"This is the author's accepted manuscript. The final published version of this work (the version of record) is published by **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, Pages 1-11, ISSN: 2168-2232, DOI: 10.1109/TSMC.2016.2645699.This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher."

Rongxin Cui received the B.Eng. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 2003 and 2008, respectively. From 2008 to 2010, he was a Research Fellow with the Centre for Offshore Research and Engineering, National University of Singapore, Singapore. He is currently a Professor with the School of Marine Science and Technology, Northwestern Polytechnical University. His current research interests include control of nonlinear systems, cooperative path planning and control for multiple robots, control and navigation for underwater vehicles, and system development. Dr. Cui servers as an Editor for the *Journal of Intelligent and Robotic Systems*.

Chenguang Yang received the B.Eng. degree in measurement and control from Northwestern Polytechnical University, Xi'an, China, in 2005, and the Ph.D. degree in control engineering from the National University of Singapore, Singapore, in 2010. He received Post-Doctoral Training with Imperial College London, London, U.K. He is a Senior Lecturer with the Zienkiewicz Centre for Computational Engineering, Swansea University, Swansea, U.K. His current research interests include robotics, automation, and computational intelligence.

Yang Li received the B.Eng. degree from Northwestern Polytechnical University, Xi'an, China, in 2011, where he is currently pursuing the Ph.D. degree with the School of Marine Science and Technology. His current research interests include cooperative path planning and control of multiple robots.

Sanjay Sharma received the B.Tech. (Hons.) degree in electrical engineering from the Indian Institute of Technology Kharagpur, Kharagpur, India, the M.Tech. (Hons.) degree in control systems from the Indian Institute of Technology (BHU) Varanasi, Varanasi, India, and the Ph.D. degree in control and systems engineering from the University of Sheffield, Sheffield, U.K. He is a Reader with t三he School of Marine Science and Engineering, University of Plymouth, Plymouth, U.K., and the Head of the Autonomous Marine Systems Research Group. He is a member of the IMechEs Marine, Informatics and Control Group and also the IFAC Technical Committee on Intelligent Autonomous Vehicles. He has authored over 70 book, journal, and refereed conference publications. His current research interests include design, development and application of artificial intelligence techniques and evolutionary algorithms in navigation, guidance, and control of marine robotics and unmanned marine craft.

Adaptive Neural Network Control of AUVs With Control Input Nonlinearities Using Reinforcement Learning

Rongxin Cui, *Member, IEEE*, Chenguang Yang, *Senior Member, IEEE*, Yang Li *Student Member, IEEE* and Sanjay Sharma

***Abstract*—In this paper, we investigate the trajectory tracking problem for a fully actuated autonomous underwater vehicle (AUV) that moves in the horizontal plane. External disturbances, control input nonlinearities and model uncertainties are considered in our control design. Based on the dynamics model derived in the discrete-time domain, two neural networks (NNs), including a critic and an action NN, are integrated into our adaptive control design. The critic NN is introduced to evaluate the long-time performance of the designed control in the current time step, and the action NN is used to compensate for the unknown dynamics. To eliminate the AUV's control input nonlinearities, a compensation item is also designed in the adaptive control. Rigorous theoretical analysis is performed to prove the stability and performance of the proposed control law. Moreover, the robustness and effectiveness of the proposed control method are tested and validated through extensive numerical simulation results.**

***Index Terms*—Autonomous underwater vehicle (AUV); Trajectory tracking; Neural network; Adaptive control**

I. Introduction

Currently, underwater vehicles, including AUVs, remote operated vehicles (ROVs), and underwater gliders, have been widely employed in various underwater tasks [1]–[5]. In civil applications, they have been widely used in seafloor mapping, for pipeline checking for the oil and gas industry and to find missing airplane wreckage in air rescue operations. In military applications, they have been extensively applied to surveillance and reconnaissance missions, mine countermeasures, oceanography, payload delivery, and other time-critical tasks. AUVs have also been involved in scientific investigations of the ocean, the ocean floor and the lakes. Precise motion control of an AUV is crucial when performing underwater tasks. However, it is a challenging because of the model nonlinearity, coupling and time-varying hydrodynamic coefficients of the dynamics, which need to be further studied.

AUVs usually move in 3D space with 6 degree of freedom (DOF) and it involves coupled dynamics between its planar and diving motions. In most studies, AUV models are always decoupled, enabling possible application of various control methods [3], [6], [7]. There are several approaches that have been proposed for AUV trajectory tracking in 3D space, specifically for planner motion or diving. The nonlinear AUV model is usually linearized first, and then, the controller can be designed based on this linear model [8], [9]. With the decoupled model, the diving control of AUVs was analyzed in [6], and a differentiator was employed to enhance the noise attenuation performance so that active disturbance rejection could be achieved. By decoupling the depth and course motions, a fuzzy depth PD controller was designed in [10]. Moreover, an output feedback control was proposed in [8] for AUVs that move in the vertical plane by transforming the path-following errors into a Serret-Frenet frame and linearizing the error dynamics. For the planner motion control of AUVs, a nonlinear control for both fully actuated and underactuated configurations was proposed in [7]. They analyzed the effectiveness of the side-slip angle of AUVs in detail. Moreover, a tilting thruster configuration was proposed in [3], and a selective switching control was designed separately for two decoupled 3-DOF (degree-of-freedom) subsystems. In [11], both a current-induced vessel model and a general vehicle model were considered, where the former model accounted for the main current loads. Cascaded system theory and observer backstepping were then employed to design a nonlinear Luenberger observer and a controller for AUVs. Moreover, these results showed that the model-based controller performed better than the conventional PD control. In this case, the model dynamics in the controller should be revised in the case of divergences.

Optimal control was also studied in [12]–[14] based on a AUV dynamical model. In [12], an optimal control was designed to control the AUV trajectory at the kinematics level, and the cost function was described as the kinetic energy cost. An appropriate Hamiltonian was then designed based on the maximum principle, and an optimal solution was finally obtained. A nonlinear suboptimal control was presented for a non-affine AUV model, and the state-dependent Riccati equation controller was applied to the point-to-point tracking of the NPS II AUV [13]. Treating uncertainty bounds as one item in the cost function, an optimal control problem was then obtained by transforming the original robust control problem; then, an indirect robust depth control was presented [14].

The hydrodynamic parameters of AUVs are always obtained by computational fluid dynamics (CFD) methods or towed experiment identification. However, due to time-varying environmental and state changes that occur during underwater tasks, the obtained hydrodynamic parameters are not invariant [15]. Thus, both external disturbance and model parameter uncertainties should be considered in designing an appropriate

This work was supported by the National Natural Science Foundation of China (NSFC) under grant 61472325, and the Natural Science Basic Research Plan in Shaanxi Province of China under grant 2015JM5254.

R. Cui (Corresponding author) and Y. Li are with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China. (email: r.cui@nwpu.edu.cn).

C. Yang is with the Zienkiewicz Centre for Computational Engineering, Swansea University, SA1 8EN, UK. (email: cyang@theiet.org).

S. Sharma is with the School of Marine Science and Engineering, Plymouth University, UK. (email: sanjay.sharma@plymouth.ac.uk).

controller [16]–[23]. To resolve the model parameter uncertainties, Mamdani Fuzzy rule-based PID parameter adjustment has been employed in [24], and then, the control design was decoupled into two channels of heading and depth. A discrete time-delay control was presented in [25], in which the dynamics of an AUV was estimated directly and model uncertainties were compensated for by the time-delay estimation.

The velocity of an AUV can be measured by a Doppler velocity log (DVL), which usually has a slow update rate of new data. To enhance the robustness of the unmodeled dynamics and external disturbances for an AUV that uses a DVL, an integral sliding-mode control was introduced in [26]. In [27], a novel method to compensate for bounded external disturbances and model uncertainties was given, where the integral of the error sign control structure was presented, and the semiglobal asymptotic tracking performance could be established through Lyapunov stability analysis. Sliding mode control and backstepping were combined in [28] to design a trajectory tracking controller for an AUV with parameter uncertainties and external disturbances.

To address external disturbances, a disturbance force measurement method was introduced to measure the forces/moments acting on AUVs in [2]; then, a feed-forward control was employed in the vehicle based on the predicted response of the dynamic models. A disturbance observer is another main method that has been used to compensate for unknown external disturbances [11], [20], [29]–[32]. The low-frequency motion and wave frequency motion of AUVs were estimated by nonlinear observers in [20], and nonlinear tracking control was designed for AUV motion subject to shallow wave disturbances. For the purpose of controlling vehicles in the near space, a type of sliding mode tracking control was applied in [32] based on a disturbance observer. In addition, an adaptive tracking control for fully actuated surface vessels employing a disturbance observer was designed in [33].

Due to the function approximation ability of NNs, fuzzy approximators, NNs and fuzzy-control-based algorithms have been widely studied to compensate for the environmental disturbances and model uncertainties of AUVs [34]–[41]. NN approximation was employed to compensate for unknown model parameters and external disturbances, which were induced by ocean currents and waves, in [35], and the uniform ultimate boundedness of the tracking errors was achieved. NNs were used to address the model uncertainties of AUVs, and dynamic surface control was also applied in the control design in [36]. The nonlinear uncertainties of the AUV dynamics were approximated by a two-layer NN in [38]. To control the diving of AUVs, an adaptive control based on a stable NN was proposed in [42]. NN adaptive control was presented for multiple unmanned surface vessels in [43], and the unmeasured states were estimated by a local observer. A radial Basis function (RBF) NN was presented to derive the adaptive controller for systems subject to external disturbances and unknown hysteresis in [44]. In a recent work [45], a nonaffine pure-feedback discrete-time nonlinear system subject to input dead-zone was considered. To compensate for the dead-zone, an adaptive compensative term and an n-step-ahead predictor was constructed by transforming the original system.

Practical control systems for AUVs are usually implemented on an embedded computer in a digital manner with samplers. Thus, the continues-time controller needs to be transformed into a discrete-time version [46]. By using the discrete-time model directly, we develop a trajectory tracking control in the presence of external disturbances, model parameter uncertainties, and control input nonlinearities. It should be noted that there already exists a number of methods to address input nonlinearity problems such as input dead-zone and saturation [47]–[52]. Based on the back-stepping method and Lyapunov analysis, an adaptive trajectory tracking controller was designed to overcome the model parameter uncertainties in [51], where a saturation function was utilized to resolve the actuator saturation problem. To prevent velocity constraint violations, a robust adaptive controller was proposed in [48] for a remotely operated vehicle, and the barrier Lyapunov function was used in the Lyapunov syntheses. In [52], a novel dynamic surface control (DSC) was proposed for a pure-feedback system with unknown input dead-zone. The complexity was clearly reduced due to the dynamic surface control. A novel NN-based adaptive control was presented for a MIMO nonlinear system, therein considering the unknown dead-zone and control directions. In addition, reinforcement learning has been studied and utilized in many fields such as machine learning and artificial intelligence [53]–[55]. Reinforcement learning was first surveyed from a computer science perspective in [53]. In [54], the "keepers" of one soccer team were trained to learn when to hold or pass the ball. In addition, deep Q-learning was presented to solve more than 20 simulated tasks successfully with a continuous control space in [55]. In this paper, motivated by the work in [45], [56], [57], we propose a reinforcement learning technique to achieve optimal trajectory tracking for AUVs by employing two NNs. The unknown nonlinearities and disturbances are approximated by the action NN; simultaneously, the tracing evaluation of the tracking performance is approximated by the critic NN. In addition, an adaptive compensation for the control input nonlinearities is considered. The preliminary results of this work were presented in [58], and extensions have been made by considering not only actuator dead-zone and saturation but also the nonlinear relationship between the nominal and actual force/moment. Moreover, a compensation policy for this nonlinearity is proposed, as will be discussed later.

The remainder of the paper is organized as follows. We present the nonlinear model of AUVs in Section II. The two adaptive NNs are designed in Section III. Simulation studies and conclusions are presented in Sections IV and V, respectively.

II. Problem Formulation

A. Motion Equations of an AUV

As described in Section I, an AUV usually moves in a 3D space with 6 DOFs, leading to coupled dynamics in its planner and diving motions. To facilitate control design, the model is usually decoupled, whereas the designed control will be

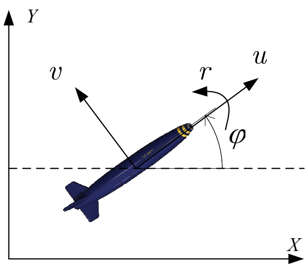


Fig. 1. AUV movement in the horizontal plane.

validated using the coupled nonlinear dynamics. We consider the planar motion of an AUV with 3 DOFs, as shown in Fig. 1. Let us denote the position coordinate of an AUV as (x,y), the yaw as (ψ) in the inertial frame, the velocity as (u) in surge, v in sway and r in yaw in the AUV body coordinates. Furthermore, let us denote the matrix of the inertia of the AUV as M and the matrices for the Coriolis and centripetal acceleration and damping as $C(\nu)$ and $D(\nu)$, respectively. In addition, we denote the forces and moments generated by gravity and buoyancy as $g(\eta)$. Consider that unknown external disturbances and model parameter uncertainties exist; then, the AUV dynamics can be given as follows:

$$\dot{\eta} = R(\psi)\nu \tag{1}$$

$$M\dot{\nu} + C(\nu)\nu + D(\nu)\nu + g(\eta) + \Delta(\eta,\nu) = \tau \tag{2}$$

where $\Delta(\eta,\nu)$ is the model uncertainty vector, which is induced by the unmodeled dynamics and external disturbances $\Delta(\eta,\nu)$, and $R(\psi)$ describes the rotation from the AUV body coordinates to the earth coordinates with 3 DOFs. The control inputs of the AUV are defined by $\tau \in \mathbb{R}^3$.

The elements in M, $m_{ij},\ i,j = 1,2,3$; the functions in $D(\nu)$, $d_{ij}(\nu),\ i,j = 1,2,3$; and the element of the disturbance vector $\Delta_i(\eta,\nu),\ i = 1,2,3$ are all unavailable for control design. To facilitate the control design, we assume that their is a nominal value of the unknown mass matrix M, which is defined by M_0. In addition, M_0 is known *a priori*. This assumption is feasible because the mass and added mass of an AUV are mainly determined by its physical shape. Control design in this work is focused on the 3-DOF model. We can extend the control policy to 6 DOFs conveniently due to the fully actuated model of AUVs used in this paper. In other words, the controller designed in this paper can also be applied to the vertical plane.

B. Dynamics Model in Discrete-Time Domain

In this subsection, we transform the continuous-time model into discrete time for subsequent control design. Eqs. (1) and (2) can be rewritten as the following equations.

$$\begin{aligned}\dot{\nu} =& -M^{-1}\left[(C(\nu)+D(\nu))\nu + g(\eta) + \Delta(\eta,\nu)\right] \\ &+ M^{-1}\tau \\ \dot{\eta} =& R(\psi)\nu\end{aligned} \tag{3}$$

If the sampling time of the embedded computer for the AUV control is selected as T_s, through the first-order Taylor expansion, we can obtain an approximative discrete-time model calculated from (3) as

$$\begin{aligned}\nu(k+1) &= \nu(k) + f_2\big(\eta(k),\nu(k)\big) + M^{-1}\tau(k) \\ \eta(k+1) &= \eta(k) + f_1\big(\eta(k)\big)\nu(k)\end{aligned} \tag{4}$$

where $\eta(k)$, $\nu(k)$ and $\tau(k)$ are the sampled values of η, ν and τ at the $k-$th sampling time, respectively. The nonlinear functions

$$\begin{aligned}f_1(\eta) =& T_s R(\psi) \in \mathbb{R}^{3\times 3} \\ f_2(\eta,\nu) =& -T_s M^{-1}\big[(C(\nu)+D(\nu))\nu \\ &+ g(\eta) + \Delta(\eta,\nu)\big] \in \mathbb{R}^3\end{aligned} \tag{5}$$

The following definitions are introduced for convenience of future design.

$$\begin{aligned}f_{11}\big(\bar{x}(k)\big) &\triangleq f_1\big(\eta(k) + f_1(\eta(k))\nu(k)\big) \\ &= f_1(\eta(k+1)) \in \mathbb{R}^3\end{aligned} \tag{6}$$

where $\bar{x}(k) \triangleq \left[\eta^\top(k), \nu^\top(k)\right]^\top$.

By using the procedure presented in [59], we can derive the following equations from (4):

$$\begin{aligned}\eta(k+2) =& \eta(k+1) + f_1(\eta(k+1))\nu(k+1) \\ =& \eta(k) + f_1(\eta(k))\nu(k) + f_{11}(\bar{x}(k))\nu(k) \\ &+ f_{11}(\bar{x}(k))f_2(\bar{x}(k)) + f_{11}(\bar{x}(k))M^{-1}\tau\end{aligned} \tag{7}$$

It is easy to check that $\frac{1}{T_s^2} f_{11}^\top(\bar{x}(k)) f_{11}(\bar{x}(k)) = R(\psi(k+1))R^\top(\eta(k+1)) = I$. Now, we define

$$\begin{aligned}f(\bar{x}(k)) \triangleq& f_{11}(\bar{x}(k))f_2(\bar{x}(k)) \in \mathbb{R}^3 \\ h(\bar{x}(k)) \triangleq& x_1(k) + f_1(\eta(k))\nu(k) \\ &+ f_{11}(\bar{x}(k))\nu(k) \in \mathbb{R}^3 \\ M_f(\bar{x}(k)) =& f_{11}(\bar{x}(k))M^{-1}f_{11}^\top(\bar{x}(k)) \in \mathbb{R}^{3\times 3} \\ \tau_f =& \frac{1}{T_s^2} f_{11}(\bar{x}(k))\tau \in \mathbb{R}^3\end{aligned}$$

Then, (7) can be written as

$$\eta(k+2) = h(\bar{x}(k)) + f(\bar{x}(k)) + M_f(\bar{x}(k))\tau_f \tag{8}$$

It is noted that $f_{11}(\bar{x}(k)) = f_1\big(\eta(k) + f_1(\eta(k))\nu(k)\big)$ is known, and then, at each time instant k, $R(\psi(k+1)) = f_{11}(\bar{x}(k))$ can be calculated so that $h(\bar{x}(k))$ is also known. The function $f(\bar{x}(k))$ is not available; thus, it has to be well considered in our design. Furthermore, the property of the M matrix results in a positive-definite matrix $M_f(\bar{x}(k))$, which is also unknown.

Actuator dead-zone and saturation inevitable exist in any physical system. In this work, we consider that the actuator has nonlinearities, including both saturation and dead-zone, as shown in Fig. 2. Let us define $\tau_{ideal}(k)$ as the ideal control input generated by the proposed controller, and the nominal force/moment acting on the vehicle can be described as

$$\overrightarrow{D}(\tau_{ideal}(k)) = \overrightarrow{m}(\tau_{ideal}(k))\tau_{ideal}(k) + \overrightarrow{b}(k) \tag{9}$$

where $\overrightarrow{D}(\tau_{ideal}(k))$ is a continuous first-order differentiable function. This formulation (9) has been widely used to describe actuator nonlinearities with both saturation and dead-zone [45], [60].

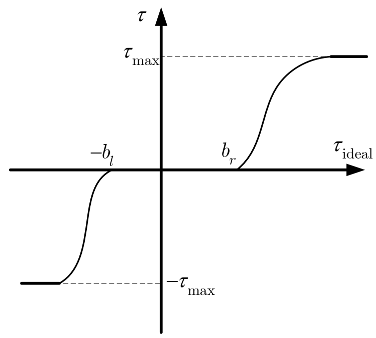


Fig. 2. Illustration of actuator nonlinearity.

The objective of this work, based on the discrete-time model (8) and the control input nonlinearities (9), is to develop a torque control input τ that makes the trajectory of an AUV, $\eta = [x, y, \psi]^\top$, follow the user-defined trajectory $\eta_d = [x_d, y_d, \psi_d]^\top$ asymptotically, i.e., $\lim_{k\to\infty}(\eta - \eta_d) = 0$.

C. NN Function Approximation

NNs and fuzzy systems are typical function approximators widely used in the control community. We choose the RBF NN to compensate for the unknown items in the dynamics for our control design, and the following RBF NNs are introduced. An unknown function $h(z) : \mathbb{R}^m \to \mathbb{R}$ can be approximated by

$$\phi(W, z) = W^\top S(z) \tag{10}$$

where the weight of the NN is denoted as $W = [w_1, w_2, \cdots, w_l]^\top \in \mathbb{R}^l$, the input vector is denoted as $z \in \Omega_z \subset \mathbb{R}^m$, and the node number of the NN is denoted as l.

In this work, the basis function of an NN is defined as $S(z) = [s_1(z), \cdots, s_l(z)]^\top$, and its element $s_i(z),\ i = 1, \ldots, l$ are selected as a Gaussian function.

Let us define $\mu_i = [\mu_{i1}, \cdots, \mu_{im}]^\top$ as the centers of the receptive domain of the NN, and let us define σ_i as the width of the Gaussian function. Then, the elements can be written as

$$s_i(z) = \exp\left[\frac{-(z-\mu_i)^\top(z-\mu_i)}{\sigma_i^2}\right], i = 1, 2, ..., l \tag{11}$$

Conventionally, W^* denotes the ideal constant weights of an NN. In [61], it was established that any continuous function can be approximated by an RBF NN (10) over a compact set $\Omega_z \subset \mathbb{R}^m$ as

$$\phi(z) = W^{*\top} S(z) + \varepsilon_z, \ \forall z \in \Omega_z \tag{12}$$

where ε_z is the approximation error.

The ideal weight W^* is required for stability analysis, and it minimizes $|\varepsilon_z|$ for all $z \in \Omega_z$, i.e.,

$$W^* \overset{def}{=} arg \min_{W' \in \mathbb{R}^l} \left\{ \sup |h(z) - W'^\top S(z)| \right\}, \quad z \in \Omega_z$$

In this work, the user-defined trajectories for an AUV, η_d and ν_d, are assumed to be suitably defined such that if $z_d = [\eta_d^\top, \nu_d^\top]^\top$ is presented as the input to an RBF NN, then $S(z_d)$ will satisfy the persistent excitation condition [62], i.e.,

$$\alpha_{\min} I_{l\times l} \leq \int_{k_0}^{k_0+k_f} S(\pi) S^\top(\pi) \mathrm{d}\pi \leq \alpha_{\max} I_{l\times l}, \quad \forall k_0 \tag{13}$$

where $\alpha_{\min}, \alpha_{\max}$ and k_f are positive constants and I is the identity matrix.

III. Adaptive NN Control Design

We derived the discrete-time model of an AUV in Section II, and we noted that there are some unknown functions in the dynamics model. Thus, the technical challenges in the control design of an AUV include the external disturbances, the partial unknown dynamics, and the input nonlinearities. In this section, we propose the trajectory control for an AUV using an NN. Two NNs are employed. The first critic NN is used to evaluate the long-time performance of the control in the current time step. Note that only $f(\bar{x}(k))$ and M_f are unknown and thus are not unavailable. Therefore, an action NN is used to approximate $f(\bar{x}(k))$ and M_f. In other words, the action NN could compensate for the effect caused by the unknown dynamics. We will design the control τ to adjust $\eta(k)$ to track the desired trajectory $\eta_d(k)$. Define

$$e(k) \triangleq \eta(k) - \eta_d(k) \in \mathbb{R}^3 \tag{14}$$

and $e_\nu(k) = \nu(k) - \nu_d(k)$ as the position tracking error and the velocity tracking error, respectively.

Define

$$\varepsilon(k) \triangleq \sum_{i=1}^{3} w_i \left|\lambda_i e_i(k) + e_{\nu,i}(k)\right| \tag{15}$$

as a new weighted tracking error, where the superscript i denotes the i-the element of the vector, λ_i is a constant weighting the position tracking error and velocity error, and w_i is a constant weight associated with the error in each channel. w_i can also be viewed as the weights of the impacts of e_i and $e_{\nu,i}$. Generally, it is chosen to achieve the normalization of e_i and $e_{\nu,i}$.

The definition in (15) is motivated by the results in [56], [57], and both position and velocity errors are considered in this work. The control objective can then be described as $\lim_{k\to\infty} \varepsilon(k) = 0$. Thus, $\varepsilon(k)$ can be viewed as a strategic utility function used to observe the instant tracking performance. Then, the long-term performance measure with a future time horizon N can be defined as [56]

$$Q(k) = \sum_{i=1}^{N-k} \alpha^{N+1-i} \varepsilon(k+i) \tag{16}$$

where the scaling factor α is defined by the user and satisfies $0 < \alpha < 1$.

The long-term performance $Q(k)$ was first introduced in [56] to denote the tracking performance including all history information. It utilized the binary system performance index $p_i(k) \in \mathbb{R}$. In addition, $\varepsilon(k) = 0$ when the tracking error is within the limits of a given boundary; otherwise, $\varepsilon(k) = 1$. In this paper, however, we use the weighted error to denote the performance rather than the binary utility function. Moreover, this measure is also similar to the standard Bellman equation [63], [64].

In the following, we design the tracking controller using two NNs. The utility function $Q(k)$ (16) is approximated by one critic NN, and the unknown item $f(\bar{x}(k))$ in (8) is approximated by the other NN.

A. Critic NN Design

As mentioned, two NNs are employed in our control design. In this subsection, we use a critic NN to approximate the unknown strategic utility function $Q(k)$. Following the techniques used in [65], we calculate its estimation. Rewriting $Q(k)$ using the NN, it can then be formulated as follows:

$$Q(k) = W_c^* S_c(k) + \mu_Q(k) \tag{17}$$

where $W_c^* \in \mathbb{R}^l$ is the optimal neural weight, $\mu_Q(k)$ denotes the NN approximation error, and $S_c(z(k)) \in \mathbb{R}^l$ denotes the activation vector. Then, the estimation of $Q(k)$ can be given by

$$\hat{Q}(k) = \hat{W}_c^\top(k) S_c(k) \tag{18}$$

where $\hat{W}_c(k)$ is the neural weight.

The input vector $z(k)$ is given by

$$\begin{aligned} z(k) \quad = \quad & [\bar{x}^\top(k), \eta_d^\top(k), , \cdots, \\ & \eta_d^\top(k+N), \nu_d^\top(k), \cdots, \nu_d^\top(k+N)]^\top \end{aligned} \tag{19}$$

Now, we can obtain the prediction error

$$e_c(k) = -\alpha \hat{Q}(k-1) + \hat{Q}(k) + \alpha^{N+1} \varepsilon(k) \tag{20}$$

Then, a critic NN is designed to minimize the objective function

$$E_c(k) = \frac{1}{2} e_c^2(k) \tag{21}$$

A simple method to update the critic NN is to use the conventional gradient-based adaption as follows:

$$\hat{W}_c(k+1) = \hat{W}_c(k) + \Delta \hat{W}_c(k) \tag{22}$$

In (22), the recurrence of $\hat{W}_c(k)$ is given by

$$\begin{aligned} &\hat{W}_c(k+1) = \hat{W}_c(k) - \alpha_c S_c(k) \\ &\times \left[\hat{W}_c^\top(k) S_c(k) - \alpha \hat{W}_c^\top(k-1) S_c(k-1) + \alpha^{N+1} p(k) \right] \end{aligned} \tag{23}$$

where $\alpha_c \in \mathbb{R}$ denotes the parameter gain of the NN. Eq. (23) shows that the weights are adjusted in accordance with the reinforcement learning signal and past critic NN output values with a discount.

B. Action NN Design

A critic NN has been used to approximate the performance evaluation function. In this subsection, an action-NN-based adaptive control considering the above-mentioned technical challenges is presented for the AUV as follows:

$$\tau_{ideal}(k) = \tau_f(k) + \hat{\xi}(k) + \beta(k) e(k) \tag{24}$$

where β is a user-defined scaling factor satisfying $|\beta(k)| \le \bar{\beta} < 1$ and τ_f is the ideal control input defined as

$$\tau_f = \hat{W}_a^\top(k) \langle \cdot \rangle S_a(\bar{z}(k)) - M_i(\bar{x}(k))[h(\bar{x}(k)) + y_d(k+2)] \tag{25}$$

in which $\xi(k)$ is an input compensation signal that will be introduced later. $\hat{\xi}, \xi^*$ and $\tilde{\xi}$ are the real compensation, the optimal compensation and the error of ξ, i.e., $\tilde{\xi} = \xi^* - \hat{\xi}$. $\hat{W}_a^\top(k)\langle \cdot \rangle S_a(\bar{z}(k))$ is the NN-based approximation of the unknown item $f_n(\bar{x}(k)) \in \mathbb{R}^3$.

$$\begin{aligned} f_n(\bar{x}(k)) = & -\left[M_f^{-1}(\bar{x}(k)) - M_i(\bar{x}(k)) \right] h(\bar{x}(k)) \\ & - M_f^{-1}(\bar{x}(k)) f(\bar{x}(k)) \end{aligned} \tag{26}$$

where $M_i(\bar{x}(k))$ is an estimation of $M_f^{-1}(\bar{x}(k))$ and

$$M_i(\bar{x}(k)) = f_{11}(\bar{x}(k)) M_0 f_{11}^\top(\bar{x}(k)) \tag{27}$$

If the nominal value of M_0 equals M exactly, then $M_i(\bar{x}(k)) M_f(\bar{x}(k)) = I$ and the first item of $f_n(\bar{x}(k))$ in (26) equals zero. In practice, M_0 is obtained according to the designer's experience. Although M_0 could never be precisely equivalent to M, the distance between M and M_0 will be compensated by the action NNs. In addition, the closer M_0 is to M, the less convergence time is needed.

According to universal approximation theory, there are $S(\bar{x}(k)) \in \mathbb{R}$ and ideal weights $W_a^{*\top}$ satisfying

$$f(\bar{x}(k)) = \mu(\bar{x}(k)) + W_a^{*\top} \langle \cdot \rangle S(\bar{x}(k)) \tag{28}$$

$$S_a(\bar{x}(k)) \in \mathbb{R}^{l \times 3}, \ \forall \bar{x}(k) \in \Omega_{\bar{x}} \tag{29}$$

where $\mu(\bar{x}(k))$ denotes the approximation error by the designed NN.

The update law of the compensation item $\xi(k)$ is designed as

$$\hat{\xi}(k+2) = \hat{\xi}(k) - \left[\gamma_\xi \hat{\xi}(k) + \alpha_\xi e(k+2) \right] \tag{30}$$

where α_ξ and γ_ξ are parameters to be specified by the designer.

Substituting the designed control (24) into the dynamics (7) results in

$$e_a(k+2) = M_f(\bar{x}(k)) \left[\tilde{D}(\nu(k)) + \tilde{W}_a^\top(k) S_a(k) \right] + d_s^*(k) \tag{31}$$

where $d_s^*(k) = M_f \mu(k)$, and the augmented error of $\tilde{D}(\nu(k))$ is defined as $\tilde{D}(\nu(k)) = D^*(\nu(k)) - D(\nu(k))$, with the superscript "$*''$ denoting the ideal value. Eq. (8) is a two-step predictor form based on the n-step formation in [66], which only involves current states and input.

Now, we need to minimize the following objective function.

$$E_a(k) = \frac{1}{2} e_a^\top(k) e_a(k) + \frac{1}{2} \hat{Q}^2(k) \tag{32}$$

where $e_a^\top(k) e_a(k)$ describes the performance of the approximation between the action NN and $f_n(\bar{x}(k))$. In addition, $\hat{Q}^2(k)$ describes the tracking performance, including all history information. Defining $k_1 = k - n$, we can now design the update law for the action NN.

$$\hat{W}_a(k+1) = \Delta \hat{W}_a(k_1) + \hat{W}_a(k_1) \tag{33}$$

where $\Delta\hat{W}_a$ can be calculated using the gradient descent method, which results in

$$\hat{W}_a(k+1) = \hat{W}_a(k_1) - \alpha_a S_a(k_1)[-\text{sign}(e(k))\hat{Q}(k)\alpha^{N+1} + e(k)] \quad (34)$$

where $\alpha_a \in \mathbb{R}$ is the adaption gain of the NN.

Lemma 1: [45] Let $V(k) = \sum_{i=1}^{n} V_i(k)$, and $V_i(k) \geq 0, k \in \mathbb{Z}^+$. If $V(k+1) \leq \sum_{i=1}^{n} p(k-n+1)V_i(k-n+1) + q(k-n+1)$, $|p(k-n+1)| \leq \bar{p} < 1$ and $|q(k-n+1)| \leq \bar{q}$. Then, we have

$$V(k) \leq \bar{V}(0) + \frac{\bar{q}}{1-\bar{p}} \quad (35)$$

Moreover, we have

$$\lim_{k \leftarrow \infty} \sup V(k) \leq \frac{\bar{q}}{1-\bar{p}} \quad (36)$$

where $\bar{V}(0) = \max_{0 \leq i \leq n-1}\{V(j)\}$.

Now, we can arrive at the following theorem, which summarizes the stability result of the developed control.

Theorem 1: If the dynamics of the vehicle can be described by (7), using the adaptive control (24), compensation parameter adaptation law (30), and NN weight update laws (22) and (33), the AUVs are able to follow the desired trajectory with bounded error when the design parameters of the control satisfy $\alpha_c\|S_c(k)\|^2 < 1$, $\alpha_a\|S_a(k)\|^2 < 1$, $\gamma_\xi < 1/2$, $0 < \alpha < \frac{\sqrt{2}}{2}$ and $\bar{\beta} + 3\alpha_\xi < \frac{2}{M_m(\bar{x}(k))}$.

Proof. Using similar techniques as employed in [45], [56]–[58], we choose a positive definite

$$V(k) = \sum_{i=1}^{4} V_i(k) \quad (37)$$

where

$$\begin{aligned} V_1(k) &= \frac{1}{\alpha_c}\text{tr}\left[\tilde{W}_c^\top(k)\tilde{W}_c(k)\right] \\ V_2(k) &= \frac{1}{\gamma_c}\|\zeta_c(k-1)\|^2, \text{ and} \\ V_3(k) &= \frac{1}{\gamma_a\alpha_a}\sum_{j=0}^{n}\text{tr}\left[\tilde{W}_a^\top(k-n+j)\tilde{W}_a(k-n+j)\right] \end{aligned} \quad (38)$$

where $\zeta_c(k) = \tilde{W}_c^\top(k)S_c(k)$ and the parameters $\gamma_c, \gamma_a > 0$. The important item $V_4(k)$ defined in (38) is introduced to handle the control input nonlinearity items and is selected as

$$V_4(k) = \frac{1}{\alpha_\xi}\tilde{\xi}^\top(k)\tilde{\xi}(k) \quad (39)$$

According to the results in [58], we can obtain

$$\Delta V_i(k+1) = V_i(k+1) - V_i(k) < 0, \ i = 1,2,3 \quad (40)$$

The details of the proof can be found in [58]. Based on (38), we obtain

$$\begin{aligned} V_4(k+2) =& \frac{\tilde{\xi}^\top(k+2)\tilde{\xi}(k+2)}{\alpha_\xi} \\ =& \frac{\tilde{\xi}^\top(k)\tilde{\xi}(k)}{\alpha_\xi} - 2\tilde{\xi}(k)^\top e(k+2) - \frac{2\gamma_\xi\tilde{\xi}^\top(k)\hat{\xi}(k)}{\alpha_\xi} \\ & + 2\gamma_\xi e(k+2)^\top\hat{\xi}(k) + \frac{\gamma_\xi^2\hat{\xi}(k)}{\alpha_\xi} \\ & + \alpha_\xi e^\top(k+2)e(k+2) \end{aligned} \quad (41)$$

Because

$$\begin{aligned} & M_f(\bar{x}(k))D(v(k)) \\ =& M_f(\bar{x}(k))\left\{-b(k) + m(k)\left[\tau_f(k) + \hat{\xi}(k) + \beta e(k)\right]\right\} \\ =& -M_f(\bar{x}(k))b(k) + M_f(\bar{x}(k))m(k)\tau_f(k)\hat{\xi}(k) \\ & + M_f(\bar{x}(k))m(k) + M_f(\bar{x}(k))m(k)\beta e(k) \\ =& -M_f(\bar{x}(k))b(k) + M_f(\bar{x}(k))m(k)\tau_f(k) \\ & - M_f(\bar{x}(k))m(k)\tilde{\xi}(k) + M_f(\bar{x}(k))m(k)\xi^* \\ & + M_f(\bar{x}(k))m(k)\beta e(k) \end{aligned}$$

Now, we can rewrite (31) as

$$\begin{aligned} e(k+2) =& M_f(\bar{x}(k))(m(k)-1)\tau_f(k) + M_f(\bar{x}(k))\tilde{W}_a^\top(k)S_a(k) \\ & - M_m(\bar{x}(k))\tilde{\xi}(k) + M_m(\bar{x}(k))\beta e(k) + M_m(\bar{x}(k))\xi^* \\ & - M_f(\bar{x}(k))b(k) + d_s^*(k) \\ =& M_m(\bar{x}(k))\left[\tilde{\xi}(k) + \beta e(k)\right] + H(k) \end{aligned} \quad (42)$$

where

$$\begin{aligned} H(k) =& M_f(\bar{x}(k))(m(k)-1)\tau_f(k) + M_m(\bar{x}(k))\xi^* \\ & - M_f(\bar{x}(k))b(k) + M_f(\bar{x}(k))\tilde{W}_a^\top(k)S_a(k) + d_s^*(k) \end{aligned} \quad (43)$$

Let us multiply by $e(k+2)$ on both sides of (42) to obtain $e^\top(k+2)e(k+2) = e^\top(k+2)H(k) + M_m(\bar{x}(k))[e^\top(k+2)\tilde{\xi}(k) + e^\top(k+2)\beta e(k)]$. Thus, we have

$$\begin{aligned} -2e^\top(k+2)\tilde{\xi}(k) =& -\frac{2e^\top(k+2)e(k+2)}{M_m(\bar{x}(k))} + \frac{2e^\top(k+2)H(k)}{M_m(\bar{x}(k))} \\ & + 2\beta e^\top(k+2)e(k) \end{aligned} \quad (44)$$

It is easy to obtain the following equations or inequalities.

$$\begin{aligned} & 2\tilde{\xi}^\top(k)\hat{\xi}(k) = -\xi^{*\top}\xi^* + \hat{\xi}^\top(k)\hat{\xi}(k) + \tilde{\xi}^\top(k)\tilde{\xi}(k) \\ & 2\gamma_\xi e^\top(k+2)\hat{\xi}(k) \leq \alpha_\xi e^\top(k+2)e(k+2) + \frac{\gamma_\xi^2\hat{\xi}^\top(k)\hat{\xi}(k)}{\alpha_\xi} \\ & \frac{2H^\top(k)e(k+2)}{M_m(\bar{x}(k))} \leq \alpha_\xi e^\top(k+2)e(k+2) + \frac{\bar{H}^\top(k)\bar{H}(k)}{\alpha_\xi} \\ & 2\beta e^\top(k+2)e(k) \leq \bar{\beta}\left[e^\top(k+2)e(k+2) + e^\top(k)e(k)\right] \end{aligned} \quad (45)$$

where $\bar{H}(k)$ is the upper bound of $H(k)$. It is noted that the first equation is a direct conclusion of $2a^\top b < a^\top a + b^\top b$.

Substituting (45) into (41), we obtain

$$\begin{aligned}
V_4(k+2) \leq & \frac{1}{\alpha_\xi}\tilde{\xi}^\top(k)\tilde{\xi}(k) - \frac{2}{M_m(\bar{x}(k))}e^\top(k+2)e(k+2) \\
& + \alpha_\xi e^\top(k+2)e(k+2) + \frac{1}{\alpha_\xi}\bar{H}^\top(k)\bar{H}(k) + \bar{\beta}e^\top(k)e(k) \\
& + \bar{\beta}e^\top(k+2)e(k+2) - \frac{\gamma_\xi}{\alpha_\xi}\tilde{\xi}^\top(k)\tilde{\xi}(k) - \frac{\gamma_\xi}{\alpha_\xi}\hat{\xi}^\top(k)\hat{\xi}(k) \\
& + \frac{\gamma_\xi}{\alpha_\xi}\xi^{*\top}\xi^* + 2\alpha_\xi e^\top(k+2)e(k+2) + \frac{2\gamma_\xi^2\hat{\xi}^\top(k)\hat{\xi}^\top(k)}{\alpha_\xi} \\
= & (1-\gamma_\xi)\frac{\tilde{\xi}^\top(k)\tilde{\xi}(k)}{\alpha_\xi} + \gamma_\xi(2\gamma_\xi - 1)\frac{\hat{\xi}^\top(k)\hat{\xi}(k)}{\alpha_\xi} \\
& + \left[\bar{\beta} + 3\alpha_\xi - \frac{2}{M_m(\bar{x}(k))}\right]e^\top(k+2)e(k+2) + \bar{\beta}e^\top(k)e(k) \\
& + \frac{\bar{H}^\top(k)\bar{H}(k)}{\alpha_\xi} + \frac{\gamma_\xi}{\alpha_\xi}\xi^{*\top}\xi^* \\
\leq & (1-\gamma_\xi)V_4(k) + \gamma_\xi(2\gamma_\xi - 1)\frac{\hat{\xi}^\top(k)\hat{\xi}(k)}{\alpha_\xi} \\
& + \left[\bar{\beta} + 3\alpha_\xi - \frac{2}{M_m(\bar{x}(k))}\right]e^\top(k+2)e(k+2) + \bar{q}
\end{aligned} \tag{46}$$

where $\bar{q} = \frac{\bar{H}^\top(k)\bar{H}(k)}{\alpha_\xi} + \frac{\gamma_\xi}{\alpha_\xi}\xi^{*\top}\xi^* + \bar{\beta}e^\top(k)e(k)$. If the parameters are selected to satisfy $\gamma_\xi < 1/2$, and $\bar{\beta} + 3\alpha_\xi < 2/M_m(\bar{x}(k))$, then we can obtain

$$V_4(k+2) \leq (1-\gamma_\xi)V_4(k) + \bar{\beta}e^\top(k)e(k) + \bar{q} \tag{47}$$

In addition, because $\gamma_\xi > 0$, $\bar{\beta} < 1$, and $\Delta V_i(k) < 0$, $i = 1,2,3$, we obtain $V(k+2) \leq \beta_{v1}V_1(k) + \beta_{v2}V_2(k) + \beta_{v3}V_3(k) + (1-\gamma_\xi)V_4(k) + \bar{\beta}e^\top(k)e(k) + \bar{q}$. Thus, we can conclude that $V(k) \leq V(0) + \frac{\bar{q}}{1-\bar{p}}$ according to Lemma 1, where the constant $\bar{p} = \max\left\{1-\gamma_\xi, \beta_{v1}, \beta_{v2}, \beta_{v3}\bar{\beta}\right\}$. This completes the proof.

IV. Simulation Studies

To evaluate the performance of the proposed control, we perform a numerical simulation based on the model of a fully actuated AUV in this section, as has been used in [67] with success. The parameters of the model are adapted from [67] as

$$M = \begin{bmatrix} 25.8 & 0 & 0 \\ 0 & 24.6612 & 0 \\ 0 & 0 & 2.76 \end{bmatrix} \tag{48}$$

$d_{11} = (0.7225 + 1.3274|u| + 5.8664v^2)kg/s$, $d_{22} = (0.8612 + 36.2823|v| + 8.05|r|)kg/s$, $d_{23} = (-0.1079 + 0.845|v| + 3.45|r|)kg/s$, $d_{32} = (-0.1052 - 5.0437|v| - 0.13|r|)kg/s$, and $d_{33} = (1.9 - 0.08|v| + 0.75|r|)kg/s$. The desired trajectory is given by

$$\begin{aligned} x_d(k) &= k \\ y_d(k) &= 4\sin(k/7) \end{aligned} \tag{49}$$

Thus, the desired yaw of the vehicle can be calculated as $\psi_d(k) = \arctan(\frac{4}{7}\cos(k/7))$ to achieve a smooth trajectory tracking. For convenience, we set the initial condition as $\eta(0) = [-2, 10, -\frac{\pi}{8}]^\top$ and $\nu(0) = 0_{3\times 1}$ for the AUV in the simulation. We introduce the time-varying and state-dependent disturbance in the earth coordinates as

$$f_c(k) = \begin{bmatrix} 0.1\sin(\nu(3)) \\ 0.0\nu(1)\eta(1) + 0.5 \\ -0.1\nu(2)\cos(\eta(3)) + 0.1\sin(\nu(2)) \end{bmatrix}$$

Accordingly, the disturbance acting on the AUV in the body-fixed frame can be written as

$$w(k) = R^\top(\psi)f_c(k)$$

The constant parameters in the controller are selected as $\alpha = 0.6$, $\alpha_0 = 0.5$, $\alpha_\alpha = 0.2$, $\alpha_\eta = 0.25$, $\alpha_c = 0.8$, $\gamma_\eta = 0.25$, and $T_s = 0.01$, which are chosen empirically. The sampling time T_s is chosen to satisfy Shannon's law. The basis functions of the NN are defined as

$$S_i(Z) := \frac{\mu_i(Z)}{\sum_{j=1}^{3^8}\mu_j(Z)}, \ \mu_i := \prod_{j=1}^{8}\nu_j, \ j = 1, \ldots, 3^8 \tag{50}$$

where the function ν_j can be selected from the $j-$th sets of $1/\left(1+e^{(-a_{3j}(Z_j-b_{3j}))}\right)$, $1/\left(1+e^{a_{1j}(Z_j+b_{1j})}\right)$, and $e^{-a_{2j}|Z_j-b_{2j}|^2}$. The maximum control signal in each channel is set as $|\tau_1| \leq 35N$, $|\tau_2| \leq 35N$, and $|\tau_3| \leq 7Nm$.

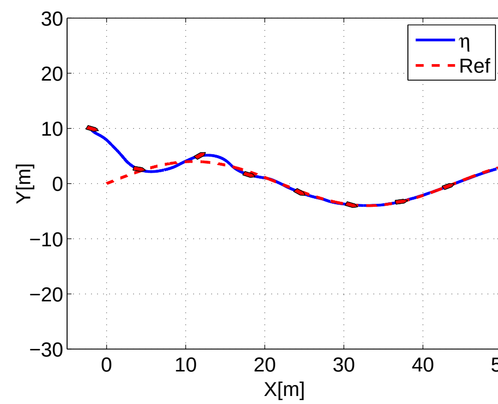


Fig. 3. Trajectory of an AUV in the simulation.

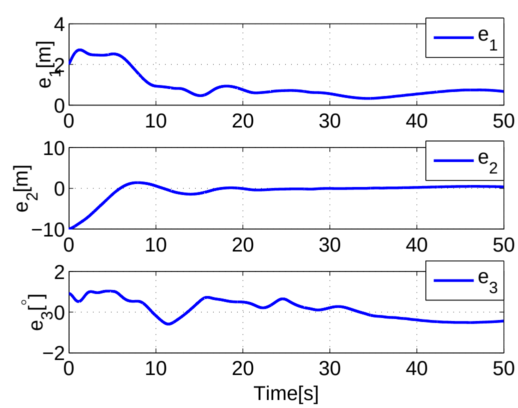


Fig. 4. Tracking error along the trajectory.

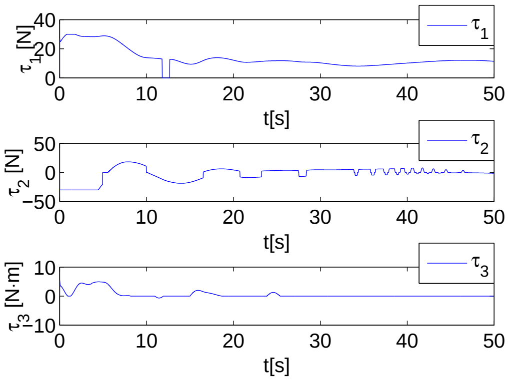


Fig. 5. Control inputs of the AUV.

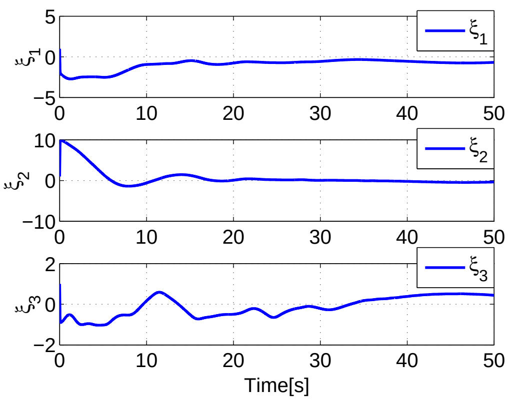


Fig. 6. Compensation control inputs ξ.

The primary simulation results are provided in Figs. 3−9 and demonstrate that the performance with our proposed control is reasonably good. It can be observed that the reference trajectory is tracked in 20 sec in Fig. 3. The tracking error is shown in Fig. 4 with a small boundary. It is clear that the norms of the NN weights as well as the control inputs are bounded in Figs. 7 and 6. In addition, Fig. 9 provides the reinforcement learning signal. It can be observed that Q is bounded near zero, which means that the weighted tracking

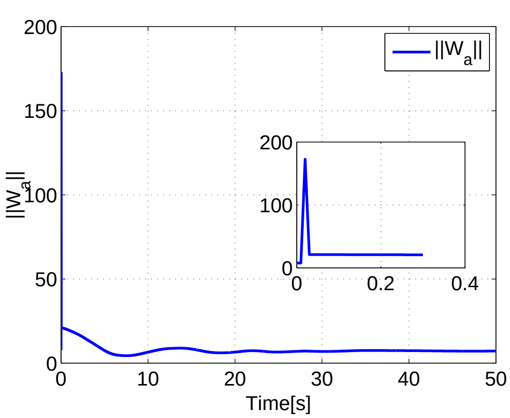


Fig. 7. Norm of action neural weights $\|\hat{W}_a\|$.

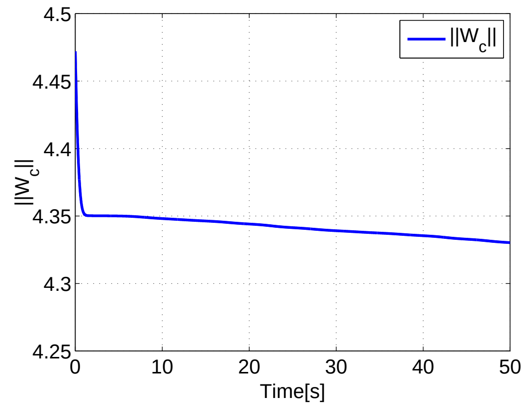


Fig. 8. Norm of critic neural weights $\|\hat{W}_c\|$.

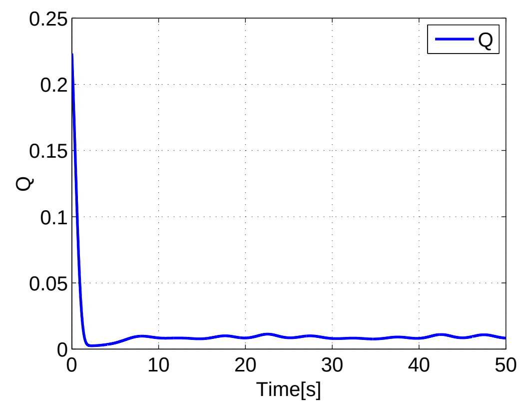


Fig. 9. Reinforcement learning signal.

error is also bounded to zero. From the figures, we can find that the performance of the AUV trajectory tracking is satisfactory, despite the unknown dynamics, control input nonlinearities and time-varying disturbance.

A. Compared with General NN Control

To validate our proposed adaptive control with reinforcement learning, we also present a comparison between the general NN control and PD control. The results are shown in Figs. 10-13. In Fig. 10, we see that the AUVs track the desired trajectory in a more effective manner with our proposed adaptive reinforcement learning control, i.e., the reference trajectory is tracked well enough when $x = 15$ by our control; however, $x = 25$ under the general NN control. Fig. 11 shows the comparison of the error, from which we see that our control achieves faster convergence. This means that the learning time needed by the NNs can be reduced by our control.

B. Compared with PD control

We choose two different PD control parameters in the simulation, $K_{p1} = [2, 2, 0.1]^\top$ $K_{d1} = [10, 10, 5]^\top$ and $K_{p2} = [10, 10, 0.5]^\top$ $K_{d2} = [1, 1, 0.5]^\top$. The results obtained with PD control are presented in Figs. 12 and 13. We see that η

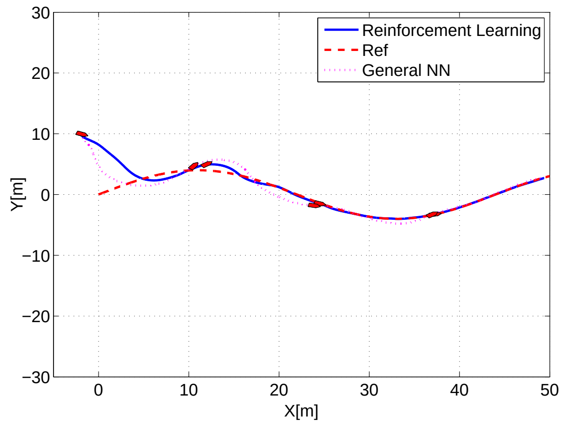


Fig. 10. Trajectory of an AUV compared with general NNs.

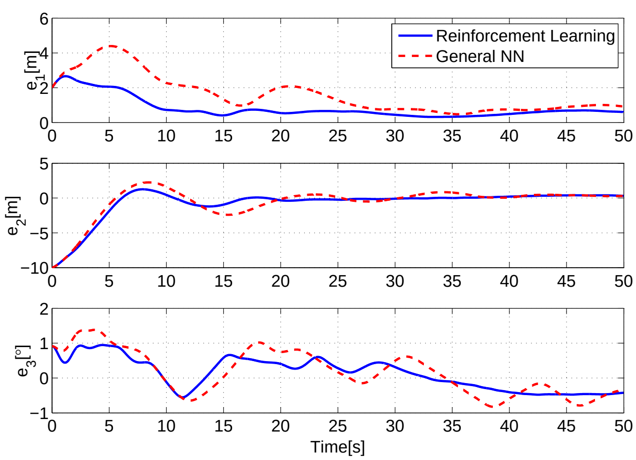


Fig. 11. Tracking error compared with general NNs.

has achieves a steady-state error because of the absence of the integrating in the PD control. The parameters in the PD control are chosen empirically and are generally difficult to choose in real applications. On the other hand, the parameters of the PD control strongly influence the effectiveness of the AUV tracking control system. The main difference is that the steady-state error is smaller when the latter parameters are selected. However, both parameters exhibit poorer performance compared with our adaptive control.

V. Conclusion

In this work, an adaptive trajectory tracking control law using NN approximation for a fully actuated AUV has been developed in the discrete-time domain. An NN-based reinforcement learning algorithm has been used to address unknown disturbances, parameter uncertainties and control input nonlinearities. Two NNs are embedded in the proposed controller: the first critic NN is used to evaluate the long-time performance of the control in the current time step, and the second action NN is used to compensate for the unknown dynamics. Rigorous theoretical analysis and extensive simulation studies have been performed to demonstrate the robustness and effectiveness of

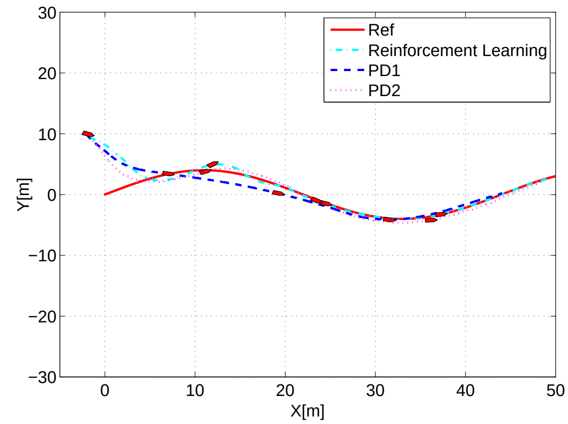


Fig. 12. Trajectory of an AUV compared with PD controller.

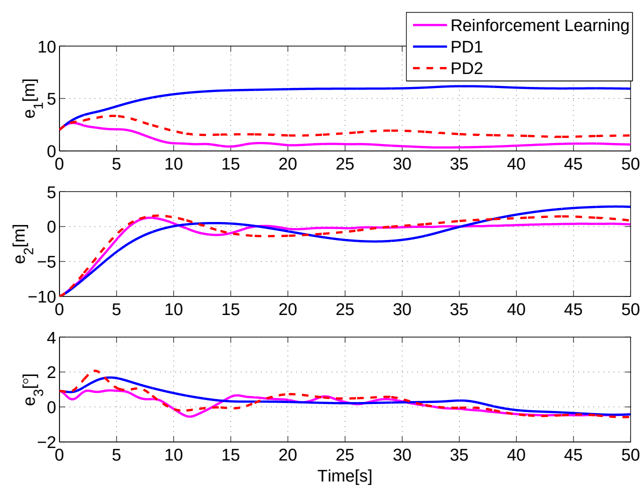


Fig. 13. Tracking error compared with PD controller. $Kp = [2,2,0.1]^\top$ $Kd = [10,10,5]^\top$ and $Kp = [10,10,0.5]^\top$ $Kd = [1,1,0.5]^\top$.

the proposed approach. A future research direction is to apply the proposed control to practical systems.

References

[1] R. Kobayashi and S. Okada, "Development of hovering control system for an underwater vehicle to perform core internal inspections," *Journal of Nuclear Science and Technology*, vol. 53, pp. 566–573, Apr 2016.

[2] J. M. Selvakumar and T. Asokan, "Station keeping control of underwater robots using disturbance force measurements," *Journal of Marine Science and Technology*, vol. 21, no. 1, pp. 70–85, 2016.

[3] S. Jin, J. Kim, J. Kim, and T. Seo, "Six-degree-of-freedom hovering control of an underwater robotic platform with four tilting thrusters via selective switching control," *IEEE/ASME Transactions on Mechatronics*, vol. 20, pp. 2370–2378, Oct 2015.

[4] Z. Peng, J. Wang, and D. Wang, "Containment maneuvering of marine surface vehicles with multiple parameterized paths via spatial-temporal decoupling," *IEEE/ASME Transactions on Mechatronics*, 2016. doi:10.1109/TMECH.2016.2632304.

[5] C. S. Chin, M. W. S. Lau, and E. Low, "Supervisory cascaded controller design: experiment test on a remotely operated vehicle," *Proceedings of the Institution of Mechanical Engineers Part C-Journal of Mechanical Engineering Science*, vol. 225, no. C3, pp. 584–603, 2011.

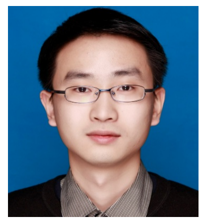
[6] Y. Shen, K. Shao, W. Ren, and Y. Liu, "Diving control of autonomous underwater vehicle based on improved active disturbance rejection control approach," *Neurocomputing*, vol. 173, Part 3, pp. 1377 – 1385, 2016.

[7] X. Xiang, C. Yu, Q. Zhang, and G. Xu, "Path-following control of an AUV: Fully actuated versus under-actuated configuration," *Marine Technology Society Journal*, vol. 50, no. 1, pp. 34–47, 2016.
[8] B. Subudhi, K. Mukherjee, and S. Ghosh, "A static output feedback control design for path following of autonomous underwater vehicle in vertical plane," *Ocean Engineering*, vol. 63, pp. 72–76, 2013.
[9] W. He and S. S. Ge, "Vibration control of a flexible beam with output constraint," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 5023–5030, 2015.
[10] I. S. Akkizidis, G. N. Roberts, P. Ridao, and J. Batlle, "Designing a Fuzzy-like PD controller for an underwater robot," *Control Engineering Practice*, vol. 11, no. 1, pp. 471–480, 2003.
[11] J. E. Refsnes, A. J. Sørensen, and K. Y. Pettersen, "Model-based output feedback control of slender-body underactuated AUVs: Theory and experiments," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 5, pp. 930–946, 2008.
[12] J. Biggs and W. Holderbaum, "Optimal kinematic control of an autonomous underwater vehicle," *IEEE Transactions on Automatic Control*, vol. 54, no. 7, pp. 1623–1626, 2009.
[13] B. Geranmehr and S. R. Nekoo, "Nonlinear suboptimal control of fully coupled non-affine six-DOF autonomous underwater vehicle using the state-dependent riccati equation," *Ocean Engineering*, vol. 96, pp. 248–257, 2015.
[14] H. Pan and M. Xin, "Depth control of autonomous underwater vehicles using indirect robust control method," *International Journal of Control*, vol. 85, no. 1, pp. 98–113, 2012.
[15] Y.-H. Eng, C.-S. Chin, and M. W.-S. Lau, "Added mass computation for control of an open-frame remotely-operated vehicle: Application using wamit and matlab," *Journal of Marine Science And Technology-Taiwan*, vol. 22, pp. 405–416, Aug 2014.
[16] C. S. Chin and S. H. Lum, "Rapid modeling and control systems prototyping of a marine robotic vehicle with model uncertainties using xPC Target system," *Ocean Engineering*, vol. 38, pp. 2128–2141, Dec 2011.
[17] W. He and S. Zhang, "Control design for nonlinear flexible wings of a robotic aircraft," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 1, pp. 351–357, 2017.
[18] T. Zhang and X. Xia, "Adaptive output feedback tracking control of stochastic nonlinear systems with dynamic uncertainties," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 9, pp. 1282–1300, 2015.
[19] J. Li and P. Lee, "Design of an adaptive nonlinear controller for depth control of an autonomous underwater vehicle," *Ocean Engineering*, vol. 32, pp. 2165–2181, 2005.
[20] S. Liu, D. Wang, and E. Pho, "Non-linear output feedback tracking control for AUVs in shallow wave disturbance condition," *International Journal of Control*, vol. 81, no. 11, pp. 930–946, 2008.
[21] W. He, Y. Ouyang, and J. Hong, "Vibration control of a flexible robotic manipulator in the presence of input deadzone," *IEEE Transactions on Industrial Informatics*, 2017. doi: 10.1109/TII.2016.2608739.
[22] S.-L. Dai, M. Wang, C. Wang, and L. Li, "Learning from adaptive neural network output feedback control of uncertain ocean surface ship dynamics," *International Journal of Adaptive Control and Signal Processing*, vol. 28, no. 3–5, pp. 341–365, 2014.
[23] J. Na, Q. Chen, X. Ren, and Y. Guo, "Adaptive prescribed performance motion control of servo mechanisms with friction compensation," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 486–494, 2014.
[24] M. H. Khodayari and S. Balochian, "Modeling and control of autonomous underwater vehicle (AUV) in heading and depth attitude via self-adaptive fuzzy PID controller," *Journal of Marine Science and Technology*, vol. 20, no. 3, pp. 559–578, 2015.
[25] R. P. Kumar, C. S. Kumar, D. Sen, and A. Dasgupta, "Discrete time-delay control of an autonomous underwater vehicle: Theory and experimental results," *Ocean Engineering*, vol. 36, no. 1, pp. 74–81, 2009.
[26] J. Kim, H. Joe, S. C. Yu, J. S. Lee, and M. Kim, "Time-delay controller design for position control of autonomous underwater vehicle under disturbances," *IEEE Transactions on Industrial Electronics*, vol. 63, pp. 1052–1061, Feb 2016.
[27] N. Fischer, D. Hughes, P. Walters, E. M. Schwartz, and W. E. Dixon, "Nonlinear rise-based control of an autonomous underwater vehicle," *IEEE Transactions on Robotics*, vol. 30, pp. 845–852, Aug 2014.
[28] J. Xu, M. Wang, and L. Qiao, "Dynamical sliding mode control for the trajectory tracking of underactuated unmanned underwater vehicles," *Ocean Engineering*, vol. 105, pp. 54 – 63, 2015.
[29] C. L. P. Chen, G. X. Wen, Y. J. Liu, and Z. Liu, "Observer-based adaptive backstepping consensus tracking control for high-order nonlinear semi-strict-feedback multiagent systems," *IEEE Transactions on Cybernetics*, vol. 46, pp. 1591–1601, July 2016.
[30] M. Chen, P. Shi, and C.-C. Lim, "Robust constrained control for mimo nonlinear systems based on disturbance observer," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3281–3286, 2015.
[31] J. Na, X. Ren, and D. Zheng, "Adaptive control for nonlinear pure-feedback systems with high-order sliding mode observer," *IEEE Transactions on Neural Networks and Learning systems*, vol. 24, no. 3, pp. 370–382, 2013.
[32] M. Chen and J. Yu, "Disturbance observer-based adaptive sliding mode control for near-space vehicles," *Nonlinear Dynamics*, vol. 82, no. 4, pp. 1671–1682, 2015.
[33] N. Wang, C. Qian, J.-C. Sun, and Y.-C. Liu, "Adaptive robust finite-time trajectory tracking control of fully actuated marine surface vehicles," vol. 24, no. 4, pp. 1454–1462, 2016.
[34] W. He, Y. Chen, and Z. Yin, "Adaptive neural network control of an uncertain robot with full-state constraints," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 620–629, 2016.
[35] K. Shojaei and M. M. Arefi, "On the neuro-adaptive feedback linearising control of underactuated autonomous underwater vehicles in three-dimensional space," *IET Control Theory and Applications*, vol. 9, pp. 1264–1273, MAY 2015.
[36] B. S. Park, "Neural network-based tracking control of underactuated autonomous underwater vehicles with model uncertainties," *Journal of Dynamic Systems Measurement and Control-Transactions of The ASME*, vol. 137, Feb 2015.
[37] Z. Peng, D. Wang, H. Zhang, and G. Sun, "Distributed neural network control for adaptive synchronization of uncertain dynamical multiagent systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 8, pp. 1508–1519, 2014.
[38] J.-H. Li, P.-M. Lee, and S.-J. Lee, "Motion control of an auv using a neural network adaptive controller," in *Proceedings of the 2002 International Symposium on Underwater Technology, 2002*, pp. 217–221, 2002.
[39] Y.-J. Liu, L. Tang, S. Tong, C. Chen, and D.-J. Li, "Reinforcement learning design-based adaptive tracking control with less learning parameters for nonlinear discrete-time mimo systems," *IEEE Transactions on Neural Networks and Learning Systems,*, vol. 26, no. 1, pp. 165–176, 2015.
[40] W. He, A. O. David, Z. Yin, and C. Sun, "Neural network control of a robotic manipulator with input deadzone and output constraint," *IEEE Transactions on Systems Man and Cybernetics-Systems*, vol. 46, pp. 759–770, Jun 2016.
[41] Z. Peng, D. Wang, Y. Shi, H. Wang, and W. Wang, "Containment control of networked autonomous underwater vehicles with model uncertainty and ocean disturbances guided by multiple leaders," *Information Sciences*, vol. 316, pp. 163–179, 2015.
[42] J. H. Li, P. M. Lee, and B. H. Jun, "Application of a robust adaptive controller to autonomous diving control of an auv," in *30th IEEE Annual Conference on Industrial Electronics Society, 2004. IECON 2004*, vol. 1, pp. 419–424, Nov 2004.
[43] Z. Peng, D. Wang, Z. Chen, and X. Hu, "Adaptive dynamic surface control for formations of autonomous surface vehicles with uncertain dynamics," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 2, pp. 513–520, 2013.
[44] M. Chen and S. S. Ge, "Adaptive neural output feedback control of uncertain nonlinear systems with unknown hysteresis using disturbance observer," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7706–7716, 2015.
[45] Y.-J. Liu and S. Tong, "Adaptive nn tracking control of uncertain nonlinear discrete-time systems with nonaffine dead-zone input," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 497–505, 2015.
[46] Y.-J. Liu, C. P. Chen, G.-X. Wen, and S. Tong, "Adaptive neural output feedback tracking control for a class of uncertain discrete-time nonlinear systems," *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1162–1167, 2011.
[47] W. He, Y. Dong, and C. Sun, "Adaptive neural impedance control of a robotic manipulator with input saturation," *IEEE Transactions on Systems Man and Cybernetics: Systems*, vol. 46, no. 3, pp. 334–344, 2016.
[48] Z. Li, C. Yang, N. Ding, S. Bogdan, and T. Ge, "Robust adaptive motion control for underwater remotely operated vehicles with velocity constraints," *International Journal of Control, Automation and Systems*, vol. 10, no. 2, pp. 421–429, 2012.

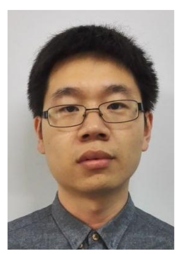
[49] W. He and S. S. Ge, "Cooperative control of a nonuniform gantry crane with constrained tension," *Automatica*, vol. 66, pp. 146–154, 2016.
[50] Y. J. Liu and S. C. Tong, "Barrier lyapunov functions for nussbaum gain adaptive control of full state constrained nonlinear systems," *Automatica*, vol. 76, pp. 143–152, 2017.
[51] F. Rezazadegan, K. Shojaei, F. Sheikholeslam, and A. Chatraei, "A novel approach to 6-DOF adaptive trajectory tracking control of an AUV in the presence of parameter uncertainties," *Ocean Engineering*, vol. 107, pp. 246 – 258, 2015.
[52] T. P. Zhang and S. S. Ge, "Adaptive dynamic surface control of nonlinear systems with unknown dead zone in pure feedback form ," *Automatica*, vol. 44, no. 7, pp. 1895–1903, 2008.
[53] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, no. 1, pp. 237–285, 1996.
[54] P. Stone and R. S. Sutton, "Scaling reinforcement learning toward robocup soccer," in *In Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 537–544, Morgan Kaufmann, 2001.
[55] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *Computer Science*, vol. 8, no. 6, p. A187, 2015.
[56] P. He and S. Jagannathan, "Reinforcement learning neural-network-based controller for nonlinear discrete-time systems with input constraints," *IEEE Transactions on System, Man and Cybernetics, Part B*, vol. 37, no. 2, pp. 425–436, 2007.
[57] B. Xu, C. Yang, and Z. Shi, "Reinforcement learning output feedback NN control using deterministic learning technique," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 635–641, 2014.
[58] R. Cui, C. Yang, Y. Li, and S. Sharma, "Neural network based reinforcement learning control of autonomous underwater vehicles with control input saturation," in *2014 International Conference on Control (UKACC'2014)*, pp. 50–55, 2014.
[59] C. Yang, S. S. Ge, C. Xiang, T. Y. Chai, and T. H. Lee, "Output feedback nn control for two classes of discrete-time systems with unknown control directions in a unified approach," *IEEE Transactions on Neural Networks*, vol. 19, no. 11, pp. 1873–1886, 2008.
[60] S. Ibrir, F. Xie, and Su, "Adaptive tracking of nonlinear systems with non-symmetric dead-zone input," *Automatica*, vol. 43, no. 3, pp. 522–530, 2007.
[61] S. S. Ge, T. H. Lee, and C. J. Harris, *Adaptive Neural Network Control of Robotic Manipulators*. London: World Scientific, 1998.
[62] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive System*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[63] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs.," *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 997–1007, 1997.
[64] J. Si and Y. T. Wang, "On-line learning control by association and reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 264–76, 2001.
[65] P. He and S. Jagannathan, "Discrete-time neural network control of nonlinear systems in non-strict feedback form," in *Proceedings. 42nd IEEE Conference on Decision and Control, 2003.*, vol. 6, pp. 5703–5708, IEEE, 2003.
[66] S. S. Ge, C. Yang, and T. H. Lee, "Adaptive predictive control using neural network for a class of pure-feedback systems in discrete time," *IEEE Transactions on Neural Networks*, vol. 19, pp. 1599–1614, SEP 2008.
[67] R. Cui, S. S. Ge, V. E. B. How, and Y. S. Choo, "Leader-follower formation control of underactuated autonomous underwater vehicles," *Ocean Engineering*, vol. 37, no. 17–18, pp. 1491–1502, 2010.

Rongxin Cui (M'09) received B.Eng. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 2003 and 2008, respectively. From August 2008 to August 2010, he worked as a Research Fellow at the Centre for Offshore Research & Engineering, National University of Singapore, Singapore. Currently, he is a Professor with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China. His current research interests are control of nonlinear systems, cooperative path planning and control for multiple robots, control and navigation for underwater vehicles, and system development.

Dr. Cui servers as an Editor for the Journal of Intelligent & Robotic Systems.



Chenguang Yang (M'10-SM'16) received the B.Eng. degree in measurement and control from Northwestern Polytechnical University, Xi'an, China, in 2005, and the Ph.D. degree in control engineering from the National University of Singapore, Singapore, in 2010. He received postdoctoral training at Imperial College London, UK. He is a senior lecturer with Zienkiewicz Centre for Computational Engineering, Swansea University, UK. His research interests lie in robotics, automation and computational intelligence.



Yang Li received B.Eng. degree from Northwestern Polytechnical University, Xi'an, China, in 2011.

He is currently pursuing the Ph.D. degree at the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China. His research interests are cooperative path planning and control of multiple robots.

Sanjay Sharma is a Reader in the School of Marine Science and Engineering at the University of Plymouth and is Head of the Autonomous Marine Systems (AMS) Research Group. He completed his BTech (Hons) in Electrical Engineering from Indian Institute of Technology, Kharagpur (IITKGP), MTech (Hons) in Control Systems from Institute of Technology- BHU, India and PhD in Control and Systems Engineering from University of Sheffield, UK. He is a member of the IMechEs Marine, Informatics and Control group (MICG) and also a member of the IFAC Technical Committee on Intelligent Autonomous Vehicles. Dr. Sharma has extensive experience in the design, development and application of artificial intelligence techniques and evolutionary algorithms in navigation, guidance and control of marine robotics and unmanned marine craft and is the author of over 70 book, journal and refereed conference publications.