sensors

MDPI

Article

Visual-Based SLAM Configurations for Cooperative Multi-UAV Systems with a Lead Agent: An Observability-Based Approach

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
[2] Department of Automatic Control, Technical University of Catalonia UPC, 08034 Barcelona, Spain; edmundo.guerra@upc.edu (E.G.); antoni.grau@upc.edu (A.G.)
* Correspondence: rodrigo.munguia@academicos.udg.mx or rodrigo.munguia@upc.edu
† These authors contributed equally to this work.

check for updates

Abstract: In this work, the problem of the cooperative visual-based SLAM for the class of multi-UA systems that integrates a lead agent has been addressed. In these kinds of systems, a team of aerial robots flying in formation must follow a dynamic lead agent, which can be another aerial robot, vehicle or even a human. A fundamental problem that must be addressed for these kinds of systems has to do with the estimation of the states of the aerial robots as well as the state of the lead agent. In this work, the use of a cooperative visual-based SLAM approach is studied in order to solve the above problem. In this case, three different system configurations are proposed and investigated by means of an intensive nonlinear observability analysis. In addition, a high-level control scheme is proposed that allows to control the formation of the UAVs with respect to the lead agent. In this work, several theoretical results are obtained, together with an extensive set of computer simulations which are presented in order to numerically validate the proposal and to show that it can perform well under different circumstances (e.g., GPS-challenging environments). That is, the proposed method is able to operate robustly under many conditions providing a good position estimation of the aerial vehicles and the lead agent as well.

Keywords: cooperative SLAM; aerial robots; state estimation; observability

1. Introduction

In recent years, advances in technology and miniaturization of flight control systems have contributed to the growth of interest for the Unmanned Aerial Vehicles (UAVs). The UAVs are very versatile in their movements making them suitable for a large number of applications [1,2]. For many applications, it is necessary that two or more UAVs perform a flight formation with respect to a lead agent, which can be another UAV, a human, or another kind of mobile robot or vehicle (multi-UAV systems with a lead agent). In general, in these kinds of systems, the lead agent will be the only member of the formation that can move freely through its environment. To control the flight formation, it is necessary to have knowledge about the states (localization) of the aerial robots as well as the state of the lead agent.

In different scenarios, autonomous navigation capability is a primary mission requirement, which in outdoor environments it is typically fulfilled by integrating a Global Positioning System (GPS). However, several mission profiles require a multi-UAV system to operate in GPS-challenging or GPS-denied environments. In these kinds of environments (e.g., natural or urban canyons, or mixed indoor-outdoor scenarios) multipath or shadowing of the GPS satellite signal creates serious difficulties

for GPS receivers to yield a reliable position. Therefore, a sort of additional sensory information (e.g., visual information) should be integrated into the system in order to improve accuracy and robustness. In a practical scenario, it is clear that during a flight mission there can exist periods of time where the whole set of (or some) UAVs moves indistinctly through a GPS-challenging or a GPS-denied environment.

In the above context, more works have appeared focusing on the use of cameras in order to develop navigation systems based on visual information that can operate when GPS is partially available or denied. Moreover, even when the position sensor is available, the visual information can be fused in order to improve the accuracy and robustness of the system. Also, cameras are well suited for their use in embedded systems. In this work, the use of a cooperative visual-based SLAM (Simultaneous Localization and Mapping) scheme is studied for addressing the problem of estimating at the same time the location of both the aerial robots and the lead agent. The general idea is that the aerial vehicles, flying in formation with respect to the lead agent, carry out the task of self-localization as well as locating the lead agent. The above scheme allows flexibility and freedom for the lead agent to carry out its own mission without the restriction of fulfilling the task of localizing itself (See Figure 1).
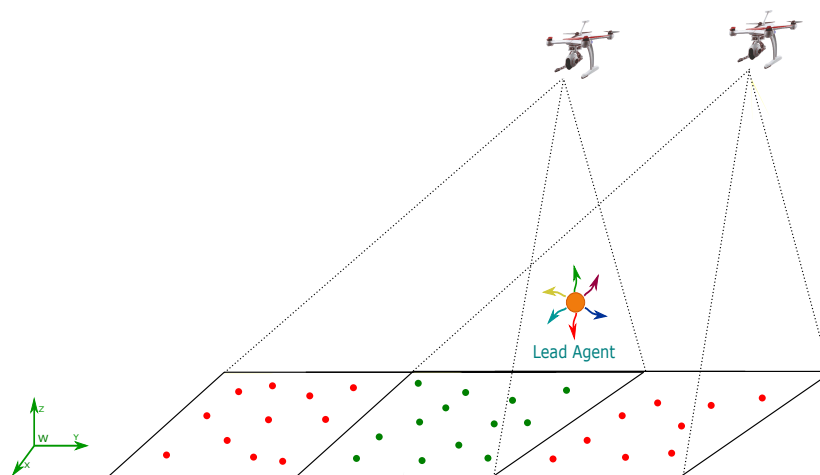

Figure 1. Multi-UAV SLAM system with a dynamic lead agent.

1.1. Related Work

The visual-based SLAM methods make use of visual features as landmarks and can be used for addressing the problem of the state estimation of robots. In this scenario, the robot operates in a priori unknown environment using only angular measurements obtained from cameras, to simultaneously building a map of its surroundings which it is used at the same time to track its position. Currently, there are two main approaches for implementing vision-based SLAM systems: (i) Filtering-based methods ([3–5]) and (ii) the optimization-based methods ([6,7]).

In the literature, there are several works that present SLAM methods applied to UAVs, such as in [8], where a Visual SLAM method is developed in outdoors, in partially structured environments. In [9], the estimation of the UAV state obtained through a Visual SLAM method is used as feedback to the control laws that stabilize the aircraft. In [10], a Visual SLAM method is developed for GPS-challenging environments. In [11], the estimation of the UAV state obtained through a Visual SLAM method is used to carry out emergency landings. In [12], a sensor fusion method for Visual SLAM through the integration of a monocular camera and a 1D-laser range-finder is presented. In [13], an approach based on graph-SLAM and loop closure detection for online mapping of unknown outdoor environments, using a small UAV, is proposed. In [14], a Inertial-Visual SLAM is proposed. In this method, inertial measurements are integrated into the system in order to recover the metric scale of the estimations. On the other hand, the performance of this method can be affected

by the dynamic error bias which is common to low-cost MEMS sensors. In [15], an EKF-based method is proposed in order to perform visual odometry with an unmanned aircraft. This method makes use of 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, another approach that addresses the problem of visual-based navigation in GPS-denied environments ([16–18]) has been appearing.

Multi-robot systems have also received great attention from the robotics research community. This attention is motivated by the inherent versatility that those systems are performing tasks that could be difficult to realize by a single robot. The use of several robots can have advantages like cost reductions, more robustness, better performance and efficiency ([19,20]). In the case of the SLAM problem, in [21,22] a centralized architecture is used where all vehicles send their sensor data to a unique Kalman filter. In [15,23,24] the idea of combining monocular SLAM with cooperative, multi-UAV information to improve navigation capabilities in GPS-challenging environments is presented. In [25], the idea of combining monocular SLAM with cooperative, human-robot information to improve navigation capabilities is presented. In [26], a visual-based cooperative localization method is proposed. According to the analysis presented in this work, the proposed system is completely observable. However, in this case, only distances and the relative orientations between the robots are estimated. This fact can be a clear drawback for applications where the global measurements of the system are required (e.g., absolute position).

As will be seen later, an important UAV technology that it is needed in order to implement the architectures proposed in this work is the cooperative visual-based detection and tracking of mobile targets. In this case, there are several works that deal with these kinds of problems. For instance, [27] presents a vision-based target detection and localization system, that makes use of different capabilities of aerial and ground unmanned vehicles as a cooperative team. In [28], a visual-based approach that allows a UAV to detect and track a cooperative flying vehicle autonomously using a monocular camera is presented. The algorithms are based on template matching and morphological filtering. In [29], a sophisticated vision-aided flocking system for UAVs is presented, which is able to operate in GPS-denied unknown environments, for missions of exploring and searching. In [30], it is presented a cooperative vision-based estimation and tracking system for objects of interest that are located on or near the ground. In the work in [31], an automatic cooperative tracking of targets system, using two quadrotors UAVs equipped with stereo vision systems, is presented. The system includes vision-based algorithms for searching and detecting of the target on the video stream. The research in [32] addresses the topic of vision-based target detection and tracking using a team of UAVs for maritime border surveillance. In this work, a method on how to integrate the perception into the control loop using two distinct teams of UAVs that are cooperatively tracking the same target is presented.

1.2. Objectives and Contributions

In this work, three configurations of cooperative SLAM for multi-UAV systems containing a lead agent are presented and analyzed. In all cases, each aerial robot is equipped with a monocular camera on board. The main differences between the system configurations have to do with the set of sensors used in each case (additional to the monocular cameras), as well as the circumstances they operate in a suitable manner.

The idea of presenting different system configurations, instead of a single "integrated" proposal, has to do the easiness of the mathematical analysis. However, also, this will allow a higher flexibility and modularity in the system implementation, since the features of each configuration are complementary and not exclusive. In other words, the authors' proposal should facilitate the implementation of an integrated system according to requirements (or availability) of hardware and the circumstances where the system will operate.

The configurations of the cooperative SLAM system proposed in this paper are based on the standard EKF-based SLAM methodology. In this context, it is extremely important to provide the proposed system with properties such as observability and robustness to errors in initial conditions since the properties mentioned above have a fundamental role in the convergence of the filter, as shown in [33,34]. Therefore, an extensive nonlinear observability test is carried out in order to analyze the system configurations. In this case, novel and important theoretical results and considerations are presented from the observability analysis. Also, in this work, the initialization process of new map features is carried out in a cooperative way. The 3D position of the new map features is estimated by means of a pseudo-stereo system formed by monocular cameras mounted on a pair of UAVs that observe common landmarks. This allows the landmark to be initialized with less uncertainty and lesser error. The pseudo-stereo system allows to initialize landmarks at distances farther than using stereo systems with a rigid baseline [35] or delayed monocular initialization methods. The above feature allows to the proposed cooperative system to have better performance in environments where the landmarks are far from the measurement system, contrary to SLAM approaches based on depth cameras, standard stereo systems or sonars.

1.3. Paper Outline

The document is organized in the following manner: Section 2 presents the general system specifications and mathematical models. Section 3 introduces the system configurations proposed in this work. Section 4 presents the nonlinear observability analysis. In Section 5 the proposed method is described. Section 7 shows the numerical simulations results, and finally, in Section 8 the conclusions of this work are presented.

2. System Specification

In this section, the different mathematical models used in this work are introduced. Those models are: the model used for representing the dynamics of a camera carried by a quadcopter, the model used for representing the dynamics of the lead agent, the representation of the landmarks as map features, the camera projection, and the GPS, altimeter, and range measurement models.

2.1. Dynamics of the System

In applications like aerial vehicles, the attitude and heading (roll, pitch, and yaw) estimation is well handled by available systems (e.g., [36,37]). In particular, in this work is assumed that the orientation of the camera always points toward the ground. In practice, the foregoing assumption can be easily addressed with the use of a servo-controlled camera gimbal. Considering the above aspects, the system state can be simplified by removing the variables related to attitude and heading which are provided by the attitude and heading reference system (AHRS). Therefore, the problem will be focused on the position estimation. Let consider the following continuous-time model describing the dynamics of the proposed system (see Figure 2):

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{x}}_h \\ \dot{\mathbf{v}}_h \\ \dot{\mathbf{x}}_c^j \\ \dot{\mathbf{v}}_c^j \\ \dot{\mathbf{x}}_a^i \end{bmatrix} = \begin{bmatrix} \mathbf{v}_h \\ \mathbf{0}_{3\times1} \\ \mathbf{v}_c^j \\ \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times1} \end{bmatrix} \tag{1}$$

where the state vector **x** is defined by:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_h & \mathbf{v}_h & \mathbf{x}_c^j & \mathbf{v}_c^j & \mathbf{x}_a^i \end{bmatrix}^T \tag{2}$$

with $i = 1, ..., n_1$ and $j = 1, ..., n_2$, where n_1 and n_2 are respectively the number of landmarks included into the map and the number of UAV-camera systems. In this work, the term *landmarks* will be used to refer to natural features of the environment which are detected and tracked from the images acquired by the cameras.
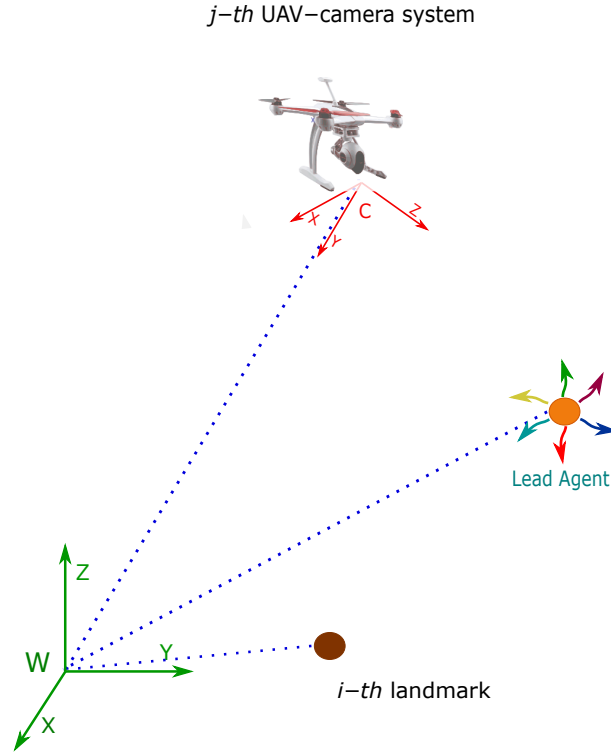

Figure 2. Coordinate reference systems.

Additionally, let $\mathbf{x_h} = \begin{bmatrix} x_h & y_h & z_h \end{bmatrix}^T$ represent the position (in meters) of the lead agent, with respect to the reference system W. Let $\mathbf{x_c}^j = \begin{bmatrix} x_c^j & y_c^j & z_c^j \end{bmatrix}^T$ represent the position (in meters) of the reference system C of the j-th camera, with respect to the reference system W. Let $\mathbf{v_h} = \begin{bmatrix} \dot{x}_h & \dot{y}_h & \dot{z}_h \end{bmatrix}^T$ represent the linear velocity (in $\frac{m}{s}$) of the lead agent. Let $\mathbf{v_c}^j = \begin{bmatrix} \dot{x}_c^j & \dot{y}_c^j & \dot{z}_c^j \end{bmatrix}^T$ represent the linear velocity (in $\frac{m}{s}$) of the j-th camera. Finally, let $\mathbf{x_a}^i = \begin{bmatrix} x_a^i & y_a^i & z_a^i \end{bmatrix}^T$ be the position of the i-th landmark (in meters) with respect to the reference system W, defined by its Euclidean parameterization. In (1), each UAV-camera, as well as the lead agent, is assumed to move freely in the three-dimensional space. Let note that a non-acceleration model is assumed for the UAV-camera systems and the lead agent. Also note that the landmarks are assumed to remain static.

2.2. Camera Measurement Model for the Projection of the Landmarks

Let consider the projection of a single landmark over the image plane of a camera. Using the pinhole model [38] (see Figure 3) the following expression can be defined:

$$
{}^i\mathbf{z_c}^j = {}^i\mathbf{h_c}^j = \begin{bmatrix} {}^iu_c^j \\ {}^iv_c^j \end{bmatrix} = \frac{1}{{}^iz_d^j} \begin{bmatrix} \frac{f_c^j}{d_u^j} & 0 \\ 0 & \frac{f_c^j}{d_v^j} \end{bmatrix} \begin{bmatrix} {}^ix_d^j \\ {}^iy_d^j \end{bmatrix} + \begin{bmatrix} c_u^j + d_{ur}^j + d_{ut}^j \\ c_v^j + d_{vr}^j + d_{vt}^j \end{bmatrix} \tag{3}
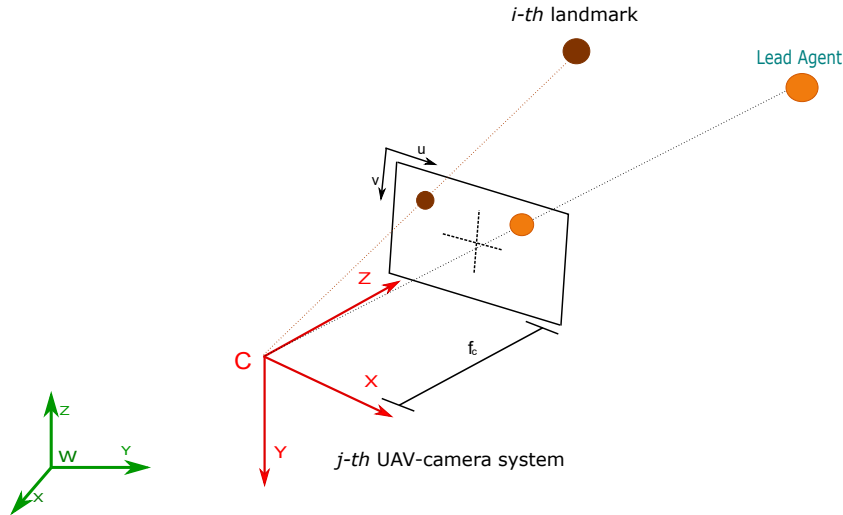$$

Figure 3. Pinhole camera projection model.

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 ${}^{i}\mathbf{p_{d}}^{j} = \begin{bmatrix} {}^{i}x_{d}^{j} & {}^{i}y_{d}^{j} & {}^{i}z_{d}^{j} \end{bmatrix}^{T}$ represent the position (in meters) of the *i*-th landmark with respect to the coordinate reference system *C* of the *j*-th camera.

Additionally,

$${}^{i}\mathbf{p_{d}}^{j} = {}^{W}\mathbf{R_{c}}^{j}(\mathbf{x_{a}}^{i} - \mathbf{x_{c}}^{j}) \tag{4}$$

where ${}^{W}\mathbf{R_{c}}^{j} \in SO3$ is the rotation matrix, that transforms from the world coordinate reference system *W* to the coordinate reference system *C* of the *j*-th camera. Recall that the rotation matrix ${}^{W}\mathbf{R_{c}}^{j}$ is known and constant, by the assumption of the use of the servo-controlled camera gimbal.

2.3. Camera Measurement Model for the Projection of the Lead Agent

Let consider the projection of the lead agent over the image plane of a camera. In this case, it is assumed that some visual feature points can be extracted from the lead agent by means of some available computer vision algorithm like [39–42] or [43].

Using the pinhole model (see Figure 3) the following expression can be defined:

$$
{}^{h}\mathbf{z_{c}}^{j} = {}^{h}\mathbf{h_{c}}^{j} = \begin{bmatrix} {}^{h}u_{c}^{j} \\ {}^{h}v_{c}^{j} \end{bmatrix} = \frac{1}{{}^{h}z_{d}^{j}} \begin{bmatrix} \frac{f_{c}^{j}}{d_{u}^{j}} & 0 \\ 0 & \frac{f_{c}^{j}}{d_{v}^{j}} \end{bmatrix} \begin{bmatrix} {}^{h}x_{d}^{j} \\ {}^{h}y_{d}^{j} \end{bmatrix} + \begin{bmatrix} c_{u}^{j} + d_{ur}^{j} + d_{ut}^{j} \\ c_{v}^{j} + d_{vr}^{j} + d_{vt}^{j} \end{bmatrix} \tag{5}
$$

Let ${}^{h}\mathbf{p_{d}}^{j} = \begin{bmatrix} {}^{h}x_{d}^{j} & {}^{h}y_{d}^{j} & {}^{h}z_{d}^{j} \end{bmatrix}^{T}$ represent the position (in meters) of the lead agent with respect to the coordinate reference system *C* of the *j*-th camera.

Additionally,

$${}^{h}\mathbf{p_{d}}^{j} = {}^{W}\mathbf{R_{c}}^{j}(\mathbf{x_{h}} - \mathbf{x_{c}}^{j}) \tag{6}$$

2.4. GPS Measurement Model

Let consider a GPS carried by the lead agent. From the GPS, measurements of the global position of the lead agent are obtained, therefore it can be defined:

$$\mathbf{z_g} = \mathbf{h_g} = \mathbf{x_h} \tag{7}$$

2.5. Altimeter Measurement Model

Let consider an altimeter carried by the *j*-th quadcopter. From the altimeter, measurements of the altitude of the *j*-th quadcopter are obtained, therefore it is defined:

$$z_a^j = h_a^j = z_c^j \tag{8}$$

2.6. Range Measurement Model

Let consider a range sensor, from which measurements of the relative distance of the *j*-th quadcopter respect to the lead agent can be obtained. Therefore, it is defined:

$${}^h z_r^j = {}^h h_r^j = \sqrt{(x_h - x_c^j)^2 + (y_h - y_c^j)^2 + (z_h - z_c^j)^2} \tag{9}$$

For a practical implementation, several techniques like [44] or [45] can be used in order to obtain these kinds of range measurements.

3. System Configurations

In this section, three different configurations of cooperative visual-based SLAM for Multi-UAV systems with a lead agent are introduced:

3.1. First Configuration

The first configuration takes into account the possibility of visual contact of at least one of the robots with the lead agent (See Figure 1). The system in its first configuration uses as inputs the following measurements: (i) the measurements obtained of the GPS carried by the lead agent, (ii) the monocular measurements obtained from the projection of the landmarks over each UAV-camera system, and (iii) the monocular measurements obtained from the projection of the lead agent over an UAV-camera system that has visual contact with it.

3.2. Second Configuration

The second configuration takes into account a scenario where there is no visual contact from any robot with the lead agent. In this case, the system is assumed to have the following sensor measurements: (i) the measurements obtained of the GPS carried by the lead agent, (ii) the monocular measurements obtained from the projection of the landmarks over each UAV-camera system, (iii) the measurements of the range of a robot with respect to the lead agent, and iv) the measurements obtained of an altimeter carried by one of the robots.

3.3. Third Configuration

The third configuration takes into account the scenarios where the GPS measurements are unavailable (GPS-denied environments). In Section 4, it will be analysed that in the absence of GPS the system will be partially observable for any possible configuration. Therefore, the objective will be to reduce the number of unobservable variables and modes of the system. In this case, in order to maximize the observability property of the system, in the absence of GPS, the third configuration is obtained by combining the first two configurations. The system in its third configuration is composed of the following measurements: (i) the monocular measurements obtained from the projection of the landmarks over each UAV-camera system, (ii) the monocular measurements obtained from the projection of the lead agent over an UAV-camera system that has visual contact with it, (iii) the measurements of the range of a robot with respect to the lead agent, and (iv) the measurements obtained by an altimeter carried by one of the robots.

Table 1 shows a summary of the three configurations according to the kind of measurements assumed to be available in each case.

Table 1. System configurations summary.

Configuration	Monocular Measurements of Landmarks	Monocular Measurements of Lead Agent	Altimeter on UAV	Range to Lead Agent	GPS on Lead Agent
First	✓	✓	✗	✗	✓
Second	✓	✗	✓	✓	✓
Third	✓	✓	✓	✓	✗

Regarding the practical scenarios where each configuration can be useful the following can be summarized: The first or second configuration can be considered for any scenario where the GPS measurements (even with low quality) are available, for instance outdoor or medium-dense environments. In particular, the second configuration will be useful in periods or circumstances where the lead agent is not seen by the monocular cameras carried by the UAVs. On the other hand, the third configuration will be more useful in any scenario or circumstance where there is no GPS availability, for instance in indoor or highly dense urban areas.

It is important to note that in a practical scenario, as it has been said before, some or all the UAVs can evolve in environments with low GPS coverage or even with no coverage. In this sense, from the implementation point of view, it is very straightforward to conceive an integrated system by simply including all the sensory inputs whenever they are available. If the minimum requirements established for each configuration are accomplished, then the integrated system will cope indistinctly with GPS-challenging or GPS-denied environments, with a performance subject to the observability criteria defined for each scenario.

4. Observability Analysis

As was previously mentioned, observability is a property of dynamic systems, because it has an important role in the convergence of the EKF. In this section, the observability properties of each cooperative SLAM configuration defined in Section 3 are studied. In the work of Hermann and Krener [46], it is demonstrated that a non-linear system is *locally weakly observable* if the observability rank condition $rank(\mathcal{O}) = dim(\mathbf{x})$ is verified, where $\mathcal{O}$ is the observability matrix.

4.1. Observability Matrices

In the case of the first configuration, the observability matrix $\mathcal{O}_1$ can be computed from:

$$\mathcal{O}_1 = \begin{bmatrix} \frac{\partial(\mathcal{L}_{\mathbf{f}}^0(^i\mathbf{h_c}^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1(^i\mathbf{h_c}^j))}{\partial \mathbf{x}} & \cdots & \frac{\partial(\mathcal{L}_{\mathbf{f}}^0(^h\mathbf{h_c}^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1(^h\mathbf{h_c}^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^0(\mathbf{h_g}))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1(\mathbf{h_g}))}{\partial \mathbf{x}} \end{bmatrix}^T \quad (10)$$

where $\mathcal{L}_{\mathbf{f}}^s\mathbf{h}$ is the *s*-th-order Lie Derivative [47], of the scalar field $\mathbf{h}$ with respect to the vector field $\mathbf{f}$. In (10) the zero-order and first-order Lie Derivatives are used for each measurement.

In the case of the measurements given by a monocular camera, for the projections of the landmarks, according to (3) and (1), the following zero-order Lie derivative can be defined:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^0(^i\mathbf{h_c}^j))}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{2\times6} & | & \mathbf{0}_{2\times6(j-1)} & -{}^i\mathbf{H_c}^{j\mathbf{W}}\mathbf{R_c}^{j} & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times6(n_2-j)} & | & \mathbf{0}_{2\times3(i-1)} & {}^i\mathbf{H_c}^{j\mathbf{W}}\mathbf{R_c}^{j} & \mathbf{0}_{2\times3(n_1-i)} \end{bmatrix} \quad (11)$$

where

$$^i\mathbf{H_c}^j = \frac{f_c^j}{(^iz_d^j)^2} \begin{bmatrix} \frac{{}^iz_d^j}{d_u^j} & 0 & -\frac{{}^ix_d^j}{d_u^j} \\ \\ 0 & \frac{{}^iz_d^j}{d_v^j} & -\frac{{}^iy_d^j}{d_v^j} \end{bmatrix} \quad (12)$$

For the same kind of measurement, the following first-order Lie Derivative can be defined:

$$\frac{\partial(\mathcal{L}_f^1({}^i\mathbf{h_c}^j))}{\partial\mathbf{x}} = \begin{bmatrix} \mathbf{0}_{2\times6} & | & \mathbf{0}_{2\times6(j-1)} & {}^i\mathbf{H_{dc}}^j & -{}^i\mathbf{H_c}^{j\,\mathbf{W}}\mathbf{R_c}^j & \mathbf{0}_{2\times6(n_2-j)} & | & \mathbf{0}_{2\times3(i-1)} & -{}^i\mathbf{H_{dc}}^j & \mathbf{0}_{2\times3(n_1-i)} \end{bmatrix} \tag{13}$$

where

$${}^i\mathbf{H_{dc}}^j = \begin{bmatrix} {}^i\mathbf{H_1}^j & {}^i\mathbf{H_2}^j & {}^i\mathbf{H_3}^j \end{bmatrix} \left({}^\mathbf{W}\mathbf{R_c}^j\right)^2 \mathbf{v_c}^j \tag{14}$$

and

$${}^i\mathbf{H_1}^j = \frac{f_c^j}{d_u^j({}^iz_d^j)^2}\begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad {}^i\mathbf{H_2}^j = \frac{f_c^j}{d_v^j({}^iz_d^j)^2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad {}^i\mathbf{H_3}^j = \frac{f_c^j}{({}^iz_d^j)^3}\begin{bmatrix} -\frac{{}^iz_d^j}{d_u^j} & 0 & \frac{2{}^ix_d^j}{d_u^j} \\ 0 & -\frac{{}^iz_d^j}{d_v^j} & \frac{2{}^iy_d^j}{d_v^j} \end{bmatrix} \tag{15}$$

In the case of the measurement given by a monocular camera, for the projections of the lead agent, according to (5) and (1), the following zero-order Lie derivative can be defined:

$$\frac{\partial(\mathcal{L}_f^0({}^h\mathbf{h_c}^j))}{\partial\mathbf{x}} = \begin{bmatrix} {}^h\mathbf{H_c}^{j\,\mathbf{W}}\mathbf{R_c}^j & \mathbf{0}_{2\times3} & | & \mathbf{0}_{2\times6(j-1)} & -{}^h\mathbf{H_c}^{j\,\mathbf{W}}\mathbf{R_c}^j & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times6(n_2-j)} & | & \mathbf{0}_{2\times3n_1} \end{bmatrix} \tag{16}$$

where

$${}^h\mathbf{H_c}^j = \frac{f_c^j}{({}^hz_d^j)^2}\begin{bmatrix} \frac{{}^hz_d^j}{d_u^j} & 0 & -\frac{{}^hx_d^j}{d_u^j} \\ 0 & \frac{{}^hz_d^j}{d_v^j} & -\frac{{}^hy_d^j}{d_v^j} \end{bmatrix} \tag{17}$$

For the same kind of measurement, the following first-order Lie Derivative can be defined:

$$\frac{\partial(\mathcal{L}_f^1({}^h\mathbf{h_c}^j))}{\partial\mathbf{x}} = \begin{bmatrix} -{}^h\mathbf{H_{dc}}^j & {}^h\mathbf{H_c}^{j\,\mathbf{W}}\mathbf{R_c}^j & | & \mathbf{0}_{2\times6(j-1)} & {}^h\mathbf{H_{dc}}^j & -{}^h\mathbf{H_c}^{j\,\mathbf{W}}\mathbf{R_c}^j & \mathbf{0}_{2\times6(n_2-j)} & | & \mathbf{0}_{2\times3n_1} \end{bmatrix} \tag{18}$$

where

$${}^h\mathbf{H_{dc}}^j = \begin{bmatrix} {}^h\mathbf{H_1}^j & {}^h\mathbf{H_2}^j & {}^h\mathbf{H_3}^j \end{bmatrix} \left({}^\mathbf{W}\mathbf{R_c}^j\right)^2 \left(\mathbf{v_c}^j - \mathbf{v_h}\right) \tag{19}$$

and

$${}^h\mathbf{H_1}^j = \frac{f_c^j}{d_u^j({}^hz_d^j)^2}\begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad {}^h\mathbf{H_2}^j = \frac{f_c^j}{d_v^j({}^hz_d^j)^2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad {}^h\mathbf{H_3}^j = \frac{f_c^j}{({}^hz_d^j)^3}\begin{bmatrix} -\frac{{}^hz_d^j}{d_u^j} & 0 & \frac{2{}^hx_d^j}{d_u^j} \\ 0 & -\frac{{}^hz_d^j}{d_v^j} & \frac{2{}^hy_d^j}{d_v^j} \end{bmatrix} \tag{20}$$

In the case of the measurement given by a GPS carried for the lead agent, according to (7) and (1), the following zero-order Lie derivative can be defined:

$$\frac{\partial(\mathcal{L}_f^0(\mathbf{h_g}))}{\partial\mathbf{x}} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & | & \mathbf{0}_{3\times6n_2} & | & \mathbf{0}_{3\times3n_1} \end{bmatrix} \tag{21}$$

where $\mathbf{I}$ is the identity matrix. For the same kind of measurement, the following first-order Lie Derivative can be defined:

$$\frac{\partial(\mathcal{L}_f^1(\mathbf{h_g}))}{\partial\mathbf{x}} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{I}_3 & | & \mathbf{0}_{3\times6n_2} & | & \mathbf{0}_{3\times3n_1} \end{bmatrix} \tag{22}$$

For the second configuration, the observability matrix $\mathcal{O}_\mathbf{2}$ can be computed as follows:

$$\mathcal{O}_\mathbf{2} = \begin{bmatrix} \frac{\partial(\mathcal{L}_f^0({}^i\mathbf{h_c}^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1({}^i\mathbf{h_c}^j))}{\partial\mathbf{x}} & \cdots & \frac{\partial(\mathcal{L}_f^0({}^h\mathbf{h}_r^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1({}^h\mathbf{h}_r^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^0(\mathbf{h}_a^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(\mathbf{h}_a^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^0(\mathbf{h_g}))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(\mathbf{h_g}))}{\partial\mathbf{x}} \end{bmatrix}^T \tag{23}$$

In this case, from the measurement given by a range sensor, according to (9) and (1), the following zero-order Lie derivative can be defined:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^0({}^h h_r^j))}{\partial \mathbf{x}} = \begin{bmatrix} {}^h\mathbf{H_r}^j & \mathbf{0}_{1\times 3} & | & \mathbf{0}_{1\times 6(j-1)} & -{}^h\mathbf{H_r}^j & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 6(n_2-j)} & | & \mathbf{0}_{1\times 3n_1} \end{bmatrix} \quad (24)$$

where

$$ {}^h\mathbf{H_r}^j = \begin{bmatrix} \frac{x_h - x_c^j}{{}^h h_r^j} & \frac{y_h - y_c^j}{{}^h h_r^j} & \frac{z_h - z_c^j}{{}^h h_r^j} \end{bmatrix} \quad (25)$$

For the same kind of measurement, the following first-order Lie Derivative can be defined:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^1({}^h h_r^j))}{\partial \mathbf{x}} = \begin{bmatrix} {}^h\mathbf{H_{dr}}^j & {}^h\mathbf{H_r}^j & | & \mathbf{0}_{1\times 6(j-1)} & -{}^h\mathbf{H_{dr}}^j & -{}^h\mathbf{H_r}^j & \mathbf{0}_{1\times 6(n_2-j)} & | & \mathbf{0}_{1\times 3n_1} \end{bmatrix} \quad (26)$$

where

$$\left({}^h\mathbf{H_{dr}}^j\right)^T = \frac{1}{{}^h h_r^j}\left[\mathbf{I}_3 - \left({}^h\mathbf{H_r}^j\right)^T {}^h\mathbf{H_r}^j\right]\left(\mathbf{v_h} - \mathbf{v_c}^j\right) \quad (27)$$

In the case of the measurement given by an altimeter carried by the *j*-th quadcopter, according to (8) and (1), the following zero-order Lie derivative can be defined:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^0(h_a^j))}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{1\times 6} & | & \mathbf{0}_{1\times 6(j-1)} & \mathbf{0}_{1\times 2} & 1 & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 6(n_2-j)} & | & \mathbf{0}_{1\times 3n_1} \end{bmatrix} \quad (28)$$

For the same kind of measurement, the following first-order Lie Derivative can be defined:

$$\frac{\partial(\mathcal{L}_{\mathbf{f}}^1(h_a^j))}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0}_{1\times 6} & | & \mathbf{0}_{1\times 6(j-1)} & \mathbf{0}_{1\times 5} & 1 & \mathbf{0}_{1\times 6(n_2-j)} & | & \mathbf{0}_{1\times 3n_1} \end{bmatrix} \quad (29)$$

The third configuration of the system takes into account circumstances where the GPS can be unavailable. In this case, the observability matrix $\mathcal{O}_\mathbf{3}$ can be computed as follows:

$$\mathcal{O}_\mathbf{3} = \begin{bmatrix} \frac{\partial(\mathcal{L}_{\mathbf{f}}^0({}^i\mathbf{h_c}^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1({}^i\mathbf{h_c}^j))}{\partial \mathbf{x}} & \cdots & \frac{\partial(\mathcal{L}_{\mathbf{f}}^0({}^h\mathbf{h_c}^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1({}^h\mathbf{h_c}^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^0({}^h h_r^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1({}^h h_r^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^0(h_a^j))}{\partial \mathbf{x}} & \frac{\partial(\mathcal{L}_{\mathbf{f}}^1(h_a^j))}{\partial \mathbf{x}} \end{bmatrix}^T \quad (30)$$

4.2. Theoretical Results

For the sake of presentation, a single observability matrix $\mathcal{O}_t$ is expanded including all the Lie derivatives obtained from each system configuration. In this case, in order to differentiate the Lie derivatives that belong to each kind of system configuration, the rows of the matrix (31) are differentiated by colors: The rows that are common to all configurations are indicated in blue. The rows that are common to the first and second configuration are indicated in red. The rows that are common to the first and third configuration are indicated in orange. The rows that are common to the second and third configuration are indicated in black. For instance, according to the above, the observability matrix $\mathcal{O}_\mathbf{2}$ (second configuration) will be composed of selecting the blue and red rows from matrix $\mathcal{O}_t$.

$$
\boldsymbol{\mathcal{O}}_t = \left[\begin{array}{cccccccccccc} \frac{\partial(\mathcal{L}_f^0(^i\mathbf{h}_c{}^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(^i\mathbf{h}_c{}^j))}{\partial\mathbf{x}} & \cdots & \frac{\partial(\mathcal{L}_f^0(\mathbf{h_g}))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(\mathbf{h_g}))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^0(^h\mathbf{h}_c{}^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(^h\mathbf{h}_c{}^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^0(^hh_r^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(^hh_r^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^0(h_a^j))}{\partial\mathbf{x}} & \frac{\partial(\mathcal{L}_f^1(h_a^j))}{\partial\mathbf{x}} \end{array} \right]^T
$$

$$
= \begin{bmatrix}
\mathbf{0}_{2\times6} & \vline & \mathbf{0}_{2\times6(j-1)} & -{}^i\mathbf{H_c}{}^j\mathbf{WR_c}{}^j & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times6(n_2-j)} & \vline & \mathbf{0}_{2\times3(i-1)} & {}^i\mathbf{H_c}{}^j\mathbf{WR_c}{}^j & \mathbf{0}_{2\times3(n_1-i)} \\[2pt]
\mathbf{0}_{2\times6} & \vline & \mathbf{0}_{2\times6(j-1)} & {}^i\mathbf{H_{dc}}{}^j & -{}^i\mathbf{H_c}{}^j\mathbf{WR_c}{}^j & \mathbf{0}_{2\times6(n_2-j)} & \vline & \mathbf{0}_{2\times3(i-1)} & -{}^i\mathbf{H_{dc}}{}^j & \mathbf{0}_{2\times3(n_1-i)} \\
\vdots & & & \vdots & & & & & \vdots & \\[4pt]
\mathbf{I}_3 \quad \mathbf{0}_3 & \vline & & \mathbf{0}_{3\times6n_2} & & & \vline & & \mathbf{0}_{3\times3n_1} & \\[4pt]
\mathbf{0}_3 \quad \mathbf{I}_3 & \vline & & \mathbf{0}_{3\times6n_2} & & & \vline & & \mathbf{0}_{3\times3n_1} & \\[4pt]
{}^h\mathbf{H_c}{}^j\mathbf{WR_c}{}^j \ \ \mathbf{0}_{2\times3} & \vline & \mathbf{0}_{2\times6(j-1)} & -{}^h\mathbf{H_c}{}^j\mathbf{WR_c}{}^j & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times6(n_2-j)} & \vline & & \mathbf{0}_{2\times3n_1} & \\[4pt]
-{}^h\mathbf{H_{dc}}{}^j \ \ {}^h\mathbf{H_c}{}^j\mathbf{WR_c}{}^j & \vline & \mathbf{0}_{2\times6(j-1)} & {}^h\mathbf{H_{dc}}{}^j & -{}^h\mathbf{H_c}{}^j\mathbf{WR_c}{}^j & \mathbf{0}_{2\times6(n_2-j)} & \vline & & \mathbf{0}_{2\times3n_1} & \\[4pt]
{}^h\mathbf{H_r}{}^j \ \ \mathbf{0}_{1\times3} & \vline & \mathbf{0}_{1\times6(j-1)} & -{}^h\mathbf{H_r}{}^j & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times6(n_2-j)} & \vline & & \mathbf{0}_{1\times3n_1} & \\[4pt]
{}^h\mathbf{H_{dr}}{}^j \ \ {}^h\mathbf{H_r}{}^j & \vline & \mathbf{0}_{1\times6(j-1)} & -{}^h\mathbf{H_{dr}}{}^j & -{}^h\mathbf{H_r}{}^j & \mathbf{0}_{1\times6(n_2-j)} & \vline & & \mathbf{0}_{1\times3n_1} & \\[4pt]
\mathbf{0}_{1\times6} & \vline & \mathbf{0}_{1\times6(j-1)} & \mathbf{0}_{1\times2} \ 1 \ \mathbf{0}_{1\times3} & \mathbf{0}_{1\times6(n_2-j)} & & \vline & & \mathbf{0}_{1\times3n_1} & \\[4pt]
\mathbf{0}_{1\times6} & \vline & \mathbf{0}_{1\times6(j-1)} & \mathbf{0}_{1\times5} \ 1 \ \mathbf{0}_{1\times6(n_2-j)} & & & \vline & & \mathbf{0}_{1\times3n_1} &
\end{bmatrix} \tag{31}
$$

According to (10) and (31), the rank of the observability matrix $\boldsymbol{\mathcal{O}}_1$ is $rank(\boldsymbol{\mathcal{O}}_1) = 3n_1 + 6n_2 + 6$, where n_1 is the number of landmarks being measured, n_2 is the number of robots and 6, is the number of states of the lead agent. In this case, n_1 is multiplied by 3, since this is the number of states per landmark given by the Euclidean parametrization, n_2 is multiplied by 6, since this is the number of states per robot given by its global position and its derivatives. Given that $dim(\mathbf{x}) = 3n_1 + 6n_2 + 6$, then, the system under its first configuration is *locally weakly observable*, because $rank(\boldsymbol{\mathcal{O}}) = dim(\mathbf{x})$.

Now, according to (23) and (31), the rank of the observability matrix $\boldsymbol{\mathcal{O}}_2$ is $rank(\boldsymbol{\mathcal{O}_2}) = 3n_1 + 6n_2 + 6$. Therefore, the proposed system under the second configuration is also *locally weakly observable*.

For the third configuration, according to (30) and (31), the maximum rank of the observability matrix $\boldsymbol{\mathcal{O}_3}$ is $rank(\boldsymbol{\mathcal{O}_3}) = (3n_1 + 6n_2 + 6) - 2$. Therefore, $\boldsymbol{\mathcal{O}_3}$ will be rank deficient ($rank(\boldsymbol{\mathcal{O}_3}) < dim(\mathbf{x})$). In this case, the unobservable modes are spanned by the right nullspace basis of the observability matrix $\boldsymbol{\mathcal{O}_3}$, therefore:

$$
\mathbf{N} = null(\boldsymbol{\mathcal{O}_3}) = \begin{bmatrix} \mathbf{S} \\ \mathbf{0}_{3\times2} \\ \hline \mathbf{S} \\ \mathbf{0}_{3\times2} \\ \vdots \\ \mathbf{S} \\ \mathbf{0}_{3\times2} \\ \hline \mathbf{S} \\ \vdots \\ \mathbf{S} \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{32}
$$

It is straightforward to verify that the right nullspace basis of $\boldsymbol{\mathcal{O}_3}$ spans for $\mathbf{N}$, (i.e., $\boldsymbol{\mathcal{O}_3}\mathbf{N} = \mathbf{0}$).

From (32) 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 in x and y of the robots, the landmarks

and the lead agent; 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.

It is very important to note, that in the absence of GPS measurements the system cannot be completely observable, given the absence of global position measurements within the system. This result is consistent with the one that can be expected from any SLAM system, which is world-centric but does not include global measurements. Therefore, in the case of the third configuration, the objective was to obtain the best possible observability result, that is, to reduce to a greater extent the number of unobservable states. It is worth noting that the first and second configurations obtain a less favorable result with respect to the third configuration in terms of the observability property when GPS measurements are not considered (see Table 2).

Table 2. Results of observability in the absence of GPS.

	Unobservable Modes	Unobservable States	Observable States
Configuration 1	4	$\mathbf{x}$	-
Configuration 2	6	$\mathbf{x_h}, \mathbf{v_h}, x_c^j, y_c^j, x_a^i, y_a^i$	$z_c^j, z_a^i, \mathbf{v_c}^j$
Configuration 3	2	$x_h, y_h, x_c^j, y_c^j, x_a^i, y_a^i$	$z_h, z_c^j, z_a^i, \mathbf{v_h}, \mathbf{v_c}^j$

Some important remarks about the former observability analysis can be extracted:

- In the case of the first and third configurations, a single robot having visual contact with the lead agent represents a sufficient condition to obtain the previous results (see Figure 4).
- In the case of the second and third configurations, only a single range sensor, used for measuring the relative distance of the *j*-th quadcopter respect to the lead agent, is needed in order to obtain the previous results (see Figure 4).
- In the case of the second and third configurations, only a single altimeter sensor, used for measuring the altitude of the *j*-th quadcopter, is needed in order to obtain the previous results (see Figure 4).
- In the case of the third configuration, if two or more robots have visual contact with the lead agent, range measurements are not necessary in order to obtain the previous results.
- For all the configurations, in order to obtain the previous results, it is necessary to link the members of the multi-UAV system through the measurements of the landmarks (see Figure 4). In other words, a robot needs to share the observation of at least two landmarks with another robot.
- In the case of the third configuration, adding Lie Derivatives of higher order to the observability matrix does not improve the results.
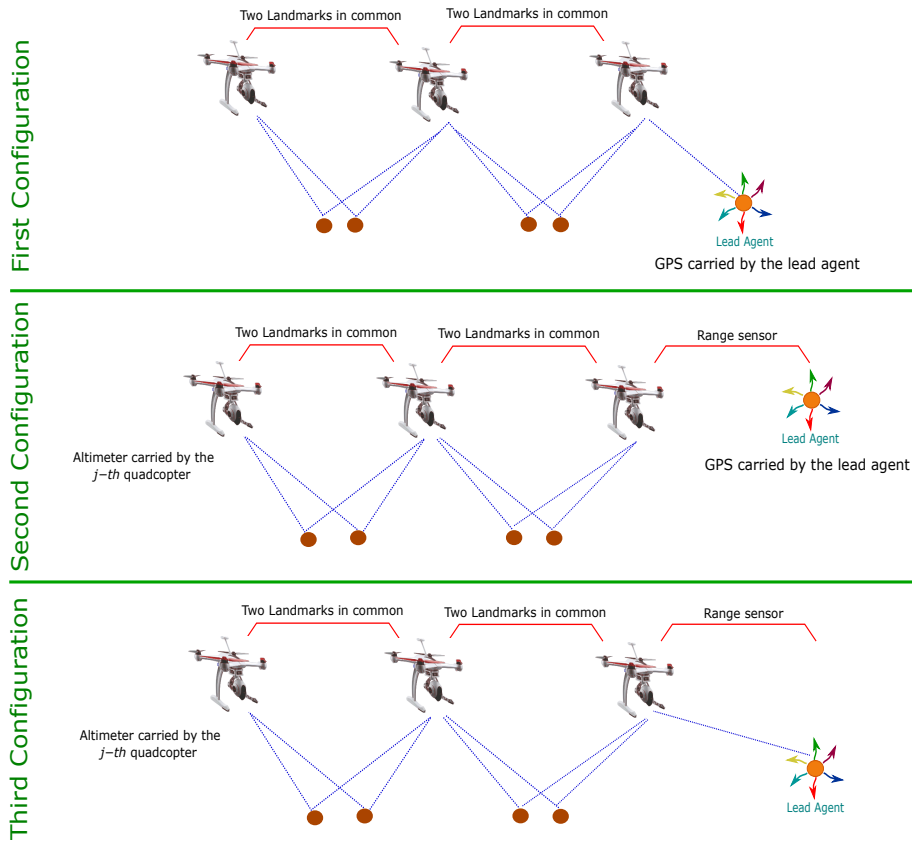
Figure 4. System configurations and minimum requirements for obtaining the results of the observability analysis.

5. EKF-Cooperative Monocular SLAM

The main goal of the proposed method is to estimate the system state using the EKF-based SLAM methodology. Figure 5 shows the architecture of the proposed system under its three configurations.
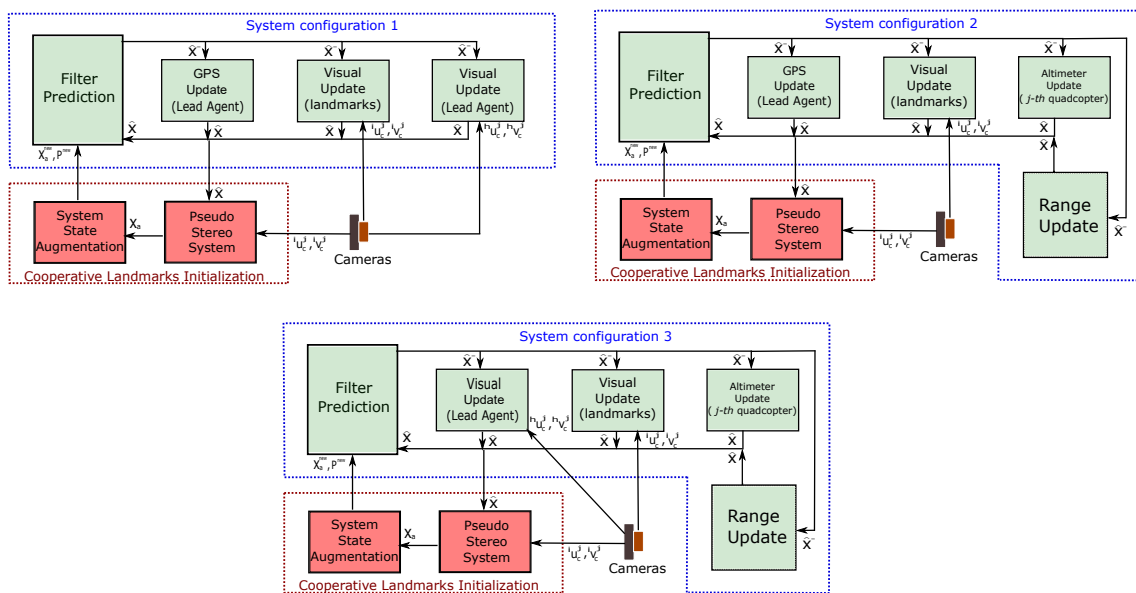

Figure 5. EKF-Cooperative SLAM: Block diagram showing the architecture of the system under its three configurations.

5.1. EKF-SLAM

Using (1), the following discretized system state model a can be defined:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{n}_{k-1}) = \begin{bmatrix} \mathbf{x}_{\mathbf{h}k} \\ \mathbf{v}_{\mathbf{h}k} \\ \mathbf{x}_{c_k}^j \\ \mathbf{v}_{c_k}^j \\ \mathbf{x}_{a_k}^i \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{\mathbf{h}k-1} + (\mathbf{v}_{\mathbf{h}k-1})\Delta t \\ \mathbf{v}_{\mathbf{h}k-1} + \boldsymbol{\zeta}_{\mathbf{h}k-1} \\ \mathbf{x}_{c_{k-1}}^j + (\mathbf{v}_{c_{k-1}}^j)\Delta t \\ \mathbf{v}_{c_{k-1}}^j + \boldsymbol{\eta}_{c_{k-1}}^j \\ \mathbf{x}_{a_{k-1}}^i \end{bmatrix} \tag{33}$$

$$\mathbf{n}_k = \begin{bmatrix} \boldsymbol{\zeta}_{\mathbf{h}k} \\ \boldsymbol{\eta}_{c_k}^j \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{\mathbf{h}}\Delta t \\ \boldsymbol{\alpha}_{\mathbf{c}}^j \Delta t \end{bmatrix} \tag{34}$$

For configuration 1, the system measurements are defined according to (3), (5) and (7), as:

$$\mathbf{z}_{1k} = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_{1k}) = \begin{bmatrix} {}^i\mathbf{h}_{c_k}^j + {}^i\mathbf{r}_{c_k}^j \\ {}^h\mathbf{h}_{c_k}^j + {}^h\mathbf{r}_{c_k}^j \\ \mathbf{h}_{g_k} + \mathbf{r}_{g_k} \end{bmatrix} \tag{35}$$

$$\mathbf{r}_{1k} = \begin{bmatrix} {}^i\mathbf{r}_{c_k}^j \\ {}^h\mathbf{r}_{c_k}^j \\ \mathbf{r}_{g_k} \end{bmatrix} \tag{36}$$

For configuration 2, the system measurements are defined according to (3), (7), (8) and (9), as:

$$\mathbf{z}_{2k} = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_{2k}) = \begin{bmatrix} {}^i\mathbf{h}_{c_k}^j + {}^i\mathbf{r}_{c_k}^j \\ \mathbf{h}_{g_k} + \mathbf{r}_{g_k} \\ h_{ak}^j + r_{ak}^j \\ {}^h h_{rk}^j + {}^h r_{rk}^j \end{bmatrix} \tag{37}$$

$$\mathbf{r}_{2k} = \begin{bmatrix} {}^i\mathbf{r}_{c_k}^j \\ \mathbf{r}_{g_k} \\ r_{ak}^j \\ {}^h r_{rk}^j \end{bmatrix} \tag{38}$$

For configuration 3, the system measurements are defined according to (3), (5), (8) and (9), as:

$$\mathbf{z}_{3k} = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_{3k}) = \begin{bmatrix} {}^i\mathbf{h}_{c_k}^j + {}^i\mathbf{r}_{c_k}^j \\ {}^h\mathbf{h}_{c_k}^j + {}^h\mathbf{r}_{c_k}^j \\ h_{ak}^j + r_{ak}^j \\ {}^h h_{rk}^j + {}^h r_{rk}^j \end{bmatrix} \tag{39}$$

$$\mathbf{r}_{3k} = \begin{bmatrix} {}^i\mathbf{r}_{c_k}^j \\ {}^h\mathbf{r}_{c_k}^j \\ r_{ak}^j \\ {}^h r_{rk}^j \end{bmatrix} \tag{40}$$

where $\mathbf{a_h}$ and $\boldsymbol{\alpha_c}^j$ represent unknown linear accelerations that are assumed to have Gaussian distribution with zero mean. Moreover, let $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$, $\mathbf{r}_{1k} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{1k})$, $\mathbf{r}_{2k} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{2k})$ and $\mathbf{r}_{3k} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{3k})$ be the noise vectors that affect the state and the measurements, which are assumed to be mutually uncorrelated. Let Δt be the differential of time and k the sample step. In this work,

a Gaussian random process is used for propagating the velocity of the vehicle. The proposed scheme is independent of the kind of aircraft and therefore is not restricted by the use of a specific dynamic model.

The prediction stage of the EKF is defined by:

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{0}) \tag{41}$$

$$\mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k \mathbf{Q}_{k-1} \mathbf{W}_k^T \tag{42}$$

The correction stage of the EKF is defined by:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-, \mathbf{0})) \tag{43}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k)\mathbf{P}_k^- \tag{44}$$

with

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^T)^{-1} \tag{45}$$

and

$$
\begin{aligned}
\mathbf{A}_k &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k-1}, \mathbf{0}) & \mathbf{C}_k &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_k^-, \mathbf{0}) \\
\mathbf{W}_k &= \frac{\partial \mathbf{f}}{\partial \mathbf{n}}(\hat{\mathbf{x}}_{k-1}, \mathbf{0}) & \mathbf{V}_k &= \frac{\partial \mathbf{h}}{\partial \mathbf{r}}(\hat{\mathbf{x}}_k^-, \mathbf{0})
\end{aligned}
\tag{46}
$$

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

5.2. Initialization of Map Features

The initialization process of new map features is carried out in a cooperative way. The 3D position of the new map features is estimated by means of a pseudo-stereo system formed by the monocular cameras mounted on a pair of UAVs that observe common landmarks. In this case, the process of initialization is carried out when a new potential landmark is observed by two cameras; if this condition is fulfilled then the landmark can be initialized by means of a linear triangulation.

The state of the new feature is computed using the *a posteriori* values obtained in the correction stage of the EKF. According to (3), the following expression can be defined in homogeneous coordinates:

$$
{}^i\gamma_c^j \begin{bmatrix} {}^i u_c^j \\ {}^i v_c^j \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{T_c}^j & \mathbf{0}_{3\times1} \end{bmatrix} \hat{\mathbf{E}}_c^j \begin{bmatrix} \mathbf{x_a}^i \\ 1 \end{bmatrix} \tag{47}
$$

where ${}^i\gamma_c^j$ is a scale factor. Additionally, it is defined:

$$
\hat{\mathbf{E}}_c^j = \begin{bmatrix} {}^W\hat{\mathbf{R}}_c^j & \hat{\mathbf{x}}_c^j \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix} \qquad \mathbf{T_c}^j = \begin{bmatrix} \frac{f_c^j}{d_u^j} & 0 & c_u^j + d_{ur}^j + d_{ut}^j \\ 0 & \frac{f_c^j}{d_v^j} & c_v^j + d_{vr}^j + d_{vt}^j \\ 0 & 0 & 1 \end{bmatrix} \tag{48}
$$

Using (47), and considering the projection onto two any UAV cameras, a linear system can be formed in order to estimate $\mathbf{x_a}^i$:

$$\mathbf{D}^i \mathbf{x_a}^i = \mathbf{b}^i \qquad \mathbf{x_a}^i = \mathbf{D}^{i\dagger} \mathbf{b}^i \tag{49}$$

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

$$\mathbf{D}^i = \begin{bmatrix} k_{31}^j \, {}^i u_c^j - k_{11}^j & k_{32}^j \, {}^i u_c^j - k_{12}^j & k_{33}^j \, {}^i u_c^j - k_{13}^j \\ k_{31}^j \, {}^i v_c^j - k_{21}^j & k_{32}^j \, {}^i v_c^j - k_{22}^j & k_{33}^j \, {}^i v_c^j - k_{23}^j \end{bmatrix} \quad \mathbf{b}^i = \begin{bmatrix} k_{14}^j - k_{34}^j \, {}^i u_c^j \\ k_{24}^j - k_{34}^j \, {}^i v_c^j \end{bmatrix} \tag{50}$$

with

$$\begin{bmatrix} \mathbf{T_c}^j & \mathbf{0}_{3\times 1} \end{bmatrix} \hat{\mathbf{E}}_\mathbf{c}^j = \begin{bmatrix} k_{11}^j & k_{12}^j & k_{13}^j & k_{14}^j \\ k_{21}^j & k_{22}^j & k_{23}^j & k_{24}^j \\ k_{31}^j & k_{32}^j & k_{33}^j & k_{34}^j \end{bmatrix} \tag{51}$$

when a new landmark is initialized, the system state $\mathbf{x}$ is augmented by: $\mathbf{x} = \begin{bmatrix} \mathbf{x_h} & \mathbf{v_h} & \mathbf{x_c}^j & \mathbf{v_c}^j & \mathbf{x_a}^i & \mathbf{x_a}^{new} \end{bmatrix}^T$. And the new covariance matrix $\mathbf{P}_{new}$ is computed by:

$$\mathbf{P}_{new} = \Delta\mathbf{J} \begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & {}^i\mathbf{R}^j \end{bmatrix} = \Delta\mathbf{J}^T \tag{52}$$

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

6. Control Flight Formation

This section presents the high-level control scheme that allows to carry out the formation of the UAVs with respect to the lead agent. The kinematic model is based on the leader-follower scheme presented in [48].

The kinematic model of the *j*-th UAV can be described as:

$$\begin{aligned} \dot{x}^j &= v_x^j \cos(\psi^j) - v_y^j \sin(\psi^j) \\ \dot{y}^j &= v_x^j \sin(\psi^j) + v_y^j \cos(\psi^j) \\ \dot{z}^j &= v_z^j \\ \dot{\psi}^j &= \omega^j \end{aligned} \tag{53}$$

Let $\mathbf{x_q}^j = \begin{bmatrix} x^i & y^i & z^i \end{bmatrix}^T$ be the position (in meters) of the *j*-th UAV, with respect to the coordinate reference system W (See Figure 6). Let v_x^j and v_y^j be the linear velocity (in m/s) components in the x and y directions of the *j*-th UAV, with respect to the coordinate reference system Q (located in the center of mass of the aerial vehicle). Let v_z^j be the linear velocity (in m/s) component in the z direction of the *j*-th UAV, with respect to the coordinate reference system W. Let ψ^j be the yaw angle (in *radians*) of the *j*-th UAV, with respect to the coordinate reference system W. Let ω^j be the angular velocity (in *radians/s*) for the yaw angle of the *j*-th UAV, with respect to the coordinate reference system W.

It is desired to maintain the *j*-th UAV to a distance λ_x^j, λ_y^j from the lead agent (See Figure 6). Let λ_x^j, λ_y^j be the coordinates of the lead agent with respect to the coordinate reference system Q. It is also desired to maintain the *j*-th UAV at an altitude differential λ_z^j from the lead agent. Given all these considerations, the following can be defined:

$$\begin{aligned} \lambda_x^j &= (x_h - x^j)\cos(\psi^j) + (y_h - y^j)\sin(\psi^j) \\ \lambda_y^j &= -(x_h - x^j)\sin(\psi^j) + (y_h - y^j)\cos(\psi^j) \\ \lambda_z^j &= z^j - z_h \end{aligned} \tag{54}$$

Differentiating (54) with respect to time, using (1), (53) and (54), the following can be obtained:

$$
\begin{aligned}
\dot{\lambda}_x^j &= \omega^j \lambda_y^j + \dot{x}_h \cos(\psi^j) + \dot{y}_h \sin(\psi^j) - v_x^j \\
\dot{\lambda}_y^j &= -\omega^j \lambda_x^j - \dot{x}_h \sin(\psi^j) + \dot{y}_h \cos(\psi^j) - v_y^j \\
\dot{\lambda}_z^j &= v_z^j - \dot{z}_h
\end{aligned}
\tag{55}
$$

Finally, the dynamics of the formation can be defined as follows:

$$
\dot{\lambda} = \mathbf{g} + \mathbf{u}
\tag{56}
$$

where

$$
\lambda = \begin{bmatrix} \lambda_x^j \\ \lambda_y^j \\ \lambda_z^j \\ \psi^j \end{bmatrix}
\qquad
\mathbf{g} = \begin{bmatrix} \omega^j \lambda_y^j + \dot{x}_h \cos(\psi^j) + \dot{y}_h \sin(\psi^j) \\ -\omega^j \lambda_x^j - \dot{x}_h \sin(\psi^j) + \dot{y}_h \cos(\psi^j) \\ -\dot{z}_h \\ 0 \end{bmatrix}
\qquad
\mathbf{u} = \begin{bmatrix} -v_x^j \\ -v_y^j \\ v_z^j \\ \omega^j \end{bmatrix}
\tag{57}
$$

In the following, it is assumed that there exist uncertainty and disturbances coupled to the input of the system. Therefore, Equation (56) can be defined as:

$$
\dot{\lambda} = \mathbf{g} + \Delta_g + \mathbf{u}
\tag{58}
$$

where Δ_g is a bounded and unknown uncertainty and disturbances term satisfying $\Delta_g \leq \epsilon$, being ϵ a positive constant. The control scheme is designed in order to allow that the proposed dynamics of the formation converges to the desired values. In this case, the sliding mode control technique [49] is proposed to be used for the development of a robust controller. The state values required by the control system are obtained from the estimator described in Section 5. The yaw angle of the *j*-th UAV is obtained through an AHRS onboard the aerial vehicle. To obtain the control laws by means of an analysis in continuous time it is assumed that the estimated value is passed through a zero order hold (ZOH) (See Figure 7).
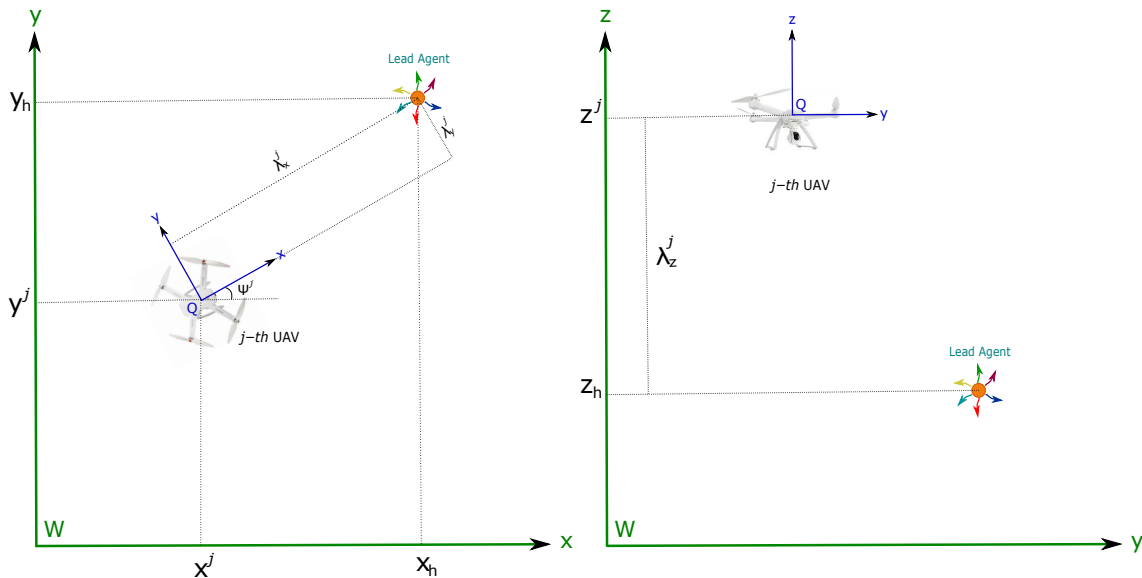

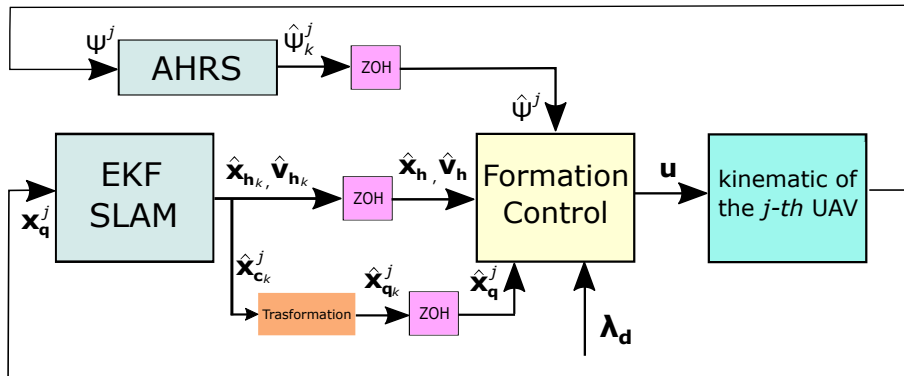**Figure 6.** UAVs-lead agent flight formation.

Figure 7. Control scheme.

Since the estimated state of the position of the UAV is defined with respect to the reference system C (see Sections 2 and 5), it is necessary to apply a transformation to the estimated state $\hat{\mathbf{x}}_{\mathbf{c}}^{j}$ for obtaining $\hat{\mathbf{x}}_{\mathbf{q}}^{j}$, which in turn, is necessary to obtain the control laws. Therefore, the following equation is defined:

$$\hat{\mathbf{x}}_{\mathbf{q}}^{j} = \hat{\mathbf{x}}_{\mathbf{c}}^{j} - {}^{\mathbf{q}}\mathbf{d}_{\mathbf{c}}{}^{j} \tag{59}$$

where ${}^{\mathbf{q}}\mathbf{d}_{\mathbf{c}}{}^{j}$ is the translation vector (in meters) from the coordinate reference system Q to the coordinate reference system C. Please note that ${}^{\mathbf{q}}\mathbf{d}_{\mathbf{c}}{}^{j}$ is assumed to be known and constant.

Firstly, the next equation is defined:

$$s_\lambda = e_\lambda + \mathbf{K_1} \int_0^t e_\lambda dt \tag{60}$$

where $e_\lambda = \hat{\lambda} - \lambda_d$, and λ_d are the desired values and $\mathbf{K_1}$ is a positive definite diagonal matrix of adequate dimensions. Then, deriving (60) and substituting the dynamics defined in (58) as well as taking **u** as the control input, the following equation can be obtained:

$$\dot{s}_\lambda = -\dot{\lambda}_d + \mathbf{K_1} e_\lambda + \hat{\mathbf{g}} + \Delta_g + \mathbf{u} \tag{61}$$

The following control law is proposed:

$$\mathbf{u} = +\dot{\lambda}_d - \mathbf{K_1} e_\lambda - \hat{\mathbf{g}} - \mathbf{K_2}\text{sign}(s_\lambda) \tag{62}$$

where $\mathbf{K_2}$ is a positive definite diagonal matrix of adequate dimensions.

To prove the stability of the flight formation system dynamics, the following Lyapunov candidate function is proposed:

$$V_\lambda = \frac{1}{2}s_\lambda{}^T s_\lambda \tag{63}$$

with derivative

$$\dot{V}_\lambda = s_\lambda{}^T \dot{s}_\lambda = s_\lambda{}^T \left(-\dot{\lambda}_d + \mathbf{K_1} e_\lambda + \hat{\mathbf{g}} + \Delta_g + \mathbf{u} \right) \tag{64}$$

substituting (62) in (64), it is obtained:

$$\dot{V}_\lambda = s_\lambda{}^T \left(\Delta_g - \mathbf{K_2}\text{sign}(s_\lambda) \right) \leq \parallel s_\lambda \parallel \epsilon - \parallel s_\lambda \parallel \alpha \leq \parallel s_\lambda \parallel (\epsilon - \alpha) \tag{65}$$

where $\alpha = \lambda_{min}(\mathbf{K_2})$. So if α is chosen such that $\alpha > \epsilon$, it is assured that $\dot{V}_\lambda$ is negative definite. Therefore, the formation system dynamics will reach and remain in the surface $s_\lambda = 0$ in a finite time.

7. Computer Simulations Results

7.1. Simulation Setup

The performance of the proposed navigation architecture has been assessed by means of numerical simulations. To this aim, a simulation environment has been developed, The environment is composed of 3*D* landmarks, randomly distributed over the ground (See Figure 8). To execute the tests, two Quadcopters, equipped with the set of sensors required by the proposed method, are emulated to follow (flying in formation) a lead agent moving freely in the environment. For the first set of tests, only the estimations obtained from the SLAM system will be evaluated, and thus, it will assumed that a control system exists capable of maintaining a perfect flying formation of the aerial robots with respect to the lead agent. For the second set of tests, the system state estimated by the SLAM will be used as feedback to the control scheme proposed in Section 6, in order to evaluate the closed-loop performance of the system.
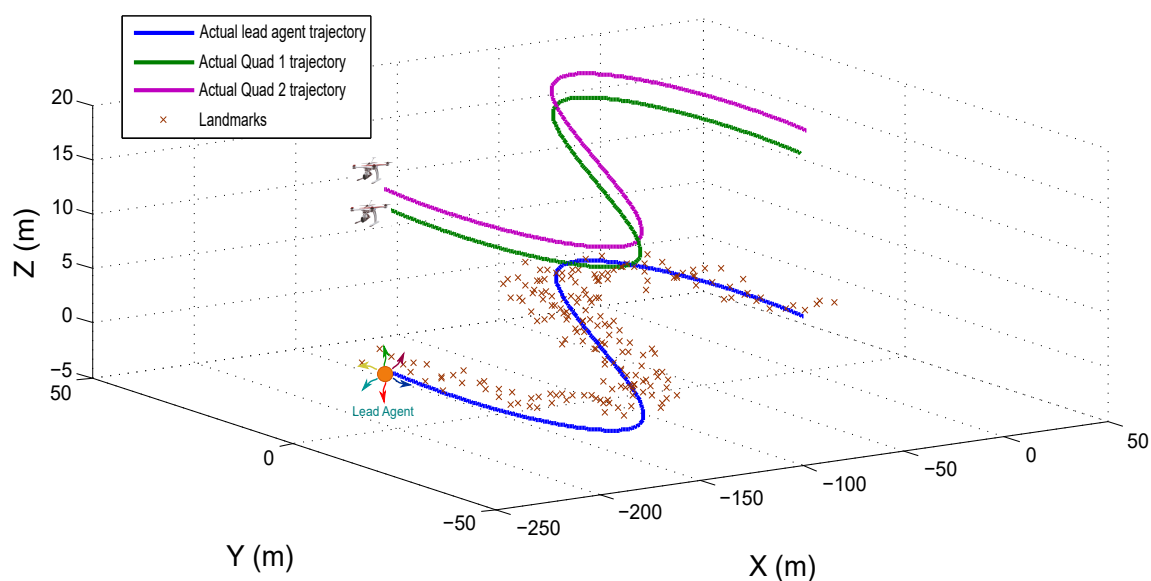

Figure 8. Simulation environment and trajectories followed by the lead agent, the Quad 1 and the Quad 2.

In computer simulations, it is assumed that the initial condition of the UAVs and the lead agent states are known with certainty. The measurements from the sensors are emulated to be 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 $c_u^j = c_v^j = 500$. In simulations, it is also assumed that there exist enough landmarks in the environment that allow a subset of them to be observed in common by the cameras of the UAVs.

To emulate the system uncertainty, the following Gaussian noise is added to measurements: Gaussian noise with $\sigma_c = 3$ pixels is added to the measurements given by the cameras. Gaussian noise with $\sigma_a = 50$ cm is added to the measurements given by the altimeter and Gaussian noise with $\sigma_r = 50$ cm is added to the measurements given by the range sensor. Gaussian noise with $\sigma_h = 0.05$ radians is added to the measurements given by the AHRS system. In the case of GPS measurements, a Gaussian noise with $\sigma_g = 1.5$ m is added. It is important to note that the noise considered for emulating GPS measurements is approximately three times bigger than the typical magnitude of the noise of a real GPS. In this case, since a local coordinate reference system is used, instead of a Geo-referenced one, then the most important source of error of the GPS that affects the local estimations is the random noise. The reason for emulating such extremely poor GPS performance is for showing that the proposed system is robust enough for dealing with very uncertain position data.

In practical applications, there are several factors that affect and degrade the performance of the Visual SLAM systems, the problem of data association is one of the main problems that appear when implementing the Visual SLAM algorithms. In addition, in cooperative visual systems, the data association problem is extended from the single-image case to the multiple-image case. Also, the gimbals used for stabilizing the cameras can be subject to small errors in its operation.

To take into account the above practical considerations, the following aspects are also considered in computer simulations: (i) outliers for the visual data association in each camera; (ii) outliers for the cooperative visual data association; (iii) perturbations in the orientation of the cameras. 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 measurement error $e_m = \sqrt{e_u^2 + e_v^2}$ pixels. In this case, e_u and e_v are the errors in the coordinate u and v (in pixels) of the projection of the landmark over the image of the camera, given by the false value in the correspondence with respect to the real value in the correct correspondence, and $e_m \in [0, 15]$ with a continuous uniform distribution. The errors in cameras orientation, due to the gimbal assumption, are emulated by perturbation of the actual orientation by means of the following sinusoidal function $e_r = 0.04 \sin(t \cdot 0.3)$ radians. Table 3 shows the number of failures introduced into the simulation due to the data association problem.

Table 3. Number of failures introduced into the simulation.

	Visual Outliers (Quad 1)	Visual Outliers (Quad 2)	Visual Outliers (Cooperative)
Multi-UAV system	9002	8400	1706
Single-UAV system	9535	-	-

7.2. SLAM Simulation Results

In the tests presented in this section, only the estimation problem is addressed, and thus, it is assumed that a control system exists capable of maintaining the flying formation of the aerial robots with respect to the lead agent.

In this case, the trajectories followed for the Quad 1, the Quad 2 and the lead agent are given by the following parametric function:

$$x_c^j(t) = x_h(t) = \frac{100\cos(t \cdot 0.015)}{1 + \sin^2(t \cdot 0.015)}$$
$$y_c^j(t) = y_h(t) = \frac{100\sin(t \cdot 0.015)\cos(t \cdot 0.015)}{1 + \sin^2(t \cdot 0.015)} \qquad (66)$$
$$z_c^j(t) = z_h(t) = 2\sin(t \cdot 0.03)$$

with the following initial conditions: $\mathbf{x_c}^1 = \begin{bmatrix} -1.5 & 0 & 15 \end{bmatrix}^T$, $\mathbf{x_c}^2 = \begin{bmatrix} 1.5 & 0 & 17 \end{bmatrix}^T$ and $\mathbf{x_h} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$. Figure 8 shows the trajectories followed by the lead agent, the Quad 1 and the Quad 2.

7.2.1. Comparative Study

A comparative study has been carried out in order to have an insight into the performance of the proposed system under its three configurations. In this case, the comparison is carried out with respect to a single-UAV SLAM system. Therefore, the study also allows to observe the advantages and drawbacks of multi-UAV systems compared with single robot systems. The computer simulation setup for the three proposed cooperative SLAM configurations, both for the case of multi-robot and single robot are described below:

1. For the first system configuration, the features initialization process is carried out as shown in Section 5, having Quad 1 visual contact with the lead agent. In the case of the single-UAV system, only one quadcopter (Quad 1) is used for obtaining measurements instead of two. In this latter case, the features initialization process is carried out as shown in [50]. In both systems, the GPS is carried by the Quad 1.

2. For the second system configuration, the simulation setup is similar to the one used for testing the first configuration, but in this case, the lead agent is maintained outside the field of view of the robots. In addition, range measurements are obtained from the Quad 1 with respect to the lead agent, and an altimeter is carried by the Quad 1.

3. For the third system configuration, the simulation setup is similar to the one used for testing the second system configuration, but in this case, no GPS measurements are considered. Instead, visual measurements of the lead agent obtained from the Quad 1 are considered.

Figure 9 shows the real and estimated trajectory obtained from the Multi-UAV system and the single-UAV system. For better comparison purposes note that the flying trajectory is divided into three stages: (i) during the time 0 to 70 s, the system is tested under its first configuration, (ii) during the time 70 to 140 s, the system is tested under its second configuration, and (iii) during the time 140 to 210 s, the system is tested in under the third configuration. Please note that, in this case, for the sake of clarity, for the case of the proposed Multi-UAV system, only the estimated trajectory of the Quad 1 is presented. The results of the estimated state of the Quad 2 are very closely similar to those presented for the Quad 1.
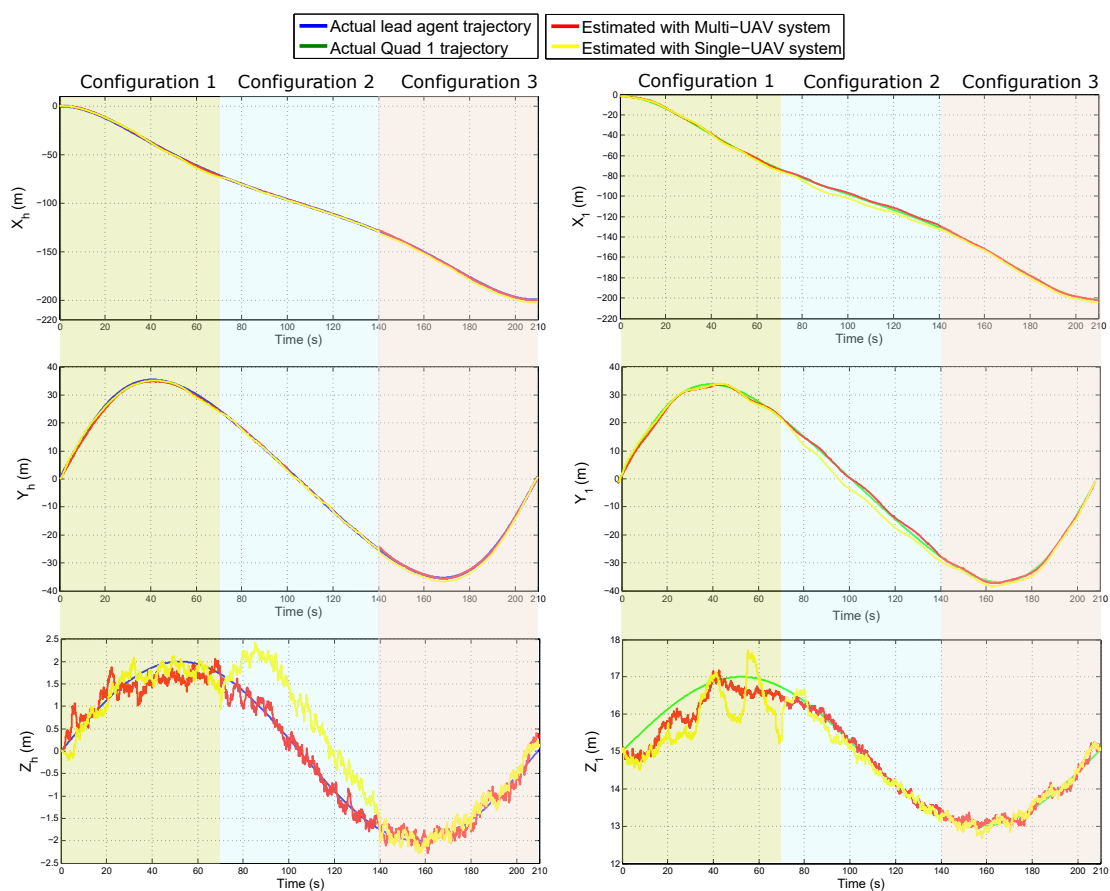

Figure 9. Estimated state of the lead agent (**left plots**) and of the Quad 1 (**right plots**), for both kind of system, under the three configurations.

Figure 10 shows the evolution over time of the absolute value of the error for the estimated trajectory of the lead agent (e^h) and the Quad 1 (e^1) with respect to the global reference system W (i.e., $e^h_x = x_h - \hat{x}_h$, $e^h_y = y_h - \hat{y}_h$, $e^h_z = z_h - \hat{z}_h$, $e^1_x = x^1_c - \hat{x}^1_c$, $e^1_y = y^1_c - \hat{y}^1_c$ and $e^1_z = z^1_c - \hat{z}^1_c$.
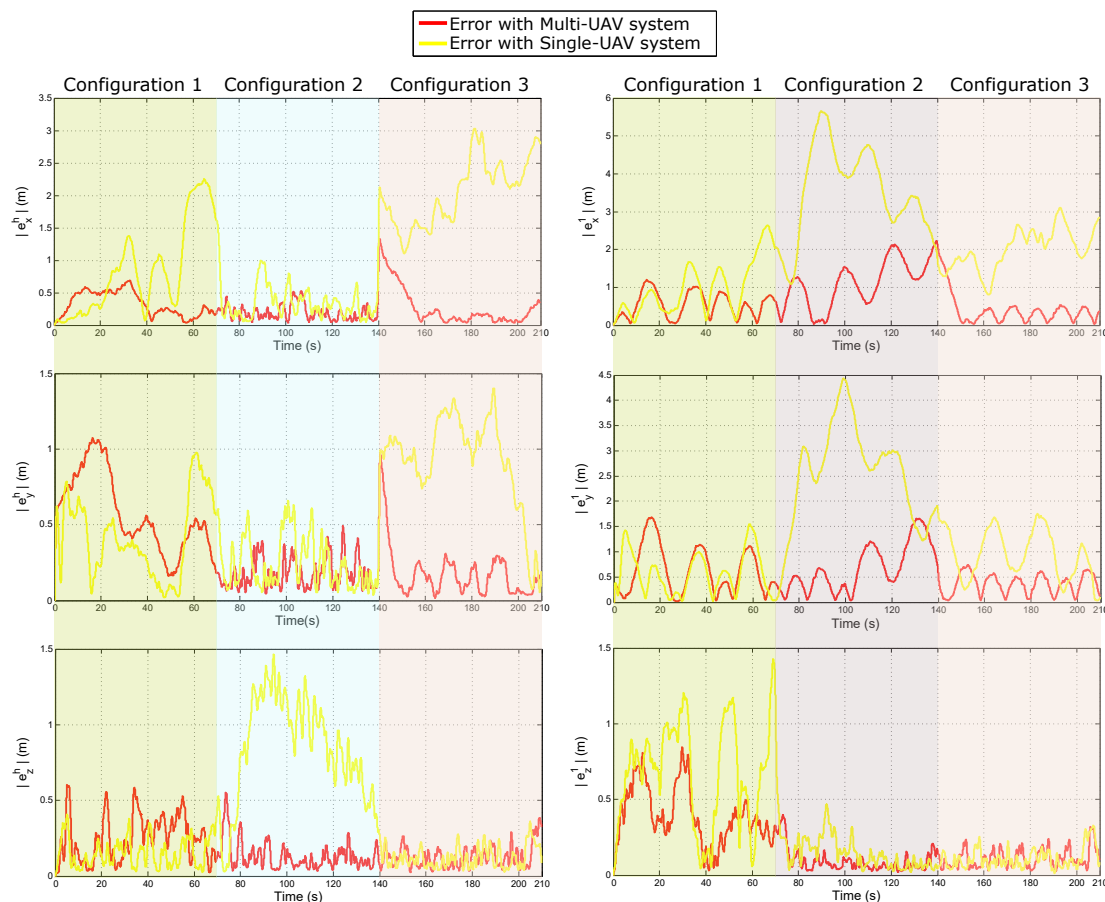

Figure 10. Comparison of the quality of the estimated trajectory obtained with the Multi-UAV system respect to the trajectory obtained with the Single-UAV system. **Left column** of plots illustrates the errors obtained for the lead agent. **Right column** of plots illustrates the errors obtained for the Quad 1.

Table 4 summarizes the Mean Squared Error (MSE) for the estimated position in the three axes of the lead agent and of the Quad 1. In this case, Table 4 shows the MSE obtained during each system configuration and the MSE obtained for the whole trajectory. According to the above results, it can be seen that the Multi-UAV system presents a better performance. Table 5 provides an insight into the performance of the proposed method for estimating the location of the 3D landmarks composing the environment. In this case, the total (sum of all) of the Mean Squared Errors for the estimated position of the landmarks is presented. Also, the total of the Mean Squared Errors for the initial estimated position of the landmarks is presented. Please note that the results are presented for each coordinate of the reference frame W. Please note that Tables 4 and 5 show the errors obtained during each system configuration as well as the errors obtained for the whole trajectory. The results show that the proposed Multi-UAV method has a much better performance than the single-UAV system, regarding the error obtained in the estimation of landmarks position.

Table 4. Mean Squared Error for the estimated position of the lead agent ($MSEX_h$, $MSEY_h$, $MSEZ_h$) and the estimated position of Quad 1 ($MSEX_1$, $MSEY_1$, $MSEZ_1$).

		$MSEX_h(m)$	$MSEY_h(m)$	$MSEZ_h(m)$	$MSEX_1(m)$	$MSEY_1(m)$	$MSEZ_1(m)$
Config. 1	**Multi-UAV system**	0.1454	0.3943	0.0873	0.4063	0.6115	0.1923
	Single-UAV system	1.1873	0.2456	0.0362	1.4233	0.5061	0.5849
Config. 2	**Multi-UAV system**	0.0538	0.0457	0.0292	1.5706	0.6709	0.0163
	Single-UAV system	0.2047	0.0935	0.8267	13.8187	7.3878	0.0409
Config. 3	**Multi-UAV system**	0.1670	0.0627	0.0241	0.3093	0.1718	0.0180
	Single-UAV system	4.4490	0.9560	0.0231	4.3048	1.1514	0.0184
Total	**Multi-UAV system**	0.1221	0.1676	0.0468	0.7621	0.4847	0.0755
	Single-UAV system	1.9472	0.4317	0.2953	6.5140	3.0151	0.2148

Table 5. Total Mean Squared Error in the position estimation of the landmarks ($MSEX_a$, $MSEY_a$, $MSEZ_a$). Total Mean Squared Error in the initial position estimation of the landmarks ($MSEX_{ai}$, $MSEY_{ai}$, $MSEZ_{ai}$).

		$MSEX_a(m)$	$MSEY_a(m)$	$MSEZ_a(m)$	$MSEX_{ai}(m)$	$MSEY_{ai}(m)$	$MSEZ_{ai}(m)$
Config. 1	**Multi-UAV system**	0.2794	0.4699	1.1571	0.9293	1.1608	4.3764
	Single-UAV system	2.5504	2.1782	18.5759	10.1012	7.9064	57.0486
Config. 2	**Multi-UAV system**	0.7379	0.4515	1.5301	2.0737	1.6544	8.8474
	Single-UAV system	1.2181	2.5109	13.2295	5.0821	9.0028	47.9993
Config. 3	**Multi-UAV system**	0.4126	0.4884	0.8722	0.8332	0.8006	1.5070
	Single-UAV system	1.5092	1.2707	2.6635	3.2802	3.7381	16.2915
Total	**Multi-UAV system**	0.4702	0.4683	1.2120	1.2452	1.2080	4.9143
	Single-UAV system	1.8385	2.0462	12.6057	6.5731	6.9910	42.2073

To evaluate the statistical consistency of each method, the average NEES (Normalized Estimation Error Squared [51]) over n_3 Monte Carlo runs was computed, as it is proposed in [52]. The NEES is estimated as follows:

$$\boldsymbol{\epsilon}_k = \left[\ \mathbf{x}_k - \hat{\mathbf{x}}_k\ \right]^T \mathbf{P}_k^{-1} \left[\ \mathbf{x}_k - \hat{\mathbf{x}}_k\ \right] \tag{67}$$

In addition, the average NEES is computed from:

$$\bar{\boldsymbol{\epsilon}}_k = \frac{1}{n_3} \sum_{r=1}^{n_3} \boldsymbol{\epsilon}_k^r \tag{68}$$

Figure 11 shows the average NEES over 50 Monte Carlo runs obtained for each system (multi-UAV and Single-UAV). The average NEES is calculated taking into account the eighteen variables that define the complete state of the UAVs and the lead agent (position and linear velocity). It is very interesting to note how the consistency of the filter considerably degenerates in the case of the single-robot system under its third configuration, this effect happens when the system is not fully observable. On the other hand, for the multi-UAV system case, the consistency of the filter remains practically stable. Given the previous results, the proposed Multi-UAV system presents a good performance in its three configurations, despite all the failures and disturbances introduced into the system. The above study provides a good insight into the robustness of the proposed system.
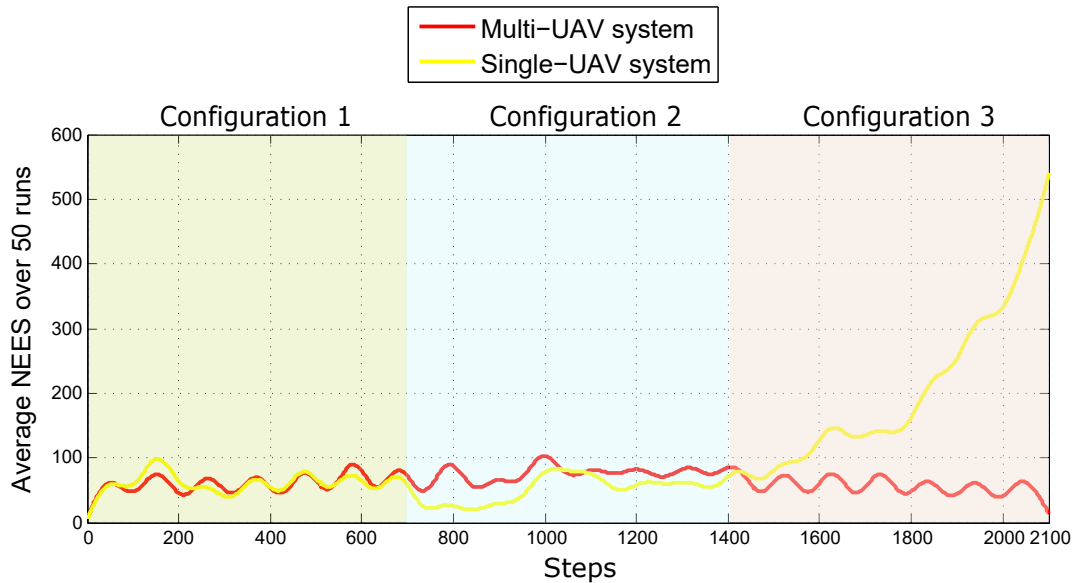
Figure 11. Average NEES obtained with both systems.

7.2.2. GPS-Denied Test

A set of simulations tests were carried out with the intention of evaluating the performance of each system configuration under GPS-denied environments. For this test, the flying trajectory is divided into two stages: (i) during time period 0 to 105 s, the vehicles move with GPS availability, (ii) during time period 105 to 210 s, the vehicles move with without GPS availability. Throughout the trajectory, the state of the system is estimated independently using each system configuration.

The following aspects are considered in this test: For configuration 1, altimeter measurements are also included, since, in practice, almost every UAV is equipped with this kind of sensor. For the configuration 3, during the first stage of the trajectory, GPS measurements are also included. These previous considerations will allow evaluating the performance of configuration 3 both in GPS-available environments and in GPS-denied environments.

Figure 12 shows for each configuration the evolution of the error over the time for the estimated position of the Quad 1 and the lead agent too, (i.e., $e_r^h = \sqrt{(x_h - \hat{x}_h)^2 + (y_h - \hat{y}_h)^2 + (z_h - \hat{z}_h)^2}$ and $e_r^1 = \sqrt{(x_c^1 - \hat{x}_c^1)^2 + (y_c^1 - \hat{y}_c^1)^2 + (z_c^1 - \hat{z}_c^1)^2}$). Given the above results, it can be observed that the three configurations yield good results in GPS-available environments, and it is also confirmed that configuration 3 provides good results in GPS-denied environments. However, the performance of the configuration 1 and 2 clearly degenerates when there is no GPS availability. Although configuration 1 uses at least two monocular measurements of the lead agent allowing to obtain three-dimensional information from the lead agent, is interesting to note that its performance is not so good in GPS-denied environments as expected. On the other hand, the better three-dimensional information is obtained by the sensory fusion of a monocular measurement and a range measurement with respect to the lead agent (as in configuration 3). The above results are produced by the fact that the 3D reconstruction from monocular measurements is more sensitive to noise.
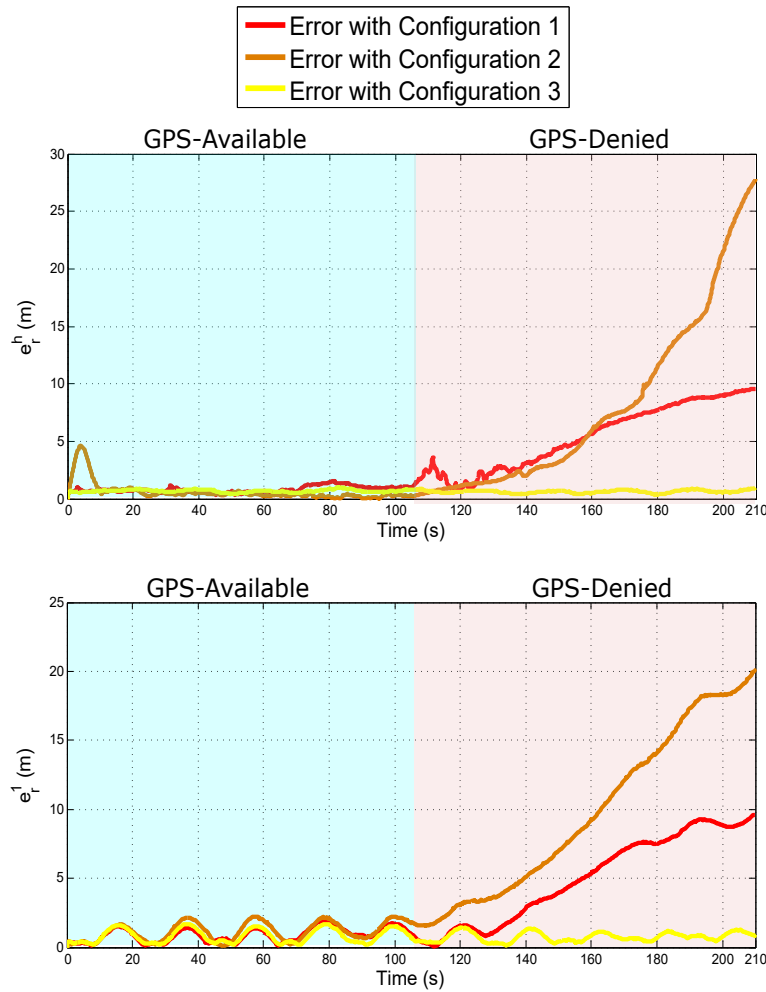
Figure 12. Comparison of the quality in the estimation of the position of the lead agent (**upper plot**) and the Quad 1 (**bottom plot**) under GPS-denied environments.

7.3. Control Simulations Results

A set of simulations was also carried out, where the SLAM estimates are used as feedback to the control laws involved in commanding the formation of the UAVs with respect to the lead agent (see Section 6). This will allow observing the performance of both the estimation and control in a closed-loop manner.

Based on the same simulation setup used previously, now, the trajectory followed by the UAVs is given by a kinematic model with control inputs. The initial conditions of the position of the quadcopters are the same used previously. In the case of the yaw angle of the aerial vehicles, the initial conditions are as follows: $\psi^1 = 1$ and $\psi^2 = \frac{1}{2}$. The vector λ_d, that defines the desired values for the flight formation is: $\lambda_d^1 = \begin{bmatrix} 2.5, & 1, & 14, & \arctan(\frac{\hat{y}_h}{\hat{x}_h}) \end{bmatrix}^T$ and $\lambda_d^2 = \begin{bmatrix} 0.5, & -1, & 16, & \arctan(\frac{\hat{y}_h}{\hat{x}_h}) \end{bmatrix}^T$.

In this test, the trajectory is divided into three stages, each one for testing each system configuration. Figure 13 shows the trajectories of the three elements composing the flight formation.

Figure 14 shows the evolution of the error with respect to the desired values λ_d which define the flight formation.
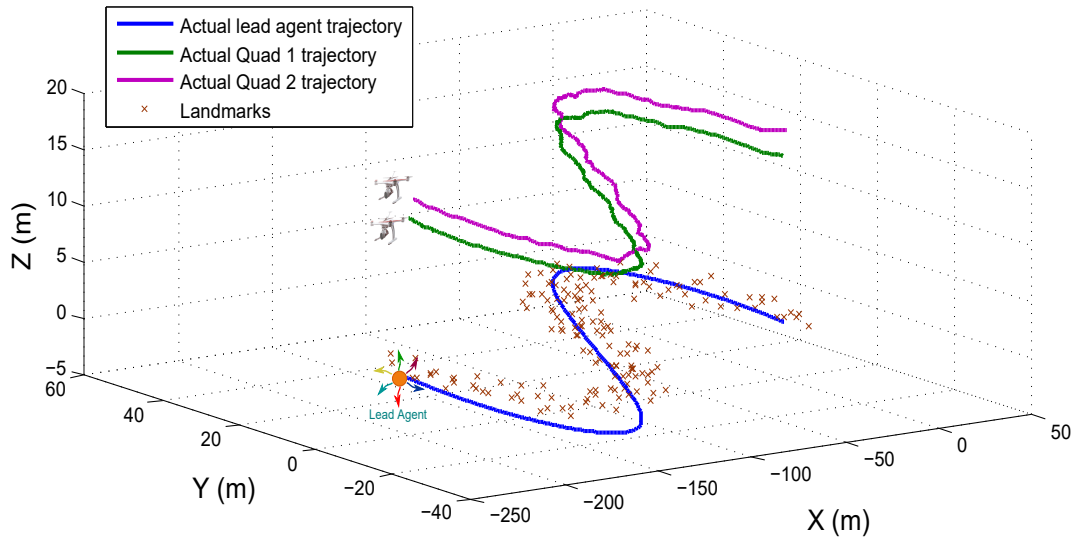
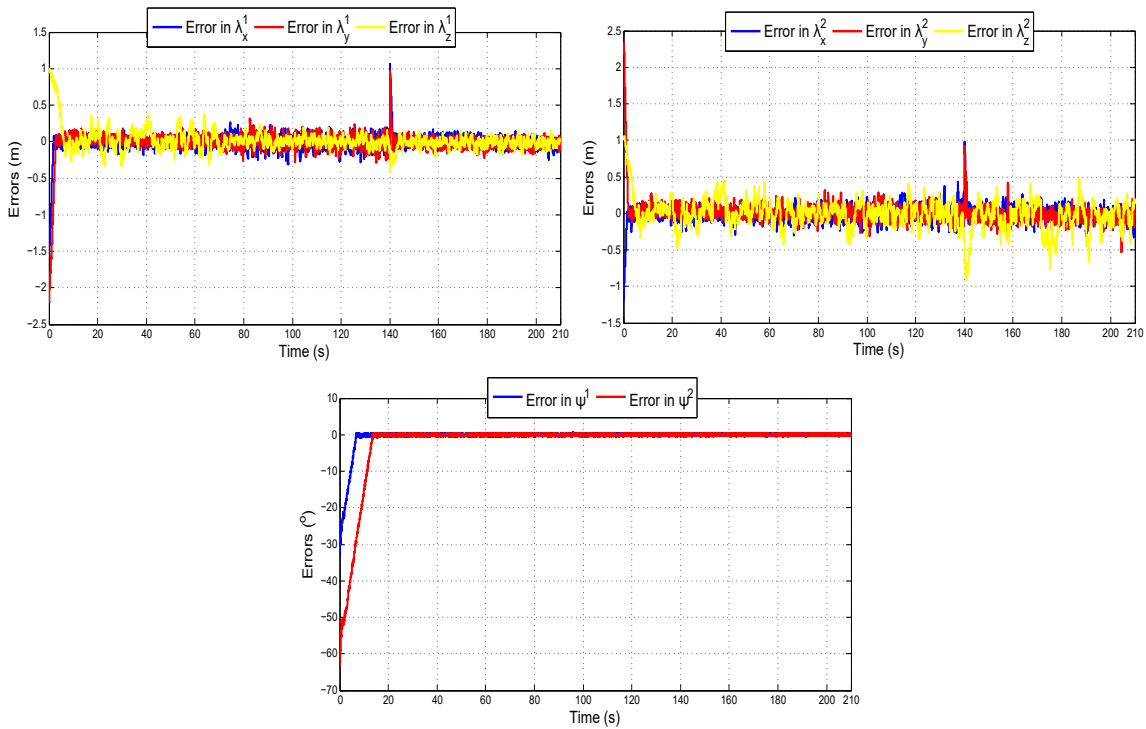**Figure 13.** Trajectories followed by the lead agent, the Quad 1 and the Quad 2 with the system in closed-loop.

Figure 14. Errors in λ_d during the flight trajectories.

Figure 15 shows, for each configuration, the evolution of the error over the time for the estimated position of the Quad 1 as well as the lead agent, when these estimates are used as feedback to the control laws.

Given the previous results, it can be observed that the estimation with the three configurations and the control have good performance in closed-loop.
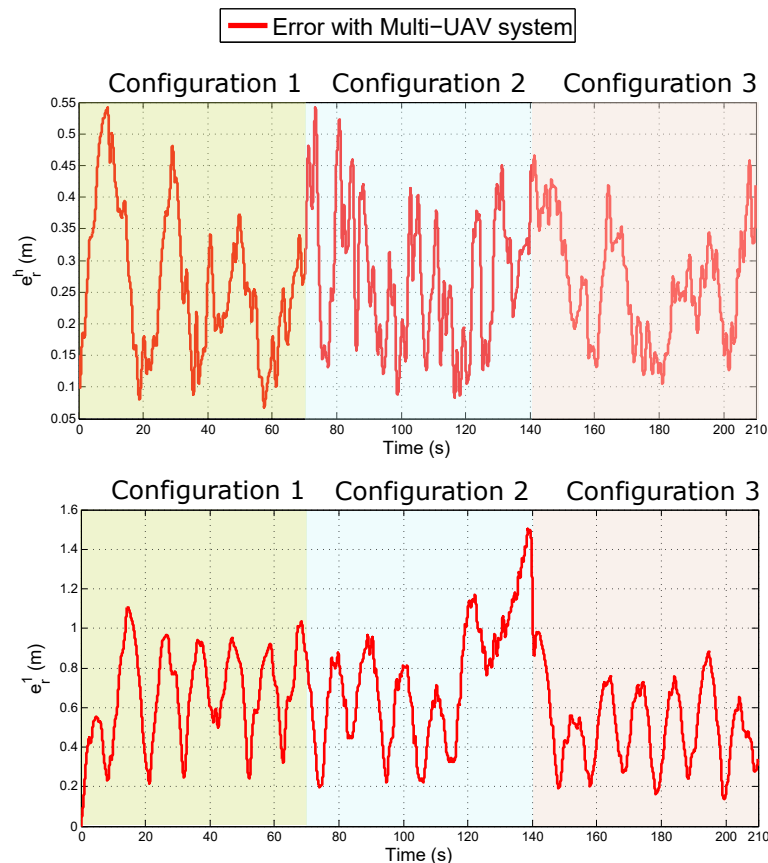
Figure 15. Comparison of the quality of the estimated position of the lead agent (**upper plot**) and the Quad 1 (**bottom plot**) in closed-loop.

8. Conclusions

In this work, the problem of the cooperative visual-based SLAM for the class of multi-UAV systems that integrate a lead agent has been addressed. For this purpose, three system configurations were proposed and investigated by means of an intensive nonlinear observability analysis. The main differences between the system configurations have to do with the set of sensors used in each case, additional to the monocular cameras carried by each aerial vehicle, as well as the circumstances they operate in a suitable manner. The objective of each configuration is to ensure that the system can be able to maximize the property of observability under different scenarios. In this case, several theoretical results were obtained from the analysis; for instance, sufficient conditions required for obtaining the observability results were established.

The first or second configuration can be considered for any scenario where the GPS measurements (even low quality measurements) are available. In particular, the second configuration will be useful in circumstances where the lead agent is not seen by cameras carried by the UAVs. The third configuration will be more useful in any scenario or circumstance where there is no GPS availability. The proposed structure based on different system configurations facilitates the mathematical and experimental analysis. However, from the implementation point of view, it will be more convenient to use an integrated system which includes all the sensory inputs whenever they are available.

The proposed system configurations were implemented using the standard EKF-based SLAM methodology. In this case, the initialization process of new map features is carried out in a cooperative way. In addition, a high-level control scheme was proposed allowing the control of the UAVs formation with respect to the lead agent.

An extensive set of computer simulations was performed in order to validate the proposed scheme, for instance: (i) the proposed multi-UAV system was compared against a single-UAV system, (ii) the

consistency of the filter was investigated by means of the average NEES test, (iii) the performance of each system configuration were evaluated under GPS-denied environments, and (iv) the performance of both the estimation and the control were evaluated by analyzing the behavior of each system configuration in closed-loop.

In the above cases, several aspects regarding applicability in real scenarios of the proposed approach were considered, for instance, the association problem of visual data for each camera and the pseudo-stereo matching as well, and also the errors induced by the assumption of the camera gimbal were emulated in simulations.

Based on the results of the simulations, it can be observed how the proposed method improves the estimation of the state by considering the multi-robot cooperative scheme. Also, it is shown that the estimates produced by the proposed SLAM system can be used directly as feedback to the high-level control laws that command the flight formation. The simulation results also show that the proposed scheme is able to offer a good performance, even in the absence of GPS measurements.

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 could be focused on developing experiments with real data in order to fully evaluate the applicability of the proposed approach.

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 for Integration of Perception with Control and Navigation for Resource-Limited, Highly Dynamic, Autonomous System, Sydney, Australia, 9–13 June 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. [CrossRef]
3. 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. [CrossRef]
4. 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 International Convention Center (TICC), Taipei, Taiwan, 7–10 April 2014; pp. 1–5.
5. Schmidt, A. *Computer Vision and Graphics: International Conference, ICCVG 2014, Warsaw, Poland, September 15–17 2014. Proceedings*; Springer International Publishing: Cham, Switzerland, 2014; Chapter The EKF-Based Visual SLAM System with Relative Map Orientation Measurements, pp. 570–577.
6. Meilland, M.; Comport, A. 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 Big Sight, Tokyo, Japan, 3–7 November 2013; pp. 3677–3683.
7. Lu, Y.; Song, D. Visual navigation using heterogeneous landmarks and unsupervised geometric constraints. *IEEE Trans. Robot.* **2015**, *31*, 736–749. [CrossRef]
8. 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. [CrossRef]
9. Weiss, S.; Scaramuzza, D.; Siegwart, R. Monocular SLAM based navigation for autonomous micro helicopters GPS-denied environments. *J. Field Robot.* **2011**, *28*, 854–874. [CrossRef]
10. Munguía, R.; Urzua, S.; Bolea, Y.; Grau, A. Vision-Based SLAM System for Unmanned Aerial Vehicles. *Sensors* **2016**, *16*, 372. [CrossRef] [PubMed]

11. Yang, T.; Li, P.; Zhang, H.; Li, J.; Li, Z. Monocular Vision SLAM-Based UAV Autonomous Landing in Emergencies and Unknown Environments. *Electronics* **2018**, *7*, 73. [CrossRef]

12. Zhang, Z.; Zhao, R.; Liu, E.; Yan, K.; Ma, Y. Scale Estimation and Correction of the Monocular Simultaneous Localization and Mapping (SLAM) Based on Fusion of 1D Laser Range Finder and Vision Data. *Sensors* **2018**, *18*, 1948. [CrossRef] [PubMed]

13. Annaiyan, A.; Olivares, M.; Voos, H. Real-time graph-based SLAM in unknown environments using a small UAV. In Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), Miami, FL, USA, 13–16 June 2017.

14. 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. [CrossRef]

15. Chowdhary, G.; Johnson, E.N.; Magree, D.; Wu, A. GPS-denied Indoor and Outdoor Monocular Vision Aided Navigation and Control of Unmanned Aircraft. *J. Field Robot.* **2013**, *30*, 415–438. [CrossRef]

16. 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 Marriott Tech Center, Denver, CO, USA, 9–12 June 2015.

17. 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. [CrossRef]

18. 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. [CrossRef]

19. Pappas, H.; Tanner, G.; Kumar, V. Leader-to-formation stability. *IEEE Trans. Robot. Autom.* **2004**, *20*, 443–455.

20. Zhu, Z.; Roumeliotis, S.; Hesch, J.; Park, H.; Venable, D. Architecture for Asymmetric Collaborative Navigation. In Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium, Myrtle Beach Marriott, Myrtle Beach, SC, USA, 23–26 April 2012.

21. 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, Massachusetts Institute of Technology, Cambridge, MA, USA, 8–11 June 2005.

22. 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, 10–17 May 2002.

23. 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. AIAA Information Systems-AIAA Infotech @ Aerospace, AIAA SciTech Forum, (AIAA 2017-0879). Available online: https://doi.org/10.2514/6.2017-0879 (accessed on 3 December 2018).

24. 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, Key Bridge Marriott, Arlington, VA, USA, 7–10 June 2016.

25. Guerra, E.; Munguia, R.; Grau, A. Monocular SLAM for Autonomous Robots with Enhanced Features Initialization. *Sensors* **2014**, *14*, 6317–6337. [CrossRef] [PubMed]

26. 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.

27. Minaeian, S.; Liu, J.; Son, Y.J. Vision-based target detection and localization via a team of cooperative UAV and UGVs. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *46*, 1005–1016. [CrossRef]

28. Opromolla, R.; Fasano, G.; Accardo, D. A Vision-Based Approach to UAV Detection and Tracking in Cooperative Applications. *Sensors* **2018**, *18*, 3391. [CrossRef] [PubMed]

29. Tang, Y.; Hu, Y.; Cui, J.; Liao, F.; Lao, M.; Lin, F.; Teo, R.S. Vision-Aided Multi-UAV Autonomous Flocking in GPS-Denied Environment. *IEEE Trans. Ind. Electron.* **2019**, *66*, 616–626. [CrossRef]

30. Bethke B.; Valenti M.; How, J. Cooperative Vision Based Estimation and Tracking Using Multiple UAVs. In *Advances in Cooperative Control and Optimization*; Pardalos, P.M., Murphey, R., Grundel, D., Hirsch, M.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 369.

31. Pirshayan, A.; Seyedarabi, H.; Haghipour, S. Cooperative Machine-Vision-Based Tracking using Multiple Unmanned Aerial Vehicles. *Adv. Comput. Sci.* **2014**, *3*, 118–123.

32. Amaral, G.; Silva, H.; Lopes, F.; Ribeiro, J.P.; Freitas, S.; Almeida, C. UAV cooperative perception for target detection and tracking in maritime environment. In Proceedings of the OCEANS 2017, Aberdeen Exhibition & Conference Centre—BHGE Arena, Aberdeen, UK, 19–22 June 2017; pp. 1–6.

33. Reif, K.; Günther, S.; Yaz, E.; Unbehauen, R. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Trans. Autom. Control* **1999**, *44*, 714–728. [CrossRef]

34. Kluge, S.; Reif, K.; Brokate, M. Stochastic stability of the extended Kalman filter with intermittent observations. *IEEE Trans. Autom. Control* **2010**, *55*, 514–518. [CrossRef]

35. Batzdorfer, S.; Bestmann, U.; Becker, M.; Schwithal, A. 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, HI, USA, 23–25 April 2013.

36. 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, Nice, France, 22–26 September 2008; pp. 340–345.

37. Munguia, R.; Grau, A. A Practical Method for Implementing an Attitude and Heading Reference System. *Int. J. Adv. Robot. Syst.* **2014**. [CrossRef]

38. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, MA, USA, 2003.

39. Srisamosorn, V.; Kuwahara, N.; Yamashita, A.; Ogata, T. Human-Tracking System Using Quadrotors and Multiple Environmental Cameras for Face-Tracking Application. *Int. J. Adv. Robot. Syst.* **2017**. [CrossRef]

40. Benezeth, Y.; Emile, B.; Laurent, H.; Rosenberger, C. Vision-Based System for Human Detection and Tracking in Indoor Environment. *Int. J. Soc. Robot.* **2010**, *2*, 41–52. [CrossRef]

41. Olivares-Mendez, M.A.; Fu, C.; Ludivig, P.; Bissyandé, T.F.; Kannan, S.; Zurad, M.; Annaiyan, A.; Voos, H.; Campoy, P. Towards an Autonomous Vision-Based Unmanned Aerial System against Wildlife Poachers. *Sensors* **2015**, *15*, 31362–31391. [CrossRef] [PubMed]

42. Briese, C.; Seel, A.; Andert, F. Vision-based detection of non-cooperative UAVs using frame differencing and temporal filter. In Proceedings of the International Conference on Unmanned Aircraft Systems, Pearl Street Dallas, TX, USA, 12–15 June 2018.

43. Mejías L.; McNamara, S.; Lai, J. Vision-based detection and tracking of aerial targets for UAV collision avoidance. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 87–92.

44. Alavi, B.; Pahlavan, K. Modeling of the TOA-based distance measurement error using UWB indoor radio measurements. *IEEE Commun. Lett.* **2006**, *10*, 275–277. [CrossRef]

45. Lanzisera, S.; Zats, D.; Pister, K.S.J. Radio Frequency Time-of-Flight Distance Measurement for Low-Cost Wireless Sensor Localization. *IEEE Sens. J.* **2011**, *11*, 837–845. [CrossRef]

46. Hermann, R.; Krener, A. Nonlinear controllability and observability. *IEEE Trans. Autom. Control* **1977**, *22*, 728–740. [CrossRef]

47. Slotine, J.E.; Li, W. *Applied Nonlinear Control*; Prentice-Hall Englewood Cliffs: Upper Saddle River, NJ, USA, 1991.

48. Guilietti, F.; Pollini, L.; Innocenti, M. Autonomous formation flight. *IEEE Control Syst. Mag.* **2000**, *20*, 30–44.

49. Utkin, V.I. Sliding Mode Control Design Principles and Applications to Electric Drives. *IEEE Trans. Ind. Electron.* **1993**, *40*, 23–36. [CrossRef]

50. 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, 16–19 August, 2006.

51. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation*; John Wiley and Sons: Hoboken, NJ, USA, 2001.

52. Bailey, T.; Nieto, J.; Guivant, J.; Stevens, M.; Nebot, E. Consistency of the ekf-slam algorithm. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Beijing, China, 4–6 December 2006.