Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3115

ANTI-INTERFERENCE TRAJECTORY TRACKING CONTROL OF QUADROTOR UAV BASED ON REINFORCEMENT LEARNING

PingAn Ren[1], EnHui Ren[2], YiFan Qu[3*]
[1]*School of Intelligent Engineering, Xiangtan Institute of Technology, Xiangtan 411100, Hunan, China.*
[2]*School of Computer Science and Technology, Zhoukou Normal University, Zhoukou 466001, Henan, China.*
[3]*School of Automation, Harbin University of Science and Technology, Harbin 150080, Heilongjiang, China.*
[*]*Corresponding Author: YiFan Qu*

Abstract: This study presents a quadrotor UAV tracking control algorithm that addresses the issues of parameter uncertainty and external disturbances during trajectory tracking. The algorithm combines the Q-Learning reinforcement learning algorithm with a nonsingular terminal sliding mode controller. Firstly, a four-rotor UAV model based on tracking error is defined, and the coupling and external interference between channels are converted into lumped interference. Extended state observers are designed for estimation and compensation of lumped interference by the outer loop position subsystem and the inner loop attitude subsystem. At the same time, a fast non-singular terminal sliding mode UAV controller is constructed, which includes an outer loop position controller and an inner loop attitude controller. Then, a Q-learning algorithm based on fuzzy strategy is proposed to realize the adaptive adjustment of the key parameters of the controller and the observer. By designing the reward function, the Q values of the UAV under different flight states are iteratively optimized and the Q table is constantly updated. Finally, the trained Q table is used for drone control. The algorithm can not only save the complicated process of manual parameter adjustment, but also realize the adaptive adjustment of key parameters in the face of different flight environments and flight states. The simulation and comparison experiments show that the proposed algorithm has a higher degree of fit with the reference trajectory in the trajectory tracking control process and has good robustness.
Keywords: Quadrotor UAV; Trajectory tracking control; Fast nonsingular terminal sliding mode; Reinforcement learning; Extended state observer

1 INTRODUCTION

Four-rotor UAVs are widely used in both military and civilian fields due to their low cost, vertical take-off capability, and high flexibility [1-2]. However, their highly nonlinear and underactuated nature makes precise trajectory tracking challenging under external disturbances, parameter variations, and unmodeled dynamics [3-4].

Various control methods have been developed to address these issues. PID-based controllers are simple but lack robustness against strong disturbances [5-6]. Sliding mode control (SMC) offers better performance in nonlinear and uncertain systems: fuzzy SMC improves adaptability [7], while non-singular terminal SMC reduces chattering but with slower convergence [8]. Predictive SMC with disturbance observers further enhances robustness [9]. Active disturbance rejection control (ADRC) has also been widely adopted for its ability to actively suppress disturbances [10], with improvements such as fixed-time SMC and extended state observers (ESO) providing faster convergence and stronger anti-disturbance capabilities [11-12]. Nevertheless, most methods still rely heavily on manual parameter tuning, which is experience-dependent and condition-sensitive [13-14].

Reinforcement learning (RL), especially Q-learning, offers a promising way to achieve adaptive control through autonomous learning [15-16]. In this paper, a Q-learning-based RL algorithm is designed to optimize fast non-singular terminal sliding mode control (FNTSMC) for trajectory tracking. The main contributions are: An error-based FNTSMC is developed for the outer-loop position and inner-loop attitude subsystems, enabling faster and more accurate convergence than conventional methods [8]; Coupling and external disturbances are treated as lumped disturbances and compensated via an extended state observer, significantly improving anti-disturbance performance compared to existing designs [9]; Controller and observer parameters are adaptively tuned using Q-learning, which reduces reliance on prior knowledge, avoids local optima, and is easy to implement [12].

2 FOUR-ROTOR UAV MODELING AND PROBLEM DESCRIPTION

Quadrotor UAVs are underactuated nonlinear systems controlled by four motor speeds. The model is established using ground and body coordinate systems, with their transformation shown in Figure 1.

Figure 1 Coordinate Conversion and Rotor Structure of Four-Rotor UAV

Since UAV modeling has been well established in literature [5], this paper focuses on controller design without repeating the modeling process. The following assumptions are made to simplify the dynamic modeling and controller design: The body is rigid, symmetrical, and its geometric center coincides with the center of mass. The quadrotor cancels the gyroscopic effect through the counter-rotation of adjacent motors. Disturbances in the system are bounded, slowly varying, and satisfy $\lim_{t\to\infty}\dot{d}_i=0, i=1,2,\cdots,6$.

The UAV dynamics model is established by Newton-Euler equation [9-10]:

$$\begin{aligned}\ddot{x}&=-\frac{U_1}{m}(\cos\psi\sin\theta\cos\phi+\sin\psi\sin\phi)\\&\quad-\frac{K_1\dot{x}}{m}+d_1\\\ddot{y}&=-\frac{U_1}{m}(\sin\psi\sin\theta\cos\phi-\cos\psi\sin\phi)\\&\quad-\frac{K_2\dot{y}}{m}+d_2\end{aligned}\tag{1}$$

$$\begin{aligned}\ddot{z}&=-\frac{U_1}{m}(\cos\theta\cos\phi)-g-\frac{K_3\dot{z}}{m}+d_3\\\ddot{\phi}&=\frac{l}{I_x}(lU_2+(I_y-I_z)\dot{\theta}\dot{\psi})-\frac{lK_4}{I_x}\dot{\phi}+d_4\\\ddot{\theta}&=\frac{l}{I_y}(lU_3+(I_z-I_x)\dot{\phi}\dot{\psi})-\frac{lK_5}{I_x}\dot{\theta}+d_5\\\ddot{\psi}&=\frac{1}{I_z}(U_4+(I_x-I_y)\dot{\theta}\dot{\phi})-\frac{K_6}{I_x}\dot{\psi}+d_6\end{aligned}\tag{2}$$

Formula (1) represents the position subsystem, and formula (2) represents the attitude subsystem.x, yandz are the body position states of the UAV.ϕ Roll Angle, θ pitch Angle, ψ yaw Angle are the three Euler angles of UAV respectively.U_1, U_2, U_3 andU_4 are the control input torques of the 4 brushless motors respectively. mis the mass of the drone. g is the gravitational acceleration. I_x, I_y, I_z are the moment of inertia of the X_b, Y_b, Z_b axis respectively.l is the distance from the center of the propeller to the center of gravity of the UAV. K_i is the coefficient of air resistance. d_i is the external disturbance of the system, where$i\in\{1,2,3,4,5,6\}$.

Since the UAV has only four actual control inputs, the amount of control is less than the system state variable. To achieve the tracking of expected x_r, y_r and z_r, a virtual control quantity is introduced into the position subsystem of formula (1) :

$$\begin{aligned}\&u_x&=U_1(\cos\psi\sin\theta\cos\phi+\sin\psi\sin\phi)/m\\\&u_y&=U_1(\sin\psi\sin\theta\cos\phi-\cos\psi\sin\phi)/m\\\&u_z&=U_1(\cos\theta\cos(\phi)/m-g\end{aligned}\tag{3}$$

In this paper, given ψ_r,the expected attitude Angleϕ_r, θ_r and expected lift forceF_r obtained by solving equation (3) are:

$$\begin{aligned}\&\phi_r&=\arcsin\left(\frac{m}{u_{\cos\psi_r u_x\sin\psi_r y}F}\right)\\\&\theta_r&=\arctan\left(\frac{u_x\cos\psi_r+}{u_{\sin\psi_r y}u_z+g}\right)\\\&F_r&=m\sqrt{u_x^2+u_y^2+(u_z+g)^2}\}\end{aligned}\tag{4}$$

From formula (1) to formula (4), it can be seen that the UAV can track the target trajectory by changing the position and attitude Angle. Based on this, the UAV tracking problem considering internal parameter uncertainty and external interference can be described as follows: Design a Q-learning reinforcement learning and sliding mode control algorithm to make the system state variable $[x,y,z,\psi]^T$ asymptotically converge to the desired signal $[x_r,y_r,z_r,\psi_r]^T$ in a finite time.

3 CONTROL ALGORITHM DESIGN AND STABILITY ANALYSIS

The UAV trajectory tracking control algorithm is composed of many parts. In this section, we first designed an extended state observer for position and attitude loops respectively to estimate lumped interference, and then designed an error-based fast non-singular terminal sliding mode controller for position and attitude subsystems respectively according to the obtained total disturbance estimates. Finally, Q-learning reinforcement learning algorithm was designed to realize adaptive adjustment of key parameters of the controller and observer. Realize the active disturbance rejection tracking of UAV in complex environment. The control structure is shown in Figure 2.

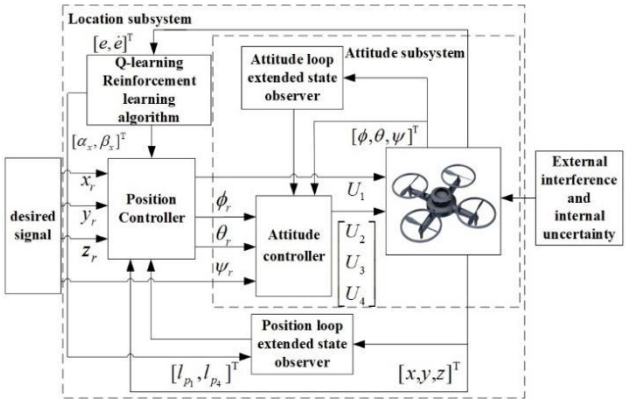


Figure 2 Four-Rotor UAV Control Structure

3.1 Design of Position and Attitude Extended State Observer

Firstly, the 3D position tracking error and attitude Angle tracking error of UAV are defined as:

$$\begin{bmatrix} \&e_x \\ \&e_y \\ \&e_z \end{bmatrix} = \begin{bmatrix} \&x-x_r \\ \&y-y_r \\ \&z-z_r \end{bmatrix} \tag{5}$$

$$e_\Theta=\Theta-\Theta_r=[\phi-\phi_r,\theta-\theta_r,\psi-\psi_r]^T \tag{6}$$

where,$\Theta=[\phi,\theta,\psi]^T$is the attitude Angle,$\Theta_r=[\phi_r,\theta_r,\psi_r]^T$is the desired attitude Angle.

According to formula (1) and (3), the position error model is obtained by bringing into formula (5) :

$$\begin{aligned} \ddot{e}_x&=u_x-\frac{K_1\dot{x}}{m}+d_1-\ddot{x}_r \\ \ddot{e}_y&=u_y-\frac{K_2\dot{y}}{m}+d_2-\ddot{y}_r \\ \ddot{e}_z&=u_z-\frac{K_3\dot{z}}{m}+d_3-\ddot{z}_r \end{aligned} \tag{7}$$

According to equations (2) and (6), the attitude Angle tracking error model is as follows:

$$\ddot{e}_\Theta=B_i u_i^1+f_i(\Theta)+D_i \tag{8}$$

where, $B_i=[l/I_x,l/I_y,1/I_z]^T$, $u_i^1=[U_2,U_3,U_4]^T$, $f_i(\Theta)=\begin{bmatrix} \&f_1 \\ \&f_2 \\ \&f_3 \end{bmatrix}=\begin{bmatrix} \&((I_y-I_z)\dot{\theta}\dot{\psi}-lK_4\dot{\phi})/I_x \\ \&((I_z-I_x)\dot{\phi}\dot{\psi}-lK_5\dot{\theta})/I_y \\ \&((I_x-I_y)\dot{\theta}\dot{\phi}-lK_5\dot{\psi})/I_z \end{bmatrix}$, $D_i=\begin{bmatrix} \&D_1 \\ \&D_2 \\ \&D_3 \end{bmatrix}=\begin{bmatrix} \&d_4-\ddot{\phi}_r \\ \&d_5-\ddot{\theta}_r \\ \&d_6-\ddot{\psi}_r \end{bmatrix}$, D_i is the lumped interference in attitude loop.

In order to estimate the external interference d_1 ,d_2 and d_3 of the position tracking error model (7), the extended state observer of the position subsystem is designed:

$$\begin{aligned} \&\dot{v}_{x_1}&=-\frac{K_1}{m}\dot{x}-\ddot{x}_r+u_x+v_{x_2}-l_{p_1}(v_{x_1}-\dot{e}_x) \\ \&\dot{v}_{x_2}&=-l_{p_2}(v_{x_1}-\dot{e}_x) \\ \&\hat{d}_1&=v_{x_2} \\ \&\dot{v}_{y_1}&=-\frac{K_2}{m}\dot{y}-\ddot{y}_r+u_y+v_{y_2}-l_{p_3}(v_{y_1}-\dot{e}_y) \\ \&\dot{v}_{y_2}&=-l_{p_4}(v_{x_1}-\dot{e}_y) \\ \&\hat{d}_2&=v_{y_2} \\ \&\dot{v}_{z_1}&=-\frac{K_3}{m}\dot{z}-\ddot{z}_r+u_z+v_{z_2}-l_{p5}(v_{z_1}-\dot{e}_z) \\ \&\dot{v}_{z_2}&=-l_{p6}(v_{z_1}-\dot{e}_z) \\ \&\hat{d}_3&=v_{z_2} \end{aligned} \tag{9}$$

Where,v_{x_1} , v_{x_2}, v_{y_1}, v_{y_2}, v_{z_1} , v_{z_2} are the state variable of the position extended state observer,l_{p_1}, l_{p_2}, l_{p_3}, l_{p_4}, l_{p_5}, l_{p_6} are the position extended state observer gain,$\hat{d}_1$, $\hat{d}_2$, $\hat{d}_3$ are the estimates of external interference.

In order to estimate the lumped disturbance D_1, D_2 and D_3 ,of the attitude subsystem, the following extended state observer for the attitude Angle tracking error model (8) is designed:

$$\begin{aligned}
\dot{v}_{\phi 1}&=k_{p1}\dot{e}_{\phi}-k_{p1}v_{\phi 1}+v_{\phi 2}+B_1u_1+f_1\\
\dot{v}_{\phi 2}&=k_{p2}\dot{e}_{\phi}-k_{p2}v_{\phi 1}\\
\dot{\hat{D}}_1&=v_{\phi 2}\\
\dot{v}_{\theta 1}&=k_{p3}\dot{e}_{\theta}-k_{p3}v_{\theta 1}+v_{\theta 2}+B_2u_2+f_2\\
\dot{v}_{\theta 2}&=k_{p4}\dot{e}_{\theta}-k_{p4}v_{\theta 1}\\
\dot{\hat{D}}_2&=v_{\theta 2}\\
\dot{v}_{\psi 1}&=k_{p5}\dot{e}_{\psi}-k_{p5}v_{\psi 1}+v_{\psi 2}+B_3u_3+f_3\\
\dot{v}_{\psi 2}&=k_{p6}\dot{e}_{\psi}-k_{p6}v_{\psi 1}\\
\dot{\hat{D}}_3&=v_{\psi 2}
\end{aligned} \tag{10}$$

Where, v_{ϕ_1} , v_{ϕ_2}, v_{θ_1}, v_{θ_2}, v_{ψ_1} , $v_{\psi 2}$ are the state variable of the attitude extended state observer, k_{p_1}, k_{p_2}, k_{p_3}, k_{p_4}, k_{p_5}, k_{p_6} are the observer gain, $\hat{D}_1$, $\hat{D}_2$, $\hat{D}_3$ are the estimates of the lumped interference of the attitude subsystem.

3.2 Fast Non-Singular Terminal Sliding Mode Controller Design

The traditional sliding mode control cannot make the system error converge in a finite time, which obviously cannot meet the control requirements of the UAV system. Therefore, the terminal sliding mode control is proposed based on it, which solves the problem that the system error can only converge gradually. In order to further solve the problem of possible strange phenomena, the following non-singular terminal sliding mode surface is designed:

$$S_i=e_i+\frac{1}{\beta_i}\dot{e}_i^{p_i/q_i},i=x,y,z,\phi,\theta,\psi \tag{11}$$

where, $\beta_i>0$,p_i, $q_i$are positive odd numbers.

It can be seen from the above formula that the non-singular terminal sliding mode surface is composed of a nonlinear function, the existence of which improves the velocity of the system error in the approach stage. When the system error is approaching, the convergence rate is faster. After the system error reaches the sliding mode surface, the convergence rate of the non-singular terminal sliding mode surface composed of nonlinear functions is slower than that of the linear function. Therefore, a fast term is added to ensure that the non-singular terminal sliding mode control strategy maintains a faster global speed during convergence:

$$S_i=e_i+\frac{1}{\alpha_i}e_i^{g_i/h_i}+\frac{1}{\beta_i}\dot{e}_i^{p_i/q_i},i=x,y,z,\phi,\theta,\psi \tag{12}$$

Where,$\alpha_i>0$， $\beta_i>0$,g_i, h_i, p_i, q_i are positive odd numbers, the following conditions are met:$\frac{p_i}{q_i}<\frac{g_i}{h_i}$， $1<\frac{p_i}{q_i}<2$.

To facilitate writing, the virtual control quantity of the attitude subsystem is redefined $u_\phi=B_1u_1^1$, $u_\theta=B_2u_2^1$, $u_\psi=B_3u_3^1$.

Theorem 1 According to the position subsystem tracking error model (7) and attitude Angle tracking error model (8), the following fast non-singular terminal sliding mode controller is designed:

$$u_i=-\beta_i\cdot\frac{q_i}{p_i}\left[\dot{e}_i^{2-p_i/q_i}\left(1+\frac{1}{\alpha_i}\cdot\frac{g_i}{h_i}\cdot e_i^{g_i/h_i-1}\right)\right]+M_i-k_{i1}\cdot S_i^{m_i/n_i}-k_{i2}\cdot\ \mathrm{sign}(S_i) \tag{13}$$

The tracking errors e_x, e_y, e_z, e_ϕ, e_θ, e_ψ of the position subsystem and the attitude subsystem can be guaranteed to converge in finite time, where:

$$M_i=\begin{bmatrix}
\dfrac{K_1\dot{x}}{m}-\hat{d}_1+\ddot{x}_r\\
\dfrac{K_2\dot{y}}{m}-\hat{d}_2+\ddot{y}_r\\
\dfrac{K_3\dot{z}}{m}-\hat{d}_3+\ddot{z}_r\\
-f_1-\hat{D}_1\\
-f_2-\hat{D}_2\\
-f_3-\hat{D}_3
\end{bmatrix},i=x,y,z,\phi,\theta,\psi \tag{14}$$

m_i is positive odd number, satisfy $\frac{m_i}{n_i}\geq 1$; k_{i1} , k_{i2} are positive real numbers, satisfy$k_{i1}>k_{i2}$,$k_{x2}>\left|d_1-\hat{d}_1\right|_{max}$,$k_{y2}>\left|d_2-\hat{d}_2\right|_{max}$,$k_{z2}>\left|d_3-\hat{d}_3\right|_{max}$,$k_{\phi 2}>\left|D_1-\hat{D}_1\right|_{max}$,$k_{\theta 2}>\left|D_2-\hat{D}_2\right|_{max}$,$k_{\psi 2}>\left|D_3-\hat{D}_3\right|_{max}$.

3.3 Stability Analysis of Closed Loop System

If theorem 1 is proved, the attitude Angle tracking error model equation (8) is introduced from the fast non-singular terminal sliding mode controller model equation (13), and the control input torques U_1, U_2, U_3 and U_4 of the four brushless motors can be obtained after deduction to control the motor speed, which can ensure that the tracking error of the UAV converges to 0 in a finite time. That is, the state variable$[x,y,z,\psi]^T$converges to the desired signal $[x_r,y_r,z_r,\psi_r]^T$. Since the controller designed in this paper has the same structure in the position control loop and the attitude control

loop, the stability of other channel controllers can be shown by proving the stability of any channel of any control loop controller. Taking the ψ channel of UAV yaw Angle as an example, the stability of the controller designed in this paper is proved.

(1) It is proved that the tracking error of the system can reach the sliding mode surface in a limited time.

The controller (13) is substituted into the attitude Angle tracking error model (8), and the closed-loop system error dynamics are obtained:

$$\ddot{e}_\psi=-\frac{\beta_\psi q_\psi}{p_\psi}\left[\dot{e}_\psi^{2-p_\psi/q_\psi}\left(1+\frac{g_\psi e_\psi^{g_\psi/h_\psi-1}}{\alpha_\psi h_\psi}\right)\right]-k_{\psi1}\cdot S_\psi^{m_\psi/n_\psi}-k_{\psi2}\cdot \operatorname{sign}(S_\psi)+D_3-\hat{D}_3 \tag{15}$$

select the Lyapunov function for S_ψ:

$$V_\psi=\frac{1}{2}S_\psi^2 \tag{16}$$

take the derivative of V_ψ:

$$\dot{V}_\psi=-\frac{S_\psi p_\psi}{\beta_\psi q_\psi}\dot{e}_\psi^{p_\psi/q_\psi-1}\left[k_{\psi2}\cdot \operatorname{sign}(S_\psi)-D_3+\hat{D}_3\right]-\left(\frac{k_{\psi1}p_\psi}{\beta_\psi q_\psi}S_\psi^{m_\psi/n_\psi+1}\right)\dot{e}_\psi^{p_\psi/q_\psi-1} \tag{17}$$

where, according to formula (12), p_ψ, q_ψ are positive odd and $1<\frac{p_\psi}{q_\psi}<2$, according to formula (13), m_ψ and n_ψ also are positive odd, $\frac{k_{\psi1}p_\psi}{\beta_\psi q_\psi}S_\psi^{m_\psi/n_\psi+1}>0, \dot{e}_\psi^{p_\psi/q_\psi-1}>0$ can be obtained. From equation (12), it can be obtained:

$$\dot{V}_\psi\le-\frac{S_\psi p_\psi}{\beta_\psi q_\psi}\dot{e}_\psi^{p_\psi/q_\psi-1}\left[k_{\psi2}\cdot \operatorname{sign}(S_\psi)-D_3+\hat{D}_3\right] \tag{18}$$

from $k_{\psi2}>\left|D_3-\hat{D}_3\right|_{max}$ and $k_{\psi1}>k_{\psi2}$, in combination with formula (18), it can be obtained:;

$$\dot{V}_\psi\le-\frac{\sqrt{2}(k_{\psi1}-k_{\psi2})p_\psi}{\beta_\psi q_\psi}\dot{e}_\psi^{p_\psi/q_\psi-1}\cdot V_\psi^{1/2} \tag{19}$$

the normal number of ε is satisfied:

$$\frac{\sqrt{2}(k_{\psi1}-k_{\psi2})p_\psi}{\beta_\psi q_\psi}\dot{e}_\psi^{p_\psi/q_\psi-1}>\varepsilon \tag{20}$$

from equation (18), it can be obtained:

$$\dot{V}_\psi\le-\varepsilon V_\psi^{1/2} \tag{21}$$

this proves that V_ψ can converge in a finite time, and thus S_ψ can converge to 0 in a finite time.

(2) It is proved that the tracking error converges in finite time after reaching the sliding mode surface.

Assuming that when $t=t_s$, the tracking error of ψ control loop reaches the sliding mode surface, then when $t>t_s$, the tracking error of ψ control loop dynamically changes from equation (12) to:

$$\dot{e}_\psi=-e_\psi^{q_\psi/p_\psi}\cdot \beta_\psi^{q_\psi/p_\psi}\left(1+\frac{1}{\alpha_\psi}e_\psi^{g_\psi/h_\psi-1}\right)^{q_\psi/p_\psi} \tag{22}$$

from the above formula:

$$\left(\frac{1}{e_\psi}\right)^{q_\psi/p_\psi}de_\psi=-\beta_\psi^{q_\psi/p_\psi}\left(1+\frac{1}{\alpha_\psi}e_\psi^{g_\psi/h_\psi-1}\right)^{q_\psi/p_\psi}dt \tag{23}$$

if we integrate both sides of equation (23), we get:

$$\int_{e_\psi(t_s)}^{e_\psi(t_s+\Delta t)}\left(\frac{1}{e_\psi}\right)^{q_\psi/p_\psi}de_\psi=-\int_{t_s}^{t_s+\Delta t}\beta_\psi^{q_\psi/p_\psi}\left(1+\frac{1}{\alpha_\psi}e_\psi^{g_\psi/h_\psi-1}\right)^{q_\psi/p_\psi}dt\le-\int_{t_s}^{t_s+\Delta t}\beta_\psi^{q_\psi/p_\psi}dt \tag{24}$$

from the above formula:

$$e_\psi(t_s+\Delta t)^{1-q_\psi/p_\psi}\le e_\psi(t_s)^{1-q_\psi/p_\psi}-\frac{p_\psi-q_\psi}{p_\psi}\beta_\psi^{q_\psi/p_\psi}\Delta t \tag{25}$$

where, $1-q_\psi/p_\psi>0, e_\psi(t_s+\Delta t)^{(p_\psi-q_\psi)/p_\psi}$ decreases monotonously with Δt. When $e_\psi(t_s+t_w)^{(p_\psi-q_\psi)/p_\psi}=0$, time t_w satisfies:

$$t_w\le\frac{p_\psi\cdot (t_s)^{1-q_\psi/p_\psi}}{p_\psi-q_\psi}e_\psi\left(\frac{1}{\beta_\psi}\right)^{q_\psi/p_\psi} \tag{26}$$

that is, ψ channel tracking error e_ψ can converge to 0 in a finite time.

To sum up, the stability of other channel controllers can be obtained according to the same proof process, which will not be analyzed one by one in this paper. The controller (12) can ensure the finite time convergence of the tracking error of the input instruction.

3.4 Q-Learning Algorithm Design Based on Fuzzy Strategy

This paper adopts a Q-learning reinforcement learning strategy to adaptively adjust the parameters of the fast non-singular terminal sliding mode (FNTSM) controller and the extended state observer (ESO), enabling the controller to self-learn and improve performance in varying environments.

The design focuses on the position loop, as it is more vulnerable to external disturbances, to enhance trajectory tracking in complex scenarios. Using the x-channel as an example, the same applies to the y and z channels. Position error e and its change rate \dot{e} are defined for the x-channel.

3.4.1 Basic model of reinforcement learning

The reinforcement learning algorithm operates without an explicit environmental model. Through interaction with the environment, it collects state information, evaluates actions using reward functions, and continuously improves the state -action mapping strategy. A Q-table is generated and updated iteratively during training, as illustrated in Figure 3.

3.4.2 The state space and parameter domain are defined based on fuzzy policy

Q-learning reinforcement learning algorithm can only learn discrete variables, so it needs to discretized the state space and parameter domain, divide the state space according to fuzzy control strategy, take$m_1=n_1=7$, A total of 7 levels of {NB,NM,NL,Z,PL,PM,PB} were used to divide the position error e and the error change rate ė of the UAV, so as to obtain different state Spaces of $m_1 \cdot n_1=49$ and define the range of each level, as shown in Table 1.

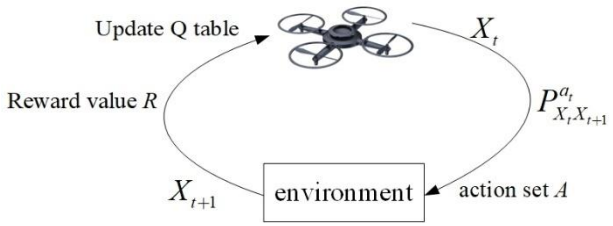


Figure 3 The Basic Framework of Reinforcement Learning

Table 1 State Interval Division Table

	e	ė
NB	[-2，-1)	[-200，-100)
NM	[-1，-0.4)	[-100，-50)
NL	[-0.4，-0.1)	[-50，-10)
Z	[-0.1，0.1]	[-10，10]
PL	(0.1，0.4]	(10，50]
PM	(0.4，1]	(50，100]
PB	(1，2]	(100，200]

Select the main controller parameters α_x, β_x , and select the extended state observer parameters l_{p1}, l_{p4}. According to debugging experience, select a reasonable parameter range and set$\alpha_x \in [1,7]$, $\beta_x \in [1,7]$, $l_{p_1} \in [10,30]$, $l_{p_4} \in [10,30]$. If $p_1=p_2=7$, $p_3=p_4=20$ are the number of selected actions, the number of available actions in action set$A=[\alpha_x,\beta_x,l_{p_1},l_{p_4}]$ is $p_1 \cdot p_2 \cdot p_3 \cdot p_4=19600$. Therefore, the Q table is a matrix of dimension 19600×49 and the expression of the Q table is:

$$Q=\begin{bmatrix} Q(X_1,a_1) & Q(X_2,a_1) & \cdots & Q(X_{A_M},a_1) \\ Q(X_1,a_2) & Q(X_2,a_2) & \cdots & Q(X_{A_M},a_2) \\ \vdots & \vdots & \ddots & \vdots \\ Q(X_1,a_{49}) & Q(X_2,a_{49}) & \cdots & Q(X_{A_M},a_{49}) \end{bmatrix}_{A_M\times 49} \tag{27}$$

where, M=19600.

3.4.3 Q-learning Algorithm learning and updating process

The position error e and error change rate ė of the current UAV control system are taken as the state X_t in the reinforcement learning process, and the combination of main controller parameters and extended state observer parameters is selected as the action set A . The reward function R, attenuation factor γ, and learning rate η are designed. The specific steps of Q-learning reinforcement learning algorithm are as follows:

Step 1: Algorithm initialization, including initializa-tion of UAV parameter information and Q-learning related parameters.

Step 2: Define the domain of position error e and error rate ė of change. The domains of e andė are represented by E_1 and E_2, respectively.

Step 3: Initializes the state space and action set A. Reasonable choice of action interval,$A_i \in [Aimax_{imin}$, p_i values were uniformly selected in these four groups, i=1,2,3,4.

$$A_{it}=A_{i\min}+(t-1)\frac{A_{i\max}-A_{i\min}}{p_i},t=1,2,\cdots p_i \tag{28}$$

Step 4: Design state transition matrix $P^{a_t}_{X_tX_{t+1}}$. Select actions based on the state transition probability, and use the ε−greedy to get the state transition probability:

$$\pi\left(a\middle|X\right)=\begin{cases}\dfrac{\varepsilon}{A(X)+1-\varepsilon}, & a=\arg\max_a Q(X,a)\\ \dfrac{\varepsilon}{A(X)}, & else\end{cases} \tag{29}$$

Step 5: Perform action a in the current state to obtain a new stateXand rewardR. The performance functionJis used to design the reward functionR, to enhance the rationality of Q table. The reward function is designed as follows:

$$J=\sum_{\tau=t}^{\tau=t+20} e^2(\tau)/20$$
$$R=\begin{cases}\&15, \&J<\rho\\ \&1, \&J=\rho\\ -15, \ J>\rho\end{cases} \tag{30}$$

where，ρ defines the parameters for the set reward.

Step 6: Update value function. The value function is designed as follows:

$$Q(X_t,A_t)=Q(X_t,A_t)+\eta[\gamma\max_a Q(X_{t+1},a)+R_{t+1}-Q(X_t,A_t)] \tag{31}$$

Step 7: When the termination condition is reached, the learning ends; Otherwise, return to Step 4. The training termination conditions are as follows:

The current training ends when the control process reaches A stable state,$|e|<0.001\&\&|\dot{e}|<0.01$.

The position error in the control is too large, and it is not meaningful to continue to iterate according to this state, so the current training is ended when$|e|>2$.

The number of iterative learning training of Q-learning is designed to be N. When N =2000, the learning ends and the trained Q-table can be obtained, which can be used for UAV trajectory tracking control. Q table is used to select the optimal action in the current state for control.

Simulation experiment

In order to verify the control performance of the algorithm proposed in this paper, this section uses the data of a small UAV model to conduct simulation experiments of the corresponding algorithm through computer simulation software, and sets the fixed parameter values required by the simulation experiments.

The main parameters of the position loop controller are adjusted adaptively by reinforcement learning algorithm, and the other parameters are consistent with the attitude loop. The system initial values are set as follows: $[x,y,z]^T=[2,2,0]^T$,$[\phi,\theta,\psi]^T=[0,0,0]^T$.

The expected trajectory is set as follows:$x_r=\sin(0.5t)$，$y_r=0.5\cos(0.5t)$，$z_r=1.5+0.2t$ $\psi_r=60^\circ\sin(0.5t)$.

In order to get close to the real flight environment and simulate the turbulent wind field, time-varying wind field and gust wind field contrary to the flight direction under the actual work scene, external disturbances are set. The external disturbances in the position subsystem are set as a function of the mass of the quadrotor UAV, so that the external disturbances it receives are at the same order of magnitude as the UAV. The external disturbances in the position subsystem are as follows:

$$d_i=\begin{cases}\&1.2m, & 0\le t<10\\ \&0.5m+m\cdot\sin(0.5t), & 10\le t<20\\ \&-1.5m, & t\ge 20\end{cases} \tag{32}$$

where, i=1,2,3.

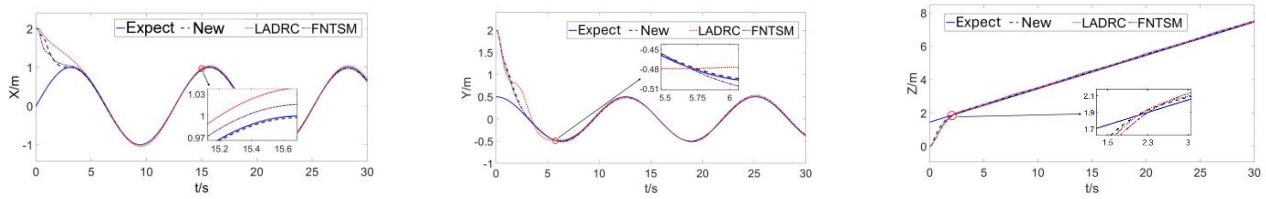
The external disturbance Settings of the attitude subsystem are as follows:

$$d_\lambda=\begin{cases}\&0.4, & 0\le t<10\\ \&0.5+0.5\sin(0.5t), & 10\le t<20\\ \&-0.6, & t\ge 20\end{cases} \tag{33}$$

where, λ=4,5,6.

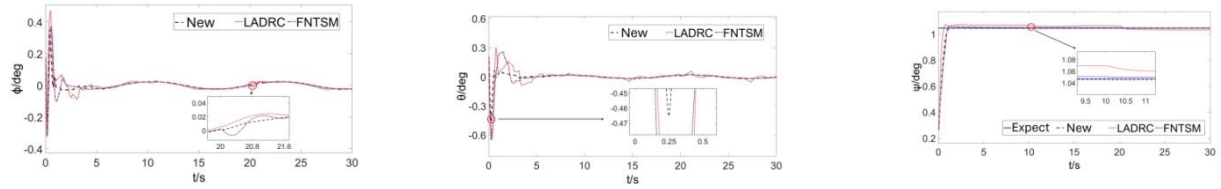
To validate the algorithm, comparisons were made with fixed-parameter FNTSM and LADRC. The proposed method demonstrates superior trajectory tracking in both position (Figure 4) and yaw angle (Figure 5(c)), with smaller steady-state error. While LADRC shows overshoot, our approach achieves faster, overshoot-free convergence, maintaining tracking error below 0.01m.

For roll (ϕ) and pitch (θ) angles, which should remain minimal during flight, the proposed algorithm provides quicker stabilization with less overshoot and better disturbance rejection compared to FNTSM and LADRC (Figure 5(a)-(b)). Overall, the Q-learning enhanced FNTSM controller outperforms both conventional LADRC and fixed-parameter sliding mode control.



(a)X-Direction Trajectory Tracking Response (b)Y-Direction Trajectory Tracking Response (c)Z-Direction Trajectory Tracking Response

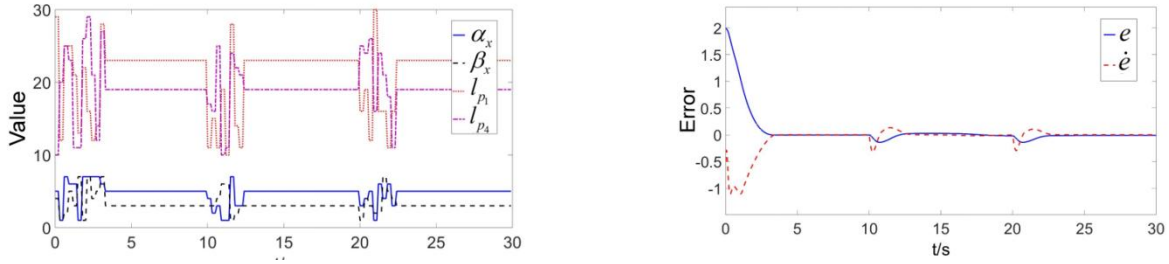
Figure 4 Position Control System Trajectory Tracking Response



(a) Roll Angle (b) Pitch Angle (c) Yaw Angle

Figure 5 Attitude Control System Trajectory Tracking Response

Figure 6 shows the adaptive parameter adjustment of the Q-learning enhanced FNTSM controller based on tracking error (e) and error rate (ė). After training, the algorithm uses a Q-table to select optimal control parameters in real-time. The figure shows that parameters adjust actively when e and ė are large, and remain stable once e and ė converge near zero. This demonstrates the controller's capability for autonomous parameter tuning and improved control performance.



(a) Adaptive Adjustment of Main Control Parameters (b) Tracking Error and Tracking Error Derivative

Figure 6 The Control Parameter Adaptively Adjusts with the Tracking Error

4 CONCLUSION

Aiming at the situation of parameter uncertainty and external interference of quadrotor UAV, this design adopts the control idea of reinforcement learning, and proposes a UAV control method that integrates Q-learning reinforcement learning with fast non-singular terminal sliding mode control. Q-learning is used to train controller and observer parameters. It makes up for the shortcomings of traditional sliding mode control, which is unable to adapt the parameters of controller and observer, and compares the simulation effect with the traditional method. The experimental results show that the method designed in this paper has excellent performance in trajectory tracking and anti-interference, and the response curve fits the reference trajectory to a high degree, which verifies the effectiveness of the algorithm. Next, in order to better adapt to complex flight scenarios and consider more practical factors, the algorithm can be further optimized in future work, and on this basis, optimization algorithms such as neural networks can be added to combine with deep learning methods to further improve the performance of the control algorithm.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Bennaceu S, Azouz N. Modelling and control of a quadrotor with flexible arms. Alexandria Engineering Journal, 2023, 65: 209-231.

[2] Xu Z, Fan L, Qiu W, et al. A robust disturbance rejection controller using model predictive control for quadrotor UAV in tracking aggressive trajectory. Drones, 2023, 7(6): 369.

[3] Yesmin A, Sinha A. Sliding mode controller for quadcopter UAVs: a comprehensive survey. Drones, 2025, 9(9): 625.

[4] Quan Q. Design and control of multi-rotor UAV. Beijing: Electronic Industry Press, 2017: 225-275.

[5] Zuo Z Y. Trajectory tracking control design with command-filtered compensation for quadrotor. IET Control Theory & Applications, 2010, 4(11): 2343-2355.
[6] Pounds P E I, Bersak D R, Dollar A M. Stability of small-scale UAV helicopters and quadrotors with added payload mass under PID control. Autonomous Robots, 2012, 33(1-2): 129-142.
[7] Wang C Y, Hu S B, Gao F H. Double fuzzy sliding mode control of four-rotor UAV with hanging load. Computer Era, 2021(12): 14-21.
[8] Zhao Z H, Li T, Jiang B. Compound continuous fast non-singular terminal sliding mode control for four-rotor UAV attitude system. Control Theory & Applications, 2023, 40(3): 459-467.
[9] Cheng X, Tang G, Wang P, et al. Predictive sliding mode control for attitude tracking of hypersonic vehicles using fuzzy disturbance observer. Mathematical Problems in Engineering, 2023, 2023(1): 1-8.
[10] Abadi El, Amraoui A, Mekki H, et al. Robust tracking control of quadrotor based on flatness and active disturbance rejection control. IET Control Theory & Applications, 2020, 14(8): 1057-1068.
[11] Lhayani M, Abbou A, El Houm Y. Trajectory tracking control of quadrotors using nonsingular fast fixed time sliding mode and fixed time extended state observer. International Journal of Intelligent Engineering and Systems, 2025, 18(10): 123-134.
[12] Lin Z, Li P. Nonsingular fast terminal sliding mode attitude control for a quadrotor based on fuzzy extended state observer//Proceedings of the 33rd Chinese Control and Decision Conference. Nanchang: IEEE, 2019: 1717-1723.
[13] Nikitin D A. Large scale systems control. Automation and Remote Control, 2021, 80(9): 1717-1733.
[14] Yoo J, Jang D, Kim H J, et al. Hybrid reinforcement learning control for a micro quadrotor flight. IEEE Control Systems Letters, 2021, 5: 155-182.
[15] Li M, Cai Z, Zhao J, et al. Disturbance rejection and high dynamic quadrotor control based on reinforcement learning and supervised learning. Neural Computing and Applications, 2012, 34(13): 1141-1161.
[16] Zhang Z, Yang H, Fei Y, et al. Control of UAV quadrotor using reinforcement learning and robust controller. IET Control Theory & Applications, 2021, 15(10): 1599-1615.