Received: 19 October 2022 | Revised: 9 May 2023 | Accepted: 17 May 2023

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12252

ORIGINAL RESEARCH

Car-following strategy of intelligent connected vehicle using extended disturbance observer adjusted by reinforcement learning

Ruidong Yan[1] | **Penghui Li**[1] | **Hongbo Gao**[2] | **Jin Huang**[3] | **Chengbo Wang**[4]

[1]School of Traffic and Transportation, The Beijing Jiaotong University, Beijing, China

[2]Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei, China

[3]School of Vehicle and Mobility, The Tsinghua University, Beijing, China

[4]Liverpool Logistics, Offshore and Marine Research Institute (LOOM), Liverpool John Moores University, Liverpool, UK

Correspondence

Hongbo Gao.
Email: ghb48@ustc.edu.cn

Funding information

State Key Laboratory of Automotive Safety and Energy, Grant/Award Number: KFY2208; National Natural Science Foundation of China, Grant/Award Numbers: U2013601, U20A20225; Key Research and Development Plan of Anhui Province, Grant/Award Number: 202004a05020058

Abstract

Disturbance observer-based control method has achieved good results in the car-following scenario of intelligent and connected vehicle (ICV). However, the gain of conventional extended disturbance observer (EDO)-based control method is usually set manually rather than adjusted adaptively according to real time traffic conditions, thus declining the car-following performance. To solve this problem, a car-following strategy of ICV using EDO adjusted by reinforcement learning is proposed. Different from the conventional method, the gain of proposed strategy can be adjusted by reinforcement learning to improve its estimation accuracy. Since the "equivalent disturbance" can be compensated by EDO to a great extent, the disturbance rejection ability of the car-following method will be improved significantly. Both Lyapunov approach and numerical simulations are carried out to verify the effectiveness of the proposed method.

KEYWORDS

adaptive system, autonomous vehicle, intelligent control

1 | INTRODUCTION

Intelligent and connected vehicle (ICV) is expected to revolutionise road transportation and is an inevitable trend in future [1, 2]. Car-following is one of the basic traffic scenarios. Up to now, various control methods have been applied to car-following scenario [3–5], but the car-following performance is limited by control theory itself in complex traffic environments. Different from the conventional control theory based methods, the disturbance observer-based control (DOBC) has achieved good results in the car-following scenario of ICV.

Because the external disturbance and parameter uncertainty. Widely exist in car-following system, it will decrease the control performance. If we know the accurate value of disturbance, the influence of the disturbance can be compensated. However, the accurate value of disturbance is usually difficult to measure directly. In order to achieve this goal, researchers have independently developed different types of anti-disturbance methods [6]. For convenience, these methods are named as DOBC. The basic idea of DOBC is to use the disturbance estimation value to compensate the effects of disturbance, and use the conventional controller to ensure the stability of the closed-loop system, so as to achieve the desired goal [7].

Disturbance observer is the key of DOBC. The conventional non-linear disturbance observer is introduced firstly. A type of non-linear disturbance observer was proposed to estimate the disturbance [8, 9], where the estimated value of the

disturbance asymptotically approaches the value of real disturbance. Then, the non-linear gain is applied to non-linear disturbance observer, and the analysis shows that the non-linear gain can improve the performance significantly. Next, it is pointed out that when the first derivative of the disturbance is not zero [10], there is a bounded estimation error in non-linear disturbance observer, and an appropriate gain can be selected to make the error sufficiently small.

Besides the conventional non-linear disturbance observer, the high-order disturbance observer is also concerned. The integral term was introduced to construct a higher-order disturbance observer, where the higher-order derivative of the disturbance is assumed to be zero [11]. For systems considering both measurement noise and inner disturbance, the integral observer and proportional integral observer were designed by using multiple integrals [12]. Then, the non-linear equation was introduced into the higher-order linear extended state observer [13]. Subsequently, the first-order non-linear disturbance observer was extended to the high-order one [14–16], further improving the accuracy of disturbance estimation. In addition, some sliding mode disturbance observers also improve the estimation accuracy by introducing higher-order item [17, 18].

However, the coefficients of the disturbance observer are usually set manually based on experience. The coefficients cannot be adaptively adjusted with the real-time traffic situation. This conventional observer based method will reduce the following performance in the complex traffic environment. At the same time, more and more artificial intelligence methods, represented by neural network [19] and reinforcement learning [20–22], have been introduced into the research of car-following scenario. This is because in the training process, the reinforcement learning algorithm can maximise the cumulative reward by interacting with the environment, so as to obtain the optimal strategy according to the changes of the environment [23].

To solve the above problem, reinforcement learning is applied to the car-following scenario under the framework of DOBC. In this paper, a car-following strategy is proposed for ICV. The main contribution of this paper is as follows:

(a) A type of car-following strategy is proposed theoretically by Lyapunov approach under the framework of DOBC, where EDO is introduced into car-following scenario to improve the car-following performance.
(b) Different from the conventional method, EDO gain of the proposed strategy is adaptively adjusted by reinforcement learning rather than set manually based on experience.
(c) Simulation results in terms of two scenarios show that the car-following performance of both first and second following vehicles are improved significantly.

The rest of this study is organised as follows. Section 2 is the vehicle model and problem formulation. The extended distributed observer design for ICV is presented in Section 3. Car-following control scheme is proposed in Section 4. Numerical simulations and conclusion are given in Sections 5 and 6.

2 | VEHICLE MODEL AND PROBLEM FORMULATION

In this section, the ICV model is given and then problem formulation is presented.

2.1 | Vehicle model

This study considers ICV using V2V communication technology without considering pedestrians. The leading vehicle will drive along a single lane and not affected by following vehicles, the leading vehicle is described as $x_0 = v_0 t + x_0(0)$, where $x_0(0)$, x_0, v_0 represent the initial position, position, and velocity respectively. According to ref. [24], the model of autonomous driving vehicle is given as follows:

$$\begin{cases} \dot{x}_i = y_i \\ \dot{v}_i = a_i \\ \dot{a}_i = -a_i/k + u_i + \omega_i \end{cases} \tag{1}$$

where x_i, v_i, a_i, u_i, and ω_i denote the position, velocity, acceleration, control torque, and external disturbance of the i th following vehicle at time t respectively. k is the dynamics parameter of autonomous driving vehicle.

2.2 | Problem formulation

Considering system model in terms of Equation (1), the objective of this study is to design a car-following scheme u_i of the ith following ICV to improve the car-following performance only using V2V communication among adjacent vehicles.

3 | EXTENDED DISTURBANCE OBSERVER DESIGN FOR INTELLIGENT CONNECTED VEHICLE

In this section, the extended disturbance observer is introduced into the design of car-following control scheme. It is designed to estimate the effects of the "equivalent disturbance" on the car-following system of ICV. Then observer gains are tuning by reinforcement learning.

3.1 | Extended disturbance observer design

Let $y_i = x_{i+1} - x_i - x_{veh} - x_{safe}$ denote the deviation between safe space x_{safe} and actual space of two adjacent vehicles. Let $y_v = v_d - v_i$ denote the deviation between the velocity of ego-vehicle and leading vehicle. Here, we chose a sliding surface as follows:

$$\begin{aligned} s_i &= y_i + c_1 y_v \\ &= y_i + c_1 (v_d - v_i) \end{aligned} \tag{2}$$

where c_1 is a positive controller gain.

The derivative of s_i can be derived as follows:

$$\dot{s}_i = \dot{y}_i + c_1 \dot{y}_v \tag{3}$$

Substituting Equation (1) and y_i into Equation (3) yields

$$\begin{aligned} \dot{s}_i &= v_{i+1} - v_i + c_1(a_0 - a_i) \\ &= v_{i+1} - v_i + c_1 a_0 + c_1 k\dot{a}_i - c_1 k u_i - c_1 k w_i \end{aligned} \tag{4}$$

Define $d_i = c_1 a_0 + c_1 k\dot{a}_i - c_1 k w_i$ as an "equivalent disturbance" and Equation (5) will be expressed as follows:

$$\dot{s}_i = v_{i+1} - v_i - c_1 k u_i + d_i \tag{5}$$

Then Equation (5) can be rewritten as follows:

$$d_i = \dot{s}_i - v_{i+1} + v_i + c_1 k u_i \tag{6}$$

According to Equation (6), the structure of d_i can be derived indirectly via the measurement of the previous frame of u_i, v_{i+1}, v_i, a_i and the calculation of $\dot{s}_i$. By defining $z_1 = d_i$, $z_2 = \dot{d}_i$, the "equivalent disturbance" can be rewritten as follows:

$$\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_2 = \ddot{d}_i \end{cases} \tag{7}$$

Furthermore, Equation (7) can be rewritten as follows by defining $A = [0,1; 0,0]$, $B = [0,1]^{\mathrm{T}}$, $z = [z_1,z_2]^{\mathrm{T}}$.

$$\dot{z} = Az + B\ddot{d}_i \tag{8}$$

Therefore, Equation (6) an be taken as the output of system Equation (8). By defining $C = [1,0]$, the output equation can be written as follows:

$$d_i = Cz \tag{9}$$

Since the pair (A, C) is observable, the extended disturbance observer for autonomous driving vehicle can be designed according to ref. [16] as follows by defining $\hat{d}_i = C\hat{z}$.

Theorem 1. *For system* Equations (8) and (9), *given an assumption that the second derivative of equivalent disturbance d_i is bounded and satisfies $\ddot{d}_i \le \delta_1$, EDO can be designed as follows:*

$$\begin{cases} \hat{z} = p + Ls_i \\ \dot{p} = (A - LC)p + ALs_i - L[CLs_i + v_{i+1} - v_i - c_1 k u_i] \end{cases} \tag{10}$$

Hence, the disturbance estimate error $\tilde{z} = z - \hat{z}$ can converge into a neighbourhood of the origin by choosing appropriate gains $L = [l_1, l_2]^{\mathrm{T}}$ and auxiliary variable p.

Proof. Substituting the second line of Equation (10) into the derivate of first line of Equation (10) we obtain

$$\dot{\hat{z}} = (A - LC)\hat{z} + AL[\dot{s}_i - v_{i+1} + v_i + c_1 k u_i] \tag{11}$$

Substituting Equations (6), (9) and $\hat{d}_i = C\hat{z}$ into Equation (11) yields

$$\dot{\hat{z}} = A\hat{z} + LC(z - \hat{z}) \tag{12}$$

By defining, Equation (8) minus Equation (12) yields

$$\dot{\tilde{z}} = (A - LC)\tilde{z} + B\ddot{d}_i \tag{13}$$

By choosing appropriate gains, $\overline{A} = A - LC$ will be Hurwitz since the pair of (A, C) is observable. Hence, given any matrix $Q > 0$, there exists a unique matrix $p > 0$ such that

$$P\overline{A} + \overline{A}^{\mathrm{T}}P = -Q \tag{14}$$

Choose Lyapunov function as

$$V_1 = \tilde{z}^T P \tilde{z} \tag{15}$$

Taking the time derivative of Equation (15) along Equations (13) and (14) yields

$$\begin{aligned} V_1 &= \dot{\tilde{z}}^T P\tilde{z} + \tilde{z}^T P\dot{\tilde{z}} = \tilde{z}^T\left(P\overline{A} + \overline{A}^T P\right)\tilde{z} + 2\tilde{z}^T PB\ddot{d}_i \\ &= -\tilde{z}^T Q\tilde{z} + 2\tilde{z}^T PB\ddot{d}_i \end{aligned} \tag{16}$$

Let λ_p and λ_q be maximum eigenvalue of P and minimum eigenvalue of Q respectively and yields

$$\begin{aligned} \dot{V}_1 &\le -\lambda_q ||\tilde{z}||_2^2 + 2||\tilde{z}||_2 ||P||_F ||B||_2 \left|\ddot{d}_i\right| \\ &\le -\lambda_q ||\tilde{z}||_2^2 + 2\lambda_q ||\tilde{z}||_2 \delta_1 \end{aligned} \tag{17}$$

where $|.|$, $\|.\|_2$, $\|.\|_F$ denote the absolute value of a variable, 2-norm of a vector and Frobenius norm of a matrix respectively.It can be derived that $\dot{V}_1 < 0$ if $||\tilde{z}||_2 > 2\lambda_p\delta_1/\lambda_q$. Then, the decrease of V_1 drives the system trajectory into $||\tilde{z}||_2 \le 2\lambda_p\delta_1/\lambda_q$. Thus, the system trajectory will converge into a bounded origin by choosing appropriate gains $L = [l_1, l_2]^{\mathrm{T}}$.

Remark 1. The gain of traditional EDO is set manually based on experience. However, the disturbance is always changing in traffic environments so that the fixed gain of EDO can not meet the requirements of the complex traffic scenarios, thus reducing the accuracy of disturbance estimation.

3.2 | Observer gain adjusted by reinforcement learning

For the proposed car-following scheme, EDO gain is adjusted by reinforcement learning. Reinforcement learning can

maximise the return through trial and error, so as to obtain an optimal strategy. Deep deterministic policy gradient (DDPG) is a typical reinforcement learning algorithm and suitable for continuous action space. Hence, DDPG algorithm in ref. [20] is used in this study to adjust EDO gains. For the convenience of using reinforcement learning, two car-following scenarios are chosen as shown in Figure 1.

Scenario 1: There are only one front vehicle and one following vehicle. Vehicle speed, position and other information are exchange between two vehicles through V2V communication. As shown in Figure 1, the car-following scheme proposed in this paper is adopted by the vehicle in the blue box, where EDO gain is updated in real time through reinforcement learning.

Markov decision process is modelled firstly. The action of DDPG is chosen as the gain of EDO of following vehicle. The state space is chosen as

$$\{y_i, \Delta v_i, y_i, \Delta v_i, v_i, a_i\} \tag{18}$$

where $y_i = x_i - x_0 - x_{veh} - x_{safe}$ is the space headway between following vehicle and front vehicle. x_{safe} and x_{veh} are safe space and vehicle length. $\Delta v_i = v_i - v_0$ is the velocity error between following vehicle and front vehicle. v_i and a_i are velocity and acceleration of following vehicle.

The reward function is chosen as follows:

$$r = r_{\Delta v_i} + r_{jerki} \tag{19}$$

$$r_{\Delta v_i} = -\omega_1 \frac{|\Delta v_i|}{v_{max}} \tag{20}$$

$$r_{jerk} = -\omega_2 \frac{|a_{i\_k} - a_{i\_k-1}|}{2a_{max}\,\Delta T} \tag{21}$$

where a_{i_k} is the acceleration of ith following vehicle at kth frame. $v_{\max}$, $a_{\max}$, ΔT are maximum velocity, maximum acceleration, time step respectively. $|.|$ represents the abstract value of a variable. ω_1 and ω_2 are positive coefficients.

Scenario 2: There are one front vehicle and two following vehicles. As shown in Figure 1, the car-following strategy proposed in this paper is adopted by following vehicles 1 and 2 in the red box, where EDO gain is updated in real time through reinforcement learning.

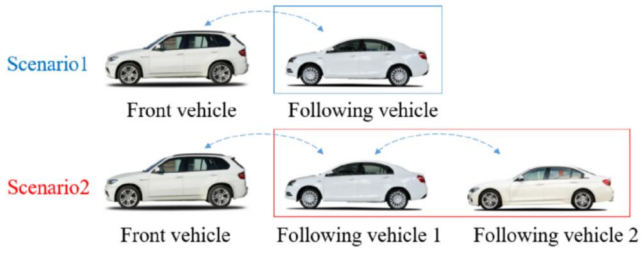

FIGURE 1 Two scenarios of CF scheme with EDO adjusted by reinforcement learning.

We construct Markov decision process for scenario 2. The action of DDPG is chosen as EDO gain for both two following vehicles at the same time. The state space is given as follows:

$$\{y_i, \Delta v_i, y_{i+1}, \Delta v_{i+1}, v_i, v_{i+1}\} \tag{22}$$

where $y_i = x_i - x_0 - x_{veh} - x_{safe}$ is the space headway between following vehicle 1 and front vehicle. $\Delta v_i = v_i - v_0$ is the velocity error between following vehicle 1 and front vehicle $y_{i+1} = x_{i+1} - x_0 - 2(x_{veh} - x_{safe})$ is the space headway between following vehicle 2 and front vehicle. $\Delta v_{i+1} = v_{i+1} - v_0$ is the velocity error between following vehicle2 and front vehicle. v_i and v_{i+1} a are velocity of following vehicles 1 and 2.

The reward function under scenario 2 is the sum of Equation (19) for following vehicles 1 and 2 as follows.

$$r = \sum_{i=1}^{2} r_{\Delta vi} + r_{jerki} \tag{23}$$

Remark 2. Different from the traditional EDO that adjusts gains through experience, the gains of EDO in this study are adjusted through reinforcement learning. The essence of reinforcement learning is to constantly try and error in the training process and obtain the strategy of maximum cumulative reward, so as to meet the requirements of disturbance adaptive estimation in complex traffic scenarios. Therefore, EDO gains that best fit the environment can be obtained in terms of different state spaces through reinforcement learning, so as to improve the accuracy of disturbance estimation.

Remark 3. When there are more than three vehicles, EDO gain will be adjusted by combining scenarios 1 and 2. For example, when there are four ICVs, we apply scenario 1 to vehicles 1 and 2, where vehicle 2 is manipulated by the proposed method. Meanwhile, we apply scenario 2 to vehicles 2, 3 and 4, where vehicles 3 and 4 are manipulated by the proposed method. Considering the risk of communication delay, the proposed car-following scheme only considers V2V communication among adjacent vehicles, so that it is only suitable for small-scale platoon.

4 | CAR-FOLLOWING STRATEGY USING EXTENDED DISTURBANCE OBSERVER

In this section, a car-following strategy using extended disturbance observer is proposed by combining extended disturbance observer and conventional sliding model control method as follows.

Theorem 2. *For system (1), the trajectory of system error will be guaranteed to converge into a neighbourhood of the origin by designing the following distributed controller with appropriate parameters.*

$$u_i = \frac{1}{c_1 k}\left[\left(v_{i+1} - v_i\right) + c_2 s_i + c_3 sign(s_i) + \hat{d}_i\right] \tag{24}$$

where u_i, v_i, k denote control torque, velocity, vehicle dynamics parameter; $\hat{d}$ is disturbance estimation in the form of Equation (10); s_i is in the form of Equation (4); c_1, c_2, c_3 are controller gains.

Proof. Substituting Equation (24) into Equation (5) we obtain

$$\dot{s}_i = -c_2 s_i - c_3 sign(s_i) + \hat{d}_i \tag{25}$$

Choose Lyapunov equation V_2 as

$$V_2 = \frac{1}{2}s_i^2 + V_1 \tag{26}$$

Taking the derivative of V_2 along *Theorem 1* and Equation (24) gives

$$\begin{aligned} \dot{V}_2 &= s_i \dot{s}_i + \dot{V}_1 \\ &\le s_i \dot{s}_i - \lambda_q \left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_q \left|\left|\tilde{z}\right|\right|_2 \delta_1 \end{aligned} \tag{27}$$

Substituting Equation (25) into Equation (27) yields

$$\begin{aligned} \dot{V}_2 &\le -c_2 |s_i|^2 - c_3 |s_i| + s_i \tilde{d}_i - \lambda_q \left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_q \left|\left|\tilde{z}\right|\right|_2 \delta_1 \\ &\le -c_2 |s_i|^2 - c_3 |s_i| + \left|\left|\tilde{z}\right|\right|_2 |s_i| - \lambda_q \left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_q \left|\left|\tilde{z}\right|\right|_2 \delta_1 \end{aligned} \tag{28}$$

Given $\delta_2 > 0$, $\left|\left|\tilde{z}\right|\right|_2 \le \delta_2$ can be obtained according to *Theorem 1*. Then Equation (28) can be rewritten as

$$\begin{aligned} \dot{V}_2 &\le -\left(\sqrt{c_2}\,|s_i| - \frac{1}{2\sqrt{c_2}}\left|\left|\tilde{z}\right|\right|_2\right)^2 - c_3 |s_i| + \frac{1}{4c_2}\left|\left|\tilde{z}\right|\right|_2^2 \\ &\quad -\lambda_q \left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_p \left|\left|\tilde{z}\right|\right|_2 \delta_1 \\ &\le -c_3 |s_i| - \left(\lambda_q - \frac{1}{4c_2}\right)\left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_p \left|\left|\tilde{z}\right|\right|_2 \delta_1 \\ &\le -c_3 |s_i| - \left(\lambda_q - \frac{1}{4c_2}\right)\left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_p \delta_1 \delta_2 \end{aligned} \tag{29}$$

Let $\alpha = \min\{c_3, (4c_2\lambda_q - 1)/4c_2\}$, $\beta = 2\lambda_p \delta_1 \delta_2$. Equation (29) can be rewritten as follows:

$$\begin{aligned} \dot{V}_2 &\le -c_3 \frac{1}{2}s_i^2 - \frac{4c_2\lambda_q - 1}{4c_2}\left|\left|\tilde{z}\right|\right|_2^2 + 2\lambda_p \delta_1 \delta_2 \\ &\le -c_3 V_2 + \beta \end{aligned} \tag{30}$$

Multiply both sides of Equation (30) by $e^{-\varepsilon t}$ and we obtain

$$e^{-\varepsilon t}\left(\dot{V}_2 + \alpha V_2\right) \le e^{-\varepsilon t}\beta \tag{31}$$

By integrating Equation (31) with the initial value $V_2(0)$ of $V_2(t)$ yields

$$V_2(t) \le [V_2(0) - \beta/\alpha]e^{-\varepsilon t} + \beta/\alpha \tag{32}$$

From Equation (32), it is known that $V_2(t)$ is bounded due to α, β, $V_2(0)$ are bounded. Hence, the trajectory of the closed-loop system will converge into a bounded region.

Remark 4 The bound of the region depends on the ratio β/α. Meanwhile, β and α are determined by parameters c_2, c_3, λp, λq, δ_1, δ_2. Due to $\ddot{d}_i \le \delta_1$ and $\left|\left|\tilde{z}\right|\right|_2 \le \delta_2$, both the disturbance characteristics and estimation error of EDO will effect the accuracy of the proposed distributed controller. According to *Theorem 1* and *Remark 2*, EDO gains $L = [l_1, l_2]^{\mathrm{T}}$ are adjusted by reinforcement learning in this study can improve the accuracy of disturbance estimation, so that it also improve the accuracy of the proposed distributed controller. This is because that "equivalent disturbance" can be compensated by EDO to a great extent, so that the disturbance rejection ability of the composite controller is improved significantly in the presence of complex traffic scenarios.

Remark 5 To avoid chattering problem, sign(x) is replaced with saturation function sat(x) = sign(x) × min{| x |/ς_1, 1}, $0 < \varsigma_1 < 1$. Hence, the proposed car-following controller in the form of Equation (24) can be rewritten as follows:

$$u_i = \frac{1}{c_1 k}\left[\left(v_{i+1} - v_i\right) + c_2 s_i + c_3 sat(s_i)\right] + \hat{d}_i \tag{33}$$

For Equation (33), $\hat{d}_i$ in the form of Equation (10) can be rewritten as follows:

$$\begin{cases} \hat{d}_i = l_1 s_i + p_1 \\ \dot{p}_1 = -l_1 \hat{d}_i - l_1(v_{i+1} - v_i - c_1 k u_i) + \hat{z}_2 \end{cases} \tag{34}$$

$$\begin{cases} \hat{z}_2 = l_2 s_i + p_2 \\ \dot{p}_2 = -l_2 \hat{d}_i - l_2(v_{i+1} - v_i - c_1 k u_i) \end{cases} \tag{35}$$

Furthermore, Equation (33) will decade into the conventional car-following controller without considering disturbance compensation as follows:

$$u_i = \frac{1}{c_1 k}\left[\left(v_{i+1} - v_i\right) + c_2 s_i + c_3 sat(s_i)\right] \tag{36}$$

5 | NUMERICAL SIMULATIONS

In order to verify the effectiveness of the proposed scheme, the simulations are carried out on the single lane without considering lane changing. Two scenarios proposed in the paper are considered. Each scenario selects three different time

intervals, and compares reward value, speed difference and acceleration in terms of statistics. For comparison, simulations are carried out through the conventional car following (CF) scheme and the proposed CF scheme under the same conditions.

We combine Python and Simulink for simulation. For the reinforcement learning training, the velocity of front vehicle is chosen from real vehicle experiment [25]. For the reinforcement learning testing, two scenarios in terms of six velocity segments of front vehicle were used.

The safe space and vehicle length are chosen as $x_{safe} = 6$m and $x_{veh} = 4$m. The initial conditions of the following vehicles are set as $y_i = x_{safe} + x_{veh}$ and $v_i(0) = v_0(0)$. Gaussian distributed noise is chosen as measurement noise. Time simulation step is chosen as 0.2s l_1 is adjusted by reinforcement learning and l_2 is 0.01. Parameters o DDPG are listed in Table 1 and change of them are often small. Controller parameters are listed in Table 2.

5.1 | Two vehicles under scenario 1

Figure 2 shows the velocity of front vehicle and following vehicle using two methods between 0 and 40s. The yellow solid line is the velocity of front vehicle. The velocity of following vehicle using conventional CF is the blue dotted line. The velocity of following vehicle using the proposed CF is the red dotted line. The front vehicle velocity fluctuates greatly during 5–30s. There is an overshoot by using the conventional CF. On the contrary, the overshoot is smaller by using the proposed CF. When the velocity of front vehicle fluctuates small, two methods have similar results. $\sum_k r_k(t)$ is the reward of following vehicle. $\sum_k |\Delta v_{i\_k}|$ is velocity error between front vehicle and following vehicle. $\sum_k |a_k(t)|$ is the acceleration of following vehicle. As shown in Table 3, the car-following performance of the proposed CF is better than that of the conventional CF during 0–40s.

As shown in Figure 3, the car-following performance of the proposed CF is much better than that of the conventional CF. When the velocity of front vehicle has a relatively large variation, the velocity of following vehicle via conventional CF has a large overshot while overshot via the proposed CF is

TABLE 1 Parameters of DDPG.

Parameter	Value
Actor learning rate	0.0001
Critic learning rate	0.001
Experience replay buffer size	500000
Batch size	32
τ	0.001
γ	0.99

TABLE 2 Parameters of CF.

Parameter	Value
c_1	2
c_2	0.8
c_3	0.1
ς_1	0.01
k	0.8
$a_{\max}$	2

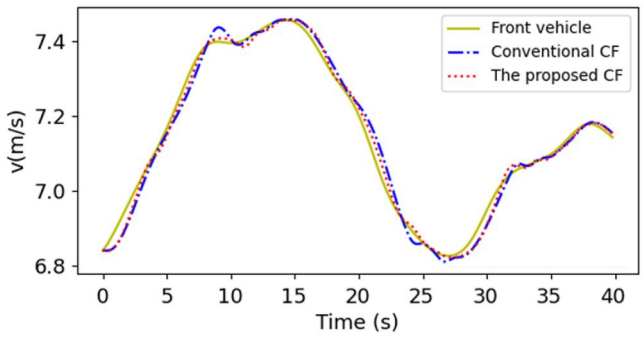

FIGURE 2 Velocity comparison during 0–40s under scenario 1.

TABLE 3 Comparison during 0–40s under scenario 1.

Item	Conventional CF	The proposed CF
$\sum_k r_{i\_k}(t)$	−3.8570	−3.4927
$\sum_k \lvert\Delta v_{i\_k}\rvert$ (m/s)	3.8261	2.8099
$\sum_k \lvert a_{i\_k}(t)\rvert (\mathrm{m/s^2})$	8.7825	8.3434

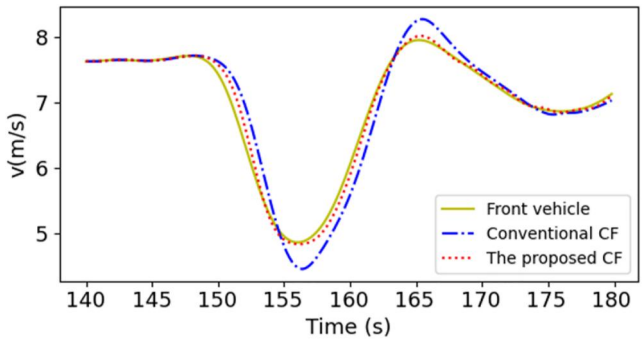

FIGURE 3 Velocity comparison during 140–180s under scenario 1.

TABLE 4 Comparison during 140–180s under scenario 1.

Item	Conventional CF	The proposed CF
$\sum_k r_{i\_k}(t)$	−14.9999	−6.5214
$\sum_k \lvert\Delta v_{i\_k}\rvert$ (m/s)	33.9340	10.3677
$\sum_k \lvert a_{i\_k}(t)\rvert (\mathrm{m/s^2})$	43.0888	37.5942

small. When there is almost no change in velocity of front vehicle during 140–145s, the conventional CF and proposed CF are similar. Table 4 shows that the performance of the proposed CF better than that of traditional CF during 140–180s.

As shown in Figure 4, the performance of the conventional CF and the proposed CF is similar during 340–355s. The maximum velocity of front vehicle is 10 m/s and minimum one is 3 m/s during 340–380s. Since the influence of velocity variation can be compensated by EDO, the proposed CF shows good car-following performance. Table 5 shows that the proposed CF offers better car-following performance than that of the conventional one during 340–380s.

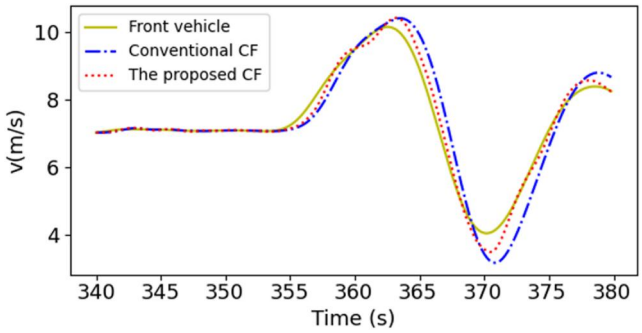

FIGURE 4 Velocity comparison during 340–380s under scenario 1.

TABLE 5 Comparison during 340–380s under scenario 1.

Item	Conventional CF	The proposed CF
$\sum_k r_{i\_k}(t)$	−25.9836	−13.2925
$\sum_k \lvert\Delta v_{i\_k}\rvert$ (m/s)	65.1875	29.5822
$\sum_k \lvert a_{i\_k}(t)\rvert$(m/s^2)	82.6754	79.9208

5.2 | Three vehicles under scenario 2

Figure 5 shows the velocity of one front vehicle and two following vehicles. Curves shown in Figure 5a are velocities of the first following vehicle using the conventional CF (blue doted line) and the proposed CF (red doted line). Figure 5b shows the comparison results of the second following vehicle. Similar to scenario 1, the conventional CF shows large overshot but the performance of the proposed CF is much better. $\sum_k r_{i\_k}(t)$ is the reward of two following vehicles. $\sum_i \sum_k \lvert\Delta v_{i\_k}\rvert$

TABLE 6 Comparison during 0–40s under scenario 2.

Item	Conventional CF	The proposed CF
$\sum_k r_{i\_k}(t)$	−8.1135	−7086
$\sum_i \sum_k \lvert\Delta v_{i\_k}\rvert$ (m/s)	8.7647	5.9180
$\sum_i \sum_k \lvert a_{i\_k}(t)\rvert$(m/s^2)	17.7031	16.5863

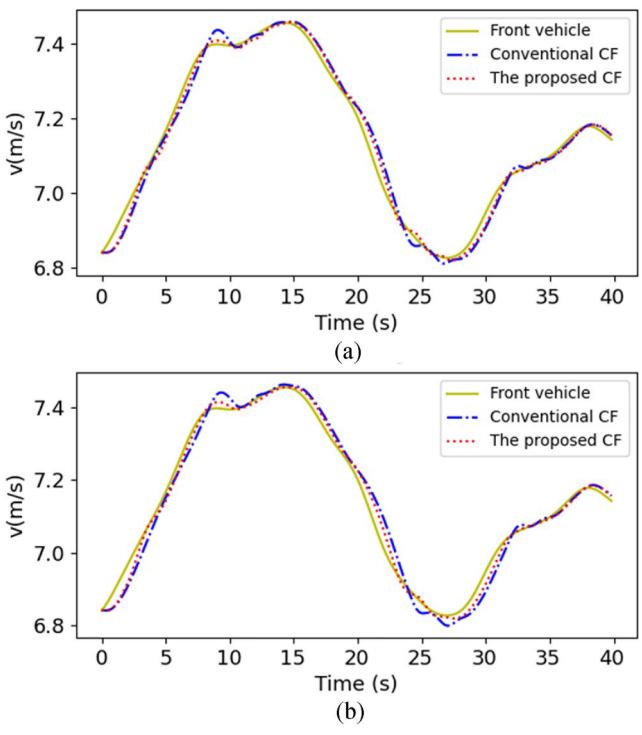

FIGURE 5 Velocity comparison during 0–40s under scenario 2. (a) First following vehicle. (b) Second following vehicle.

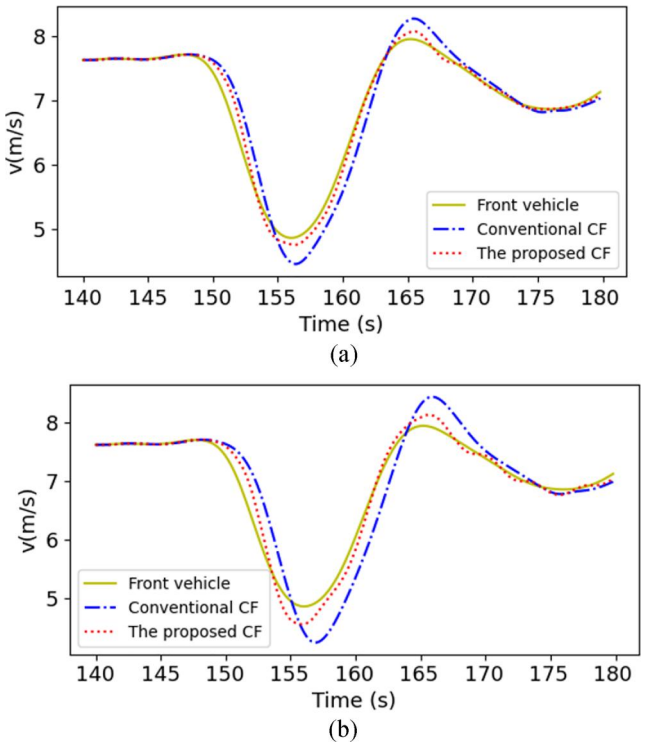

FIGURE 6 Velocity comparison during 140–180s under scenario 2. (a) First following vehicle. (b) Second following vehicle.

TABLE 7 Comparison during 140–180s under scenario 2.

Item	Conventional CF	The proposed CF
$\sum_k r_{i\_k}(t)$	−37.0577	−17.4786
$\sum_i \sum_k \lvert\Delta v_{i\_k}\rvert$ (m/s)	87.4507	33.0584
$\sum_i \sum_k \lvert a_{i\_k}(t)\rvert$(m/s^2)	92.2617	81.0890

TABLE 8 Comparison during 340–380s under scenario 2.

Item	Conventional CF	The proposed CF
$\sum_k r_{i\_k}(t)$	−65.3479	−33.3994
$\sum_i \sum_k \lvert\Delta v_{i\_k}\rvert$ (m/s)	167.7266	78.1936
$\sum_i \sum_k \lvert a_{i\_k}(t)\rvert$(m/s^2)	171.2841	161.5966

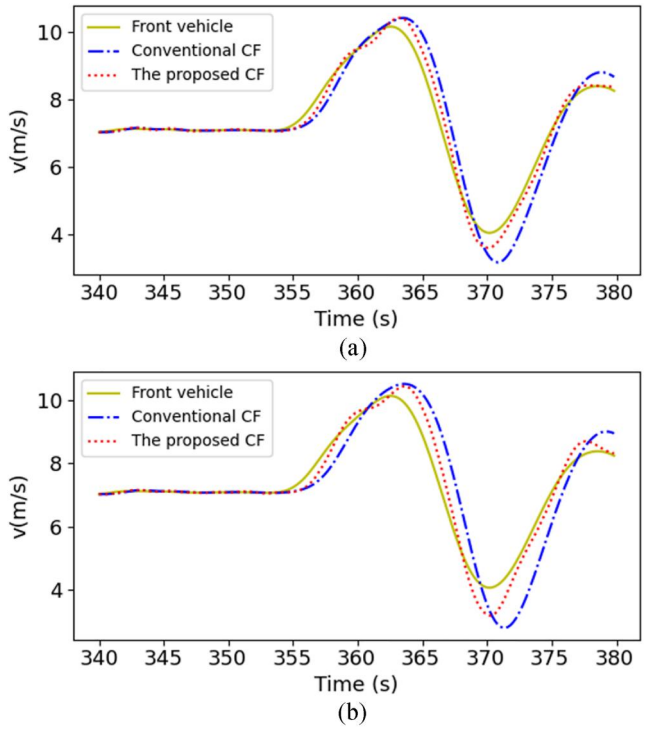

FIGURE 7 Velocity comparison during 340–380s under scenario 2. (a) First following vehicle. (b) Second following vehicle.

is the velocity error of two following vehicles. $\sum_i \sum_k \lvert a_{i\_k}(t)\rvert$ is the acceleration of two following vehicles. Table 6 shows that the proposed CF offers better car-following performance than that of the conventional one during 0–40s.

As shown in Figure 6a and Figure 6b, compared with the conventional CF, the performance of the first and second following vehicles using the proposed CF are is better than that of the conventional method. When there is a large velocity vibration of front vehicle, there is an overshot for the velocity of both first and second following vehicles using the conventional CF as compared with that of the proposed CF. Table 7 shows that the proposed CF offers better car-following performance than that of the conventional one during 140–180s.

As shown in Figure 7, when the velocity of the front vehicle fluctuates greatly during 355–375s, the conventional CF shows a large overshoot. However, the overshoot is reduced by using the proposed CF because EDO can effectively compensate for the influence of equivalent disturbance. When the velocity of the front vehicle fluctuates small, the car-following performance between the conventional CF and the proposed CF is not great. As can be seen from Table 8, the car-following performance of the proposed CF is better than that of the conventional CF during 340–380s.

6 | CONCLUSION

This paper proposes a type of car-following strategy of ICV using EDO, where the gain of EDO is adjusted by reinforcement learning to adapt to real time traffic environments. Stability of the proposed car-following method has been proven by Lyapunov approach. Meanwhile, numerical simulations in terms of two scenarios are carried out to verify the effectiveness of the proposed method. As compared with the conventional car-following method, it shows that the performance of both first and second following vehicles are improved significantly by using the proposed car-following method under same conditions if reinforcement learning is well trained.

ACKNOWLEDGEMENTS

This work was partly sponsored by the State Key Laboratory of Automotive Safety and Energy (Grant No. KFY2208), the Key Research and Development Plan of Anhui Province (Grant No. 202004a05020058), the Natural Science Foundation of Hefei, China (Grant No. 2021032), and the National Natural Science Foundation of China (Grant Nos. U20A20225 and U2013601).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ORCID

Ruidong Yan https://orcid.org/0000-0002-9479-6489
Hongbo Gao https://orcid.org/0000-0002-5271-1280

REFERENCES

1. Yang, D.G., et al.: Intelligent and connected vehicles: current status and future perspectives. Sci. China Technol. Sci. 61(10), 1446–1471 (2018). https://doi.org/10.1007/s11431-017-9338-1
2. Omeiza, D., et al.: Explanations in Autonomous Driving: A Survey (2021). arXiv:2103.05154
3. Guo, J.H., Luo, Y.G., Li, K.Q.: Integrated adaptive dynamic surface car-following control for nonholonomic autonomous electric vehicles. Sci. China 8, 1–10 (2017). https://doi.org/10.1007/s11431-016-9081-1

4. Li, S.B., et al.: Model predictive multi-objective vehicular adaptive cruise control. IEEE Trans. Control Syst. Technol. 19(3), 556–566 (2011). https://doi.org/10.1109/tcst.2010.2049203
5. Shaout, A., Jarrah, M.A.: Cruise control technology review. Comput. Electr. Eng. 23(4), 259–271 (1997). https://doi.org/10.1016/s0045-7906(97)00013-x
6. Chen, W.H., et al.: Disturbance observer based control and related methods — an overview. IEEE Trans. Ind. Electron. 63(2), 1083–1095 (2015). https://doi.org/10.1109/tie.2015.2478397
7. Guo, L., Chen, W.H.: Disturbance attenuation and rejection for systems with nonlinearity via DOBC approach. Int. J. Robust Nonlinear Control 15(3), 109–125 (2010). https://doi.org/10.1002/rnc.978
8. Chen, W.H.: Nonlinear disturbance observer enhanced dynamic inversion control of missiles. J. Guid. Control Dynam. 26(1), 161–166 (2003). https://doi.org/10.2514/2.5027
9. Zhou, C.Q., et al.: Robust temporal logic motion control via disturbance observers. IEEE Trans. Ind. Electron. 20, 1–10 (2022)
10. Chen, W.H., Guo, L.: Analysis of disturbance observer based control for nonlinear systems under disturbances with bounded variation. Proc. Control (2004)
11. Jiang, G., Song, W., Wang, S.: Design of observer with integrators for linear systems with unknown input disturbances. Electron. Lett. 36(13), 1168–1169 (2000). https://doi.org/10.1049/el:20000799
12. Ho, D.W., Gao, Z.: Proportional multiple-integral observer design for descriptor systems with measurement output disturbances. IEE Proc. Control Theor. Appl. 151(3), 279–288 (2004). https://doi.org/10.1049/ip-cta:20040437
13. Liang, Q., et al.: A unified observer for high order disturbances in time series expansion. In: Proceedings of the 26th Chinese Control and Decision Conference. pp, 2505–2510 (2014)
14. Qian, C., et al.: Robustly string stable longitudinal control for vehicle platoons under communication failures: a generalized extended state observer-based control approach. IEEE Trans. Intell. Veh. 25, 1–10 (2022)
15. Ginoya, D., Shendge, P.D., Phadke, S.B.: Sliding mode control for mismatched uncertain systems using an extended disturbance observer. IEEE Trans. Ind. Electron. 61(4), 1983–1992 (2014). https://doi.org/10.1109/tie.2013.2271597
16. Yan, R.D., Wu, Z.: Attitude stabilization of flexible spacecrafts via extended disturbance observer based controller. Acta Astronaut. 133, 73–80 (2017). https://doi.org/10.1016/j.actaastro.2017.01.004
17. Yan, R.D., et al.: Feedforward compensation-based finite-time traffic flow controller for intelligent connected vehicle subject to sudden velocity changes of leading vehicle. IEEE Trans. Intell. Transport. Syst. 21(8), 3357–3365 (2020). https://doi.org/10.1109/tits.2019.2926443
18. Yan, R.D., et al.: Distributed car-following control for intelligent connected vehicle using improved super-twisting compensator subject to sudden velocity changes of leading vehicle. IEEE Trans. Intell. Transport. Syst. 99, 1–10 (2021)
19. Jiang, W., Li, D., Ge, S.S.: Micro flapping-wing vehicles formation control with attitude estimation. IEEE Trans. Cybern. 52(2), 1035–1047 (2022). https://doi.org/10.1109/tcyb.2020.2988911
20. Lillicrap, T.P., et al.: Continuous Control with Deep Reinforcement Learning (2015). arXiv:1509.02971
21. Mu, C.X., et al.: Adaptive composite frequency control of power systems using reinforcement learning. CAAI Trans. Intell. Technol. 7(4), 671–684 (2022). https://doi.org/10.1049/cit2.12103
22. Ren, J.K., et al.: Deep reinforcement learning using least-squares truncated temporal-difference. CAAI Transactions on Intelligence Technology, 1–15 (2023). https://doi.org/10.1049/cit2.12202
23. Zhu, M.X., et al.: Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving. Transport. Res. C Emerg. Technol. 117, 1–14 (2020). https://doi.org/10.1016/j.trc.2020.102662
24. Liu, H.Q., Jiang, R.: Improving comfort level in traffic flow of CACC vehicles at lane drop on two-lane highways. Phys. Stat. Mech. Appl. 575, 1–9 (2021). https://doi.org/10.1016/j.physa.2021.126055
25. Stern, R.E., et al.: Dissipation of stop-and-go waves via control of autonomous vehicles: field experiments. Transport. Res. C Emerg. Technol. 89, 205–221 (2018). https://doi.org/10.1016/j.trc.2018.02.005

How to cite this article: Yan, R., et al.: Car-following strategy of intelligent connected vehicle using extended disturbance observer adjusted by reinforcement learning. CAAI Trans. Intell. Technol. 9(2), 365–373 (2024). https://doi.org/10.1049/cit2.12252