

Practical Formation Acquisition Mechanism for Nonholonomic Leader-follower Networks

Kader Monhamady Kabore^a and Samet Güler^b

Dept. Electrical and Electronics Eng., Abdullah Gül University, Barbaros, Kayseri, 38080, Turkey

Keywords: Multi-robot Formation Control, Directed Graphs, Convolutional Neural Networks.

Abstract: A grand challenge lying ahead of the realization of multi-robot systems is the lack of an adequate coordination mechanism with reliable localization solutions. In some workspaces, external infrastructure needed for precise localization may not be always available to the MRS, e.g., GPS-denied environments, and the robots may need to rely on their onboard resources without explicit communication. We address the practical formation control of nonholonomic ground robots where external localization aids are not available. We propose a systematic framework for the formation maintenance problem that is composed of a localization module and a control module. The onboard localization module relies on heterogeneity in sensing modality comprised of ultrawideband, 2D LIDAR, and camera sensors. Particularly, we apply deep learning-based object detection algorithm to detect the bearing between robots and fuse the outcome with ultrawideband distance measurements for precise relative localization. Integration of the localization outcome into a distributed formation acquisition controller yields high performance. Furthermore, the proposed framework can eliminate the magnetometer sensor which is known to produce unreliable heading readings in some environments. We conduct several realistic simulations and real world experiments whose results validate the competency of the proposed solution.


1 INTRODUCTION


There has been a long-standing interest in multi-robot systems (MRS) to address complex tasks beyond the ability of a single robot. While equipping a single robot with excessive load and highly capable sensor suits could be a viable option, swarming offers a promising direction to handle challenging applications. A recent survey highlights the advances in aerial swarms for critical tasks such as security and surveillance, collaborative transportation, and environment monitoring (Chung et al., 2018; Almadhoun et al., 2016; Abdelkader et al., 2021). For instance, in a dramatic search and rescue operation, a group of drones can rapidly map and provide insights about an area which may be hard to access (Tian et al., 2020). In addition, multiple agents can easily operate in environments which can be dangerous for humans. For instance, an aerial robot team can inspect safely electrical power lines or wind turbines (Silano et al., 2021). Therefore, MRS will positively impact multiple work fields. However, it is worth noticing that a safe re-

alization of these applications with MRS entails to solve challenges in several subfields such as formation control and multi-agent localization. Decades of research have contributed to these fields, yet there is still a long road ahead to reaching a complete decentralized MRS.

On the formation control side, several studies have focused on maintaining the desired shape pattern. References (Anderson et al., 2008; Oh et al., 2015) summarize various techniques on the formation acquisition using the distances between agents only. In the same direction, the authors of (Zhao and Zelazo, 2019; Trinh et al., 2018) demonstrate control laws based on the bearing vectors among agents. Various stable and convergent algorithms have been proposed for single- or double-integrator kinematic agents in primitive simulation environments by assuming almost perfect agent dynamics and sensing modalities. However, these assumptions usually do not reflect the real life experimental scenarios, particularly when the robots employ onboard sensors only.

When it comes to localization, the primary objective is to design a decentralized system and localization methods without the reliance on infrastruc-

^a  <https://orcid.org/0000-0001-5388-9649>

^b  <https://orcid.org/0000-0002-9870-166X>

tures such as Global Positioning System (GPS) or motion capture (mocap) cameras. Such infrastructures can provide a MRS with positioning solutions with millimeter-level precision. However, the MRS may be desired to operate in critical environments, and the positioning system should not stand as a barrier to the operation. An agent should be able to maneuver safely indoors and outdoors with the proposed localization method. To address this challenge, previous research have proposed to use additional markers as a pattern for relative localization. The study in (Saska et al., 2017) illustrates that a circular marker with a predefined diameter can be used with a flood and fill image processing technique to estimate the relative position between two agents. Similarly, special color markers could be placed on a specific agent to be identified to extract the relative positions between the agents (Fidan et al., 2012; Lin et al., 2020). Likewise, this work (Walter et al., 2018) introduces ultraviolet (UV) instead of markers and uses a UV recognizable camera to process the blinking signals and compute the relative positions. The marker approach is proven effective to identify the relative location but offers a solution solely constrained by the unique design of the marker. Additionally, this approach scales poorly: When the number of agents increases, different patterns need to be presented to identify each agent in the swarm (Saska et al., 2017; Fidan et al., 2012).

A novel approach presents an infrastructure-free and completely onboard alternative for relative localization for a leader-follower scheme (Guler et al., 2020; Nguyen et al., 2020). The follower is equipped ultrawideband (UWB) sensors with known position and estimates its relative position to a leader with a single UWB sensor. The UWB distance measurements are fused in a filtering algorithm to extract an estimate of the relative position. Besides range sensors, cameras and the prospect of vision have been implemented as well (Kabore and Güler, 2021a; Kabore and Güler, 2021b; Liu et al., 2018). The authors in (Vrba and Saska, 2020) apply a convolutional neural network (CNN) model to detect a leader drone and use the camera intrinsic parameters to project the object from 2D to 3D and obtain a moderate estimation of the relative positions. These results demonstrate the effectiveness of the distance and vision sensors for reliable relative localization.

Currently, there is no generic practical localization framework that can enable MRS to operate in any environmental conditions. The proposed theoretical solutions usually remain insufficient to reflect the real-time characteristics of robotic systems. On the other hand, most of the current practical solutions rely on a central unit for computation and com-

munication among robots (De Queiroz et al., 2019; Choi and Kim, 2021), which prevents to implement further completely distributed formation control algorithms. This work aims at proposing a relative localization algorithm for MRS that does not depend on a central computation unit and explicit communication among robots. Thus, the proposed solution allows a distributed implementation of the entire system. Particularly, we apply a heterogeneous sensing model on three non-holonomic ground robots with three sensor types, namely, UWB, camera, and 2D LIDAR, which yields two sensing modalities. We implement a CNN-based object detection to generate the relative bearing between a group of robots. We show that the heterogeneity of the sensing modalities leads to high performance in the formation acquisition objective with the leader-first follower-second follower constraint graph. The main contributions of this work are as follows:

- We propose a systematic framework comprising a relative localization and distributed formation control modules for MRS without explicit communication and exchange of velocities between robots.
- The proposed relative localization system with heterogeneous sensing modalities relies on the onboard units only, and no external infrastructure or ground station is incorporated.
- We demonstrate the practicality of the framework on a three-robot system in both realistic simulations and real experiments.
- We propose a control method which eliminates the magnetometer sensor which may perform unreliably in some environments.

The rest of the work is organized as follows. We define our system model and the problem considered in Section 2. We propose the localization and control modules in Section 3 and Section 4, respectively. Section 5 demonstrates the simulation and experimental results, and Section 6 is on conclusions and future works.

2 SYSTEM MODELING

In this section, we illustrate the mathematical model of the robots used in our system, the underlying graph in the network, and the main objective.

2.1 Preliminaries

We consider a swarm of non-holonomic ground robots $S = \{R_0, R_1, \dots, R_N\}$, $N \geq 2$. Each robot pos-

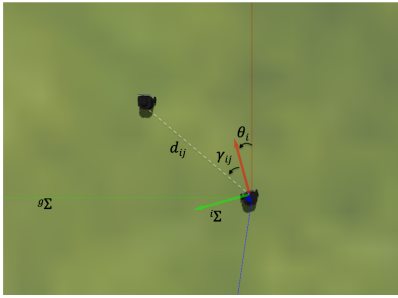


Figure 1: Illustration of agent R_i and the frame constraints.

sesses the unicycle kinematics model as follows:

$$\begin{aligned} \dot{x}_i(t) &= v_i(t) \cos(\theta_i(t)) \\ \dot{y}_i(t) &= v_i(t) \sin(\theta_i(t)) \\ \dot{\theta}_i(t) &= \omega_i(t) \end{aligned} \quad (1)$$

where x_i, y_i are the Cartesian coordinates and $\theta_i(t)$ is the heading angle of robot R_i , v_i and ω_i are the control inputs for linear velocity and the angular velocity about the inertial z -axis, respectively (Fig. 1). Each robot R_i comprises onboard sensing and computational units, and the robots do not interact or communicate between each other or with a ground station except for active sensing (e.g., distance sensing by ultrawideband (UWB) sensors).

The local sensing and communication interactions among agents in a MRS can be described with a graph. We define a leader-follower directed graph $G = (V, E)$ among the agents of the swarm S . Each vertex $i \in V$ represents an agent R_i in S while each directed edge $(i, j) \in E$ from i to j implies the distance constraint d_{ij} between R_i and R_j . Particularly, we constrain our study to the minimally persistent graphs which are minimally rigid and constraint consistent. Such a directed graph can be generated using the Henneberg construction method such that no vertex has more than two outgoing edges (Anderson et al., 2008).

Denote a formation framework $F = (S, G, Q)$ for the swarm S along with the underlying graph structure G and the desired (constraint consistent) distance configuration $Q = \{q_{ij} | (i, j) \in E\}$. We assume a leader-

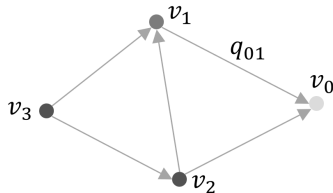


Figure 2: The underlying minimally persistent constraint graph used in this work. The leader node v_0 has no distance constraint to satisfy whereas the first follower node v_1 has one constraint and the second follower node v_2 has two constraints.

first follower formation such that vertex v_0 has no outgoing edge, and thus it does not have any restriction on the mobility. The vertex v_1 has one outgoing edge toward v_0 which implies that it can reside at any point on the circle with the center at v_0 and radius q_{01} (Fig. 2). Finally, each vertex v_i , $i \in \{2, \dots, N\}$ has exactly two outgoing edges toward its leader set $\{v_{i-2}, v_{i-1}\}$, which implies that it has two distance constraints $q_{i-2,i}, q_{i-1,i}$ to satisfy. The graph G constructed by this operation constitutes a minimally persistent graph where the entire vertex set can move up to translation and rotation as long as all distance constraints in Q are satisfied. We assign robot R_0 as the leader, R_1 as the first follower, and the remaining robots R_i , $i \in \{2, \dots, N\}$ as the second followers. Therefore, in the formation $F = (S, G, Q)$, if the follower robots R_i , $i \in \{1, \dots, N\}$ satisfy the distance constraints $q_{ij} \in Q$, the entire formation will move as a whole forming the shape dictated by Q . Notably, the underlying graph G does not depend on the robot motion model. Indeed, the original rigid formation control algorithms are designed for point agent kinematics, e.g., holonomic agents. Here, we will modify these control algorithms to apply on non-holonomic agents.

2.2 Problem Formulation

We formulate our main objective as follows. Consider the swarm $S = \{R_0, R_1, \dots, R_N\}$, $N \geq 2$, and the leader-first follower formation framework $F = (S, G, Q)$ defined above. We assume that all agents in the swarm rely on their onboard resources solely, there is no ground station, and explicit inter-robot communication is not available. Particularly, robot R_0 is equipped with a UWB anchor sensor whereas robot R_1 is mounted with a monocular camera and a UWB sensor. Each of the remaining robots $\{R_2, \dots, R_N\}$ is equipped with a 2D LIDAR sensor. Thus, the sensing mechanism of the robots consists of heterogeneity. The goal is to design (i) a relative localization solution for the given system and (ii) a distributed control law for the robots to converge from suitable initial conditions to a desired formation configuration imposed by $F = (S, G, Q)$. In the sequel, we consider three-robot case ($N = 2$) as a prototype model noting that the proposed framework is scalable to more robotic agents. In Section 5 we discuss the practical details for extension to more number of robots.

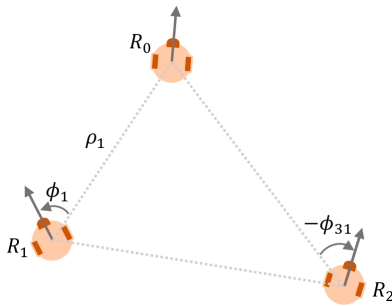


Figure 3: The top-view of three-robot system with the polar dynamics variables.

3 SENSOR NETWORK

We now present the relative localization solution and study the features of the sensors mounted on the robots, namely ultrawideband (UWB) sensors, monocular camera, and 2D LIDAR sensors. We model the hybrid usage of these sensors in our localization system which is solely dependent on onboard sensors and does not exploit any external infrastructure.

Distance and vision measurements can be exploited in a complementarily and efficiently manner to acquire the characteristics of the localized target. Particularly, UWB sensors and cameras can act as dual sensors to provide a solid framework for multi-robot localization. On one side, UWB sensors operate with high bandwidth capability and generate accurate omnidirectional range measurements utilizing the time-of-flight (TOF) method. On the other side, a monocular camera empowered with the recent advancements in computer vision and deep learning techniques can be applied for the visual detection of a leader robot by a follower robot. Particularly, by utilizing the convolutional neural network (CNN)-based detection methods, one can accurately extract a relative bearing between two robots. Moreover, since most commercial versions of these sensors are lightweight, they can be implemented on many unmanned ground vehicles (UGVs) and unmanned aerial vehicles (UAVs) by maintaining feasibility.

We use a pair of UWB sensors, one anchor on robot R_0 and one tag on robot R_1 , together with a monocular camera on R_1 to provide an efficient localization system for the follower robot R_1 . We acquire the distance q_{01} between R_0 and R_1 from the UWB sensors while the monocular camera is used to extract the bearing angle ϕ_1 (Fig. 3).

Furthermore, we propose a CNN model to successfully detect the leader robot R_0 with a monocular camera on robot R_1 . Recently, deep learning tech-

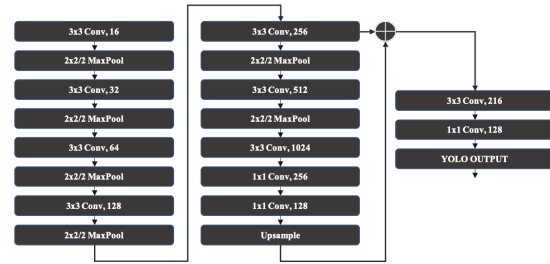


Figure 4: The YOLO Tiny architecture used in the training.

niques have far surpassed the traditional computer vision methods. For instance, in a target tracking task, the classical feature invariant extraction followed by key point matching may not yield precise detection results, particularly when the object is viewed from only one angle and the sensor and the target object are in motion (O'Mahony et al., 2019). In contrast, CNN has been shown to be more effective in dynamics scenarios. We particularly utilize the YOLO v3-Tiny¹, which is an improvement over the previous CNN architectures such as Faster-RCNN (Ren et al., 2015) and YOLO-v3 (Redmon and Farhadi, 2018) with its remarkably light and robust structure adapted for dynamic environments. Consequently, this network is more suitable for real-time detection with on-board limited resources. We redesign the structure of the network to accommodate our single class output and tune the network hyperparameters (Fig. 4). By combining the UWB distance measurements with the bearing data generated by the YOLO v3-Tiny detection, we generate the relative position p_{01} in the body frame of robot R_1 . The object detection module yields bounding boxes of the detected objects, and the image size can be extracted from the monocular camera. Using this intuition, we derive the relative bearing from the target using the middle pixel of the bounding box (Fig. 5a).

2D LIDARs offer a practical detection solution for small-sized robots with limited payload and low en-

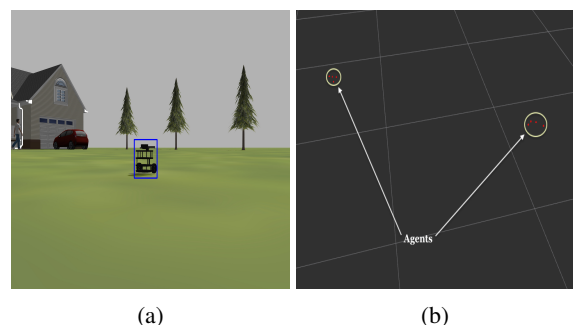


Figure 5: Simulation images captured in Gazebo: (a) Detection in the simulation environment, (b) 2D LIDAR active localization of two agents from point clouds.

¹<https://pjreddie.com/darknet/yolo/>

energy consumption. Moreover, since the motion capability of nonholonomic robots is restrained by nonlinear constraints, large field-of-view of the 2D LIDARs provides a vital advantage on this type of robots. Therefore, we opt for a 2D LIDAR sensor on the second follower robots R_i , $i \in \{2, \dots, N\}$. A second follower robot R_i is to maintain the distances toward its leaders $\{R_{i-2}, R_{i-1}\}$. Thus, it calculates the relative location toward $\{R_{i-2}, R_{i-1}\}$ using the laser scan point cloud and an iterative cluster algorithm. The aim is to highlight the target agent within the range scan of the sensor and find the relative positions p_{02} , p_{12} in the body frame of robot R_i . We first convert the raw scan data to point cloud in 2D to differentiate the robots R_0, R_1 from the surroundings. Then we separate the robots with a threshold found empirically to segment the robot point clouds into clusters and eliminate the remaining laser scans (Fig. 5b).

4 DISTRIBUTED FORMATION CONTROL SCHEME

We have presented a relative localization system for the robots to compute the relative position toward the leader agents using onboard sensors. We aim at putting these techniques into practice in two distributed formation control schemes. The first exploits the onboard fusion of odometer data to be able to control the swarm on a Euclidean plane whereas the second generates a control mechanism to satisfy the requirements on the relative quantities by excluding the need for the magnetic sensing aids.

4.1 Generic Control Scheme

In contrast to the holonomic mobility, the agent model (1) inherits nonholonomic constraints on the speeds. Inspired by the control scheme of (Fidan et al., 2012), we design our controller for the holonomic kinematics then convert the speed commands to a form which can be applied in the nonholonomic model. Denote the desired control input required to steer a holonomic agent toward its desired location by $u^d = [v_x^d, v_y^d]^T$. Then, one can adjust the nonholonomic control mechanism as follows:

$$u_i^d(t) = v_i(t) \begin{bmatrix} \cos(\theta_i(t)) \\ \sin(\theta_i(t)) \end{bmatrix}, \quad (2)$$

$$\omega_i(t) = k_\omega(\theta_i(t) - \theta_i^d(t)) + \dot{\theta}_i^d(t), \quad (3)$$

where $k_\omega < 0$ is a proportional gain, $\theta_i^d(t) = \text{atan}(v_y^d(t)/v_x^d(t))$ is the desired heading of robot R_i ,

$\text{atan}(\cdot)$ is the four-quadrant arc-tangent function, and

$$\dot{\theta}_i^d = \begin{cases} \frac{\dot{v}_y^d v_x^d - \dot{v}_x^d v_y^d}{(v_x^d)^2 + (v_y^d)^2} & \text{if } u^d \neq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

By using the relation in (2), that is $\|u_i^d(t)\| = v_i(t)$, we can choose the linear speed $v_i(t)$ as a proportion of the desired speed for the holonomic agent, while the angular speed can be directly used as in (3). The term $\dot{\theta}_i^d$ adds a derivative gain action effect in (3). However, in our real experiments, we did not observe significant effect of the term $\dot{\theta}_i^d$. Thus, we set $\dot{\theta}_i^d(t) = 0$ in our evaluations. We now derive the control inputs for the three types of robots in the swarm.

Leader. Since out-degree of the leader robot R_0 is two, it has two degrees of freedom for mobility and is not confined to any distance constraint toward any other robot in the swarm. The leader is stationary during a formation acquisition task. For flocking tasks, it can be assigned a constant control input.

First Follower. With the out-degree equals one, robot R_1 is to maintain a desired distance toward the leader R_0 . Since we use two sensing modalities, which are UWB distance sensing and vision sensing, robot R_1 is able to sense and control both distance and bearing toward R_0 , which both maintains rigidity and fixes the formation orientation. We propose a two-step control law as follows. Since the monocular camera on robot R_1 is restrained on the field of view (FOV), it may not be able to detect the leader R_0 when it is out of the FOV. In such a scenario, we rotate R_1 with constant angular velocity (and zero linear velocity) until it sees the leader within the range, assuming that the initial distance between robot R_0 and R_1 allows detection. Upon detection from the CNN model, the robot R_1 moves to a point which satisfies the distance constraint $d_{01} = \|p_0 - p_1\|$. Denote the distance error by

$$e_{01} = \|q_{01}\| - d_{01}, \quad (5)$$

and the relative position between robots R_0 and R_1 by

$$z_{01} = [p_{01}^x, p_{01}^y]^T, \quad (6)$$

where p_{01}^x and p_{01}^y are respectively the relative displacement on the x and y axes in the local frame of R_1 . We propose the following control input for R_1 :

$$u_1^d(t) = k_v e_{01}(t) z_{01}(t), \quad (7)$$

$$\theta_1^d(t) = \theta_1(t) - \phi_{01}(t) + \theta_{off}, \quad (8)$$

where $k_v < 0$ is the proportional gain, $\phi_{01} \in [-\pi, \pi]$ is the bearing angle derived from the object detection,

and $\theta_{off} \in [-\pi, \pi)$ is a constant value that represents the offset between the robot body frame and the camera frame. The offset parameter θ_{off} allows the sensor to have a different orientation than the robot body frame.

Second Followers. The second follower robots have two constraints in the constraint graph G . In our three-robot system, the second follower R_2 tries to maintain the distance constraints d_{02} and d_{12} toward robots R_0 and R_1 . We use the following control law for R_i , $i \in 2, \dots, N$:

$$u_i^d(t) = k_v \sum_{j \in \mathcal{N}_i} e_{ij} z_{ij} \quad (9)$$

$$\theta_i^d(t) = \sum_{j \in \mathcal{N}_i} (\theta_i(t) - \phi_{ij}(t)) + \theta_{off} \quad (10)$$

where $k_v < 0$ is the proportional gain, \mathcal{N}_i is the neighbors of R_i in the graph G , $\phi_{ij} \in [-\pi, \pi)$ is the bearing angle between R_i and R_j measured in the body frame of the second follower R_i , $e_{ij} = \|q_{ij}\| - d_{ij}$ is the distance error, and

$$z_{ij} = [p_{ij}^x, p_{ij}^y]^\top \quad (11)$$

is the relative position between R_i and R_j in the body frame of R_i .

4.2 Control Law without Magnetic Sensing

In the previous section, we have derived a distributed control mechanism for the three types of robots in the MRS S . The control laws for robots R_1 and R_2 heavily rely on the odometer data which is produced by fusing the IMU and magnetometer measurements. However, in some workspace environments, magnetometer sensors can be affected by strong electromagnetic variations, thus the readings may be unreliable and have low SNR. In such scenarios, the entire system may yield poor localization and control performance. To avoid this issue, one needs to derive the control laws directly based on the exteroceptive sensors.

Assuming the leader robot R_0 is stationary, consider the polar dynamics between robots R_0 and R_1 as follows:

$$\begin{aligned} \dot{\rho}_1 &= -\bar{v}_1 \cos(\phi_1) \\ \dot{\phi}_1 &= \frac{\bar{v}_1}{\rho_1} \sin(\phi_1) + \omega_1 \end{aligned} \quad (12)$$

where $\rho_1 = \|q_{01}\|$ and $\phi_1 = \text{atan}(x_0 - x_1, y_0 - y_1)$ is the relative bearing angle (Fig. 3). Thus, the dynamics (12) allows to control the distance and bearing angle directly by adjusting the control inputs \bar{v}_1, ω_1 at

the expense of reducing the controllable degree-of-freedom to two. Thus, the formation objective can be achieved, but the robot cannot be steered to any location on the plane. However, since formation acquisition can be satisfied by setting only the distance and bearing to their desired values, the dynamics (12) remains a suitable model for our main objective. We aim at devising a control law so that the distance and bearing toward the leader converge to their desired values.

We propose the following control law for the first follower:

$$\begin{aligned} \bar{v}_1 &= \text{sat}(K_v |e_\rho|, 0, v_{\text{up}}), \\ \omega_1 &= \text{sat}(K_\omega e_\phi, -\omega_{\text{up}}, \omega_{\text{up}}), \end{aligned} \quad (13)$$

where $K_v, K_\omega, v_{\text{up}}, \omega_{\text{up}} > 0$ are design constants, $\text{sat}(a, \bar{a}, \underline{a})$ is a saturation function such that if $a \geq \bar{a}$, then $a = \bar{a}$, and if $a \leq -\bar{a}$, then $a = -\bar{a}$, and

$$e_\rho = \rho_1 - \rho_1^{\text{des}}, \quad e_\phi = \phi_1 - \phi_1^{\text{des}},$$

with $\rho_{\text{des}} > 0$ and ϕ_{des} being the desired distance and bearing values, respectively. If $\phi_1(t) \in [-\pi/2, \pi/2)$, the controller (13) steers R_1 toward the desired location imposed by the constraints $\rho_1^{\text{des}}, \phi_1^{\text{des}}$. We assume that initially robot R_0 resides in the FOV of the camera of robot R_1 . If such an initialization is not possible, one can purely rotate R_1 with a constant angular velocity in one direction until robot R_0 enters the FOV of the camera of robot R_1 then apply (13) for formation acquisition. Thus, we expect that the robot converges to its desired location, leaving the stability and convergence analysis to future works.

5 EVALUATION

In this section, we demonstrate our CNN training experience and the results of our simulations and real world experiments. Finally, we discuss the practical aspects.

5.1 CNN Training

We trained an object detection model for each of the simulation and real experiments. While standard objects can be found in the existing pool of dataset such as COCO (Lin et al., 2014), we did not have a collection of image dataset for our particular object of interest, Turtlebot3. Therefore, we first constructed a dataset for the simulation and the real experiment. We collected 337 images of the Turtlebot3 from the simulation and 501 images from the real robot. Then, we conducted the manual labeling process, i.e., assigning individual bounding boxes to the Turtlebot3

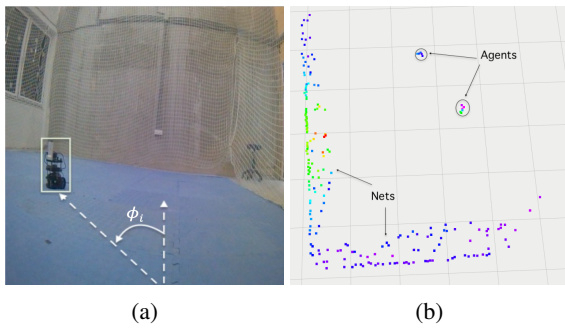


Figure 6: Sensors view in experiment: a) A detection from Jetson nano: A view from R_1 , b) A scan from the LDS LIDAR seen by robot R_2 .

sections of the images. During the training phase, we randomly split the dataset into the training, validation, and testing sets. Then, we trained the network by selecting the best observed hyperparameters, which are a batch size of 64, a momentum of 0.9, and a learning rate of 0.001. We obtained the final results on the testing set of mean average precision (mAP@75) of 96% on the experiments, which enabled the high precision during experimental tests. Sample detection images from the CNN can be seen in Fig. 5a (simulation) and 6a (experiment).

5.2 Simulation

We simulated our proposed system in Gazebo, a realistic and powerful environment that allows to visualize robotic systems with similar physical properties. We used three Turtlebot3 burger robots and onboard sensors with the following specifications: A monocular RGB camera with FOV of 2.45 radians on R_1 and a 2D LIDAR with 360 FOV and a range of 6.28 meters on R_2 . We processed the LIDAR data using the point cloud package from robot operating system (ROS) and used OpenCV² for post-processing the YOLO model inference. We mimic the UWB sensor measurements with the distance values obtained from

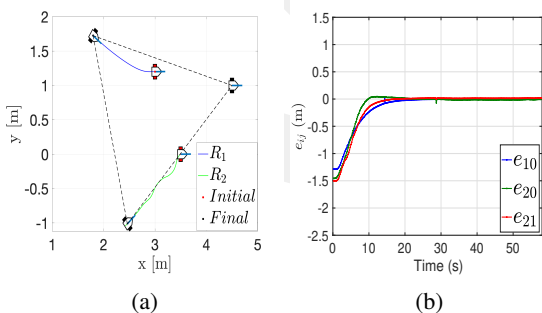


Figure 7: Formation acquisition by expansion simulation: a) Trajectories of the agents, b) the inter-robot distance errors.

²<https://opencv.org/opencv-4-4-0/>

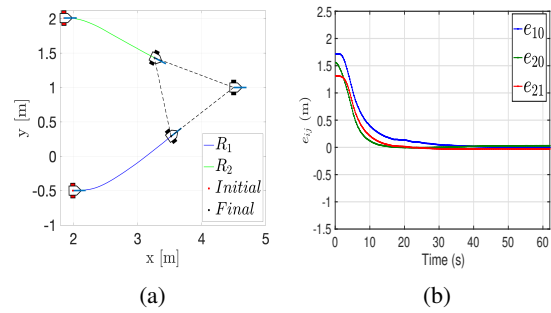


Figure 8: Formation acquisition by contraction simulation: a) Trajectories of the agents, b) the inter-robot distance errors.

Gazebo with added zero-mean Gaussian noise, i.e., $q_{ij}(t) = \bar{q}_{ij}(t) + \eta_{ij}(t)$ where $\eta_{ij} \sim N(0, \sigma^2)$ with σ the standard deviation.

In the simulations, we evaluated the setup for two formation objectives: Contraction and expansion. We used the generic formation control scheme using odometry data. These two key actions aimed at demonstrating the flexibility of the agents to adapt to different scenarios independent of the initial conditions. The first assessment requires the robots to build a formation with smaller relative distances than the initial values whereas the second evaluation starts the formation from a configuration where the distances are smaller than their desired values. The trajectories and the relative distance errors are depicted in Fig. 8 and Fig. 7. We observed an exponential convergence to the desired configuration and fast decrement of the distance errors e_{ij} . Also, we obtained a CNN detection rate greater than 90% on R_1 in all simulations, which proved efficiency of the proposed localization method.

5.3 Experiments

We performed a series of tests in an indoor lab environment where GPS is not available. The three-robot

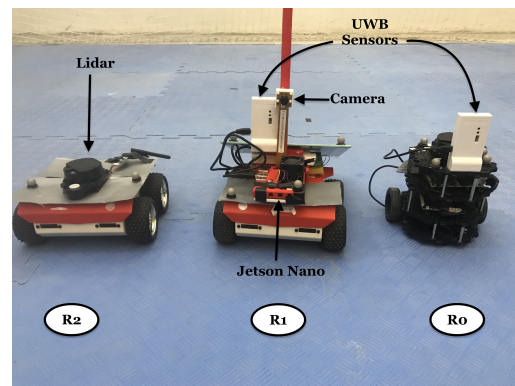


Figure 9: The robots and hardware used in the experiment.

system comprised a Turtlebot3 as the leader (R_0) and two Rosbots as the first-follower (R_1) and the second-follower (R_2), respectively (Fig. 9). We equipped the robots with the sensor network explained in Section 3. Particularly, we mounted a pair of DWM1001 UWB modules on the leader and the first follower to produce the relative distance between them. Also, we mounted a CSI camera through an Intel Jetson Nano computer on the first follower for bearing detection. We calibrated the camera to remove the distortions and reduced the image size from (3280×2464) to (416×416) which is the input to the CNN model. This resizing operation eliminates the need for extra pre-processing operations on the image before the detection, and eventually, reduces the overall detection time. With its GPU cores, Jetson Nano performed the model inference using the accelerated TensorRT and produced a detection speed of up to 60 ms which was reasonable for our platform. Finally, we equipped the second Rosbot with a 2D LDS LIDAR with the FOV of 360 degrees and the maximum range of 3.5 meters. We recorded the ground truth pose data of the robots with a VICON motion capture system for illustration purposes. The motion capture system data was not used in any part of our algorithms.

We evaluated the proposed framework on the three-robot system defined above relying on their onboard resources. The results of a contraction experiment is illustrated in Fig. 10. In this experiment, the robots were initiated from $p_0(0) = [0.02, 0.62, 0.0]^T$ meters, $p_1(0) = [1.14, -1.33, 0.0]^T$ meters, and $p_2(0) = [-1.13, -1.41, 0.0]^T$ meters with different heading angles which are also denoted in Fig. 10. Initially, the leader robot R_0 was not in the FOV of the robot R_1 . Thus, R_1 first rotated with zero linear speed and waited for the leader to enter its camera FOV. Once R_1 is detected with the CNN at the 36th second, it started approaching R_0 based on the control rule (7). On the other hand, robot R_2 segmented the LIDAR scans to generate clusters for robots R_0 and R_1 . After building the clusters, robot R_2 generated the distance estimates q_{02}, q_{12} . We assumed that initially the robots R_0 and R_1 were present within the laser range. We depict a sample point cloud view of the scans in Fig. 6b. There were nets on the left and rear sides of R_2 which can be clearly seen in Fig. 6b, which we removed with a simple pre-processing in the algorithm. We observed that the robots converged to the desired configuration which are $d_{ij} = 0.5$ meters for all i, j (Fig. 10b), and the error remained bounded for sufficiently long time (around 400 seconds). We conjecture that the relatively small error in the distances were due to the sensor noises. For instance, the UWB distance readings from this experiment with

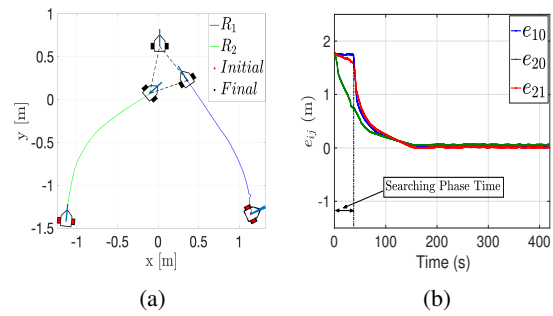


Figure 10: Contraction experiment: a) Trajectories of the agents, b) the inter-robot distance errors.

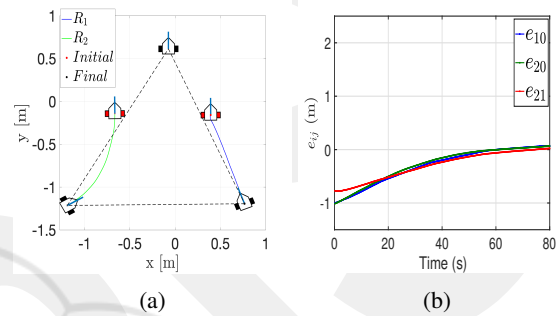


Figure 11: Expansion Experiment . a) Trajectories of the agents, b) The relative distance errors.

its ground truth are given in Fig. 12b, which shows a noise with a variance of around 0.05 meters. Nevertheless, the framework exhibited high performance in terms of formation acquisition. A similar performance can be seen in the formation expansion experiment in Fig. 11.

Furthermore, we evaluated the case where the magnetometer is not available for the first follower (Fig. 12a). In this experiment, robot R_1 used the controller (13) and was able to converge toward the desired configuration. This result reflects the flexibility of the proposed framework for magnetometer-less settings. Although not presented here, another controller which does not depend on magnetometer sensor can be designed for the second follower robot as well, which is one of our future objectives.

We now discuss some practical details which were important when transitioning from simulations to experiments. In our framework, the second follower robot R_2 relies on LIDAR solely. However, when LIDAR detects another object in the workplace, the two clusters for localizing the robots R_0 and R_1 may not be generated precisely, which may deteriorate the performance. Therefore, we emphasize the assumption that the workplace needs to be free of any other object. If such a condition cannot be met, further post-processing of LIDAR measurements would be needed. Furthermore, although we have provided a general framework for $N + 1$ -robot swarms, we have

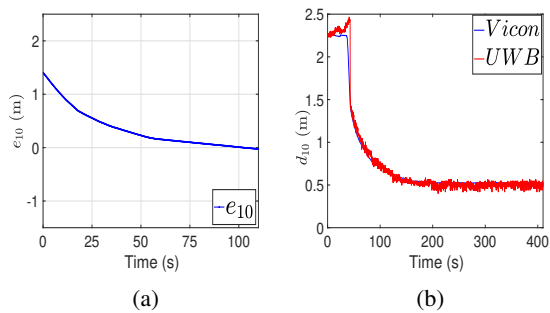


Figure 12: a) The error on R_1 during a formation contraction without magnetometers, b) A Comparison of UWB sensor data with ground truth from Vicom capture system.

focused on a three-robot setup in evaluations. When there are more than one second-follower robots, i.e., when robots $R_i, i \in \{3, \dots, N\}$ are added to the system, a more advanced LIDAR post-processing would be necessary to discriminate the two leader robots for each second-follower.

Moreover, although UWB sensors provide accurate distance data in most environments, they may generate low signal-to-noise ratio if not calibrated properly. We suggest finding the correct bias and noise variance values of the UWB sensors with a proper calibration. In our experiments, we used a basic calibration which requires to place the UWB pair at locations with certain distances and collecting data for a certain time. However, UWB distance measurements may vary significantly with varying angular placements of the sensors and ambient factors. We emphasize the importance of UWB calibration for more precise evaluations.

6 CONCLUSIONS

We have investigated the practical formation control problem for MRS using onboard sensors. We have proposed a systematic framework including a localization module and a control module. The localization module consists of a bearing angle derivation based on a YOLOv3-Tiny architecture with a monocular camera and a systematic clustering for a 2D LIDAR. The control module consists of a set of distributed formation control laws for three types of robots in the leader-follower hierarchy. Furthermore, we have proposed a control method which does not use of a magnetometer sensor which is known to generate unreliable sensor readings in some environments. We have evaluated the framework on non-holonomic ground robots for the formation contraction and expansion objectives in both simulations and experiments. The high performance obtained in a series of experiments indicates the applicability of the

proposed solution. Future research may include extension of the framework for dynamic path planning and obstacle avoidance for the whole swarm using on-board resources. Also, LIDAR data post-processing to discriminate the leader robots in the FOV needs to be studied to scale up the swarm size.

ACKNOWLEDGEMENTS

This paper has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TÜBİTAK (Project No: 118C348). However, the entire responsibility of the paper belongs to the owner of the paper. The financial support received from TÜBİTAK does not mean that the content of the publication is approved in a scientific sense by TÜBİTAK.

REFERENCES

- Abdelkader, M., Güler, S., Jaleel, H., and Shamma, J. S. (2021). Aerial swarms: Recent applications and challenges. *Current Robotics Reports*, 2(3):309–320.
- Almadhoun, R., Taha, T., Seneviratne, L., Dias, J., and Cai, G. (2016). A survey on inspecting structures using robotic systems. *International Journal of Advanced Robotic Systems*, 13(6):1729881416663664.
- Anderson, B. D., Yu, C., Fidan, B., and Hendrickx, J. M. (2008). Rigid graph control architectures for autonomous formations. *IEEE Control Systems Magazine*, 28(6):48–63.
- Choi, Y. H. and Kim, D. (2021). Distance-based formation control with goal assignment for global asymptotic stability of multi-robot systems. *IEEE Robotics and Automation Letters*, 6(2):2020–2027.
- Chung, S.-J., Paranjape, A. A., Dames, P., Shen, S., and Kumar, V. (2018). A survey on aerial swarm robotics. *IEEE Transactions on Robotics*, 34(4):837–855.
- De Queiroz, M., Cai, X., and Feemster, M. (2019). *Formation control of multi-agent systems: A graph rigidity approach*. John Wiley & Sons.
- Fidan, B., Gazi, V., Zhai, S., Cen, N., and Karataş, E. (2012). Single-view distance-estimation-based formation control of robotic swarms. *IEEE Transactions on Industrial Electronics*, 60(12):5781–5791.
- Guler, S., Abdelkader, M., and Shamma, J. S. (2020). Peer-to-peer relative localization of aerial robots with ultrawideband sensors. *IEEE Transactions on Control Systems Technology*, pages 1–16.
- Kabore, K. M. and Güler, S. (2021a). Deep learning based formation control of drones. In *Deep Learning for Unmanned Systems*, pages 383–413. Springer.
- Kabore, K. M. and Güler, S. (2021b). Distributed formation control of drones with onboard perception. *IEEE/ASME Transactions on Mechatronics*.

- Lin, J., Miao, Z., Zhong, H., Peng, W., Wang, Y., and Fierro, R. (2020). Adaptive image-based leader–follower formation control of mobile robots with visibility constraints. *IEEE Transactions on Industrial Electronics*, 68(7):6010–6019.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.
- Liu, X., Ge, S. S., and Goh, C.-H. (2018). Vision-based leader–follower formation control of multiagents with visibility constraints. *IEEE Transactions on Control Systems Technology*, 27(3):1326–1333.
- Nguyen, T.-M., Qiu, Z., Cao, M., Nguyen, T. H., and Xie, L. (2020). Single landmark distance-based navigation. *IEEE Transactions on Control Systems Technology*, 28(5):2021–2028.
- Oh, K.-K., Park, M.-C., and Ahn, H.-S. (2015). A survey of multi-agent formation control. *Automatica*, 53:424–440.
- O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., and Walsh, J. (2019). Deep learning vs. traditional computer vision. In *Advances in Intelligent Systems and Computing*, pages 128–144. Springer International Publishing.
- Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.
- Saska, M., Baca, T., Thomas, J., Chudoba, J., Preucil, L., Krajnik, T., Faigl, J., Loianno, G., and Kumar, V. (2017). System for deployment of groups of unmanned micro aerial vehicles in gps-denied environments using onboard visual relative localization. *Autonomous Robots*, 41(4):919–944.
- Silano, G., Baca, T., Penicka, R., Liuzza, D., and Saska, M. (2021). Power line inspection tasks with multi-aerial robot systems via signal temporal logic specifications. *IEEE Robotics and Automation Letters*, 6(2):4169–4176.
- Tian, Y., Liu, K., Ok, K., Tran, L., Allen, D., Roy, N., and How, J. P. (2020). Search and rescue under the forest canopy using multiple uavs. *The International Journal of Robotics Research*, 39(10-11):1201–1221.
- Trinh, M. H., Zhao, S., Sun, Z., Zelazo, D., Anderson, B. D., and Ahn, H.-S. (2018). Bearing-based formation control of a group of agents with leader-first follower structure. *IEEE Transactions on Automatic Control*, 64(2):598–613.
- Vrba, M. and Saska, M. (2020). Marker-less micro aerial vehicle detection and localization using convolutional neural networks. *IEEE Robotics and Automation Letters*, 5(2):2459–2466.
- Walter, V., Saska, M., and Franchi, A. (2018). Fast mutual relative localization of UAVs using ultraviolet LED markers. In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE.
- Zhao, S. and Zelazo, D. (2019). Bearing rigidity theory and its applications for control and estimation of network systems: Life beyond distance rigidity. *IEEE Control Systems Magazine*, 39(2):66–83.