

# Event-Triggered Optimal Formation Tracking Control Using Reinforcement Learning for Large-Scale UAV Systems

Ziwei Yan<sup>1</sup>, Liang Han<sup>1\*</sup>, Xiaoduo Li<sup>1</sup>, Jinjie Li<sup>2</sup> and Zhang Ren<sup>2</sup>

**Abstract**—Large-scale UAV switching formation tracking control has been widely applied in many fields such as search and rescue, cooperative transportation, and UAV light shows. In order to optimize the control performance and reduce the computational burden of the system, this study proposes an event-triggered optimal formation tracking controller for discrete-time large-scale UAV systems (UASs). And an optimal decision - optimal control framework is completed by introducing the Hungarian algorithm and actor-critic neural networks (NNs) implementation. Finally, a large-scale mixed reality experimental platform is built to verify the effectiveness of the proposed algorithm, which includes large-scale virtual UAV nodes and limited physical UAV nodes. This compensates for the limitations of the experimental field and equipment in real-world scenario, ensures the experimental safety, significantly reduces the experimental cost, and is suitable for realizing large-scale UAV formation light shows.

## I. INTRODUCTION

In recent years, enormous studies on cooperative control for multi-UAV system are overwhelming. Among them, multi-UAV formation tracking control is a common and fundamental problem, and the formation scale has changed from a few UAV groups previously to large-scale formations nowadays. With the exponential growth of formation scale, the system performance is more rigorously required. Hence, researchers proposed optimal control that optimize system performance and formation cost to reduce the system burden.

Optimal control is an effective strategy to balance the system performance and computational consumption. For discrete-time systems, the optimal control can be achieved by solving the Bellman equation [1]–[4]. However, since the analytic solution of the Bellman nonlinear equation is difficult to be derived directly, all studies above consider the linear systems. In order to approximate optimal controllers for nonlinear systems, reinforcement learning (RL) has been introduced in [5]–[10], where the actor-critic neural network (NN) is the commonly used RL architecture. The critic NN gives feedback to optimize the actions by evaluating the system performance index, while the actor NN issues optimized control commands to improve the system behaviors.

This work was supported in part by the Science and Technology Innovation 2030 Key Project of "New Generation Artificial Intelligence" under Grant 2018AAA0102305, and in part by the National Natural Science Foundation of China under Grant 61803014 and Grant 61873011.

\*Corresponding author: L. Han, liang\_han@buaa.edu.cn.

<sup>1</sup>Z. Yan, L. Han and X. Li are with the Sino-French Engineer School, Beihang University, Beijing 100191, China. yanziwei@buaa.edu.cn; liang\_han@buaa.edu.cn; lixiaoduo@sjtu.edu.cn.

<sup>2</sup>J. Li and Z. Ren are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, China. lijinjie@buaa.edu.cn; renzhang@buaa.edu.cn.

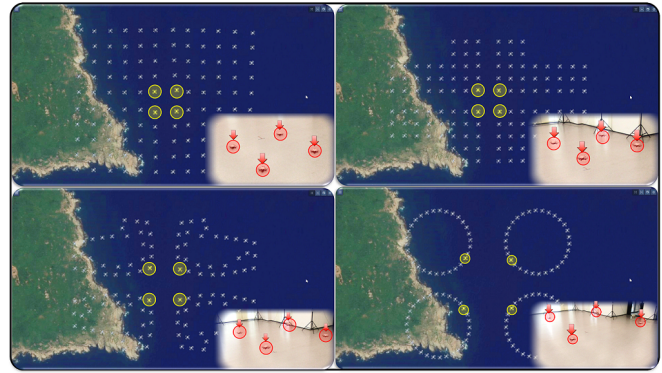


Fig. 1. Hundreds of virtual UAVs and four physical UAVs in a large-scale mixed reality experimental platform to realize a UAV formation light show.

Last few years, RL architectures have been widely used in optimal cooperative control. In [11], a learning-based adaptive dynamic programming algorithm is used to solve the formation tracking problem for multi-UAV systems with model uncertainty in obstacle environments. In [12], for the multi-agent tracking control with actuator faults, an adaptive optimal fault-tolerant tracking controller is designed by combining the RL algorithm and the backstepping method. However, the studies above consider the time-triggered mechanism, which will continuously or periodically update the controller and NNs. This will inevitably cause a huge computational consumption when the formation size is large.

This study introduces the event-triggered mechanism where the UAV updates the controller and actor NN only when the specific condition is triggered, which allows the UAV system (UAS) to update control commands on demand, aperiodically rather than continuously and indiscriminately. At the same time, it means that the event-triggered mechanism can effectively reduce the system computational burden compared to time-triggered cases. For the RL-based optimal cooperative problems, a number of studies have introduced the event-triggered mechanism as an improvement, whether for continuous-time systems [13]–[18] and discrete-time systems [19]–[24]. However, these studies only implement numerical simulations without verifying the algorithm effectiveness for engineering applications.

In order to verify the efficiency of the proposed algorithm, a large-scale mixed reality experimental platform is established in this study. A large-scale UAV light show is realized using a limited number of physical UAVs and hundreds of virtual UAVs (Fig. 1). In view of the main limitations of large-scale formation show, such as high requirement for experimental field, expensive experimental equipment, and

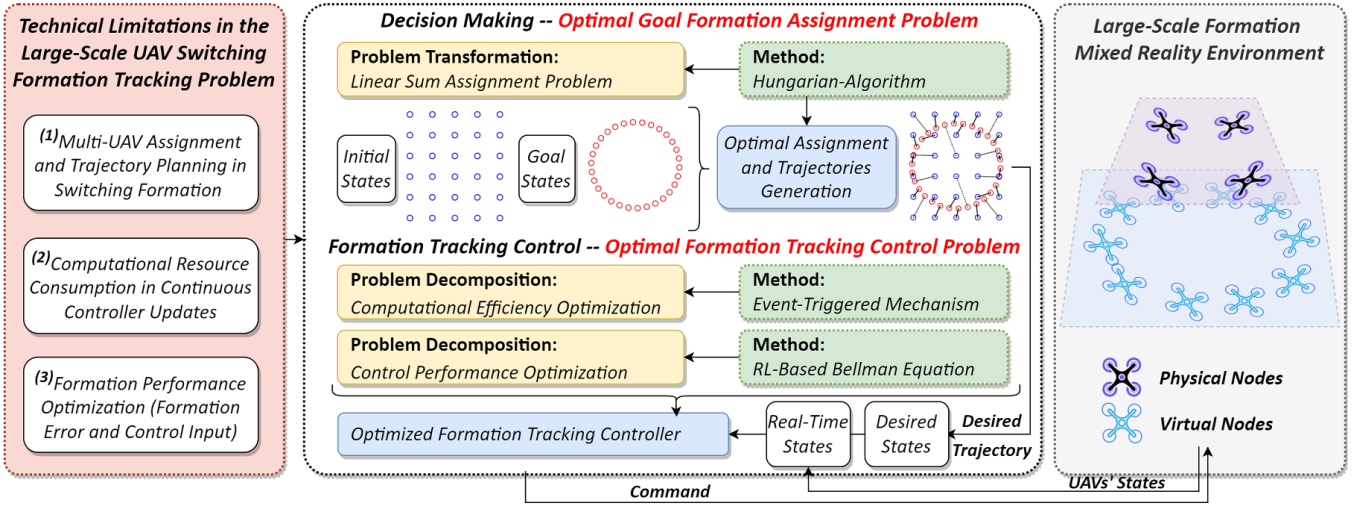


Fig. 2. The algorithm overview of this study. To solve the large-scale UAV switching formation tracking problem, for the decision layer, the Hungarian algorithm is used to solve the goal formation assignment problem. For the control layer, the event-triggered mechanism and the RL-based Bellman equation are used to optimize the computational efficiency and the control performance of the system, respectively.

poor safety factor, the established experimental platform has the advantages of low experimental cost and high safety factor. And the experimental results show that this platform only requires one room and four UAVs to achieve the desired effect of hundreds of UAVs in the large airspace.

This paper proposes an event-triggered optimal formation tracking controller for discrete-time large-scale UASs, and an online actor-critic NN structure is established to implement the control. The main contributions of this study include:

- 1) Considering the main difficulties involved in the decision and control layers for large-scale switching formations, an algorithm framework integrating optimal formation assignment and optimal control is established.
- 2) The event-triggered online actor-critic NN is proposed to approximate the local performance index and learn controller. Compared with [5]–[12], actor NN weights and controller only update at triggering instants.
- 3) This study highlights the engineering application of the proposed algorithm, establishes a mixed reality experimental platform for large-scale UAV formation, and verifies the proposed algorithm on this platform.

## II. ALGORITHM OVERVIEW

As shown in Fig. 2, an algorithm framework is designed to solve the large-scale UAV optimal formation tracking problem. And the algorithms for optimal assignment and optimal control are summarized in Algorithm 1 and Algorithm 2.

### Algorithm 1 Decision Layer - Optimal Assignment

**Input:** 1) Initial position set  $\vec{\mathbf{G}}$ ; 2) Expected formation is  $\vec{\mathbf{S}}$ ; 3) UAV number  $N$ .

**Output:** 1) Target position set  $\vec{\mathbf{F}}$ ; 2) Optimal assignment list  $\chi^*$ ; 3) Desired switching trajectory  $h(k)$ .

1:  $(\vec{\mathbf{F}}, \chi^*, h(k)) = \text{Optimal\_Assignment}(\vec{\mathbf{G}}, \vec{\mathbf{S}}, N)$ .

### Algorithm 2 Control Layer - Optimal Control

**Input:** 1) Desired trajectory  $h(k)$  2) Real-time states  $x(k)$ .

**Output:** 1) Optimized controller  $u(k)$ ; 2) Triggering instants sequence  $\{k_i^i\}_{i \in \mathbb{Z}^+}$ .

- 1: Initialize the actor-critic NN;
- 2: Calculate triggering condition  $f_i(k)$  with  $h_i(k)$  and  $x_i(k)$ ;
- 3: **if**  $f_i(k) > 0$  **then**
- 4:     Using actor NN to learn the controller  $u_i(k)$ ;
- 5:     Calculate and update the actor NN;
- 6:     The corresponding instant  $k$  is recorded in  $\{k_i^i\}_{i \in \mathbb{Z}^+}$ .
- 7: **else**
- 8:      $u_i(k+1) = u_i(k)$ ;
- 9:     The actor NN remains the same as before.
- 10: **end if**
- 11: Calculate and update the critic NN;
- 12: Approximate the system performance index.

## III. PRELIMINARIES

### A. Graph Theory

For a large-scale UAS with a virtual leader and  $N$  followers, the communication network can be described as a graph  $\mathcal{G} = \{\mathcal{A}, \mathcal{E}, \mathcal{W}\}$ , where  $\mathcal{A} = \{a_1, \dots, a_N\}$  denotes the node set,  $\mathcal{E}$  is the edge set and  $e_{ij} = (a_i, a_j) \in \mathcal{E}$ , and  $\mathcal{W}$  is the adjacency matrix with elements  $w_{ij}$ , for  $e_{ij} \in \mathcal{E}$ ,  $i \neq j$ ,  $w_{ij} = 1$ , otherwise  $w_{ij} = 0$ . Define  $N_i$  as the neighbor set of UAV  $a_i$ . Then, the in-degree matrix  $\mathcal{D} = \text{diag}_{i \in [1, N]} \{\sum_{j \in N_i} w_{ij}\}$ . Laplacian matrix is  $L = \mathcal{D} - \mathcal{W}$ . Further, a diagonal matrix  $B = \text{diag}_{i \in [1, N]} \{b_i\}$  describes the communication between followers and the virtual leader. Assume that the leader can communicate with at least one follower, and there exists a spanning tree with the leader as the root in the directed group  $\hat{\mathcal{G}} = \{\hat{\mathcal{A}}, \hat{\mathcal{E}}\}$  with  $\hat{\mathcal{A}} = \{a_0, \dots, a_N\}$  and  $\hat{\mathcal{E}} \in \hat{\mathcal{A}} \times \hat{\mathcal{A}}$ .

**Assumption 1.** The graph  $\mathcal{G}$  is connected and the directed graph  $\hat{\mathcal{G}}$  contains a spanning tree.

## B. Radial Basis Function (RBF) NNs

An unknown nonlinear function  $\zeta(\Omega)$  can be approximated by a RBF NN,  $\zeta(\Omega) = W^{*T}\psi(\Omega)$ , where  $W^* \in \mathbb{R}^s$  is a weight vector,  $\psi(\Omega) \in \mathbb{R}^s$  is a basis function vector with  $\psi(\Omega) = [\psi_1(\Omega), \dots, \psi_s(\Omega)]^T$ , and  $s \in \mathbb{Z}^+$  represents the number of NN nodes in the hidden layer. In this paper,  $\psi_i(\Omega)$  is chosen as the Gaussian Basis Function

$$\psi_i(\Omega) = \exp(-(\Omega - v_i)^T(\Omega - v_i)/2t_i^2), \quad i = 1, \dots, s. \quad (1)$$

where  $v_i$  is the center of the Gaussian kernel function and  $t_i$  is the width parameter of the function. For the approximation of the RBF NN, it is always possible to find an ideal weight  $W^* = \arg_{\min W} \{\sup_{\Omega} \|\zeta(\Omega) - W^T\psi(\Omega)\|\}$ . As a result, the unknown function  $\zeta(\Omega)$  can be written as

$$\zeta(\Omega) = W^{*T}\psi(\Omega) + \sigma(\Omega). \quad (2)$$

where  $\sigma(\Omega)$  is the approximation error with  $|\sigma(\Omega)| \leq \sigma_T$ , and  $\sigma_T$  is a constant.

## C. Problem Description

Consider a large-scale UAS with  $N$  followers and a virtual leader, each UAV  $a_i$  ( $i \in \{0, \dots, N\}$ ) is modeled as:

$$\begin{cases} p_i(k+1) = p_i(k) + v_i(k)T \\ v_i(k+1) = v_i(k) + u_i(k)T + f_i(p_i(k), v_i(k)). \end{cases} \quad (3)$$

where  $p_i(k) \in \mathbb{R}^m$  and  $v_i(k) \in \mathbb{R}^m$  are the position and velocity of the UAV  $a_i$ ,  $u_i(k) \in \mathbb{R}^m$  is the control input with  $m$  the movement dimension of UAV, and  $f_i(\cdot)$  is the unknown bounded nonlinear function. For the virtual leader  $a_0$ ,  $u_0(k)$  is known and  $f_0(\cdot) = 0$ .

Then, for each follower UAV  $a_i$ , the position tracking error  $\xi_{pi} \in \mathbb{R}^m$  and velocity tracking error  $\xi_{vi} \in \mathbb{R}^m$  are defined as

$$\begin{cases} \xi_{pi}(k) = p_i(k) - p_0(k) - \eta_{pi}(k) \\ \xi_{vi}(k) = v_i(k) - v_0(k) - \eta_{vi}(k). \end{cases} \quad (4)$$

where  $\eta_i(k) = [\eta_{pi}^T(k), \eta_{vi}^T(k)]^T \in \mathbb{R}^{2m}$  is the desired state difference between the leader  $a_0$  and the follower  $a_i$ . Then, define the disagreement error for UAV  $a_i$  as

$$\begin{cases} \varepsilon_{pi}(k) = \sum_{j \in N_i} w_{ij}(\xi_{pi}(k) - \xi_{pj}(k)) + b_i \xi_{pi}(k) \\ \varepsilon_{vi}(k) = \sum_{j \in N_i} w_{ij}(\xi_{vi}(k) - \xi_{vj}(k)) + b_i \xi_{vi}(k). \end{cases} \quad (5)$$

Let  $\varepsilon(k) = [\varepsilon_1^T(k), \dots, \varepsilon_N^T(k)]^T$  with  $\varepsilon_i(k) = [\varepsilon_{pi}^T(k), \varepsilon_{vi}^T(k)]^T$ . Define the local performance index function (PIF) as

$$V_i(\varepsilon_i(k), u_i(k)) = \sum_{l=k}^{\infty} \beta^{l-k} R_i(\varepsilon_i(l), u_i(l)), \quad (6)$$

where  $R_i(\varepsilon_i(k), u_i(k)) = \varepsilon_i^T(k)\varepsilon_i(k) + u_i^T(k)u_i(k)$  is the utility function.

## IV. OPTIMAL ASSIGNMENT AND OPTIMAL CONTROL

In this section, the algorithm overview shown in Fig. 2 is expanded and detailed separately in three parts: optimal formation assignment, event-triggered optimal controller design, and online actor-critic NNs implementation.

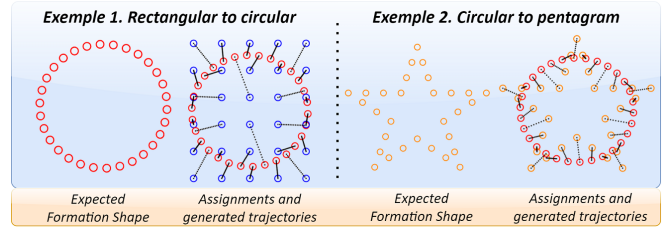


Fig. 3. Two examples of switching formations, with the expected formation shape on the left and the assignments and generated trajectories on the right.

## A. Optimal Assignment for Switching Formations

For a system with  $N$  UAVs of radius  $r$ , suppose its initial position set is  $\vec{G} = \{\vec{g}_i^T\} \in \mathbb{R}^{N \times m}$ , the expected formation is  $\vec{S} = \{\vec{s}_j^T\} \in \mathbb{R}^{N \times m}$ , and the target position set is  $\vec{F} = \{\vec{f}_j^T\} \in \mathbb{R}^{N \times m}$  with  $\vec{f}_j = \rho \vec{s}_j + \vec{d}$ , where  $\rho \in \mathbb{R}^+$  is the scale factor and  $\vec{d}$  is the translation vector.

**Lemma 1.** [25] *The optimal scale factor  $\rho^*$  and translation vector  $\vec{d}^*$  can be calculated as*

$$\begin{aligned} \rho^* &= \frac{\sum_{i=1}^N \vec{g}_i (\sum_{j=1}^N \vec{s}_j)^T + N \kappa^*}{\sum_{j=1}^N \vec{s}_j (\sum_{j=1}^N \vec{s}_j)^T - N \sum_{j=1}^N \vec{s}_j \vec{s}_j^T}, \\ \vec{d}^* &= \frac{\sum_{i=1}^N \vec{g}_i - \rho^* \sum_{j=1}^N \vec{s}_j}{N}. \end{aligned} \quad (7)$$

## Algorithm 3 [25] Pseudo-Cost Assignment Algorithm

**Input:** 1) Initial position set  $\vec{G}$ ; 2) Expected formation is  $\vec{S}$ .

**Output:** 1) Optimal assignment list  $\chi^* \in \mathbb{Z}^N$ ; 2) Minimum cost matrix  $\kappa^* \in \mathbb{R}^{N \times N}$ .

- 1:  $\kappa_{ij} = -\vec{g}_i \vec{s}_j^T$ .
- 2:  $(\chi^*, \kappa^*) = \text{Hungarian-LSAP}(\kappa)$ .

Algorithm 3 is introduced to enable UAVs to reach the target positions simultaneously without collision. Define the formation switching duration as  $t_s = \max_{i \in [1, N]} \|\vec{g}_i - \vec{f}_{\chi(i)}\|_2 / v_{\max}$ , where  $\chi(i)$  is the position index assigned to UAV  $a_i$ , and  $v_{\max}$  is the maximum allowable velocity. Then, the desired switching trajectory for UAV  $a_i$  is

$$h_i(k) = \vec{g}_i + \frac{\vec{f}_{\chi(i)} - \vec{g}_i}{t_s} kT, \quad k \in [0, \frac{t_s}{T}]. \quad (8)$$

And the collision avoidance conditions of UAVs are noted  $\|\vec{g}_i - \vec{g}_j\|_2 \geq 2\sqrt{2}r$  and  $\|\vec{s}_i - \vec{s}_j\|_2 \geq 2\sqrt{2}r$ .

## B. Event-Triggered Optimal Controller Design

Define the triggering instants sequence for the UAV  $a_i$  as  $\{k_l^i\}_{l \in \mathbb{Z}^+}$ , which is determined by the following condition,

$$f_i(k) = \|e_i^j(k)\|^2 - \frac{(1 - 2\kappa^2)}{2\kappa^2} \|\varepsilon_i(k)\|^2, \quad (9)$$

where  $\kappa \in (\frac{1}{2}, \frac{\sqrt{2}}{2})$ ,  $\varepsilon_i(k)$  is the current disagreement error and the controller  $u_i(k)$  only updates at the triggering instant  $k_l^i = \inf\{k > k_{l-1}^i, f_i(k) > 0\}$ , that is,  $\hat{u}_i(k) = u_i(\varepsilon_i(k_l^i))$ , where  $\varepsilon_i(k_l^i)$  is the triggering error. The zero-order holder (ZOH) is used until next triggering instant occurs, one has

$$u_i(k) = \hat{u}_i(k) = u_i(k_l^i), \quad k \in [k_l^i, k_{l+1}^i). \quad (10)$$

**Lemma 2.** [20] There exists a positive constant  $\kappa$  satisfying the inequality as follows,

$$\|\varepsilon_i(k+1)\| = \|g_i(\varepsilon_i(k), u_i(k_i^j))\| \leq \kappa\|\varepsilon_i(k)\| + \kappa\|e_i^j(k)\| \quad (11)$$

where  $e_i^j(k) = \varepsilon_i(k_i^j) - \varepsilon_i(k)$ ,  $k \in [k_i^j, k_{i+1}^j)$ . And when the event is triggered,  $e_i^j(k) = 0$ .

**Theorem 1.** Considering the large-scale UAS (3) following the Assumption 2 and the triggering condition (9), the disagreement error is asymptotically stable and formation tracking control for the large-scale UAS (3) is achieved.

*Proof.* (a) For  $\forall k \in [k_i^j, k_{i+1}^j)$ , the controller  $u_i(k)$  remains a constant,  $u_i(k) = u_i(k+1)$ . Design the Lyapunov function

$$\mathcal{V}_{ei}(k) = \varepsilon_i^T(k)\varepsilon_i(k) + u_i^T(k)u_i(k). \quad (12)$$

With the Lemma 2 and the triggering condition (9),  $\Delta\mathcal{V}_{ei}(k) = \mathcal{V}_{ei}(k+1) - \mathcal{V}_{ei}(k)$  can be derived as

$$\Delta\mathcal{V}_{ei}(k) \leq 2\kappa^2\|\varepsilon_i(k)\|^2 + 2\kappa^2\|e_i^j(k)\|^2 - \|\varepsilon_i(k)\|^2 \leq 0. \quad (13)$$

(b) For  $k \in \{k_i^j\}_{l \in \mathbb{Z}^+}$ , the controller  $u_i(k)$  will be updated, and the Lyapunov function is designed as

$$\mathcal{V}_{ei}(k) = \beta^k V_i(\varepsilon_i(k)). \quad (14)$$

Then, the first difference of the  $\mathcal{V}_{ei}$  is

$$\Delta\mathcal{V}_{ei} = -\beta^k R_i(\varepsilon_i(k), u_i(k_i^j)) \leq 0. \quad (15)$$

Overall,  $\varepsilon_i(k)$  converges asymptotically, which means that event-triggered formation tracking control is achieved.  $\square$

### C. Implementation with Online Actor-Critic NNs

The optimal controller  $u_i^*(k)$  is the unique solution of (6), and  $u_i^*(k)$  can be derived by the equation  $\frac{\partial V_i(\varepsilon_i(k))}{\partial u_i(k)} = 0$ ,

$$u_i^*(k) = -\frac{\beta\gamma_i T}{2} \frac{\partial V_i^*(\varepsilon_{vi}(k+1))}{\partial \varepsilon_{vi}(k+1)}. \quad (16)$$

where  $\gamma_i = \sum_{j \in N_i} w_{ij} + b_i$ . Let  $\frac{\partial V_i^*(\varepsilon_{vi}(k+1))}{\partial \varepsilon_{vi}(k+1)} = \frac{2\alpha_p \varepsilon_{pi}(k)}{\beta\gamma_i T} + \frac{2\alpha_v \varepsilon_{vi}(k)}{\beta\gamma_i T} + \frac{V_i^0(\varepsilon_i)}{\beta\gamma_i T}$ , where constants  $\alpha_p > 0$ ,  $\alpha_v > 0$ , and  $V_i^0(\varepsilon_i) = -2\alpha_p \varepsilon_{pi}(k) - 2\alpha_v \varepsilon_{vi}(k) + \beta\gamma_i T \frac{\partial V_i^*(\varepsilon_{vi}(k+1))}{\partial \varepsilon_{vi}(k+1)}$ . Then one has

$$u_i^*(k) = -\alpha \varepsilon_i(k) - \frac{V_i^0(\varepsilon_i)}{2}, \quad (17)$$

where  $\alpha = [\alpha_p, \alpha_v] \otimes I_m$ . And in order to implement the approximation of the optimal controller and the optimized PIF, actor-critic NNs are introduced.

1) *Critic NN Design:* The critic NN, which is utilized to approximate the the optimized PIF,

$$\frac{\partial \hat{V}_i(\varepsilon_i(k+1))}{\partial \varepsilon_i(k+1)} = \frac{2\alpha \varepsilon_i(k)}{\beta\gamma_i T} + \frac{1}{\beta\gamma_i T} \hat{W}_{ci}^T(k) \psi_i(\varepsilon_i), \quad (18)$$

where  $\hat{W}_{ci} \in \mathbb{R}^{s \times m}$  is the estimation of  $W_{ci}^*$ , which is trained by the critic NN update law

$$\hat{W}_{ci}(k+1) = \hat{W}_{ci}(k) - \mu_{ci} T \psi_i(\varepsilon_i) \psi_i^T(\varepsilon_i) \hat{W}_{ci}(k), \quad (19)$$

where  $\mu_{ci}$  is the constant gain for the critic NN.

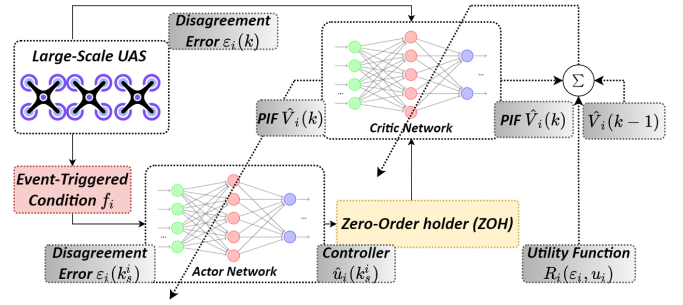


Fig. 4. A framework for implementation with event-triggered online actor-critic NNs for formation tracking control.

2) *Actor NN Design:* The actor NN, which is utilized to approximate the formation tracking controller (17),

$$u_i^*(k) = \begin{cases} -\alpha \varepsilon_i(k) - \frac{1}{2} \hat{W}_{ai}^T(k) \psi_i(\varepsilon_i(k)), & k \in \{k_i^j\}_{l \in \mathbb{Z}^+}, \\ u_i^*(k_i^j), & k \in [k_i^j, k_{i+1}^j). \end{cases} \quad (20)$$

where  $\hat{W}_{ai} \in \mathbb{R}^{s \times m}$  is the estimation of  $W_{ai}^*$ , which is trained by the actor NN update law

$$\hat{W}_{ai}(k+1) = \begin{cases} \hat{W}_{ai}(k) - \mu_{ai} T \psi_i(\varepsilon_i) \psi_i^T(\varepsilon_i) \\ \quad \times (\hat{W}_{ai}(k) - \hat{W}_{ci}(k)), & k \in \{k_i^j\}_{l \in \mathbb{Z}^+}, \\ \hat{W}_{ai}(k), & k \in [k_i^j, k_{i+1}^j). \end{cases} \quad (21)$$

where  $\mu_{ai}$  is the constant gain for actor NN.

**Definition 1.** A system is said to be ultimately uniformly bounded (UUB) if there exist positive constants  $\delta$  and  $\rho$ ,  $\forall x(k_0) < \delta$ ,  $\exists T(\rho, \delta) \geq 0$  such that

$$\forall k > k_0 + T, \quad \|x(k)\| < \rho.$$

**Theorem 2.** For any UAV  $a_i$  in the large-scale UAS (3) with bounded initial states, the optimized event-triggered formation tracking controller (20) is utilized, where the iterative update laws of the actor-critic NNs are (19) and (21), and the constant gains involved satisfy the conditions,

$$\mu_{ci} > \mu_{ai} > 0, \quad 0 < T < \frac{\mu_{ci} - \mu_{ai}}{\mu_{ci}^2 \lambda_{\max}^{\psi_i \psi_i^T}}, \quad (22)$$

where  $\lambda_{\max}^{\psi_i \psi_i^T}$  is the maximum eigenvalues of  $\psi_i(k) \psi_i^T(k)$ . Then, the disagreement error  $\varepsilon_i(k)$ , and the estimation errors  $\tilde{W}_{ai}(k)$ ,  $\tilde{W}_{ci}(k)$  of the actor-critic NN weights are UUB, where  $\tilde{W}_{ai}(k) = \hat{W}_{ai}(k) - W_{ai}^*$  and  $\tilde{W}_{ci}(k) = \hat{W}_{ci}(k) - W_{ci}^*$ .

*Proof.* The stability analysis process will be divided into two parts according to the trigger instants  $k \in \{k_i^j\}_{l \in \mathbb{Z}^+}$  and the trigger instant intervals  $k \in [k_i^j, k_{i+1}^j)$ .

(a) For  $\forall k \in [k_i^j, k_{i+1}^j)$ , the triggering condition  $f_i(k) \leq 0$  is satisfied, then design the Lyapunov function as follows,

$$\mathcal{V}(k) = \sum_{i=1}^N \varepsilon_i^T(k) \varepsilon_i(k) + \frac{1}{2} \sum_{i=1}^N \text{Tr} \{ \tilde{W}_{ai}^T(k) \tilde{W}_{ai}(k) \} + \frac{1}{2} \sum_{i=1}^N \text{Tr} \{ \tilde{W}_{ci}^T(k) \tilde{W}_{ci}(k) \} > 0. \quad (23)$$

For the first difference of  $\mathcal{V}(k)$ ,

$$\begin{aligned} \Delta \mathcal{V}(k) &\leq -\sum_{i=1}^N \frac{\mu_{ai} T}{2} \text{Tr}(\tilde{W}_{ai}^T(k) \psi_i \psi_i^T \tilde{W}_{ai}(k)) \\ &\quad - \sum_{i=1}^N \frac{(\mu_{ci} - \mu_{ai}) T}{2} \text{Tr}(\tilde{W}_{ci}^T(k) \psi_i \psi_i^T \tilde{W}_{ci}(k)) \quad (24) \\ &\quad + \sum_{i=1}^N \frac{\mu_{ci} T}{2} \text{Tr}(W_{ci}^* \psi_i \psi_i^T W_{ci}^*), \end{aligned}$$

where  $\sum_{i=1}^N \frac{\mu_{ci} T}{2} \text{Tr}(W_{ci}^* \psi_i \psi_i^T W_{ci}^*) \leq \tau$ ,  $\tau > 0$ . and  $\zeta = \min_{i \in [1, N]} \left\{ \mu_{ai} T \lambda_{\max}^{\psi_i \psi_i^T}, (\mu_{ci} - \mu_{ai}) T \lambda_{\max}^{\psi_i \psi_i^T} \right\}$ . Then, one has  $\Delta \mathcal{V}(k) \leq -\zeta \mathcal{V}_l(k) + \tau$ . Then, with  $k$  iterations, one has

$$\begin{aligned} \mathcal{V}(k) &\leq (1 - \zeta)^k \mathcal{V}_l(0) + \frac{\tau}{\zeta} \left( 1 - (1 - \zeta)^k \right) \\ &\quad - \mathcal{V}_l(0) + \sum_{i=1}^N \varepsilon_i^T(0) \varepsilon_i(0). \quad (25) \end{aligned}$$

(b) For  $\forall k \in \{k_j^i\}_{i \in \mathbb{Z}^+}$ , the controller  $u_i^*(k_j^i)$  is updated, then design the Lyapunov function as follows,

$$\begin{aligned} \mathcal{V}'(k) &= \sum_{i=1}^N \beta^k V_i(\varepsilon_i(k)) + \frac{1}{2} \sum_{i=1}^N \text{Tr} \{ \tilde{W}_{ai}^T(k) \tilde{W}_{ai}(k) \} \\ &\quad + \frac{1}{2} \sum_{i=1}^N \text{Tr} \{ \tilde{W}_{ci}^T(k) \tilde{W}_{ci}(k) \} > 0. \quad (26) \end{aligned}$$

Similarly to the previous case,  $\mathcal{V}'(k)$  is bounded. For both cases, the disagreement error  $\varepsilon(k)$  of the system and all the estimation errors  $\tilde{W}_{ai}(k)$ ,  $\tilde{W}_{ci}(k)$  are UUB.  $\square$

## V. MIXED REALITY EXPERIMENT

In this section, a mixed reality experimental platform is constructed to verify the proposed optimal assignment and optimal control algorithms for large-scale UAV formation.

### A. Mixed Reality Experimental Platform

In order to verify the effectiveness of the algorithm, the mixed reality large-scale UAV experimental platform shown in Fig. 5 is established, including a real experimental scenario and a virtual experimental scenario. The real scenario contains only several physical UAV nodes limited by the experimental site, and their state information is obtained by the motion capture system – OptiTrack, and transmitted to the formation controller via a router. In contrast, the virtual scenario can contain hundreds of virtual UAV nodes, whose dynamics model is designed as a second-order model. The formation controller uses the obtained state information of all nodes to calculate the control input of each node and send it back to each node for formation control. And to visualize the formation effect, the state information of physical nodes is also sent to the virtual scene for real-time display.

In this experiment, 4 physical UAVs and 116 virtual UAVs, a total of 120 UAVs, are used to achieve a large-scale formation. The Laplacian matrix used in the system is

$$L = \begin{bmatrix} 4 & -1 & -1 & 0 & \cdots & 0 & -1 & -1 \\ -1 & 4 & -1 & -1 & 0 & \cdots & 0 & -1 \\ -1 & -1 & 4 & -1 & -1 & 0 & \cdots & 0 \\ \cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \\ -1 & -1 & 0 & \cdots & 0 & -1 & -1 & 4 \end{bmatrix} \in \mathbb{R}^{120 \times 120}.$$

The diagonal matrix  $B = \text{diag}(\text{mod}(N, 2))$ . And the other parameters are listed in Table I.

### B. Results Analysis

As shown in Fig. 6, the large-scale UAS successfully generates collision-free optimal formation switching trajectories between multiple formations using the optimal assignment algorithm. With the optimized formation tracking controller (20), the large-scale mixed reality UAV nodes can follow the generated expected trajectories without collision. And the

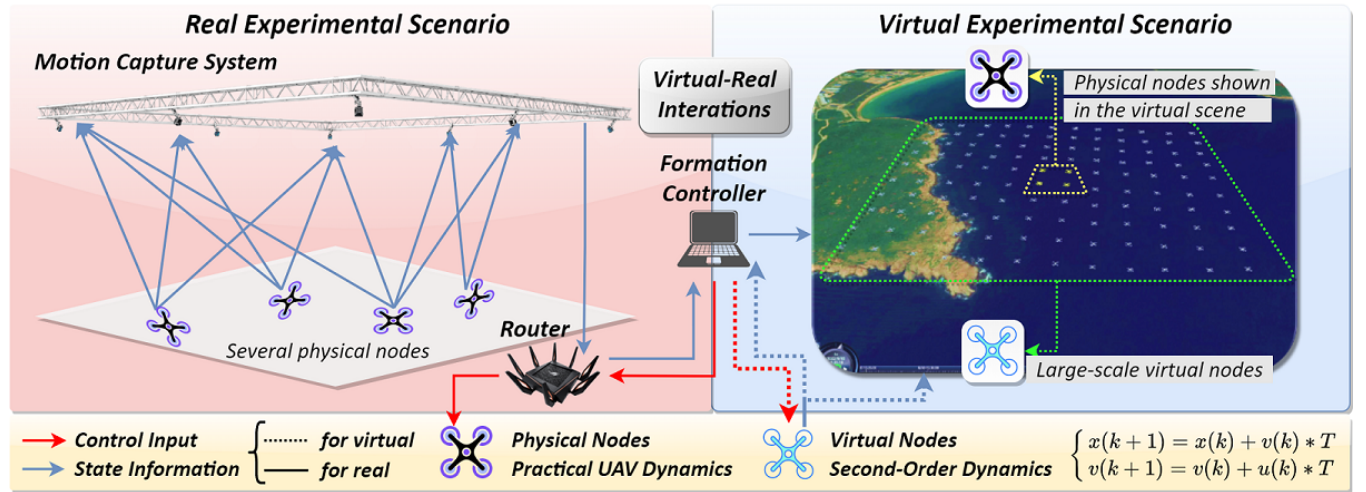
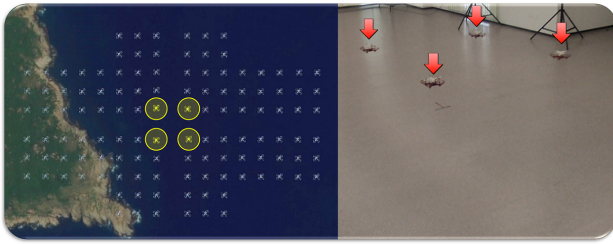


Fig. 5. Mixed reality experimental platform framework. To implement the large-scale UAV switching formation control, several physical UAV nodes and large-scale virtual UAV nodes are used to achieve mixed reality interactions.

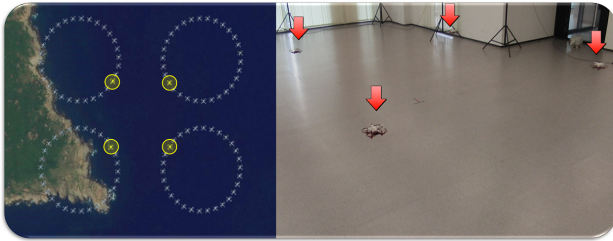
TABLE I  
EXPERIMENTAL SETUP

Parameter	Value	Description
$r$	0.14	Radius of UAV (m)
$s$	60	RBF NN nodes
$t_i$	1	Gaussian Basis Function width
$v_i$	[-3,3]	Gaussian Basis Function center.
$\hat{W}_{ai}(0)$	$\{0.3\}_{60 \times 2}$	Initial actor NN weights
$\hat{W}_{ci}(0)$	$\{0.3\}_{60 \times 2}$	Initial critic NN weights
$\mu_{ai}$	6	Actor NN gain
$\mu_{ci}$	8	Critic NN gain
$\alpha$	[6,4]	Parameter in controller $u(k)$
$T$	0.01	Sampling period (s)

formation tracking error  $\xi(k)$  is bounded as shown in Fig. 7. With the event-triggered mechanism, the controller update frequency per UAV is significantly reduced. As shown in Fig. 8, the event-triggered algorithm can reduce the update frequency to 27% of the time-triggered algorithm, which updates the controller continuously or periodically. And the triggering instants sequence for UAVs 1–30 in the last 1000 iterations is also illustrated, where the triggering instants for each UAV are aperiodically and on-demand. Fig. 9 illustrates that the actor and critic NN weights are UUB. And the optimal formation tracking control problem is solved.



(a) Cross-shaped mixed reality large-scale formation



(b) Circular mixed reality large-scale formation

Fig. 6. The large-scale UAV switching formations in the experiment.

### C. Summary

The mixed reality platform built in this study combines the strong engineering applicability of physical experimental platforms with the high flexibility and low cost of virtual simulation platforms. On the one hand, compared with the fully virtual simulations, it can more accurately verify the effectiveness of the proposed algorithm for practical UAV dynamics and real-world experimental scenes. On the other hand, compared with the fully physical experiments, the platform has significant advantages in experimental cost and field requirements, which can achieve a heavyweight UAV formation show with lightweight experimental equipment.

## VI. CONCLUSION

For the large-scale switching formation tracking problem, this study introduced the optimal formation assignment algorithm and proposed an optimal controller, which broken the main technical limits in decision and control layers of large-scale UASs. Further, compared with the time-triggered controller, the designed event-triggered algorithm reduced the update rate for controller and actor NN to 27%. More notably, this paper integrated the strong engineering applicability of practical experiments and high flexibility of virtual platforms to practically achieve large-scale switching formation. However, this study ignored the spatial constraints within the practical scene and the dynamics accuracy of the virtual UAV. In future work, autonomous obstacle avoidance in large-scale switching formation will be realized, and the virtual UAV dynamics will be improved in more detail.

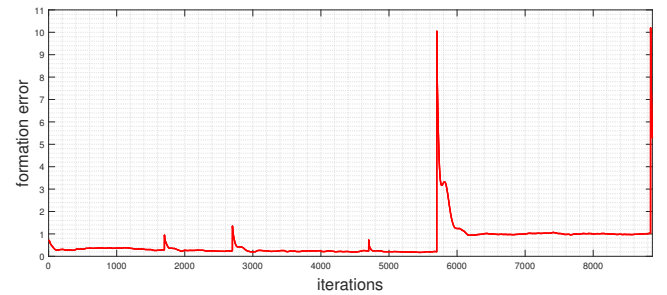


Fig. 7. The norm of formation tracking error  $\|\xi(k)\|$ . The moments when the tips appear in the curve are the switch from one formation to another.

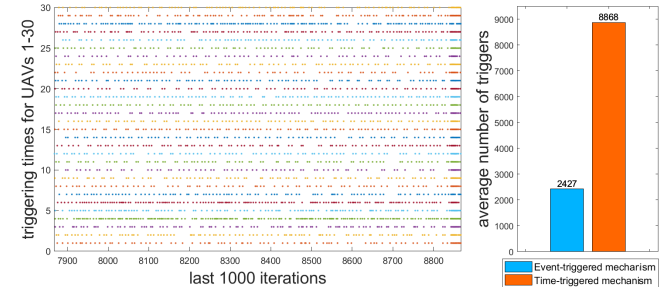


Fig. 8. The triggering instants sequence and the mean triggers for the event-triggered mechanism compared to the time-triggered mechanism.

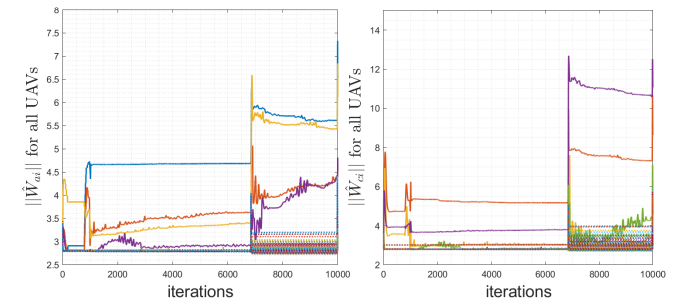


Fig. 9. Actor-critic NN weight norm for all UAVs, where the solid lines represent the physical UAVs and the dotted lines represent the virtual UAVs.

## REFERENCES

- [1] H. Zhang, T. Feng, H. Liang, et al, "LQR-Based Optimal Distributed Cooperative Design for Linear Discrete-Time Multiagent Systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 599-611, Mar. 2017.
- [2] C. O. Aguilar and A. J. Krener, "Numerical Solutions to the Bellman Equation of Optimal Control," *J. Optim. Theory Appl.*, vol. 160, no. 2, pp. 527-552, Feb. 2014.
- [3] B. Kiumarsi, F. L. Lewis, H. Modares, et al, "Reinforcement -learning for optimal tracking control of linear discrete-time systems with unknown dynamics," *Automatica*, vol. 50, no. 4, pp. 1167-1175, Apr. 2014.
- [4] H. J. Kappen, "Optimal control theory and the linear Bellman equation," in *Bayesian Time Series Models*, 1st ed., D. Barber, A. T. Cengil, and S. Chiappa, Eds. Cambridge University Press, 2011, pp. 363-387.
- [5] R. Kamalapurkar, J. A. Rosenfeld, and W. E. Dixon, "Efficient model-based reinforcement learning for approximate online optimal control," *Automatica*, vol. 74, pp. 247-258, Dec. 2016.
- [6] B. Pang, Z. P. Jiang, and I. Mareels, "Reinforcement learning for adaptive optimal control of continuous-time linear periodic systems," *Automatica*, vol. 118, p. 109035, Aug. 2020.
- [7] A. Perrusqua and W. Yu, "Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: An overview," *Neurocomputing*, vol. 438, pp. 145-154, May 2021.
- [8] J. Zhang, Z. Wang, and H. Zhang, "Data-Based Optimal Control of Multiagent Systems: A Reinforcement Learning Design Approach," *IEEE Trans. Cybern.*, vol. 49, no. 12, pp. 4441-4449, Dec. 2019.
- [9] H. Liu, Q. Meng, F. Peng, et al, "Heterogeneous formation control of multiple UAVs with limited-input leader via reinforcement learning," *Neurocomputing*, vol. 412, pp. 63-71, Oct. 2020.
- [10] W. Zhao, H. Liu and F. L. Lewis, "Robust Formation Control for Cooperative Underactuated Quadrotors via Reinforcement Learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4577-4587, Oct. 2021.
- [11] Y. Guo, G. Chen, and T. Zhao, "Learning-based collision-free coordination for a team of uncertain quadrotor UAVs," *Aerosp. Sci. Technol.*, vol. 119, p. 107127, Dec. 2021.
- [12] H. Li, Y. Wu, and M. Chen, "Adaptive Fault-Tolerant Tracking Control for Discrete-Time Multiagent Systems via Reinforcement Learning Algorithm," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1163-1174, Mar. 2021.
- [13] M. Long, H. Su, and Z. Zeng, "Model-Free Event-Triggered Consensus Algorithm for Multiagent Systems Using Reinforcement Learning Method," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 52, no. 8, pp. 5212-5221, Aug. 2022.
- [14] X. Guo, W. Yan, and R. Cui, "Event-Triggered Reinforcement Learning-Based Adaptive Tracking Control for Completely Unknown Continuous-Time Nonlinear Systems," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3231-3242, Jul. 2020.
- [15] B. Yan, P. Shi, and C. C. Lim, "Robust Formation Control for Nonlinear Heterogeneous Multiagent Systems Based on Adaptive Event-Triggered Strategy," *IEEE Trans. Autom. Sci. Eng.*, pp. 1-13, Aug. 2021.
- [16] V. Narayanan and S. Jagannathan, "Event-Triggered Distributed Control of Nonlinear Interconnected Systems Using Online Reinforcement Learning With Exploration," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2510-2519, Sep. 2018.
- [17] X. Yang, H. He, and D. Liu, "Event-Triggered Optimal Neuro-Controller Design With Reinforcement Learning for Unknown Nonlinear Systems," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 49, no. 9, pp. 1866-1878, Sep. 2019.
- [18] X. Yang and H. He, "Decentralized Event-Triggered Control for a Class of Nonlinear-Interconnected Systems Using Reinforcement Learning," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 635-648, Feb. 2021.
- [19] H. Li, Y. Wu, M. Chen, et al, "Adaptive Multigradient Recursive Reinforcement Learning Event-Triggered Tracking Control for Multiagent Systems," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1-13, Jul. 2021.
- [20] Z. Peng, R. Luo, J. Hu, et al, "Distributed Optimal Tracking Control of Discrete-Time Multiagent Systems via Event-Triggered Reinforcement Learning," *IEEE Trans. Circuits Syst. Regul. Pap.*, pp. 3689-3700, Jun. 2022.
- [21] W. Bai, T. Li, Y. Long, et al, "Event-Triggered Multigradient Recursive Reinforcement Learning Tracking Control for Multiagent Systems," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1-14, Jul. 2021.
- [22] J. Lu, Q. Wei, Y. Liu, et al, "Event-Triggered Optimal Parallel Tracking Control for Discrete-Time Nonlinear Systems," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 52, no. 6, pp. 3772-3784, Jun. 2022.
- [23] S. Zhao, J. Wang, H. Wang, et al, "Goal representation adaptive critic design for discrete-time uncertain systems subjected to input constraints: The event-triggered case," *Neurocomputing*, vol. 492, pp. 676-688, Jul. 2022.
- [24] F. Tang, B. Niu, G. Zong, et al, "Periodic event-triggered adaptive tracking control design for nonlinear discrete-time systems via reinforcement learning," *Neural Netw.*, vol. 154, pp. 43-55, Oct. 2022.
- [25] S. Agarwal and S. Akella, "Simultaneous Optimization of Assignments and Goal Formations for Multiple Robots," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, Brisbane, QLD, pp. 6708-6715, May 2018.