aerospace

MDPI

Article

Image-Based Visual Servoing for Quadrotor Formation Encirclement and Tracking of Unknown Targets

Hanyu Guo [1], Tao Song [1], Jianchuan Ye [1,*], Yusra Abdulrahman [2,*], Xuechen Gu [1], Tao Jiang [3] and Yihao Dong [4]

[1] School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China
[2] Advanced Research and Innovation Center (ARIC), Department of Aerospace Engineering, Khalifa University, Abu Dhabi 127788, United Arab Emirates
[3] School of Automation, Chongqing University, Chongqing 400044, China
[4] Center for Autonomous Robotic Systems, Khalifa University of Science and Technology, Abu Dhabi 127788, United Arab Emirates
* Correspondence: yejianchuan@yeah.net (J.Y.); yusra.abdulrahman@ku.ac.ae (Y.A.)

Abstract

A target state estimation method based on multiple quadrotors is proposed for unknown maneuvering targets, and a distributed formation control method Image-Based Visual Servoing (IBVS) is also proposed to achieve encirclement tracking of unknown maneuvering targets. In the tracking control, collision avoidance constraints for nodes within the formation are also introduced, and based on the shared position information within the formation, the positions of other nodes within the Field of View (FOV) of each node are predicted for detecting unknown targets. Firstly, an Interacting Multiple Model (IMM) was designed based on multiple motion modes to estimate the position and velocity of the target. A virtual camera coordinate system containing translational and yaw rotations was established between the formation and the target based on the estimated values. Then, a distributed control method based on IBVS was further designed by combining image deviation. At the same time, a safe distance between nodes within the formation was introduced, and collision avoidance constraints of the Control Barrier Function (CBF) were designed. Finally, the position of the formation nodes within the FOV was predicted. The simulation results demonstrate that, utilizing the proposed estimation method, the estimation accuracy for target velocity improves by 26.5% in terms of Root Mean Square Error (RMSE) compared to existing methods. Furthermore, the proposed control method enables quadrotor formations to successfully achieve encirclement tracking of unknown maneuvering targets, significantly reducing tracking errors in comparison to conventional approaches.

Keywords: quadrotor formation; image-based visual servoing; distributed control; the interacting multiple model filter; a control barrier function

Check for updates

Academic Editor: Gokhan Inalhan

Received: 19 December 2025
Revised: 12 January 2026
Accepted: 21 January 2026
Published: 30 January 2026

Copyright: © 2026 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.

1. Introduction

With the further enhancement of the intelligence and autonomy of quadrotors, as well as the increasing exploration of the potential of multi-agent systems across various industries, low-cost and efficient autonomous control methods for quadrotor formations have attracted growing attention from enthusiasts and researchers [1–5]. In scenarios involving the tracking of unknown aerial maneuvering targets, accurate perception of the target's motion state is indispensable [6,7]. For lightweight small-scale quadrotors, lightweight onboard strapdown cameras are often used for environmental and target state

perception [8]. However, a single camera only provides two-dimensional information of the target, making it difficult for the quadrotors to obtain an accurate estimate of the target's three-dimensional motion. Moreover, the maneuverability of the target can easily cause it to move out of the quadrotor's camera FOV, thereby hindering effective autonomous tracking [9]. The use of stereo cameras can partially alleviate this issue [10,11], yet, constrained by small quadrotor platforms, accurately estimating the motion state of highly maneuverable and distant targets remains challenging [12]. Employing formations of small quadrotors can better adapt to the aforementioned scenarios [13]. Theoretically, the perception of the target is no longer limited by the size of a single quadrotor platform, enabling the tracking of highly maneuverable targets. Furthermore, by designing specific formations [11], the target can be more effectively observed by multiple onboard cameras—for instance, by positioning the target's image near the center of the FOV of cameras oriented from different directions and locations.

Accurately estimating the target velocity based on observation data from quadrotor formations is essential for formation tracking control. Common methods include direct differentiation of position measurements [14] and the design of Extended Kalman Filters (EKF) [15]. While direct differentiation may be feasible for estimating low-speed targets with stable motion, it is susceptible to noise interference [16]. The EKF can satisfy most application scenarios but may introduce significant estimation errors when the target's motion mode changes [17]. For targets exhibiting abrupt motion changes, the IMM filter may offer better adaptability for state estimation [18]. The essence of the IMM is similar to that of the EKF. However, when the target's motion mode changes, the IMM adaptively adjusts the estimation gain coefficients based on the matching degree of different motion models, thereby reducing estimation errors caused by target motion variations [19].

For target tracking scenarios, lightweight and small quadrotors often employ vision-based servo control methods, which are primarily categorized into Position-Based Visual Servoing (PBVS) and IBVS [20]. PBVS constructs the three-dimensional positional relationship relative to the target based on two-dimensional image information from the camera and subsequently designs the control law using this established relative positional relationship [21]. Consequently, the design of its control law is generally not constrained by the limitations of two-dimensional image errors, and it often disregards FOV constraints. However, its control accuracy and stability can be affected by modeling uncertainties, camera calibration errors, and noise [22]. In contrast, IBVS directly designs the control law based on two-dimensional image errors, thereby reducing the influence of modeling uncertainties, camera calibration errors, and noise to a relatively lesser extent [22], while also more readily guiding the target toward the center of the FOV.

In general, the primary objective of formation control schemes is to enable a group of agents to achieve predefined relative positions, distances, and maintain a desired configuration (typically geometric) under certain inter-agent constraints [23]. For instance, Wang [24] designed a circular formation comprising three quadrotors around a target to facilitate localization and motion planning for micro-drones. Karras [25] proposed an encirclement formation of four quadrotors for capturing maneuvering targets. Rothe [26] devised a specialized quadrotor formation for net-based capture of targets. Wu [27] developed a specific quadrotor formation for maritime strike and defense operations. Montijano [28] designed several formation configurations for quadrotors to investigate the influence of formations on observation performance. The authors further noted that a commonly employed distributed formation control strategy is consensus-based control, in which states such as position and orientation of each agent converge collectively to their respective desired values. This approach offers robustness against a variety of interaction topologies within the team.

We propose a distributed formation control method for quadrotors based on IBVS, designed to generate an approaching and encircling formation for tracking an unknown maneuvering target. This method enables stable tracking of the target while preventing collisions among formation members during the process. The main contributions of this work are as follows:

Multi-View Target Estimation via IMM Filtering.

Using onboard strapdown cameras on multiple quadrotors, the unknown target is localized from multiple viewpoints. Considering both constant velocity (CV) and constant acceleration (CA) motion modes, an IMM filter is designed to estimate the target's position and velocity, and the estimated states are utilized to compensate for the control input.

Distributed IBVS Control with Virtual Camera Coordinates.

By establishing a virtual camera coordinate system that incorporates the quadrotor's own attitude, the influence of pitch and roll motions on the target's image features is eliminated. Based on the virtual image error and the estimated target states, a distributed IBVS control law is designed using graph connectivity theory. The stability of the proposed method is analyzed via Lyapunov-based techniques, enabling stable formation-based encircling and tracking of the unknown target while driving the target's image toward the center of the FOV.

Collision Avoidance and Formation Prediction of Quadrotors.

Taking into account the minimum safe distance among quadrotors within the formation, CBFs are incorporated with the motion model to derive linearized collision avoidance constraints. A quadratic cost function is formulated, leading to an optimal control law that satisfies all constraints. Moreover, by integrating the estimated target states with each quadrotor's own attitude, the image positions of other formation members in the camera field are predicted, thereby preventing misidentification of fellow quadrotors as the target.

Numerical Simulation and Validation.

Numerical simulations are conducted for formation-based encircling and tracking of an unknown maneuvering target. The results demonstrate that the adopted IMM estimator improves target velocity estimation accuracy by 23% compared to existing methods. The proposed control scheme achieves stable encircling and tracking of the target by a quadrotor formation, with significantly reduced tracking errors relative to state-of-the-art approaches.

The remainder of this paper is organized as follows. Section 2 introduces the necessary preliminaries. Section 3 details the design process of the estimator and controller. Simulation experiments are presented in Section 4, followed by the conclusion in the final section.

2. Preliminaries

2.1. Coordinate Frame Description

As illustrated in Figure 1, the system involves multiple coordinate frames. Transformations between different frames m and n are performed via the rotation matrix $\mathbf{R}_m^n$ The Earth-fixed coordinate system (ECS) is denoted as frame $\{O_e - x_e y_e z_e\}$. The body-fixed coordinate system (BCS) is denoted as frame $\{O_b - x_b y_b z_b\}$. The camera coordinate system (CCS), denoted as frame $\{O_c - x_c y_c z_c\}$, is typically assumed to coincide with the BCS for simplicity of description. Frame $\{O_v - x_v y_v z_v\}$ represents the virtual camera coordinate system (VCS). Its origin coincides with O_c and it is established to simplify the description of the projection relationship between the translational motion of the target in the ECS and the image, without being influenced by camera attitude. The position of the target in the ECS is denoted as T. The coordinates (u_{vt}, v_{vt}) represent the position of T in the virtual image, while (u_{rt}, v_{rt}) represents the position of T in the real image.

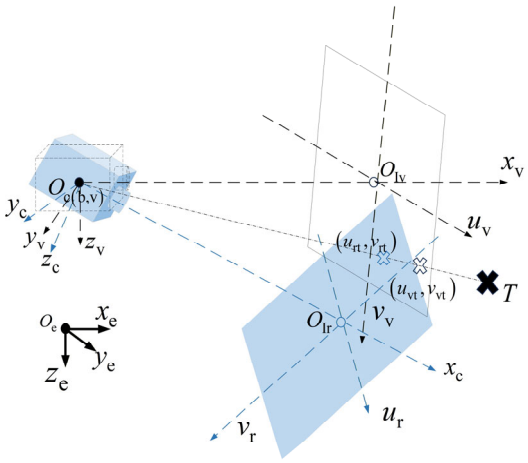


Figure 1. System framework and coordinate system diagram.

2.2. Quadrotor Model

$$\begin{aligned} \dot{\boldsymbol{p}} &= \boldsymbol{v} \\ \dot{\boldsymbol{v}} &= \boldsymbol{g} + \tfrac{1}{m}(\boldsymbol{F} + \boldsymbol{F}_{\mathrm{d}}) \\ \dot{\boldsymbol{R}}_{\mathrm{b}}^{\mathrm{e}} &= \boldsymbol{R}_{\mathrm{b}}^{\mathrm{e}}\boldsymbol{\omega}_{\times} \\ \boldsymbol{J}\dot{\boldsymbol{\omega}} &= \boldsymbol{\tau} - \boldsymbol{\omega} \times \boldsymbol{J}\boldsymbol{\omega} + \boldsymbol{\tau}_d \end{aligned} \tag{1}$$

Here, $\boldsymbol{p} = \begin{bmatrix} p_x & p_y & p_z \end{bmatrix}^{\mathrm{T}}$ and $\boldsymbol{v} = \begin{bmatrix} v_x & v_y & v_z \end{bmatrix}^{\mathrm{T}}$ denote the velocity and position of the quadrotor in the Earth-fixed coordinate system, respectively, satisfying $\boldsymbol{v} = \boldsymbol{R}_b^e \mathrm{v_b}$, where $\boldsymbol{v}_{\mathrm{b}} = \begin{bmatrix} v_{\mathrm{b}x} & v_{\mathrm{b}y} & v_{\mathrm{b}z} \end{bmatrix}^{\mathrm{T}}$ represents the velocity of the quadrotor in the BCS. $\boldsymbol{g} = \begin{bmatrix} 0 & 0 & g_G \end{bmatrix}^{\mathrm{T}}$ and g_G denotes the gravitational constant. $\boldsymbol{R}_b^e \in \mathbb{R}^{3\times3}$ denotes the rotation matrix from the BCS to the ECS. $\boldsymbol{\omega} = \begin{bmatrix} \omega_x & \omega_y & \omega_z \end{bmatrix}^{\mathrm{T}}$ represents the angular velocity of the quadrotor, and $\boldsymbol{\omega}_{\times}$ denotes the skew-symmetric matrix generated by $\boldsymbol{\omega}$. $\boldsymbol{F} \in \mathbb{R}^3$ and $\boldsymbol{\tau} \in \mathbb{R}^3$ denote the thrust force and torque, respectively, while $\boldsymbol{F}_{\mathrm{d}} \in \mathbb{R}^3$ and $\boldsymbol{\tau}_{\mathrm{d}} \in \mathbb{R}^3$ represent the disturbance force and disturbance torque, respectively. Additionally, $m \in \mathbb{R}$ and $\boldsymbol{J} \in \mathbb{R}^{3\times3}$ denote the mass and moment of inertia of the quadrotor, respectively.

2.3. IBVS Framework

For clarity of presentation, it is assumed that the CCS coincides with the BCS of the quadrotor, leading to the relation $\boldsymbol{v}_{\mathrm{c}} = \boldsymbol{v}_{\mathrm{b}}$. A pinhole camera model is adopted, where the image-space position of quadrotor i is given by $\boldsymbol{s} = \begin{bmatrix} u & v \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} f p_{\mathrm{c}y}/p_{\mathrm{c}x} & f p_{\mathrm{c}z}/p_{\mathrm{c}x} \end{bmatrix}^{\mathrm{T}}$. Here, f denotes the camera focal length, and $\boldsymbol{p}_{\mathrm{c}} = \begin{bmatrix} p_{\mathrm{c}x} & p_{\mathrm{c}y} & p_{\mathrm{c}z} \end{bmatrix}^{\mathrm{T}}$ represents the relative position in the camera frame. Based on its position in the ECS, the dynamics of the image feature $\dot{\boldsymbol{s}}$ can be expressed via the interaction matrix $\boldsymbol{L}_{\mathrm{sr}} \in \mathbb{R}^{2\times6}$ [29] and the velocity vector $\boldsymbol{v}_{\mathrm{sr}} = \begin{bmatrix} v_{\mathrm{c}}^{\mathrm{T}} & \omega^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ as follows:

$$\dot{\boldsymbol{s}} = \boldsymbol{L}_{\mathrm{sr}}\boldsymbol{v}_{\mathrm{sr}} \tag{2}$$

The specific form of $\boldsymbol{L}_{\mathrm{sr}}$ is given as follows:

$$\boldsymbol{L}_{\mathrm{sr}} = \begin{bmatrix} \frac{u}{p_{\mathrm{c}x}} & -\frac{f}{p_{\mathrm{c}x}} & 0 & v & -\frac{f^2+u^2}{f} & \frac{uv}{f} \\ \frac{v}{p_{\mathrm{c}x}} & 0 & -\frac{f}{p_{\mathrm{c}x}} & -u & \frac{f^2+v^2}{f} & -\frac{uv}{f} \end{bmatrix}$$

The desired relative position in the ECS corresponds to the desired position $\boldsymbol{s}^*$ in the image space. The IBVS control law is typically designed as $\boldsymbol{v}_{\mathrm{rd}} = -\gamma \hat{\boldsymbol{L}}_{sr}^{+}(\boldsymbol{s} - \boldsymbol{s}^*)$, where $\gamma > 0$ is a

gain coefficient, $\boldsymbol{L}_{sr}^{+}$ denotes the Moore–Penrose inverse of $\boldsymbol{L}_{\mathrm{sr}}$, and the depth information p_{cx} required in $\boldsymbol{L}_{\mathrm{sr}}$ needs to be estimated. Even under large initial error conditions, this control law can achieve local asymptotic stability of the closed-loop system [30], i.e., $\boldsymbol{s} \to \boldsymbol{s}^{*}$.

Since the quadrotor is an underactuated system, to decouple the coupling effects of pitch and roll motions on translational motion, the following image model is established based on a virtual camera system:

$$\dot{\boldsymbol{s}}_{\mathrm{v}} = \begin{bmatrix} \dot{u}_{\mathrm{v}} \\ \dot{v}_{\mathrm{v}} \end{bmatrix} = \boldsymbol{L}_{\mathrm{sv}} \boldsymbol{v}_{sv} = \begin{bmatrix} -\frac{u_{\mathrm{v}}}{p_{\mathrm{v}x}} & \frac{f}{p_{\mathrm{v}x}} & 0 & \frac{f^2+u_{\mathrm{v}}^2}{f} \\ -\frac{v_{\mathrm{v}}}{p_{\mathrm{v}x}} & 0 & \frac{f}{p_{\mathrm{v}x}} & \frac{u_{\mathrm{v}} v_{\mathrm{v}}}{f} \end{bmatrix} \begin{bmatrix} v_{\mathrm{v}x} \\ v_{\mathrm{v}y} \\ v_{\mathrm{c}z} \\ \omega_z \end{bmatrix} \tag{3}$$

where $\boldsymbol{p}_{\mathrm{v}} = \begin{bmatrix} p_{\mathrm{v}x} & p_{\mathrm{v}y} & p_{\mathrm{v}z} \end{bmatrix}^{\mathrm{T}}$ denotes the relative position vector in the VCS, and velocity vector $\boldsymbol{v}_{\mathrm{sv}} = \begin{bmatrix} \boldsymbol{v}_{\mathrm{v}}^{\mathrm{T}} & \omega_z \end{bmatrix}^{\mathrm{T}}$, where $\boldsymbol{v}_{\mathrm{v}}$ represents the velocity in the VCS. Therefore, $\boldsymbol{v}_{\mathrm{vd}} = -\gamma \hat{\boldsymbol{L}}_{\mathrm{sv}}^{+}(\boldsymbol{s} - \boldsymbol{s}^{*})$ can be adopted as the control law based on the selected image model.

2.4. Basic Concepts of Graph Theory

A graph $\mathcal{G}$ consists of a vertex set $\mathcal{V}(\mathcal{G})$ and an edge set $\mathcal{E}(\mathcal{G})$, representing the communication relationships among nodes in this paper, where each node corresponds to a quadrotor. The set of neighboring nodes of vertex i is denoted as $\mathcal{N}_i(\mathcal{G}) \triangleq \{j \in \mathcal{V} | (i,j) \in \mathcal{E}(\mathcal{G})\}$. The Laplacian matrix of graph $\mathcal{G}$ is defined as $\boldsymbol{\mathcal{L}}(\mathcal{G}) \triangleq \boldsymbol{D} - \boldsymbol{A} \in \mathbb{R}^{N \times N}$, where $\boldsymbol{A} = \{a_{ij}\}_{N \times N}$ denotes the adjacency matrix of $\mathcal{G}$ with elements $a_{ij} = 1$ if $i \neq j$ and $(i,j) \in \mathcal{E}(\mathcal{G})$, and $a_{ij} = 0$ otherwise. $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_N)$ is the degree matrix with $d_i \triangleq \sum_{j=1}^{N} a_{ij}$. Since the graph considered in this paper is undirected, $\boldsymbol{\mathcal{L}}(\mathcal{G})$ is a symmetric positive semidefinite matrix [30].

Consider a multi- quadrotor system composed of N quadrotors (nodes), whose dynamic model is given by Equation (1). The joint state vector formed by the positions of all nodes is denoted as $\boldsymbol{p}_q = \begin{bmatrix} \boldsymbol{p}_1^{\mathrm{T}} & \ldots & \boldsymbol{p}_N^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$, and the distance between any two nodes i and j is given by $\delta_{ij} = \left\| \boldsymbol{p}_i - \boldsymbol{p}_j \right\|$.

Assumption 1. *The graph $\mathcal{G}$ considered in this paper is connected.*

2.5. Problem Statement

This paper aims to employ multiple quadrotors, modeled as in Equation (1), to achieve formation-based encirclement tracking control of an unknown maneuvering target.

In the image space, based on the IBVS control law, the image position of the maneuvering target is driven to converge to the center of the camera's FOV, i.e., $\boldsymbol{s} \to \boldsymbol{s}^{*}$. Simultaneously, the image positions of quadrotors are predicted and marked within the FOV to prevent misidentifying a quadrotor as the maneuvering target.

In the ECS, through a distributed control scheme, the quadrotors are guided to form a uniformly distributed encircling formation around the target, i.e., $\delta_{ij} \to p_{\mathrm{d}}$, where $p_{\mathrm{d}} \in \mathbb{R}^{+}$ denotes the desired inter-agent distance. Meanwhile, collisions between quadrotors must be avoided, i.e., $\delta_{ij} > h_{\mathrm{c}}$, where $h_{\mathrm{c}} > 0$ represents the safety distance.

3. Control Method Design

The proposed control framework is illustrated in Figure 2.

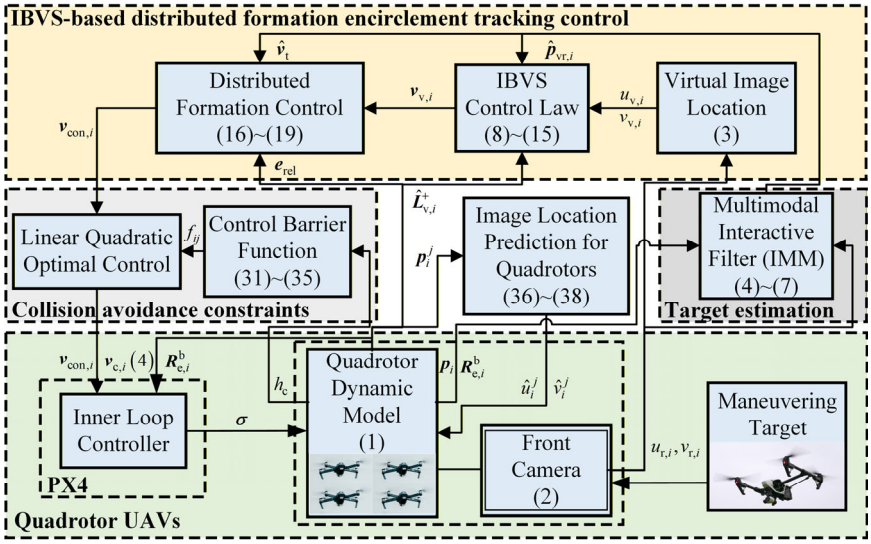


Figure 2. Schematic diagram of control system framework.

3.1. Target Velocity Estimation via IMM Filter

An Interacting Multiple Model (IMM) filter is a hybrid estimation framework that incorporates m kinematic models, each corresponding to an independent Kalman Filter. The switching among the models is governed by a Markov chain, with model probabilities μ_k^j updated recursively over time [17].

The IMM filter generally operates in two phases: the filtering phase and the mixing phase. In the filtering phase, each filter performs prediction and update (the Kalman filtering process) on the state $\hat{\boldsymbol{x}}_{k-}^i$ based on its corresponding dynamic model and the current measurement $\boldsymbol{z}_k$. In the mixing phase, the estimation results from all models are interactively fused according to their probabilities $\mu_k^{i|j}$. For each model j, the mixed initial state estimate $\hat{\boldsymbol{x}}_k^j$ and covariance $\boldsymbol{P}_k^j$ are computed, preparing the initialization for the next filtering step. The specific procedure is given in Equation (4):

$$\begin{aligned} \hat{\boldsymbol{x}}_k^j &= \sum_{i=1}^{m} \mu_k^{i|j} \hat{\boldsymbol{x}}_{k-}^i \\ \boldsymbol{P}_k^j &= \sum_{i=1}^{m} \mu_k^{i|j} \left(\boldsymbol{P}_{k-}^j + \left[\hat{\boldsymbol{x}}_k^i - \hat{\boldsymbol{x}}_k^j \right] \left[\hat{\boldsymbol{x}}_k^i - \hat{\boldsymbol{x}}_k^j \right]^{\mathrm{T}} \right) \end{aligned} \tag{4}$$

Based on the likelihoods μ_k^j of each model, the overall state estimate $\hat{\boldsymbol{x}}_k$ and overall covariance $\boldsymbol{P}_k$ are obtained by performing weighted fusion of the estimates from all models, as shown in Equation (5):

$$\begin{aligned} \hat{\boldsymbol{x}}_k &= \sum_{j=1}^{m} \mu_k^j \hat{\boldsymbol{x}}_k^j \\ \boldsymbol{P}_k &= \sum_{j=1}^{m} \mu_k^j \left(\boldsymbol{P}_k^j + \left[\hat{\boldsymbol{x}}_k^j - \hat{\boldsymbol{x}}_k \right] \left[\hat{\boldsymbol{x}}_k^i - \hat{\boldsymbol{x}}_k \right]^{\mathrm{T}} \right) \end{aligned} \tag{5}$$

For further details regarding the IMM framework, please refer to [31]. In this work, we adopt two motion models: the CV model and the CA model. The stochastic discrete-time state-space representation of the kinematic models is given by Equation (6):

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{n}_{R,k}, \boldsymbol{n}_{R,k} \sim \mathcal{N}(0, \boldsymbol{R}) \tag{6}$$

where $\boldsymbol{x}$ is the state vector of the target, $\boldsymbol{A}$ denotes the state transition matrix, $\boldsymbol{n}_{R,k}$ represents zero-mean Gaussian process noise with covariance matrix $\boldsymbol{R}$. For the measurement of the

target position, multiple onboard cameras are employed for cooperative localization. A multi-view geometric linearized triangulation technique [32] is adopted, and the target position is estimated via least-squares fitting. The stochastic discrete-time state-space representation of the measurement model is given by Equation (7):

$$\boldsymbol{z}_{k+1} = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{n}_{Q,k}, \boldsymbol{n}_{Q,k} \sim \mathcal{N}(0, \boldsymbol{Q}) \tag{7}$$

where $\boldsymbol{z} \in \mathbb{R}^3$ is the measurement vector, $\boldsymbol{H}$ denotes the measurement matrix, and $\boldsymbol{n}_{Q,k}$ represents zero-mean Gaussian measurement noise with covariance matrix $\boldsymbol{Q} = \sigma_q \boldsymbol{I}_3$. The CV and CA motion models can be derived from Equation (6), which follows a classical state transition process [17].

3.2. Formation Encirclement Controller Based on IBVS

As stated in the problem description in Section 2.5, a control law must be designed to achieve the control objectives in both the image space and the ECS.

$$\boldsymbol{p}_{\mathrm{vr},i} = \boldsymbol{R}^{\mathrm{v}}_{\mathrm{e},i} \hat{\boldsymbol{p}}_{\mathrm{r},i}, i \in \begin{bmatrix} 1 & \dots & N \end{bmatrix} \tag{8}$$

where $\boldsymbol{p}_{\mathrm{vr},i}$ denotes the relative position in the VCS, the subscript i represents the i-the quadrotor, $\boldsymbol{R}^{\mathrm{v}}_{\mathrm{e},i}$ signifies the rotation matrix from the ECS to the VCS, and $\hat{\boldsymbol{p}}_{\mathrm{r},i} = \boldsymbol{p}_i - \hat{\boldsymbol{p}}_{\mathrm{t}}$ indicates the relative position in the ECS. The virtual image coordinates $u_{\mathrm{v},i}$, $v_{\mathrm{v},i}$ can be obtained through Equation (9):

$$\begin{aligned} u_{\mathrm{v},i} &= \frac{f}{\boldsymbol{p}_{\mathrm{vr},i}(1)} \boldsymbol{p}_{\mathrm{vr},i}(2) \\ v_{\mathrm{v},i} &= \frac{f}{\boldsymbol{p}_{\mathrm{vr},i}(1)} \boldsymbol{p}_{\mathrm{vr},i}(3) \end{aligned} \tag{9}$$

The virtual image error is constructed as shown in Equation (10):

$$\boldsymbol{e}_{\mathrm{img},i} = \begin{bmatrix} u_{\mathrm{vd}} - u_{\mathrm{v},i} \\ v_{\mathrm{vd}} - v_{\mathrm{v},i} \end{bmatrix} \tag{10}$$

where u_{vd} and v_{vd} denote the desired virtual image coordinates. The depth-direction error in the virtual camera coordinate system is constructed as Equation (11):

$$e_{\mathrm{dis},i} = \frac{2\boldsymbol{p}_{\mathrm{vr},i}(1) - \sqrt{2} p_{\mathrm{d}}}{\sqrt{2} p_{\mathrm{d}}} \tag{11}$$

where p_{d} is the desired position error, which is designed here as a time-varying quantity. Let p_{d0} and p_{d1} denote the initial and final desired position errors, respectively, and let ξ_{t} represent the time decay rate. The expression for p_{d} is given by $p_{\mathrm{d}} = (p_{\mathrm{d0}} - p_{\mathrm{d1}}) \exp(-\xi_{\mathrm{t}} t) + p_{\mathrm{d1}}$.

The yaw-angle error is constructed as Equation (12):

$$e_{\psi,i} = \psi_{\mathrm{d},i} - \psi_i \tag{12}$$

Combining (9)–(12), the error vector $\boldsymbol{e}_i$ is obtained as shown in Equation (13):

$$\boldsymbol{e}_i = \begin{bmatrix} e_{\mathrm{dis},i} & \boldsymbol{e}^{\mathrm{T}}_{\mathrm{img},i} & e_{\psi,i} \end{bmatrix}^{\mathrm{T}} \tag{13}$$

Based on (3) and (13), the control law in the image space is designed as follows:

$$\boldsymbol{v}_{\mathrm{v},i} = -\lambda_{\mathrm{v},i} \hat{\boldsymbol{L}}^{+}_{\mathrm{v},i} \boldsymbol{e}_i \tag{14}$$

where $\lambda_{v,i} \in \mathbb{R}^+$ is the control gain coefficient, $v_{v,i}(4)$ denotes the fourth element of vector $v_{v,i}$, representing the desired yaw angular rate that can be directly used as the input to the attitude control loop. $\hat{L}_{v,i}^{+}$ denotes the Moore–Penrose pseudoinverse of $\hat{L}_{v,i}$, whose explicit form is given by Equation (15):

$$\hat{L}_{v,i} = \begin{bmatrix} \frac{-\sqrt{2}}{p_d} & 0 & 0 & \frac{-\sqrt{2}p_{vr,i}(2)}{p_d} \\ \frac{-u_{v,i}}{\hat{p}_{vr,i}(1)} & \frac{f}{\hat{p}_{vr,i}(1)} & 0 & \frac{f^2+u_{v,i}^2}{f} \\ \frac{-v_{v,i}}{\hat{p}_{vr,i}(1)} & 0 & \frac{f}{\hat{p}_{vr,i}(1)} & \frac{u_{v,i}v_{v,i}}{f} \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{15}$$

From Equation (15), it can be observed that $\hat{L}_{v,i}$ is non-singular. Inspired by [33], the position error of each node is formulated as Equation (16):

$$e_{rel,i} = \sum_{j \in \mathcal{N}_i(\mathcal{G})} \left(\left\| p_j - p_i \right\| - \frac{p_d}{\sqrt{2}} \right) \cdot \frac{p_j - p_i}{\left\| p_j - p_i \right\|} \tag{16}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. By combining the expression of p_q from Section 2.2 with Equation (13), we further derive Equation (17):

$$e_{rel} = (\mathcal{L} \otimes I_3) p_q - \frac{\sqrt{2}}{2} D\left(p_q\right) \tag{17}$$

where $e_{rel} = \begin{bmatrix} e_{rel,1} & \dots & e_{rel,N} \end{bmatrix}^T$ is the relative position error vector of all nodes, and $\otimes$ denotes the Kronecker product. Combining with Equation (14), the control law is designed as (18):

$$v_{c,i} = -\lambda_{v,i} \hat{L}_{v,i}^{+}(1:3,1:3) e_i(1:3) + \lambda_{c,i} R_{e,i}^{v} e_{rel,i} \tag{18}$$

where $\lambda_{c,i} \in \mathbb{R}^+$ is the control gain coefficient, $\hat{\mathcal{L}}_{v,i}^{+}(1:3,1:3)$ denotes the submatrix formed by the first three rows and three columns of $\hat{L}_{v,i}^{+}$, and $e_i(1:3)$ represents the vector containing the first three elements of e_i. Incorporating the estimated target velocity $\hat{v}_t$ obtained from the IMM filter in Section 3.1 and performing coordinate transformation, the desired control input is given by:

$$v_{con,i} = R_v^e (v_{c,i} + R_e^v \hat{v}_t) \tag{19}$$

Lemma 1. *According to LaSalle's invariance principle, if a system $\dot{x} = f(x)$ is locally Lipschitz and there exists a continuously differentiable positive-definite function $V(x)|\mathcal{D} \subset \mathbb{R}^n$ such that $\dot{V}(x) \leq 0$ for all $x \in \mathcal{D}$. Let $\Omega = \left\{ x \in \mathcal{D} \middle| \dot{V}(x) = 0 \right\}$ and let $\mathcal{M}$ be the largest invariant set in Ω. Then every trajectory $x(t)$ starting in $\mathcal{D}$ approaches $\mathcal{M}$ as $t \to \infty$.*

Let $V_{rel} = {}^1/_4 \sum_{i \neq j} \left(\left(p_j - p_i \right) - p_d \right)^2$ and $H_{ij} = \partial^2 V_{rel} / \partial p_i \partial p_j$. The Hessian matrix $\mathcal{H} \in \mathbb{R}^{3m \times 3m}$ is defined as symmetric and positive semidefinite, where its block $(i,j)_{i,j \in [1,N]}$ is given by H_{ij}. Let $h_{min} > 0$ and $h_{max} > 0$ denote the minimum eigenvalue $\lambda_{min}(\mathcal{H})$ and the maximum eigenvalue $\lambda_{max}(\mathcal{H})$ of $\mathcal{H}$, respectively. Furthermore, let $\mathcal{L} = \mathrm{diag}\left(L_{s,1} \quad \dots \quad L_{s,m} \right)$ and $\mathcal{L}^{-1} = \mathrm{diag}\left(L_{s,1}^{-1} \quad \dots \quad L_{s,m}^{-1} \right)$, where $L_{s,i}$ represents $\hat{L}_{v,i}^{+}(1:3,1:3)$.

Assumption 2. *The norms of the block matrix* $\boldsymbol{\mathcal{L}}$ *and its pseudo-inverse* $\boldsymbol{\mathcal{L}}^{-1}$*, denoted as* $A = \|\mathcal{L}\|$*and* $B = \left\|\boldsymbol{\mathcal{L}}^{-1}\right\|$*, respectively, satisfy* $AB > 1$*. Furthermore, there exists* $\eta \in [(q + \sqrt{q^2 + 4pg})/2p, min\left(\alpha/h_{max}B \quad \beta h_{min}/A\right)]$*, where* $p = h_{max}(AB - 1)$*,* $g = \alpha\beta(h_{max}AB - h_{min})$*,* $q = \alpha(A + h_{max}) + \beta h_{max}B(h_{min} + A)$*,* $\alpha > 0$ *and* $\beta > 0$*.*

Theorem 1. *Under Assumptions 1 and 2, the control law given by Equation (19) simultaneously achieves convergence of the image error in the image space and convergence of the formation error in the ECS. The proof is as follows:*

Proof. Combining Equations (14) and (16), let $\boldsymbol{a}_i = \boldsymbol{e}_i(1:3)$ and $\boldsymbol{b}_i = \boldsymbol{e}_{\mathrm{rel},i}$. Substituting these into Equation (18) yields:

$$\dot{\boldsymbol{a}}_i = -\lambda_{\mathrm{v}}\boldsymbol{a}_i - \lambda_{\mathrm{c}}\boldsymbol{L}_{s,i}\boldsymbol{b}_i \tag{20}$$

Combining $V_{rel} = {}^1/_4\sum_{i\neq j}\left(\left\|\boldsymbol{p}_j - \boldsymbol{p}_i\right\| - p_d\right)^2$, we have $\boldsymbol{b}_i = -\nabla_{p_i}V_{\mathrm{rel}}$. Thus, it follows that:

$$\dot{\boldsymbol{b}}_i = -\nabla^2_{\boldsymbol{p}_i}V_{\mathrm{rel}}\dot{\boldsymbol{p}}_i - \sum_{j\neq i}\frac{\partial^2 V_{\mathrm{rel}}}{\partial \boldsymbol{p}_i \partial \boldsymbol{p}_j}\dot{\boldsymbol{p}}_j \tag{21}$$

Linearizing around the equilibrium point, $\dot{\boldsymbol{b}}_i$ can be approximated as:

$$\dot{\boldsymbol{b}}_i \approx -\lambda_{\mathrm{v}}\boldsymbol{H}_{ii}\boldsymbol{L}_{s,i}^{-1}\boldsymbol{a}_i - \lambda_{\mathrm{c}}\boldsymbol{H}_{ii}\boldsymbol{b}_i - \lambda_{\mathrm{v}}\sum_{j\neq i}\boldsymbol{H}_{ij}\boldsymbol{L}_{s,j}^{-1}\boldsymbol{a}_j - \lambda_{\mathrm{c}}\sum_{j\neq i}\boldsymbol{H}_{ij}\boldsymbol{b}_j \tag{22}$$

Let $\boldsymbol{a} = \begin{bmatrix}\boldsymbol{a}_1 & \dots & \boldsymbol{a}_N\end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{b} = \begin{bmatrix}\boldsymbol{b}_1 & \dots & \boldsymbol{b}_N\end{bmatrix}^{\mathrm{T}}$. Based on the definition of the block matrix $\boldsymbol{\mathcal{L}}$, and according to Equations (20)–(22), the following dynamic relationship holds:

$$\dot{\boldsymbol{a}} = -\lambda_{\mathrm{v}}\boldsymbol{a} - \lambda_{\mathrm{c}}\boldsymbol{\mathcal{L}}\boldsymbol{b}$$

$$\dot{\boldsymbol{b}} = -\lambda_{\mathrm{v}}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{L}}^{-1}\boldsymbol{a} - \lambda_{\mathrm{c}}\boldsymbol{\mathcal{H}}\boldsymbol{b}$$

Consider the following Lyapunov function candidate:

$$V(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{2}\alpha\|\boldsymbol{a}\|^2 + \frac{1}{2}\beta\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{H}}^{-1}\boldsymbol{b} + \eta\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$$

From $V(\boldsymbol{a},\boldsymbol{b})$, it can be observed that when $V(\boldsymbol{a},\boldsymbol{b}) = 0$, it is equivalent to $\boldsymbol{e} = \boldsymbol{0}$ and $\boldsymbol{e}_{rel} = \boldsymbol{0}$, which implies that both the image-space error and the formation position error converge to zero.

According to the positive-definiteness condition:

$$\begin{bmatrix}\alpha\boldsymbol{I} & \eta\boldsymbol{I} \\ \eta\boldsymbol{I} & \beta\boldsymbol{\mathcal{H}}^{-1}\end{bmatrix} > 0$$

This requires satisfying:

$$\beta > \frac{\eta^2}{\alpha}h_{\min} \tag{23}$$

Computing the derivative of $V(\boldsymbol{a},\boldsymbol{b})$ yields:

$$\dot{V}(\boldsymbol{a},\boldsymbol{b}) = \alpha\boldsymbol{a}^{\mathrm{T}}\dot{\boldsymbol{a}} + \beta\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{H}}^{-1}\dot{\boldsymbol{b}} + \eta\dot{\boldsymbol{a}}^{\mathrm{T}}\boldsymbol{b} + \eta\boldsymbol{a}^{\mathrm{T}}\dot{\boldsymbol{b}}$$

Substituting Equations (26) and (27) and rearranging into quadratic form:

$$\begin{aligned}\dot{V}(\boldsymbol{a},\boldsymbol{b}) = &-\lambda_{\mathrm{v}}\alpha\|\boldsymbol{a}\|^2 - \lambda_{\mathrm{c}}\beta\|\boldsymbol{b}\|^2 - \eta\lambda_{\mathrm{v}}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{L}}^{-1}\boldsymbol{a} - \eta\lambda_{\mathrm{c}}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{L}}^{\mathrm{T}}\boldsymbol{b} - \\ &\lambda_{\mathrm{c}}\alpha\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mathcal{L}}^{\mathrm{T}}\boldsymbol{b} - \lambda_{\mathrm{v}}\beta\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mathcal{L}}^{-\mathrm{T}}\boldsymbol{a} - \eta\lambda_{\mathrm{v}}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} - \eta\lambda_{\mathrm{c}}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mathcal{H}}\boldsymbol{b}\end{aligned} \tag{24}$$

Denoting $A = \|\mathcal{L}\|$ and $B = \left\|\mathcal{L}^{-1}\right\|$, an upper bound of $\dot{V}(\boldsymbol{a},\boldsymbol{b})$ can be obtained:

$$\begin{aligned}\dot{V}(\boldsymbol{a},\boldsymbol{b}) \leq &-\lambda_{\mathrm{v}}(\alpha - |\eta|h_{\max}B)\|\boldsymbol{a}\|^2 - \lambda_{\mathrm{c}}(\beta h_{\min} - |\eta|A)\|\boldsymbol{b}\|^2 + \\ &[\lambda_{\mathrm{c}}\alpha A + \lambda_{\mathrm{v}}\beta h_{\max}B + |\eta|(\lambda_{\mathrm{v}} + \lambda_{\mathrm{c}}h_{\max})]\|\boldsymbol{a}\|\|\boldsymbol{b}\|\end{aligned} \tag{25}$$

Let $p_{11} = \lambda_{\mathrm{v}}(\alpha - |\eta|h_{\max}B)$, $p_{22} = \lambda_{\mathrm{c}}(\beta h_{\min} - |\eta|A)$, $p_{12} = \frac{1}{2}[\lambda_{\mathrm{c}}\alpha A + \lambda_{\mathrm{v}}\beta h_{\max}B + |\eta|(\lambda_{\mathrm{v}} + \lambda_{\mathrm{c}}h_{\max})]$,

$$\dot{V}(\boldsymbol{a},\boldsymbol{b}) \leq -\begin{bmatrix}\|\boldsymbol{a}\| \\ \|\boldsymbol{b}\|\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}p_{11} & -p_{12} \\ -p_{12} & p_{22}\end{bmatrix}\begin{bmatrix}\|\boldsymbol{a}\| \\ \|\boldsymbol{b}\|\end{bmatrix} \tag{26}$$

From Equation (26), it is required that conditions $\alpha > |\eta|h_{\max}B$ and $\beta h_{\min} > |\eta|A$ are satisfied. To ensure condition $p_{11}p_{22} > p_{12}^2$, the following must hold:

$$c_1c_2\lambda_{\mathrm{v}}\lambda_{\mathrm{c}} > \frac{1}{4}(a\lambda_{\mathrm{v}} + b\lambda_{\mathrm{c}})^2 \tag{27}$$

where $c_1 = \alpha - |\eta|h_{\max}B > 0$, $c_2 = \beta h_{\min} - |\eta|A > 0$, $a = \beta h_{\max}B + |\eta|$, $b = \alpha A + |\eta|h_{\max}$. Let $r = \lambda_{\mathrm{v}}/\lambda_{\mathrm{c}}$, a quadratic inequality with respect to r can be derived:

$$a^2r^2 + (2ab - 4c_1c_2)r + b^2 < 0 \tag{28}$$

According to the discriminant of the quadratic equation and Assumption 2, the feasible range of r is

$$\frac{4c_1c_2 - 2ab - 4\sqrt{c_1c_2(c_1c_2 - ab)}}{2a^2} < r < \frac{4c_1c_2 - 2ab + 4\sqrt{c_1c_2(c_1c_2 - ab)}}{2a^2} \tag{29}$$

Rearranging the discriminant $(c_1c_2 - ab)$ into a quadratic equation in terms of $|\eta|$

$$p|\eta|^2q|\eta| - g > 0 \tag{30}$$

According to the positive-definiteness conditions $\alpha > |\eta|h_{\max}B$ and $\beta h_{\min} > |\eta|A$, together with Assumption 2, there exist constants $\eta \in [(q + \sqrt{q^2 + 4pg})/2p$, and $\min\left(\alpha/h_{\max}B \quad \beta h_{\min}/A\right)]$ such that inequality (30) holds. Consequently, a feasible solution for r in (29) exists, ensuring that (26) is negative semidefinite. Thus, Theorem 1 is proved. Furthermore, by invoking Lemma 1, the system is uniformly asymptotically stable. □

3.3. CBF-Based Collision Avoidance for Quadrotors

As mentioned in Section 2.3, during motion, collision avoidance among quadrotors must be guaranteed, i.e., $\delta_{ij} > h_{\mathrm{c}}$ must hold, where h_{c} denotes the minimum safe distance. Inspired by [34], we introduce a collision-avoidance constraint based on the CBF h_{ij}, with $\boldsymbol{s}_i = \begin{bmatrix}\boldsymbol{p}_i^{\mathrm{T}} & \boldsymbol{v}_i^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$, and subsequently derive the admissible input set:

$$\mathcal{S}_{ij} = \left\{(\boldsymbol{s}_i, \boldsymbol{s}_j)\middle|h_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) \geq 0\right\} \tag{31}$$

where the specific form of h_{ij} is given by Equation (32):

$$h_{ij} = \sqrt{4u_m\left(\left\|\Delta \boldsymbol{p}_{ij}\right\| - h_{\mathrm{c}}\right)} + \frac{\Delta \boldsymbol{p}_{ij}^{\mathrm{T}}}{\left\|\Delta \boldsymbol{p}_{ij}\right\|}\Delta \boldsymbol{v}_{ij} \tag{32}$$

where u_m denotes the maximum allowable velocity. The corresponding linear CBF constraint can then be derived as:

$$-\Delta \boldsymbol{p}_{ij}^{\mathrm{T}}\Delta \boldsymbol{u}_{ij}^{\mathrm{T}} \leq \alpha_1 h_{ij}^3\left\|\Delta \boldsymbol{p}_{ij}\right\| - \frac{\left(\Delta \boldsymbol{v}_{ij}^{\mathrm{T}}\Delta \boldsymbol{p}_{ij}\right)^2}{\left\|\Delta \boldsymbol{p}_{ij}\right\|^2} + \left\|\Delta \boldsymbol{v}_{ij}\right\|^2 + \frac{2u_m\Delta \boldsymbol{v}_{ij}^{\mathrm{T}}\Delta \boldsymbol{p}_{ij}}{\sqrt{4u_m\left(\left\|\Delta \boldsymbol{p}_{ij}\right\| - h_{\mathrm{c}}\right)}} \tag{33}$$

where $\Delta \boldsymbol{u}_{ij} = \boldsymbol{v}_{con,i} - \boldsymbol{v}_{con,j}$. Substituting f_{ij} for the right-hand side expression in inequality (33), it follows that $f_{ij} = f_{ji}$ must be satisfied. Constraint (33) can thus be rewritten as:

$$-\Delta \boldsymbol{p}_{ij}^{\mathrm{T}}\Delta \boldsymbol{u}_{ij}^{\mathrm{T}} \leq f_{ij} \tag{34}$$

which can be further allocated to nodes i and j, respectively, as:

$$-\Delta \boldsymbol{p}_{ij}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{con},i} \leq \frac{f_{ij}}{2}, -\Delta \boldsymbol{p}_{ij}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{con},j} \leq \frac{f_{ji}}{2} \tag{35}$$

Based on the linear constraint in (35), the optimal control law for node i is obtained as:

$$\begin{aligned} \boldsymbol{v}_{\mathrm{con},i} = \arg\min_{j\in\mathcal{N}_i} \ & \tfrac{1}{2}\boldsymbol{v}_{\mathrm{con},i}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{con},i} \\ \text{s.t.} \quad & \mathrm{C1}: \dot{\boldsymbol{s}}_i = \boldsymbol{A}\boldsymbol{s}_i + \boldsymbol{B}\boldsymbol{v}_{\mathrm{oc},i} \\ & \mathrm{C2}: -\Delta \boldsymbol{p}_{ij}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{con},i} \leq \tfrac{f_{ij}}{2}, j \in \mathcal{N}_i \\ & \mathrm{C3}: \left\|\boldsymbol{v}_{\mathrm{con},i}\right\| \leq u_m \end{aligned}$$

Assumption 3. *The quadrotor formation satisfies constraint $\delta_{ij} > h_c$ at the initial time $t = 0$ s.*

3.4. Image-Coordinate Prediction of Quadrotors in the Field of View

During the tracking of a maneuvering target, the formation motion may cause a member of the formation to appear within the FOV of a node, leading to potential misidentification as the target. Such misidentification can interfere with accurate target tracking control and degrade the precision and stability of state estimation. Therefore, it is necessary to incorporate the prediction and marking of quadrotor positions into the control process to avoid misidentification. Inspired by [35], we project the positions of cooperative quadrotor onto the image plane for labeling and eliminate potential false detections based on overlap analysis. Each node receives the position information of other cooperating nodes via the communication network. The procedure for calculating the image-plane position of the j-th neighboring quadrotor on the onboard camera of the i -th quadrotor is described as follows:

$$\boldsymbol{p}_i^j = \boldsymbol{R}_{\mathrm{e}}^{\mathrm{c},i}\left(\boldsymbol{p}_j - \boldsymbol{p}_i\right) \tag{36}$$

where $\boldsymbol{p}_i^j$ denotes the position of quadrotor j in the CCS of quadrotor i, and $\boldsymbol{R}_{\mathrm{e}}^{\mathrm{c},i}$ is the rotation matrix from the ECS to the CCS of quadrotor i Combining with Equation (9), the image-plane coordinates u_i^j and v_i^j can be further obtained through the camera model as:

$$\begin{bmatrix} \hat{u}_i^j \\ \hat{v}_i^j \end{bmatrix} = \begin{bmatrix} f\frac{p_i^j(2)}{p_i^j(1)} & f\frac{p_i^j(3)}{p_i^j(1)} \end{bmatrix}^{\mathrm{T}} \tag{37}$$

Similarly, the position of the target in the CCS i, denoted as $\boldsymbol{p}_i^{\mathrm{t}}$, and its corresponding image-plane coordinates u_i^{t} and v_i^{t} can be obtained as:

$$\begin{bmatrix} u_i^{\mathrm{t}} \\ v_i^{\mathrm{t}} \end{bmatrix} = \begin{bmatrix} f\frac{p_i^{\mathrm{t}}(2)}{p_i^{\mathrm{t}}(1)} & f\frac{p_i^{\mathrm{t}}(3)}{p_i^{\mathrm{t}}(1)} \end{bmatrix}^{\mathrm{T}} \tag{38}$$

Taking the projected image coordinates u_i^j, v_i^j as the center, we define a bounding box with a threshold radius d_N. When a moving point appears within the bounding box of radius d_N, it is labeled as a misidentified quadrotor. In special cases where the same moving point appears simultaneously in two intersecting bounding boxes, such scenarios are not analyzed or processed here but will be addressed in detail in future work.

4. Simulation Experiments

4.1. Position and Velocity Estimation of the Maneuvering Target

The initial position of the target is $\boldsymbol{p}_{\mathrm{t}} = \begin{bmatrix} 50\ \mathrm{m} & 0\ \mathrm{m} & -20\ \mathrm{m} \end{bmatrix}^{\mathrm{T}}$, and its initial velocity is $\boldsymbol{v}_{\mathrm{t}} = \begin{bmatrix} 2\ \mathrm{m/s} & 1\ \mathrm{m/s} & 0.5\ \mathrm{m/s} \end{bmatrix}^{\mathrm{T}}$. At time $t = 6.67$ s, the target acceleration is $\boldsymbol{a}_{\mathrm{t}} = \begin{bmatrix} 3\ \mathrm{m/s^2} & 2\ \mathrm{m/s^2} & 1\ \mathrm{m/s^2} \end{bmatrix}^{\mathrm{T}}$. At time $t = 13.33$ s, the target acceleration switches to $\boldsymbol{a}_{\mathrm{t}} = \begin{bmatrix} -2\ \mathrm{m/s^2} & -2\ \mathrm{m/s^2} & -2\ \mathrm{m/s^2} \end{bmatrix}^{\mathrm{T}}$.

In Table 1, MPE denotes mean positional error, MVE denotes mean velocity error, MAPE denotes mean absolute positional error, and MAVE denotes mean absolute velocity error. MPE and MVE are computed as the RMSE, while MAPE and MAVE are calculated as the mean absolute error.

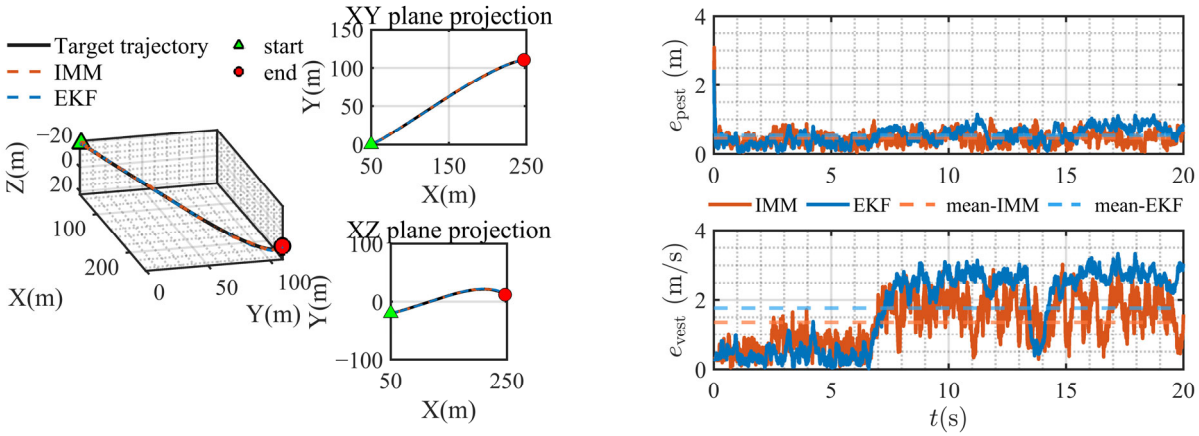
Table 1. Root mean square error and mean absolute error of the estimation.

Estimation Methods	MPE (m)	MVE (m/s)	MAPE (m/s)	MAVE (m/s)
IMM	0.4998	1.5195	0.4579	1.3535
EKF	0.5840	2.0660	0.4583	1.7637

Figure 3a illustrates the estimated trajectory of the target, where the triangle marker denotes the starting point and the circle marker indicates the endpoint. Both the EKF and the IMM filter achieve stable estimation of the target's changing position. As shown in Figure 3b, the IMM filter outperforms the EKF in terms of both position and velocity estimation errors. Based on the MPE and MVE data in Table 1, it is observed that for target position estimation, the IMM reduces the estimation bias by 14.4% compared to the EKF, while for target velocity estimation, the accuracy improvement reaches 26.5%. The accuracy metric employed here is the root mean square error (RMSE) between the estimated and actual values. The percentage improvement in estimation accuracy is calculated as per Equation (39).

$$p_{\mathrm{im}} = \frac{RMSE_{IMM} - RMSE_{EKF}}{RMSE_{EKF}} \cdot 100\% \tag{39}$$

where p_{im} denotes the percentage of improvement, and $RMSE_{IMM}$ and $RMSE_{EKF}$ represent the RMSE corresponding to the IMM and EKF estimation methods, respectively.



(**a**) Variation in position estimation (**b**) Variation in estimation error

Figure 3. EKF and IMM estimation of target location.

Table 2 presents the instantaneous absolute estimation errors at the time instants of 1 s, 5 s, 10 s, 15 s, and 20 s. Table 3 summarizes the average absolute estimation errors over the intervals of 0–5 s, 5–10 s, 10–15 s, and 15–20 s.

Table 2. Instantaneous absolute error.

t (s)	Positional (m)		Velocity (m/s)	
	IMM	EKF	IMM	EKF
1	0.3808	0.2536	0.3223	0.4065
5	0.4898	0.2720	0.6200	0.1376
10	0.6591	0.6965	2.0633	2.9200
15	0.2455	0.5201	0.7774	2.4286
20	0.5531	0.8371	1.5813	3.0512

Table 3. Average absolute error per interval.

t (s)	Positional (m)		Velocity (m/s)	
	IMM	EKF	IMM	EKF
0–5	0.4459	0.4464	0.6754	0.4346
5–10	0.4221	0.4235	1.3766	1.5855
10–15	0.4750	0.4761	1.6717	2.3067
15–20	0.4885	0.4901	1.6901	2.7280

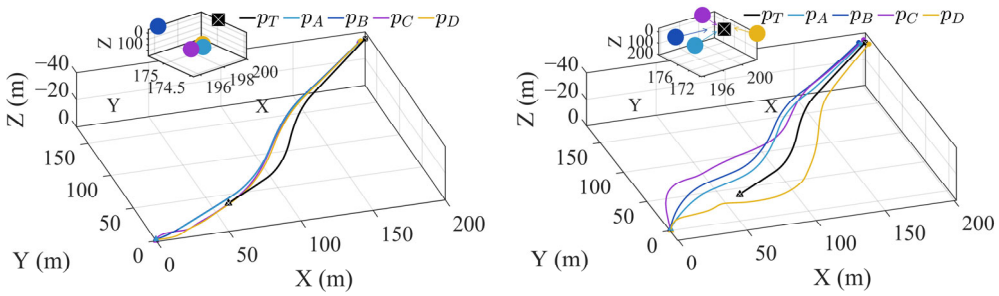
As shown in Tables 2 and 3, the estimation performance of the IMM filter is comparable to that of the EKF when the target undergoes uniform linear motion. However, when the target exhibits accelerated motion, the IMM filter demonstrates superior estimation accuracy. These observations are further corroborated by the results presented in Figure 3b.

4.2. Simulation of Encirclement and Tracking Control of Mobile Targets

The target has an initial position $\boldsymbol{p}_t = \begin{bmatrix} 50\,\mathrm{m} & 0\,\mathrm{m} & -20\,\mathrm{m} \end{bmatrix}^T$ and initial velocity $\boldsymbol{v}_t = \begin{bmatrix} 3\,\mathrm{m/s} & 1.5\,\mathrm{m/s} & -0.6\,\mathrm{m/s} \end{bmatrix}^T$. At time $t = 10$ s, the target acceleration is $\boldsymbol{a}_t = \begin{bmatrix} 0\,\mathrm{m/s^2} & 0.5\,\mathrm{m/s^2} & 0\,\mathrm{m/s^2} \end{bmatrix}^T$. At $t = 20$ s, the acceleration changes to $\boldsymbol{a}_t = \begin{bmatrix} 0\,\mathrm{m/s^2} & -0.3\,\mathrm{m/s^2} & 0.06\,\mathrm{m/s^2} \end{bmatrix}^T$ From $t = 30$ s onward, the target maintains a constant velocity motion.

Four quadrotors, denoted as Q_A, Q_B, Q_C, and Q_D, are employed to perform encircling tracking of the target. Their respective initial positions are $\boldsymbol{p}_A = \begin{bmatrix} 0\,\text{m} & 0\,\text{m} & -1\,\text{m} \end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{p}_B = \begin{bmatrix} 0.1\,\text{m} & 1\,\text{m} & -1\,\text{m} \end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{p}_C = \begin{bmatrix} -0.1\,\text{m} & 0\,\text{m} & 0\,\text{m} \end{bmatrix}^{\mathrm{T}}$, and $\boldsymbol{p}_D = \begin{bmatrix} 0.2\,\text{m} & 1\,\text{m} & 0\,\text{m} \end{bmatrix}^{\mathrm{T}}$, all with the same initial velocity 0 m/s. The desired initial position offset $p_{\mathrm{d}0}$ and final position offset $p_{\mathrm{d}1}$ are set to 1 m and 0.3 m, respectively. The initial yaw angles of all quadrotors are 0 rad, while the desired yaw angles are $\psi_{\mathrm{d}A} = -\pi/4$ rad, $\psi_{\mathrm{d}B} = \pi/4$ rad, $\psi_{\mathrm{d}C} = -3\pi/4$ rad, $\psi_{\mathrm{d}D} = 3\pi/4$ rad.

Figure 4 illustrates the tracking performance of the target using control methods based on PBVS and IBVS, respectively. Curves of different colors in the figure denote the flight trajectories of different quadrotors, with each color representing a single vehicle. Circular markers in the same color as a curve indicate the final position of the corresponding quadrotor. As shown in Figure 4, both methods are capable of achieving stable tracking of the target. The triangle in the figure represents the starting point of the formation and the target, the circle represents the final position of the formation, and the black square represents the final position of the target. From the enlarged view in Figure 4a, it can be observed that the formation successfully tracks the target. However, due to the high susceptibility of the PBVS method to noisy environments and its strong dependence on model accuracy, the formation fails to achieve a well-rounded encircling formation around the target during the tracking process, which deviates from our control objective. From Figure 4b, it can be observed that compared with the PBVS method, the IBVS-based control approach results in initially divergent trajectories among the quadrotors. This behavior arises because the IBVS-based control law is solved in the image space. Referring to the control law design process in Section 3.2, it can be noted that changes in yaw directly affect the error in the image space. Based on the current yaw angle and image error, a unique solution for the relative position with respect to the target in the ECS can be obtained. Consequently, while achieving the desired image position and yaw angle, an encircling formation around the target is generated. In the terminal tracking phase, the enlarged view in the figure shows the position distribution between the formation and the target, with arrows indicating the yaw direction of the quadrotor. This indicates that during the tracking process, the formation establishes and maintains an encircling configuration around the target.

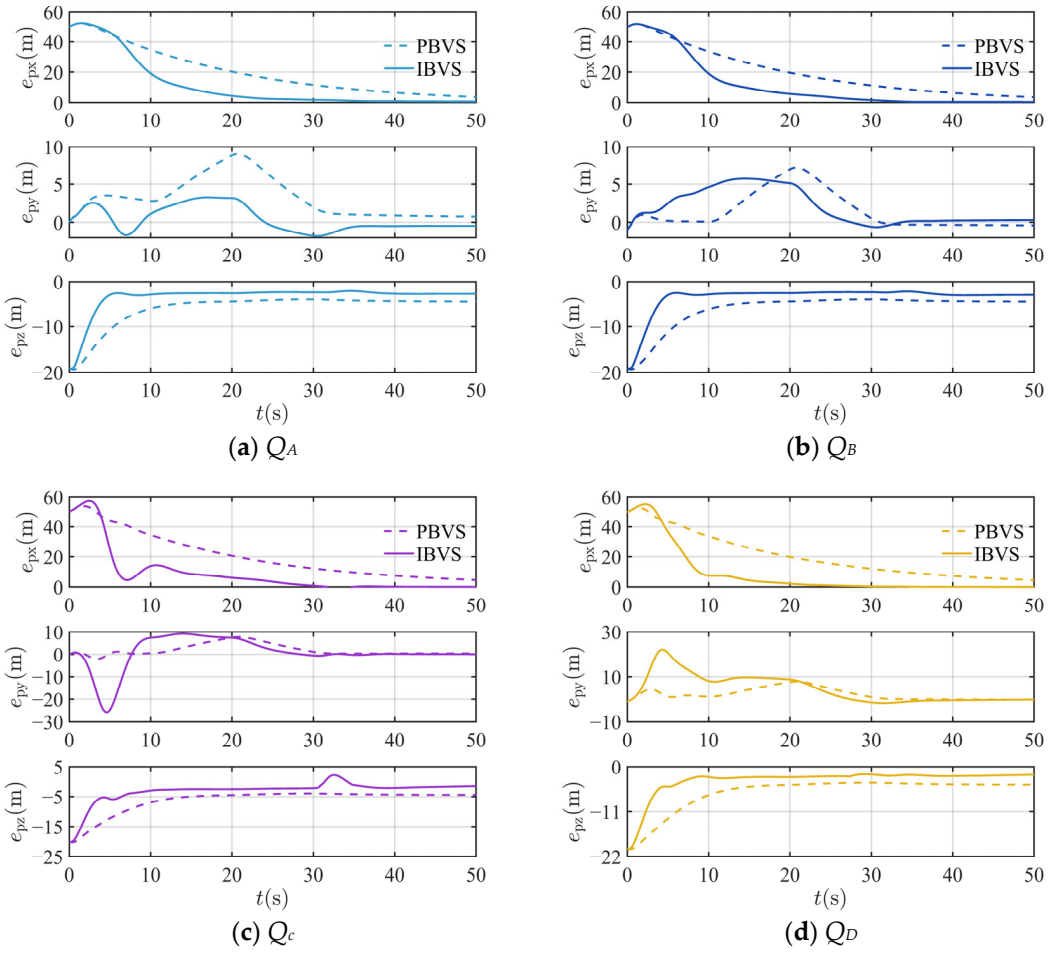


(**a**) PBVS (**b**) IBVS

Figure 4. Encirclement and tracking engineering of a maneuvering target.

Figure 5 shows the variation in tracking error $\boldsymbol{e}_p$ for quadrotors Q_A, Q_B, Q_C, and Q_D relative to the target T using the PBVS and IBVS methods, respectively. The figure indicates that both methods achieve stable convergence of the tracking error. At $t = 10$ s and $t = 20$ s, the tracking error exhibits noticeable fluctuations, which are attributed to changes in the target's maneuverability at those instants. Within the same tracking period, the tracking error under the IBVS method is significantly smaller than that under PBVS. This is because

the IBVS method relies directly on the image-space error, offering better robustness to noise and model uncertainties. In contrast, the PBVS method reconstructs a 3D relative position with respect to the target from 2D image information and then designs the control law based on this positional relationship. Consequently, noise and model uncertainties present in the 2D image space are amplified in the 3D space, directly degrading control accuracy. As seen in Figure 5c,d, during the initial tracking stage, the error along the y-axis e_{py} is relatively large when using the IBVS method, which aligns with the formation-divergence behavior observed in Figure 4b. This occurs because the designed control law, which is based on the current image error and yaw angle, leads to an encircling motion trajectory around the target before the yaw angle converges to its desired value.



(**a**) Q_A (**b**) Q_B

(**c**) Q_c (**d**) Q_D

Figure 5. Position tracking error of maneuvering target.

Figure 6 illustrates the variation in relative distances among quadrotors within the formation, both before and after the introduction of CBF-based collision-avoidance constraints. In the figure, the dashed lines labeled $(i,j), i,j \in \{A \quad B \quad C \quad D\}, i \neq j$ represent the relative distance curves between quadrotors Q_i and Q_j without CBF, while the solid lines labeled $i_c j_c$ denote the corresponding curves after CBF is applied. The red dotted line near the bottom indicates the minimum safe distance h_c defined in the CBF framework: a collision is considered to occur if the relative distance $\delta_{ij} < h_c$ between any two quadrotors falls below h_c. From the figure, it can be seen that in the final stage of encirclement tracking, the introduction of CBF effectively ensures that the distance between all quadrotors remains above the safety threshold h_c. In contrast, the curve BC without CBF shows the occurrence of collisions (as shown in bold between 35 s and 40 s in the figure).

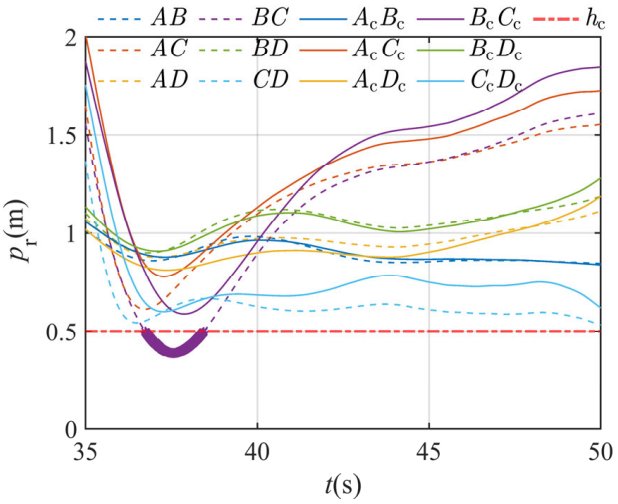
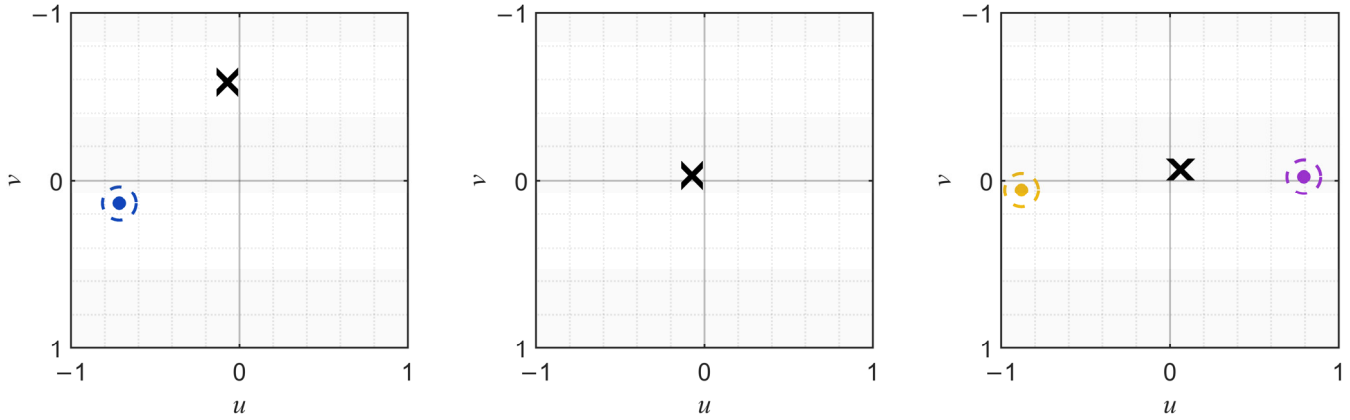


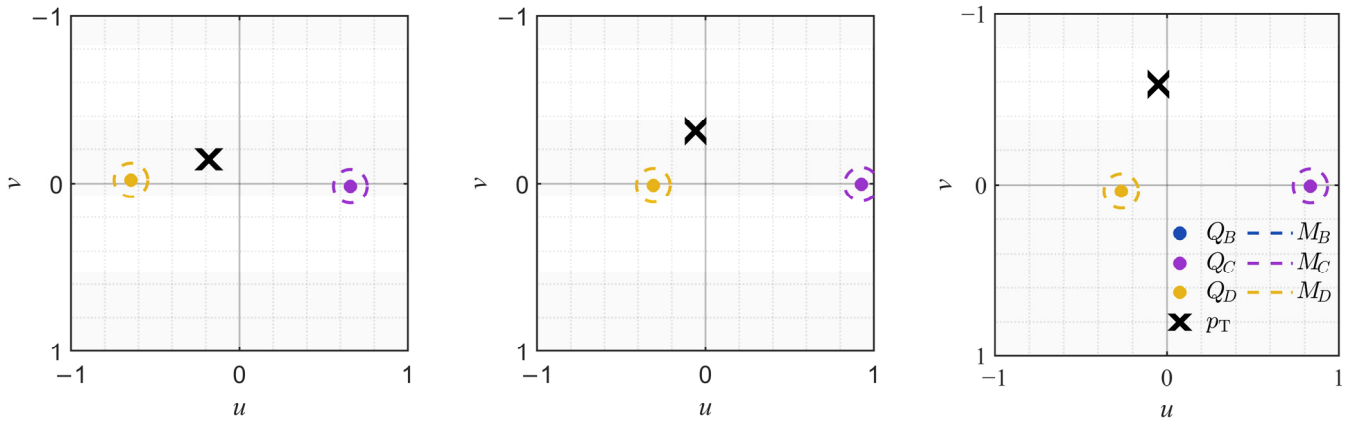
Figure 6. Changes in relative position before and after introducing collision avoidance constraints.

Figure 7 illustrates the FOV schematic of the onboard strapdown camera mounted on Q_A. Subfigures (a)–(f) show the FOV status at $t = 1$ s. 9 s, 19 s, 29 s, 39 s and 49 s, respectively. The dots in the figure denote the quadrotors of the formation visible in the FOV, labeled as Q_i, $i \in \{A \quad B \quad C \quad D\}$. The cross symbol $\times$ marks the position of the target within the FOV. The dashed circles represent the identification markers for the formation quadrotors, denoted as M_i. As can be observed, during the tracking process, each quadrotor can predict the image positions of other formation members in its own FOV by utilizing its own attitude and the communicated position information of other quadrotors. This allows the target to be distinguished from multiple moving objects when they appear simultaneously in the FOV, thereby preventing estimation errors and control disruptions caused by misidentification. Furthermore, the figure confirms that the proposed image-space-based control method successfully maintains the target near the center of the FOV throughout the tracking process.



(**a**) $t = 1$s (**b**) $t = 9$s (**c**) $t = 19$s

Figure 7. *Cont.*



(**d**) $t = 29$s (**e**) $t = 39$s (**f**) $t = 49$s

Figure 7. Dynamic changes in the position of moving objects within Q_A's FOV.

5. Conclusions

This paper addresses the formation-based encirclement control of an unknown maneuvering target by proposing an IMM filter for target state estimation and a distributed IBVS formation control method for target encircling and tracking. Based on the RMSE metric, the simulation results show that the proposed estimation method achieves a 26.5% improvement in target state estimation accuracy compared to existing methods. The stability of the proposed control scheme is rigorously analyzed using Lyapunov theory. Compared to existing control methods, the distributed IBVS-based formation control not only achieves stable encircling and tracking of the target but also further reduces tracking errors. The introduced collision-avoidance constraints effectively prevent inter-agent collisions during formation control, thereby enhancing operational safety. Furthermore, the integrated prediction of other formation members within the FOV enables reliable distinction between the target and cooperative agents, mitigating estimation errors and control disruptions caused by misidentification. Meanwhile, the proposed control law successfully drives the image position of the target toward the center of the FOV.

Author Contributions: Conceptualization, J.Y.; Methodology, H.G.; Software, H.G.; Validation, Y.D.; Formal analysis, Y.A.; Investigation, Y.A.; Resources, T.S.; Data curation, X.G.; Writing—original draft, H.G.; Writing—review & editing, T.J.; Visualization, X.G. and Y.D.; Supervision, T.S., J.Y. and T.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Sichuan-Chongqing Science and Technology Innovation Cooperation Program Project, grant number CSTB2022TIAD-CUX0015.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study.

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

Nomenclature

T	the position of the target	$\boldsymbol{A}$	the state transition matrix
$\boldsymbol{R}_{\mathrm{b}}^{\mathrm{e}} \in \mathbb{R}^{3\times3}$	the rotation matrix	$\boldsymbol{n}_{R,k}$	zero-mean Gaussian process noise
g_G	the gravitational constant	$\boldsymbol{R}$	the process covariance matrix
$(u_{\mathrm{v}}, v_{\mathrm{v}})$	virtual image coordinates	$\boldsymbol{H}$	the measurement matrix
$(u_{\mathrm{r}}, v_{\mathrm{r}})$	real image coordinates	$\boldsymbol{n}_{Q,k}$	zero-mean Gaussian measurement noise
$\boldsymbol{p}$	the position of the quadrotor	$\boldsymbol{\sigma}_q$	Standard error

$\boldsymbol{v}$	the velocity of the quadrotor	$\boldsymbol{Q}$	the measurement noise covariance
$\boldsymbol{v}_{\mathrm{b}}$	the velocity of the quadrotor in the BCS	$\boldsymbol{p}_{vr,i}$	the relative position in the VCS
$\boldsymbol{\omega}$	the angular velocity	$\hat{\boldsymbol{p}}_{r,i}$	the relative position in the ECS
$\boldsymbol{\omega}_{\times}$	the skew-symmetric matrix generated by ω	e_{dis}	the depth-direction error
$\boldsymbol{F} \in \mathbb{R}^3$	the thrust force	p_{d0}	the initial desired position errors
$\boldsymbol{\tau} \in \mathbb{R}^3$	the torque	p_{d1}	the final desired position errors
$\boldsymbol{F}_{\mathrm{d}} \in \mathbb{R}^3$	the disturbance force	ξ_{t}	the time decay rate
$\boldsymbol{\tau}_{\mathrm{d}} \in \mathbb{R}^3$	the disturbance torque	e_{ψ}	the yaw-angle error
$m_m \in \mathbb{R}$	the mass of the quadrotor	ψ_{d}	the desired yaw-angle
$\boldsymbol{J} \in \mathbb{R}^{3\times3}$	the moment of inertia of the quadrotor	ψ	the yaw-angle
$\boldsymbol{v}_{\mathrm{c}}$	the velocity of the quadrotor in the CCS	$\boldsymbol{e}_{\mathrm{img}}$	the virtual image error
$\boldsymbol{s}$	the image feature	$\lambda_{\mathrm{v},i} \in \mathbb{R}^+$	the control gain coefficient
f	the camera focal length		
$\boldsymbol{p}_{\mathrm{c}}$	the relative position in the camera frame	$\hat{\boldsymbol{L}}_{\mathrm{v}}^{+}$	the Moore–Penrose pseudoinverse of $\hat{\boldsymbol{L}}_{\mathrm{v}}$
$\boldsymbol{L}_{\mathrm{sr}} \in \mathbb{R}^{2\times6}$	the interaction matrix	$\boldsymbol{e}_{\mathrm{rel}}$	the relative position error vector of all nodes
$\boldsymbol{v}_{\mathrm{sr}}$	$\boldsymbol{v}_{\mathrm{sr}} = \begin{bmatrix} v_{\mathrm{c}}^{\mathrm{T}} & \boldsymbol{\omega}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$	$\lambda_{\mathrm{c},i} \in \mathbb{R}^+$	the control gain coefficient
$\boldsymbol{s}^*$	the desired image feature	$\hat{\boldsymbol{v}}_{\mathrm{t}}$	the estimated target velocity
$\boldsymbol{v}_{\mathrm{rd}}$	the IBVS control law	V_{rel}	a positive-definite function
$\gamma > 0$	a gain coefficient	$\boldsymbol{\mathcal{H}} \in \mathbb{R}^{3m\times3m}$	the Hessian matrix
$\boldsymbol{L}_{sr}^{+}$	the Moore–Penrose inverse of L_{sr}	$\lambda_{\min}(\boldsymbol{\mathcal{H}})$	the minimum eigenvalue
$\boldsymbol{p}_{\mathrm{v}}$	the relative position vector	$\lambda_{\max}(\boldsymbol{\mathcal{H}})$	the maximum eigenvalue
$\boldsymbol{v}_{\mathrm{sv}}$	the velocity vector	$\mathcal{L}$	$\boldsymbol{\mathcal{L}} = \mathrm{diag}\begin{pmatrix} \boldsymbol{L}_{s,1} & \dots & \boldsymbol{L}_{s,m} \end{pmatrix}$
$\hat{\boldsymbol{L}}_{\mathrm{sv}}$	the interaction matrix	$h_{\min} > 0$	$h_{\min} = \lambda_{\min}(\mathcal{H})$
$\hat{\boldsymbol{L}}_{\mathrm{sv}}^{+}$	the Moore–Penrose inverse of $\hat{\boldsymbol{L}}_{\mathrm{sv}}$	$h_{\max} > 0$	$h_{\max} = \lambda_{\max}(\mathcal{H})$
$\boldsymbol{v}_{\mathrm{v}}$	the velocity in the VCS	η	$\eta \in [(q + \sqrt{q^2 + 4pg})/2p,$ $\min\begin{pmatrix} \alpha/h_{\max}B & \beta h_{\min}/A \end{pmatrix}]$
$\boldsymbol{v}_{\mathrm{vd}}$	the control law based on virtual camera model	A	$A = \|\mathcal{L}\|$
$\mathcal{G}$	A graph	B	$B = \left\|\mathcal{L}^{-1}\right\|$
$\mathcal{V}(\mathcal{G})$	a vertex set	p	$p = h_{\max}(AB - 1)$
$\mathcal{E}(\mathcal{G})$	an edge set	g	$g = \alpha\beta(h_{\max}AB - h_{\min})$
$\mathcal{N}_i(\mathcal{G})$	$\mathcal{N}_i(\mathcal{G}) \triangleq \{j \in \mathcal{V} \mid (i,j) \in \mathcal{E}(\mathcal{G})\}$	q	$q = \alpha(A + h_{\max}) +$ $\beta h_{\max}B(h_{\min} + A)$
$\boldsymbol{\mathcal{L}}(\mathcal{G}) \in \mathbb{R}^{N\times N}$	the Laplacian matrix	α	$\alpha > 0$
$\boldsymbol{A} = \left\{a_{ij}\right\}_{N\times N}$	the adjacency matrix	β	$\beta > 0$
$\boldsymbol{D}$	the degree matrix	$\boldsymbol{a}_i$	$\boldsymbol{a}_i = \boldsymbol{e}_i(1:3)$
N	the number of quadrotors	$\boldsymbol{b}_i$	$\boldsymbol{b}_i = \boldsymbol{e}_{\mathrm{rel},i}$
$\boldsymbol{p}_q$	the positions of all nodes	$\nabla_{\boldsymbol{p}_i}$	gradient
δ_{ij}	the distance between any two nodes i and j	$\boldsymbol{a}$	$\boldsymbol{a} = \begin{bmatrix} a_1 & \dots & a_N \end{bmatrix}^{\mathrm{T}}$
t	time	$\boldsymbol{b}$	$\boldsymbol{b} = \begin{bmatrix} b_1 & \dots & b_N \end{bmatrix}^{\mathrm{T}}$
$\boldsymbol{p}_i^{\mathrm{t}}$	the position of the target in the CCS i	$V(\boldsymbol{a},\boldsymbol{b})$	the Lyapunov function candidate
u_i^{t}	image-plane coordinates	$\boldsymbol{I}$	identity matrix
v_i^{t}	image-plane coordinates	p_{11}	$p_{11} = \lambda_{\mathrm{v}}(\alpha - \|\eta\|h_{\max}B)$
$\boldsymbol{p}_{\mathrm{t}}$	the position of the target	p_{22}	$p_{22} = \lambda_{\mathrm{c}}(\beta h_{\min} - \|\eta\|A)$
$\boldsymbol{v}_{\mathrm{t}}$	its initial velocity	p_{12}	$p_{12} = \frac{1}{2}[\lambda_{\mathrm{c}}\alpha A + \lambda_{\mathrm{v}}\beta h_{\max}B$ $+\|\eta\|(\lambda_{\mathrm{v}} + \lambda_{\mathrm{c}}h_{\max})]$
$\boldsymbol{a}_{\mathrm{t}}$	the target acceleration	c_1	$c_1 = \alpha - \|\eta\|h_{\max}B > 0$
$\boldsymbol{e}_p$	the variation in tracking error	c_2	$c_2 = \beta h_{\min} - \|\eta\|A > 0$
$p_{\mathrm{d}} \in \mathbb{R}^+$	the desired interagent distance	a	$a = \beta h_{\max}B + \|\eta\|$
$h_{\mathrm{c}} > 0$	the safety distance	b	$b = \alpha A + \|\eta\|h_{\max}$

https://doi.org/10.3390/aerospace13020138

$\boldsymbol{x}$	the state vector	r	$r = \lambda_{\mathrm{v}}/\lambda_{\mathrm{c}}$
μ_k^j	model probabilities likelihoods	$\Delta \boldsymbol{u}_{ij}$	$\Delta \boldsymbol{u}_{ij} = \boldsymbol{v}_{con,i} - \boldsymbol{v}_{con,j}$
$\hat{\boldsymbol{x}}_{k-}^i$	A priori state vector	u_m	the maximum allowable velocity
$\boldsymbol{z} \in \mathbb{R}^3$	the measurement vector	f_{ij}	the right-hand side expression in (33)
$\mu_k^{i\|j}$	The likelihood of model i conditional on model j.	d_N	a threshold radius
$\hat{\boldsymbol{x}}_k^j$	the mixed initial state estimate	p_{im}	the percentage of improvement
$\boldsymbol{P}_k^j$	estimation error covariance	Q_A, Q_B, Q_C, Q_D	quadrotors
$\hat{\boldsymbol{x}}_k$	the overall state estimate	$\boldsymbol{p}_A, \boldsymbol{p}_B, \boldsymbol{p}_C, \boldsymbol{p}_D$	initial positions
$\boldsymbol{P}_k$	the overall covariance	$\psi_{\mathrm{d}A}, \psi_{\mathrm{d}B}, \psi_{\mathrm{d}C}, \psi_{\mathrm{d}D}$	the desired yaw angles

References

1. Li, Q.; Lu, X.; Wang, Y. A Novel Hierarchical Distributed Robust Formation Control Strategy for Multiple Quadrotor Aircrafts. *IEEE Trans. Syst. Man Cybern. Syst.* **2025**, *55*, 6450–6462. [CrossRef]
2. Liu, K.; Yang, W.; Jiao, L.; Yuan, Z.; Wen, C.Y. Fast Fixed-Time Distributed Neural Formation Control-based Disturbance Observer for Multiple Quadrotor UAVs Under Unknown Disturbances. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 13137–13155. [CrossRef]
3. Huang, Y.; Xu, X.; Meng, Z.; Sun, J. A Smooth Distributed Formation Control Method for Quadrotor UAVs Under Event-Triggering Mechanism and Switching Topologies. *IEEE Trans. Vehic. Technol.* **2025**, *74*, 10081–10091. [CrossRef]
4. Liu, H.; Li, B.; Ahn, C.K. Asymptotically stable learning-based formation control for multi-quadrotor UAVs. *IEEE Internet Things J.* **2025**, *12*, 24985–24995. [CrossRef]
5. Doakhan, M.; Kabganian, M.; Azimi, A. Aerial Payload Transportation with Energy-Efficient Flexible Formation Control of Quadrotors. *Eur. J. Control* **2025**, *86*, 101406. [CrossRef]
6. Upadhyay, J.; Rawat, A.; Deb, D. Multiple drone navigation and formation using selective target tracking-based computer vision. *Electronics* **2021**, *10*, 2125. [CrossRef]
7. Seidaliyeva, U.; Ilipbayeva, L.; Taissariyeva, K.; Smailov, N.; Matson, E.T. Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review. *Sensors* **2024**, *24*, 125. [CrossRef] [PubMed]
8. García, M.; Caballero, R.; González, F.; Viguria, A.; Ollero, A. Autonomous drone with ability to track and capture an aerial target. In *Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020*; IEEE: New York, NY, USA, 2020; pp. 32–40.
9. Yang, L.; Wang, X.; Liu, Z.; Shen, L. Robust Online Predictive Visual Servoing for Autonomous Landing of a Rotor UAV. *IEEE Trans. Intell. Veh.* **2025**, *10*, 2818–2835. [CrossRef]
10. Li, W.; Ma, Y.; Zhang, Y.; Li, B.; Shi, Y.; Chu, L. A Multiangle Observation and Imaging Method for UAV Swarm SAR Based on Consensus Constraints. *IEEE Sens. J.* **2025**, *25*, 19776–19793. [CrossRef]
11. Yang, K.; Bai, C.; Quan, Q. Multiview Target Estimation for Multicopter Swarm Interception. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 5023312. [CrossRef]
12. Tao, Y.; Chen, S.; Huo, X. Image-Based Visual Servo Control of a Quadrotor under Field of View Constraints Using a Pan-Tilt Camera. In *Proceedings of the 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020*; IEEE: New York, NY, USA, 2020; pp. 3518–3523.
13. Lin, F.; Peng, K.; Dong, X.; Zhao, S.; Chen, B.M. Vision-based formation for UAVs. In *Proceedings of the 11th IEEE International Conference on Control & Automation (ICCA), Taichung, Taiwan, 18–20 June 2014*; IEEE: New York, NY, USA, 2014; pp. 1375–1380.
14. Vrba, M.; Stasinchuk, Y.; Báča, T.; Spurný, V.; Petrlík, M.; Heřt, D.; Žaitlík, D.; Saska, M. Autonomous capture of agile flying objects using UAVs: The MBZIRC 2020 challenge. *Rob. Auton. Syst.* **2022**, *149*, 103970. [CrossRef]
15. Ma, J.; Chen, P.; Xiong, X.; Zhang, L.; Yu, S.; Zhang, D. Research on vision-based servoing and trajectory prediction strategy for capturing illegal drones. *Drones* **2024**, *8*, 127. [CrossRef]
16. Su, Y.X.; Zheng, C.H.; Müller, P.C.; Duan, B.Y. A simple improved velocity estimation for low-speed regions based on position measurements only. *IEEE Trans. Control Syst. Technol.* **2006**, *14*, 937–942. [CrossRef]
17. Pliska, M.; Vrba, M.; Báča, T.; Saska, M. Towards safe mid-air drone interception: Strategies for tracking & capture. *IEEE Robot. Autom. Lett.* **2024**, *10*, 8810–8817. [CrossRef]
18. Fan, X.; Wang, G.; Han, J.; Wang, Y. Interacting multiple model based on maximum correntropy Kalman filter. *IEEE Trans. Circuits Syst. II Exp. Briefs* **2021**, *68*, 3017–3021. [CrossRef]
19. Yunita, M.; Suryana, J.; Izzuddin, A. Error performance analysis of IMM-Kalman filter for maneuvering target tracking application. In *Proceedings of the 2020 6th International Conference on Wireless and Telematics (ICWT), Yogyakarta, Indonesia, 3–4 September 2020*; IEEE: New York, NY, USA, 2020; pp. 1–6.
20. Leomanni, M.; Ferrante, F.; Dionigi, A.; Costante, G.; Valigi, P.; Fravolini, M.L. Quadrotor Control System Design for Robust Monocular Visual Tracking. *IEEE Trans. Control Syst. Technol.* **2024**, *32*, 1995–2008. [CrossRef]

21. Lin, J.; Miao, Z.; Wang, Y.; Wang, H.; Wang, X.; Fierro, R. Vision-Based Safety-Critical Landing Control of Quadrotors with External Uncertainties and Collision Avoidance. *IEEE Trans. Control Syst. Technol.* **2024**, *32*, 1310–1322. [CrossRef]
22. Zhao, W.; Liu, H.; Lewis, F.L.; Valavanis, K.P.; Wang, X. Robust Visual Servoing Control for Ground Target Tracking of Quadrotors. *IEEE Trans. Control Syst. Technol.* **2019**, *28*, 1980–1987. [CrossRef]
23. Bastourous, M.; Al-Tuwayyij, J.; Guérin, F.; Guinand, F. Image based visual servoing for multi aerial robots formation. In *Proceedings of the 28th Mediterranean Conference on Control and Automation (MED), Saint-Raphaël, France, 15–18 September 2020*; IEEE: New York, NY, USA, 2020; pp. 115–120.
24. Wang, C.; Sun, Y.; Ma, X.; Chen, Q.; Gao, Q.; Liu, X. Multi-agent dynamic formation interception control based on rigid graph. *Complex Intell. Syst.* **2024**, *10*, 5585–5598. [CrossRef]
25. Karras, G.C.; Bechlioulis, C.P.; Fourlas, G.K.; Kyriakopoulos, K.J. Formation control and target interception for multiple multi-rotor aerial vehicles. In *Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020*; IEEE: New York, NY, USA, 2020; pp. 85–92.
26. Rothe, J.; Strohmeier, M.; Montenegro, S. Autonomous Multi-UAV Net Defense System for Aerial Drone Interception. In *Proceedings of the 2025 10th International Conference on Control and Robotics Engineering (ICCRE), Nagoya, Japan, 9–11 May 2025*; IEEE: New York, NY, USA, 2025; pp. 171–177.
27. Wu, A.; Yang, R.; Li, H.; Lv, M. A specified-time cooperative optimal control approach to unmanned aerial vehicle swarms. In *Proceedings of the 2023 9th International Conference on Control, Automation and Robotics (ICCAR), Beijing, China, 21–23 April 2023*; IEEE: New York, NY, USA, 2023; pp. 163–169.
28. Montijano, E.; Cristofalo, E.; Zhou, D.; Schwager, M.; Sagues, C. Vision-based distributed formation control without an external positioning system. *IEEE Trans. Robot.* **2016**, *32*, 339–351. [CrossRef]
29. Chaumette, F.; Hutchinson, S. Visual servo control. I. Basic approaches. *IEEE Robot. Autom. Mag.* **2006**, *13*, 82–90. [CrossRef]
30. Chavez-Aparicio, E.I.; Becerra, H.M.; Hayet, J.B. A Novel Consensus-Based Formation Control Scheme in the Image Space. *IEEE Control Syst. Lett.* **2024**, *8*, 2769–2774. [CrossRef]
31. Li, X.R.; Jilkov, V.P. Survey of maneuvering target tracking. Part V. Multiple-model methods. *IEEE Trans. Aerosp. Electron. Syst.* **2005**, *41*, 1255–1321. [CrossRef]
32. Lee, J.; Lee, D.; Wang, Z.; Wong, K.C.; So, S. Vision-based position and velocity estimation of an intruder drone by multiple detector drones. In Proceedings of the Asia Pacific International Symposium On Aerospace Technology (APISAT 2024), Adelaide, South Australia, 5–8 November 2024; pp. 1748–1754.
33. Xu, Y.; Qu, Y.; Luo, D.; Duan, H.; Guo, Z. Distributed predefined-time estimator-based affine formation target-enclosing maneuver control for cooperative underactuated quadrotor UAVs with fault-tolerant capabilities. *Chin. J. Aeronaut.* **2025**, *38*, 103042. [CrossRef]
34. Wang, L.; Ames, A.D.; Egerstedt, M. Safety barrier certificates for collisions-free multirobot systems. *IEEE Trans. Robot.* **2017**, *33*, 661–674. [CrossRef]
35. Zheng, C.; Mi, Y.; Guo, H.; Chen, H.; Zhao, S. Vision-Based Cooperative MAV-Capturing-MAV. *arXiv* **2025**, arXiv:2503.06412. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.