

Article

Cooperative Optimization of UAVs Formation Visual Tracking

Nicola Lissandrini * , Giulia Michieletto , Riccardo Antonello , Marta Galvan, Alberto Franco and Angelo Cenedese 

Department of Information Engineering, University of Padova, 35131 Padova, Italy

* Correspondence: nicola.lissandrini@studenti.unipd.it

Received: 23 May 2019; Accepted: 5 July 2019; Published: 7 July 2019



Abstract: The use of unmanned vehicles to perform tiring, hazardous, repetitive tasks, is becoming a reality out of the academy laboratories, getting more and more interest for several application fields from the industrial, to the civil, to the military contexts. In particular, these technologies appear quite promising when they employ several low-cost resource-constrained vehicles leveraging their coordination to perform complex tasks with efficiency, flexibility, and adaptation that are superior to those of a single agent (even if more instrumented). In this work, we study one of said applications, namely the visual tracking of an evader (target) by means of a fleet of autonomous aerial vehicles, with the specific aim of focusing on the target so as to perform an accurate position estimation while concurrently allowing a wide coverage over the monitored area so as to limit the probability of losing the target itself. These clearly conflicting objectives call for an optimization approach that is here developed: by considering both aforementioned aspects and the cooperative capabilities of the fleet, the designed algorithm allows controlling in real time the single fields of view so as to counteract evasion maneuvers and maximize an overall performance index. The proposed strategy is discussed and finally assessed through the realistic Gazebo-ROS simulation framework.

Keywords: UAVs; visual tracking; coverage; multi-agent formation; optimization

1. Introduction

Presently, unmanned aerial vehicles (UAVs) constitute an attractive research topic in robotics due to the high potential opportunities offered by these autonomous vehicles arising as a key technology in urban, rural, manufacturing and military contexts [1–6]. Quadrotors, in particular, represent the most exploited UAV platforms and their current applications range from the classical visual sensing tasks (e.g., surveillance and aerial photography [7]) to the modern environment exploration and physical interaction (e.g., search and rescue operations [8,9], grasping and manipulation tasks [10,11]). Besides their high popularity, standard quadrotors are challenged by their highly non-linear, strongly coupled and under-actuated dynamics that implies several limitations regarding both the executable maneuvers set and the efficiently achievable tasks. To overcome these drawbacks, the focus of aerial robotics community is currently moving toward cooperative approaches where the task execution is needed for a formation of UAV platforms, rather than a single improved vehicle [12–15]. According to the multi-agent paradigm, the idea is indeed to combine the limited single vehicle capabilities in order to efficiently solve complex tasks, as, for instance, multiple targets localization over a vast area [16], cumbersome payload transportation [17], and communication relay chain establishment [18], to cite a few.

In this spirit, hereafter we account for a quadrotors formation required to track a ground robot freely moving in the environment. Within this context, the advantage deriving from the employment of a multi-agent system (rather than a single, although large and complex, UAV) mainly relies on the presence of multiple visual sensors, i.e., on-board calibrated cameras. This, indeed, may results in

increasing the monitored area and thus in decreasing the probability to lose the target. In general, the use of UAVs swarm to achieve surveillance task is supported by the fact that a collaborative approach, resting upon the use of multiple sensors, actuators and other resources/capabilities simultaneously, may significantly increase the performance of individual agent, as well as, of the overall group, and improve the robustness of the swarm operation with respect to external disturbances, single failures, or malicious attacks.

Related works—Previous considerations motivate the wide existing literature related to UAVs cooperative surveillance issues, involving monitoring [19], events detection [20] and target tracking [21]. In all the aforementioned cases, the involved aerial vehicles are typically supposed to be equipped with calibrated cameras so that one main and crucial issue in the surveillance task can be found in the optimization of the monitored area (see, e.g., [22–24] and the references therein). In this context, current state-of-art techniques can be divided in coverage and object detection and tracking strategies. In the first case, the idea is to optimize the monitor area accounting for the space geometry [25] and popular approaches rest upon Voronoi area partitioning [26,27] and/or SLAM (Simultaneous Localization And Mapping) solutions [28,29]. On the other hand, object detection and tracking solutions discriminate according to the image data processes [30]: the visual information deriving from all the cameras in the multi-agent system can be combined to detect the target [31], or each visual sensor can perform the object detection resting only on its information and then sharing the estimate with the other group components [32]. However exploited in the application perspective, it clearly appears that the advantages deriving from the coordination among UAVs and the combination of multiple views are well established [33,34].

Contributions—In this paper, we focus on the target tracking problem accounting for a formation of quadrotors equipped with calibrated cameras. In particular, we assume that the group is able to move by maintaining a certain configuration (a similar work about UAVs formation preservation while tracking a target is addressed in [35]) and we propose an optimization procedure for the vehicles orientation. In doing so, we decouple the problem of formation control and maintenance (not considered here) from that of visual tracking optimization, to better focus on this latter, towards the definition of a distributed solution that may constitute the fast inner control loop of a regulation strategy that includes the formation objectives on a slower timescale. The original aspect of our solution is constituted by the trade-off between the maximization of the monitored area and the maximization of the cameras overlapping portion, which implies the increase of the consistency and the robustness of the retrieved information at the cost of a more limited monitored area. In actual facts, the two opposed goals are motivated on the one side by the minimization of the probability of losing the target while facing the ability of the target to perform evasion maneuvers, and on the other by the minimization of the target position estimate error, namely an optimal target detection capability of the UAVs formation. In order to lead to a practically treatable optimization strategy, the proposed approach adopts a pin-hole camera model to derive an approximation of the projection of the image plane onto the plane where the target is supposed to move. This last one, in particular, is suitably sampled so that each camera view can be numerically optimized by accounting for the problem geometry through the definition of proper matrices that describe the visible area.

Paper structure—The rest of the paper is organized as follows. Section 2 is devoted to the description of the considered scenario, focusing first on the single quadrotor dynamics and on the graph-based formation representation, and then to the formalization of the cooperative tracking problem. A strategy to solve this task is illustrated in the main Section 4: this consists of the determination of the optimal trade-off between the minimization of the target position estimation error and the target loss probability accounting for the approximation of the monitored area derived in Section 3. In Section 5 the validity of the proposed method is validated and assessed through the results achieved in the Gazebo simulative environment. The main conclusions and some future research directions are figured out in Section 6.

2. Scenario Description

The considered scenario—depicted in Figure 1—involves a group of $n \geq 3$ quadrotors with the task of tracking a mobile ground robot, referred as *target*. To achieve this goal, each UAV composing the formation is assumed to be equipped with a (calibrated) camera, a position sensor and a communication interface with its siblings. In the rest of the section, we formalize the tracking problem (Section 2.3) through the introduction of suitable models for the quadrotor dynamics (Section 2.1) and for the whole formation (Section 2.2).

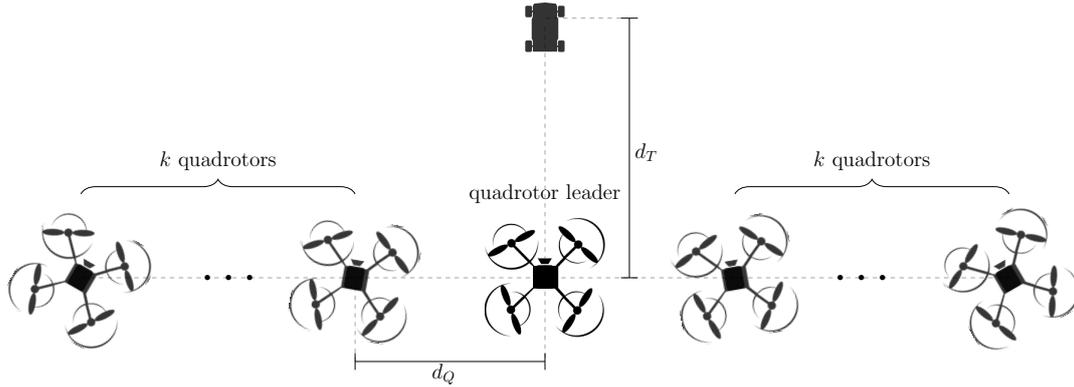


Figure 1. Considered scenario (top view): a formation of $n = 2k + 1$, $k \in \mathbb{N}$, quadrotors is required to track a ground robot. The UAVs group is supposed to move rigidly on the xy -plane of the world frame \mathcal{F}_W preserving its shape, maintaining a fixed distance d_T from the target and ensuring that the quadrotor leader is always pointing at the target.

2.1. Quadrotor Model

A quadrotor is an under-actuated flying system with four rotating propellers whose spinning rate can be regulated in order to control the six degrees of freedom (dofs) of the platform. Its dynamics can be described through the Euler-Newton model for rigid-body motion as explained in the following.

First, we introduce two reference frames: the inertial *world frame* $\mathcal{F}_W = \{O_W, (x_W, y_W, z_W)\}$, whose axes direction is identified by the (unit) vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ of the canonical basis of \mathbb{R}^3 and the *body frame* $\mathcal{F}_{B_i} = \{O_{B_i}, (x_{B_i}, y_{B_i}, z_{B_i})\}$ attached to the i -th quadrotor, $i \in \{1 \dots n\}$, so that its origin O_{B_i} coincides with the vehicle center of mass (c.o.m.). The position of O_{B_i} in the world frame is identified by the vector $\mathbf{p}_{O_{B_i}}^W \in \mathbb{R}^3$, while the orientation of \mathcal{F}_{B_i} regarding \mathcal{F}_W is described by the rotation matrix ${}^W\mathbf{R}_{B_i}$ belonging to the three-dimensional Special Orthogonal space $SO(3)$. In particular, we assume that the matrix ${}^W\mathbf{R}_{B_i} \in SO(3)$ derives from the composition of three consecutive rotations around the axes of the body frame, according to a given sequence. Hence, it follows that ${}^W\mathbf{R}_{B_i} = \mathbf{R}(\phi_i, \theta_i, \psi_i)$ where $\phi_i \in (-\frac{\pi}{2}, \frac{\pi}{2}]$, $\theta_i \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\psi_i \in (-\pi, \pi]$ are the *roll*, *pitch* and *yaw* angles identifying respectively a rotation around the x_{B_i} , y_{B_i} and z_{B_i} -axis of \mathcal{F}_{B_i} . Hereafter, we assume that only the yaw angle ψ_i and the (whole) position $\mathbf{p}_{O_{B_i}}^W$ are directly and independently controllable through the assignment of a certain *control moment* $\tau_i^W \in \mathbb{R}$ and a *control force* $\mathbf{f}_i^W \in \mathbb{R}^3$, respectively, both depending on the propellers spinning rate. Given these premises and introducing the time dependence, the equations governing the i -th quadrotor dynamics, $i \in \{1 \dots n\}$, are

$$m\ddot{\mathbf{p}}_{O_{B_i}}^W(t) = -mg\mathbf{e}_3 + \mathbf{f}_i^W(t), \tag{1}$$

$$J_z\dot{\psi}_i(t) = \tau_i^W(t), \tag{2}$$

where $g \in \mathbb{R}^+$ is the gravity constant, whereas $m \in \mathbb{R}^+$ and $J_z \in \mathbb{R}^+$ denote the mass and the coefficient of inertia around z_{B_i} -axis of the body frame which are supposed to be the same for all the UAVs in the formation.

From now on, we assume that each quadrotor is able to track both reference position and yaw angle with negligible error by implementing an effective control law (e.g., based on PID regulators [36,37], MPC approach [38,39], geometric control scheme [40,41]) and resting upon the available measurements. To this end, we assume that each i -th UAV is supplied with a high-precision position sensor (as, for instance, a GNSS sensor) allowing the platform to localize itself in the world frame and regarding the other quadrotors, and a communication interface, i.e., a (ideal) radio module with communication range (communication radius) larger than the formation size. Furthermore, we suppose that each vehicle is equipped with a calibrated camera rigidly attached in its c.o.m. so that the direction of the camera optical axis is given by the yaw angle. Formally, introducing the camera frame $\mathcal{F}_{C_i} = \{O_{C_i}, (x_{C_i}, y_{C_i}, z_{C_i})\}$, we assume that its origin O_{C_i} coincides with O_{B_i} and its orientation regarding \mathcal{F}_{B_i} is time-invariant, namely the angle $\beta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ highlighted in Figure 2 is fixed over time, hence the direction of the optical axis, i.e., z_{C_i} -axis, can be adjusted by rotating around z_{B_i} -axis of the i -th quadrotor body frame. In Figure 2, we also report the image plane \mathcal{I}_i associated with the i -th camera: according to the pinhole camera model [42], the image plane is placed at a certain distance (usually named focal length) regarding the camera center O_{C_i} along the optical axis direction. The center O_{I_i} of \mathcal{I}_i identifies the origin of the (bi-dimensional) image plane frame $\mathcal{F}_{I_i} = \{O_{I_i}, (x_{I_i}, y_{I_i})\}$ that will be used in Section 3.

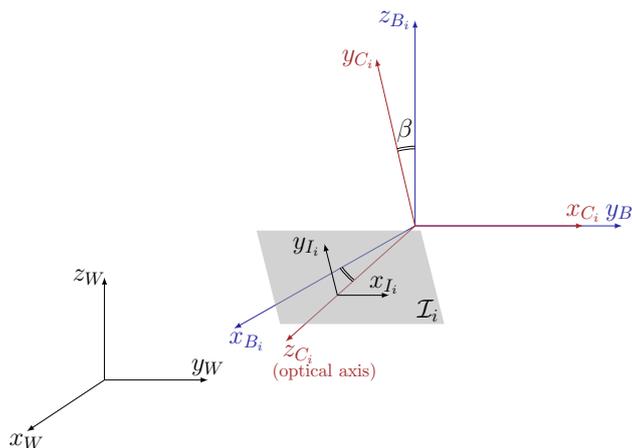


Figure 2. Representation of the reference frames introduced in Section 2, namely the world frame $\mathcal{F}_W = \{O_W, (x_W, y_W, z_W)\}$, the body frame $\mathcal{F}_{B_i} = \{O_{B_i}, (x_{B_i}, y_{B_i}, z_{B_i})\}$, the camera frame $\mathcal{F}_{C_i} = \{O_{C_i}, (x_{C_i}, y_{C_i}, z_{C_i})\}$ and the image plane frame $\mathcal{F}_{I_i} = \{O_{I_i}, (x_{I_i}, y_{I_i})\}$, $i \in \{1 \dots n\}$.

2.2. Formation Model

According to most accepted model, a n -agent formation can be described resting upon graph theory: each i -th component of the group, $i \in \{1 \dots n\}$, is associated with the vertex v_i in set \mathcal{V} (vertex set) with the cardinality $|\mathcal{V}| = n$, while the interaction between i -th and j -th agents is represented by the edge e_{ij} belonging to the set \mathcal{E} (edge set). In the considered scenario, we distinguish between the communication graph $\mathcal{G}_c = (\mathcal{V}, \mathcal{E}_c)$ and the ad-hoc defined visibility update graph $\mathcal{G}_v = (\mathcal{V}, \mathcal{E}_v)$ associated with the formation: the former accounts for the capabilities of the vehicles to exchange information, while the latter is defined in order to optimize the formation total field of view (f.o.v.).

Because of the hypothesis about the communication radius in Section 2.1, each pair of quadrotors in the formation can always communicate. Hence, \mathcal{G}_c results to be a time-invariant undirected complete graph so that for each pair of vertices in the graph there exists a not oriented edge. This implies the possibility to recover reliable relative bearing measurements among all the quadrotors in the group

by exploiting their position sensors, which in turn ensures the opportunity to control the formation through a bearing-rigidity approach (see, e.g., [43]). Under these hypotheses, we can assume that a suitable controller is implemented on each quadrotor so that the formation is able to perform global roto-translation on a plane parallel to the $(x_W y_W)$ -plane of the world frame by maintaining its shape. In particular, we suppose that at each time instant all the UAVs in the group are aligned along a certain direction with a fixed distance among them and maintaining a constant height regarding the ground. Formally, for each $i \in \{1 \dots n\}$, it holds that $\mathbf{p}_{O_{B_i}}^W(t) = p_i \mathbf{d}(t) + \zeta \mathbf{e}_3$ for $t \geq 0$ with $|p_i - p_j| = d_Q$, $j = \max\{0, i - 1\}$, $d_Q \in \mathbb{R}^+$, $\mathbf{d}(t) = [d_x(t) \ d_y(t) \ 0]^\top \in \mathbb{R}^3$, $\|\mathbf{d}(t)\| = 1$ and $\zeta \in \mathbb{R}^+$.

To describe the properties of the visibility update graph \mathcal{G}_v , some premises are needed. First, we assume that $n = 2k + 1$, $k \in \mathbb{N}$, i.e., the number of UAVs composing the formation is odd, and then we label the formation components so that the ℓ -th agent, $\ell = k + 1$, is identified as the formation *quadrotor leader*. This platform is characterized by the fact that its camera is always pointing through the target and its position projection on $(x_W y_W)$ -plane of \mathcal{F}_W , corresponding to $\mathbf{p}_{O_{B_\ell}}^W(t) = p_\ell \mathbf{d}(t) \in \mathbb{R}^3$, is always at a certain fixed distance $d_T \in \mathbb{R}^+$ regarding the target position. In order to optimize the formation total f.o.v., the visibility update graph is assumed as in Figure 3: \mathcal{G}_v is an oriented time-invariant graph so that $\mathcal{E}_v = \{e_{ij}, j = i + 1, i \in 1 \dots \ell - 1\} \cup \{e_{ij}, j = i - 1, i \in \ell + 1 \dots n\}$, in other words we are considering neighboring camera pairs with potentially overlapping f.o.v.. The agents interaction set \mathcal{E}_v constitutes an arbitrary choice in the optimization framework introduced in Section 4, although this represents a minimum configuration in terms of edges number.

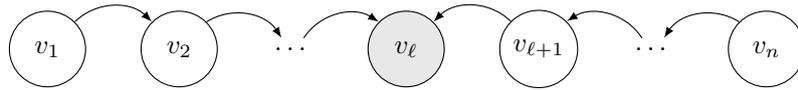


Figure 3. Assumed visibility update graph: ℓ -th node corresponds to quadrotor leader.

Please note that both the inter-agents distance d_Q and the target distance d_T affect the quality of the target position reconstruction: higher values for d_Q and lower values for d_T reduce the error in depth reconstruction, while if the UAVs are too far apart or d_T is too small issues may arise concerning the visibility of the target. Nonetheless, we assume that d_Q and d_T are selected in order to guarantee the possibility to reconstruct the target position with negligible error and adopting a distributed estimation paradigm (as, for instance, [44]).

2.3. Problem Statement

Under the assumptions stated in Sections 2.1 and 2.2, the problem of tracking a mobile ground robot through the described quadrotors formation boils down to the definition of an optimal control law regulating the UAVs orientation. In particular, the idea is to adjust the quadrotors yaw angles, and consequently the attitude of their embedded cameras, by providing reference values that aim at optimizing the total f.o.v. of the formation, resulting from the combination of the single camera views, with the twofold intent of minimizing the error on the target position estimate and the probability of losing the target during two consecutive orientation updates. Next observations highlight the challenges of the required task.

1. Assuming that the target position is estimated through triangulation algorithms, the quality of target position estimate is related to the number of cameras that are able to see the target [45]. If the f.o.v. of all the cameras are overlapping, i.e., if their optical axes are pointing towards the target, all of them can contribute to some extent with visual information regarding the target and then the quality of the estimate is increased. In the light of this fact, we assume that the more the camera f.o.v. are overlapping, the more it is likely that the target is simultaneously visible from a larger number of cameras.

- On the other hand, the target loss probability decreases as the area of the total formation f.o.v., corresponding to the monitored region, increases. Indeed, when the covered area is minimal, it may happen that the target rapidly drifts outside of the view of any camera. Conversely, if the views are less overlapping, it is more likely that if one camera loses the target, it falls into the view of another camera that is looking in a slightly different direction.

Clearly, the minimization of both the target position estimate error and its loss probability results to be conflicting requirements: the former tends to maximize the overlapping of the cameras f.o.v. (Figure 4), whereas the latter favors the coverage entailing the maximization of the formation total f.o.v. (Figure 5). Our aim is to derive a procedure to attain an optimal trade-off between the two purposes by adopting a distributed paradigm.

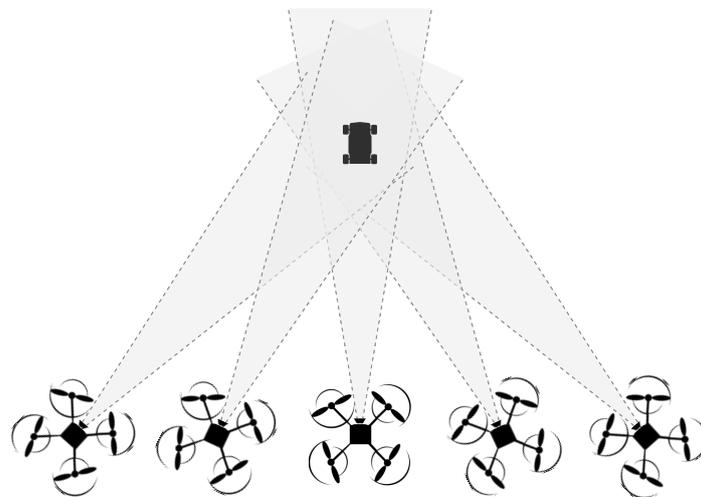


Figure 4. Maximum overlapping among cameras f.o.v.

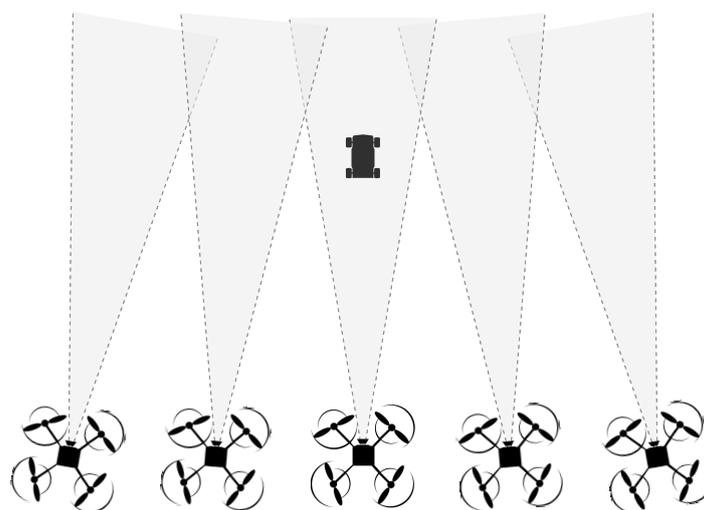


Figure 5. Minimum overlapping among cameras f.o.v.

3. Computation of the Formation Total f.o.v.

Accounting for the observations in Section 2.3, one can trivially realize that the tracking task achievement depends on the formation total f.o.v., which derives from the combination of the single cameras views (depending on the quadrotors attitude). For this reason this section is structured in

two parts: firstly (in Section 3.1) we propose an approximation of the perspective projection of the image plane of a single camera on the $(x_W y_W)$ -plane of the world frame, hereafter referred as *target plane* \mathcal{T} , and then the computation of the formation total f.o.v. is obtained by exploiting the derived approximation (Section 3.2).

Please note that in most of this section we drop out the time dependence to allow for a more compact notation, however time will be explicitly re-introduced at the end of Section 3.2 to highlight the variables dependency.

3.1. Single Camera f.o.v. Approximation

In this first part, the attention is focused on the i -th single camera. The perspective projection of its image plane \mathcal{I}_i on the target plane depends on the corresponding UAV attitude, i.e., on the ψ_i angle. Such region on the plane \mathcal{T} has always a trapezoidal shape and, in the following, we exploit this fact to derive an approximation of the the i -th camera f.o.v. through a suitable discretization of the target plane.

To estimate the area visible from the i -th camera, we first determine the trapezoid vertices by projecting the corners of its image plane \mathcal{I}_i onto the target plane as depicted in Figure 6. For this purpose, we introduce the matrix $\mathbf{P}_i \in \mathbb{R}^{3 \times 4}$ that summarizes the intrinsic and extrinsic parameters of the i -th camera. It is well-known that, accounting for the pinhole camera model, the projection on the image plane \mathcal{I}_i of a generic point Q in \mathcal{F}_W is given by $\tilde{\mathbf{p}}_Q^{I_i} = \mathbf{P}_i \tilde{\mathbf{p}}_Q^W$, where $\tilde{\mathbf{p}}_Q^{I_i} \in \mathbb{R}^3$ and $\tilde{\mathbf{p}}_Q^W \in \mathbb{R}^4$ identify the position of Q in \mathcal{F}_{I_i} and in \mathcal{F}_W , respectively, expressed in homogeneous coordinates. Nonetheless, the problem we need to solve is the inverse: given the position $\tilde{\mathbf{p}}_{Q^*}^{I_i} \in \mathbb{R}^3$ on the image plane, determine the position (expressed in \mathcal{F}_W) of the point Q^* on the target plane such that $\tilde{\mathbf{p}}_{Q^*}^{I_i} = \mathbf{P}_i \tilde{\mathbf{p}}_{Q^*}^W$. In this context, we observe that the projection map \mathbf{P}_i defines a not surjective function on \mathcal{I}_i : all the points of the ray $\mathcal{R}(O_{C_i}, \overline{O_{C_i}Q^*})$, starting at O_{C_i} and oriented as the segment $\overline{O_{C_i}Q^*}$, are described by the same coordinates in \mathcal{F}_{I_i} , i.e., their projection onto the image plane corresponds to the same point. Hence, we proceed by defining a parametrization for the ray $\mathcal{R}(O_{C_i}, \overline{O_{C_i}Q^*})$ and then by computing its intersection with the plane \mathcal{T} .

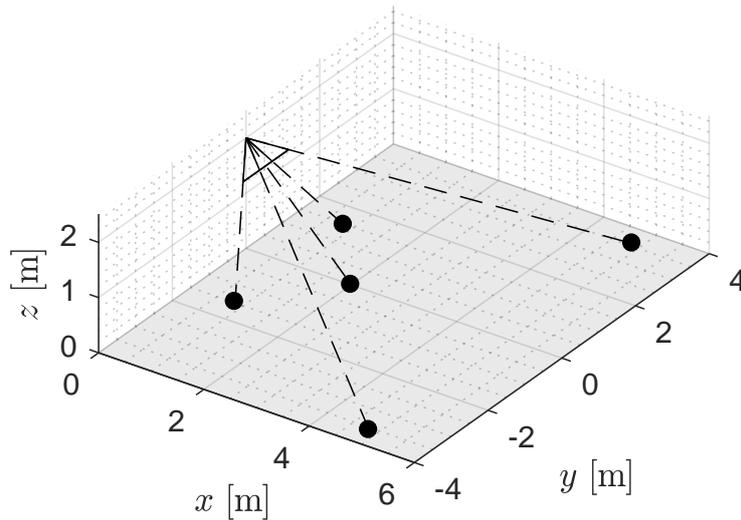


Figure 6. Projected image plane corners and center onto target plane.

We first observe that, accounting for the left Moore-Penrose pseudoinverse of the camera matrix \mathbf{P}_i , it is possible to determine the position (expressed in \mathcal{F}_W in homogeneous coordinates) of a generic point Q_g belonging to $\mathcal{R}(O_{C_i}, \overline{O_{C_i}Q^*})$. Indeed, it holds that $\tilde{\mathbf{p}}_{Q_g}^W = \mathbf{P}_i^\dagger \tilde{\mathbf{p}}_{Q_g}^{I_i}$, where $\mathbf{P}_i^\dagger = \mathbf{P}_i^\top (\mathbf{P}_i \mathbf{P}_i^\top)^{-1}$. Therefore, the direction of the segment $\overline{O_{C_i}Q^*} = \overline{O_{C_i}Q_g}$ in \mathcal{F}_W is identified by the (unit) vector

$$\mathbf{v} = \frac{\mathbf{p}_{Q_s}^W - \mathbf{p}_{O_{B_i}}^W}{\|\mathbf{p}_{Q_s}^W - \mathbf{p}_{O_{B_i}}^W\|} \in \mathbb{R}^3, \quad (3)$$

where $\mathbf{p}_{Q_s}^W \in \mathbb{R}^3$ and $\mathbf{p}_{O_{B_i}}^W \in \mathbb{R}^3$ denote the position of Q and O_{C_i} in the world frame (in non-homogeneous coordinates), respectively. As a consequence, a suitable parametrization of the points on the ray $\mathcal{R}(O_{C_i}, \overline{O_{C_i}Q^*})$ is given by

$$\mathbf{p}_Q^W(\lambda) = \mathbf{p}_{O_{B_i}}^W + \lambda \mathbf{v}, \quad (4)$$

with $\lambda \in \mathbb{R}^+$. The value of λ determines the position (in non-homogeneous coordinates) of a point along the direction \mathbf{v} in \mathcal{F}_W . The coordinates of the required point Q^* are thus obtained by determining the value λ^* such that $\mathbf{p}_Q^W(\lambda^*)$ identifies a position on the target plane \mathcal{T} . To do so, we consider the third component of $\mathbf{p}_Q^W(\lambda^*)$, namely $z_Q^W(\lambda^*) = \mathbf{e}_3^\top \mathbf{p}_Q^W(\lambda^*) \in \mathbb{R}$. This must be zero since the target plane coincides with the $(x_W y_W)$ -plane of \mathcal{F}_W , hence we impose $z_Q^W(\lambda^*) = \mathbf{e}_3^\top \mathbf{p}_{O_{B_i}}^W + \lambda^* \mathbf{e}_3^\top \mathbf{v} = 0$. It thus follows that $\lambda^* = -(\mathbf{e}_3^\top \mathbf{p}_{O_{B_i}}^W) / (\mathbf{e}_3^\top \mathbf{v})$ and the position of $Q^* \in \mathcal{T}$ in the world frame is given by

$$\mathbf{p}_{Q^*}^W = \mathbf{p}_Q^W(\lambda^*) = \mathbf{p}_{O_{B_i}}^W - \frac{\mathbf{e}_3^\top \mathbf{p}_{O_{B_i}}^W}{\mathbf{e}_3^\top \mathbf{v}} \mathbf{v}. \quad (5)$$

Please note that the computation (5) requires the knowledge of $\mathbf{p}_{O_{B_i}}^W$: in the considered scenario where O_{C_i} coincides with the i -th UAV c.o.m. O_{B_i} , it is reasonable to assume that the position of O_{B_i} can be estimate (with negligible error) thanks to the position sensor of the i -th quadrotor.

Through the described procedure, each i -th camera can derive the projection on the target plane of its image plane corners $V_{i,l}, l \in \{1 \dots 4\}$, and center O_{I_i} , all depending on the optical axis orientation, namely on the ψ_i angle of the i -th quadrotor, and worthwhile to determine an approximation of the whole i -th camera f.o.v.. The core idea, indeed, is to determine the two region of the plane delimited by the vertices and exploit the projected image center to mathematically discern the visible one. Hereafter we neglect the third component both in the coordinates and position vector definition when this is zero, i.e., when we deal with points on the target plane.

We first identify the four lines passing through subsequent vertices of the projected image plane. We denote by $\mathbf{p}_{V_{i,l}}^W \in \mathbb{R}^2, l \in \{1 \dots 4\}$, and by $\mathbf{p}_{O_{I_i}}^W \in \mathbb{R}^2$ the position of the projection of the image plane corners $V_{i,l}, l \in \{1 \dots 4\}$, and center O_{I_i} on the target plane and we suppose that the corners points are ordered such that $V_{i,l}$ and $V_{i,(l+1) \bmod 4}$ are subsequential. Therefore the line $r_{i,l}(x, y) = a_{i,l}x + b_{i,l}y + c_{i,l} = 0$ passing through $\mathbf{p}_{V_{i,l}}^W$ and $\mathbf{p}_{V_{i,(l+1) \bmod 4}}^W$ is identified by the coefficient vector $\theta_{i,l} = [a_{i,l} \ b_{i,l} \ c_{i,l}] \in \mathbb{R}^3$ that solves the system

$$\begin{bmatrix} (\mathbf{p}_{V_{i,l}}^W)^\top & 1 \\ (\mathbf{p}_{V_{i,(l+1) \bmod 4}}^W)^\top & 1 \end{bmatrix} \begin{bmatrix} a_{i,l} \\ b_{i,l} \\ c_{i,l} \end{bmatrix} = \mathbf{0}. \quad (6)$$

After the computation of $\theta_{i,l}$ for all $l \in \{1 \dots 4\}$, the plane \mathcal{T} turns out to be divided into four pairs of half-planes defined by $H_{i,l}^+ = \{(x, y) \in \mathcal{T} \mid r_{i,l}(x, y) > 0\}$ and $H_{i,l}^- = \{(x, y) \in \mathcal{T} \mid r_{i,l}(x, y) < 0\}$ as depicted in Figure 7. It is then easy to check whether a point $Q \in \mathcal{T}$ placed at (x_Q^W, y_Q^W) lies in the same region of the projection of O_{I_i} on the target plane by evaluating if

$$r_{i,l}(x_Q^W, y_Q^W) r_{i,l}(x_{O_{I_i}}^W, y_{O_{I_i}}^W) > 0, \quad \forall l = 1 \in \{1 \dots, 4\} \quad (7)$$

being $\mathbf{p}_{O_{I_i}}^W = [x_{O_{I_i}}^W \ y_{O_{I_i}}^W] \in \mathbb{R}^2$ the vector identifying the position of the projection of O_{I_i} on \mathcal{T} . Note that the condition (7) accounts for the intersection of the four half-planes.

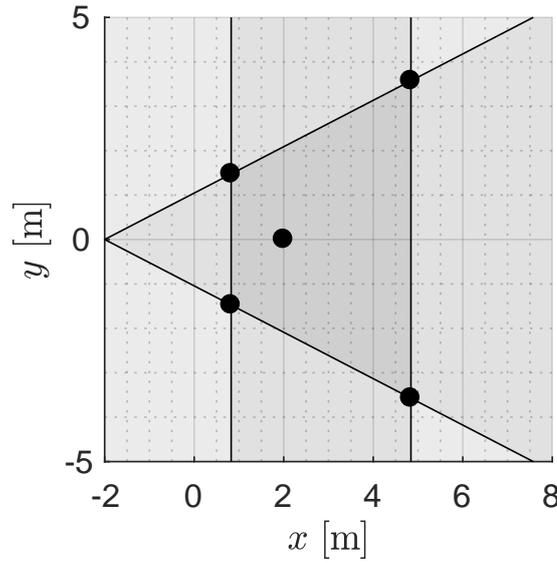


Figure 7. Lines $r_{i,\iota}$, $\iota \in \{1 \dots 4\}$ and resulting half-planes.

We introduce now a discretization of the target plane resorting on a proper sampling rate $\delta \in \mathbb{R}^+$. Formally, we consider the sampled version of \mathcal{T} , i.e.,

$$\mathcal{T}_\delta = \{(x, y) \in [x_m, x_M] \times [y_m, y_M] \mid x = x_m + h\delta, y = y_m + k\delta, h, k \in \mathbb{N}\} \quad (8)$$

where the limits $x_m, x_M, y_m, y_M \in \mathbb{R}$ have to be properly chosen according to the considered scenario. We assume that whether a point $Q \in \mathcal{T}$ placed at (x_Q^W, y_Q^W) is visible from any i -th camera, then all the points in $[x_Q^W - \frac{1}{2}\delta, x_Q^W + \frac{1}{2}\delta] \times [y_Q^W - \frac{1}{2}\delta, y_Q^W + \frac{1}{2}\delta]$ are considered visible as well by the same camera. This implies that each sample of \mathcal{T}_δ (a square of side δ) is assumed to be visible from the i -th camera only if its center is visible.

These premises lead to the introduction for each i -th camera of the matrix $\mathbf{M}_i(\psi_i) \in \mathbb{R}^{r \times c}$ with $r = \lfloor \frac{x_M - x_m}{\delta} \rfloor$, $c = \lfloor \frac{y_M - y_m}{\delta} \rfloor$ so that

- each entry $m_{h,k}$, $h \in [1, r]$, $k \in [1, c]$, refers to the point $Q_{h,k}$ placed at $(x_m + h\delta, y_m + k\delta)$ in the sampled plane \mathcal{T}_δ , i.e., such that its position in the world frame is identified by the vector $\mathbf{p}_{Q_{h,k}}^W = [x_m + h\delta \quad y_m + k\delta]^T \in \mathbb{R}^2$;
- any non-zero value of $m_{h,k}$ indicates that the sample centered in $Q_{h,k}$ is not visible from the i -th camera while a zero value means that this is not visible. More formally, if condition (7) holds, then, the value of $m_{h,k}$ is set to $m^* \in \mathbb{R}/\{0\}$, otherwise it is set to zero. The definition of m^* may be context-dependent: although the easiest choice is to impose $m^* = 1$, spatial information can be incorporate in this value, as will be explained later.

It is straightforward that the choice of the sampling rate δ influences the accuracy of the visibility matrix at the expense of its dimensions. As a consequence, a proper settling is mandatory based on the considered scenario in terms of dimensions of the involved vehicles. An example of result of the procedure described in this section is reported in Figure 8.

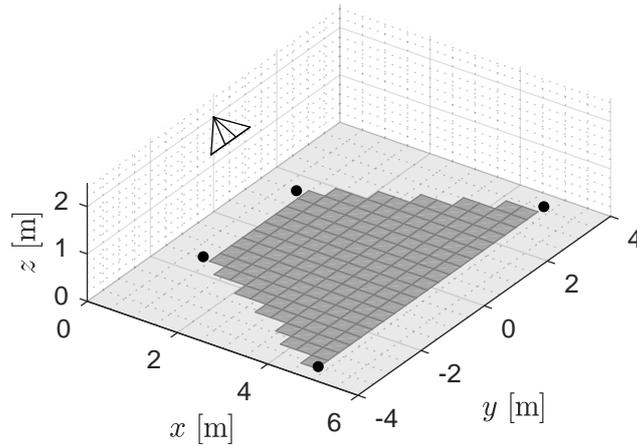


Figure 8. Approximated i -th camera f.o.v. with $\delta = 0.35$ m.

3.2. Formation Total f.o.v. Approximation

In Section 3.1 we have introduced a method to compute an approximation of the i -th single camera visible region of the target plane. This second part is devoted to the derivation of the formation total f.o.v., accounting for the visibility update graph reported in Figure 3: the idea is to combine the information summarized by the matrices $\{\mathbf{M}_i(\psi_i)\}_{i=1}^n$ by exploiting the agents interaction described by the set \mathcal{E}_v .

In this direction, for each i -th quadrotor in the formation, namely for each i -th vertex in the graph \mathcal{G}_v , we introduce the neighboring set $\mathcal{N}_i = \{v_j \in \mathcal{V} \mid e_{i,j} \in \mathcal{E}_v\}$ that allows defining the *visibility matrix* $\mathbf{M}'_i \in \mathbb{R}^{r \times c}$ as follows

$$\mathbf{M}'_i = \mathbf{M}_i(\psi_i) + \sum_{j \in \mathcal{N}_i} \mathbf{M}_j(\psi_j). \tag{9}$$

This new matrix encodes the information about the overlapping between the f.o.v. of neighboring cameras: for instance, when the value of m^* is set to 1, the samples of \mathcal{T}_δ in the overlapping area corresponds to the matrix entries with a value larger than 1; in particular such a value indicates the number of cameras to which the corresponding sample is visible. In Figure 9 we report the case of two neighboring cameras, namely the i -th and j -th cameras such that the edge $e_{ij} \in \mathcal{E}_c$ is oriented from v_i to v_j . The darker region corresponds to the overlapping area between the two views, i.e., to the samples of the target plane whose corresponding entries in the visibility matrix \mathbf{M}'_i have value greater than 1.

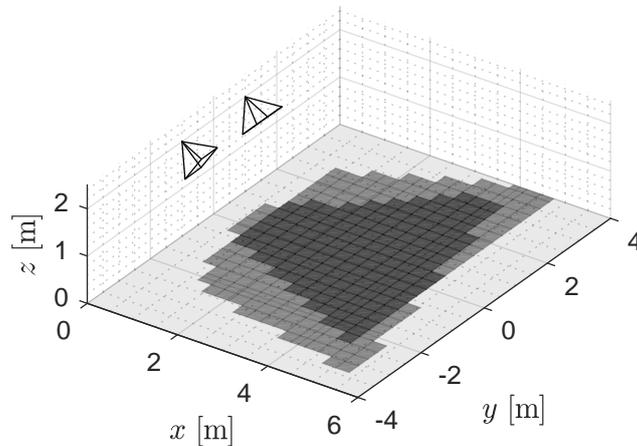


Figure 9. Projected f.o.v. of two neighboring cameras.

To conclude, we can observe that the formation total f.o.v. is described by the *total visibility matrix* $\mathbf{M}' = \sum_{i=1}^n \mathbf{M}'_i$. Please note that $\mathbf{M}' = \mathbf{M}'(\{\psi_i(t)\})$, namely at each time instant $t \geq 0$ the total visibility matrix depends on the orientation of all the quadrotor in the formation, while each matrix $\mathbf{M}'_i, i \in \{1 \dots n\}$, depends on the current orientation of the i -th agent and of all its neighboring agents nonetheless to simplify the notation, hereafter, we explicit only the dependence on $\psi_i(t)$ by setting $\mathbf{M}'_i = \mathbf{M}'_i(\psi_i(t))$.

4. Optimization of Quadrotors Attitude

As figured out in Section 2.3, the problem of tracking a mobile ground robot through the quadrotors formation described in Section 2.2 consists of the definition of an update law for the reference yaw angle of all the UAVs in the group. In particular, the intent is to account for the target position error estimation (Section 4.1) and for the target loss probability (Section 4.2) by deriving a suitable trade-off between the minimization of these two quantities (Section 4.3) that ensures the tracking task achievement.

4.1. Minimization of Target Position Estimation Error

According to the first observation carried out in Section 2.3, the error on the target position estimate decreases as far as the dimension of the cameras f.o.v. overlapping area increases. In particular, the most advantageous scenario corresponds to the case wherein all the cameras optimal axes are pointed towards the target and the total formation f.o.v. turns out to have minimum size. The minimization of the formation total f.o.v. resulting from the combination of the single cameras views is addressed in this section, by exploiting the approximation derived in Section 3.2. In detail, we propose a solution based on the spatial information that can be inferred from the visibility matrices.

Consider two neighboring cameras i and j , such that $(i, j) \in \mathcal{E}_v$: the two corresponding f.o.v. on the target plane, approximated with the algorithm described in the previous section, define a region of the space where they are overlapping. The projection of the optical axis of camera i on $\mathcal{T}(\mathcal{T}_\delta)$ will split the overlapping region into two parts, highlighted with red and blue colors in Figure 10. We notice that if the overlapping region is mostly laying on the right side (i.e., the blue cells are more than red ones in Figure 10) then a counterclockwise rotation around its z_{B_i} -axis of \mathcal{F}_{B_i} will increase the overlapping region, and vice-versa. Motivated by this fact, we assign a positive value to each cell placed on the left side (red cells in Figure 10), and a negative value to those on the right side (blue cells). Specifically, for the i -th camera, $i \in \{1 \dots n\}$, let $|m'_{h,k}| \in \mathbb{R}$ be the value of the (h,k) -entry of the visibility matrix $\mathbf{M}'_i(\psi_i(t))$ in (9) corresponding to the sample of \mathcal{T}_δ centered in $Q_{h,k}$. We consider the following quantity

$$N_i(\psi_i(t)) = \sum_{|m'_{h,k}| > 1} f_{h,k}(\psi_i(t)), \tag{10}$$

with

$$f_{h,k}(\psi_i(t)) = \begin{cases} +1 & \text{for } \mathbf{e}_2^\top \mathbf{p}_{Q_{h,k}}^{B_i}(t) \geq 0, \\ -1 & \text{otherwise.} \end{cases} \tag{11}$$

In (11), the vector $\mathbf{p}_{Q_{h,k}}^{B_i}(t) \in \mathbb{R}^3$ corresponds to the position of $Q_{h,k}$ in \mathcal{F}_{B_i} at time instant t , computed as $\mathbf{p}_{Q_{h,k}}^{B_i}(t) = ({}^W \mathbf{R}_{B_i}(t))^\top (\mathbf{p}_{Q_{h,k}}^W(t) - \mathbf{p}_{O_{B_i}}^W(t))$. Evaluating the second component of $\mathbf{p}_{Q_{h,k}}^{B_i}(t)$ it is possible to check whether the sample of \mathcal{T}_δ centered in $Q_{h,k}$ and visible from both i -th and j -th cameras at time instant t ($|m'_{h,k}| > 1$) is placed on the left ($f_{h,k}(\psi_i(t)) = +1$) or on the right ($f_{h,k}(\psi_i(t)) = -1$) i -th camera f.o.v. portion. Therefore, from Figure 10, which reports the action of function (11), one can realize that the absolute value of $N_i(\psi_i(t))$ represents the *imbalance factor* between the right and left i -th camera f.o.v. portion.

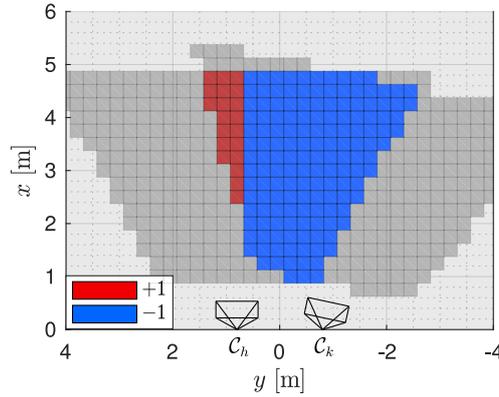


Figure 10. Overlapping samples: those in the left portion (red) of camera h are counted positively while the others (blue) are counted with a negative sign.

In this context, one can perceive that the minimization of the formation total f.o.v. (i.e., of the target position estimation error) relies on the balance between left and right f.o.v. portions for each camera, hence on the minimization of $|N_i(\psi_i(t))|$, for all $i \in \{1 \dots n\}$.

Formally, we aim at solving the following optimization problem for $i \in \{1 \dots n\}$:

$$\min_{\psi_i} \frac{1}{2} (N_i(\psi_i))^2. \quad (12)$$

Since the cost function in (12) is derived through a numerical approximation, we propose an iterative greedy solution based on the fact that when $N_i(\psi_i(t)) < 0$, then the i -th camera overlapping region is unbalanced on its right side, so the i -th quadrotor is required to rotate clockwise around its z_{B_i} -axis in order to minimize the index and vice-versa. It is then reasonable to choose the optimizing direction of the yaw angles according to the sign of $N_i(\psi_i(t))$ and define, for each agent, the following update rule:

$$\dot{\psi}_i(t) = k_\psi \text{sign}(N_i(\psi_i(t))) |N_i(\psi_i(t))|, \quad (13)$$

where $k_\psi \in \mathbb{R}^+$ is a constant positive gain influencing the convergence speed of the proposed procedure. Please note that we assume to simultaneously regulate both the position and the attitude of the formation vehicles resting upon the fact that their two pose components are independently controllable as explained in Section 2.1.

4.2. Target Loss Probability Definition

The second observation reported in Section 2.3 supports the maximization of the area monitored by the formation, basing on the fact that the target loss probability decreases as far as the dimension of the total formation f.o.v. increases. Although this requirement is in contrast regarding the motivations justifying the angles updating strategy (29), in this section we focus on this issue introducing a suitable dynamic model for the target, figuring out a Kalman-based solution to predict its position over time (similarly to what proposed in [46]), and then, computing the probability that is not visible from any camera according to the derived position prediction. Such a probability is hereafter referred as *target loss probability*.

In the considered scenario, any assumption is made on the mobile ground robot tracked by the quadrotors formation. The second-order model is thus adopted to describe its dynamics: denoting by $\mathbf{p}_T^W(t) \in \mathbb{R}^2$ and $\mathbf{u}_T^W(t) \in \mathbb{R}^2$ the position of the target on the plane \mathcal{T} and its control input vector at time instant t , respectively, we assume that $\dot{\mathbf{p}}_T^W(t) = \mathbf{u}_T^W(t)$ for $t \geq 0$. This choice is motivated by the fact that second-order systems well approximate the behavior of physical robotic agents and that controlling acceleration (rather than velocity) generally allows a vehicle to realize smooth movements.

In order to derive a prediction law for the target position $\mathbf{p}_T^W(t)$ over time, we introduce the vector $\mathbf{q}_T^W(t) = [(\mathbf{p}_T^W(t))^\top \ (\dot{\mathbf{p}}_T^W(t))^\top]^\top \in \mathbb{R}^4$ whose dynamics is regulated by the following equation

$$\dot{\mathbf{q}}_T^W(t) = \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{I}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \end{bmatrix} \mathbf{q}_T^W(t) + \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_{2 \times 2} \end{bmatrix} \mathbf{u}_T^W(t) = \mathbf{F} \mathbf{q}_T^W(t) + \mathbf{G} \mathbf{u}_T^W(t). \quad (14)$$

Then, to exploit a Kalman-based approach, we consider the discrete version of (14) derived by accounting for the sample period $T_s \in \mathbb{R}^+$ so that the matrices $\mathbf{F} \in \mathbb{R}^{4 \times 4}$ and $\mathbf{G} \in \mathbb{R}^{4 \times 2}$ are substituted by

$$\mathbf{F}_{T_s} = \exp(\mathbf{F}T_s) = \begin{bmatrix} \mathbf{I}_{2 \times 2} & T_s \mathbf{I}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{I}_{2 \times 2} \end{bmatrix} \quad \text{and} \quad \mathbf{G}_{T_s} = \int_0^{T_s} \exp(\mathbf{F}t) \mathbf{G} dt = \begin{bmatrix} \frac{T_s^2}{2} \mathbf{I}_{2 \times 2} \\ T_s \mathbf{I}_{2 \times 2} \end{bmatrix}. \quad (15)$$

Moreover, we account for the fact that each camera can retrieve a noisy measurement of the current target position by introducing a normalized Gaussian noise vector with zero mean and unit variance. A suitable model for the Kalman filter is thus given by

$$\mathbf{q}_T^W(s+1) = \mathbf{A} \mathbf{q}_T^W(s) + \mathbf{B} \mathbf{n}_T^W(s) \quad (16)$$

$$\mathbf{p}_T^W(s) = \mathbf{C} \mathbf{q}_T^W(s) + \mathbf{D} \mathbf{n}_T^W(s) \quad (17)$$

where $\mathbf{q}_T^W(s+1) = \mathbf{q}_T^W((s+1)T_s)$ and $\mathbf{n}_T^W(s) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{4 \times 4})$. In (16) we have also introduced the following matrices depending on the parameters $\sigma_u \in \mathbb{R}^+$ and $\sigma_p \in \mathbb{R}^+$ which account for the variance on the target dynamic model and on the target position measurements

$$\mathbf{A} = \mathbf{F}_{T_s} \in \mathbb{R}^{4 \times 4} \quad \mathbf{B} = [\mathbf{0}_{4 \times 2} \quad \sigma_u \mathbf{G}_{T_s}] \in \mathbb{R}^{4 \times 4} \quad (18)$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} \end{bmatrix} \in \mathbb{R}^{2 \times 4} \quad \mathbf{D} = [\sigma_p \mathbf{I}_{2 \times 2} \quad \mathbf{0}_{2 \times 2}] \in \mathbb{R}^{2 \times 4} \quad (19)$$

Denoting by $\hat{\mathbf{q}}_T^W(s) \in \mathbb{R}^4$ the estimate of $\mathbf{q}_T^W(s)$ in (16) computed at a certain time sample $s \geq 0$ having properly initialized the related Kalman filter, the (minimum error) prediction of the state after $r \in \mathbb{R}^+$ steps, i.e., $\hat{\mathbf{q}}_T^W(s+r) \in \mathbb{R}^4$, is given by

$$\hat{\mathbf{q}}_T^W(s+r) = \mathbf{A}^r \hat{\mathbf{q}}_T^W(s). \quad (20)$$

Furthermore, indicating with $\Sigma(s) \in \mathbb{R}^{4 \times 4}$ the variance matrix of $\hat{\mathbf{q}}_T^W(s)$, the variance of the prediction (20) results to be

$$\mathbf{Q}(s+r) = \mathbf{A}^r \Sigma(s) (\mathbf{A}^\top)^r + \sum_{v=0}^{r-1} \mathbf{A}^v \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^v \in \mathbb{R}^{4 \times 4} \quad (21)$$

As a consequence, the target position prediction $\hat{\mathbf{p}}_T^W(s+r) \in \mathbb{R}^2$ after r steps can be computed at each time sample $s \geq 0$ as

$$\hat{\mathbf{p}}_T^W(s+r) = \mathbf{C}^r \hat{\mathbf{q}}_T^W(s+r), \quad (22)$$

with variance

$$\mathbf{P}(s+r) = \mathbf{C}^r \mathbf{Q}(s+r) (\mathbf{C}^\top)^r + \sum_{v=0}^{r-1} \mathbf{C}^v \mathbf{D} \mathbf{D}^\top (\mathbf{C}^\top)^v \in \mathbb{R}^{2 \times 2} \quad (23)$$

Please note that, since the target position is supposed to be driven by a Gaussian noise according to (16), at each time sample $s \geq 0$ it is possible to retrieve the complete description of its prediction distribution, i.e., the function $f_T : \mathbb{R}^+ \times \mathcal{T}_\delta \rightarrow \mathbb{R}^+$ associating to each point $(x_m + h\delta, y_m + k\delta)$ of the (sampled) target plane the time-dependent infinitesimal probability that $\hat{\mathbf{p}}_T^W(s+r) = [x_m + h\delta \quad y_m + k\delta]^\top$ (an example of $f_T(\cdot)$ is depicted in Figure 11). Then, given the

distribution $f_T(\cdot)$ of the target position prediction, at any time instant t we can compute the probability that the target falls outside the view of any camera in the formation, namely the so-called target loss probability. In this direction, we impose that $f_T(t, \cdot, \cdot) = f_T(s, \cdot, \cdot)$ for $t \in [sT_s, (s + 1)T_s)$, namely the prediction distribution is supposed constant for the entire sample period.

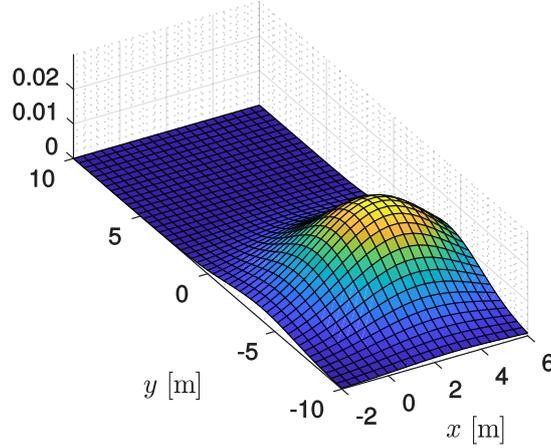


Figure 11. Example of resulting prediction distribution $f_T(\cdot)$.

Indicating with $\mathcal{P}_i(t) \subset \mathcal{T}$ the region of the target plane visible by i -th camera, $i \in \{1 \dots n\}$, at time t , we consider the total formation f.o.v. $\mathcal{P}(t) = \cup_{i=1}^n \mathcal{P}_i(t)$ described by the total visibility matrix $\mathbf{M}'(\{\psi_i(t)\})$ introduced at the end of Section 3.2 so that each non-zero entry corresponds to a sample of the target plane \mathcal{T}_δ visible by at least a camera in the formation. The target loss probability $p_T(t) \in [0, 1]$ at time instant t can thus be computed as

$$p_T(t) = \iint_{\mathcal{T} \setminus \mathcal{P}(t)} f_T(t, x, y) dx dy = 1 - \iint_{\mathcal{P}(t)} f_T(t, x, y) dx dy. \tag{24}$$

Nonetheless, in the following we provide a numerical approximation of (24) based on the f.o.v. approximation illustrated in Section 3.

To this end, we account for the fact that each i -th agent, $i \in \{1 \dots n\}$, can compute its visibility matrix $\mathbf{M}'_i(\psi_i(t))$ according to (9). Moreover, because of the hypothesis on the communication graph, we can assume that at each time instant t the matrix $\mathbf{M}'(\{\psi_i(t)\})$ is known to all the agents of the formation. As a consequence, each quadrotor can compute the following approximation for (24)

$$p_T(t) \simeq 1 - \sum_h \sum_k f_I(m'_{h,k}) f_T(t, x_m + h\delta, y_m + k\delta) \delta^2 \tag{25}$$

depending on the indicator function

$$f_I(m'_{h,k}) = \begin{cases} 0 & \text{if } m'_{h,k} \neq 0, \\ 1 & \text{otherwise,} \end{cases} \tag{26}$$

where, with abuse of notation regarding the previous sections, $m'_{h,k}$ denotes the (h, k) entry of $\mathbf{M}'(\{\psi_i(t)\})$. Please note that the function (26) considers all the non visible samples in the target plane \mathcal{T}_δ .

According to the approximation (25), the target loss probability is usually distributed as in the example in Figure 12 and explicitly depends on all the quadrotor orientations, namely $p_T(t) = p_T(\{\psi_i(t)\})$, as it will appear in the optimization framework defined in the following.

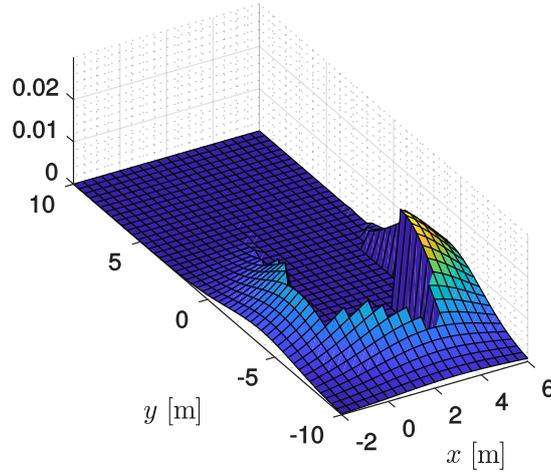


Figure 12. Example of the distribution of the target loss probability.

4.3. Regulated Solution

In Section 4.1 we have introduced the optimization strategy (29) aiming at maximizing the overlapping area of the formation cameras f.o.v. motivated by the purpose of minimizing the target position estimation error.

We now propose the full optimization problem, accounting for both the estimation error and target-loss probability minimization. Dropping the time dependence for the ease of readability, for each agent $i \in \{1 \dots n\}$ we aim at solving the following:

$$\min_{\psi_i} \frac{1}{2} (N'_i(\psi_i))^2, \quad N'_i(\psi_i) = N_i(\psi_i) + k_\rho \sum_{|m'_{h,k}| > 1} \rho_i(p_T(\{\psi_i\})), \quad (27)$$

with

$$\rho_i(p_T(\{\psi_i\})) = \text{sign}(i - \ell) p_T(\{\psi_i\}), \quad (28)$$

where the sign function allows discriminating the left/right neighbors of the quadrotor leader. Please note that the constant gain $k_\rho \in \mathbb{R}^+$ in (27) constitutes a regulator factor that allows tuning the trade-off between the minimization of the target position estimation error (low value for k_ρ) and of the target loss probability (high value for k_ρ).

As the proposed solution to (27), we consider the following update law for steering the yaw angles trajectory

$$\dot{\psi}_i(t) = k'_\psi \text{sign}(N'_i(\psi_i(t))) |N'_i(\psi_i(t))|, \quad (29)$$

with tuning parameter $k'_\psi \in \mathbb{R}^+$.

5. Simulation Results

The method figured out in Section 4 is based on several heuristics that have no theoretical guarantees of convergence, hence to assess the validity of the proposed solution an extensive simulation campaign has been carried out within a realistic simulation environment resting upon Gazebo multi-robot simulator and ROS (Robotic Operating System) architecture. In detail, Gazebo is employed to simulate in a realistic way the agents dynamics (both the UAVs and the ground robot one) with the ODE physics engine. Thanks to sparse matrix methods from the *Eigen* library, the simulation was performed with a real-time factor above 0.8. The on-line performance are optimized also by modeling

all the agents as (standalone) ROS nodes running C++ code in a parallel fashion on a multi-core processor. In particular, each UAV-ROS-node executes the (distributed) optimization algorithm based on the index (27) except for the leader UAV-ROS-node that is responsible only for the computation of the target loss probability. Please note that the latter requirement is more expensive regarding the attitude optimization needing the determination of the total formation view: for this reason, it is reasonable to assume that the camera mounted on the leader quadrotor is characterized by powerful hardware.

In the following we report the results of a specific simulation (Scenario #1) wherein a formation of five quadrotors is required to track a target that follows a trajectory accounting for two evasion maneuvers that consist in sharp curves, as depicted in Figure 13. In Figure 14 we show the two target position components as regards time in order to highlight their different movements along the x_W -axis: the blue area corresponds to a linear motion with constant velocity, the orange area corresponds to a sinusoidal path and the green area corresponds still to a sinusoidal path but with increased amplitude. Please note that the two last phases aim at simulating to two evasion maneuvers with increasing velocity.

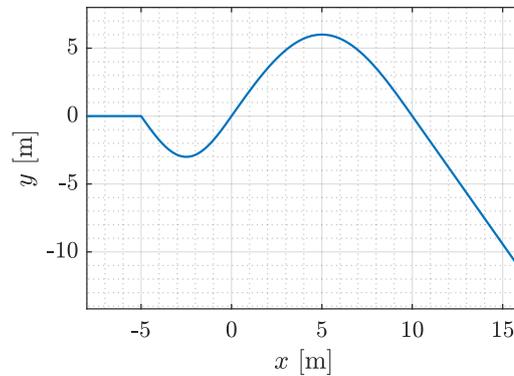


Figure 13. Target trajectory on the (x_W, y_W) plane.

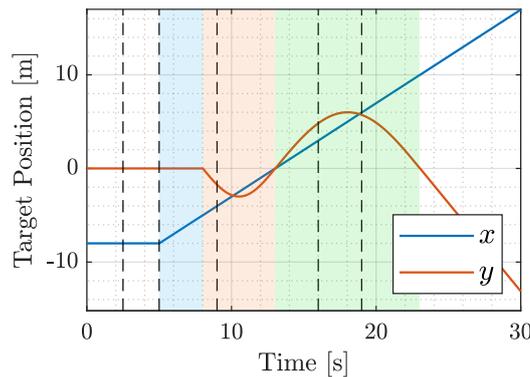


Figure 14. Target trajectory highlighting the position components trend regarding time and the different motion phases.

According to the assumptions stated in Section 2.2, at each time instant $t \geq 0$ the UAVs team is arranged so that the vehicles are aligned along a certain direction $\mathbf{d}(t)$ with a fixed distance d_Q among them and a constant height ζ regarding the ground. In particular, we impose $d_Q = 1$ m, $\zeta = 2$ m and $\mathbf{d}(t) = \mathbf{e}_2$ for all $t \geq 0$. Moreover, regarding the quadrotor leader, namely the ℓ -th agent with $\ell = 3$, we impose $\|\tilde{\mathbf{p}}_{O_{B_\ell}}^W(t) - \mathbf{p}_T^W(t)\| = d_T = 2$ m for all $t \geq 0$. On the other hand, we set $k_\psi = 1.65 \cdot 10^{-4}$ and $k_\rho = 0.5$: note that high values for the former gain improve the angle optimization convergence,

although this is limited by the dynamics of the underlying position controller, whereas high values for the latter gain entails the total formation f.o.v. increase. As regards the target loss probability computation, we chose $T_s = 0.01$ s and $r = 20$: the choice of the prediction horizon (resulting from the product rT_s) is critical and has to take into account the velocity of all the agents involved in the tracking task.

Figure 15a shows the initial conditions for the simulation, highlighting the positions of the UAVs that are supposed to be reached after an initial takeoff and formation stabilization phase. At the beginning of the test, the target is stopped and visible from all the cameras in the group, hence the predicted target loss probability is almost zero. This implies the possibility to progressively increase the overlapping area of all the cameras views, thus reducing the target position estimate error. This phenomenon is observable in Figure 15b,c: note that the target remains stopped and visible from all the cameras whose views overlapping is maximized slightly before $t = 5$ s. At $t = 5$ s, the target starts moving along the x_W -axis direction with constant velocity (blue phase in Figure 14). The UAVs formation reacts: all the quadrotors accelerate in the same direction and, after a transient phase, the target is still visible from all the cameras. At $t = 8$ s the target executes the first evasion maneuver following a sinusoidal path (orange phase in Figure 14). To track this drifting movement, all the quadrotors are required to increase their roll angles, causing a temporary deformation of their f.o.v. projection on the target plane. During this phase, the target loss probability increases until its maximum value as reported in Figure 16a-top, nonetheless it is reasonable to assume that this value is higher regarding the real one since the Kalman filter is not capable of handling abrupt changes in the cameras f.o.v. approximation. Besides this fact, the formation is still able to track the target moving at a low speed. At $t = 13$ s the target perform a second evasion maneuver continuing to follow a sine path but with augmented velocity (green phase in Figure 14). As a consequence, the target loss probability increases again and correspondingly the total formation f.o.v.. To observe this fact, it is suitable to compare Figure 16a-top and Figure 16b-bottom reporting the target loss probability trend and the value of the monitored area, respectively. The total formation f.o.v. results to be maximized at $t = 16$ s, when the target loss probability boils down to (almost) zero. At the same time instant the target decelerates and the cameras views increases their overlapping area (Figure 15e). At $t = 20$ s the target increases its speed again and keeps it constant: the total formation f.o.v. is finally increased until a balance between coverage and tracking is met, according to the chosen parameters and the target speed. A video of the described sequence is available at [47].

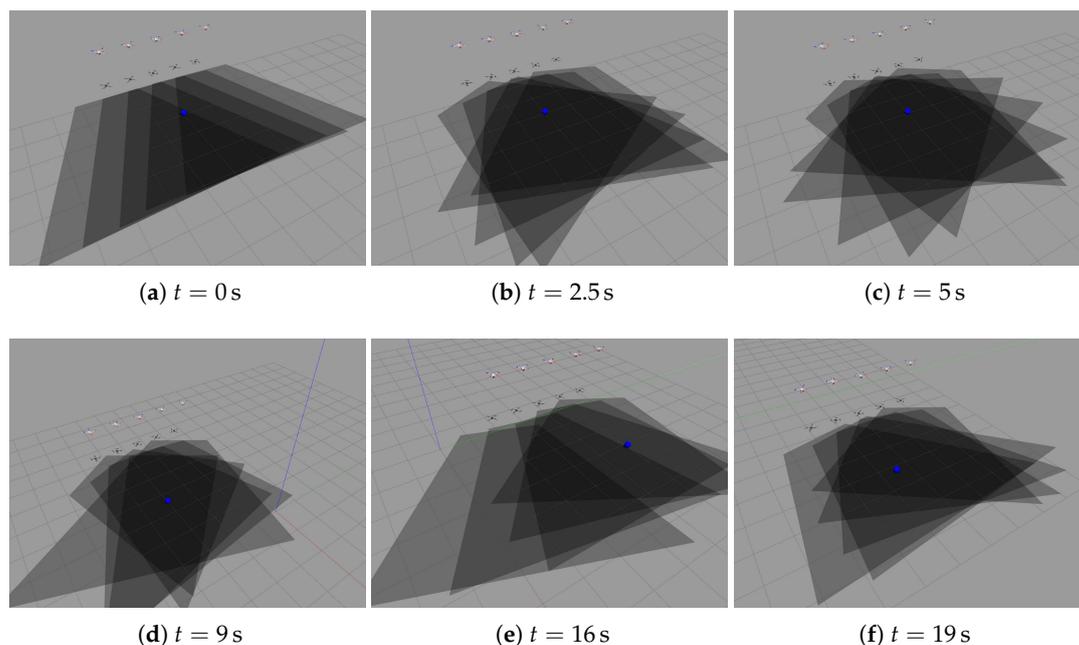


Figure 15. Gazebo simulation snapshots acquired in correspondence to the dashed lines in Figure 14.

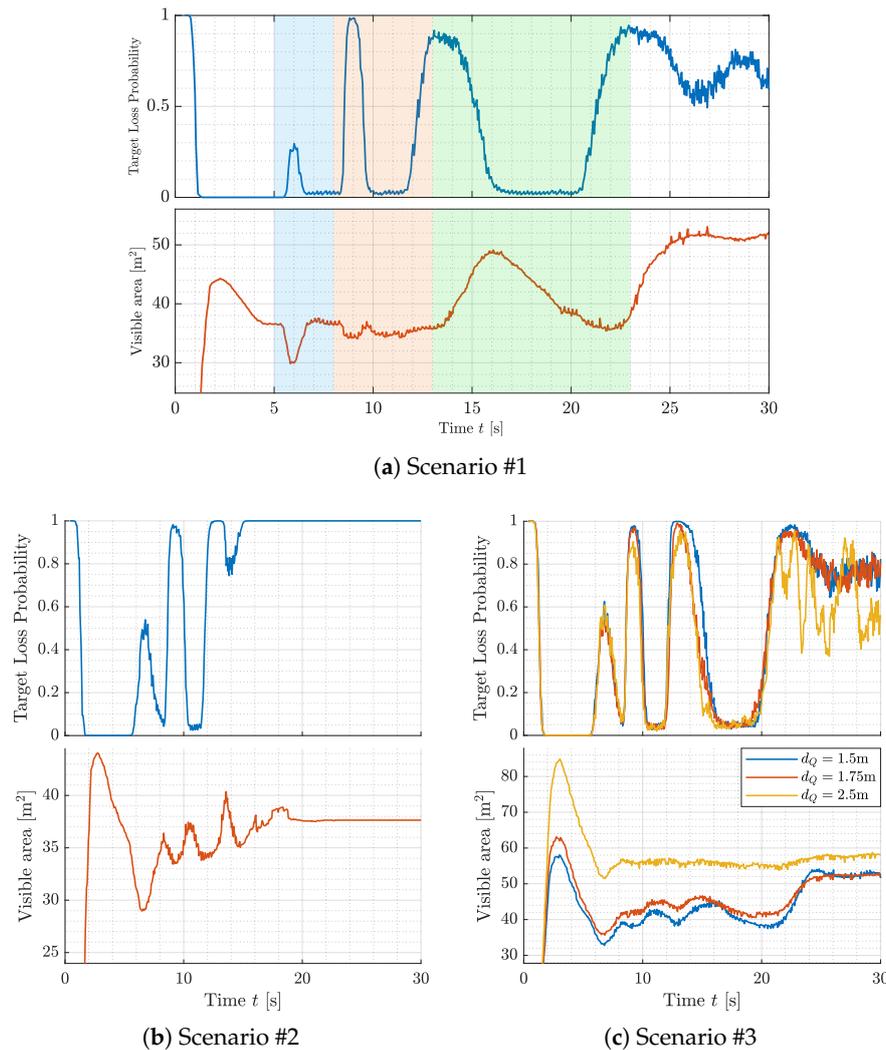


Figure 16. The three considered scenarios: (a) Target evasion maneuver with the proposed multi-objective optimization; (b) Target evasion maneuver without multi-objective optimization; (c) Target evasion maneuver with different formation configurations. All panels report the target loss probability (top) and the total visible area (bottom).

Comparative Experiments

In order to validate the proposed algorithm, two other scenarios have been considered.

Firstly, the same evasion maneuver of Scenario #1 has been tested without the optimization of the yaw angle - that is, with all the cameras looking towards the target and therefore with maximum overlapping (Scenario #2). The results of this case are reported in Figure 16b. It occurs that at about $t = 13$ s, in correspondence to the second strong evasion movement, the target is lost and hence without a reference the formation stops. In doing this, the UAVs perform a rolling movement that accidentally will allow the target to slightly enter some f.o.v. for a few instants: this motivates the small drop in the target-loss probability right afterwards. However, this is caused by the vehicles stopping dynamics and there is no way to exploit this behavior to recover the tracking of the moving target. Conversely, by comparing Figure 16b with Figure 16a it can be noticed that around the same time the loss probability increases also in the optimized case but in this case the algorithm is able to manage the issue and keep tracking the target.

Secondly, in Figure 16c we show the results of the proposed algorithms with different linear formation configurations (Scenario #3). Indeed in this context we remark the fact that formation control and cooperative visual tracking are aspects of the same problem that can be managed through nested

control loops in order to solve the complexity of the distributed solution: given a specific formation in terms of mutual relative positions among agents and vertical distances from the target plane (slow outer loop, not considered in this work), visual tracking optimization can be obtained by leveraging a local control action (fast inner loop, discussed here). In particular, Figure 16c reports the behaviors for the target loss probability and total visible area and it can be well appreciated how the formation features enter the performance of the visual tracking task: while with $d_Q < 2$ m (blue and red curves) a dynamics of the visible area is well in relation to the target loss probability, if the configuration is more spread ($d_Q = 2.5$ m, dark yellow curve) the target loss probability exhibits a more nervous trend while the visible area is basically constant and maximized. This is suggestive of a more static behavior of the cameras f.o.v. and hence of a less accurate *focus on event* ability, being the number of cameras facing the event low and constant; actually, by decoupling formation control from visual optimization, we gain a better control on this feature and we can regulate formation so as to control the information flow from the scene.

6. Conclusions

This work presents a cooperative strategy to optimize the overall view performance of a UAVs formation that is tracking a moving target, accounting for target pose reconstruction quality and, at the same time, for the probability of keeping the target in view. Being the two objectives in opposition, a method to obtain a trade-off has been introduced to favor one of the two only when necessary.

The procedure is based on an approximation of the visible region on the ground plane of each agent, which requires the knowledge of the camera projection matrix, resting upon the pin-hole camera model. For the purpose of reconstruction quality, the views must be as much overlapping as possible, and this is obtained by considering, for each agent, a neighboring camera and then minimizing a suitable defined index, namely the imbalance factor, with the aim of having the overlapping visible portion on the right and on the left side regarding the projection of the optical axis of each camera on the target plane. A larger monitored area reduces the probability that the target is able to evade the surveillance of all the cameras. However, this is needed only when the target is moving fast on the boundary of the total formation f.o.v.. To this aim, a simple double integrator has been considered as the dynamics model for the target, which allows the use of a Kalman based predictor to estimate the probability distribution of the target position after a certain period and, consequently, the so-called target loss probability. The imbalance factor is then corrected by an amount proportional to the latter, in a way that, when the probability is high, the minimum of the cost term is achieved when the neighboring cameras are not fully overlapping.

Since the optimization method is based on several heuristics and approximations, the assessment on a realistic simulation is required. The simulation campaign is performed within the Gazebo framework and the algorithms are implemented using ROS, validating the efficacy of the proposed solution, even in case of unexpected evasion maneuvers of the target and in comparison with a non-optimized approach where the target occurs to be lost.

Author Contributions: Conceptualization, N.L., M.G., A.F. and A.C.; Formal Analysis, N.L., G.M., R.A., M.G., A.F. and A.C.; Methodology, N.L., M.G. and A.F.; Project Administration, A.C.; Software, N.L.; Supervision, G.M. and A.C.; Writing—Original Draft, N.L. and G.M.; Writing—Review & Editing, R.A. and A.C.

Funding: This research received no external funding.

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

References

1. Nonami, K.; Kendoul, F.; Suzuki, S.; Wang, W.; Nakazawa, D. *Autonomous Flying Robots: Unmanned Aerial Vehicles and Micro Aerial Vehicles*; Springer Science & Business Media: Berlin, Germany, 2010.
2. Austin, R. *Unmanned Aircraft Systems: UAVS Design, Development And Deployment*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 54.

3. Gupte, S.; Mohandas, P.I.T.; Conrad, J.M. A survey of quadrotor unmanned aerial vehicles. In Proceedings of the 2012 Proceedings of IEEE Southeastcon, Orlando, FL, USA, 15–18 March 2012; pp. 1–6.
4. Mahony, R.; Kumar, V.; Corke, P. Multirotor aerial vehicles: Modeling, estimation, and control of quadrotor. *IEEE Robot. Autom. Mag.* **2012**, *19*, 20–32. [[CrossRef](#)]
5. Kumar, V.; Michael, N. Opportunities and challenges with autonomous micro aerial vehicles. *Int. J. Robot. Res.* **2012**, *31*, 1279–1291. [[CrossRef](#)]
6. Valavanis, K.P.; Vachtsevanos, G.J. *Handbook of Unmanned Aerial Vehicles*; Springer: Berlin, Germany, 2015.
7. Zaheer, Z.; Usmani, A.; Khan, E.; Qadeer, M.A. Aerial surveillance system using UAV. In Proceedings of the 2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN), Hyderabad, Telangana State, India, 21–23 July 2016; pp. 1–7.
8. Sun, J.; Li, B.; Jiang, Y.; Wen, C.y. A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes. *Sensors* **2016**, *16*, 1778. [[CrossRef](#)] [[PubMed](#)]
9. Al-Kaff, A.; Gómez-Silva, M.J.; Moreno, F.M.; de la Escalera, A.; Armingol, J.M. An Appearance-Based Tracking Algorithm for Aerial Search and Rescue Purposes. *Sensors* **2019**, *19*, 652. [[CrossRef](#)] [[PubMed](#)]
10. Khamseh, H.B.; Janabi-Sharifi, F.; Abdessameud, A. Aerial manipulation—A literature survey. *Robot. Auton. Syst.* **2018**, *107*, 221–235. [[CrossRef](#)]
11. Ruggiero, F.; Lippiello, V.; Ollero, A. Aerial manipulation: A literature review. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1957–1964. [[CrossRef](#)]
12. Sahingoz, O.K. Mobile networking with UAVs: Opportunities and challenges. In Proceedings of the 2013 International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta, GA, USA, 28–31 May 2013; pp. 933–941.
13. Hou, Z.; Wang, W.; Zhang, G.; Han, C. A survey on the formation control of multiple quadrotors. In Proceedings of the 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, South Korea, 28 June–1 July 2017; pp. 219–225.
14. Chung, S.J.; Paranjape, A.A.; Dames, P.; Shen, S.; Kumar, V. A survey on aerial swarm robotics. *IEEE Trans. Robot.* **2018**, *34*, 837–855. [[CrossRef](#)]
15. Liu, Y.; Bucknall, R. A survey of formation control and motion planning of multiple unmanned vehicles. *Robotica* **2018**, *36*, 1019–1047. [[CrossRef](#)]
16. Gu, J.; Su, T.; Wang, Q.; Du, X.; Guizani, M. Multiple moving targets surveillance based on a cooperative network for multi-UAV. *IEEE Commun. Mag.* **2018**, *56*, 82–89. [[CrossRef](#)]
17. Tan, Y.H.; Lai, S.; Wang, K.; Chen, B.M. Cooperative control of multiple unmanned aerial systems for heavy duty carrying. *Ann. Rev. Control* **2018**, *46*, 44–57. [[CrossRef](#)]
18. Li, B.; Jiang, Y.; Sun, J.; Cai, L.; Wen, C.Y. Development and testing of a two-UAV communication relay system. *Sensors* **2016**, *16*, 1696. [[CrossRef](#)] [[PubMed](#)]
19. Kanistras, K.; Martins, G.; Rutherford, M.J.; Valavanis, K.P. Survey of unmanned aerial vehicles (UAVs) for traffic monitoring. *Handbook of Unmanned Aerial Vehicles*; Springer: Dordrecht, the Netherlands. **2015**; pp. 2643–2666.
20. Yanmaz, E. Event detection using unmanned aerial vehicles: Ordered versus self-organized search. In *International Workshop on Self-Organizing Systems*; Springer: Berlin, Germany, 2009; pp. 26–36.
21. Zhao, J.; Xiao, G.; Zhang, X.; Bavirisetti, D.P. A Survey on Object Tracking in Aerial Surveillance. In *International Conference on Aerospace System Science and Engineering*; Springer: Berlin, Germany, 2018; pp. 53–68.
22. Schwager, M.; Julian, B.J.; Rus, D. Optimal coverage for multiple hovering robots with downward facing cameras. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 3515–3522.
23. Doitsidis, L.; Weiss, S.; Renzaglia, A.; Achtelik, M.W.; Kosmatopoulos, E.; Siegwart, R.; Scaramuzza, D. Optimal surveillance coverage for teams of micro aerial vehicles in GPS-denied environments using onboard vision. *Auton. Robots* **2012**, *33*, 173–188. [[CrossRef](#)]
24. Saska, M.; Chudoba, J.; Přeučil, L.; Thomas, J.; Loianno, G.; Třešňák, A.; Vonásek, V.; Kumar, V. Autonomous deployment of swarms of micro-aerial vehicles in cooperative surveillance. In Proceedings of the 2014 International Conference on Unmanned Aircraft Systems (ICUAS), Orlando, FL, USA, 27–30 May 2014; pp. 584–595.

25. Mavrinnac, A.; Chen, X. Modeling coverage in camera networks: A survey. *Int. J. Comput. Vis.* **2013**, *101*, 205–226. [[CrossRef](#)]
26. Ganguli, A.; Cortés, J.; Bullo, F. Maximizing visibility in nonconvex polygons: nonsmooth analysis and gradient algorithm design. *SIAM J. Control Optim.* **2006**, *45*, 1657–1679. [[CrossRef](#)]
27. Chevet, T.; Maniu, C.S.; Vlad, C.; Zhang, Y. Voronoi-based UAVs Formation Deployment and Reconfiguration using MPC Techniques. In Proceedings of the 2018 International Conference on Unmanned Aircraft Systems (ICUAS), Dallas, TX, USA, 12–15 June 2018; pp. 9–14.
28. Munguía, R.; Urzua, S.; Bolea, Y.; Grau, A. Vision-based SLAM system for unmanned aerial vehicles. *Sensors* **2016**, *16*, 372. [[CrossRef](#)] [[PubMed](#)]
29. Schmuck, P.; Chli, M. Multi-uav collaborative monocular slam. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3863–3870.
30. Robin, C.; Lacroix, S. Multi-robot target detection and tracking: Taxonomy and survey. *Auton. Robots* **2016**, *40*, 729–760. [[CrossRef](#)]
31. Evans, M.; Osborne, C.J.; Ferryman, J. Multicamera object detection and tracking with object size estimation. In Proceedings of the 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, Krakow, Poland, 27–30 August 2013; pp. 177–182.
32. Xu, B.; Bulan, O.; Kumar, J.; Wshah, S.; Kozitsky, V.; Paul, P. Comparison of early and late information fusion for multi-camera HOV lane enforcement. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, Spain, 15–18 September 2015; pp. 913–918.
33. Coates, A.; Ng, A.Y. Multi-camera object detection for robotics. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 3–7 May 2010; pp. 412–419.
34. Del Rosario, J.R.B.; Bandala, A.A.; Dadios, E.P. Multi-view multi-object tracking in an intelligent transportation system: A literature review. In Proceedings of the 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, Philippines, 1–3 December 2017; pp. 1–4.
35. Poesi, F.; Cavallaro, A. Distributed vision-based flying cameras to film a moving target. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 2453–2459.
36. Pounds, P.; Mahony, R.; Corke, P. Modelling and control of a large quadrotor robot. *Control Eng. Pract.* **2010**, *18*, 691–699. [[CrossRef](#)]
37. Michael, N.; Mellinger, D.; Lindsey, Q.; Kumar, V. The GRASP multiple micro-UAV testbed. *IEEE Robot. Autom. Mag.* **2010**, *17*, 56–65. [[CrossRef](#)]
38. Kalabić, U.; Gupta, R.; Di Cairano, S.; Bloch, A.; Kolmanovsky, I. MPC on Manifolds with an Application to SE (3). In Proceedings of the 2016 American Control Conference (ACC), Boston, MA, USA, 6–8 July 2016; pp. 7–12.
39. Garcia, G.A.; Kim, A.R.; Jackson, E.; Kashmiri, S.S.; Shukla, D. Modeling and flight control of a commercial nano quadrotor. In Proceedings of the 2017 International Conference on Unmanned Aircraft Systems (ICUAS), Miami, FL, USA, 13–16 June 2017; pp. 524–532.
40. Lee, T.; Leok, M.; McClamroch, N.H. Geometric tracking control of a quadrotor UAV on SE(3). In Proceedings of the 49th IEEE conference on decision and control (CDC), Atlanta, GA, USA, 15–17 December 2010; pp. 5420–5425.
41. Lee, T. Geometric tracking control of the attitude dynamics of a rigid body on SO(3). In Proceedings of the 2011 American Control Conference, San Francisco, CA, USA, 29 June–1 July 2011; pp. 1200–1205.
42. Ma, Y.; Soatto, S.; Kosecka, J.; Sastry, S.S. *An Invitation to 3-D Vision: From Images to Geometric Models*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 26.
43. Schiano, F.; Franchi, A.; Zelazo, D.; Giordano, P.R. A rigidity-based decentralized bearing formation controller for groups of quadrotor UAVs. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, South Korea, 9–14 October 2016; pp. 5099–5106.
44. Kamthe, A.; Jiang, L.; Dudys, M.; Cerpa, A. Scopes: Smart cameras object position estimation system. In *European Conference on Wireless Sensor Networks*; Springer: Berlin, Germany, 2009; pp. 279–295.
45. Masiero, A.; Cenedese, A. On triangulation algorithms in large scale camera network systems. In Proceedings of the 2012 American Control Conference (ACC), Montreal, QC, Canada, 27–29 June 2012; pp. 4096–4101.

46. Fu, X.; Liu, K.; Gao, X. Multi-UAVs communication-aware cooperative target tracking. *Appl. Sci.* **2018**, *8*, 870. [[CrossRef](#)]
47. Lissandrini, N.; Michieletto, G.; Antonello, R.; Galvan, M.; Franco, A.; Cenedese, A. Cooperative Optimization of UAVs Formation Visual Tracking. 2019. Available online: <https://youtu.be/MXV0cQ4qmRk> (accessed on 25 June 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).