality property of a rotation matrix is exploited and knowing the intrinsic parameters of the camera, ${}^j\mathbf{R}_q^n$ and ${}^j\mathbf{d}_q^n$ can be computed from (15) as follows:

$${}^j\mathbf{R}_q^n = \begin{bmatrix} {}^j\mathbf{r}_1^n & {}^j\mathbf{r}_2^n & {}^j\mathbf{r}_1^n \times {}^j\mathbf{r}_2^n \end{bmatrix} \quad (17)$$

$${}^j\mathbf{d}_q^n = {}^j\mathbf{r}_3^n \quad (18)$$

with:

$$\begin{bmatrix} {}^j\mathbf{r}_1^n & {}^j\mathbf{r}_2^n & {}^j\mathbf{r}_3^n \end{bmatrix} = \frac{1}{j\delta_c^n} (\mathbf{T}_c^j)^{-1} {}^j\mathbf{S}_c^n \quad (19)$$

and:

$$j\delta_c^n = \left\| (\mathbf{T}_c^j)^{-1} {}^j\mathbf{s}_1^n \right\| = \left\| (\mathbf{T}_c^j)^{-1} {}^j\mathbf{s}_2^n \right\| \quad (20)$$

Finally, the distance between the j -th camera and the n -th camera can be computed from:

$${}^j\mathbf{z}_r^n = {}^j\mathbf{h}_r^n = \mathbf{x}_c^n - \mathbf{x}_c^j = ({}^W\mathbf{R}_c^j)^T ({}^j\mathbf{R}_q^n \mathbf{d}_c^n + {}^j\mathbf{d}_q^n) \quad (21)$$

where \mathbf{d}_c^n is the translation vector of the reference system Q to the reference system C of the n -th UAV-camera system. This parameter is obtained by a camera-robot calibration process. The rotation matrix ${}^W\mathbf{R}_c^j$ can be obtained from an Attitude and Heading Reference System (AHRS) or an inertial measurement unit [35,36] mounted on board the j -th UAV-camera system.

3. Observability Analysis

In this section, the nonlinear observability properties of an aerial multi-robot system are studied. Observability is an inherent property of a dynamic system and has an important role in the accuracy of its estimation process; moreover, this fact has important consequences in the context of SLAM.

A system is defined as observable if the initial state \mathbf{x}_0 , at any initial time t_0 , can be determined given the state transition and observation models of the system and observations $\mathbf{z}[t_0, t]$, from time t_0 to a finite time t . In [37], it is demonstrated that a non-linear system is locally weakly observable if the observability rank condition $\text{rank}(\mathcal{O}) = \dim(\mathbf{x})$ is verified, where \mathcal{O} is the observability matrix.

As previously mentioned, 6DOF-monocular SLAM represents a kind of partially-observable system with a high number of unobservable modes and states that can be applied to both the single-robot case and the multi-robot case. The following references are examples of works where the problem of the observability of 6DOF-monocular SLAM systems has also been studied, such as [26–29].

For the analysis developed in this work, and for the sake of simplicity, the system (1) is redefined as:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) = \begin{bmatrix} \dot{\mathbf{x}}_c^j \\ \dot{\lambda}_c^j \\ \dot{\mathbf{v}}_c^j \\ \dot{\boldsymbol{\omega}}_c^j \\ \dot{\mathbf{x}}_a^i \end{bmatrix} = \begin{bmatrix} \mathbf{v}_c^j \\ \boldsymbol{\omega}_c^j \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \end{bmatrix} \quad (22)$$

Let $\boldsymbol{\lambda}_c^j = [\phi_c^j \ \theta_c^j \ \psi_c^j]^T$ be the Euler angles of the j -th camera with respect to the coordinate system W .

The observability matrix \mathcal{O} can be computed as:

$$\mathcal{O} = \begin{bmatrix} \frac{\partial(\mathcal{L}_f^0(\mathbf{h}_c^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(\mathbf{h}_c^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_f^0(\mathbf{h}_r^n))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(\mathbf{h}_r^n))}{\partial \mathbf{x}} \end{bmatrix}^T \quad (23)$$

Let $\mathcal{L}_f^s \mathbf{h}$ be the s -th-order Lie derivative [38] of the scalar field \mathbf{h} with respect to the vector field \mathbf{f} . For example, in (23), the zero-order and first-order Lie derivatives are used for each measurement.

For the measurement given by a monocular camera, according to (4) and (22), the following zero-order Lie derivative can be defined:

$$\frac{\partial(\mathcal{L}_f^0(i\mathbf{h}_c^j))}{\partial \mathbf{x}} = \left[\mathbf{0}_{2 \times 12(j-1)} \quad \mathbf{H}_x^j \quad \mathbf{0}_{2 \times 12(n_2-j)} \quad | \quad \mathbf{0}_{2 \times 3(i-1)} \quad \mathbf{H}_a^i \quad \mathbf{0}_{2 \times 3(n_1-i)} \right] \quad (24)$$

where:

$$\mathbf{H}_x^j = \mathbf{H}_c \begin{bmatrix} -{}^W\mathbf{R}_c^j & [{}^i\mathbf{p}_d^j \times] & \mathbf{0}_{2 \times 6} \end{bmatrix} \quad (25)$$

$$\mathbf{H}_a^i = \mathbf{H}_c {}^W\mathbf{R}_c^j \quad (26)$$

and:

$$\mathbf{H}_c = \frac{f_c^j}{i z_d^j} \begin{bmatrix} i z_d^j & 0 & -i x_d^j \\ 0 & i z_d^j & -i y_d^j \end{bmatrix} \quad (27)$$

Note that $[\cdot \times]$ denotes the antisymmetric matrix of the vector (\cdot) . The first-order Lie derivative can also be defined in the following:

$$\frac{\partial(\mathcal{L}_f^1(i\mathbf{h}_c^j))}{\partial \mathbf{x}} = \left[\mathbf{0}_{2 \times 12(j-1)} \quad \mathbf{H}_{dx}^j \quad \mathbf{0}_{2 \times 12(n_2-j)} \quad | \quad \mathbf{0}_{2 \times 3(i-1)} \quad \mathbf{H}_{da}^i \quad \mathbf{0}_{2 \times 3(n_1-i)} \right] \quad (28)$$

where:

$$\mathbf{H}_{dx}^j = \begin{bmatrix} \mathbf{X}^j & \mathbf{\Psi}^j & -\mathbf{H}_c {}^W\mathbf{R}_c^j & \mathbf{H}_c [{}^i\mathbf{p}_d^j \times] \end{bmatrix} \quad (29)$$

$$\mathbf{H}_{da}^i = \begin{bmatrix} \mathbf{H}_c^1 & \mathbf{H}_c^2 & \mathbf{H}_c^3 \end{bmatrix} {}^W\mathbf{R}_c^j (-{}^W\mathbf{R}_c^j \mathbf{v}_c^j + [{}^i\mathbf{p}_d^j \times] \boldsymbol{\omega}_c^j) - \mathbf{H}_c [\boldsymbol{\omega}_c^j \times] {}^W\mathbf{R}_c^j \quad (30)$$

with:

$$\mathbf{X}^j = -\mathbf{H}_{da}^i \quad (31)$$

$$\mathbf{\Psi}^j = \begin{bmatrix} \mathbf{H}_c^1 & \mathbf{H}_c^2 & \mathbf{H}_c^3 \end{bmatrix} [{}^i\mathbf{p}_d^j \times] (-{}^W\mathbf{R}_c^j \mathbf{v}_c^j + [{}^i\mathbf{p}_d^j \times] \boldsymbol{\omega}_c^j) - \mathbf{H}_c [{}^W\mathbf{R}_c^j \mathbf{v}_c^j \times] - \mathbf{H}_c [\boldsymbol{\omega}_c^j \times] [{}^i\mathbf{p}_d^j \times] \quad (32)$$

and:

$$\mathbf{H}_c^1 = \frac{f_c^j}{i z_d^j} \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad (33)$$

$$\mathbf{H}_c^2 = \frac{f_c^j}{i z_d^j} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (34)$$

$$\mathbf{H}_c^3 = \frac{f_c^j}{i z_d^j} \begin{bmatrix} -i z_d^j & 0 & 2i x_d^j \\ 0 & -i z_d^j & 2i y_d^j \end{bmatrix} \quad (35)$$

Considering the case where relative measurements of the distance between robots are available, the following statement can be defined from (21) and (22):

For the zero-order Lie derivative, if $j < n$ (the index of the observing robot is lesser than the index of the observed robot):

$$\frac{\partial(\mathcal{L}_f^0(j\mathbf{h}_r^n))}{\partial \mathbf{x}} = \left[\mathbf{0}_{3 \times 12(j-1)} \quad -\mathbf{M}_x^j \quad \mathbf{0}_{3 \times 12(n-j-1)} \quad \mathbf{M}_x^n \quad \mathbf{0}_{3 \times 12(n_2-n)} \quad | \quad \mathbf{0}_{3n_1} \right] \quad (36)$$

On the other hand, if $j > n$, then:

$$\frac{\partial(\mathcal{L}_f^0(j\mathbf{h}_r^n))}{\partial \mathbf{x}} = \left[\mathbf{0}_{3 \times 12(n-1)} \quad \mathbf{M}_x^n \quad \mathbf{0}_{3 \times 12(j-n-1)} \quad -\mathbf{M}_x^j \quad \mathbf{0}_{3 \times 12(n_2-j)} \quad | \quad \mathbf{0}_{3n_1} \right] \quad (37)$$

and:

$$\mathbf{M}_{\mathbf{x}}^{j,n} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3 \times 9} \end{bmatrix} \quad (38)$$

where \mathbf{I} is the identity matrix.

For the first-order Lie derivative, if $j < n$:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^1(j\mathbf{h}_{\mathbf{r}}^n))}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{3 \times 12(j-1)} & -\mathbf{M}_{\mathbf{dx}}^j & \mathbf{0}_{3 \times 12(n-j-1)} & \mathbf{M}_{\mathbf{dx}}^n & \mathbf{0}_{3 \times 12(n_2-n)} & | & \mathbf{0}_{3n_1} \end{bmatrix} \quad (39)$$

On the other hand, if $j > n$ (the index of the observing robot is higher than the index of the observed robot), then:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^1(j\mathbf{h}_{\mathbf{r}}^n))}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{3 \times 12(n-1)} & \mathbf{M}_{\mathbf{dx}}^n & \mathbf{0}_{3 \times 12(j-n-1)} & -\mathbf{M}_{\mathbf{dx}}^j & \mathbf{0}_{3 \times 12(n_2-j)} & | & \mathbf{0}_{3n_1} \end{bmatrix} \quad (40)$$

with

$$\mathbf{M}_{\mathbf{dx}}^{j,n} = \begin{bmatrix} \mathbf{0}_{3 \times 6} & \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} \quad (41)$$

With the above considerations, the observability matrix for the proposed system (22) can be defined as follows:

$$\mathcal{O} = \begin{bmatrix} \mathbf{0}_{2 \times 12(j-1)} & \mathbf{H}_{\mathbf{x}}^j & \mathbf{0}_{2 \times 12(n_2-j)} & | & \mathbf{0}_{2 \times 3(i-1)} & \mathbf{H}_{\mathbf{a}}^i & \mathbf{0}_{2 \times 3(n_1-i)} \\ \mathbf{0}_{2 \times 12(j-1)} & \mathbf{H}_{\mathbf{dx}}^j & \mathbf{0}_{2 \times 12(n_2-j)} & | & \mathbf{0}_{2 \times 3(i-1)} & \mathbf{H}_{\mathbf{da}}^i & \mathbf{0}_{2 \times 3(n_1-i)} \\ \mathbf{0}_{3 \times 12(j-1)} & -\mathbf{M}_{\mathbf{x}}^j & \mathbf{0}_{3 \times 12(n-j-1)} & \mathbf{M}_{\mathbf{x}}^n & \mathbf{0}_{3 \times 12(n_2-n)} & | & \mathbf{0}_{3n_1} \\ \mathbf{0}_{3 \times 12(j-1)} & -\mathbf{M}_{\mathbf{dx}}^j & \mathbf{0}_{3 \times 12(n-j-1)} & \mathbf{M}_{\mathbf{dx}}^n & \mathbf{0}_{3 \times 12(n_2-n)} & | & \mathbf{0}_{3n_1} \\ & & & \vdots & & & \\ \mathbf{0}_{3 \times 12(n-1)} & \mathbf{M}_{\mathbf{x}}^n & \mathbf{0}_{3 \times 12(j-n-1)} & -\mathbf{M}_{\mathbf{x}}^j & \mathbf{0}_{3 \times 12(n_2-j)} & | & \mathbf{0}_{3n_1} \\ \mathbf{0}_{3 \times 12(n-1)} & \mathbf{M}_{\mathbf{dx}}^n & \mathbf{0}_{3 \times 12(j-n-1)} & -\mathbf{M}_{\mathbf{dx}}^j & \mathbf{0}_{3 \times 12(n_2-j)} & | & \mathbf{0}_{3n_1} \end{bmatrix} \quad (42)$$

The maximum rank of the observability matrix (42) is $\text{rank}(\mathcal{O}) = 3n_1 + 12n_2 - 3$, where n_1 is the number of landmarks being measured and n_2 is the number of robots. n_1 is multiplied by three, since this is the number of states per landmark given by the Euclidean parametrization. n_2 is multiplied by 12, since this is the number of states per robot given by its global position, orientation (Euler angles) and its derivatives. Therefore, \mathcal{O} will be rank deficient ($\text{rank}(\mathcal{O}) < \dim(\mathbf{x})$).

The unobservable modes are spanned by the right nullspace basis of the observability matrix \mathcal{O} ; therefore:

$$\mathbf{N} = \text{null}(\mathcal{O}) = \begin{bmatrix} \mathbf{I}_3^1 \\ \mathbf{0}_{9 \times 3}^1 \\ \vdots \\ \mathbf{I}_3^j \\ \mathbf{0}_{9 \times 3}^j \\ \hline \mathbf{I}_3^1 \\ \vdots \\ \mathbf{I}_3^i \end{bmatrix} \quad (43)$$

It is straightforward to verify that the right nullspace basis of \mathcal{O} spans for \mathbf{N} (i.e., $\mathcal{O}\mathbf{N} = \mathbf{0}$).

From (43), it can be seen that the system is partially observable and that the unobservable modes cross with the states that correspond to the global position of the robots and the landmarks; these states are unobservable. An important conclusion is that all the vectors of the right null space basis are orthogonal with the rest of the states, and therefore, these states are completely observable.

The results of the observability analysis are summarized in Table 1.

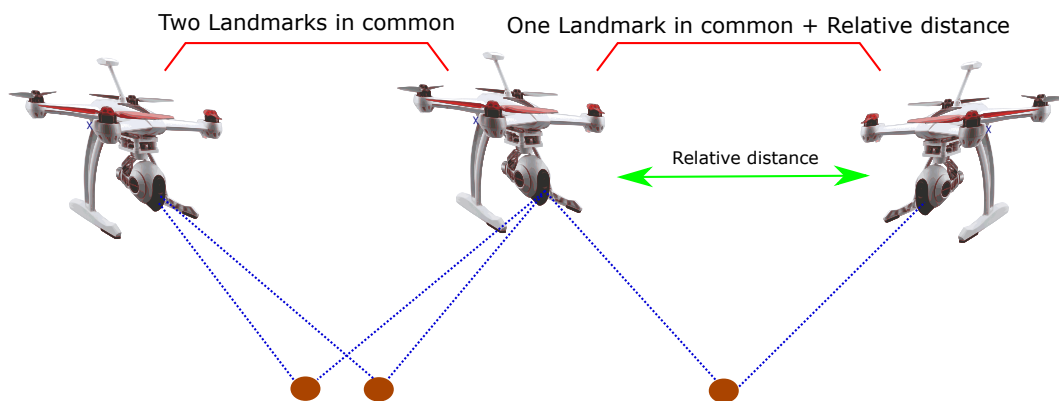
Table 1. Results of the non-linear observability analysis.

	Unobservable Modes	Unobservable States	Observable States
Monocular	5	$\mathbf{x}_c^j, \mathbf{v}_c^j, \mathbf{x}_a^i, \psi_c^j, \psi_c^i$	$\phi_c^j, \phi_c^i, \theta_c^j, \theta_c^i$
Cooperative	3	$\mathbf{x}_c^j, \mathbf{x}_a^i$	$\mathbf{v}_c^j, \lambda_c^j, \omega_c^j$

Some important remarks on the analysis can be extracted:

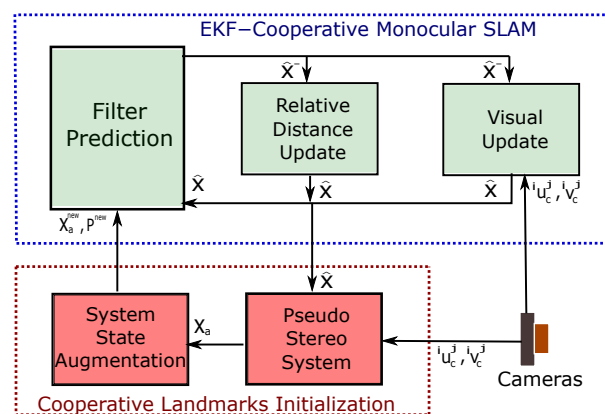
- In order to obtain the previous results, it is necessary to link the members of the multi-UAV system through the measurements (see Figure 4). In other words, (i) a robot needs to share the observation of at least two landmarks with another robot or (ii) a robot needs to measure its relative distance with respect to another robot in addition to both observing one landmark in common.
- A single measurement of the relative distance between two robots represents a sufficient condition to obtain the previous results (see Figure 4).
- Adding Lie derivatives of higher order to the observability matrix does not improve the results.

From the above results, it can be concluded that the proposed cooperative system, although still partially observable, considerably reduces the unobservable modes and states with respect to the 6DOF-monocular SLAM system. This contribution represents an advantage to improve the accuracy and consistency in the estimation process.

**Figure 4.** Requirements to obtain the results of the observability analysis for the proposed system.

4. EKF-Cooperative Monocular SLAM

In this section, the proposed monocular cooperative SLAM algorithm, based on an Extended Kalman Filter (EKF), is presented. Figure 5 shows the architecture of the proposed system.

**Figure 5.** Block diagram showing the architecture of the system: EKF-cooperative monocular SLAM.

4.1. EKF-SLAM

According to (1), the discrete system state to be estimated is defined by:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{n}_{k-1}) = \begin{bmatrix} \mathbf{x}_{c_k}^j \\ \mathbf{q}_{c_k}^j \\ \mathbf{v}_{c_k}^j \\ \boldsymbol{\omega}_{c_k}^j \\ \mathbf{x}_{a_k}^i \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{c_{k-1}}^j + (\mathbf{v}_{c_{k-1}}^j) \Delta t \\ \mathbf{q}_{c_{k-1}}^j \times \mathbf{q}((\boldsymbol{\omega}_{c_{k-1}}^j) \Delta t) \\ \mathbf{v}_{c_{k-1}}^j + \boldsymbol{\zeta}_{c_{k-1}}^j \\ \boldsymbol{\omega}_{c_{k-1}}^j + \boldsymbol{\eta}_{c_{k-1}}^j \\ \mathbf{x}_{a_{k-1}}^i \end{bmatrix} \quad (44)$$

$$\mathbf{n}_k = \begin{bmatrix} \boldsymbol{\zeta}_{c_k}^j \\ \boldsymbol{\eta}_{c_k}^j \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{c_k}^j \Delta t \\ \boldsymbol{\alpha}_{c_k}^j \Delta t \end{bmatrix} \quad (45)$$

with system measurements defined according to (4) and (21), as:

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k) = \begin{bmatrix} {}^i\mathbf{h}_{c_k}^j + {}^i\mathbf{r}_{c_k}^j \\ {}^j\mathbf{h}_{r_k}^n + {}^j\mathbf{r}_{e_k}^n \end{bmatrix} \quad (46)$$

$$\mathbf{r}_k = \begin{bmatrix} {}^i\mathbf{r}_{c_k}^j \\ {}^j\mathbf{r}_{e_k}^n \end{bmatrix} \quad (47)$$

Let $\mathbf{a}_{c_k}^j$ and $\boldsymbol{\alpha}_{c_k}^j$ represent unknown linear and angular accelerations that are assumed to have a Gaussian distribution with zero mean. Let $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ and $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ be the noise vectors that affect the state and the measurement, which are assumed to be mutually uncorrelated. Let Δt be the differential of time and k the sample step. Note that in this work, for simplicity, a Gaussian random process is used for propagating the velocity of the vehicle. However, a feasible alternative could be to use the dynamical model of the aircraft instead. However, this approach commonly requires having considerable knowledge of the specific physics of each aerial vehicle where the proposed method would have to be applied.

The prediction stage of the EKF is defined by:

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{0}) \quad (48)$$

$$\mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k \mathbf{Q}_{k-1} \mathbf{W}_k^T \quad (49)$$

The correction stage of the EKF is defined by:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-, \mathbf{0})) \quad (50)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k) \mathbf{P}_k^- \quad (51)$$

with:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^T)^{-1} \quad (52)$$

and:

$$\begin{aligned} \mathbf{A}_k &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k-1}, \mathbf{0}) & \mathbf{C}_k &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_k^-, \mathbf{0}) \\ \mathbf{W}_k &= \frac{\partial \mathbf{f}}{\partial \mathbf{n}}(\hat{\mathbf{x}}_{k-1}, \mathbf{0}) & \mathbf{V}_k &= \frac{\partial \mathbf{h}}{\partial \mathbf{r}}(\hat{\mathbf{x}}_k^-, \mathbf{0}) \end{aligned} \quad (53)$$

\mathbf{P} is the covariance matrix of the system state, and \mathbf{K} is the Kalman gain.

4.2. Initialization of Map Features

Taking advantage of the multi-UAV cooperative system, the initialization process of new map features is carried out through a pseudo-stereo system composed of two different UAV cameras that observe common landmarks. This fact allows initializing the landmarks with less uncertainty since 3D information of the position of the landmarks is gathered from the beginning. The three-dimensional data obtained by the pseudo-stereo system can improve the information obtained by other sensors. For example, the traditional fixed stereo system has a limited operating range due to the fixed baseline between the cameras.

The process of initialization is carried out when a new landmark is observed by two cameras, and if this condition is fulfilled, then the landmark can be initialized by means of a linear triangulation. In this case, the measurement is computed using the *a posteriori* values obtained in the correction stage of the EKF.

According to (4) and (6), the following expression can be defined in homogeneous coordinates:

$${}^i\gamma_c^j \begin{bmatrix} {}^iu_c^j \\ {}^iv_c^j \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_c^j & \mathbf{0}_{3 \times 1} \end{bmatrix} \hat{\mathbf{E}}_c^j \begin{bmatrix} \mathbf{x}_a^i \\ 1 \end{bmatrix} \quad (54)$$

where:

$$\hat{\mathbf{E}}_c^j = \begin{bmatrix} {}^w\hat{\mathbf{R}}_c^j & \hat{\mathbf{x}}_c^j \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (55)$$

Using (54) and considering the projection onto two any UAV cameras, a linear system can be formed in order to estimate \mathbf{x}_a^i :

$$\mathbf{D}^i \mathbf{x}_a^i = \mathbf{b}^i \quad \mathbf{x}_a^i = \mathbf{D}^{i^\dagger} \mathbf{b}^i \quad (56)$$

where \mathbf{D}^{i^\dagger} is the Moore–Penrose right pseudo-inverse matrix of \mathbf{D}^i , and:

$$\mathbf{D}^i = \begin{bmatrix} k_{31}^j {}^iu_c^j - k_{11}^j & k_{32}^j {}^iu_c^j - k_{12}^j & k_{33}^j {}^iu_c^j - k_{13}^j \\ k_{31}^j {}^iv_c^j - k_{21}^j & k_{32}^j {}^iv_c^j - k_{22}^j & k_{33}^j {}^iv_c^j - k_{23}^j \end{bmatrix} \quad (57)$$

$$\mathbf{b}^i = \begin{bmatrix} k_{14}^j - k_{34}^j {}^iu_c^j \\ k_{24}^j - k_{34}^j {}^iv_c^j \end{bmatrix}$$

with:

$$\begin{bmatrix} \mathbf{T}_c^j & \mathbf{0}_{3 \times 1} \end{bmatrix} \hat{\mathbf{E}}_c^j = \begin{bmatrix} k_{11}^j & k_{12}^j & k_{13}^j & k_{14}^j \\ k_{21}^j & k_{22}^j & k_{23}^j & k_{24}^j \\ k_{31}^j & k_{32}^j & k_{33}^j & k_{34}^j \end{bmatrix} \quad (58)$$

When a new landmark is initialized, the system state \mathbf{x} is augmented by: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_c^j & \mathbf{q}_c^j & \mathbf{v}_c^j & \boldsymbol{\omega}_c^j & \mathbf{x}_a^i & \mathbf{x}_a^{new} \end{bmatrix}^T$.

The new covariance matrix \mathbf{P}_{new} is computed by:

$$\mathbf{P}_{new} = \Delta \mathbf{J} \begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & {}^i\mathbf{R}^j \end{bmatrix} = \Delta \mathbf{J}^T \quad (59)$$

where $\Delta \mathbf{J}$ is the Jacobian for the initialization function and ${}^i\mathbf{R}^j$ is the measurement noise covariance matrix for $({}^iu_c^j, {}^iv_c^j)$.

Map Management

The real-time feasibility of EKF-based visual SLAM systems has been proven since early works like [39]. Nevertheless, it is well known that due to the nature of the Kalman filter, in SLAM, the system state can always reach a size that will make it impossible to maintain a real-time performance for a given hardware. In this sense, 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 relatively near the origin of the navigation frame. Therefore, old features can be removed from the system state and covariance matrix, to prevent the system state from reaching a size that affects the computational performance.

On the other hand, although large-scale SLAM and loop-closing are not considered in this work, it is important to note that 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 [12].

5. Computer Simulations Results

In this section, computer simulation results are presented. The computer simulations were performed in order to validate the performance of the proposed method. A MATLAB® implementation was used for this purpose.

With the intention of making an exhaustive analysis of the performance of the proposed system, a comparison is carried out with respect to the other three typical single-robot SLAM configurations. The comparison allows one to note the advantages and drawbacks of multi-UAV systems compared with single robot systems.

For the computer simulations setup, two quadcopters equipped with an onboard monocular camera are simulated, while moving maintaining a stable flight formation. In this case, a Quadcopter (Quad 2) navigates over the other (Quad 1) at an arbitrary relative distance. In the computer simulations, it is considered that Quad 1 remains all the time inside the visual field of Quad 2. It is also assumed that there exist some landmarks observed in common by the cameras of both quadcopters.

The characteristics of the three SLAM configurations used for the comparison are described below:

1. The first configuration to be compared is monocular SLAM. In this case, the estimates are obtained from the monocular camera carried by Quad 1. The Monocular SLAM approach used to implement this configuration is based on the method proposed in [40]. In this method, the map features are parametrized with the inverse depth parametrization. Both the initialization and update process are performed by means of the monocular measurements. The metric scale of the estimates cannot be retrieved when only monocular vision is used. For this reason, for this configuration, it is assumed that the position of the landmarks seen in the first frame (at the beginning of the flight trajectory) is perfectly known.
2. The second configuration to be compared is stereo SLAM. In this case, the estimates are obtained from a stereo system, with a baseline of 15 cm, carried by Quad 1. In this method, the map features are parametrized with the Euclidean parametrization. The feature initialization process is carried out directly by means of the 3D information provided by the stereo system. The state update is also performed using the stereo measurements.
3. The third configuration to be compared is a hybrid system stereo-monocular SLAM. In this case, the estimates are obtained from a stereo system, with a baseline of 15 cm, carried by Quad 1. In this method, the map features are parametrized with the Euclidean parametrization. The features initialization process is carried out directly by means of the 3D information provided by the stereo system. Unlike the second configuration, in this case, the state update is performed through monocular measurements obtained from one of the cameras of the stereo system.

In computer simulations, it is assumed that the initial condition of the quadcopter states is known with certainty. In order to emulate uncertainty, Gaussian noise with $\sigma_c = 3$ pixels is added to the measurements given by the cameras. The measurements from the cameras are taken with a frequency of 10 Hz. The intrinsic parameters used for the cameras are $f_c^j/d_u^j = f_c^j/d_v^j = 200.1$ and $c_u^j = c_v^j = 500$. The environment is composed of 3D points, randomly distributed over the ground. Furthermore, it is assumed that the camera can detect and track visual features without error, avoiding the data association problem. Furthermore, the problem of the influence of the estimates on the control system was not considered. In other words, an almost perfect control over the vehicle is assumed. The trajectory followed by the vehicles begins near the ground, then it moves away from the initial position taking a higher altitude as the trajectory progresses.

The average NEES (Normalized Estimation Error Squared [41]) over n_3 Monte Carlo runs was used in order to evaluate the consistency of each method, as proposed in [42]. The NEES is estimated as follows:

$$\epsilon_k = \left[\mathbf{x}_k - \hat{\mathbf{x}}_k \right]^T \mathbf{P}_k^{-1} \left[\mathbf{x}_k - \hat{\mathbf{x}}_k \right] \quad (60)$$

The average NEES is computed from:

$$\bar{\epsilon}_k = \frac{1}{n_3} \sum_{r=1}^{n_3} \epsilon_k^r \quad (61)$$

Figure 6 shows the real and estimated trajectory obtained from the cooperative system. Figure 7 shows the real and estimated trajectory obtained with all the configurations. Note that in this case, only the trajectory of Quad 1 is presented. In this simulation, it can be seen that as the trajectory evolves, the error considerably increases for the single-robot configurations. On the other hand, for the proposed (cooperative) method, the error is better bounded.

Figure 8 shows the evolution over time of the real and estimated states (position and orientation) for Quad 1. In this case, the initial results are confirmed. The results of the estimated state of Quad 2 are not shown, but they are closely similar to those presented for Quad 1. Table 2 summarizes the Mean Squared Error (MSE) for the position in the three axes of Quad 1.

Figure 9 shows the average NEES over 50 Monte Carlo runs obtained for each method. The average NEES is calculated taking into account the twelve variables that define the complete state of the vehicle (position, orientation, linear velocity and angular velocity). It is very interesting to note how the consistency of the filter considerably degenerates in the three cases of the single-robot configurations. On the other hand, for the cooperative case, the consistency of the filter remains practically stable.

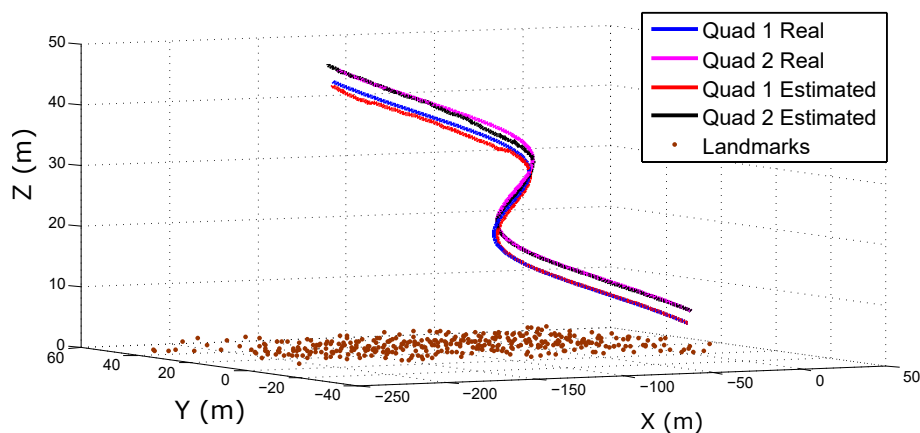


Figure 6. Estimated trajectories of the Quadcopters (Quad) obtained with the cooperative method.

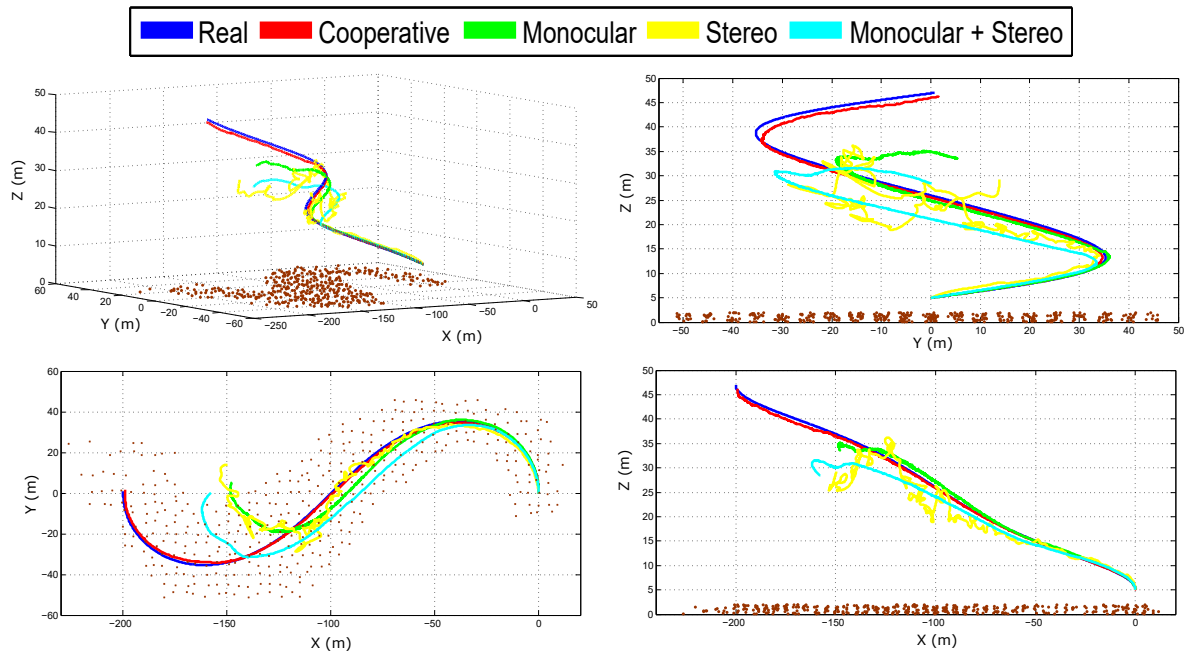


Figure 7. Estimated trajectory of Quad 1 obtained with all the configurations.

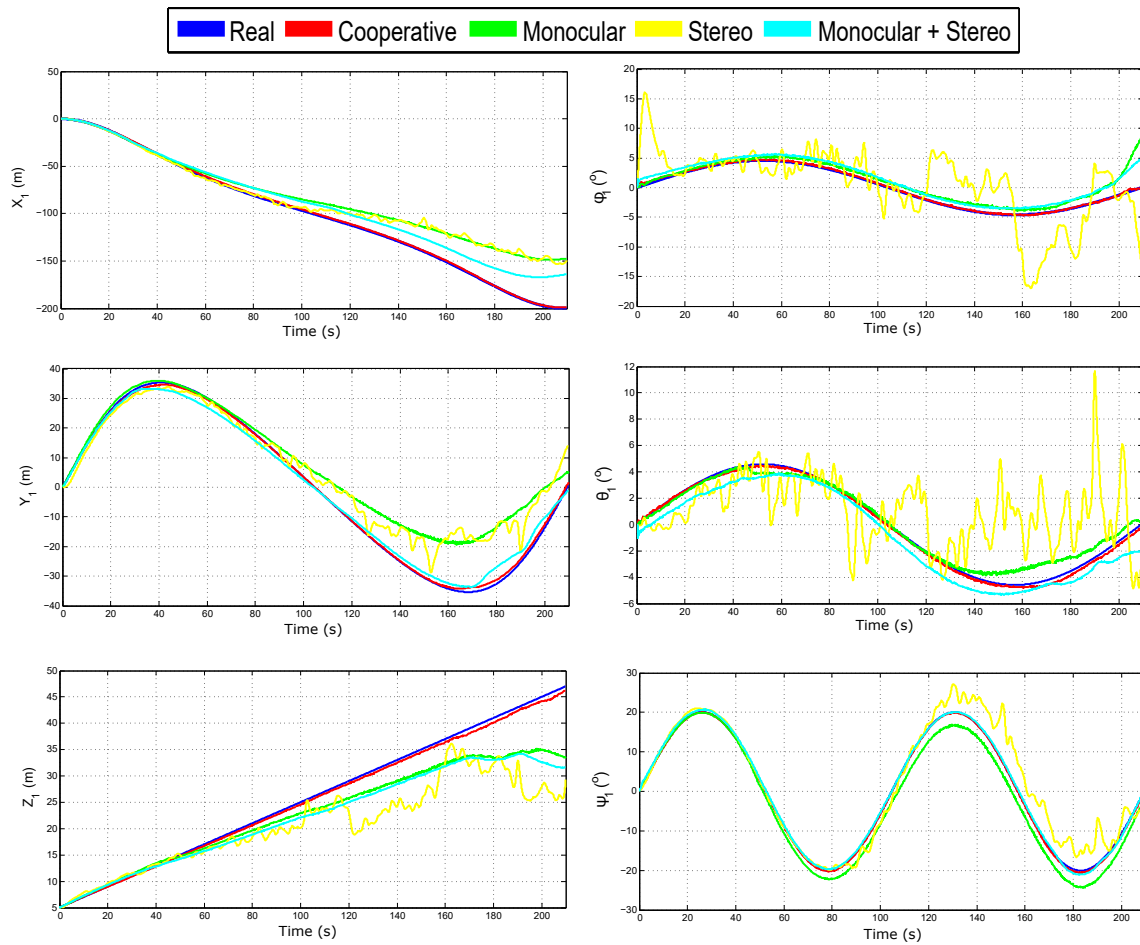


Figure 8. Estimated state of Quad 1.

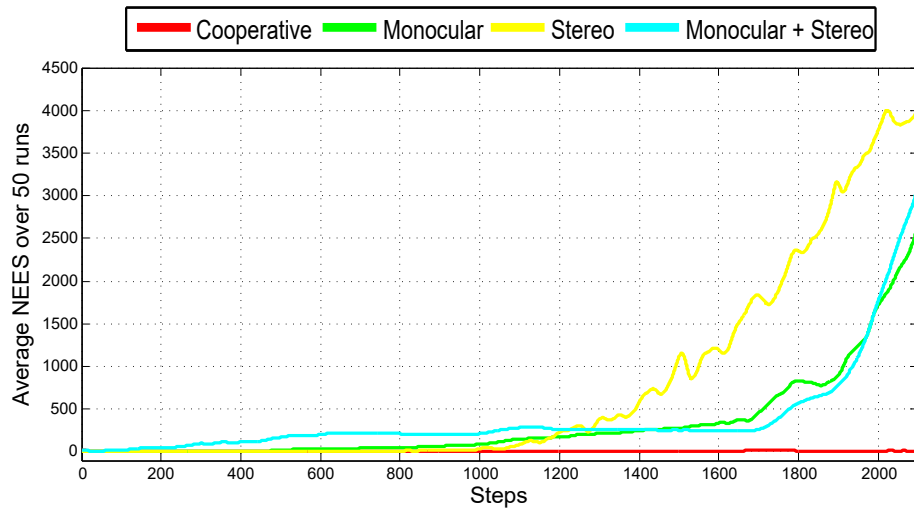


Figure 9. Average Normalized Estimation Error Squared (NEES) obtained with the four configurations.

Table 2. Mean squared error in the position estimation.

	MSE_x (m)	MSE_y (m)	MSE_z (m)
Cooperative	0.36	0.05	0.008
Monocular	30.31	6.28	2.36
Stereo	28.42	5.33	3.82
Monocular + Stereo	12.64	1.51	2.85

Figure 10 shows the relative distances (from Quad 1 to Quad 2) estimated with the method proposed in Section 2. It can be seen that these measurements are good enough to be used to update the filter (see Section 4). It is important to remark that the observability results presented in Section 3 depend on these measurements. The lower-right plot of Figure 10 shows an image frame captured from the monocular camera carried by Quad 2. In this case, the projection of the landmarks can be appreciated, as well as the projections of the four rotors of Quad 1 needed to compute the homography.

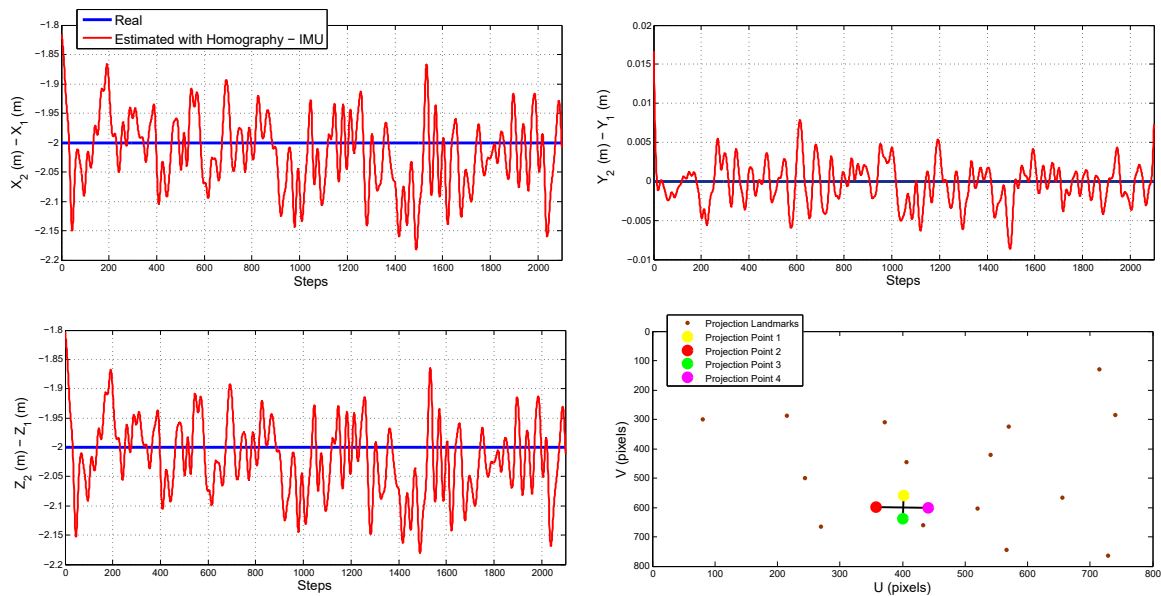


Figure 10. Estimation of the relative distances between the flying vehicles by means of homographies.

In order to compare the quality of the measurements obtained with the fixed stereo system and those obtained with the cooperative pseudo-stereo system, some computer simulations were performed. In this case, the error was computed for the estimated landmarks' positions, assuming that the position of Quad 1 was perfectly known along the flight trajectory. For the fixed stereo system, the camera-camera calibration is perfectly known. For the cooperative pseudo-stereo system, the camera-camera calibration is obtained from the homography, and therefore, it presents a certain level of error.

Figure 11 shows the absolute value of the mean error obtained for both methods. In this experiment, the same measurements were performed for both systems. In the lower-right plot, the number of measurements per frame is shown. In the case of the fixed stereo system, the accuracy of its measurements is affected by the small baseline between cameras. This is especially notorious when the vehicle moves far away from the landmarks (the altitude is increased). In the case of the cooperative pseudo-stereo system, the error in estimation is much better bounded, although the calibration of the system is not perfectly known. A suitable explanation has to do with the possibility of having an arbitrarily greater baseline between the cameras.

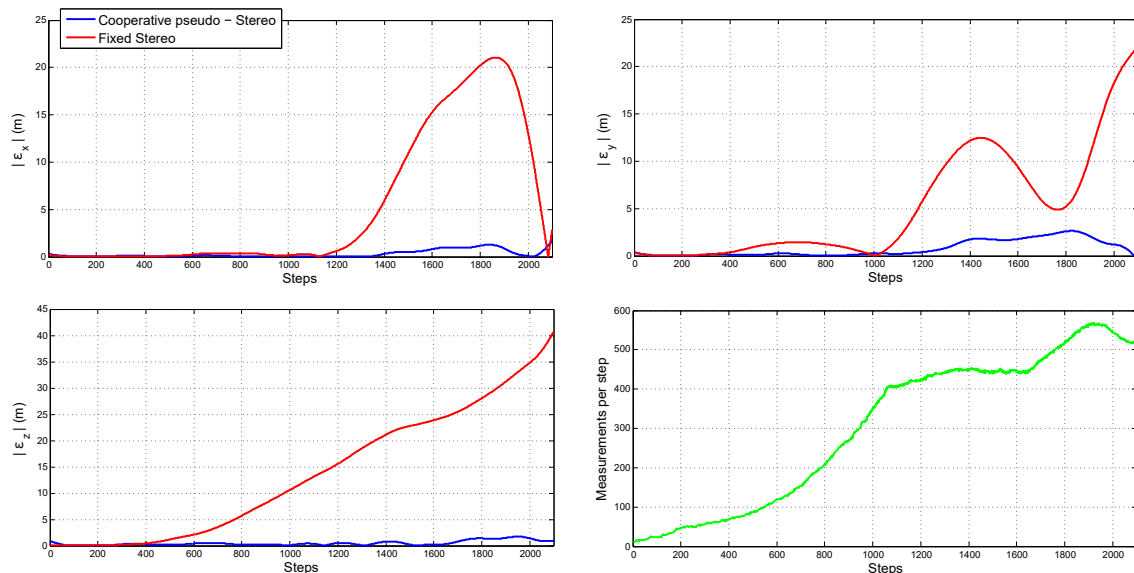


Figure 11. Comparison of the quality of the measurements obtained from a fixed stereo system and those obtained with the cooperative pseudo-stereo system.

Figure 12 illustrates the above fact. In this case, the statistical results obtained from simulating the measurement of a single landmark with (i) the cooperative pseudo-stereo system and (ii) a monocular method are presented. In the simulation, the UAV-Camera 1 system is located at $[x, y, z] = [3, 3, 25]$ at instant k . The UAV-Camera 2 system is located at $[x, y, z] = [4, 3, 30]$ at instant k . Thus, the baseline in the cooperative system is equal to 5.09 meters. A landmark is located at $[x, y, z] = [3.5, 3, 15]$. In order to model the inaccuracies associated with the cooperative pseudo-stereo approach, the estimated location of the UAV-Camera 2 system was modeled by adding a Gaussian noise with $\sigma = 50$ cm to its actual location. In order to emulate the monocular measurements, it is assumed that the UAV-Camera 1 system was moved (at some instant $k + t$) to $[x, y, z] = [3.3, 3, 25.1]$ to generate a parallax with respect to the landmark. Thus, the baseline in the monocular system is equal to 0.31 meters. The drift associated with the estimated displacement of the UAV-Camera 1 system is modeled by adding Gaussian noise with standard deviation $\sigma = 5$ cm to the actual location at instant $k + t$. In all cases, the angular measurements provided by the cameras are corrupted by Gaussian noise with $\sigma = 3$ degrees. Using the above conditions, a Monte Carlo simulation with 1000 executions has been used to estimate the landmark position with linear triangulation. In Figure 12, ellipsoids are used to illustrate

the uncertainties in the estimated positions. According to the simulation results, it is better to have a larger baseline between two cameras with greater position uncertainty (like the cooperative system) than a small baseline with small uncertainty (like monocular measurements with low parallax).

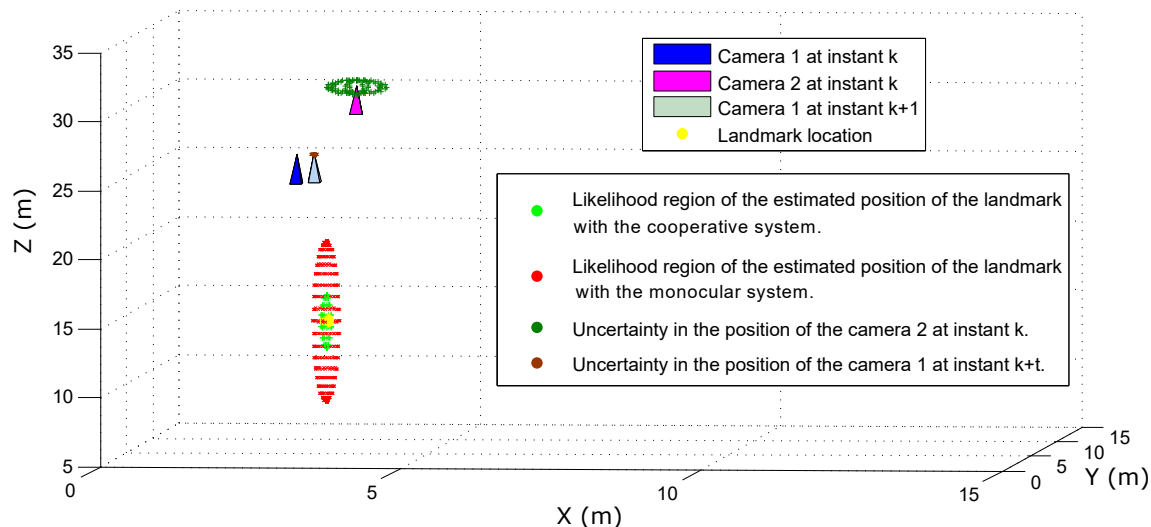


Figure 12. Measurement of a single landmark using: (i) cooperative pseudo-stereo system; and (ii) the delayed monocular initialization method.

In practical applications, there are several related factors that can severely also affect the performance of a system. For instance, in visual SLAM, the data association problem is critical for these approaches to be reliable. Although currently, there are several methods available for rejecting outliers, it is difficult to completely eliminate this problem. In addition, in cooperative visual systems, the data association problem can be extended from the single-image case to the multiple-image case. Furthermore, a problem that can arise in multi-robot systems, contrary to the mono-robot systems, is related to the communication issues between robots. This problem can cause loss of information or even make the interchange of information impossible during certain periods.

In order to take into account the above practical considerations, a set of additional computer simulations is presented. In this case, based on the same simulation setup used previously, the following aspects are now added: (i) outliers for the visual data association in each camera; (ii) outliers for the cooperative visual data association; (iii) outages of communication between robots; (iv) failures in the homography-based technique used for estimating the relative distance between robots.

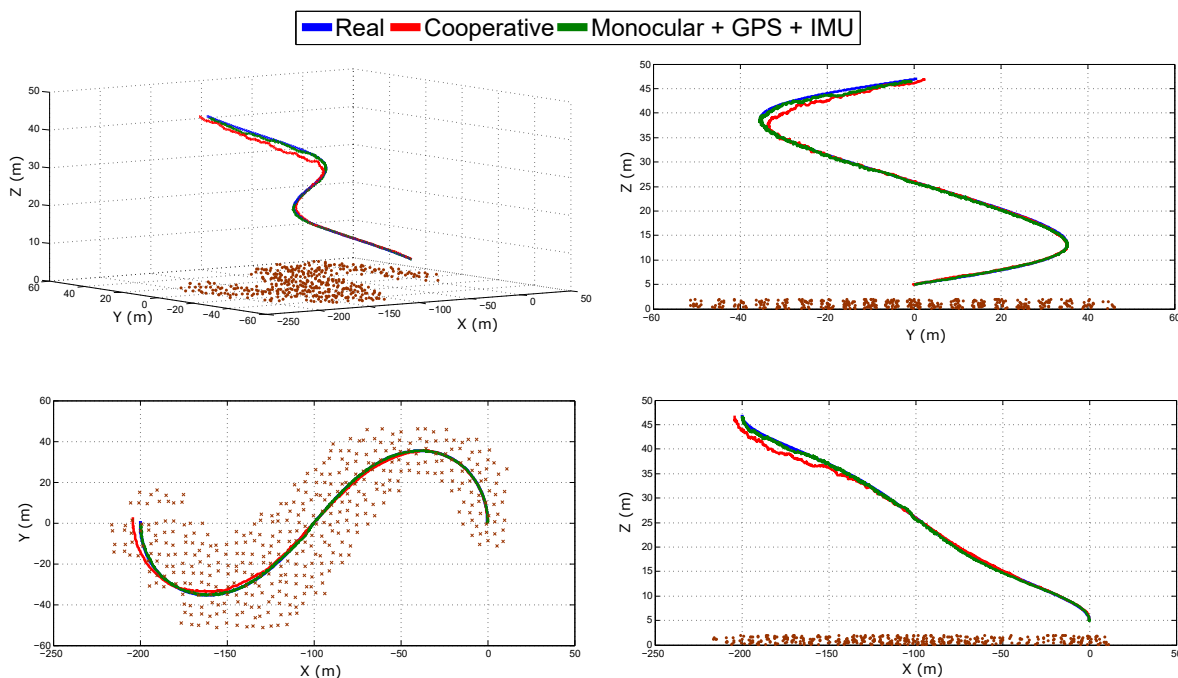
In order to emulate the failures of the visual data association process, 5% of the total number of visual correspondences are forced to be outliers in a random manner. In this case, each outlier is modeled by means of a big measurement error of $\sqrt{e_u^2 + e_v^2} = 56 \pm 14\sigma$ pixels. With the objective of having a good insight into the performance of the proposed method, under the above conditions, a comparison with a reliable general method is carried out. In this case, the method chosen is a monocular SLAM system aided by measurements of the position given by a GPS and attitude measurements obtained from an IMU (monocular SLAM + GPS + IMU).

Table 3 shows the number of failures introduced into the simulation: (i) the number of outliers introduced in the visual tracking process of Quad 1; (ii) the number of outliers introduced in the visual tracking process of Quad 2; (iii) the number of outliers introduced in the visual data association process used for cooperatively measuring the landmarks by means of Quad 1 and Quad 2; (iv) the number of outages in communication between robots, which result in filter update not being carried out with the information given by Quad 2; and (v) the number of failures in the homography-based technique, which result in the filter update not being carried out with the information given by the relative distance between the Quads.

Table 3. Number of failures introduced into the simulation.

	No. of Quad 1 Visual Outliers	No. of Quad 2 Visual Outliers	No. of Cooperative Visual Outliers	No. of Communication Outages	No. of Homography Failures
Cooperative	9002	8400	1706	210	420
Monocular + GPS + IMU	9535	-	-	-	-

Figure 13 shows the real and estimated trajectory obtained with the two configurations: (i) cooperative SLAM; and (ii) monocular SLAM + GPS + IMU. Figure 14 shows the evolution over time of the real and estimated states (position and orientation) of Quad 1 obtained with both configurations. Note that in this case, only the trajectory of Quad 1 is presented for illustration purposes, but estimates of Quad 2 are closely similar to those presented for Quad 1. Table 4 summarizes the mean squared error for the position in the three axes of Quad 1 obtained with both configurations. In this simulation, both configurations have a good performance, in the case of monocular SLAM + GPS + IMU, this result was expected, since this system has enough sources of information to determine all the states. The cooperative system shows a good performance despite all the failures introduced into the system. The above study provides a good insight about the robustness of the proposed (cooperative) system.

**Figure 13.** Estimated trajectory of Quad 1 obtained with the two configurations.**Table 4.** Mean squared error in the position estimation.

	MSE_x (m)	MSE_y (m)	MSE_z (m)
Cooperative	0.85	0.23	0.07
Monocular + GPS + IMU	0.01	0.04	0.05

Table 5 provides an insight into the performance of the proposed method for estimating the features map. In this case, the total (sum of all) of the mean squared errors for the estimated position of the landmarks is presented for both configurations. Furthermore, the total of the mean squared errors for the initial estimated position of the landmarks is presented. Note that the results are presented for each coordinate of the reference frame W . The results show that the proposed cooperative system has

a better performance than the monocular SLAM + GPS + IMU system, regarding the error obtained in the estimation of the position of the landmarks, although the latter has more sources of information provided by its sensors.

Table 5. Total mean squared error in: (i) the position estimation of the landmarks (MSE_{xm} , MSE_{ym} , MSE_{zm}); and (ii) the initial position estimation of the landmarks (MSE_{xmi} , MSE_{ymi} , MSE_{zmi}).

	MSE_{xm} (m)	MSE_{ym} (m)	MSE_{zm} (m)	MSE_{xmi} (m)	MSE_{ymi} (m)	MSE_{zmi} (m)
Cooperative	7.84	14.87	6.29	74.21	83.72	45.50
Monocular + GPS + IMU	21.80	30.31	13.27	394.05	427.91	193.99

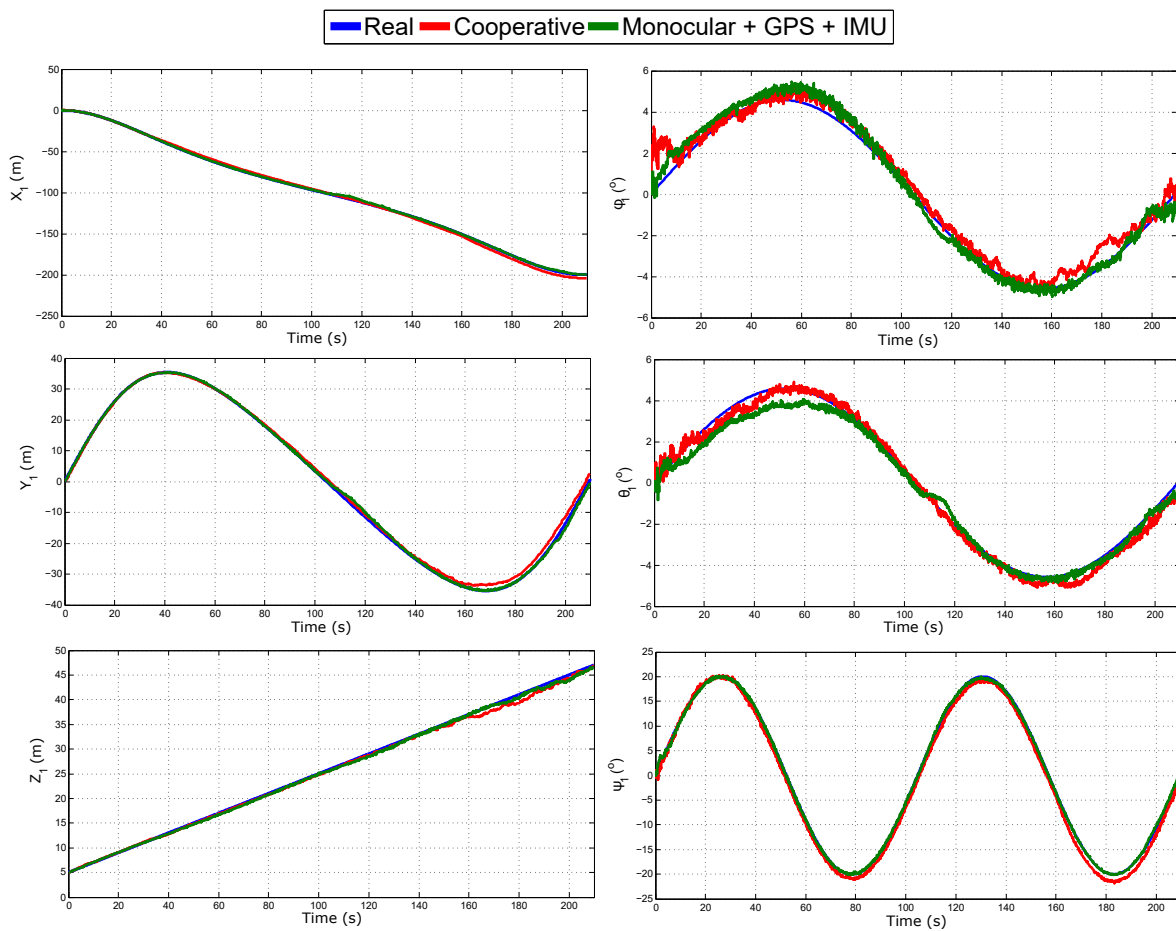


Figure 14. Estimated state of Quad 1.

6. Conclusions

In this work, a vision-based cooperative SLAM system with application to unmanned aerial vehicles has been presented. The general idea is to take advantage of a cooperative UAV scheme in order to improve the accuracy and consistency of the state estimation process of the whole system. To achieve this purpose, it was proposed to add some relative distances between the robots as system measurements for updating the EKF. These measures provide metric information to the system, unlike other configurations where the scale of the system is a problem. Through a non-linear observability analysis, it is verified that the observability of the cooperative system improves the observability obtained for a single-robot configuration. In this case, the observability of the system is improved by adding the measures of relative distances. Sufficient conditions required for obtaining the observability results were established. In order to infer the 3D knowledge of the position of the landmarks for

initializing the map features with less uncertainty, in the proposed method, pseudo-stereo systems are formed from pairs of aerial robots.

An extensive set of computer simulations was performed in order to validate the proposed method. In the computer simulations, the proposed system was compared against four single-robot configurations of visual SLAM. Based on the results of the simulations, it can be observed how the proposed method (cooperative) improves the estimation of the state with respect to the other configurations. The difference in the performance of the systems is especially notorious when the distance from the cameras to the landmarks increases. Furthermore, it was shown that the consistency of the filter is improved with the proposed method. Computer simulations also show that the accuracy of the measurements obtained from the pseudo-stereo system is better than the measurements obtained from a stereo system with a fixed small baseline.

In computer simulations, an effort has been made in order to emulate several aspects regarding applicability in real scenarios of the proposed approach. For instance, the data association problem has been considered by emulating outliers (mismatches) during the tracking of visual features on each monocular camera, as well as on the pseudo-stereo matching. Furthermore, issues for the multi-robot communication were considered, as well as failures on the homography technique used to provide measurements of the relative distance between robots. However, although computer simulations are useful for evaluating the full statistical consistency of the methods, they can still neglect important practical issues that appear when the methods are used in real scenarios. In this sense, it is important to note that future work should be focused on developing experiments with real data in order to validate the applicability of the proposed approach fully. Therefore, it should be interesting to investigate more practical aspects, like the homography-based technique or the pseudo-stereo matching process.

Acknowledgments: This research has been funded by AEROARMS EU Project H2020-ICT-2014-1-644271.

Author Contributions: R.M. and A.G. developed the theoretical framework and algorithms presented in this research; R.M. and J-C.T. conceived and designed the experiments; E.G. and J-C.T. performed the experiments and supervised data management; R.M. and A.G. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, Z.; Douillard, B.; Morton, P.; Vlaskine, V. Towards Collaborative Multi-MAV-UGV Teams for Target Tracking. In Proceedings of the 2012 Robotics: Science and Systems Workshop on Integration of Perception with Control and Navigation for Resource-Limited, Highly Dynamic, Autonomous Systems, Sydney, Australia, 9–10 July 2012.
2. Michael, N.; Shen, S.; Mohta, K.; Mulgaonkar, Y.; Kumar, V.; Nagatani, K.; Okada, Y.; Kiribayashi, S.; Otake, K.; Yoshida, K.; et al. Collaborative mapping of an earthquake-damaged building via ground and aerial robots. *J. Field Robot.* **2012**, *29*, 832–841.
3. Bachrach, S.P.; He, R.; Roy, N. RANGE-Robust autonomous navigation in GPS-denied environments. *J. Field Robot.* **2011**, *28*, 644–666.
4. Aut, F.; De la Cruz, C.; Carelli, R.; Bastos, T. Navegación Autónoma asistida basada en SLAM para una silla de ruedas robotizada en entornos restringidos. *Rev. Iberoam. Autom. Inform. Ind.* **2011**, *8*, 81–92.
5. Andert, F.; Lorenz, S.; Mejias, L.; Bratanov, D. Radar-Aided Optical Navigation for Long and Large-Scale Flights over Unknown and Non-Flat Terrain. In Proceedings of the 2016 International Conference on Unmanned Aircraft Systems (ICUAS), Arlington, VA, USA, 7–10 June 2016.
6. Vázquez-Martín, R.; Núñez, P.; Bandera, A.; Sandoval, F. Curvature Based Environment Description for Robot Navigation using Laser Range Sensors. *Sensors* **2009**, *9*, 5894–5918.
7. Zhou, H.; Zou, D.; Pei, L.; Ying, R.; Liu, P.; Yu, W. Structslam: Visual slam with building structure lines. *IEEE Trans. Veh. Technol.* **2015**, *64*, 1364–1375.
8. Pi, S.; He, B.; Zhang, S.; Nian, R.; Shen, Y.; Yan, T. Stereo visual slam system in underwater environment. In Proceedings of the OCEANS 2014, Taipei, Taiwan, 7–10 April 2014; pp. 1–5.

9. Schmidt, A. The EKF-Based Visual SLAM System with Relative Map Orientation Measurements. In Proceedings of the Computer Vision and Graphics: International Conference, ICCVG 2014, Warsaw, Poland, 15–17 September 2014; Springer International Publishing: Cham, Switzerland, 2014; pp. 570–577.
10. Meilland, M.; Comport, A.I. On unifying key-frame and voxel-based dense visual slam at large scales. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 3677–3683.
11. Lu, Y.; Song, D. Visual navigation using heterogeneous landmarks and unsupervised geometric constraints. *IEEE Trans. Robot.* **2015**, *31*, 736–749.
12. Strasdat, H.; Montiel, J.; Davison, A. Realtime monocular SLAM: Why filter? In Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, USA, 3–7 May 2010; pp. 2657–2664.
13. Artieda, J.; Sebastian, J.; Campoy, P.; Correa, J.; Mondragon, I.; Martinez, C.; Olivares, M. Visual 3-D SLAM from UAVs. *J. Intell. Robot. Syst.* **2009**, *55*, 299–321.
14. Weiss, S.; Scaramuzza, D.; Siegwart, R. Monocular SLAM based navigation for autonomous micro helicopters GPS-denied environments. *J. Field Robot.* **2011**, *28*, 854–874.
15. 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.* **2014**, *10*, 387–396.
16. 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 Robot.* **2013**, *30*, 415–438.
17. Andert, F.; Mejias, L. Improving monocular SLAM with altimeter hints for fixed-wing aircraft navigation and emergency landing. In Proceedings of the 2015 International Conference on Unmanned Aircraft Systems (ICUAS), Denver, CO, USA, 9–12 June 2015.
18. Urzua, S.; Munguía, R.; Grau, A. Vision-based SLAM system for MAVs in GPS-denied environments. *Int. J. Micro Air Veh.* **2017**, *9*, 283–296.
19. Perez-Grau, F.J.; Ragel, R.; Caballero, F.; Viguria, A.; Ollero, A. An architecture for robust UAV navigation in GPS-denied areas. *J. Field Robot.* **2018**, *35*, 121–145.
20. Pappas, H.; Tanner, G.; Kumar, V. Leader-to-formation stability. *IEEE Trans. Robot. Autom.* **2004**, *20*, 443–455.
21. Zhu, Z.; Roumeliotis, S.; Hesch, J.; Park, H.; Venable, D. Architecture for Asymmetric Collaborative Navigation. In Proceedings of the IEEE Position Location and Navigation Symposium (PLANS), Myrtle Beach, SC, USA, 23–26 April 2012.
22. Mourikis, A.I.; Roumeliotis, S.I. Performance Bounds for Cooperative Simultaneous Localisation and Mapping (C-SLAM). In Proceedings of the Robotics: Science and Systems Conference, Cambridge, MA, USA, 8–10 June 2005.
23. Fenwick, J.W.; Newman, P.M.; Leonard, J.J. Cooperative Concurrent Mapping and Localisation. In Proceedings of the IEEE International Conference on Robotics and Automation, Washington, DC, USA, 11–15 May 2002.
24. Vetrella, A.R.; Opromolla, R.; Fasano, G.; Accardo, D.; Grassi, M. Autonomous Flight in GPS-Challenging Environments Exploiting Multi-UAV Cooperation and Vision-aided Navigation. In Proceedings of the AIAA Information Systems-AIAA Infotech @ Aerospace, AIAA SciTech Forum, (AIAA 2017-0879), Grapevine, TX, USA, 9–13 January 2017; doi:10.2514/6.2017-0879.
25. Vetrella, A.R.; Fasano, G.; Accardo, D. Cooperative Navigation in GPS-Challenging Environments Exploiting Position Broadcast and Vision-based Tracking. In Proceedings of the 2016 International Conference on Unmanned Aircraft Systems—ICUAS'16, Arlington, VA, USA, 7–10 June 2016.
26. Huang, G.P.; Mourikis, A.I.; Roumeliotis, S.I. *Generalized Analysis and Improvement of the Consistency of EKF-based SLAM*; Multiple Autonomous Robotic Systems; Technical Report Number 2008-0001; University of Minnesota: Minneapolis, MN, USA, 2008.
27. Huang, G.P.; Trawny, N.; Mourikis, A.I.; Roumeliotis, S.I. On the Consistency of Multi-robot Cooperative Localization. In Proceedings of the International Conference on Robotics Science and Systems, RSS 2009, Seattle, WA, USA, 28 June–1 July 2009.
28. Kim, J.; Sukkarieh, S. Improving the Real-Time Efficiency of Inertial SLAM and Understanding its Observability. In Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, 28 September–2 October 2004.

29. Bryson, M.; Sukkarieh, S. Active Airborne Localisation and Exploration in Unknown Environments using Inertial SLAM. In Proceedings of the 2006 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2006.
30. Melnyk, I.V.; Hesch, J.A.; Roumeliotis, S.I. Cooperative Vision-aided Inertial Navigation Using Overlapping Views. In Proceedings of the International Conference on Robotics and Automation. RiverCentre, Saint Paul, MN, USA, 14–18 May 2012.
31. Barbour, N.M. Inertial Navigation Sensors. In *NATO RTO Lecture Series, RTO-EN-SET-116, Low-Cost Navigation Sensors and Integration Technology*; NATO: Washington, DC, USA, 2011.
32. Batzdorfer, S.; Bestmann, U.; Becker, M.; Schwithal, A.; Schattenbergm J.; Lang, T.; Andert, F.; Dittrich, J. Using combined IMU/Stereo Vision/cooperative GNSS System for Positioning of UxV Swarms within Catastrophic Urban Scenarios. In Proceeding of the Institute of Navigation Pacific PNT 2013, Honolulu, Hawaii, 23–25 April 2013.
33. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, MA, USA, 2003.
34. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334.
35. Euston, M.; Coote, P.; Mahony, R.; Kim, J.; Hamel, T. A complementary filter for attitude estimation of a fixed-wing UAV. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2008, Nice, France, 22–26 September 2008; pp.340–345.
36. Munguia, R.; Grau, A. A Practical Method for Implementing an Attitude and Heading Reference System. *Int. J. Adv. Robot. Syst.* **2014**, *11*, 62.
37. Hermann, R.; Krener, A. Nonlinear controllability and observability. *IEEE Trans. Autom. Control* **1977**, *22*, 728–740.
38. Slotine, J.E.; Li, W. *Applied Nonlinear Control*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1991.
39. Davison, A. Real-time simultaneous localisation and mapping with a single camera. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; Volume 2, pp. 1403–1410.
40. Montiel, J.M.M.; Civera, J.; Davison, A. Unified inverse depth parametrization for monocular SLAM. In Proceedings of the Robotics: Science and Systems Conference, Philadelphia, PA, USA, 19 August 2006.
41. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation*; John Wiley and Sons: Hoboken, NJ, USA, 2001.
42. Bailey, T.; Nieto, J.; Guivant, J.; Stevens, M.; Nebot, E. Consistency of the ekf-slam algorithm. In Proceedings of IEEE International Conference on Intelligent Robots and Systems, Beijing, China, 4–6 December 2006.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).