Multi-Agent Formation Control Using Epipolar Constraints

Pedro Roque, *Student Member, IEEE*, Pedro Miraldo, *Member, IEEE*, and Dimos V. Dimarogonas, *Fellow, IEEE*

***Abstract*—Formation control of multi-agent systems has profound applications in today's technological scene, ranging from satellite constellations, collaborative load transportation, cooperative surveillance, and distributed aperture imaging systems. Often, these applications are needed in environments where localization is challenging or inexistent, such as indoor and underground environments or extra-planetary scenarios (such as Mars or the Moon). In this letter, we propose a novel formation control scheme using image feature correspondences from widespread onboard cameras and only one range measurement between the formation leader and one of its neighbors. Then, optimal control inputs generated by a Nonlinear Model Predictive Control-based control law drive the agents toward the desired formation setting. The framework is tested both in simulation and on mobile platforms in a laboratory environment, with multiple camera types.**

***Index Terms*—Visual servoing, multi-robot systems, vision-based navigation.**

I. Introduction

COORDINATING a group of agents in GPS-denied environments is a hard task. Precise localization systems are expensive and proprioception is often not accurate enough for precise formation control or load transportation tasks [1], [2].

This work focuses on a multi-agent formation control problem for M agents as illustrated in Fig. 1, one leader and $M-1$ followers. The objective is to derive control laws for the followers to converge and maintain predefined relative poses between each other and the leader. While this is a well-studied problem, there are still several challenges related to the absence/limited availability of accurate relative pose sensing, particularly in GPS and/or heading-denied environments. To cope with this, we explore the use of image features from on-board cameras in the control module. We propose two methods for multi-agent relative pose coordination that guide the followers' agents to the desired relative position. The methods here proposed use Model Predictive Control (MPC) [3], [4], [5] and Visual Servoing [6], [7] techniques with i) five matching features in the images of each agent, ii) one range sensing measurement from the leader to one follower to achieve a desired formation geometry, and iii) control inputs from the neighbors to track the leader/followers motion.

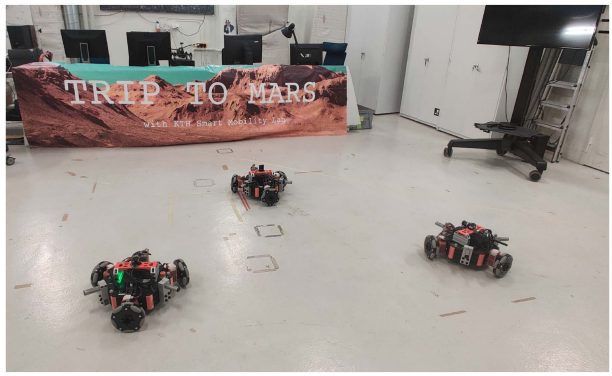


Fig. 1. Formation with 3 agents. The formation is defined as relative positions and attitudes with respect to the camera frames, which can be translated into the agent's frames.

Image-based Visual Servoing (IBVS) is a technique for driving an agent to a desire position [6], [7] using image feedback. Instead of controlling the position of the robot directly, IBVS models the agent's velocities as a function of the errors between the current and goal image features from on-board cameras. The first works exploring epipolar geometry in the IBVS are presented in [8], [9]. The authors define the errors as the distance between the image features and the epipolar lines, obtained from the current and desired relative pose, and their method drives the robot to a desired position, up to a scale factor. In this work, however, we address the problem of minimizing both rotation and translation error simultaneously and include control and state constraints to ensure the features remain in the image.

Regarding relative pose coordination using image features, [10] derives a method that provides control inputs from the epipoles computed from neighboring robots. The method reaches consensus in their orientations, without the need of directly observing each other. In [11], the authors use IMU and computer vision to obtain a rectified image [12]. The method is appropriate for aerial vehicles with down-pointing cameras. A distributed consensus scheme to deal with the translation

Manuscript received 7 March 2024; accepted 23 July 2024. Date of publication 15 August 2024; date of current version 30 October 2024. This article was recommended for publication by Associate Editor Ezio Malis and Editor Pascal Vasseur upon evaluation of the reviewers' comments. The work of Pedro Roque and Dimos V. Dimarogonas were supported in part by H2020 ERC Grant LEAFHOUND, in part by Swedish Research Council (VR), in part by Knut och Alice Wallenberg Foundation (KAW), and in part by Wallenberg AI, Autonomous Systems and Software Program (WASP) DISCOWER, funded by KAW. Pedro Miraldo is exclusively supported by MERL. *(Corresponding author: Pedro Roque.)*

Pedro Roque and Dimos V. Dimarogonas are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: padr@kth.se; dimos@kth.se).

Pedro Miraldo is with the Mitsubishi Electric Research Labs (MERL), Cambridge, MA 02139 USA (e-mail: miraldo@merl.com).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2024.3444690, provided by the authors.

Digital Object Identifier 10.1109/LRA.2024.3444690

scale is proposed. Other vision-based approaches for relative pose coordination are available: [13], [14] present methods for motion coordination and control strategy for leader-follower formations of non-holonomic vehicles, under visibility and communication constraints, as well as saturation of control inputs. [15] uses a distributed consensus of $M \geq 3$ agents for aerial-robotic teams. Using a PID-based control, each robot uses its view of a target and the relative distance from its two closest neighbors. [16] addresses the formation control of aerial vehicles with downward-facing cameras. The solution computes the control commands from the projection of a subset of ground vehicles. In [17], the authors present a vision-based method for incremental depth and relative pose estimation of ground vehicles.

In this letter, we propose a novel Image-Based Formation Control (IBFC) framework which employs visual information to drive the robotic agents to maintain a desired formation. The goal is to generate control inputs using corresponding image features from the robots' on-board cameras. In contrast to [10], [11], we i) locally obtain control inputs in the image-frame, without the need for global localization methods or heading references, such as those provided by an IMU; ii) allow for the 6 Degrees-of-Freedom (DoF) formation coordination with multiple camera types, by modeling each camera as a general projection system; and iii) use an MPC framework able to generate optimal control inputs in the image space. In [16], the authors address an entirely different problem, as they use ground vehicles as landmarks and thus leading to a largely reduced amount of features that can be used. In contrast to [15] that uses range sensing between the current and at least two neighboring agents, we only require one range distance between the leader and any other member of the formation, minimizing the amount of extra and less accurate sensors on-board the vehicles. We stress that, while some previous methods explore epipolar lines for IBVS, to the best of our knowledge, this is the first approach that exploits these constraints for multi-agent formation control. Since our method does not require inertial measurements, it is especially appealing to formations operating in microgravity or in the absence of a trustworthy magnetometer or GPS measurements. Moreover, the use of intersections from multiple epipolar constraints for $M > 2$ agents formation control is new in the literature, avoiding the use of relative position measurements for large formations, and extending considerably the application areas of our method with respect to the state of the art.

The manuscript is structured as follows: Section II details the background knowledge used to solve the problems in Section III with the methods proposed in Sections IV; simulation and experimental results are shown in Section V, and conclusions and future work are presented in Section VI.

Notation: Small bold letters represent vectors. Matrices are denoted by bold capital letters. In particular, $\mathbf{I}_n$ represents an identity matrix in $\mathbb{R}^{n\times n}$, and $\mathbf{0}_{n\times m}$ a zero matrix in $\mathbb{R}^{n\times m}$. Regular letters denote scalars. The skew-symmetric matrix representation of $\mathbf{a}$ is represented as $\mathbf{a}_\times$. Caligraphic letters denote reference frames. Rotation matrices and translation vectors from frames $\mathcal{A}$ to $\mathcal{B}$ are defined as $\mathbf{R}^{\mathcal{B}}_{\mathcal{A}} \in \mathcal{SO}(3)$ and $\mathbf{t}^{\mathcal{B}}_{\mathcal{A}} \in \mathbb{R}^3$. Features represented in a reference frame $\mathcal{A}$ are denoted as $\mathbf{a}^{\mathcal{A}}$. When the origin/target frame is the inertial frame, we omit the corresponding letter. The symbol $\sim$ denotes that the right side of an equation is equal to the left up to a scale factor. As in the MPC literature, the predicted value of variable $\mathbf{a}$ with the information available at time step k for the future time $k+n$, is written as $\mathbf{a}(k+n|k)$. A sub component r of a vector $\mathbf{a}$ is written as $\mathbf{a}_{[r]}$. Lastly, the Veronese map [18] of a vector $\mathbf{a}$ or matrix $\mathbf{A}$ is defined as ${}_\diamond\mathbf{a}$ and ${}_\diamond\mathbf{A}$, respectively.

II. Background

We consider a formation setting with M agents, led by $\mathcal{L}$, the leader of the multi-agent team, and followers $\mathcal{F}_i$. The set of all agents is defined as $\mathfrak{G} = \{\mathcal{L}, \mathcal{F}_1, \ldots, \mathcal{F}_{M-1}\}$. Fig. 1 depicts such scenario with $\mathrm{M} = 3$ agents. Moreover, we consider general camera systems in this framework, which can be applied to any general single viewpoint images.

Assume that each robotic agent in $\mathfrak{G}$ has a calibrated camera (see [12]), where each homogenous 3D feature $\mathbf{p}^{\mathcal{W}}_i \triangleq \begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T$ belonging to the inertial frame $\mathcal{W}$ is projected to the normalized image point $\mathbf{s}^{\mathcal{C}}_i \triangleq \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T$, where u_i and v_i are the normalized pixel coordinates in the camera frame $\mathcal{C}$, for each feature i. We model such cameras as a central projection system, which includes catadioptric cameras, perspective cameras, as well as several lens distortion models. Each projective ray $\mathbf{c}^{\mathcal{C}}_i \triangleq \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T$ is defined as $\mathbf{c}^{\mathcal{C}}_i = \mathbf{T}^{\mathcal{C}}_{\mathcal{W}} \mathbf{p}^{\mathcal{W}}_i$, where the matrix $\mathbf{T}^{\mathcal{C}}_{\mathcal{W}} \in \mathbb{R}^{3\times 4}$ is the camera extrinsics matrix, parametrized by $\mathbf{T}^{\mathcal{C}}_{\mathcal{W}} = {\mathbf{R}^{\mathcal{W}}_{\mathcal{C}}}^T \begin{bmatrix} \mathbf{I} & -\mathbf{t}^{\mathcal{W}}_{\mathcal{C}} \end{bmatrix}$, as in [12]. Then, we use the canonical perspective plane (CPP) model in [18], and the division (DIV) model in [19] to obtain the normalized image points $\mathbf{s}^{\mathcal{C}}_i$. Each model requires a parameter $\alpha \in (-1, 1]$ encoding the nonlinearities of general central projective systems.

A. Image-Based Visual Servoing (IBVS)

IBVS consists of controlling a camera movement based solely on image features. Consider an observed feature $\mathbf{s}_i$, and it's non-homogeneous representation $\mathbf{f}_i = \begin{bmatrix} u_i & v_i \end{bmatrix}^T$. The corresponding desired feature position in the camera frame is here denoted as $\bar{\mathbf{f}}_i$. The goal of the visual servoing task is to drive the robot to a unique position in which $\mathbf{f}_i$ converges to $\bar{\mathbf{f}}_i$, corresponding to the desired pose of the camera. The error between the desired and current feature observations is defined as

$$\tilde{\mathbf{f}}_i = \mathbf{f}_i - \bar{\mathbf{f}}_i \;\Rightarrow\; \dot{\tilde{\mathbf{f}}}_i = \dot{\mathbf{f}}_i. \tag{1}$$

Next, we introduce the interaction matrix $\mathbf{L}(\mathbf{f}_i, \alpha, z_i) \in \mathbb{R}^{2\times 6}$, referred to as $\mathbf{L}_i$, that relates the velocity of a generalized camera of a robotic agent, $\mathbf{u} \in \mathbb{R}^6$, with the movement of the observed image features in the image plane:

$$\begin{bmatrix} \dot{u}_i \\ \dot{v}_i \end{bmatrix} = \mathbf{L}_i \mathbf{u}, \tag{2}$$

where $\mathbf{L}_i \in \mathbb{R}^{2\times 6}$ [20, Eq. 13] is defined as

$$\mathbf{L}_i \triangleq \begin{bmatrix} -\frac{1+u_i^2(1-\alpha(\eta+\alpha))+v_i^2}{\rho(\eta+\alpha)} & \frac{\alpha u_i v_i}{\rho} & \frac{\eta u_i}{\rho} & \cdots & u_i v_i & -\frac{(1+u_i^2)\eta-\alpha v_i^2}{\eta+\alpha} & v_i \\ \frac{\alpha u_i v_i}{\rho} & -\frac{1+v_i^2(1-\alpha(\eta+\alpha))+u_i^2}{\rho(\eta+\alpha)} & \frac{\eta v_i}{\rho} & & -\frac{(1+v_i^2)\eta-\alpha u_i^2}{\eta+\alpha} & -u_i v_i & -u_i \end{bmatrix},$$

with $\eta = \sqrt{1+(1-\alpha^2)(u_i^2+v_i^2)}$ and $\rho = \sqrt{x_i^2+y_i^2+z_i^2}$.

In (2), $\mathbf{u}$ is defined as follows:

$$\mathbf{u} = \begin{bmatrix} \boldsymbol{\nu}^T & \boldsymbol{\omega}^T \end{bmatrix}^T, \tag{3}$$

where $\boldsymbol{\nu} \in \mathbb{R}^3$ and $\boldsymbol{\omega} \in \mathbb{R}^3$ are the linear and angular velocities, respectively. A common image-based visual servoing strategy [7] is to design the control input as $\mathbf{u} = -k\mathbf{L}_i^{\dagger}(\mathbf{f}_i - \bar{\mathbf{f}}_i)$ where $\mathbf{L}_i^{\dagger}$ is the Moore–Penrose pseudoinverse of $\mathbf{L}_i$. It can be seen in [7] that such control strategy exponentially drives the error $\mathbf{f}_i - \bar{\mathbf{f}}_i$ to the origin. It is important to note that the interaction matrix $\mathbf{L}_i$ requires an estimate of the depth z_i. This can be achieved with on-board monocular estimators, as shown in the literature [21], [22], [23].

B. Epipolar Geometry

Let j and l be two perspective cameras in a shared workspace and $\mathbf{f}_i^j$ and $\mathbf{f}_i^l$ the representation of a point $\mathbf{p}_i^{\mathcal{W}}$ in each camera image. Then, the epipolar constraint is

$$\mathbf{s}_i^{l^T} \mathbf{E}_j^l \mathbf{s}_i^j = 0, \quad j, l \in \mathfrak{G}, \tag{4}$$

where $\mathbf{E}_j^l \sim \mathbf{t}_{j_\times}^l \mathbf{R}_j^l$ - with $\sim$ meaning equal up to a scale factor - is the essential matrix that encodes the relative position $\mathbf{t}_j^l$ and attitude $\mathbf{R}_j^l$ between the two cameras, and where $\mathbf{s}_i^l$ and $\mathbf{s}_i^j$ are the normalized image features, noting that $\mathbf{s}_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T$. In the computer vision literature [12], it is common to use (4) to estimate an essential matrix $\mathbf{E}_j^l$ and extract a camera motion between two image samples.

III. Problem Statement

In this work, we aim to explore the epipolar geometry to define the formation control problem, enabling us to obtain a desired formation configuration based solely on image features $\mathbf{f}_i^j$ and $\mathbf{f}_i^l$, a single relative distance measurement $\|\mathbf{t}_j^l\|$ between two agents in $\mathfrak{G}$, and the neighbors control inputs for zero steady-state tracking error. Particularly, we consider two cases: i) the coordination of two agents, for $M = 2$, and ii) the coordination of three or more agents, for $M > 2$. We will start by defining the multi-agent formation, followed by the problem statement we aim at addressing.

A. Multi-Agent Formation System

The formation geometry is defined by the desired relative poses among the agents, that is, $\bar{\mathbf{R}}_j^l$ and $\bar{\mathbf{t}}_j^l$, encoded in an essential matrix $\bar{\mathbf{E}}_j^l$ (9), $j, l \in \mathfrak{G}$, with $j \neq l$.

We consider two types of followers. One, here denoted as $\mathcal{F}_1$, that can measure a distance to the leader $\mathcal{L}$, with state $\boldsymbol{\xi}^{\mathcal{F}_1}$ and kinematics $\dot{\boldsymbol{\xi}}^{\mathcal{F}_1}$ given by

$$\boldsymbol{\xi}^{\mathcal{F}_1} = \left[\mathbf{t}_{\mathcal{F}_1}^{\mathcal{L}^T}, \mathbf{f}_1^{\mathcal{F}_1^T}, \ldots, \mathbf{f}_5^{\mathcal{F}_1^T} \right]^T \in \mathbb{R}^{13}, \tag{5a}$$

$$\dot{\boldsymbol{\xi}}^{\mathcal{F}_1} = \begin{bmatrix} \begin{bmatrix} -\mathbf{I} & -\mathbf{t}_{\mathcal{F}_\times}^{\mathcal{L}} \end{bmatrix} \\ \mathbf{L}_1(\mathbf{f}_1^{\mathcal{F}_1}) \\ \vdots \\ \mathbf{L}_5(\mathbf{f}_5^{\mathcal{F}_1}) \end{bmatrix} \mathbf{u}^{\mathcal{F}_1} + \begin{bmatrix} \begin{bmatrix} \mathbf{R}_{\mathcal{L}}^{\mathcal{F}} & -\mathbf{t}_{\mathcal{F}_\times}^{\mathcal{L}} \mathbf{R}_{\mathcal{L}}^{\mathcal{F}} \end{bmatrix} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{u}^{\mathcal{L}}, \tag{5b}$$

where $\mathbf{t}_{\mathcal{F}_1}^{\mathcal{L}}$ is estimated from 5 feature matches with the leader by calculating the unique essential matrix and obtaining the scale with the aid of the range measurement $\|\mathbf{t}_{\mathcal{F}_1}^{\mathcal{L}}\|$. Then, $\mathbf{R}_{\mathcal{L}}^{\mathcal{F}}$ is directly obtained from the estimated essential matrix. Note that, from [24], we require at least 5 common feature observations of static world-frame points to correctly extract a camera pose (up to a scale factor). Therefore, the following assumption follows.

Assumption 1: Each agent j in the formation $\mathfrak{G}$ observes at least 5 non-colinear features in common with two formation neighbors $o, r \in \mathfrak{G} \backslash j$.

The leader $\mathcal{L}$ and followers $\mathcal{F}_2, \ldots, \mathcal{F}_{M-1}$ only measure image features and have their states $\boldsymbol{\xi}^o$ and kinematics $\dot{\boldsymbol{\xi}}^o, o \in \mathfrak{G} \backslash \mathcal{F}_1$ given as

$$\boldsymbol{\xi}^o = \begin{bmatrix} \mathbf{f}_1^o \\ \vdots \\ \mathbf{f}_5^o \end{bmatrix}^T \in \mathbb{R}^{10}, \quad \dot{\boldsymbol{\xi}}^o = \begin{bmatrix} \mathbf{L}_1(\mathbf{f}_1^o) \\ \vdots \\ \mathbf{L}_5(\mathbf{f}_5^o) \end{bmatrix} \mathbf{u}^o. \tag{6}$$

In the above equations, $\mathbf{L}_i(.)$ are the interaction matrices defined in (2). Each state $\boldsymbol{\xi}^j$ in (5) and (6) is constrained by a polytope Ξ_j where we wish the state to evolve, that is

$$\boldsymbol{\xi}^j \in \Xi^j, \quad j \in \mathfrak{G}. \tag{7}$$

In the same way, the control input is constrained as $\mathbf{u}^j \in U^j \subset \mathbb{R}^6$, where U^j is defined as

$$U^j \triangleq \{\mathbf{u}^j \in \mathbb{R}^6 : \mathbf{u}_{[r]}^{\min} \leq \mathbf{u}_{[r]}^j \leq \mathbf{u}_{[r]}^{\max}\}, r = 1, ..., 6, \tag{8}$$

with $\mathbf{u}^{\max}$ and $\mathbf{u}^{\min}$ being constant vectors such that U contains the origin, that is, $\mathbf{u}_{[r]}^{\min} \leq 0 \leq \mathbf{u}_{[r]}^{\max}$.

B. Problem Definition

Considering the multi-agent team defined in the latter section, we define the problems to be addressed in this letter in the following manner:

Problem 1 (Two Agent Coordination): Given i) an essential matrix $\mathbf{E}_j^l$, encoding the desired relative pose between two agents $j, l \in \mathfrak{G}, j \neq l$, up to a scale factor (4), ii) a relative distance measurement $\|\mathbf{t}_j^l\|$ between any two agents[1], iii) two matched feature sets $F^j = \{\mathbf{f}_1^j, \ldots, \mathbf{f}_i^j\}$ and $F^l = \{\mathbf{f}_1^l, \ldots, \mathbf{f}_i^l\}$ where $\mathbf{f}_i^j$ and $\mathbf{f}_i^l$ correspond to the same $\mathbf{p}_i$ observed by agents $j, l \in \mathfrak{G}$ respectively, with at least 5 feature correspondences, and iv) a predicted control input sequence and the distortion parameter α from a neighbor j, design $\mathbf{u}^l$ such that the agent l with state and kinematics (5), under constraints (7) and (8), is driven to a desired relative attitude $\bar{\mathbf{R}}_j^l$ and relative position $\bar{\mathbf{t}}_j^l$.

For the $M = 3$ case we consider the following problem:

Problem 2 (Formation Control): Given i) two essential matrices $\mathbf{E}_j^o$ and $\mathbf{E}_l^o$, encoding the desired and consistent relative pose between the agents $j, l, o \in \mathfrak{G}, j \neq l \neq o$, up to a scale factor (4), ii) three sets of features $F^j = \{\mathbf{f}_1^j, \ldots, \mathbf{f}_i^j\}$, $F^l = \{\mathbf{f}_1^l, \ldots, \mathbf{f}_i^l\}$ and $F^o = \{\mathbf{f}_1^o, \ldots, \mathbf{f}_i^o\}$ where each $\mathbf{f}_i^j$, $\mathbf{f}_i^l$ and $\mathbf{f}_i^o$ correspond to the same $\mathbf{p}_i$ observed by agents $j, l, o \in \mathfrak{G}$ respectively, with at least 5 feature correspondences, and iii) the predicted control input sequences and the distortion parameters α^j, α^l from a

[1]This relative distance can be measured through ultrasonic or ultra-wide band sensors.

neighbor j, l, design $\mathbf{u}^o$ such that the agent $o$with state and kinematics (6), under constraints (7) and (8), is driven to the desired relative attitudes $\bar{\mathbf{R}}_j^o$ and $\bar{\mathbf{R}}_l^o$, and relative positions $\bar{\mathbf{t}}_j^o$ and $\bar{\mathbf{t}}_l^o$.

Remark 1: It is important to note that combining the later proposed solutions to Problem 1 or Problem 2 allows us to control a formation of $M > 3$ agents, considering that one agent is in formation with a leader $\mathcal{L}$ by solving Problem 1, and the remaining agents coordinate with respect to their neighbors by solving Problem 2.

In Section IV-B and Section IV-C, two control schemes are proposed to solve the coordination problems presented in Problem 1 and Problem 2, respectively.

IV. Proposed Solution

In this work, we propose a novel solution to the formation control problems defined before. In particular, we aim to explore epipolar constraints (4) to define the formation geometry. Instead of estimating the essential matrix $\mathbf{E}_j^l$, we propose to design $\bar{\mathbf{E}}_j^l$ with a desired relative pose parametrized by a desired relative position $\bar{\mathbf{t}}_j^l$ and relative orientation $\bar{\mathbf{R}}_j^l$, such that

$$\bar{\mathbf{E}}_j^l \sim \bar{\mathbf{t}}_{j_\times}^l \bar{\mathbf{R}}_j^l. \tag{9}$$

As in [18], we use the lifting Veronese maps ${}_\diamond\mathbf{s}$ and ${}_\diamond\mathbf{E}$ to extend (4) to generalized projection systems, such that

$${}_\diamond\mathbf{s}_i^{l^T}\, {}_\diamond\bar{\mathbf{E}}_l^j\, {}_\diamond\mathbf{s}_i^j = 0, \quad j, l \in \mathfrak{G} \tag{10}$$

with ${}_\diamond\mathbf{s}_i^{l^T} \in \mathbb{R}^6$ and ${}_\diamond\bar{\mathbf{E}}_l^j \in \mathbb{R}^{6\times 6}$.

A. Model Predictive Image-Based Visual Servoing

To solve each agent's control problem we propose the use of a Nonlinear Model Predictive Controller with image feedback – hereafter referred to as Model Predictive Image-Based Visual Servoing (MP-IBVS) [25]. MP-IBVS is a Finite-Horizon Optimal Controller (FHOC) [26] which minimizes a cost function $J(\boldsymbol{\epsilon}^j, \tilde{\mathbf{u}}^j)$ depending on the error $\boldsymbol{\epsilon}^j := \psi^j(\boldsymbol{\xi}^j, \bar{\boldsymbol{\xi}}^j, \boldsymbol{\iota}^l)$, where $\boldsymbol{\xi}^j$ is the state, $\bar{\boldsymbol{\xi}}^j$ its desired value, and $\boldsymbol{\iota}^j$ is the received information vector, and control input error $\tilde{\mathbf{u}}^j$. The cost is minimized in a receding horizon of length N, while taking into account the discrete system model $g^j(\boldsymbol{\xi}^j, \mathbf{u}^j)$. The optimization problem is constrained by state and control sets Ξ^j and U^j. We will appropriately design these variables to achieve the desired formation control task. The MP-IBVS problem can then be generally written as

$$\underset{\mathbf{u}^j}{\textbf{minimize}}\ J(\boldsymbol{\epsilon}^j, \tilde{\mathbf{u}}^j) \tag{11a}$$

$$\textbf{subject to}: \boldsymbol{\xi}^j(k+n+1|k) = g^j(\boldsymbol{\xi}^j(k), \mathbf{u}^j(k)), \tag{11b}$$

$$\boldsymbol{\xi}^j(k+n|k) \in \Xi^j, \tag{11c}$$

$$\mathbf{u}^j(k+n|k) \in U^j, \tag{11d}$$

$$n = 1, \ldots, N-1, \tag{11e}$$

$$\boldsymbol{\xi}^j(0|0) = \boldsymbol{\xi}^j(0), \quad j \in \mathfrak{G}\backslash\mathcal{L}. \tag{11f}$$

Solving the optimization problem in (11) results in $N-1$ predicted control inputs $\mathbf{u}_N^j = \{\mathbf{u}^j(k|k), \ldots, \mathbf{u}^j(k+N-1|k)\}$ and predicted states $\boldsymbol{\xi}_N^j = \{\boldsymbol{\xi}^j(k+1|k), \ldots, \boldsymbol{\xi}^j(k+N|k)\}$. The MPC cost function $J(\boldsymbol{\epsilon}^j, \tilde{\mathbf{u}}^j)$ to minimize is defined as

$$J(\boldsymbol{\epsilon}^j, \tilde{\mathbf{u}}^j) = \sum_{n=0}^{N-1} l\left(\boldsymbol{\epsilon}^j(k+n|k), \tilde{\mathbf{u}}^j(k+n|k)\right) + V\left(\boldsymbol{\epsilon}^j(k+N|k)\right), \tag{12a}$$

$$l\left(\boldsymbol{\epsilon}^j, \tilde{\mathbf{u}}^j\right) = \boldsymbol{\epsilon}^j(k+n|k)^T \mathbf{Q}_e \boldsymbol{\epsilon}^j(k+n|k) + \tilde{\mathbf{u}}^j(k+n|k)^T \mathbf{Q}_u \tilde{\mathbf{u}}^j(k+n|k) \tag{12b}$$

$$V\left(\boldsymbol{\epsilon}^j\right) = \boldsymbol{\epsilon}^j(k+N|k)^T \mathbf{Q}_N \boldsymbol{\epsilon}^j(k+N|k), \tag{12c}$$

where $\mathbf{Q}_e$, $\mathbf{Q}_u$ and $\mathbf{Q}_N$ are positive-definite weighing matrices, $V(\boldsymbol{\epsilon}^j)$ is the terminal cost function, and $\tilde{\mathbf{u}}^j$ is

$$\tilde{\mathbf{u}}^j(k+n|k) = \mathbf{u}^j(k+n|k) - \bar{\mathbf{u}}^j(k+n|k),$$

where $\bar{\mathbf{u}}^j$ is the necessary control input to generate the desired system trajectory. In the next two sections, we show how to use (11) to solve Problem 1 and Problem 2.

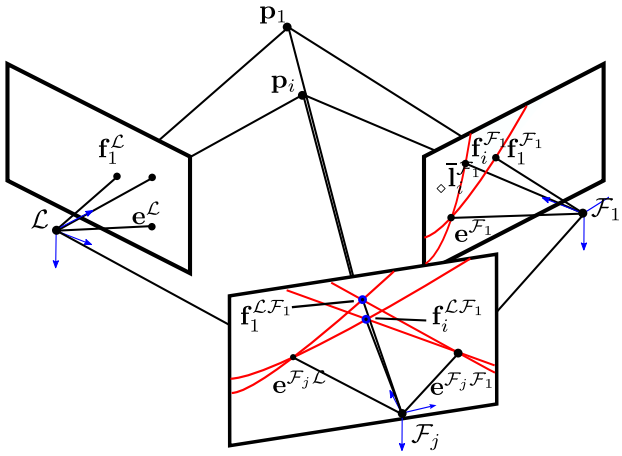


Fig. 2. Depiction of the epipolar geometry for the formation control scenario with three or more agents, where the epipolar curves defined by $\bar{\mathbf{R}}_j^o$, $\bar{\mathbf{t}}_j^o$, $\bar{\mathbf{R}}_l^o$ and $\bar{\mathbf{t}}_l^o$, $j, l, o \in \mathfrak{G}$ are seen in the agents image plane.

B. Two Agent Coordination

We first consider the problem of relative pose control between the leader $\mathcal{L}$ and one follower ($\mathcal{F}_1$), assuming that $\mathcal{F}_1$ has a relative range measurement with respect to $\mathcal{L}$, defined in Problem 1. We abbreviate $\mathcal{F}_1$ to $\mathcal{F}$ in the sequel. In this scenario, $\mathcal{F}$ receives from $\mathcal{L}$ the information vector $\boldsymbol{\iota}^L = \{\mathbf{s}_1^{\mathcal{L}}(k), \ldots, \mathbf{s}_5^{\mathcal{L}}(k), \alpha^{\mathcal{L}}, \mathbf{u}_N^{\mathcal{L}}\}$. Consider (4) and the scenario in Fig. 2, that can be extended to an arbitrary number of points $\mathbf{p}_i^{\mathcal{W}}$, with corresponding features $\mathbf{f}_i^{\mathcal{L}}$ and $\mathbf{f}_i^{\mathcal{F}}$, on $\mathcal{L}$ and $\mathcal{F}$, respectively. Furthermore, consider the five received features $\mathbf{s}_i^{\mathcal{L}}$, in the $\mathcal{L}$ frame, and five matched features $\mathbf{s}_i^{\mathcal{F}}, i = 1, \ldots, 5$, in the $\mathcal{F}$ frame. The feature-matching is done through robust and scale-invariant feature descriptors (such as SIFT) and geometric verification (with RANSAC). Given a desired essential matrix $\bar{\mathbf{E}}_{\mathcal{F}}^{\mathcal{L}}$, the desired epipolar curves ${}_\diamond\bar{\mathbf{l}}_i^{\mathcal{F}}$ are obtain through

$${}_\diamond\bar{\mathbf{l}}_i^{\mathcal{F}} = {}_\diamond\bar{\mathbf{E}}_{\mathcal{F}}^{\mathcal{L}}\, {}_\diamond\mathbf{s}_i^{\mathcal{L}}, \tag{13}$$

where ${}_{\diamond}\bar{\mathbf{E}}^{\mathcal{L}}_{\mathcal{F}}$ and ${}_{\diamond}\mathbf{s}^{\mathcal{L}}_{i}$ are the lifted representation of $\bar{\mathbf{E}}^{\mathcal{L}}_{\mathcal{F}}$ and $\mathbf{s}^{\mathcal{L}}_{i}$, respectively. Due to (9) and (10), if ${}_{\diamond}\mathbf{s}^{\mathcal{F}T}_{i}\,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}=0 \Leftrightarrow {}_{\diamond}\mathbf{s}^{\mathcal{F}T}_{i}\,{}_{\diamond}\bar{\mathbf{E}}^{\mathcal{L}}_{\mathcal{F}}\,{}_{\diamond}\mathbf{s}^{\mathcal{L}}_{i}=0$, then the two agents are in their desired relative poses given by $\bar{\mathbf{E}}^{\mathcal{L}}_{\mathcal{F}}$, up to a scale factor.

However, if ${}_{\diamond}\mathbf{s}^{\mathcal{F}}_{i}$ does not overlap with ${}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}$, then there exists an algebraic error corresponding to ${}_{\diamond}\mathbf{s}^{\mathcal{F}T}_{i}\,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}\neq 0 \Leftrightarrow {}_{\diamond}\mathbf{s}^{\mathcal{F}T}_{i}\,{}_{\diamond}\bar{\mathbf{E}}^{\mathcal{L}}_{\mathcal{F}}\,{}_{\diamond}\mathbf{s}^{\mathcal{L}}_{i}\neq 0$. It is, then, our control objective to align the locally observed features ${}_{\diamond}\mathbf{s}^{\mathcal{F}}_{i}$ with the epipolar curves ${}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}$. This algebraic error can be re-written as

$$
\begin{aligned}
{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}T}_{1}\,{}_{\diamond}\mathbf{s}^{\mathcal{F}}_{1} &= \begin{bmatrix} a & b & c & d & e & 1\end{bmatrix}\begin{bmatrix} x^2 & xy & y^2 & x & y & 1\end{bmatrix}^T \\
&= \begin{bmatrix} x & y & 1\end{bmatrix}\underbrace{\frac{1}{2}\begin{bmatrix} 2a & b & d \\ b & 2c & e \\ d & e & 2\end{bmatrix}}_{\mathbf{M}_1}\begin{bmatrix} x \\ y \\ 1\end{bmatrix} \\
&= \mathbf{s}^{\mathcal{F}T}_{1}\mathbf{M}_1\mathbf{s}^{\mathcal{F}}_{1}
\end{aligned}
$$

for a lifted curve ${}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{1}\sim\begin{bmatrix} a & b & c & d & e & 1\end{bmatrix}$ with constants $a,\dots,e\in\mathbb{R}$, as in [18], numerically obtained in runtime. For the case in which $\mathcal{L}$ is static, then the curves ${}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}$ are also static in the optimization horizon N of the FHOC in (11). However, if $\mathcal{L}$ moves in the workspace, then the features $\mathbf{f}^{\mathcal{L}}_{i}$ will move according to (2), for which we need the distortion parameter α, as in Section II-A, and the predicted control sequence $\{\mathbf{u}^{\mathcal{L}}(k|k),\dots,\mathbf{u}^{\mathcal{L}}(k+N-1|k)\}$ of the leader $\mathcal{L}$ to calculate predicted curves $\{{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}(k+1|k),\dots,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}}_{i}(k+N|k)\}$. Accordingly, the error function $\boldsymbol{\epsilon}^{\mathcal{F}}(k+n|k):=\psi^{\mathcal{F}}_{2}(\boldsymbol{\xi}^{\mathcal{F}},\bar{\boldsymbol{\xi}}^{\mathcal{F}},\boldsymbol{\iota}^{\mathcal{L}})$, at each time-step k, for the two-agent formation is defined as

$$
\boldsymbol{\epsilon}^{\mathcal{F}}(k+n|k)=\begin{bmatrix} \frac{1}{2}(\|\hat{\mathbf{t}}^{\mathcal{L}}_{\mathcal{F}}(k+n|k)\|^2-\|\bar{\mathbf{t}}^{\mathcal{L}}_{\mathcal{F}}\|^2) \\ \frac{1}{2}\mathbf{s}^{\mathcal{F}}_{1}(k+n|k)^T\mathbf{M}_1(k+n|k)\mathbf{s}^{\mathcal{F}}_{1}(k+n|k) \\ \vdots \\ \frac{1}{2}\mathbf{s}^{\mathcal{F}}_{5}(k+n|k)^T\mathbf{M}_5(k+n|k)\mathbf{s}^{\mathcal{F}}_{5}(k+n|k)\end{bmatrix}, \tag{14}
$$

with $\boldsymbol{\epsilon}^{\mathcal{F}}(k)\in E^{\mathcal{F}}\subset\mathbb{R}^6$ where $E^{\mathcal{F}}$ is a polytope, and where $\|\hat{\mathbf{t}}^{\mathcal{L}}_{\mathcal{F}}(k+n|k)\|^2-\|\bar{\mathbf{t}}^{\mathcal{L}}_{\mathcal{F}}\|^2$ represents the relative distance error, and $\mathbf{s}^{\mathcal{F}}_{i}(k+n|k)^T\mathbf{M}_i(k+n|k)\mathbf{s}^{\mathcal{F}}_{i}(k+n|k)$ represents the algebraic curve-distance errors, for $i=1,...5$ and $n=0,...,N$. The dynamics of the error (14) are derived from (5) and (13), and are given by

$$
\dot{\boldsymbol{\epsilon}}^{\mathcal{F}}=\underbrace{\begin{bmatrix} \mathbf{t}^{\mathcal{L}T}_{\mathcal{F}}\begin{bmatrix}-\mathbf{I} & -\mathbf{t}^{\mathcal{L}}_{\mathcal{F}_\times}\end{bmatrix} \\ \begin{bmatrix}\mathbf{s}^{\mathcal{F}T}_{1}\mathbf{M}_1\end{bmatrix}_{1:2}\mathbf{L}_1(\mathbf{f}^{\mathcal{F}}_{1}) \\ \vdots \\ \begin{bmatrix}\mathbf{s}^{\mathcal{F}T}_{5}\mathbf{M}_5\end{bmatrix}_{1:2}\mathbf{L}_5(\mathbf{f}^{\mathcal{F}}_{5})\end{bmatrix}}_{f^{\mathcal{F}}_{g}:=f_g(\boldsymbol{\epsilon}^{\mathcal{F}},\boldsymbol{\xi}^{\mathcal{F}},\boldsymbol{\iota}^{\mathcal{F}})}\mathbf{u}^{\mathcal{F}}+\underbrace{\begin{bmatrix}\mathbf{t}^{\mathcal{L}T}_{\mathcal{F}}\begin{bmatrix}\mathbf{R}^{\mathcal{F}}_{\mathcal{L}} & \mathbf{0}\end{bmatrix}\mathbf{u}^{\mathcal{L}} \\ \frac{1}{2}\mathbf{s}^{\mathcal{F}T}_{1}\dot{\mathbf{M}}_1\mathbf{s}^{\mathcal{F}}_{1} \\ \vdots \\ \frac{1}{2}\mathbf{s}^{\mathcal{F}T}_{5}\dot{\mathbf{M}}_5\mathbf{s}^{\mathcal{F}}_{5}\end{bmatrix}}_{f^{\mathcal{F}}_{f}:=f_f(\boldsymbol{\epsilon}^{\mathcal{F}},\boldsymbol{\xi}^{\mathcal{F}},\boldsymbol{\iota}^{\mathcal{F}})}. \tag{15}
$$

We now provide the assumptions and conditions for local stability in the neighborhood of the trajectory imposed by the leader.

Assumption 2: During the control task, $f_g(\boldsymbol{\epsilon}^{\mathcal{F}},\boldsymbol{\xi}^{\mathcal{F}},\boldsymbol{\iota}^{\mathcal{F}})$ remains full-rank.

Theorem 1: Consider (5), controlled by the FHOC in (11), where the error $\boldsymbol{\epsilon}^{j}$ is defined in (14). Let Assumptions 1 and 2 hold considering state and control constraints (7) and (8). Consider the feedback controller given by

$$
\mathbf{u}^{\mathcal{F}}_{inv}(k)=f^{\mathcal{F}T}_{g}\cdot(f^{\mathcal{F}}_{g}f^{\mathcal{F}T}_{g})^{-1}\cdot\left(-f^{\mathcal{F}}_{f}-\frac{\mathbf{S}}{h}\boldsymbol{\epsilon}(k)\right), \tag{16}
$$

where $\mathbf{S}\succ 0$ is a diagonal matrix, and $h>0$ the sampling time, such that

$$
V(g^{\mathcal{F}}(\boldsymbol{\epsilon}^{\mathcal{F}},\mathbf{u}^{\mathcal{F}}_{inv}))-V(\boldsymbol{\epsilon}^{\mathcal{F}})+l(\boldsymbol{\epsilon}^{\mathcal{F}},\mathbf{u}^{\mathcal{F}}_{inv})\leq 0,\forall\boldsymbol{\epsilon}\in\Omega,
$$

where Ω is a terminal set defined by $\Omega:=\{\boldsymbol{\epsilon}\in E:V(\boldsymbol{\epsilon})\leq\delta\}$. Then, the system asymptotically converges to $\boldsymbol{\epsilon}^{\mathcal{F}}=\mathbf{0}$ as $t\to\infty$.

The proof can be found in [27].

C. Formation Control

Assuming that at least one follower is in formation with the leader (without loss of generality, let it be $\mathcal{F}_1$) using the MP-IBVS controller proposed in Section IV-B, we propose a control scheme that is capable of driving an agent to the correct relative pose in the formation based solely on image features, addressing Problem 2. In other words, any other agent in $\mathfrak{G}\backslash\{\mathcal{L},\mathcal{F}_1\}$ can achieve their desired formation pose without the need for range measurements.

Consider the scenario in Fig. 2, where agents $\mathcal{L}$ and $\mathcal{F}_1$ are at the desired relative pose, and follower $\mathcal{F}_j$ receives image features from both agents. In this setting, we show that the relative pose of the follower $\mathcal{F}_j$ with respect to $\mathcal{L}$ and $\mathcal{F}_1$ is uniquely defined by the epipolar geometry shared by the three camera system, assuming that the closest image solution is the desired one. This property can be extended to any agent $\mathcal{F}_j, j=2,...,M-2$.

Let ${}_{\diamond}\mathbf{s}^{\mathcal{F}_j}_{i}, i=1,\dots,5$, be the set of features observed by the follower $\mathcal{F}_j, j=2,\dots,M-2$, and epipolar curves sets ${}_{\diamond}\bar{L}_{\mathcal{L}}=\{{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{i,\mathcal{L}}\}$ and ${}_{\diamond}\bar{L}_{\mathcal{F}_1}=\{{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{i,\mathcal{F}_1}\}$ relative to the leader $\mathcal{L}$ and follower $\mathcal{F}_1$, where ${}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{i,l}={}_{\diamond}\bar{\mathbf{E}}^{l}_{\mathcal{F}_j}\,{}_{\diamond}\mathbf{s}^{l}_{i},\ l\in\{\mathcal{L},\mathcal{F}_1\},\ i=1,\dots,5$. The relative pose, up to a scale factor, that minimizes the error between $\mathbf{f}^{\mathcal{F}_j}_{i}$ and the curve sets ${}_{\diamond}\bar{L}_{\mathcal{L}}$ and ${}_{\diamond}\bar{L}_{\mathcal{F}_1}$ is the one where all features in $\mathbf{f}^{\mathcal{F}_j}_{i}$ lie on top of the respective curves in ${}_{\diamond}\bar{L}_{\mathcal{L}}$ or ${}_{\diamond}\bar{L}_{\mathcal{F}_1}$. However, with two sets of curves, which represent the desired relative pose to the leader $\mathcal{L}$ and follower $\mathcal{F}_1$, the only pose that minimizes both errors is the intersection of the curves in the sets ${}_{\diamond}\bar{L}_{\mathcal{L}}$ and ${}_{\diamond}\bar{L}_{\mathcal{F}_1}$. Since such curves can have up to four intersections, in practice we must assume that the intersection we are interested in is the closest to the observed feature. This is a common assumption in IBVS approaches [7], where it is reasonable to expect that the error to the desired configuration in the image plane is kept small. Note that agents $\mathcal{L}$ and $\mathcal{F}_1$ must be in the correct formation geometry (after applying the method of Section IV-B) for the $M>2$ agent to converge to the desired relative position.

Let $\bar{F}_{\mathcal{LF}_1}=\{\bar{\mathbf{f}}^{\mathcal{LF}_1}_{1},...,\bar{\mathbf{f}}^{\mathcal{LF}_1}_{5}\}$ be the image features corresponding to the intersection of the curve sets ${}_{\diamond}\bar{L}_{\mathcal{L}}$ and ${}_{\diamond}\bar{L}_{\mathcal{F}_1}$ (in the $\mathcal{F}_j$ image-plane). In this case, we wish to minimize the error between $\mathbf{f}^{\mathcal{F}_j}_{i}$ and $\bar{\mathbf{f}}^{\mathcal{LF}_1}_{i}, i=1,\dots,5,\ j=2,\dots,M-2$. We can obtain the predicted line sets ${}_{\diamond}\bar{L}^{\mathcal{L}}_{N}=\{{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{1,\mathcal{L}}(k+1|k),\dots,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{5,\mathcal{L}}(k+1|k),\dots,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{1,\mathcal{L}}(k+N|k),\dots,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{5,\mathcal{L}}(k+N|k)\}$ and ${}_{\diamond}\bar{L}^{\mathcal{F}_1}_{N}=\{{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{1,\mathcal{F}_1}(k+1|k),\dots,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{5,\mathcal{F}_1}(k+1|k),\dots,{}_{\diamond}\bar{\mathbf{l}}^{\mathcal{F}_j}_{1,\mathcal{F}_1}$

$(k+N|k),\ldots,\bar{\mathbf{l}}^{\mathcal{F}_j}_{5,\mathcal{F}_1}(k+N|k)\}$ with the information vectors $\boldsymbol{\iota}^l(k)=\{\mathbf{s}^l_1(k),\ldots,\mathbf{s}^l_5(k),\alpha^l,\mathbf{u}^l_N\}$, $l=\mathcal{L},\mathcal{F}_1$, for two neighbors, (13) and a proper discretization of (2). In this case, the error function $\boldsymbol{\epsilon}^{\mathcal{F}_j}(k+n|k):=\psi^{\mathcal{F}_j}_3(\boldsymbol{\xi}^{\mathcal{F}_j},\bar{\boldsymbol{\xi}}^{\mathcal{F}_j},\boldsymbol{\iota}^l)$, at each time-step k, for the three agent formation is defined as

$$\boldsymbol{\epsilon}^{\mathcal{F}_j}(k+n|k)=\begin{bmatrix}\bar{\mathbf{f}}^{\mathcal{L}\mathcal{F}_1}_1(k+n|k)-\mathbf{f}^{\mathcal{F}_j}_1(k+n|k)\\ \vdots\\ \bar{\mathbf{f}}^{\mathcal{L}\mathcal{F}_1}_5(k+n|k)-\mathbf{f}^{\mathcal{F}_j}_5(k+n|k)\end{bmatrix},\quad(17)$$

where $\boldsymbol{\epsilon}^{\mathcal{F}_j}\in E^j\subset\mathbb{R}^{10}$ concatenates the image-based feature errors.

Assumption 3: There exists a terminal controller $\mathbf{u}^{\mathcal{F}_j}$ such that

$$V(g^{\mathcal{F}_j}(\boldsymbol{\epsilon}^{\mathcal{F}_j},\mathbf{u}^{\mathcal{F}_j}))-V(\boldsymbol{\epsilon}^{\mathcal{F}_j})+l(\boldsymbol{\epsilon}^{\mathcal{F}_j},\mathbf{u}^{\mathcal{F}_j})\leq 0,\forall\boldsymbol{\epsilon}^{\mathcal{F}_j}\in\Omega^{\mathcal{F}_j},$$

where $\Omega^{\mathcal{F}_j}:=\{\boldsymbol{\epsilon}^{\mathcal{F}_j}\in E:V(\boldsymbol{\epsilon}^{\mathcal{F}_j})\leq\delta^{\mathcal{F}_j}\}$ is a terminal set in the neighborhood of $\boldsymbol{\epsilon}^{\mathcal{F}_j}=\mathbf{0}$.

Remark 2: Assumption 3 is common to provide feasibility and stability results in MPC schemes [26], [28]. The assumption is fair provided that the agents have sufficient control input authority to track the formation geometry, as the intersection of the 4th-order parabolic curves is algorithmic and, to the best of our knowledge, it is not possible to derive the closed-form dynamics of its intersection.

Theorem 2: Consider (6) controlled by (11) and the error (17). Let Assumptions 1 and 3 hold considering state and control constraints (7) and (8). Then, the error (17) asymptotically converges to $\boldsymbol{\epsilon}^{\mathcal{F}_j}=\mathbf{0},j=2,\ldots,M-1$ as $t\to\infty$.

Proof: It is known [26], [28] that under validity of Assumption 3, the MPC scheme in (11) is asymptotically stable. Assumption 3 provides a terminal set $\Omega^{\mathcal{F}_j}$ in which the discrete dynamics of $\boldsymbol{\epsilon}^{\mathcal{F}_j}$ are invariant given $\mathbf{u}^{\mathcal{F}_j}$, ensuring recursive feasibility. Then, with $V(g^{\mathcal{F}_j}(\boldsymbol{\epsilon}^{\mathcal{F}_j},\mathbf{u}^{\mathcal{F}_j}))-V(\boldsymbol{\epsilon}^{\mathcal{F}_j})+l(\boldsymbol{\epsilon}^{\mathcal{F}_j},\mathbf{u}^{\mathcal{F}_j})\leq 0,\forall\boldsymbol{\epsilon}^{\mathcal{F}_j}\in\Omega^{\mathcal{F}_j}$, asymptotic stability is guaranteed, concluding the proof, and showing that (17) asymptotically converges to $\boldsymbol{\epsilon}^{\mathcal{F}_j}=\mathbf{0}$. □

V. Results

In this section, we detail the simulation and experimental results collected using the proposed control schemes in Section IV-B and Section IV-C. The algorithms were implemented in the EpiC library, available at: https://github.com/KTH-DHSG/epic.

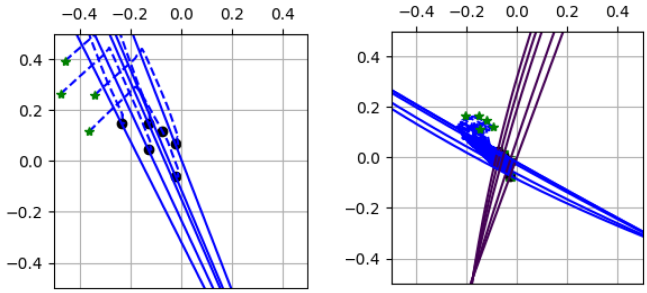
A. Algorithm

In Algorithm 1, we detail the algorithmic implementation of performing formation control with the solutions for Problems 1 and 2. We consider an arbitrary number of agents but assume, without loss of generality, that Follower 1 has range-measuring capabilities capabilities to solve Problem 1.

B. Simulation Results

The simulated results were collected using Python 3.8 in a laptop with a Core-i7 11800H @ 2.30GHz and 16GB of DDR4 memory at 3200 MHz. The results were collected with the Python script in `demo/multi_agent_dynamic.py`. The simulation starts with the leader and five followers spread close to their formation geometries but with a randomized error. The presented results showcase a scenario in which the agents need to converge to their desired relative poses while the leader is moving at 1 cm/s. The controller parameters can be found in the `epic/config` folder, while the geometry centers (poses around which we introduce a randomized error) are available in the file `demo/6_agents_formation.json`.

In Figs. 3 and 4 we show the image plane for 2 followers involved in the formation task and the relative pose error of



(a) Feature motion for for $\mathcal{F}_1$. (b) Feature motion for $\mathcal{F}_3$.

Fig. 3. The normalized image plane contains the epipolar curves for the respective neighbors, the intersection of the two sets of epipolar curves (green cross) and the observed features (black blobs). In (a) and (b) it is possible to observe the initial feature position for the agent (green stars), with an error towards the curves and their intersection, respectively. In black, is the final feature location, while the dashed line shows the motion of the corresponding feature.

Algorithm 1: Image-based Formation Control for M Agents.

Require $T\geq 0$ total simulation time
Require $M\geq 2$ two-agents minimum
Require $\mathbf{u}^{\mathcal{L}}$ leader guidance
while $t\leq T$ **do**
  for all $i\in\mathfrak{G}$ **do**
    if $i=\mathcal{L}$ **then**
      $F^{\mathcal{L}}\leftarrow$ FeatureExtraction()
      $\boldsymbol{\xi}^{\mathcal{L}}\leftarrow$ PropagateMotion($\boldsymbol{\xi}^{\mathcal{L}},\mathbf{u}^{\mathcal{L}}$)
      $\boldsymbol{\iota}^{\mathcal{L}}\leftarrow$ BroadcastInformation() {information for all followers}
    else if $i=\mathcal{F}_1$ **then**
      $F^{\mathcal{F}_1}\leftarrow$ FeatureExtraction()
      $\mathbf{u}^{\mathcal{F}_1}\leftarrow$ IBRC($F^{\mathcal{F}_1},\boldsymbol{\iota}^{\mathcal{L}}$) {Solve Problem 1}
      $\boldsymbol{\xi}^{\mathcal{F}_1}\leftarrow$ PropagateMotion($\boldsymbol{\xi}^{\mathcal{F}_1},\mathbf{u}^{\mathcal{F}_1}$)
      $\boldsymbol{\iota}^{\mathcal{F}_1}\leftarrow$ BroadcastInformation()
    else
      $j,o\leftarrow$ Neighbors() {Agent neighbors, e.g. $j\leftarrow\mathcal{L},o\leftarrow\mathcal{F}_1$}
      $F^i\leftarrow$ FeatureExtraction()
      $\mathbf{u}^i\leftarrow$ IBFC($F^i,\boldsymbol{\iota}^j,\boldsymbol{\iota}^o$) {Solve Problem 2}
      $\boldsymbol{\xi}^i\leftarrow$ PropagateMotion($\boldsymbol{\xi}^i,\mathbf{u}^i$)
      $\boldsymbol{\iota}^i\leftarrow$ BroadcastInformation()
    end if
  end for
  $t\leftarrow t+\Delta t$
end while

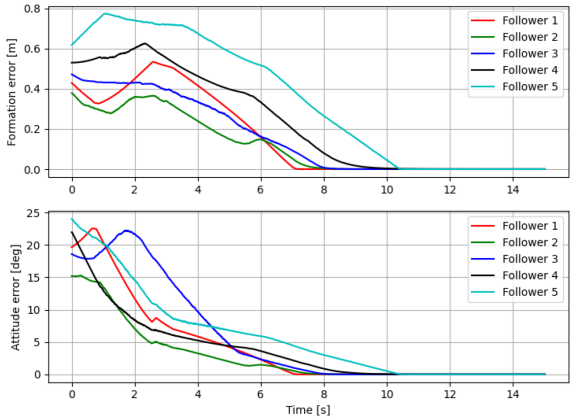


Fig. 4. Simulation results for the six-agent formation. Convergence is observed while respecting the state and input constraints. Followers 1 and 5 are perspective cameras, 2 and 4 are parabolic cameras, while follower 3 is hyperbolic.

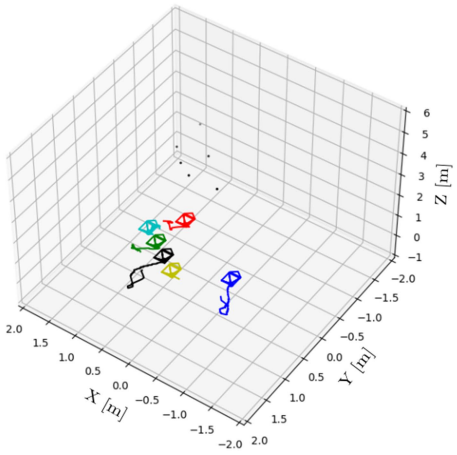
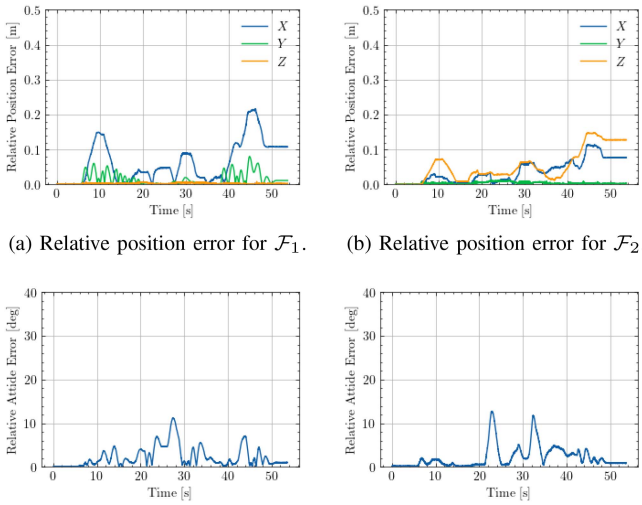


Fig. 5. Trajectory of the agents during convergence and tracking. In red, the leader $\mathcal{L}$, and in blue, the range-measuring follower $\mathcal{F}_1$. The other agents use only the image features.

the group, respectively, for a time-period of 15 seconds. The agent's trajectory is seen in Fig. 5. We observe that all agents converge to their desired formation geometry and that the tracking is achieved with zero steady-state error. The computational time was, on average, approximately $51ms$ and $16ms$ with prediction horizons of $N = 20$ and $N = 5$ for the controllers in Section IV-B and Section IV-C, respectively. An animation is available in `data/simulated_animation.mp4` in the EpiC repository. Running the demo script will provide more insight into control bounds, image plane, and CPU time. The initial state is normally randomized with a mean error of 50 cm in position and 20 deg in attitude.

C. Experimental Results

Experimental results were conducted in a laboratory environment equipped with a Motion Capture System and three holonomic mobile bases. These mobile bases have 3 Degrees-of-Freedom (DoF), which are 3 less than the simulated setup.



(a) Relative position error for $\mathcal{F}_1$. (b) Relative position error for $\mathcal{F}_2$.

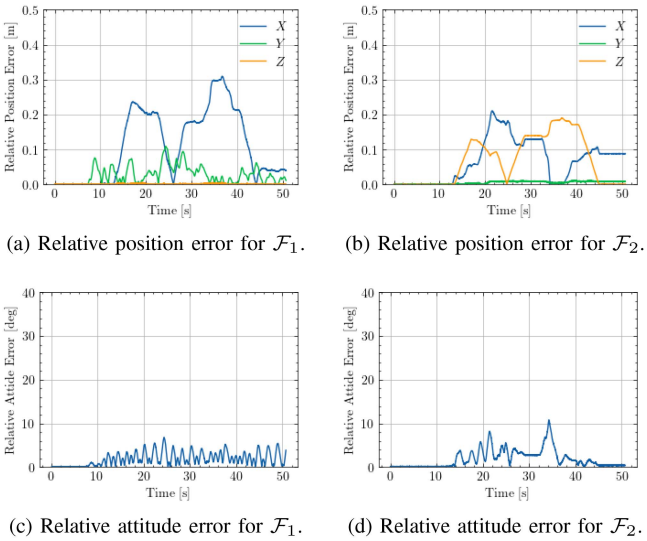
(c) Relative attitude error for $\mathcal{F}_1$. (d) Relative attitude error for $\mathcal{F}_2$.

Fig. 6. During a 55-second formation control task with no noise, we observed that the agents kept in track with the formation leader with an average attitude error of 3 degrees and an average position error of 7cm.

However, in the 2D cartesian plane, these robots can move independently in rotation and translation. The setup can be seen in Fig. 1.

On the mobile bases, we run a SIFT feature extractor and matching pipeline. In particular, each agent extracts 200 SIFT features from the environment, and then $\mathcal{F}_1$ matches 5 of these features with the leader, while $\mathcal{F}_2$ finds 5 features in common both with the leader and $\mathcal{F}_1$. A motivation to use a SIFT feature pipeline is the capability to operate in real unstructured environments and to transition toward outdoor conditions. For the range, we took advantage of the Motion Capture System and ran two scenarios, i) with perfect range estimation, and ii) with a simulated ultra-wideband ranging node, modeled by noise with 0.01[m] of mean and 0.1[m] of variance. The leader $\mathcal{L}$ was manually controlled through a number of set-points in the environment, with translations and rotations of approximately 0.4[m] and 10[deg], while followers $\mathcal{F}_1$ and $\mathcal{F}_2$ were controller with the proposed methods in Section IV-B and Section IV-C.

In Fig. 6 we can observe that formation task performed during a 50-second maneuver with a perfect range estimation. We observe that the relative attitude error is kept small during the entire time, with an average of error 3 degrees. The relative position error is kept on. In Fig. 7 however, we ran the same tests but with a noisy range measurement, equivalent to a ultra-wideband device. It is worth noting that, although this noise highly impacts the follower $\mathcal{F}_1$, the attitude error is still kept low, while the relative position error worsens. Still, in this harsh scenario, the follower $\mathcal{F}_2$ kept a lower relative position error than $\mathcal{F}_1$, as this agent also receives measurements from the leader that help with lowering the sensitivity of the controller to the harsh maneuvers of $\mathcal{F}_1$. For $\mathcal{F}_1$, the attitude error was kept below 12 degrees at all times, with an average of 4 degrees during the task execution. The `data/` folder in our repository provides videos and telemetry of both experiments.



(a) Relative position error for $\mathcal{F}_1$. (b) Relative position error for $\mathcal{F}_2$.

(c) Relative attitude error for $\mathcal{F}_1$. (d) Relative attitude error for $\mathcal{F}_2$.

Fig. 7. During a 50-second formation control task with range noise, the agents kept an average attitude error of 4 degrees and an average position error of 12cm. The range noise is characterized by a mean of 0.01[m] and a variance of 0.1[m].

VI. Conclusions and Future Work

In this letter, we derived two control laws using epipolar constraints for formation control: one for coordination among two agents, and another for formation control considering a higher number of agents, observing common points of interest in the world. To the best of our knowledge this is the first work that achieves visual coordination considering a single range measurement, in a multi-agent formation setting. Experimental results provided an insight into the performance of the proposed methods in a realistic scenario.

In the future, we will investigate the application of such strategies to higher-order nonlinear systems, and explore how to use the proposed framework in an active-vision setting.

References

[1] K. Guo, X. Li, and L. Xie, "Ultra-wideband and odometry-based cooperative relative localization with application to multi-UAV formation control," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2590–2603, Jun. 2020.

[2] D. K. Villa, A. S. Brandao, and M. Sarcinelli-Filho, "A survey on load transportation using multirotor UAVs," *J. Intell. Robot. Syst.*, vol. 98, pp. 267–296, 2020.

[3] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Nob Hill Pub., 2009.

[4] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[5] P. Roque, S. Heshmati-Alamdari, A. Nikou, and D. V. Dimarogonas, "Decentralized formation control for multiple quadrotors under unidirectional communication constraints," *IFAC Proc.*, vol. 53, pp. 3156–3161, 2020.

[6] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Automat.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.

[7] F. Chaumette and S. Hutchinson, "Visual servo control. I. Basic approaches," *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[8] P. Rives, "Visual servoing based on epipolar geometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2000, pp. 602–607.

[9] G. L. Mariottini, G. Oriolo, and D. Prattichizzo, "Image-based visual servoing for nonholonomic mobile robots using epipolar geometry," *IEEE Trans. Robot.*, vol. 23, no. 1, pp. 87–100, Feb. 2007.

[10] E. Montijano, J. Thunberg, X. Hu, and C. Sagues, "Epipolar visual servoing for multirobot distributed consensus," *IEEE Trans. Robot.*, vol. 29, no. 5, pp. 1212–1225, Oct. 2013.

[11] E. Montijano, E. Cristofalo, D. Zhou, M. Schwager, and C. Sagues, "Vision-based distributed formation control without an external positioning system," *IEEE Trans. Robot.*, vol. 32, no. 2, pp. 339–351, Apr. 2016.

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[13] S. Maniatopoulos, D. Panagou, and K. J. Kyriakopoulos, "Model predictive control for the navigation of a nonholonomic vehicle with field-of-view constraints," in *Proc. IEEE Amer. Control Conf.*, 2013, pp. 3967–3972.

[14] D. Panagou and V. Kumar, "Cooperative visibility maintenance for leader–follower formations in obstacle environments," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 831–844, Aug. 2014.

[15] F. Poiesi and A. Cavallaro, "A distributed vision-based consensus model for aerial-robotic teams," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 169–176.

[16] M. Aranda, Y. Mezouar, G. López-Nicolás, and C. Sagüés, "Scale-free vision-based aerial control of a ground formation with hybrid topology," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 4, pp. 1703–1711, Jul. 2019.

[17] R. T. Rodrigues, P. Miraldo, D. V. Dimarogonas, and A. P. Aguiar, "A framework for depth estimation and relative localization of ground robots using computer vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 3719–3724.

[18] J. P. Barreto and K. Daniilidis, "Epipolar geometry of central projection systems using veronese maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 1258–1265.

[19] A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. I–I.

[20] G. Caron, E. Marchand, and E. M. Mouaddib, "Photometric visual servoing for omnidirectional cameras," *Auton. Robots*, vol. 35, no. 2, pp. 177–193, 2013.

[21] R. Spica and P. R. Giordano, "A framework for active estimation: Application to structure from motion," in *Proc. IEEE Conf. Decis. Control*, 2013, pp. 7647–7653.

[22] D. Wofk, F. Ma, T.-J. Yang, S. Karaman, and V. Sze, "FastDepth: Fast monocular depth estimation on embedded systems," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 6101–6108.

[23] R. T. Rodrigues, P. Miraldo, D. V. Dimarogonas, and A. P. Aguiar, "Active depth estimation: Stability analysis and its applications," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2002–2008.

[24] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–770, Jun. 2004.

[25] M. Sauvée, P. Poignet, E. Dombre, and E. Courtial, "Image based visual servoing through nonlinear model predictive control," in *Proc. IEEE Conf. Decis. Control*, 2006, pp. 1776–1781.

[26] J. B. Rawlings, D. Q. Mayne, and M. Diehl, *Model Predictive Control: Theory, Computation, and Design*, vol. 2. Santa Barbara, California: Nob Hill Publishing Madison, 2017.

[27] "Multi-agent formation control using epipolar constraints - extended version," 2024. Accessed: Aug. 6th, 2020. [Online]. Available: https://pedroroque.dev/assets/pdf/RAL2024_extended.pdf

[28] D. Limón, T. Alamo, F. Salas, and E. F. Camacho, "On the stability of constrained MPC without terminal constraint," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 832–836, May 2006.