Research Article

A Q-Learning-Based Parameters Adaptive Algorithm for Formation Tracking Control of Multi-Mobile Robot Systems

Chen Zhang ¹, Wen Qin ¹, Ming-Can Fan ², Ting Wang ¹ and Mou-Quan Shen ¹

¹College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing 211816, Jiangsu, China

²School of Mathematics and Statistics, Huizhou University, Huizhou 516007, Guangdong, China

Correspondence should be addressed to Wen Qin; qinwen.wts@163.com

Received 10 September 2021; Revised 7 November 2021; Accepted 15 December 2021; Published 5 January 2022

Academic Editor: Miaomiao Wang

Copyright © 2022 Chen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes an adaptive formation tracking control algorithm optimized by Q-learning scheme for multiple mobile robots. In order to handle the model uncertainties and external disturbances, a desired linear extended state observer is designed to develop an adaptive formation tracking control strategy. Then an adaptive method of sliding mode control parameters optimized by Q-learning scheme is employed, which can avoid the complex parameter tuning process. Furthermore, the stability of the closed-loop control system is rigorously proved by means of matrix properties of graph theory and Lyapunov theory, and the formation tracking errors can be guaranteed to be uniformly ultimately bounded. Finally, simulations are presented to show the proposed algorithm has the advantages of faster convergence rate, higher tracking accuracy, and better steady-state performance.

1. Introduction

A multi-mobile robot system can present intelligent behaviours through mutual cooperation and achieve work efficiency and fault tolerance that a single individual cannot provide, so that it can complete some more difficult tasks. The coordinated formation control of multiple mobile robots has received extensive attention due to its important applications in the industrial and medical field [1]. The most existing control methods dealing with formation control problems of multiple mobile robots mainly include behaviour-based control [2], virtual structures [3], and leader-follower architecture [4–6]. As a decentralized control strategy, the leader-follower formation structure has become the preferred control strategy due to its simplicity and scalability and requires less computation and communication resources than other strategies [7]. The movement types of mobile robots are divided into omnidirectional mobile robot (OMR) (holonomic) and non-holonomic one [8]. Concerning the uncertainty and disturbance of robot docking, reference [9] proposed a novel robust containment architecture for nonholonomic

mobile robot formations with docking capabilities, which realized multirobot formation maintenance/switching, docking, and collision avoidance. In [10], a dynamic control law was developed for the cooperative target encircling problem of multiple unicycle mobile robots subject to heterogeneous input disturbances generated by the linear exogenous system.

Since nonholonomic mobile robots have fewer controllable degrees of freedom (DOFs) than holonomic mobile robots, geometric constraints are introduced on the robot's motion. Common examples with incomplete constraints are unicycles and car-like wheeled mobile robots. In contrast, holonomic mobile robots have the same number of controllable DOFs and total DOFs, which makes them very flexible and being able to move within the workspace without geometric constraints (e.g., they can perform both rotation and lateral translation). A typical example of this class of vehicles is the omnidirectional robot with mechanical wheels. More details on the types and configurations of mobile robots can be found in [11]. References [12, 13] studied the formation control problem of multiple omnidirectional mobile robots. Reference [12]

developed a distributed adaptive control law for a multi-robot system by obtaining information from moving targets through some mobile robots and using backstepping control technology in formation control. Reference [13] proposed an improved collision avoidance and formation control to configure a multirobot system optimized for omnidirectional visual simultaneous localization and mapping. However, the performance of the control schemes in [12, 13] will deteriorate when there exist uncertainties in the kinematics and dynamics of the omnidirectional mobile robots.

In the research of formation control of complex nonlinear systems, sliding mode control (SMC) is an effective robust controller for suppressing disturbances because of its excellent characteristics of being insensitive to system parameter changes and external disturbance when the system enters the sliding mode. In [14], a nonsingular fast terminal SMC was proposed, which can drive the tracking error to zero in finite time. Reference [15] investigated the leader-follower formation control for multiwheel mobile robots by combining a motion controller with a dynamic controller based on sliding mode. Considering the bounded external disturbance and parameter uncertainty of mobile robots, reference [16] proposed a dual-loop attitude tracking robust controller for mobile robots, using SMC with modified arrival law to ensure that the actual speed converges within a finite time. The main disadvantage of SMC is chattering phenomenon due to the discontinuity of the control law. In order to alleviate chattering, adaptive SMC [17] and higher order SMC [18, 19] have been proposed. However, these control methods may bring serious chattering even leading to instability when the system is exposed to a dynamic environment with large uncertainties and disturbances. Moreover, the traditional SMC is conservative to some extent since it ignores the information of uncertainties and disturbances. An effective way to solve this problem is using the disturbance estimation and compensation to decrease the conservatism and improve the control performance. Reference [20] proposed a disturbance observer and super-twisting SMC for the multirobot formation. Reference [21] designed an adaptive high-gain observer for the robot to estimate the nonlinearity that appears in the dynamics of the wheeled robot. In [22], the active disturbance rejection control technology was used to estimate the external disturbance in the inner loop of the double closed-loop strategy. On the other hand, the velocities of omnidirectional mobile robots cannot often be measured due to the lack of sensors which are needed in controller design.

Extended state observer (ESO) based active disturbance rejection control (ADRC), proposed by Han [23], is a powerful tool to cope with uncertainties and external disturbances. The key idea of ADRC is that the total disturbances (including internal uncertain dynamics, cross-coupling, and external disturbances), regarded as an extended state of the system, can be estimated by the ESO and then compensated in the control signal. As a matter of fact, the ESO is a state observer to estimate both the system states and the total disturbances. Considering the

advantage of it, ESO is adopted to estimate the total disturbances and then followed by the ESO-based controller constructed to compensate it. Reference [24] employed a nonlinear extended state observer (NESO) to estimate unknown states as well as uncertainties and designed a robust finite-time tracking control scheme to handle wheeled mobile robots with parameter uncertainties and disturbances. Reference [25] used NESO-based estimation and compensation signals into the closed-loop control method; an NESO-based decoupling control method was then proposed to solve the attitude control problem of hypersonic gliding vehicles. In [26], an NESO was used to estimate the system uncertainties, and a saturation-resistant adaptive SMC was designed based on the estimated values to achieve robust trajectory tracking for a wheeled mobile robot. However, it is difficult to find appropriate nonlinear functions to design an NESO in practice. For convenience of theoretical analysis, Gao proposed a linearized and bandwidth-based linear ADRC (LADRC) to simplify parameter tuning and standardize controller tuning [27]. Linear ESO (LESO) with nested inner and outer loops was used in [28] to actively estimate and eliminate generalized interference and improve the estimation accuracy in various practical models. The proposal of LADRC makes the design and adjustment of the controller easier and more effective, and the tracking error will be more decreased than some classic control structures [16], which greatly promotes the engineering application of ADRC. However, the LADRC is loss of design flexibility because the LADRC parameters are adjusted based on bandwidth. A general LADRC with more tuning parameters was proposed in [29], and nowadays LADRC has been employed to various cases in engineering application and becomes much more popular [30, 31].

Reinforcement learning (RL) has been developed rapidly in recent years. As one of the important algorithms of RL, Q-learning is off-policy, tabular, model-free, and based on temporal-difference methods [32]. It has the advantages of not relying on models and having good learning effects for complex systems. In order to improve the control performance, some scholars combined Q-learning with PID control and proposed many excellent control methods [33, 34]. In the autonomous underwater vehicle system, reference [35] proposed a Q-learning PID controller based on RBFNN to improve control performance, in which Q-learning neural network was used to adaptively optimize control parameters. Reference [36] combined model-based Q-learning into the predictive control setting to provide closed-loop stability in online learning and ultimately improve the performance of the limited range controller. Chen proposed an adaptive auto-disturbance-rejection controller parameter adaptation method based on Q-learning for ship heading control with multiple uncertainties due to wind, wave, and current interference [37].

Inspired by the above statements, this paper investigates the formation tracking control of a multiple omnidirectional mobile robot system. The considered mobile robots have internal modelling uncertainties and external

disturbances (considered as total disturbances). To handle the disturbances, an LESO is constructed, and the total disturbances can be effectively estimated through the ESO. Then, on the basis of distributed formation tracking control architecture, an LESO-based SMC (LESO-SMC) is designed for each OMR to ensure that the observer errors and the formation neighbourhood errors are uniformly ultimately bounded (UUB). However, LESO based control is not widely used in practice because there are not adequate methods for LESO parameter adjustment. In view of this, and furthermore for obtaining better control performance, Q-learning is employed to optimize the bandwidth parameters of LESO and the control parameters of SMC and to avoid the complex parameter tuning process. In addition, a simulation example is given to verify the effectiveness of the proposed method.

The main features of the proposed methods are summarized as follows:

- (1) An LESO is constructed to estimate the ‘total disturbances’ in real time, including both internal parameter uncertainties and external disturbances, and then an LESO-SMC based formation protocol is developed for the OMR system. The LESO provides distinctly better robustness against ‘total disturbances’ by providing accurate input variables to the control system, including the states of the mobile robot at each order, as well as the extended state representing the ‘total disturbances’. Then a corresponding improved strategy on the SMC is made to compensate the influence of the ‘total disturbances’, which ensures a faster convergence rate and decreases the conservatism of the traditional SMC.
- (2) To take full advantages of the LESO-SMC, an adaptive method of LESO-SMC parameters optimized by Q-learning algorithm is proposed in the formation tracking control of the OMR system. Q-learning is introduced to perform online parameter adaptation (including the observer, sliding mode variables, and controller parameters), which exhibits better formation tracking performance and avoids the complex parameter tuning process.

The organization of this paper is as follows. In Section 2, the dynamic model and some preliminary knowledge are outlined. In Section 3, the proposed controller design based on LESO and SMC is presented. Both the Q-learning algorithm and the Q-learning parameter tuning process are also introduced. The results of numerical simulations are discussed in Section 4, followed by the conclusion of this paper.

Notations: A^T and A^{-1} represent the transpose and inverse of matrix A , respectively. R^n represents n -dimensional real column vector set. I_n denotes an $n \times n$ unit matrix. 0_n denotes an $n \times n$ zero matrix. $\otimes$ stands for Kronecker product. $\text{sign}(\cdot)$ is the sign function. $\|\cdot\|$ represents the Euclidean norm. $\text{diag}x_1, x_2, \dots, x_n$ denotes the diagonal matrix with its diagonal entries being

$x_1, x_2, \dots, x_n$. $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ represent the maximum and the minimum eigenvalues of matrix A , respectively.

2. Problem Formulation and Preliminaries

2.1. Dynamic Model. The Euler–Lagrange equation of motion can be used to describe the dynamic behaviour of an OMR. The dynamic model of the i th OMR can be described as [31]

$$M_i \ddot{q}_i + C_i \dot{q}_i + G_i = \tau_i, \quad (1)$$

where $i = 1, \dots, n$, $q_i = [x_i, y_i, \theta_i]^T \in R^3$ represents the position and orientation angle of the i th robot in the world coordinate frame. M_i is the inertia matrix, C_i is the Coriolis and centrifugal term, G_i is the gravitational force, and τ_i denotes the control input. It is assumed that the robots move on flat ground, where gravitational force G_i is 0.

Considering the unknown dynamic disturbances and model uncertainties, a new dynamic model is obtained as follows:

$$(M_i + \Delta M_i) \ddot{q}_i + (C_i + \Delta C_i) \dot{q}_i = \tau_i + d_i, \quad (2)$$

where $\Delta M_i \in R^{3 \times 3}$ and $\Delta C_i \in R^{3 \times 3}$ denote uncertainty terms, and $d_i \in R^3$ denotes the unknown disturbance term.

The above equation can be rewritten as

$$M_i \ddot{q}_i + C_i \dot{q}_i + \Delta_i = \tau_i + d_i, \quad (3)$$

where $\Delta_i = \Delta M_i \ddot{q}_i + \Delta C_i \dot{q}_i$ denotes the uncertain term.

Assumption 1. The unknown disturbance term d_i satisfies $\|d_i^{(j)}\| \leq \lambda_1$, $j = 0, 1, 2$, where λ_1 is an unknown positive constant.

Assumption 2. The uncertain terms ΔM_i and ΔC_i are bounded. Thus, the uncertain term Δ_i satisfies $\|\Delta_i^{(j)}\| \leq \lambda_2$, $j = 0, 1, 2$, where λ_2 is an unknown positive constant.

As for the LESO design in the following section, the total disturbance is extended as a new state. Define $x_{i,1} = q_i$, $x_{i,2} = \dot{q}_i$, $x_{i,3} = f_i = M_i^{-1}(d_i - \Delta_i) \in R^3$, where $x_{i,3}$ is the extended state of the total disturbances f_i . The dynamic model (3) can be transformed into the following system:

$$\begin{cases} \dot{x}_{i,1} = x_{i,2}, \\ \dot{x}_{i,2} = x_{i,3} + M_i^{-1} \tau_i - M_i^{-1} C_i x_{i,2}, \\ \dot{x}_{i,3} = \dot{f}_i. \end{cases} \quad (4)$$

The dynamic equation of the virtual leader is

$$\begin{cases} \dot{x}_l = v_l, \\ \dot{v}_l = u_l, \end{cases} \quad (5)$$

where $x_l = [x_{lx}, x_{ly}, x_{l\theta}]^T \in R^3$ and $v_l = [v_{lx}, v_{ly}, v_{l\theta}]^T \in R^3$ indicate the position and velocity of the virtual leader, respectively. u_l is the control input of the virtual leader which can be obtained by some followers.

Assumption 3. The derivative of the total disturbances f_i is bounded by an unknown constant ρ_i , i.e., $\|\dot{f}_i\| \leq \rho_i$.

Remark 1. Note that f_i in the dynamic model (4) indicates an unknown term, such as external disturbances and model uncertainties for the mechanism of OMRs. In practical applications, the total disturbances mainly include wheel-ground sliding, modelling uncertainty due to robot load variation, etc.

Remark 2. In practice, both the speed of the DC motor which drives the OMR forward and its derivative have upper bounds, i.e., q_i , $\dot{q}_i$, and $\ddot{q}_i$ are all bounded. Δ_i is related to $\dot{q}_i$ and $\ddot{q}_i$, so we can conclude that Δ_i and its derivative are bounded. Similarly, f_i and its time derivative are bounded as well. Therefore, the assumptions of Δ_i and f_i in Assumptions 1 and 2 are reasonable.

2.2. Algebraic Graph Theory. Consider an omnidirectional mobile robot system consisting of one virtual leader and n followers. Assume that each robot is a node, and the information exchange among follower robots can be described by a directed graph G . The graph $G = (V, E, A)$ is composed of the node set $V = \{v_1, \dots, v_n\}$, the edge set $E = \{(v_i, v_j) \in V \times V\}$, and the adjacency matrix $A = (a_{ij}) \in R^{n \times n}$. If there is an edge between agents i and j , i.e., $(v_i, v_j) \in E$, then $a_{ij} > 0$; otherwise $a_{ij} = 0$. The set of neighbours of node v_i is denoted by $N = \{j: (v_i, v_j) \in E\}$. The in-degree of node v_i is defined as $d_i = \sum_{j \in N} a_{ij}$. Then, the in-degree matrix of digraph G is $D = \text{diag}\{d_1, \dots, d_n\}$, and the Laplacian matrix of digraph G is $L = D - A$. A path in graph G from v_i to v_j is a sequence of distinct vertices starting with v_i and ending with v_j such that consecutive vertices are adjacent. Graph G is connected if there is a path between any two vertices. Graph G contains a directed spanning tree if there is a vertex (the root node) which can reach all the other vertices through a directed path. The virtual leader's adjacency weight matrix $B = \text{diag}\{b_1, \dots, b_n\}$, where b_i represents the topological weight of the communication between agent i and the virtual leader, $b_i > 0$, if there is communication between the agent i and the virtual leader; otherwise, $b_i = 0$.

Assumption 4. For the considered multiple mobile robot systems (2), graph G for the n follower robots contains a directed spanning tree; i.e., there is a vertex (the root node) which can reach all the other vertices through a directed path.

Lemma 1 (see [38]). *If G is a directed graph which contains a directed spanning tree and at least the root-node agent has access to the virtual leader, then the matrix $L + B$ is of full rank.*

3. Main Results

In this section, in order to achieve better formation tracking control performance, the LESO-SMC scheme will be

designed for system (3) in the presence of the unknown disturbances and model uncertainties, such that all follower OMRs can track the virtual leader with the given formation configuration in advance and maintain the same speed with the virtual leader. Furthermore, a parameter adaptation method based on Q-learning algorithm is involved in LESO-SMC to avoid the complex parameter tuning process, which displays a better formation tracking control performance.

3.1. Linear Extended State Observer Design. In this section, we use the LESO to estimate the OMR's total disturbances f_i , which include unknown disturbances and model uncertainties.

The LESO for system (4) is designed as follows:

$$\begin{cases} \dot{\hat{x}}_{i,1} = \hat{x}_{i,2} + \beta_1(x_{i,1} - \hat{x}_{i,1}), \\ \dot{\hat{x}}_{i,2} = \hat{x}_{i,3} + M_i^{-1}\tau_i - M_i^{-1}C_i\hat{x}_{i,2} + \beta_2(x_{i,1} - \hat{x}_{i,1}), \\ \dot{\hat{x}}_{i,3} = \beta_3(x_{i,1} - \hat{x}_{i,1}), \end{cases} \quad (6)$$

where $\hat{x}_{i,1}$, $\hat{x}_{i,2}$, $\hat{x}_{i,3}$, and $\hat{h}_i$ are the estimations of $x_{i,1}$, $x_{i,2}$, $x_{i,3}$, and h_i , respectively, and β_k ($k = 1, 2, 3$) is the observer gain to be determined. $\beta_1 = 3\omega_0$, $\beta_2 = 3\omega_0^2$, $\beta_3 = \omega_0^3$, where ω_0 is the ESO bandwidth [39].

Define the estimation errors as follows: $\tilde{x}_{i,j} = x_{i,j} - \hat{x}_{i,j}$ ($j = 1, 2, 3$); then the estimation error equation is given by

$$\begin{cases} \dot{\tilde{x}}_{i,1} = \tilde{x}_{i,2} - 3\omega_0\tilde{x}_{i,1}, \\ \dot{\tilde{x}}_{i,2} = \tilde{x}_{i,3} + M_i^{-1}C_i\tilde{x}_{i,2} - 3\omega_0^2\tilde{x}_{i,1}, \\ \dot{\tilde{x}}_{i,3} = \dot{f}_i - \omega_0^3\tilde{x}_{i,1}. \end{cases} \quad (7)$$

Let $\xi_i = [\xi_{i,1}, \xi_{i,2}, \xi_{i,3}]^T = [\tilde{x}_{i,1}, \tilde{x}_{i,2}/\omega_0, \tilde{x}_{i,3}/\omega_0^2]^T \in R^3$; equation (7) can be rewritten as

$$\dot{\xi}_i = \omega_0 A_1 \xi_i + \frac{N_2 M_i^{-1} C_i \tilde{x}_{i,2}}{\omega_0} + \frac{N_1 \dot{f}_i}{\omega_0^2}, \quad (8)$$

where $A_1 = \begin{bmatrix} -3I_3 & I_3 & 0_3 \\ -3I_3 & 0_3 & I_3 \\ -I_3 & 0_3 & 0_3 \end{bmatrix} \in R^{9 \times 9}$ is Hurwitz, $N_1 = \begin{bmatrix} 0_3 \\ 0_3 \\ I_3 \end{bmatrix}$,

$$N_2 = \begin{bmatrix} 0_3 \\ I_3 \\ 0_3 \end{bmatrix}.$$

Slightly different to the proof in [40], the convergence analysis of LESO (6) is given below.

Lemma 2. *Considering the estimation error dynamics (8), the LESO (6) is bounded stable if the observer bandwidth ω_0 is designed to satisfy the condition $\omega_0 - 2\|P\| \cdot \|\kappa\omega_0\|/\omega_0 - 1 > 0$.*

Proof. Since matrix A_1 is Hurwitz, there exists a positive definite matrix P_1 which satisfies $A_1^T P_1 + P_1 A_1 = -I_9$.

$$\|N_2 M_i^{-1} C_i \tilde{x}_{i,2}\| \leq \kappa \|\tilde{x}_{i,2}\| \leq \kappa \omega_0 \|\xi_{i,2}\|, \quad (9)$$

where κ is an unknown constant.

Consider a Lyapunov function candidate for (8) as

$$V_{eso} = \sum_{i=1}^n \xi_i^T P_1 \xi_i. \quad (10)$$

Then we obtain the time derivative of V_{eso} as

$$\begin{aligned} \dot{V}_{eso} &= \sum_{i=1}^n \left(\dot{\xi}_i^T P_1 \xi_i + \xi_i^T P_1 \dot{\xi}_i \right) \\ &= \left(\sum_{i=1}^n \omega_0 \xi_i^T A_1^T P_1 \xi_i + \frac{(N_2 M_i^{-1} C_i \tilde{x}_{i,2})^T P_1 \xi_i}{\omega_0} + \frac{(N_1 \dot{f}_i)^T P_1 \xi_i}{\omega_0^2} + \omega_0 \xi_i^T P_1 A_1 \xi_i + \frac{\xi_i^T P_1 (N_2 M_i^{-1} C_i \tilde{x}_{i,2})}{\omega_0} + \frac{\xi_i^T P_1 (N_1 \dot{f}_i)}{\omega_0^2} \right) \\ &= \sum_{i=1}^n \left(-\omega_0 \xi_i^T \xi_i + 2 \frac{\xi_i^T P_1 N_2 M_i^{-1} C_i \tilde{x}_{i,2}}{\omega_0} + 2 \frac{\xi_i^T P_1 N_1 \dot{f}_i}{\omega_0^2} \right) \\ &= -\omega_0 \xi^T \xi + 2 \frac{\xi^T P \bar{N} M_i^{-1} C_i \tilde{x}_{i,2}}{\omega_0} + 2 \frac{\xi^T P N \dot{f}}{\omega_0^2} \\ &\leq -\left(\omega_0 - 2 \frac{\|P\| \cdot \|\kappa \omega_0\|}{\omega_0} \right) \|\xi\|^2 + 2 \frac{\xi^T P N \dot{f}}{\omega_0^2}, \end{aligned} \quad (11)$$

where

$$\begin{aligned} \xi &= [\xi_1^T, \dots, \xi_n^T]^T, \\ P &= 1_n \otimes P_1, \\ N &= 1_n \otimes N_1, \\ \bar{N} &= 1_n \otimes N_2, \\ 1_n &= [1, 1, \dots, 1]^T \in R^n, \\ \dot{f} &= [\dot{f}_1^T, \dots, \dot{f}_n^T]^T. \end{aligned} \quad (12)$$

Using Young's inequality, one has

$$2 \frac{\xi^T P N \dot{f}}{\omega_0^2} \leq \|\xi\|^2 + \frac{\|P N\|^2 \|\dot{f}\|^2}{\omega_0^4}. \quad (13)$$

From Assumption 3, we have $\dot{f} \leq \rho$, where ρ is an unknown constant. It can be obtained from (11) that

$$\dot{V}_{eso} \leq -\left(\omega_0 - 2 \frac{\|P\| \cdot \|\kappa \omega_0\|}{\omega_0} - 1 \right) \|\xi\|^2 + \frac{\|P N\|^2 \rho^2}{\omega_0^4}. \quad (14)$$

If $\omega_0 - 2\|P\| \cdot \|\kappa \omega_0\|/\omega_0 - 1 > 0$, then $\|P N\|^2 \rho^2/\omega_0^4$ is bounded; hence the proposed LESO is bounded stable. The proof is completed.

To achieve a better tracking control performance, in this section, an LESO-SMC-based formation control scheme will

be proposed to ensure the tracking performance based on the dynamic models introduced in Section 2. $\square$

3.2. Sliding Mode-Based Formation Controller Design. Define the system neighbourhood errors as

$$\begin{cases} e_{i,1} = \sum_{j \in N} a_{ij} ((\hat{x}_{i,1} - \delta_i) - (\hat{x}_{j,1} - \delta_j)) + b_i (\hat{x}_{i,1} - x_l - \delta_i), \\ e_{i,2} = \sum_{j \in N} a_{ij} (\hat{x}_{i,2} - \hat{x}_{j,2}) + b_i (\hat{x}_{i,2} - v_l), \end{cases} \quad (15)$$

where $\delta_i = [\delta_{ix}, \delta_{iy}, \delta_{i\theta}]^T \in R^3$ denotes the desired relative position for the i th robot and the virtual leader.

According to (15), we design the ESO-based sliding mode surface of the overall formation for the i th agent as

$$S_i = \mu e_{i,1} + e_{i,2}, \quad (16)$$

where μ is a positive constant.

The time derivative of S_i is

$$\begin{aligned} \dot{S}_i &= \mu \dot{e}_{i,1} + \dot{e}_{i,2} \\ &= \mu e_{i,2} + \sum_{j \in N} a_{ij} (\dot{\hat{x}}_{i,2} - \dot{\hat{x}}_{j,2}) + b_i (\dot{\hat{x}}_{i,2} - u_l). \end{aligned} \quad (17)$$

The formation tracking controller based on ESO-SMC algorithm for the i th agent is designed as

$$\tau_i = M_i \left(M_i^{-1} C_i \hat{x}_{i,2} - \hat{x}_{i,3} + u_l + \left(\sum_{j \in N} a_{ij} + b_i \right)^{-1} \cdot \left(\sum_{j \in N} a_{ij} \tau_j - \mu e_{i,2} - c_1 S_i - c_2 \text{sgn}(S_i) \right) \right), \quad (18)$$

FIGURE 1: The proposed control block diagram.

where c_1 and c_2 are positive constants, and u_l is the control input of the virtual leader which can be obtained by some followers.

The designed control structure block diagram is shown in Figure 1.

3.3. Stability Analysis of State Tracking Error Dynamics. We have introduced LESO to estimate the states and uncertainties of each order of OMR; next we will propose the SMC-based formation controller.

Theorem 1. Consider the OMR system composed of (4) and (5) that satisfies Assumptions 1–4. Under the adaptive formation tracking control algorithm optimized by Q-learning scheme with the LESO (6), the SMC (18) and the Q-learning-based parameter adaptive algorithm (see Section 3.4), the estimation errors of the proposed LESO, and the formation tracking errors are guaranteed to be UUB.

Similar to [41], we comprehensively consider the closed-loop system composed of the estimation errors of the observer (6) and the sliding variable (16). The stability analysis of the closed-loop system is given below.

Proof. Define the positive definite Lyapunov function as

$$V = \frac{1}{2} \sum_{i=1}^n \xi_i^T P_1 \xi_i + \frac{1}{2} \sum_{i=1}^n S_i^T S_i, \quad (19)$$

where $M = \text{diag}\{M_i\}$, $C = \text{diag}\{C_i\}$.

Substituting (21)–(23) into (20), one can obtain

$$\begin{aligned} \sum_{i=1}^n S_i^T \dot{S}_i &= S^T (\mu e_2 + (L+B) \otimes I_3 \cdot (\hat{x}_3 + M^{-1} \tau - 1_n \otimes u_l)) \\ &= -c_1 S^T S - c_2 S^T \text{sgn}(S) \\ &= -c_1 \|S\|^2 - c_2 \|S\| < 0. \end{aligned} \quad (24)$$

with the time derivative of the second term in the right-hand side of (19) being

$$\begin{aligned} \sum_{i=1}^n S_i^T \dot{S}_i &= \sum_{i=1}^n S_i^T (\mu \dot{e}_{i,1} + \dot{e}_{i,2}) \\ &= \sum_{i=1}^n S_i^T \left(\mu e_{i,2} + \sum_{j \in N} a_{ij} (\hat{x}_{i,2} - \hat{x}_{j,2}) + b_i (\hat{x}_{i,2} - u_l) \right). \end{aligned} \quad (20)$$

The system neighbourhood error (15) can be expressed in matrix form as

$$\begin{cases} e_1 = (L+B) \otimes I_3 \cdot (\hat{x}_1 - 1_n \otimes x_l - \delta), \\ e_2 = (L+B) \otimes I_3 \cdot (\hat{x}_2 - 1_n \otimes v_l), \end{cases} \quad (21)$$

where $e_1 = [e_{1,1}^T, \dots, e_{n,1}^T]^T \in R^{3n}$, $e_2 = [e_{1,2}^T, \dots, e_{n,2}^T]^T \in R^{3n}$, $\hat{x}_1 = [\hat{x}_{1,1}^T, \dots, \hat{x}_{n,1}^T]^T \in R^{3n}$, $\hat{x}_2 = [\hat{x}_{1,2}^T, \dots, \hat{x}_{n,2}^T]^T \in R^{3n}$, $\delta = [\delta_1^T, \dots, \delta_n^T]^T \in R^{3n}$, $1_n = [1, 1, \dots, 1]^T \in R^n$.

The sliding surface (16) can be rewritten as

$$S = \mu e_1 + e_2. \quad (22)$$

Then the formation tracking controller (18) can be rewritten as

$$\tau = M(M^{-1}C\hat{x}_2 - \hat{x}_3 + 1_n \otimes u_l + (L+B)^{-1} \otimes I_3 \cdot (-\mu e_2 - c_1 S - c_2 \text{sgn}(S))), \quad (23)$$

Substituting (14) and (24) into (19), then we have

$$\begin{aligned} \dot{V} &\leq -\frac{1}{2} \left(\omega_0 - 2 \frac{\|P\| \cdot \|\kappa \omega_0\|}{\omega_0} - 1 \right) \|\xi\|^2 + \frac{\|PN\|^2 \rho^2}{2\omega_0^4} \\ &\quad - c_1 \|S\|^2 - c_2 \|S\| \leq -\varphi^T \Gamma \varphi + \varsigma, \end{aligned} \quad (25)$$

where $\varphi = [\xi^T, S^T]$, $\Gamma = \begin{bmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{bmatrix}$, $\varsigma = \|PN\|^2 \rho^2 / 2\omega_0^4 - c_2 \|S\|$,
with $\Gamma_1 = \text{diag} \{ \underbrace{\eta, \dots, \eta}_n \}$, $\eta = 1/2 (\omega_0 - 2\|P\| \cdot \|\kappa\omega_0\| /$

$\omega_0 - 1)$, $\Gamma_2 = \text{diag} \{ \underbrace{c_1, \dots, c_1}_n \}$. Since Γ_1 and Γ_2 are positive definite, one has

$$\dot{V} \leq -\lambda_{\min}(\Gamma)(\|\xi\|^2 + \|S\|^2) + \varsigma \leq -\lambda_{\min}(\Gamma) \left(\frac{1}{\lambda_{\max}(P)} \xi^T P \xi + S^T S \right) + \varsigma \leq -\chi V + \varsigma, \quad (26)$$

where $\chi = 2\lambda_{\min}(\Gamma) \min\{1, 1/\lambda_{\max}(P)\}$.

Based on Gronwall's inequality, one has

$$V(t) \leq V(0)e^{-\chi t} + \frac{\varsigma}{\chi} (1 - e^{-\chi t}). \quad (27)$$

Since $V(t) \geq 1/2 \xi^T P \xi \geq 1/2 \lambda_{\min}(P) \|\xi\|^2$, one has $1/2 \lambda_{\min}(P) \|\xi\|^2 \leq V(0)e^{-\chi t} + \varsigma/\chi (1 - e^{-\chi t})$.

Hence $\|\xi\| \leq \sqrt{2[V(0)\chi e^{-\chi t} + \varsigma(1 - e^{-\chi t})]/\chi \lambda_{\min}(P)}$; then we obtain the bound of the observer error as

$$\|\xi\| \leq \sqrt{\frac{2\varsigma}{\chi \lambda_{\min}(P)}}, \quad t \rightarrow \infty. \quad (28)$$

Since $V(t) \geq 1/2 S^T S = 1/2 \|S\|^2$, one has $\|S\| \leq \sqrt{2[V(0)\chi e^{-\chi t} + \varsigma(1 - e^{-\chi t})]/\chi}$. Similarly, we can conclude that $\|S\| \leq \sqrt{2\varsigma/\chi}$, $t \rightarrow \infty$. Therefore ξ and S are UUB. By choosing appropriate parameters, the neighbourhood error can be attracted into a small stable region containing the origin. The proof is completed.

3.4. Q-Learning Based Parameter Optimization Process.

The learning process of Q-learning is of continuous environmental interaction. First, at time t , select an action value $\bar{A}_t$. Then the agent will transfer from the original state $\bar{S}_t$ to a new state $\bar{S}_{t+1}$ with a probability of $P(\bar{S}_{t+1}|\bar{S}_t, a)$. At this time, due to environmental interaction, the agent can get a feedback return R , then the time variable is updated, and the agent restarts the above steps in the new state until the optimal strategy is finally obtained. The Q-learning algorithm is shown below.

Parameters: $\alpha \in (0, 1]$, small $\varepsilon > 0$, $\gamma \in (0, 1]$.

Step 1. For each state-action pair (s, a) , $\forall s \in \bar{S}, a \in \bar{A}$, initialize the table entry $Q(s, a)$ to zero. Observe the current state s .

Step 2. Loop for each episode: Select an action a for state s using ε -greedy policy and execute it, use the same method to choose a_t for state s_t , then receive immediate reward r_t , and observe the new state s_{t+1} . Update the table entry for $Q(s_t, a_t)$ as follows:

$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$, $s_t \leftarrow s_{t+1}$, until s is the terminal state and all of $Q(s, a)$ are convergent.

Step 3. Output the final policy $\pi(s) = \arg \max Q(s, a)$.

In this paper, we consider the application of Q-learning ideas to the optimization of observers and controllers. Regarding the formation error and the derivative of the error as the state of Q-learning, dividing the controller parameters selection into a reasonable range and combining the divided values as an optional action, a Q-learning algorithm for observer and controller parameters optimization can be obtained. The specific steps of the algorithm are as follows:

- States set are composed of $\bar{S}_t(\varepsilon_i)$, $(i = 1, 2)$.
- Actions set are described as $\omega_0 \in [2, 4]$, $c_1 \in [5, 7]$, $c_2 \in [5, 7]$, $\mu \in [3, 5]$. Reasonably divide the interval of the four parameters and randomly select and combine them to obtain the actions set.
- In this paper, the stage performance index, i.e., the average of the squared errors of 10 sampling times, is used to design the reward function. Rewards are given as follows:

$$\begin{cases} J = \sum_{k=1}^{10} \frac{e_{i,1}^2(k)}{10}, \\ r = \begin{cases} 100, & J \leq \sigma, \\ -100, & J > \sigma. \end{cases} \end{cases} \quad (29)$$

Perform the learning of the value function Q according to the algorithm flow mentioned above. Then we get the learned Q table and the optimal strategy $\pi(s) = \arg \max Q(s, a)$ for online parameter selection.

There are 3 termination conditions as follows. If any one of them is fulfilled then the Q-value training is terminated.

- It is not desired in practice that the formation error in the control is too large, which will make little sense to continue to iteratively calculate the Q table. So when $|e_{i,1}| > 10$, the training is terminated and reinitialized.
- The control process has reached a steady state; then the training is terminated, i.e., $|e_{i,1}| < 0.001$ and $|e_{i,2}| < 0.01$.

- (c) The simulation time $t = 500s$; then the training is terminated.

Denote L_t as the training times. When $L_t = 800$, the training is terminated and the trained Q table can be obtained for online control.

Remark 3. Both the formation error and the error variation rate are divided into 7 linguistic values, {NB, NM, NL, Z, PL, PM, PB}, so that the number of state set elements is 49. Choose a reasonable range for each element in the parameter set $\Pi = [\omega_0 c_1 c_2 \mu]$. Based on the experience of adjusting the LESO-SMC controller parameters, set $\omega_0 \in [2, 4]$, $c_1 \in [5, 7]$, $c_2 \in [5, 7]$, $\mu \in [3, 5]$. Let $p_1 = p_2 = p_3 = p_4 = 3$; i.e., each parameter has 3 possible values; then the number of equivalent selectable actions is $\bar{A}_t = 81$. Therefore, the value function matrix $Q \in R^{81 \times 49}$ for the formation system.

Remark 4. The introduction of Q-learning avoids the process of selecting and optimizing the controller, the observer, and the sliding mode parameters. After Q-learning optimization, the convergence performance of the controller and observer can be guaranteed.

4. Numerical Simulations

In this section, numerical simulation examples are used to illustrate the previous conclusions and the effectiveness of the proposed control scheme. Consider a scenario where a multiagent systems composed of three OMRs (followers) are simultaneously tracking a preset target (the virtual leader).

The communication topology G is given in Figure 2. The corresponding Laplacian matrix L and the adjacency weight matrix B can be described as

$$L = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ -1 & 0 & 1 \end{bmatrix}, \quad (30)$$

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Each dynamics model of the three OMRs can be described by an Euler-Lagrange equation as follows:

$$M_i \ddot{q}_i + C_i \dot{q}_i + \Delta_i = \tau_i + d_i, \quad i = 1, 2, 3, \quad (31)$$

where

$$M_i = \begin{bmatrix} m & 0 & m d \sin \theta_i \\ 0 & m & -m d \cos \theta_i \\ m d \sin \theta_i & -m d \cos \theta_i & I_v \end{bmatrix},$$

$$C_i = \begin{bmatrix} 0 & 0 & m \dot{\theta}_i \cos \theta_i \\ 0 & 0 & m \dot{\theta}_i \sin \theta_i \\ 0 & 0 & 0 \end{bmatrix}.$$

Physical parameters and their values of the OMR are shown in Table 1. (x_i, y_i) and θ_i denote the positions and the orientation angle of the i th OMR in the x and y directions in the ground coordinate system, respectively. $q_i = [x_i, y_i, \theta_i]^T \in R^3$ represents the position and orientation angle of the i th robot in the world

FIGURE 2: The communication topology of the systems. The root node 0 denotes the virtual leader.

coordinate frame, $\dot{q}_i = [\dot{x}_i, \dot{y}_i, \dot{\theta}_i]^T \in R^3$ represents the linear velocity and angular velocity of the i th robot in the world coordinate frame, and $\ddot{q}_i = [\ddot{x}_i, \ddot{y}_i, \ddot{\theta}_i]^T \in R^3$ represents the linear acceleration and angular acceleration of the i th robot in the world coordinate frame.

We assume the initial position states of each OMR are randomly chosen within $[-5, 5] \times [-5, 5]$, the initial angle is 0, and the initial velocity is $[0, 0, 0]^T$. $\delta_1 = (-2, 0, 0)^T$, $\delta_2 = [1, \sqrt{3}, 0]^T$, and $\delta_3 = [1, -\sqrt{3}, 0]^T$. The comparison simulations are carried out by the proposed adaptive method of LESO-SMC parameters with and without Q-learning algorithm, respectively, denoted by 'Q-LESO-SMC' and 'LESO-SMC', respectively. The parameters of LESO-SMC are chosen as $\omega_0 = 2$, $c_1 = 6$, $c_2 = 7$, $\mu = 4$, and the parameters of Q-LESO-SMC are obtained after online optimization by the Q-learning algorithm in Section 3.4. The other parameters used in the simulation are chosen as $\alpha = 0.1$, $\gamma = 0.8$, $\sigma = 0.0001$, $\varepsilon = 0.2$.

In order to observe the performance of the two different controllers, the robot formation is implemented in two cases, including the case with constant disturbances and the case with sinusoidal disturbances.

Case A. The case with constant disturbances.

Consider the robot formation subject to constant external disturbances, denoted by step functions. The corresponding results are shown in Figures 3–9.

The estimation behaviours of the LESO in two dimensions for Case A with constant disturbances are shown in Figures 3–5. Here $\tilde{x}_{i,1x}$, $\tilde{x}_{i,2x}$, and $\tilde{x}_{i,3x}$ ($\tilde{x}_{i,1y}$, $\tilde{x}_{i,2y}$, $\tilde{x}_{i,3y}$) denote the position, velocity, and the extended state (total disturbances) estimation in x -direction (in y direction) for the i th agent, $i = 1, 2, 3$, respectively. In each figure, the error plots are locally enlarged to highlight the specific convergence time and steady errors. From the comparison results, the estimation errors of all three states converge to a small stable region, and the estimation error of Q-LESO has faster convergence time and smaller steady estimation error than that of LESO under the constant disturbances.

Figure 6 shows the trajectories of three mobile robots and the virtual leader. $\star$ and $\circ$ represent the end positions of each follower robot and the virtual leader, respectively. It can be seen from the figure that the robots form a preset formation with the virtual leader as the center.

In order to make the comparison of control performance between LESO-SMC and Q-LESO-SMC, the formation tracking performance of each follower under these two controllers is depicted in Figures 7–8. Here $e_{i,1x}$ and $e_{i,2x}$ ($e_{i,1y}$ and $e_{i,2y}$) denote the formation tracking position error

TABLE 1: Model parameters of the OMR.

Parameters	Definitions	Values
m	Mass of the OMR	4.5 kg
I_v	Moment of inertia	0.25 kg*m ²
d	Distance between the center of mass of the OMR and the intersection point of the symmetry axis of the OMR and the wheel axis	0.2 m
r	Wheel radius	0.0635 m
b	The distance between an actuated wheel and the symmetry axis	0.166 m

FIGURE 3: The position estimation error $\tilde{x}_{i,1}$ in Case A. (a) The position estimation error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The position estimation error of the y -direction of the i th agent $i = 1, 2, 3$.

and velocity error in x -direction (in y direction) for the i th agent, $i=1,2,3$, respectively. It can be seen that the convergence speed of formation tracking error with Q-LESO-SMC is faster than that with LESO-SMC method, and Q-LESO-SMC method has stronger ability to suppress disturbance by comparing the magnitudes of the steady-state errors.

The position trajectories for three OMRs are shown in Figure 9, where $\circ$ represents the starting position of each

robot. It can be seen that 3 followers quickly track the virtual leader under the proposed Q-LESO-SMC method.

Case B. The case with sinusoidal disturbances.

For simplicity of simulation, consider the robot formation subject to nonlinear external disturbances, such as sine functions, e.g., $f_i = [0.5e^{-0.02t}\cos(t), 0.2\sin(t), 0.5\cos(t)]^T$. The corresponding results are shown in Figures 10–16.

FIGURE 4: The velocity estimation error $\tilde{x}_{i,2}$ in Case A. (a) The velocity estimation error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The velocity estimation error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 5: Continued.

FIGURE 5: The extended state estimation error $\tilde{x}_{i,3}$ in Case A. (a) The disturbance estimation error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The disturbance estimation error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 6: The formation trajectories of three followers and the virtual leader in Case A.

FIGURE 7: Continued.

FIGURE 7: The formation tracking position error $e_{i,1}$ in Case A. (a) The formation tracking position error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The formation tracking position error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 8: The formation tracking velocity error $e_{i,2}$ in Case A. (a) The formation tracking velocity error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The formation tracking velocity error of the y -direction of the i th agent $i = 1, 2, 3$.

Figures 10–12 show the comparison of estimation performance between Q-LESO and LESO with sinusoidal disturbances. The comparison of disturbance estimation

behaviour is shown in Figure 12, where the ‘total disturbances’ of each follower are approximated by Q-LESO in a shorter period. In addition, the amplitude of the

FIGURE 9: The tracking trajectories of the three followers and the virtual leader in Case A. (a) The tracking trajectories of the 1st follower for the virtual leader. (b) The tracking trajectories of the 2nd follower for the virtual leader. (c) The tracking trajectories of the 3rd follower for the virtual leader.

FIGURE 10: Continued.

FIGURE 10: The position estimation error $\tilde{x}_{i,1}$ in Case B(a) The position estimation error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The position estimation error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 11: The velocity estimation error $\tilde{x}_{i,2}$ in Case B(a) The velocity estimation error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The velocity estimation error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 12: The disturbance estimation error $\tilde{x}_{i,3}$ in Case B(a) The disturbance estimation error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The disturbance estimation error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 13: The formation trajectories of three followers and the virtual leader in Case B.

steady state of the estimation error of Q-LESO is smaller than that of LESO as can be seen in the locally enlarged plot.

Similar to Figure 6, Figure 13 shows the trajectories of the three mobile robots and the virtual leader in the same way. As can be seen from the figure, the target triangular

FIGURE 14: The formation tracking position error $e_{i,1}$ in Case B(a) The formation tracking position error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The formation tracking position error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 15: Continued.

FIGURE 15: The formation tracking velocity error $e_{i,2}$ in Case B(a) The formation tracking velocity error of the x -direction of the i th agent $i = 1, 2, 3$. (b) The formation tracking velocity error of the y -direction of the i th agent $i = 1, 2, 3$.

FIGURE 16: The tracking trajectories of the three followers and the virtual leader in Case B(a) The tracking trajectories of the 1st follower for the virtual leader. (b) The tracking trajectories of the 2nd follower for the virtual leader. (c) The tracking trajectories of the 3rd follower for the virtual leader.

formation is achieved with the virtual leader at the center.

Figures 14–15 show the comparison of formation tracking error between Q-LESO-SMC and LESO-SMC with sinusoidal disturbances. As can be seen, the followers track the virtual leader faster with the Q-LESO-SMC-based formation

controller. The error steady-state part is locally enlarged, and comparing the magnitude of the steady-state error before and after optimization shows that the error amplitude of Q-LESO-SMC is smaller than that of LESO-SMC.

Similar to Figure 9 in Case A, the position trajectories for three OMRs are shown in Figure 16. It can be seen that 3

followers quickly track the virtual leader under the proposed Q-ESO-SMC method, and the tracking error is smaller than that under the LESO-SMC method.

Remark 5. In Case B, both LESO and Q-LESO methods can achieve well formation control performance. The comparison results show that the Q-LESO exhibited the advantages of faster convergence rate, smaller tracking error, and better disturbance suppression performance. In addition, to get good performance, the LESO needs larger observer gains, which may go beyond bandwidths of practical systems and make the required control energy infeasible. However, higher observer gains may lead to the bigger overshoot (the so-called peaking phenomenon); see Figures 10–12. Therefore, it is a trade-off between cost and performance.

5. Conclusions

In this paper, considering the uncertainties and external disturbances in the formation process, an LESO-SMC with Q-learning adaptive optimization are proposed to achieve formation tracking of multiple OMRs. The simulation results show that the proposed control method has the advantages of faster convergence rate, higher tracking accuracy, and good steady-state performance. However, the LESO brings large overshoot with increasing bandwidth. We will further investigate this problem in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Jiangsu Natural Science Foundation of PR China (Grant no. BK20171019), the National Natural Science Foundation of China (Grant no. 61703175), and the Natural Science Foundation of Huizhou University (Grant no. hzu201806).

References

- [1] W. Ren and R. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*, Springer, New York, NY, USA, 2008.
- [2] B. Tucker and R. Arkin, "Behavior-based formation control for multirobot teams," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 6, pp. 926–939, 1998.
- [3] C. Bai, P. Yan, W. Pan, and J. Guo, "Learning-based multi-robot formation control with obstacle avoidance," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [4] J. H. Acosta, J. P. Jiménez, C. C. Hernández, and R. M. Clark, "Leader-follower formation control of wheeled mobile robots without attitude measurements," *Applied Sciences*, vol. 11, no. 22, p. 5639, 2021.
- [5] B. Park and S. Yoo, "Connectivity-maintaining and collision-avoiding performance function approach for robust leader-follower formation control of multiple uncertain underactuated surface vessels," *Automatica*, vol. 127, Article ID 109501, 2021.
- [6] J. Lin, Z. Miao, H. Zhong, W. Peng, Y. Wang, and R. Fierro, "Adaptive image-based leader-follower formation control of mobile robots with visibility constraints," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6010–6019, 2020.
- [7] L. Cao, H. Li, and N. Wang, "Observer-based event-triggered adaptive decentralized fuzzy control for nonlinear large-scale systems," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 6, pp. 1201–1214, 2018.
- [8] E. Almasri and M. K. Uyguroğlu, "Modeling and trajectory planning optimization for the symmetrical multiwheeled omnidirectional mobile robot," *Symmetry*, vol. 13, no. 6, p. 1033, 2021.
- [9] N. Lashkari, M. Biglarbegan, and S. X. Yang, "Development of a novel robust control method for formation of heterogeneous multiple mobile robots with autonomous docking capability," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 4, pp. 1759–1776, 2020.
- [10] X. Yu, J. Ma, N. Ding, and A. Zhang, "Cooperative target enclosing control of multiple mobile robots subject to input disturbances," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3440–3449, 2021.
- [11] R. Siegwart, I. Nourbakhs, and D. Scaramuzza, *Introduction to Autonomous mobile Robots*, MIT press, Cambridge, MA, USA, 2011.
- [12] J. Ghommam, M. Saad, F. Mnif, and M. Faisal, "Finite-time circular formation around a moving target with multiple underactuated ODIN vehicles," *Mathematics and Computers in Simulation*, vol. 180, pp. 230–250, 2021.
- [13] Y.-W. Choi, J.-W. Choi, S.-G. Im, D. Qian, and S.-G. Lee, "Multi-robot avoidance control based on omni-directional visual SLAM with a fisheye lens camera," *International Journal of Precision Engineering and Manufacturing*, vol. 19, no. 10, pp. 1467–1476, 2018.
- [14] G. Zhang, Y. Wang, J. Wang, J. Chen, and D. Qian, "Disturbance observer-based super-twisting sliding mode control for formation tracking of multi-agent mobile robots," *Measurement and Control*, vol. 53, no. 5-6, pp. 908–921, 2020.
- [15] B. M. Yousuf, A. S. Khan, and A. Noor, "Multi-agent tracking of non-holonomic mobile robots via non-singular terminal sliding mode control," *Robotica*, vol. 38, no. 11, pp. 1984–2000, 2020.
- [16] R. Zhou and T. Wen, "Analysis and tuning of general linear active disturbance rejection controllers," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 7, pp. 5497–5507, 2018.
- [17] J. Baek, M. Jin, and S. Han, "A new adaptive sliding-mode control scheme for application to robot manipulators," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 6, pp. 3628–3637, 2016.
- [18] G. Bartolini, A. Ferrara, and E. Usai, "Chattering avoidance by second-order sliding mode control," *IEEE Transactions on Automatic Control*, vol. 43, no. 2, pp. 241–246, 1998.
- [19] B. He, G. Liu, H. Chen, and X. Hu, "Extended state observer-based sliding mode learning control for mechanical system," *Measurement and Control*, vol. 53, no. 7-8, pp. 1395–1403, 2020.
- [20] Y. Zhao, Y. Zhang, and J. Lee, "Lyapunov and sliding mode based leader-follower formation control for multiple mobile robots with an augmented distance-angle strategy," *International Journal of Control, Automation and Systems*, vol. 17, no. 5, pp. 1314–1321, 2019.

- [21] Y. Hui and S. Wang, "Trajectory tracking control for non-holonomic wheeled mobile robots with external disturbances and parameter uncertainties," *International Journal of Control, Automation and Systems*, vol. 18, no. 12, pp. 3015–3022, 2020.
- [22] I. Jmel, H. Dimassi, S. H. Said, and F. M'Sahil, "Adaptive observer-based sliding mode control for a two-wheeled self-balancing robot under terrain inclination and disturbances," *Mathematical Problems in Engineering*, vol. 2021, 15 pages, 2021.
- [23] J. Han, "From PID to active disturbance rejection control," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 3, pp. 900–906, 2009.
- [24] Y. Guo, L. Yu, and J. Xu, "Robust finite-time trajectory tracking control of wheeled mobile robots with parametric uncertainties and disturbances," *Journal of Systems Science and Complexity*, vol. 32, no. 5, pp. 1358–1374, 2019.
- [25] J. Chen, N. Du, and Y. Han, "Decoupling attitude control of a hypersonic glide vehicle based on a nonlinear extended state observer," *International Journal of Aerospace Engineering*, vol. 2020, 11 pages, 2020.
- [26] Y. Guo, L. Yu, and J. M. Xu, "Anti-windup adaptive sliding mode trajectory tracking control of wheeled mobile robot based on extended state observer," *Journal of Systems Science and Mathematical Sciences*, vol. 37, no. 5, pp. 1179–1193, 2017.
- [27] Z. Gao, "Scaling and bandwidth-parameterization based controller tuning," in *Proceedings of the 2003 American Control Conference*, pp. 4989–4996, Denver, CO, USA, June 2003.
- [28] Y. Huang and W. Xue, "Active disturbance rejection control: methodology and theoretical analysis," *ISA Transactions*, vol. 53, no. 4, pp. 963–976, 2014.
- [29] Z. Zheng and L. Sun, "Path following control for marine surface vessel with uncertainties and input saturation," *Neurocomputing*, vol. 177, pp. 158–167, 2016.
- [30] K. Chang, Y. Xia, K. Huang, and D. Ma, "Obstacle avoidance and active disturbance rejection control for a quadrotor," *Neurocomputing*, vol. 190, pp. 60–69, 2016.
- [31] H. Yang, S. Wang, Z. Zuo, and P. Li, "Trajectory tracking for a wheeled mobile robot with an omnidirectional wheel on uneven ground," *IET Control Theory & Applications*, vol. 14, no. 7, pp. 921–929, 2020.
- [32] S. Richard and G. Andrew, *Reinforcement Learning: An Introduction*, MIT press, Cambridge, MA, USA, 2018.
- [33] A. E. Hakim, H. Hindersah, and E. Rijanto, "Application of reinforcement learning on self-tuning PID controller for soccer robot multi-agent system," in *Proceedings of the 2013 Joint International Conference on Rural Information and Communication Technology and Electric-Vehicle Technology rICT and ICEV-T*, pp. 1–6, Bandung, Indonesia, November 2013.
- [34] V. Aghaei, A. Onat, and I. Eksin, "Fuzzy PID controller design using Q-learning algorithm with a manipulated reward function," in *Proceedings of the European Control Conference (ECC)*, pp. 2502–2507, IEEE, Linz, Austria, July, 2015.
- [35] Z. Li, W. Liu, L. Li, L. Guo, and W. Zhang, "Modeling and adaptive controlling of cable-drogue docking system for autonomous underwater vehicles," *International Journal of Adaptive Control and Signal Processing*, 2021.
- [36] L. Beckenbach, P. Osinenko, and S. Streif, "A Q-learning predictive control scheme with guaranteed stability," *European Journal of Control*, vol. 56, pp. 167–178, 2020.
- [37] Z. Chen, B. Qin, M. Sun, and Q. Sun, "Q-learning-based parameters adaptive algorithm for active disturbance rejection control and its application to ship course control," *Neurocomputing*, vol. 408, pp. 51–63, 2020.
- [38] J. Hu and Y. S. Lin, "Consensus control for multi-agent systems with double-integrator dynamics and time delays," *IET Control Theory & Applications*, vol. 4, no. 1, pp. 109–118, 2010.
- [39] Z. Qing, L. Gao, and Z. Gao, "On stability analysis of active disturbance rejection control for nonlinear time-varying plants with unknown dynamics," in *Proceedings of the 46th IEEE conference on decision and control*, pp. 3501–3506, New Orleans, LA, USA, December, 2007.
- [40] Z. Yuan, Y. Tian, Y. Yin, S. Wang, J. Liu, and L. Wu, "Trajectory tracking control of a four mecanum wheeled mobile platform: an extended state observer-based sliding mode approach," *IET Control Theory & Applications*, vol. 14, no. 3, pp. 415–426, 2019.
- [41] J. Rath, M. Defoort, H. Karimi, and K. Veluvolu, "Output feedback active suspension control with higher order terminal sliding mode," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1392–1403, 2016.