

Communication-Aware Motion Planning of AUV in Obstacle-Dense Environment: A Binocular Vision-Based Deep Learning Method

Jing Yan¹, Senior Member, IEEE, Liang Zhang, Xian Yang¹, Cailian Chen¹, Member, IEEE, and Xinpeng Guan¹, Fellow, IEEE

Abstract—Communication-aware motion planning of autonomous underwater vehicle (AUV) is regarded as an emergent requirement for marine intelligent transportation system. However, the fading acoustic channel and the complex underwater environment make it difficult to realize such task. This paper is concerned with a communication-aware motion planning issue for AUV in obstacle-dense environment. We first develop an intelligent AUV system, which includes binocular cameras for short-distance obstacle avoidance, sonars for long-distance detection, and modems for acoustic communication with buoys. For such system, the parallax angles from AUV to obstacles are utilized to construct an optimal motion planning problem by integrating our previously proposed channel estimation approach. In order to solve the above problem, a deep learning method called depth deterministic policy gradient (DDPG) is developed to minimize the cost function, such that a collision-free path can be planned for AUV while maintaining the communication quality. Note that the advantages of our solution are highlighted as: 1) balance the communication quality and motion stability over the disk model-based methods; 2) improve the collision-avoidance efficiency in path lengths and control efforts as compared with the distance-based methods. Finally, simulation and experimental studies are both provided to verify the effectiveness of our method.

Index Terms—Autonomous underwater vehicle, motion planning, binocular parallax, deep learning, obstacle avoidance.

I. INTRODUCTION

WITH the rapid developments of marine science and artificial intelligence, the research of autonomous underwater vehicle (AUV) has been a compelling topic over the past decades in maritime transportation system. As an

Manuscript received 13 January 2023; revised 5 June 2023; accepted 12 July 2023. Date of publication 25 July 2023; date of current version 29 November 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62222314, Grant 61973263, and Grant 62033011; in part by the Youth Talent Program of Hebei under Grant BJ2020031; in part by the Distinguished Young Foundation of Hebei Province under Grant F2022203001; in part by the Excellent Youth Project for NSF of Hebei Province under Grant F2021203056; and in part by the Central Guidance Local Foundation of Hebei Province under Grant 226Z3201G. The Associate Editor for this article was Y. Li. (Corresponding author: Jing Yan.)

Jing Yan and Liang Zhang are with the Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China (e-mail: jyan@ysu.edu.cn; ysd_x_zl@stunmail.ysu.edu.cn).

Xian Yang is with the Institute of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China (e-mail: xyang@ysu.edu.cn).

Cailian Chen and Xinpeng Guan are with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: cailianchen@sjtu.edu.cn; xpguan@sjtu.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TITS.2023.3296415>, provided by the authors.

Digital Object Identifier 10.1109/TITS.2023.3296415

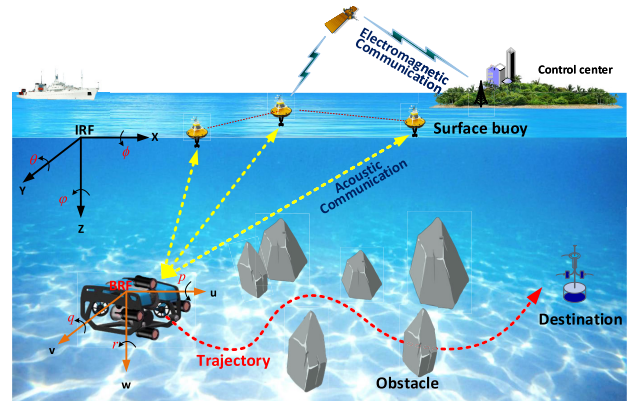


Fig. 1. Description of an AUV communication-aware motion planning system in obstacle-dense environment.

integrated underwater detection and mobility platform, AUV can perform various types of applications, such as subsea exploration [1], pollution detection [2], creature sampling [3], and map construction [4]. In order to understand and explore the ocean, AUV requires to plan a collision-free path from the start point to the destination, as depicted by Fig. 1. Thereby, a fundamental question is given as follows: Given the specific starting/ending positions and the environmental constraints, what is the optimum path of AUV and how to get there?

To answer that question, many motion planning schemes have been developed for AUV. For instance, Cui et al. [5] adopted the random tree star to develop a mutual information-based motion planning strategy for the field sampling of AUV. In [6], an adaptive motion planning strategy was provided to estimate the marine scalar field, where the sparse variation Gaussian process was utilized to capture local measurements of AUV. Yu et al. [7] investigated the path planning of AUV in target traveling missions, through which a shortest path faster algorithm was designed to accomplish task allocation and traveling order determination. Besides that, a target search algorithm was developed in [8] to generate a complete target probability map. The above works are well developed without considering the influences of obstacles in underwater environment. As far as we know, obstacles such as coral reefs, shipwrecks and sea walls are inevitably involved in underwater environment [9]. Ignoring the influence of obstacle can lead to the failure of AUV motion planning. Therefore, it is necessary to incorporate the obstacle avoidance into the motion planning of AUV in complex underwater environment.

Considering the influence of obstacle, the collision-free motion planning methods can be classified into two major categories: distance-based solutions and vision-based solutions. Particularly, the distance-based solutions collect relative distances between AUV and obstacle by using on-board active/passive sonar [10], [11]. Meanwhile, the vision-based solutions collect the depth images of obstacles through camera stereo calibration [12], such that the gray values of depth images can be used to calculate the relative distances between AUV and obstacle. Of note, the former category is traditionally the primary way for long-distance detection of AUV. However, sonar has low resolution ability due to the underwater reverberation noise and multi-path fading effect [13], which may lead to the performance degradation of collision avoidance, especially in obstacle-dense environments. For example, the side-scan sonar with a low frequency can only have low-quality acoustic data, due to the speckle noises caused by acoustic wave interferences [14], [15]. Alternatively, the latter category has the characteristics of high resolution ability and low cost, which is very suitable for the short-distance and obstacle-dense underwater environments. In view of this, we focus on the vision-based motion planning issue of AUV. More recently, some vision-based obstacle avoidance algorithms have been developed for AUV. For instance, a multi-camera collision-avoidance system was constructed in [16], and then, a repulsive potential function was designed to move AUV away from collisions. In [17], a vision-based A-star algorithm was developed for AUV to predict the regions that could lead to collisions, and hence, a set of warning signals were generated to allow AUV to perform evasive maneuvers. Nevertheless, the potential function or A-star based methods inevitably face local optimal and sensitivity to initial values. To remedy the local optimal issue, a reinforcement learning based motion planning strategy was developed in [18], whose aim was to output collision-free trajectories for AUV by integrating the convolutional neural network. In [19], the double Q learning and convolutional neural network were jointly employed to perform vision-based collision-avoidance navigation for AUV. In [20], a double deep Q learning-based global path planning strategy was developed for unmanned surface vehicle. However, the vehicle dynamics in [18] and [19] were simplified as the first-order linear differential equation. As has been pointed out in [21], AUV dynamics are with multiple freedom nonlinear property. Moreover, the double Q learning network cannot be directly applied to continuous state space since it finds action that maximizes the action-value function [22], and hence, it suffers from the difficulty of handling continuous underwater complex environment. Based on this, some behavior rule-based collision-avoidance algorithms have been proposed in [23] and [24] to resolve the collision-avoidance navigation of multiple freedom AUV, but it is not an easy task to design the behavior rules, especially for obstacle-dense environment. With regard to this, how to employ the vision measurements to design a learning-based collision-avoidance algorithm for multiple freedom AUV is still an unsolved issue.

Beyond that, we notice that the communication between AUV and surface buoy plays an important role in motion planning, since AUV requires to send its local measurement

TABLE I
COMPARISON WITH THE OTHER LITERATURES

Performance	Communication		Obstacle avoidance	Vision solution
	quality			
Refs. [5]–[8], [14], [15]	×	×	×	×
Refs. [10], [11], [20]	×	√	×	×
Refs. [12], [16]–[19]	×	√	√	√
Refs. [21], [26]–[30], [32]	√	×	×	×
Refs. [25], [31]	√	√	√	×
This paper	√	√	√	√

to control center via the relay of buoy. Nowadays, most of the existing works adopt the approximate deterministic disk model to characterize underwater channel quality [25], [26]. Specifically, it is assumed that the channel quality is ideal within a predefined communication range, otherwise, it is invalid. The above assumption is reasonable for the motion planning of terrestrial vehicle, however it is too realistic for AUV, because underwater acoustic communication is often affected by shadowing effect and path loss with geographic position [27]. For that reason, it is necessary to integrate the underwater acoustic communication quality into the motion planning process. To this end, a motion-planning approach by integrating stochastic channel estimation framework was proposed in [28] to achieve collision-free motion planning for robot. Followed by this, a communication-aware and energy-efficient motion planning algorithm was developed in [29], which can capture the realistic channel property of robot. In [30], the co-optimization of communication and motion of robot was performed under energy constraints. Nevertheless, the communication-aware motion planning algorithms in [28], [29], and [30] are not developed in the context of AUV. In our previous works [31], [32], the communication-aware motion planning algorithms were designed for AUV in obstacle environment, however they adopt the distance-based measurement to perform collision-avoidance maneuvers. Note that the distance-based solution ignores the dimensions of AUV and obstacles, which is not suitable for AUV in obstacle-dense environment. With consideration of the obstacle-dense underwater environment, it is necessary to design a vision-based motion planning method for AUV that jointly considers the communication quality in cyber side and complex AUV dynamics in physical side.

This paper develops a communication-aware motion planning solution for AUV in obstacle-dense environment. The difficulties in attacking the aforementioned issues can be concluded into two points: 1) how to sense the parallax angle of multiple obstacles by using a binocular camera? 2) how to drive AUV to the destination with collision avoidance and communication quality maintenance? In view of this, an intelligent AUV system is developed to enable it have obstacle avoidance and control abilities. Then, a depth deterministic policy gradient (DDPG) based motion planning algorithm is designed to move AUV away from obstacles and toward to destination position. The comparison with the other existing literatures is provided in Table I, through which the main features of this paper are summarized as follows.

1) *Co-optimization framework of communication and motion of AUV.* By integrating our previously proposed

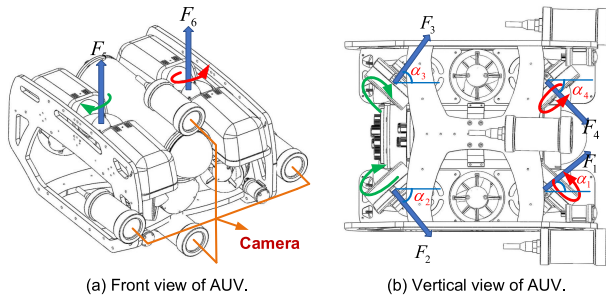


Fig. 2. Mechanical structure of AUV with modular concepts.

channel estimation approaches [32], a co-optimization framework is developed for AUV in obstacle-dense environment, which includes AUV dynamics, vision measurement, communication channel quality and destination position. Compared with the disk model-based methods in [25] and [26], the co-optimization framework in this paper can balance the communication quality and motion stability. Meanwhile, the four freedom motion model can well capture the dynamics of AUV.

2) *DDPG-based obstacle avoidance algorithm*. The parallax angles from AUV to obstacles are employed to design a DDPG-based obstacle avoidance algorithm. By considering the sizes of obstacle and AUV, it can effectively reflect the threat level of obstacles to AUV. Compared with the distance-based obstacle avoidance algorithms in [10], [11], [31], and [32], the vision-based solution in this paper can improve the collision-avoidance efficiency in path lengths and control efforts.

3) *Experimental test via the co-design of communication and control*. It is important to note that, most of the vision and deep learning based motion planning algorithms for AUV are verified by simulation results. With a viewpoint to the practical application, it is necessary and meaningful to check the effectiveness through experimental results. To this end, we design an intelligent AUV system, and more importantly, the theory results of this paper are verified in water pool.

This paper is organized as follows. Section II presents the system model and problem formulation. Section III gives the binocular vision-based path planning solution. Simulation and experimental results are performed in Section IV. Section V demonstrates the conclusions and future works.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

In order to perform communication-aware motion planning, AUV makes direct acoustic communication with surface buoys, and at the same time it plans collision-free path to the destination (see Fig. 1). The mechanical structure of AUV with modular concepts is presented by Fig. 2, which consists of a buoyancy shell, six thrusters, two pairs of binocular cameras, a communication unit, an electrical power system, sonars and control cabin. The thrusters are fixed on the AUV frame to ensure AUV can move as freely as possible. Two pairs of binocular cameras are fixed vertically and horizontally on the front of AUV, which can achieve obstacle identification and parallax angle calculation tasks. The role of sonars is

to make long-distance detection. The yaw angle of AUV is measured by gyroscope. In addition, the horizontal position information of AUV is measured by the ultra-short baseline (USBL) system from Waterlinked [33], while the depth information is measured by a depth gauge. The communication unit ensures the effective data transmission among different devices, including the acoustic communication with buoys and the electromagnetic communication with control center.

The position and orientation vector for AUV in inertial reference frame (IRF) is defined as $\eta = [x, y, z, \psi]^T$, where x , y and z denote the positions on X, Y and Z axes, respectively. In addition, ψ is the angle on yaw. The linear and angle velocity vector in body-fixed reference frame (BRF) is expressed as $\mathbf{v} = [u, v, w, r]^T$, where u , v and w are the linear velocities on surge, sway and heave, respectively. r denotes the angle velocity on yaw. Referring to [34] and [35], the motion model of AUV can be given as

$$\begin{aligned} \dot{\eta} &= \mathbf{J}(\eta)\mathbf{v}, \\ \mathbf{M}\dot{\mathbf{v}} + \mathbf{C}(\mathbf{v})\mathbf{v} + \mathbf{D}(\mathbf{v})\mathbf{v} + \mathbf{g}(\eta) &= \boldsymbol{\tau}, \end{aligned} \quad (1)$$

where $\mathbf{M} = \text{diag}(m_u, m_v, m_w, I_r)$ is the inertia matrix. m_u , m_v and m_w represent the masses on surge, sway and heave respectively, I_r is the moment of inertia on yaw. $\mathbf{C}(\mathbf{v}) = [0, 0, 0, -m_v v; 0, 0, 0, m_u u; 0, 0, 0, 0; m_v v, -m_u u, 0, 0]$ is the Coriolis centripetal matrix. $\mathbf{D}(\mathbf{v}) = \text{diag}(k_u + k_{u|u}|u|, k_v + k_{v|v}|v|, k_w + k_{w|w}|w|, k_r + k_{r|r}|r|)$ is the kinematic coefficient damping matrix, where k_u, k_v, k_w and k_r are the linear damping scales on surge, sway, heave and yaw, respectively. $k_{u|u}, k_{v|v}, k_{w|w}$ and $k_{r|r}$ are the quadratic damping scales. $\mathbf{g}(\eta) \in \mathcal{R}^4$ is the hydrostatic force vector. $\mathbf{J} = [\cos \psi, -\sin \psi, 0, 0; \sin \psi, \cos \psi, 0, 0; 0, 0, 1, 0; 0, 0, 0, 1]$ is the rotation matrix. Meanwhile, $\boldsymbol{\tau} = [\tau_u, \tau_v, \tau_w, \tau_r]^T$ is the control input vector, which is required to be designed. Of note, τ_u, τ_v, τ_w and τ_r are the control inputs on surge, sway, heave and yaw, respectively.

Remark 1: From the perspective of application, the control input vector $\boldsymbol{\tau}$ is generated by the motions of thrusters. In view of this, we take the case of six thrusters as an example. Then the relationship between control input and thruster motion can be described as $\boldsymbol{\tau} = \mathbf{H}\mathbf{F}$, where $\mathbf{F} = [F_1, F_2, F_3, F_4, F_5, F_6]^T$ is the force vector generated of thrusters (see Fig. 2). Meanwhile, \mathbf{H} is the transformation matrix, i.e.,

$$\mathbf{H} = \begin{bmatrix} -\cos \alpha_1 & -\cos \alpha_2 & -\cos \alpha_3 & -\cos \alpha_4 & 0 & 0 \\ \sin \alpha_1 & -\sin \alpha_2 & \sin \alpha_3 & -\sin \alpha_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ H_{41} & H_{42} & H_{43} & H_{44} & 0 & 0 \end{bmatrix},$$

where α_i is the angle of F_i with X axis for $i \in \{1, 2, 3, 4\}$. Particularly, $H_{41} = -\sin \alpha_1 - \cos \alpha_1$, $H_{42} = -\sin \alpha_2 - \cos \alpha_2$, $H_{43} = \sin \alpha_3 + \cos \alpha_3$, and $H_{44} = \sin \alpha_4 + \cos \alpha_4$.

As has been mentioned above, AUV needs to maintain the communication connection with surface buoys. Along with this, we note that the connectivity can be evaluated by the received signal noise ratio (SNR) [29]. In view of this, we use the SNR between AUV and buoy to characterize the communication performance, as depicted by Fig. 3. Let $\mathbf{P}_A = [x, y, z]^T$ denote the position vector of AUV, and $\mathbf{P}_{B,j} = [x_j, y_j, z_j]^T$ be the position vector of buoy $j \in \{1, \dots, m_B\}$, where m_B

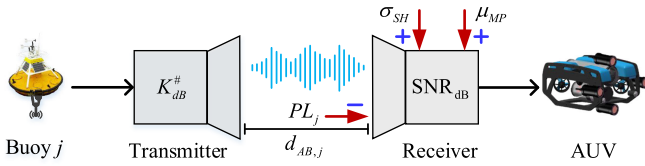


Fig. 3. Description of the SNR between AUV and buoy j , where $PL_j = 10k^\# \log_{10}(d_{AB,j}) + 10d_{AB,j} \log_{10}(\alpha(f)) + N_{dB}^0$.

represents the number of buoys. Referring to [36], the SNR between AUV and buoy j can be defined as

$$\begin{aligned} \text{SNR}_{dB}(\mathbf{P}_A, \mathbf{P}_{B,j}) &= K_{dB}^\# - 10k^\# \log_{10}(d_{AB,j}) - 10d_{AB,j} \log_{10}(\alpha(f)) \\ &\quad + \sigma_{SH}(\mathbf{P}_A, \mathbf{P}_{B,j}) + \mu_{MP}(\mathbf{P}_A, \mathbf{P}_{B,j}) - N_{dB}^0, \end{aligned} \quad (2)$$

where $K_{dB}^\#$ is the average energy consumption of transmitting one bit data in dB, $k^\#$ is the spreading coefficient, $d_{AB,j} = \|\mathbf{P}_A - \mathbf{P}_{B,j}\|$ denotes the relative distance between AUV and buoy j , $\alpha(f)$ is the acoustic absorption at frequency f , and N_{dB}^0 is the noise power spectral density in dB. Besides that, σ_{SH} and μ_{MP} are random parameters, reflecting the shadow fading and multipath fading effects, respectively.

Note that the random parameters σ_{SH} and μ_{MP} in (2) are unknown to AUV. In our previous work [31], [32], an integral reinforcement learning-based estimator was developed to capture the unknown shadowing and multipath parameters. For that reason, we employ our previous channel estimation approach to predict the SNR. Then, the predicted mean of SNR between AUV and buoy j is denoted by

$$\begin{aligned} \overrightarrow{\text{SNR}}_{dB}(\mathbf{P}_A, \mathbf{P}_{B,j}) &= \mathbf{h}\theta^* - \varepsilon + \Xi^T \Omega^{-1} (Y_{dB} - H\theta^* + \varepsilon), \end{aligned} \quad (3)$$

with

$$\mathbf{h} = [1, -10 \log_{10}(\|\mathbf{P}_A - \mathbf{P}_{B,j}\|)], \quad (4)$$

$$\varepsilon = N_{dB}^0 + 10 \log_{10}(\alpha(f)), \quad (5)$$

where θ^* denotes the estimated channel parameter. In addition, Ξ represents the basis function, H is location-related parameter, and Ω is the variance of measurement Y_{dB} .

In the following, we consider the obstacles-dense underwater environment. Generally, the off-line map of AUV monitoring region can be known by the remote sensing technique. Meanwhile, AUV is capable of online sensing obstacles by using the onboard cameras. For the above consideration, the obstacles can be covered by convex columns or spheres. Particularly, we define the ratio of length to width for obstacle as κ , and hence, the following two cases are considered: 1) If $\kappa \in [1 - \delta_\#, 1 + \delta_\#]$, we regard the obstacle as a convex sphere, where $\delta_\#$ is a positive decimal according to the requirement of the system; 2) otherwise, the obstacle is regarded as a convex column. Based on this, the m th sphere obstacle in the environment is denoted by $S_m(O_{sm}, \rho_{sm})$, where O_{sm} is its center and $\rho_{sm}(\rho_{sm} > 0)$ is its radius for $m = 1, 2, \dots$. On the other hand, the n th column obstacle is denoted by $C_n(O_{cn}, \rho_{cn}, h_{cn})$, where O_{cn} is its center, $\rho_{cn}(\rho_{cn} > 0)$ is its

radius, and h_{cn} is its height for $n = 1, 2, \dots$. In summary, the following obstacle definition is provided.

Definition 1 (Obstacle Set): For AUV at time step k , the set of detected obstacles within its sensing range can be defined as a subset $\Omega_k \subset \{S_1(O_{s1}, \rho_{s1}), \dots, S_m(O_{sm}, \rho_{sm}), \dots; C_1(O_{c1}, \rho_{c1}, h_{c1}), \dots, C_n(O_{cn}, \rho_{cn}, h_{cn}), \dots\}$.

B. Problem Formulation

The problems of communication-aware motion planning of AUV in obstacle-dense environment is detailed as follows.

Problem 1 (Co-optimization of Communication and Motion): Although communication and control usually depend on each other, they are often designed independently. For this reason, we attempt to co-optimize the communication and motion qualities for AUV. This problem is reduced to design a co-optimization framework for AUV, with considerations of AUV model in (1), SNR prediction in (3) and obstacle set Ω_k .

Problem 2 (Obstacle Avoidance With Vision Measurements): The obstacle-dense environment makes it difficult to achieve collision-free motion planning for AUV by means of distance-based measurements. In view of this, it is necessary to adopt the vision information to plan paths for AUV. This problem is reduced to design a DDPG-based obstacle avoidance algorithm for AUV by using the on-board binocular cameras.

III. MAIN RESULTS

We first construct a co-optimization framework of communication and motion planning for AUV. Then, a DDPG-based obstacle avoidance algorithm is developed. Finally, the performance analysis of our solution is presented.

A. Co-Optimization of Communication and Motion

For ease of illustration, the state of AUV is defined as $\mathbf{X} = [\eta, \mathbf{v}]^T$, and the sampling interval is $\delta \in \mathcal{R}^+$. With Taylor expansion, model (1) at time step k is rearranged as

$$\mathbf{X}(k+1) = f(\mathbf{X}(k)) + h(\mathbf{X}(k))\tau(k), \quad (6)$$

with

$$\begin{aligned} f(\mathbf{X}) &= \begin{bmatrix} \eta + \delta \mathbf{J}(\eta) \mathbf{v} \\ \mathbf{v} - \delta \mathbf{M}^{-1} [\mathbf{C}(\mathbf{v}) \mathbf{v} + \mathbf{D}(\mathbf{v}) \mathbf{v} + \mathbf{g}(\eta)] \end{bmatrix}, \\ h(\mathbf{X}) &= [0_{4 \times 4}; \delta \mathbf{M}^{-1}]^T. \end{aligned}$$

Correspondingly, the state of destination is defined as $\bar{\mathbf{X}}_{\text{target}} = [x_g, y_g, z_g, 0, 0, 0, 0]^T$, where x_g , y_g and z_g indicate the position coordinates of destination. It is assumed that $\bar{\mathbf{X}}_{\text{target}}$ is decided by control center, which can be sent to AUV by the data relay of surface buoys. Note that the learning aim of DDPG is to obtain the maximum cumulative reward function after the motion planning procedure. Then, the motion planning of AUV includes three parts, i.e., destination tracking, obstacle avoidance and communication maintenance.

To steer AUV towards the destination, the reward function of destination tracking is set as

$$r_{\text{track}}(k) = - \|\mathbf{X}(k) - \bar{\mathbf{X}}_{\text{target}}\|. \quad (7)$$

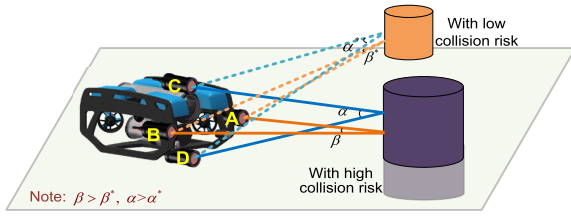


Fig. 4. The schematic diagram of binocular parallax detection.

Next, a vision-based collision-avoidance strategy is provided to construct the reward function of obstacle avoidance. We note that AUV perceives the distance from an obstacle with parallax, i.e., the angle across AUV's views from two different observation points to obstacle. When the parallax is large, it means AUV may collide with the obstacle. Otherwise, it means AUV has low risk of colliding with obstacle, as depicted by Fig. 4. Thereby, the objective of vision-based reward function is divided into two categories: 1) horizontal binocular cameras measure the horizontal parallax of obstacle, and then the reward function on horizontal direction can be defined; 2) vertical binocular cameras acquire the vertical parallax, whose aim is to define the reward function on vertical direction.

To this end, we suppose obstacle $k_{\#} \in \Omega_k$ is detected by AUV at step k . It is worth noting that we can adopt active or passive sonars to sense obstacle, through which detection procedure can be conducted to detect the existence of obstacle. As the sonar detection is not the focus of this paper, and hence, the detailed procedure of sonar detection is omitted here. Readers can refer to [37] and [38] for the detailed sonar detection process. In this way, the following steps are conducted to develop the collision-avoidance reward function.

Step 1 (Discretize the Shape of Obstacle): We first discretize the shape of obstacle $k_{\#}$. If $\kappa \in [1 - \delta_{\#}, 1 + \delta_{\#}]$, we regard obstacle $k_{\#}$ as a convex sphere, otherwise it is a convex column. Obstacle surface is discretized equally, and these discretized spatial points are mapped to the spheres or columns via orthographic projection. As shown in Fig. 5, we take a convex column as an example, where the mapped spatial point in convex column for the $i \in (1, 2, \dots)$ row and the $j \in (1, 2, \dots)$ column is denoted by P_{ij} .

Step 2 (Acquire Obstacle's Position in Pixel Coordinate): Based on the image formation principle [39], the point P_{ij} is projected onto the camera's photoreceptor by the camera lens. In the following, we transform point P_{ij} from the world coordinate system to the image coordinate system. In this way, the operations of rigid body transformation, transmission projection and matrix transformation are conducted. Based on this, the pixel coordinates of point P_{ij} with respect to cameras A, B, C and D can be respectively set as

$$\mathbf{P}_{A,i,j} = [X_{A,i,j}, Y_{A,i,j}]^T, \quad \mathbf{P}_{B,i,j} = [X_{B,i,j}, Y_{B,i,j}]^T, \quad (8)$$

$$\mathbf{P}_{C,i,j} = [X_{C,i,j}, Y_{C,i,j}]^T, \quad \mathbf{P}_{D,i,j} = [X_{D,i,j}, Y_{D,i,j}]^T, \quad (9)$$

where $X_{\Delta,i,j}$ is the horizontal coordinate of point P_{ij} in the pixel coordinate system for camera $\Delta \in \{A, B, C, D\}$, Similarly, $Y_{\Delta,i,j}$ is the vertical coordinate of point P_{ij} in the pixel coordinate system for camera $\Delta \in \{A, B, C, D\}$.

Step 3 (Calculate the Pixel Difference of Obstacle): The parallax angle can be indirectly obtained by the pixel difference. Thereby, we define X_o and Y_o as the horizontal and vertical coordinates of the center point on image coordinate system, respectively. Regarding X_o and Y_o as the benchmarks, one has the following pixel differences, i.e.,

$$\begin{aligned} d_{Ax,i,j} &= |X_{A,i,j} - X_o|, & d_{Ay,i,j} &= |Y_{A,i,j} - Y_o|, \\ d_{Bx,i,j} &= |X_{B,i,j} - X_o|, & d_{By,i,j} &= |Y_{B,i,j} - Y_o|, \\ d_{Cx,i,j} &= |X_{C,i,j} - X_o|, & d_{Cy,i,j} &= |Y_{C,i,j} - Y_o|, \\ d_{Dx,i,j} &= |X_{D,i,j} - X_o|, & d_{Dy,i,j} &= |Y_{D,i,j} - Y_o|, \end{aligned} \quad (10)$$

where $d_{\Delta x,i,j}$ denotes the pixel difference on horizontal direction for camera $\Delta \in \{A, B, C, D\}$. Meanwhile, $d_{\Delta y,i,j}$ denotes the pixel differences on vertical direction.

The relative distances between camera's optical center and projection point are denoted by $f_{A,i,j}$, $f_{B,i,j}$, $f_{C,i,j}$ and $f_{D,i,j}$. From (10), the above relative distances are given as

$$f_{A,i,j} = \sqrt{f_A^2 + d_{Ay,i,j}^2}, \quad f_{B,i,j} = \sqrt{f_B^2 + d_{By,i,j}^2}, \quad (11)$$

$$f_{C,i,j} = \sqrt{f_C^2 + d_{Cx,i,j}^2}, \quad f_{D,i,j} = \sqrt{f_D^2 + d_{Dx,i,j}^2}, \quad (12)$$

where f_{Δ} is the focal length for camera $\Delta \in \{A, B, C, D\}$.

Step 4 (Design of Parallax-Based Reward Function): Let $\theta_{\Delta,i,j}$ denote the angle between the line of camera $\Delta \in \{A, B, C, D\}$ optical center and the line of binocular camera baseline, as depicted by Fig. 5. Noting with orthogonality and parallelism, the above angles can be calculated by

$$\theta_{A,i,j} = \begin{cases} \arctan(f_{A,i,j}/d_{Ax,i,j}) & \text{if } d_{Ax,i,j} \neq 0 \\ \pi/2 & \text{else} \end{cases} \quad (13)$$

$$\theta_{B,i,j} = \begin{cases} \arctan(f_{B,i,j}/d_{Bx,i,j}) & \text{if } d_{Bx,i,j} \neq 0 \\ \pi/2 & \text{else} \end{cases} \quad (14)$$

$$\theta_{C,i,j} = \begin{cases} \arctan(f_{C,i,j}/d_{Cy,i,j}) & \text{if } d_{Cy,i,j} \neq 0 \\ \pi/2 & \text{else} \end{cases} \quad (15)$$

$$\theta_{D,i,j} = \begin{cases} \arctan(f_{D,i,j}/d_{Dy,i,j}) & \text{if } d_{Dy,i,j} \neq 0 \\ \pi/2 & \text{else} \end{cases} \quad (16)$$

Based on (13), (14), (15) and (16), the horizontal parallax angle $\theta_{H,i,j}$ and vertical parallax angle $\theta_{V,i,j}$ are

$$\begin{aligned} \theta_{H,i,j} &= \begin{cases} \pi - P_H^{\theta}, & \text{if } 2 \tan^{-1} \frac{f_A}{X_o} \leq P_H^{\theta} \leq F(B_H) \\ 0 & \text{else} \end{cases} \\ \theta_{V,i,j} &= \begin{cases} \pi - P_V^{\theta}, & \text{if } 2 \tan^{-1} \frac{f_C}{Y_o} \leq P_V^{\theta} \leq F(B_V) \\ 0 & \text{else} \end{cases} \end{aligned} \quad (17)$$

where $P_H^{\theta} = \theta_{A,i,j} + \theta_{B,i,j}$, $P_V^{\theta} = \theta_{C,i,j} + \theta_{D,i,j}$, $F(B_H) = \frac{\pi}{2} + \tan^{-1} \frac{f_A}{X_o} + \tan^{-1} \frac{X_o \Upsilon - f_A B_H}{f_A \Upsilon}$, and $F(B_V) = \frac{\pi}{2} + \tan^{-1} \frac{f_C}{Y_o} + \tan^{-1} \frac{Y_o \Upsilon - f_C B_V}{f_C \Upsilon}$. Of note, B_H and B_V represent the baseline lengths of horizontal and vertical binocular cameras, respectively. In addition, Υ is the maximum visible distance of the camera, which is a constant related to the current underwater environment. The detailed design procedure of the thresholds (i.e., P_H^{θ} and P_V^{θ}) is provided in Corollary 2.

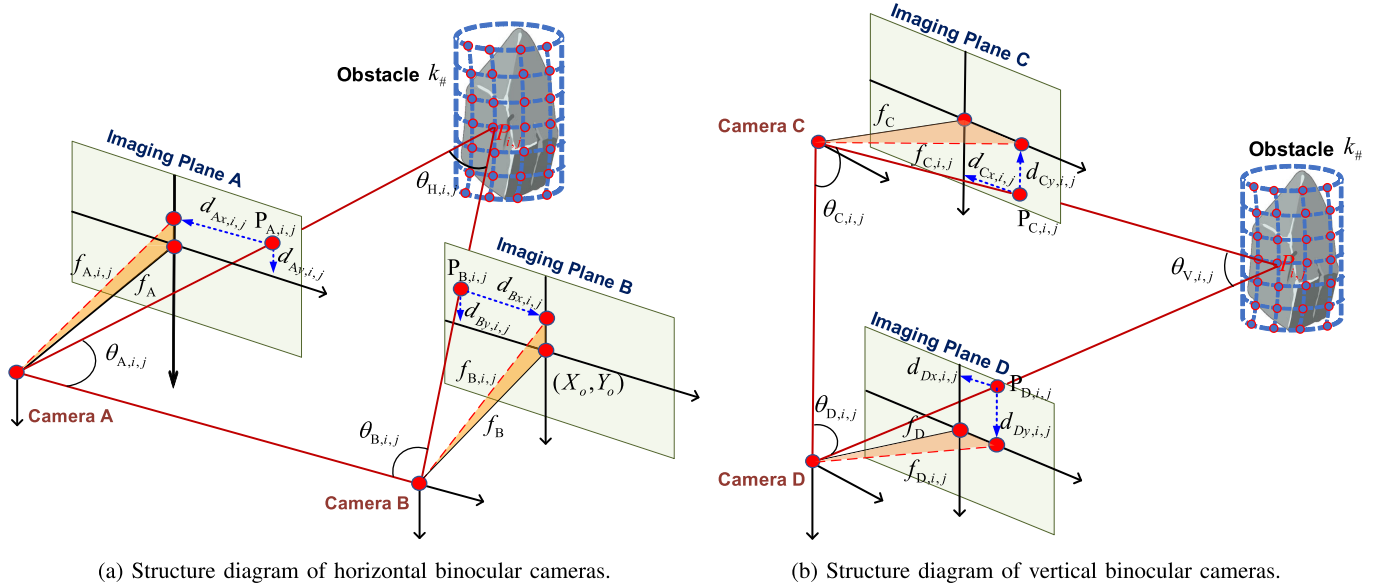


Fig. 5. Schematic diagram of the horizontal and vertical binocular cameras for obstacle $k_{\#} \in \Omega_k$.

Calculate the parallax angles for the other spatial points, and hence, the maximum horizontal/vertical parallax angles for obstacle $k_{\#}$ are represented as $\theta_{H,k_{\#}} = \max_{i,j} \theta_{H,i,j}$ and $\theta_{V,k_{\#}} = \max_{i,j} \theta_{V,i,j}$. Applying the above result to the other obstacle within the sensing range, we can easily get the parallax-based reward function, i.e., at time step k

$$r_{\text{avoid}}(k) = - \sum_{k_{\#} \in \Omega_k} (\theta_{H,k_{\#}} + \theta_{V,k_{\#}}). \quad (18)$$

Next, the reward function of communication efficiency is designed. Particularly, we measure the underwater communication quality by calculating the SNR between AUV and buoys. In order to guarantee the communication connection with surface buoys, AUV should maintain a minimum required SNR with at least one buoy. For that consideration, a threshold of the minimum required SNR is denoted as SNR_{th} , which is associated with the on-board communication equipment in AUV. If $\text{SNR}_{\text{dB}}(\mathbf{P}_A, \mathbf{P}_{B,j}) \geq \text{SNR}_{\text{th}}$, then AUV can communicate with buoy $j \in \{1, \dots, m_B\}$ while satisfying the quality of service requirement, i.e., the communication link is connected. Otherwise, the communication link is regarded to be disconnected. Then, we define the following rules:

1) If the maximum SNR received by AUV is less than SNR_{th} , then the communication quality of AUV at time step k is poor. That means AUV needs to enlarge its reward, such that the communication quality can be improved;

2) If the maximum SNR received by AUV is larger than SNR_{th} , then the communication quality of AUV at step k is good. So, AUV needs to keep its current reward, such that the current communication quality can be maintained.

Our previous channel estimation approach [31], [32] is conducted to predict the SNR between AUV and buoy $j \in \{1, \dots, m_B\}$. From (3), the maximum SNR received by AUV at time step k is defined as $\overrightarrow{\text{SNR}}_{\text{max}}(k) = \max_{j \in \{1, \dots, m_B\}} \text{SNR}_{\text{dB}}(\mathbf{X}(k), \mathbf{P}_{B,j})$. Accordingly, the reward

function of communication efficiency is constructed as

$$r_{\text{snr}}(k) = \begin{cases} 0 & \text{if } \overrightarrow{\text{SNR}}_{\text{max}}(k) \geq \text{SNR}_{\text{th}} \\ -\frac{1}{10^{\overrightarrow{\text{SNR}}_{\text{max}}(k)/10}} & \text{else} \end{cases} \\ = \frac{0.5 \text{sign}(\overrightarrow{\text{SNR}}_{\text{max}}(k) - \text{SNR}_{\text{th}}) - 0.5}{10^{(\overrightarrow{\text{SNR}}_{\text{max}}(k)/10)}, \quad (19)$$

where $\text{sign}(\cdot)$ denotes the sign function.

From (7), (18) and (19), the total reward function for AUV communication and control can be given as

$$R(\mathbf{X}(k)) = -K_1 \|\mathbf{X}(k) - \bar{\mathbf{X}}_{\text{target}}\| - K_2 \sum_{k_{\#} \in \Omega_k} (\theta_{H,k_{\#}} + \theta_{V,k_{\#}}) \\ - K_3 \frac{0.5 - 0.5 \text{sign}(\overrightarrow{\text{SNR}}_{\text{max}}(k) - \text{SNR}_{\text{th}})}{10^{(\overrightarrow{\text{SNR}}_{\text{max}}(k)/10)}, \quad (20)$$

where K_1 , K_2 and K_3 are positive weighting factors. Of note, the roles of K_1 , K_2 and K_3 are to balance the tracking, obstacle avoidance and communication efficiency.

Based on (20), the value function is defined as

$$Q(\mathbf{X}(k)) = E \left[\sum_{l=k}^{\infty} \gamma^{l-k} \cdot R(\mathbf{X}(l)) \right], \\ = E [R(\mathbf{X}(k)) + \gamma Q(\mathbf{X}(k+1))], \quad (21)$$

where $\gamma \in (0, 1]$ represents the discount factor, and its role is to reflect the impact of the future state on the present.

Hence, the optimization of (21) is to select $\tau^*(k)$, such that an optimal update policy of $\tau(k)$ can be obtained, i.e.,

$$\tau^*(k) = \underset{\tau}{\text{argmax}} (E[R(\mathbf{X}(k)) + \gamma Q(\mathbf{X}(k+1))]). \quad (22)$$

B. DDPG-Based Obstacle Avoidance Algorithm

This section designs a DDPG-based obstacle avoidance algorithm, such that the optimal solution $\tau^*(k)$ in (22) can be

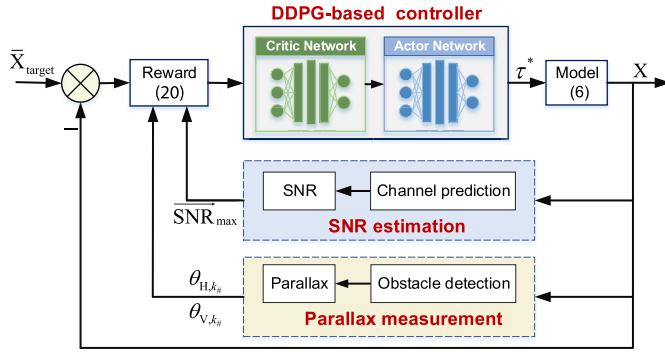


Fig. 6. Main framework of the motion planning solution.

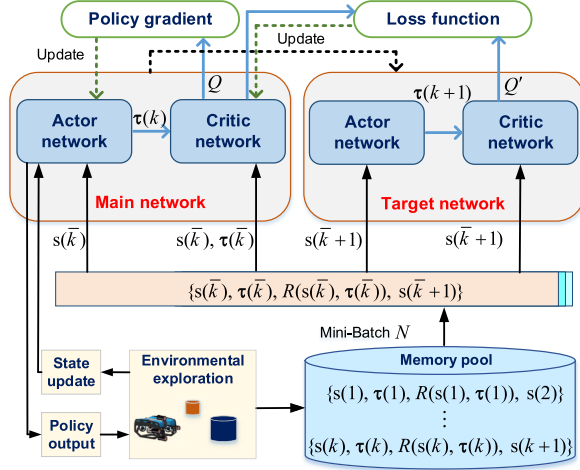


Fig. 7. Construction of DDPG-based obstacle avoidance algorithm.

derived. It is noted that DDPG is a deep deterministic policy gradient algorithm with deterministic output actions [40], [41]. Compared with the random policies, it tends to require fewer samples for gradient estimation. Besides that, it can solve the action space continuity problem, which is lacked in the classical deep Q-network (DQN) algorithm. For this purpose, we consider the dynamic model (6), whose state and action spaces are set as follows. 1) *State*: at time step k , the state space of AUV is described as $\mathbf{s}(k) = \{\mathbf{X}(k), \theta_{H,k\#}(k), \theta_{V,k\#}(k), \overline{\text{SNR}}_{\max}(k)\}$; 2) *Action*: the control input vector $\tau(k)$ is regarded as the action space of AUV. Particularly, the main framework of the motion planning solution for AUV is described by Fig. 6.

The DDPG-based obstacle avoidance algorithm includes the main network and the target network. Specifically, the main network is consist of actor network $\pi(\mathbf{s}(k)|\vartheta^\pi)$ and critic network $Q(\mathbf{s}(k), \tau(k)|w^Q)$, where the role of actor network is to select the current action $\tau(k)$ based on the current state $\mathbf{s}(k)$. Meanwhile, the role of critic network is to evaluate the value of actions. On the other hand, the target network is consist of the target actor network $\pi'(\mathbf{s}(k)|\vartheta^{\pi'})$ and the target critic network $Q'(\mathbf{s}(k), \tau(k)|w^{Q'})$. Of note, the role of $\pi'(\mathbf{s}(k)|\vartheta^{\pi'})$ is to select the next action $\tau(k+1)$ based on the next state $\mathbf{s}(k+1)$ via the memory pool sampling data, while the role of $Q'(\mathbf{s}(k), \tau(k)|w^{Q'})$ is to fit the target value function for subsequent network update. As shown in Fig. 7, the following steps are conducted to seek $\tau^*(k)$.

Step 1 (Environmental Experience Collection): Based on the current state $\mathbf{X}(k)$, one obtains the state set $\mathbf{s}(k)$, i.e., the binocular parallax angles $\theta_{H,k\#}$ and $\theta_{V,k\#}$ are acquired by (17) while $\overline{\text{SNR}}_{\max}(k)$ is estimated by (3). Subsequently, a random exploration action is generated, i.e.,

$$\tau(k) = \pi(\mathbf{s}(k)|\vartheta^\pi) + N^\#, \quad (23)$$

where $\pi(\mathbf{s}(k)|\vartheta^\pi)$ is the mapping output of actor network under state $\mathbf{s}(k)$ and weight ϑ^π . Besides, $N^\#$ is the Ornstein Uhlenbeck (OU) noise, obeying the Gaussian distribution.

From (3) and noting with $\mathbf{s}(k)$, one can easily obtain the reward $R(\mathbf{s}(k), \tau(k))$ by (20). In addition, the state space of AUV at time step $k+1$ can be obtained through (6), (17) and (3). Accordingly, the collected environmental experience at time step k is depicted as

$$\{\mathbf{s}(k_*), \tau(k_*), R(\mathbf{s}(k_*), \tau(k_*)), \mathbf{s}(k_*+1)\}_{k_*=1}^k. \quad (24)$$

Repeating the above procedure with multiple episodes, one can easily have the total environmental experience for AUV.

Step 2 (Training the Entire Network Weights): After the environmental exploration process, we need to update the network weights, such that the loss function of $Q(\mathbf{s}(k), \tau(k)|w^Q)$ can be minimized while the loss function of $\pi(\mathbf{s}(k)|\vartheta^\pi)$ can be maximized. To this end, N empirical data sets are extracted from the memory pool, and the corresponding time steps are re-labeled as $\{1_L, \dots, N_L\}$ where $1 \leq 1_L \leq k$ and $1 \leq N_L \leq k$. Particularly, an arbitrary empirical data set at time step \bar{k} is denoted by $\{\mathbf{s}(\bar{k}), \tau(\bar{k}), R(\mathbf{s}(\bar{k}), \tau(\bar{k})), \mathbf{s}(\bar{k}+1)\}$ where $\bar{k} \in \{1_L, \dots, N_L\}$. Based on this, at time step k , the loss function of $Q(\mathbf{s}(k), \tau(k)|w^Q)$ is constructed as

$$\text{Loss}(w_k^Q) = E(y(\bar{k}) - Q(\mathbf{s}(\bar{k}), \tau(\bar{k})|w_k^Q))^2, \quad (25)$$

with

$$y(\bar{k}) = R(\mathbf{s}(\bar{k}), \tau(\bar{k})) + \gamma \max_{\mathbf{s}(\bar{k}+1), \tau(\bar{k}+1)} Q'(\mathbf{s}(\bar{k}+1), \tau(\bar{k}+1)|w_k^{Q'}), \quad (26)$$

where $Q'(\mathbf{s}(\bar{k}+1), \tau(\bar{k}+1)|w_k^{Q'})$ denotes the Q-value of the target critic network $Q'(\mathbf{s}(k), \tau(k)|w_k^{Q'})$.

In order to minimize $\text{Loss}(w_k^Q)$, we employ the gradient descent to update weight w_k^Q , and hence, one has

$$\begin{aligned} \nabla_{w^Q} \text{Loss}(w_k^Q) &= \nabla_{w^Q} E(y(\bar{k}) - Q(\mathbf{s}(\bar{k}), \tau(\bar{k})|w_k^Q))^2 \\ &= -\frac{2}{N} \sum_{\bar{k}=1_L}^{N_L} (y(\bar{k}) - Q(\mathbf{s}(\bar{k}), \tau(\bar{k})|w_k^Q)) \\ &\quad \times \nabla_{w^Q} Q(\mathbf{s}(\bar{k}), \tau(\bar{k})|w_k^Q). \end{aligned} \quad (27)$$

From (27), one updates the weight w_k^Q as

$$w_{k+1}^Q = w_k^Q - \gamma_1 \nabla_{w^Q} \text{Loss}(w_k^Q), \quad (28)$$

where $\gamma_1 \in (0, 1)$ is the learning rate and w_{k+1}^Q is the weight of $Q(\mathbf{s}(k), \tau(k)|w^Q)$ at time step $k+1$.

Algorithm 1 DDPG-Based Motion Planning for AUV

```

1 Initialize  $w^Q, \vartheta^\pi$  randomly;
2  $w^{Q'} \leftarrow w^Q$  and  $\vartheta^{\pi'} \leftarrow \vartheta^\pi$ ;
3 for  $episode=1 : T_m$  do
4   set  $k = 1$ ;
5   while  $k < k_m$  do
6     Get  $\mathbf{s}(k)$  and  $\tau(k)$  by (3), (17) and by (23);
7     Calculate  $R(\mathbf{s}(k))$  and  $\tau(k)$  by (20);
8     Update  $\mathbf{s}(k+1)$  and store (24) in memory pool;
9     if  $episode > T_e$  then
10      if  $\text{MOD}(k, N_p) = 0$  then
11        Extract  $N$  data sets from memory pool;
12        Get  $y(\bar{k})$  from (26);
13        Update  $w_{k+1}^Q$  in (28) by minimizing
14          (25);
15        Update  $\vartheta_{k+1}^\pi$  in (31) by maximizing
16          (29);
17        Update target network weights  $w^{Q'}$  and
18           $\vartheta^{\pi'}$  with (32) and (33);
19       $k \leftarrow k + 1$ ;
20 Acquire the optimization value of  $\vartheta^{\pi*}$ ;
21 repeat
22   Get  $\mathbf{s}(k)$  based  $\mathbf{X}(k)$  by (3) and (17);
23   Get  $\pi(\mathbf{s}(k)|\vartheta^{\pi*})$  and  $\tau^*(k) \leftarrow \pi(\mathbf{s}(k)|\vartheta^{\pi*})$ ;
24    $\mathbf{s}(k) \leftarrow \mathbf{s}(k+1)$ ;
25 until Arrive at destination;

```

In the following, the weight ϑ^π is updated such that the loss function of $\pi(\mathbf{s}(k)|\vartheta^\pi)$ can be maximized. Along with this, the loss function of $\pi(\mathbf{s}(k)|\vartheta^\pi)$ is defined as

$$Loss(\vartheta_k^\pi) = E(Q(\mathbf{s}(\bar{k}), \pi(\mathbf{s}(\bar{k})|\vartheta_k^\pi)|w^Q)). \quad (29)$$

The gradient ascent is to update ϑ_k^π , and hence, one has

$$\begin{aligned} & \nabla_{\vartheta^\pi} Loss(\vartheta_k^\pi) \\ &= \nabla_{\vartheta^\pi} E(Q(\mathbf{s}(\bar{k}), \pi(\mathbf{s}(\bar{k})|\vartheta_k^\pi)|w^Q)) \\ &= \frac{1}{N} \left(\sum_{\bar{k}=1}^{N_L} \nabla_{\vartheta^\pi} Q(\mathbf{s}(\bar{k}), \pi(\mathbf{s}(\bar{k})|\vartheta_k^\pi)|w^Q) \right. \\ & \quad \left. \times \nabla_{\vartheta^\pi} \pi(\mathbf{s}(\bar{k})|\vartheta_k^\pi) \right). \end{aligned} \quad (30)$$

With (30), one updates the weight ϑ_k^π as

$$\vartheta_{k+1}^\pi = \vartheta_k^\pi + \gamma_2 \nabla_{\vartheta^\pi} Loss(\vartheta_k^\pi), \quad (31)$$

where $\gamma_2 \in (0, 1)$ denotes the learning rate and ϑ_{k+1}^π denotes the weight of $\pi(\mathbf{s}(k)|\vartheta^\pi)$ at time step $k+1$.

Accordingly, the weights in target network can be updated by using a soft update replicate from the main network at every N_p time steps. In view of this, one has

$$w^{Q'} = \rho w^Q + (1 - \rho) w^{Q'}, \quad (32)$$

$$\vartheta^{\pi'} = \rho \vartheta^\pi + (1 - \rho) \vartheta^{\pi'}, \quad (33)$$

where $w^{Q'}$ is the weight of target critic network $Q'(\mathbf{s}(k), \tau(k)|w^{Q'})$, and $\vartheta^{\pi'}$ is the weight of target actor network $\pi'(\mathbf{s}(k)|\vartheta^{\pi'})$. In addition, $\rho \in (0, 1)$ is the learning rate.

Repeating the above process, the optimal weights w^{Q*} and $\vartheta^{\pi*}$ can be obtained by minimizing (25) and maximizing (29).

Step 3 (Output the Optimal Action of AUV): Substituting $\vartheta^{\pi*}$ into the main actor network, one can obtain the optimal action $\tau^*(k)$ at time step k , i.e.,

$$\tau^*(k) = \pi(\mathbf{s}(k)|\vartheta^{\pi*}). \quad (34)$$

Extending $\tau^*(k)$ to the other time steps, the optimal actions of AUV are obtained. Correspondingly, the optimal states of AUV are obtained by substituting $\tau^*(k)$ into (6). For clear illustration, the complete DDPG-based motion planning procedure is shown in Algorithm 1, where T_e is the episode of exploration process, T_m is the total training episodes and k_m is the maximum time step of every episode.

C. Performance Analysis

We first study the optimization of DDPG algorithm. To support the proof, the following properties are presented [42].

Property 1 (Lipschitz Continuity): For a differentiable function $f: S \subseteq R \rightarrow R$, if there exists lipschitz constant $L > 0$, then $\forall a, b \in S$ have the following inequality

$$f(b) \leq f(a) + \nabla f(a)(b - a) + \frac{L}{2} \|b - a\|^2. \quad (35)$$

Property 2 (Strong Convexity): For a differentiable function $f: S \subseteq R \rightarrow R$ is U -strongly convex and $U > 0$, then $\forall a, b \in S$ have the following inequality

$$f(b) \geq f(a) + \nabla f(a)(b - a) + \frac{U}{2} \|b - a\|^2. \quad (36)$$

Based on this, we have the following theorem.

Theorem 1: Consider the AUV model (1) with DDPG-based motion planning algorithm. If the parameters T_m , k_m and N_p satisfy the following condition

$$\frac{T_m k_m}{N_p} \geq \log \frac{1}{1 - \partial \varpi^{-1}} \epsilon^{-1} (Loss(w_0^Q) - Loss^*(w^Q)), \quad (37)$$

where $Loss^*(w^Q)$ denotes the optimal value of loss function, $\epsilon > 0$ is a decimal, $0 < \frac{\partial}{\varpi} < 1$, ∂ and ϖ are respectively the ∂ -strongly convex and ϖ -lipschitz constants of $Loss(w^Q)$, then the loss function can converge to the optimal value.

Proof: First, we prove that the loss error is a monotonous decrease function with weight w_k^Q . To this end, we employ Property 1 to obtain the following result

$$\begin{aligned} & Loss(w_{k+1}^Q) - Loss(w_k^Q) \\ & \leq \nabla_{w^Q} Loss(w_k^Q)(w_{k+1}^Q - w_k^Q) + \frac{\varpi}{2} \|w_{k+1}^Q - w_k^Q\|^2 \end{aligned} \quad (38)$$

From (28) and (38), we can further have

$$\begin{aligned} & Loss(w_{k+1}^Q) - Loss(w_k^Q) \\ & \leq \nabla_{w^Q} Loss(w_k^Q)(w_{k+1}^Q - w_k^Q) + \frac{\varpi}{2} \|w_{k+1}^Q - w_k^Q\|^2 \\ & \leq \frac{\varpi \gamma_1^2}{2} \|\nabla_{w^Q} Loss(w_k^Q)\|^2 - \gamma_1 \|\nabla_{w^Q} Loss(w_k^Q)\|^2. \end{aligned} \quad (39)$$

Set $\gamma_1 = \varpi^{-1}$, thus (39) can be rearranged as

$$Loss(w_{k+1}^Q) - Loss(w_k^Q) \leq -\frac{1}{2\varpi} \|\nabla_{w^Q} Loss(w_k^Q)\|^2. \quad (40)$$

Similarly, based on Property 2, one has

$$\begin{aligned} & Loss(w_{k+1}^Q) - Loss(w_k^Q) \\ & \geq \nabla_{w^Q} Loss(w_k^Q)(w_{k+1}^Q - w_k^Q) + \frac{\partial}{2} \left\| w_{k+1}^Q - w_k^Q \right\|^2. \end{aligned} \quad (41)$$

We set that $w_{k+1}^Q = w_k^Q - \partial^{-1} \nabla_{w^Q} Loss(w_k^Q)$, and hence (41) can be rearranged as

$$Loss(w_{k+1}^Q) - Loss(w_k^Q) \geq -\frac{1}{2\partial} \left\| \nabla_{w^Q} Loss(w_k^Q) \right\|^2. \quad (42)$$

Note that the final iteration output is regarded as the optimal loss function $Loss^*(w^Q)$. Thus, from (42), one has

$$Loss^*(w^Q) - Loss(w_k^Q) \geq -\frac{1}{2\partial} \left\| \nabla_{w^Q} Loss(w_k^Q) \right\|^2. \quad (43)$$

Subtracting $Loss^*(w^Q)$ from both sides of (40), one has

$$\begin{aligned} & Loss(w_{k+1}^Q) - Loss^*(w^Q) \\ & \leq Loss(w_k^Q) - Loss^*(w^Q) - \frac{1}{2\partial} \left\| \nabla_{w^Q} Loss(w_k^Q) \right\|^2 \end{aligned} \quad (44)$$

Taking (43) into (44), one has

$$\begin{aligned} & Loss(w_{k+1}^Q) - Loss^*(w^Q) \\ & \leq \left(1 - \frac{\partial}{\partial}\right) (Loss(w_k^Q) - Loss^*(w^Q)). \end{aligned} \quad (45)$$

Note that $Loss(w_k^Q) - Loss^*(w^Q) \geq 0$ and $(1 - \frac{\partial}{\partial}) > 0$ always hold. Thus we extend (45) to the whole iteration sequence, i.e., from the initial iteration to the $\frac{T_m k_m}{N_p}$ th iteration during the T_m episodes, through which we have

$$\begin{aligned} & Loss(w_k^Q) \Big|_{\frac{T_m k_m}{N_p}} - Loss^*(w^Q) \\ & \leq \left(1 - \frac{\partial}{\partial}\right)^{\frac{T_m k_m}{N_p}} (Loss(w_0^Q) - Loss^*(w^Q)). \end{aligned} \quad (46)$$

We regard that $Loss(w_k^Q) \Big|_{\frac{T_m k_m}{N_p}}$ converges to $Loss^*(w^Q)$ if

$Loss(w_k^Q) \Big|_{\frac{T_m k_m}{N_p}} - Loss^*(w^Q) \leq \epsilon$, where ϵ is a positive decimal. We set $k_{th} = \log \frac{1}{1 - \frac{\partial}{\partial}} \epsilon^{-1} (Loss(w_0^Q) - Loss^*(w^Q))$. Based on this, we can further have

$$Loss(w_k^Q) \Big|_{k_{th}} - Loss^*(w^Q) \leq \epsilon, \quad (47)$$

which means $Loss(w_k^Q) \Big|_{k_{th}}$ can converge to the optimal if (37) is satisfied. Once $Loss(w_k^Q)$ reaches to the optimal value, the optimal control policy $\tau^*(k)$ can be derived subsequently. ■

Typically, the horizontal parallax angle is adopted to achieve obstacle avoidance, but it can limit the flexibility of the paths of AUV. Alternatively, this paper jointly employs the horizontal and vertical parallax angles to achieve collision-free motion planning. Then the following corollary is provided.

Corollary 1: By employing the vertical binocular cameras, the measured vertical parallax angle $\theta_{V,i,j}$ in (17) can correlate with the motion planning of AUV in Z axis.

Proof: Suppose the vertical binocular camera $\bar{\Delta} \in \{C, D\}$ has detected the target point $[X_{\bar{\Delta},i,j}, Y_{\bar{\Delta},i,j}]^T$ at depth z . When

AUV moves down (i.e., the value of z decreases), $Y_{\bar{\Delta},i,j}$ will decrease, otherwise, it will rise. Then, we have

$$Y_{\bar{\Delta},i,j} \propto (z), \text{ i.e., } Y_{\bar{\Delta},i,j} = \chi_{\bar{\Delta}}(z), \quad (48)$$

where $\chi_{\bar{\Delta}}$ is a fixed positive scale factor.

Based on (48), $d_{\bar{\Delta},i,j}$ in (10) can be rearranged as

$$d_{C,y,i,j} = |Y_{C,i,j} - Y_o| = |\chi_C(z) - Y_o|, \quad (49)$$

$$d_{D,y,i,j} = |Y_{D,i,j} - Y_o| = |\chi_D(z) - Y_o|. \quad (50)$$

Assume the horizontal position is fixed, then $f_{\bar{\Delta},i,j}$ in (12) is a fixed constant, which can be denoted as $f_{\bar{\Delta}}^\#$. From (49) and (50), we can recalculate $\theta_{V,i,j}$ as

$$\begin{aligned} \theta_{V,i,j} &= \pi - \theta_{C,i,j} - \theta_{D,i,j} \\ &= \pi - \arctan\left(\frac{f_{\bar{\Delta}}^\#}{|\chi_C(z) - Y_o|}\right) - \arctan\left(\frac{f_{\bar{\Delta}}^\#}{|\chi_D(z) - Y_o|}\right) \\ &= \pi - \arctan\left[\frac{f_{\bar{\Delta}}^\# |\chi_D(z) - Y_o| + f_{\bar{\Delta}}^\# |\chi_C(z) - Y_o|}{|\chi_C(z) - Y_o| |\chi_D(z) - Y_o| - f_{\bar{\Delta}}^\# f_{\bar{\Delta}}^\#}\right]. \end{aligned} \quad (51)$$

Clearly, $\theta_{V,i,j}$ is related to z , and the motion of AUV in Z axis requires to be adjusted during the collision-avoidance procedure, since $f_{\bar{\Delta}}^\#$ and $\chi_{\bar{\Delta}}$ are fixed constants. ■

In (17), the thresholds of P_H^θ and P_V^θ are provided. To verify the correctness, the following corollary is given.

Corollary 2: Given the maximum visible distance Υ of a binocular camera, the threshold ranges of P_H^θ and P_V^θ are related to the camera baselines, which can be set as $[2 \tan^{-1} \frac{f_A}{X_o}, \frac{\pi}{2} + \tan^{-1} \frac{f_A}{X_o} + \tan^{-1} \frac{X_o \Upsilon - f_A B_H}{f_A \Upsilon}]$ and $[2 \tan^{-1} \frac{f_C}{Y_o}, \frac{\pi}{2} + \tan^{-1} \frac{f_C}{Y_o} + \tan^{-1} \frac{Y_o \Upsilon - f_C B_V}{f_C \Upsilon}]$, respectively.

Proof: We first study the lower limit value of P_H^θ , i.e., the case when $\theta_{H,i,j}$ has the maximum value. As shown in Fig. 8 (a), one can obtain that the above case is occurred when $\theta_{A,i,j} = \theta_{B,i,j} = \tan^{-1}(f_A/X_o)$. Based on this, the lower limit value of P_H^θ in (17) can be determined as

$$P_H^\theta = \theta_{A,i,j} + \theta_{B,i,j} \geq P_{H,\min} = 2 \tan^{-1} \frac{f_A}{X_o}. \quad (52)$$

On the contrary, the upper limit value of P_H^θ can be acquired by minimizing $\theta_{H,i,j}$. Hence, the above case is occurred when point P_{ij} is on the visual boundary, as shown by Fig. 8 (b). By using inverse tangent function, one can have

$$\theta_{B1} = \frac{\pi}{2}, \quad \theta_{B2} = \tan^{-1} \frac{X_o \Upsilon - f_A B_H}{f_A \Upsilon}, \quad (53)$$

$$\theta_{A,i,j} = \tan^{-1} \frac{f_A}{X_o}, \quad (54)$$

$$\theta_{B,i,j} = \theta_{B1} + \theta_{B2} = \frac{\pi}{2} + \tan^{-1} \frac{X_o \Upsilon - f_A B_H}{f_A \Upsilon}. \quad (55)$$

Based on (53), (54) and (55), the upper limit value of P_H^θ in (17) can be determined as

$$\begin{aligned} P_H^\theta &= \theta_{A,i,j} + \theta_{B,i,j} \\ &\leq P_{H,\max} \\ &= \frac{\pi}{2} + \tan^{-1} \frac{f_A}{X_o} + \tan^{-1} \frac{X_o \Upsilon - f_A B_H}{f_A \Upsilon}. \end{aligned} \quad (56)$$

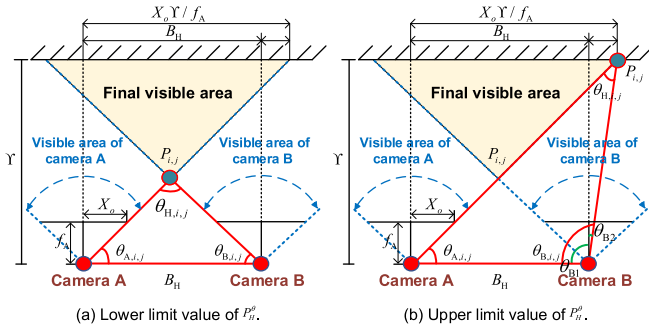


Fig. 8. Schematic diagram of the threshold calculation.

TABLE II
SOME PARAMETERS OF AUV AND DDPG NETWORK

Parameter	Value	Parameter	Value
m_u	11.5	m_v	19.6
m_w	19.6	I_r	35.3
k_u	16.0	k_v	17.8
k_w	25.6	k_r	25.4
$k_u u $	29.4	$k_v v $	28.3
$k_w w $	26.8	$k_r r $	21.6
T_m	1500	k_m	800
N	128	N_p	50
ρ	0.005	γ	0.99
γ_1	10^{-3}	γ_2	10^{-3}

Similarly, the threshold of P_V^θ can be calculated by

$$2 \tan^{-1} \frac{f_C}{Y_o} \leq P_V^\theta \leq \frac{\pi}{2} + \tan^{-1} \frac{f_C}{Y_o} + \tan^{-1} \frac{Y_o \Upsilon - f_C B_V}{f_C \Upsilon}.$$

Clearly, the thresholds of P_H^θ and P_V^θ are related to the binocular camera baselines (i.e. B_H and B_V), since f_A , f_C , X_o , Y_o and Υ are constants. That completes the proof. ■

IV. SIMULATION AND EXPERIMENTAL STUDIES

A. Simulation Studies

In this section, we perform simulation results to verify the developed motion planning solution. Consider that the motion-planning system is consist of an AUV and three surface buoys. Particularly, the initial state of AUV is $[0, 0, -3, 0, 0, 0, 0, 0]^T$, while the position vectors of buoys are $\mathbf{P}_{B,1} = [5, 2, 0]^T$, $\mathbf{P}_{B,2} = [4, 7.5, 0]^T$ and $\mathbf{P}_{B,3} = [10, 9.5, 0]^T$. The state of destination is $\mathbf{X}_{\text{target}} = [10, 9.5, -1, 0, 0, 0, 0, 0]^T$. In addition, the spherical and column obstacles are set as follows: $S_1 = [3, 7, -1.3, 1]^T$, $S_2 = [8, 3, -0.3, 1.2]^T$, $C_1 = [1, 3, 1, 2]^T$, $C_2 = [3, 0.5, 0.7, 2.7]^T$, $C_3 = [5, 5, 1, 1.7]^T$, $C_4 = [8, 3, 1.2, 1.5]^T$ and $C_5 = [8.5, 6, 1, 2.5]^T$. Some parameters of AUV and DDPG network are provided in Table II.

1) *Vision-Based Motion Planning With SNR and DDPG*: As presented in Section III, a binocular vision-based path planning solution with consideration of SNR and DDPG is designed for AUV. To verify its effectiveness, the estimated channel parameters are given as follows: $\theta^* = [-50, 3.95]^T$, $\sigma_{SH} = 3.3419$ and $\varepsilon = -100$ dBm. Based on these parameters, we employ (3) to predict the SNRs in the whole monitoring region, which can be shown in Fig. 9(a). Based on this, the motion trajectory of AUV is depicted by Fig. 9(b). Correspondingly, the motion state of AUV in the

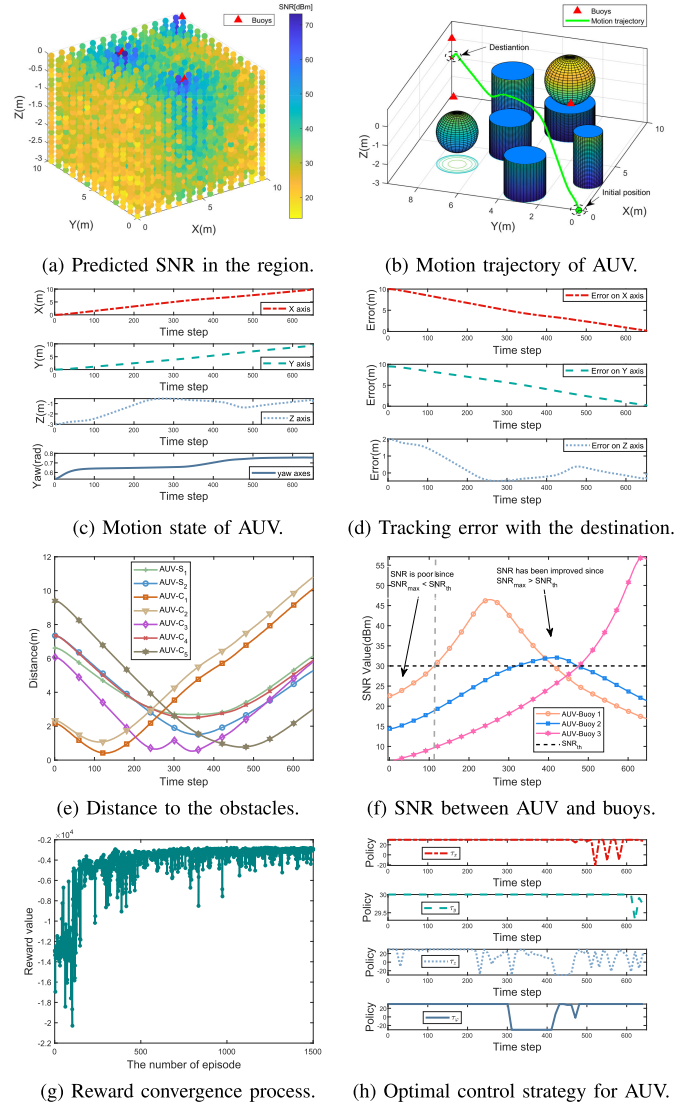


Fig. 9. Simulation results for the binocular vision-based motion planning of AUV with SNR and DDPG.

X-axis, Y-axis, Z-axis and yaw are shown in Fig. 9(c). To show more clearly, we define the tracking error vector as $e_{\text{track}} = [x - x_g, y - y_g, z - z_g]^T$, which can be shown by Fig. 9(d). From Figs. 9(b)-(d), we know that the tracking task can be achieved, since the tracking error can converge to zero.

Similarly, the relative distance between AUV and column obstacle $n \in \{1, \dots, 5\}$ is defined as $e_{cn} = \|X - O_{cn}\| - \rho_{cn}$, while the relative distance between AUV and sphere obstacle $m \in \{1, 2\}$ is $e_{sm} = \|X(k) - O_{sm}\| - \rho_{sm}$. Along with this, the relative distances between AUV and obstacles are shown in Fig. 9(e). Clearly, we find that the collision-avoidance task can be accomplished by AUV, since all the relative distances are greater than zero. Note that the communication channel quality, i.e., SNR, is incorporated into the motion planning procedure. Meanwhile, our objective is to guarantee that the SNR for at least one buoy is greater than the threshold. In view of this, Fig. 9(f) presents the SNR between AUV and buoy $j \in \{1, 2, 3\}$. Clearly, the SNR is smaller than the threshold at $k = 1$, which means the communication quality is poor at

the initial stage. After using the communication-aware motion solution, the SNR shows an increasing trend during $1 < k \leq 113$. When $k > 113$, we find that the SNR for at least one buoy is greater than the threshold, indicating the communication connection with at least one buoy can be guaranteed.

Besides that, 1500 episodes are conducted in Algorithm 1 to guarantee the convergence of loss function. For that design, the reward value in (20) can be shown in Fig. 9(g). It indicates that the network weights can converge to the optimal. Based on this, the optimal control strategy τ^* can be derived, as shown in Fig. 9 (h). These results verify the effectiveness of the vision-based motion planning solution in this paper.

2) *Comparison With Distance-Based Obstacle Avoidance*: Note that the distance-based obstacle avoidance solutions have been proposed in [10] and [11]. However, the distance-based solutions are not suitable for obstacle-dense environment. To verify the above judgement, we incorporate the distance measurements into the motion planning of AUV. For easy to compare, the SNR is removed from the reward function. In view of this, the motion trajectories of AUV based on distance and parallax measurements are shown in Fig. 10(a). Correspondingly, the values of reward functions are depicted in Fig. 10(b), while the relative distances between AUV and obstacles are shown in Figs. 10(c)-(d). From Figs. 10(b)-(d), we note that the reward functions with the above two solution can both converge to the optimal, however the vision-based solution can safely avoid collision at a closer distance.

Meanwhile, the parallax-based motion planning solution in this paper can reflect the size information of AUV. This is very important for the practical application of AUV in obstacle-dense environment. For that consideration, the following scenarios are considered: 1) A sonar sensor is installed on the middle-front region of AUV; 2) A pair of binocular cameras are installed on the horizontal region of AUV. Once an obstacle is detected by AUV, the detection result of Scenario 1 by using a narrow AUV (e.g., the width is 1 m) is shown in Fig. 10(e), while the detection result of Scenario 1 by using a wide AUV (e.g., the width is 2 m) is shown in Fig. 10(f). From Figs. 10(e)-(f), we find the detection results have the same value. On the other hand, the detection results of Scenario 2 are shown in Figs. 10(g)-(h). Clearly, we find the wider AUV can trigger a larger parallax value compared with a narrower vehicle in Scenario 2. Therefore, the parallax-based solution in this paper can indirectly reflect the size information of the AUV.

3) *Comparison With the Single Horizontal Cameras*: As mentioned above, this paper jointly employs the horizontal and vertical parallax angles to achieve collision-free motion planning, which can improve the flexibility of obstacle avoidance. To verify its merit, we consider the following two cases: 1) The horizontal parallax angle is adopted to achieve obstacle avoidance, e.g., the solution in [13]; 2) The horizontal and vertical parallax angles are jointly adopted to achieve the obstacle avoidance, i.e., the solution in this paper. Besides that, it is set that three column obstacles block the way of AUV. Accordingly, the motion trajectories of AUV in the above two cases are shown by Fig. 11(a). Correspondingly, the states of AUV in the above cases are shown by Figs. 11(b)-(c).

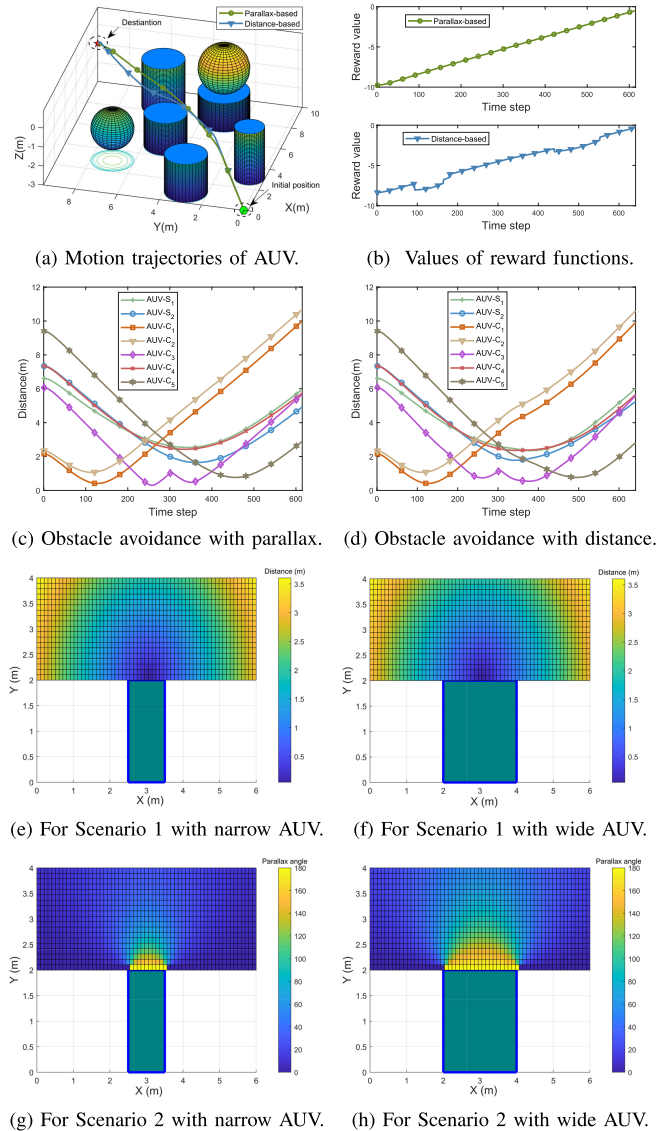


Fig. 10. Comparison with distance-based solutions, e.g., [10] and [11].

To show more clearly, we define the path length as the total distance from start to finish the obstacle avoidance. Then, the path length and total time step in the above cases are shown in Fig. 11(d). Clearly, the path length and total time step in Case 2 are shorter than the ones in Case 1, which verify the effectiveness of Corollary 1. More specifically, Case 1 runs 354 time steps to finish the tracking task, while Case 2 only runs 285 time steps. The results in Figs. 11 (b)-(d) demonstrate that the adoption of horizontal and vertical cameras is meaningful.

4) *Tradeoff of SNR During the Motion Planning*: As has mentioned above, the consideration of SNR can ensure the communication efficiency during the motion planning process. However, it may degrade the tracking performance. To reflect this situation, the following three cases are considered: 1) Communication efficiency plays a major role on the reward function, e.g., $K_1 = 0.6$, $K_2 = 0.01$ and $K_3 = 26$; 2) Tracking stability plays a major role on the reward function, e.g., $K_1 = 0.7$, $K_2 = 0.01$ and $K_3 = 10$; 3) Communication

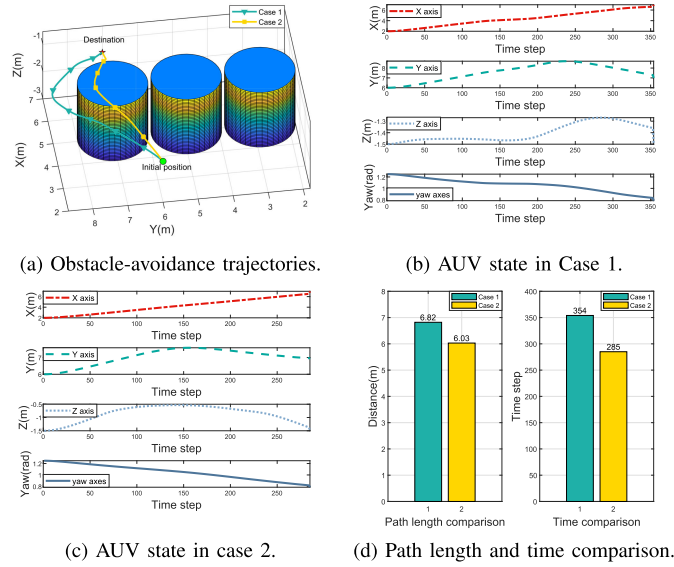


Fig. 11. Comparison with the single horizontal cameras in [13].

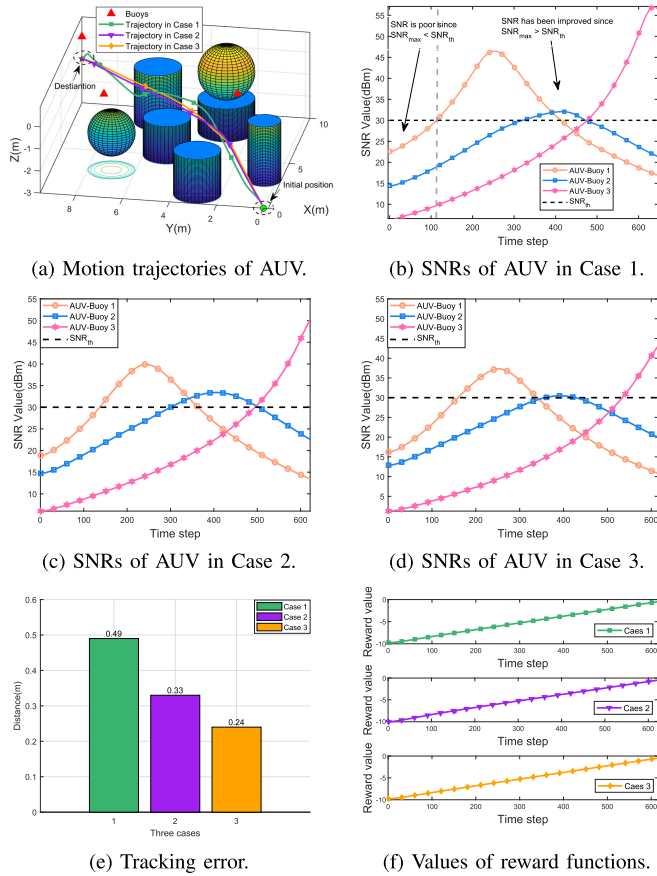


Fig. 12. Tradeoff of the SNR during the motion planning.

efficiency is ignored, e.g., $K_1 = 0.7$, $K_2 = 0.01$ and $K_3 = 0$. Based on the above design, the motion trajectories of AUV are shown in Fig. 12(a). Correspondingly, the SNRs between AUV and buoy $j \in \{1, 2, 3\}$ for the above three cases are shown in Figs. 12(b)-(d), respectively. Meanwhile, the tracking errors are presented in Fig. 12(e). From Figs. 12(a)-(e), we can find that there is a tradeoff between communication effectiveness

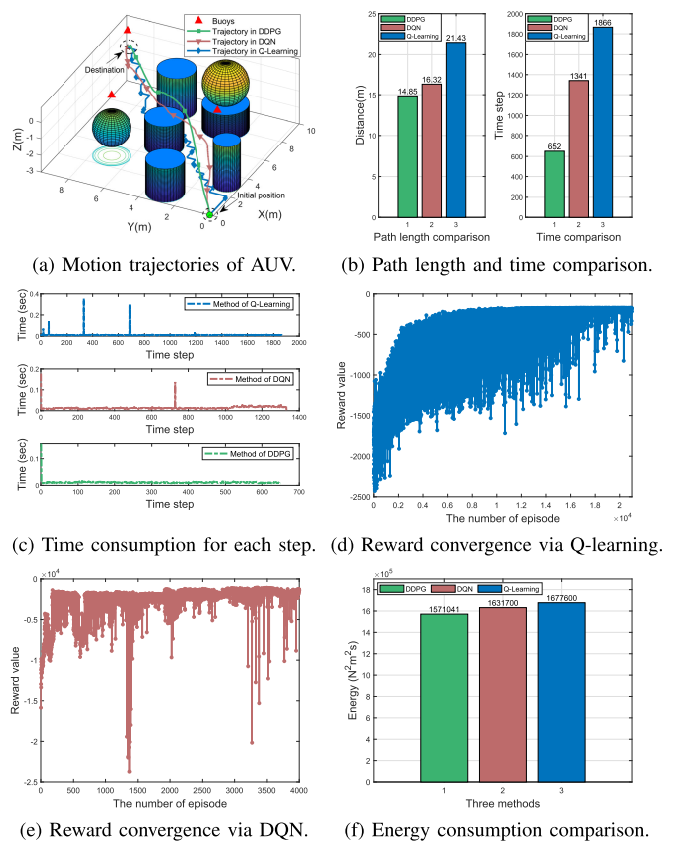
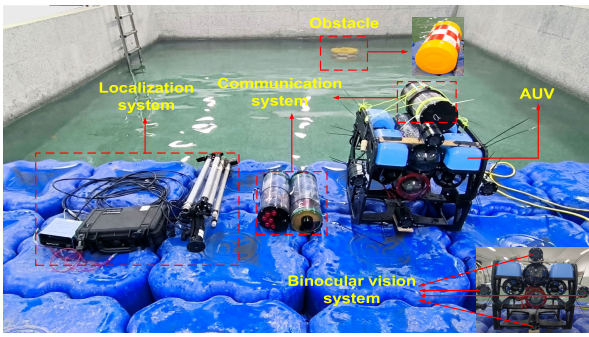


Fig. 13. Comparison with Q learning and DQN, e.g., [19] and [20].

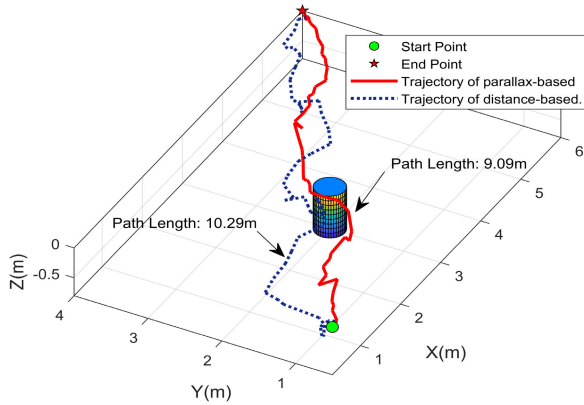
and tracking stability. Therefore, the researchers can adjust the weights according to the actual engineering requirements. In addition, the values of reward functions are depicted by Fig. 12(f), which can demonstrate the optimal of the DDPG-based motion-planning algorithm in this paper.

5) *Comparison With the Q Learning and DQN Methods:* The Q learning method was adopted in [19] to perform collision-avoidance navigation, while the DQN method was utilized in [20] to achieve global path planning. However, these methods cannot be directly applied to continuous state space since they find action that maximizes the action-value function. For comparison purpose, the Q learning and DQN methods are employed in this section to plan the path of AUV. Based on this, the motion trajectories of AUV with the above three methods are shown in Fig. 13(a). Correspondingly, the path length and time cost are shown in Fig. 13(b). The time consumption for each time step is presented in Fig. 13(c). The reward value by using Q learning and DQN methods is shown in Figs. 13(d)-(e). It is obvious that the traditional optimization methods (i.e., Q learning and DQN methods) can minimize the cost function in general. However, the path length and time cost with the DDPG method are the smallest, as compared with the ones in Q learning and DQN methods.

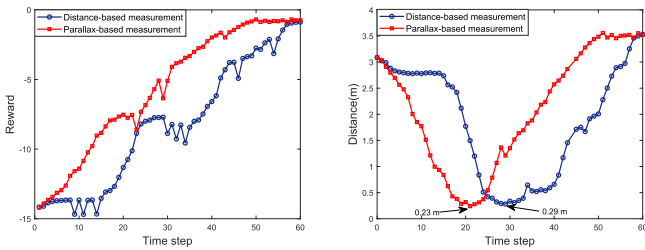
Next, the energy consumption indicator is adopted to quantitatively measure the control efforts. Referring to [43], the energy consumption indicator is defined as $E_{\text{total}} = \sum_{k=1}^{t_f} E(k)$, where $E(k) = \|\tau_u(k)\|^2 + \|\tau_v(k)\|^2 + \|\tau_w(k)\|^2 + \|\tau_r(k)\|^2$, and t_f denotes the termination step. Based on this,



(a) Experimental setup in obstacle-sparse environment.



(b) Motion trajectory of AUV with vision and distance based solutions.

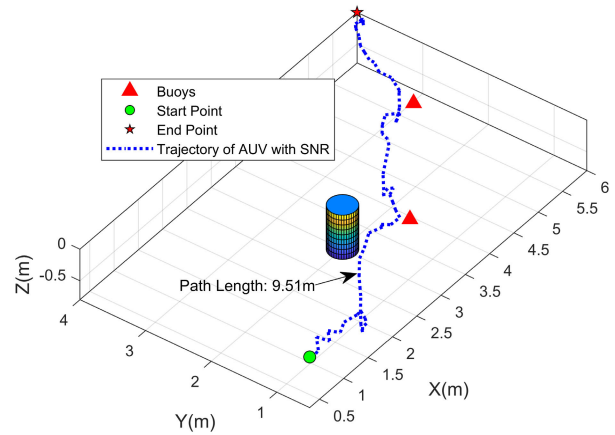


(c) Convergence of reward function. (d) Distance between AUV and obstacle. Fig. 14. Experimental setup and results when the SNR is ignored in obstacle-sparse environment.

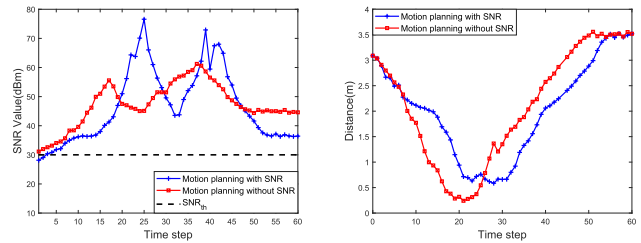
Fig. 13(f) shows the energy consumption of controller by using Q learning, DQN and DDPG methods. From Fig. 13(f), we find that the energy consumption indicator by using DDPG is the smallest as compared with the other two cases. It reflects that the DDPG-based solution in this paper can improve the collision-avoidance efficiency in control efforts.

B. Experimental Results

In this section, the experimental results are provided. Specifically, the hardware is mainly composed of four parts: 1) *Binocular vision system*. The binocular vision system is built by four monocular cameras for real-time obstacle detection. 2) *Communication system*. Acoustic transducer and acoustic modem that adopt orthogonal frequency division multiplexing (OFDM) model are employed for the wireless communication with surface buoys. Meanwhile, the video transmission is accomplished by the umbilical communication. 3) *Localization system*. The USBL is adopted for the self-localization of AUV. 4) *AUV*. BlueROV2, which is one of the most affordable high-performance underwater vehicles, acts as the major structure of AUV in our system. Besides that,



(a) Motion trajectory of AUV.



(b) Maximum SNRs between AUV and (c) Distance between AUV and buoys.

Fig. 15. Experimental results when the SNR is considered in obstacle-sparse environment.

the control cabin mainly contains STM32 microprocessor and NVIDIA microcomputer. Of note, STM32 microprocessor is used to control the motion of AUV. Meanwhile, the NVIDIA microcomputer is used to process the video streams from the vision system.

1) Experimental Results in Obstacle-Sparse Environment:

In this case, one underwater obstacle exists in the environment. For clear verification, we first ignore the influence of SNR, and hence, the experimental setup is depicted by Fig. 14(a). By employing the parallax vision-based measurement, the DDPG is performed to generate the motion trajectory of AUV, as shown in Fig. 14(b). At the same time, we also employ the distance-based measurement to plan the motion of AUV, which is also shown in Fig. 14(b). It is clear that, the parallax vision-based solution in this paper can improve the collision-avoidance efficiency in path lengths and control efforts, since the path length of distance-based solution is 10.29 m while the parallax vision-based solution is 9.09 m. Besides that, the reward functions with the above two measurements are presented in Fig. 14(c), while the relative distances between AUV and obstacles are shown in Fig. 14(d). From Figs. 14 (c)-(d), we can find that the reward functions with the above two measurements can both converge to the optimal, however the parallax vision-based solution can safely avoid collision at a more closer distance (i.e., 0.23 m). The above results demonstrate the necessary and meaningful of parallax vision-based solution in this paper. Particularly, the demo for these results is provided in the first part of [44].

In the following, we incorporate the SNR quality into the motion planning of AUV. By employing the parallax

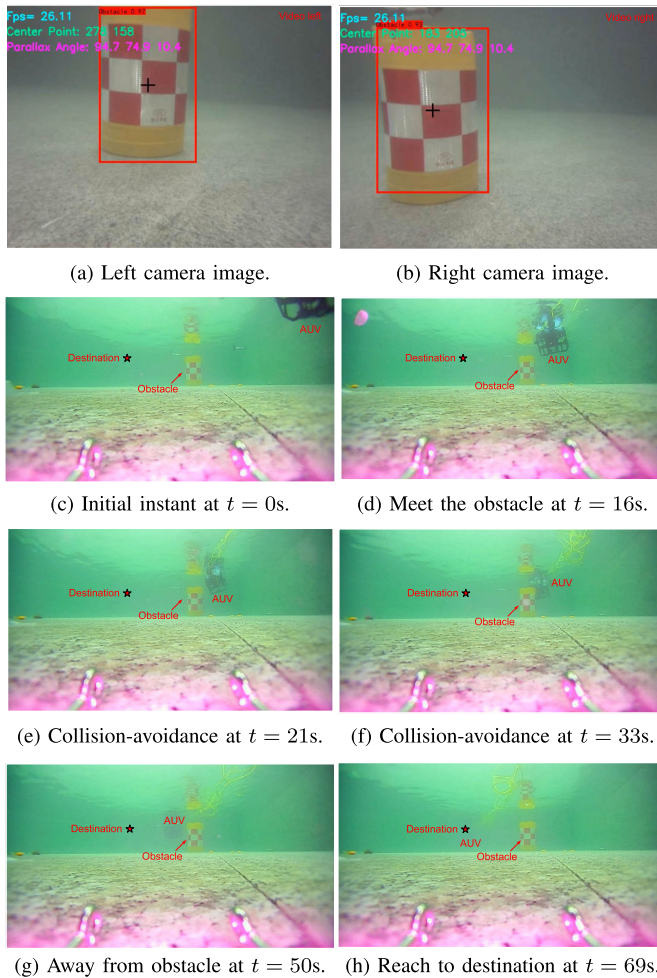
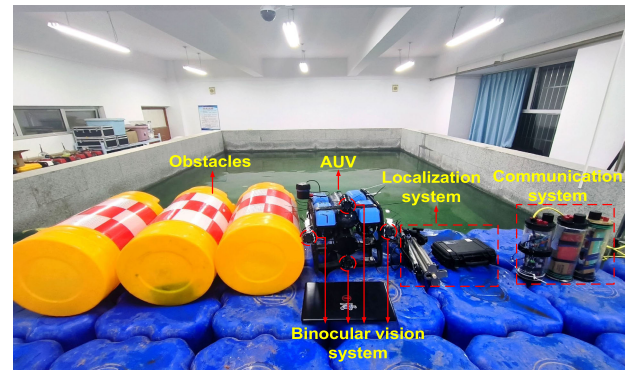
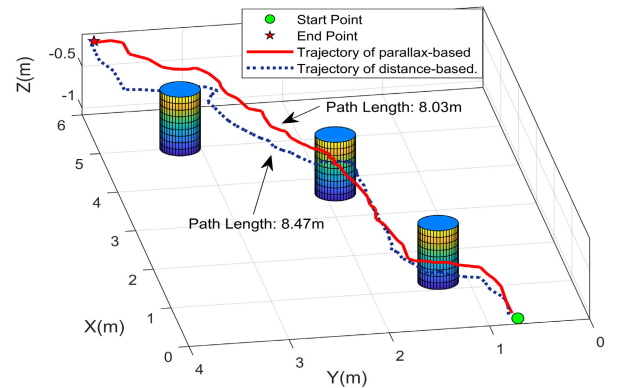


Fig. 16. Obstacle recognition and the video screenshots for the single-obstacle avoidance procedure.

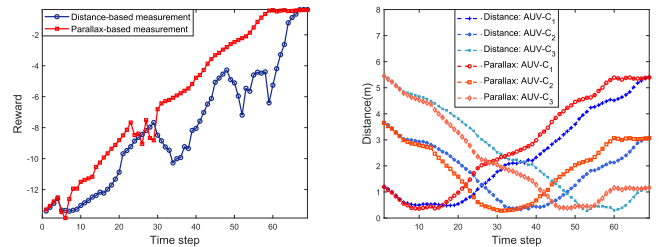
vision-based path planning measurement, the motion trajectory of AUV is shown in Fig. 15(a). Clearly, the consideration of SNR in the reward function can drive AUV approach to the surface buoy, such that the communication quality can be guaranteed even then the performance of motion stability may decrease. These results demonstrate that it is a tradeoff between communication effectiveness and motion stability. The maximum SNRs between AUV and buoys for the above two measurements are shown in Fig. 15(b), while the relative distances between AUV and obstacles are shown in Fig. 15(c). From Figs. 15 (b)-(c), we can find that both measurements can successfully avoid obstacles. However, when the SNR is considered in the reward function, the channel quality in the AUV motion planning process will be improved. Besides that, the obstacle recognition process from the left and right cameras on AUV can be depicted by Figs. 16(a)-(b), respectively. Clearly, the pixel coordinates (278, 158) of the obstacle center point can be obtained from the left camera. At the same time, the pixel coordinates (183, 205) of the obstacle center point can be obtained from the right camera. Based on this, the parallax angle is obtained as $\theta_{H,1,1} = 180^\circ - \theta_{A,1,1} - \theta_{B,1,1} = 180^\circ - 94.7^\circ - 74.9^\circ = 10.4^\circ$. Meanwhile, the video screenshots for the obstacle avoidance procedure are shown in



(a) Experimental setup in obstacle-dense environment.



(b) Motion trajectory of AUV with vision and distance based solutions.

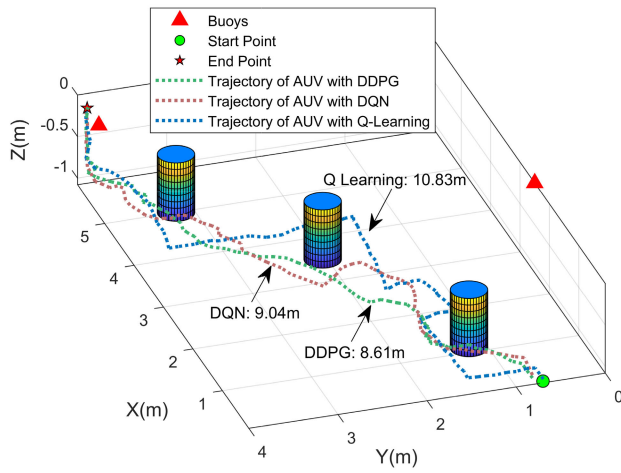


(c) Convergence of reward function. (d) Distance between AUV and obstacle.

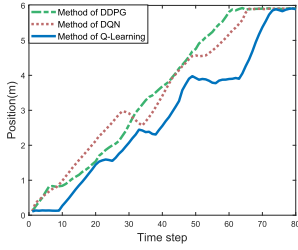
Fig. 17. Experimental results when the SNR is ignored in obstacle-dense environment.

Figs. 16(c)-(h). These results reflect the effectiveness of the collision-avoidance algorithm. Particularly, the demo for these results is provided in the second part of [44].

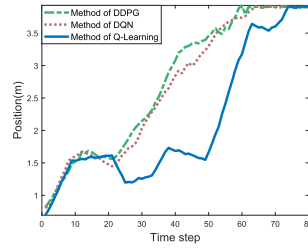
2) *Experimental Results in Obstacle-Dense Environment:* In this section, the obstacle-dense environment is considered, where three obstacles exist in the workspace, as shown by Fig. 17(a). We first ignore the influence of SNR, and hence, the motion trajectories of AUV with parallax-based and distance-based solutions are shown in Fig. 17(b). Clearly, the path length by using parallax-based solution (i.e., 8.03 m) is smaller than the one by using distance-based solution (i.e., 8.47 m). Besides that, the reward functions with the above two solutions are presented by Fig. 17(c). The relative distances between AUV and obstacles are shown in Fig. 17(d), where C1, C2 and C3 denote the obstacle 1, obstacle 2 and obstacle 3, respectively. We can see that the reward functions with the above two solutions can both converge to the optimal, meanwhile, the parallax-based solution can avoid obstacles at



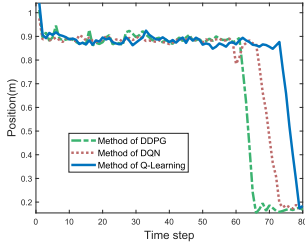
(a) Motion trajectory of AUV when SNR is considered.



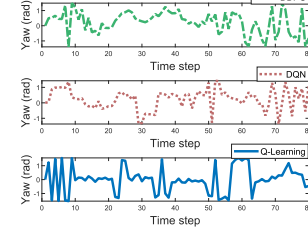
(b) Position on X axis.



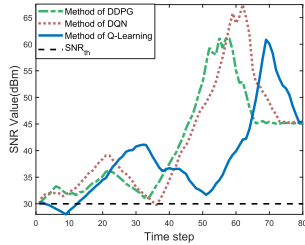
(c) Position on Y axis.



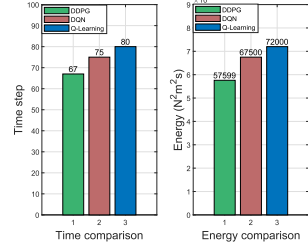
(d) Position on Z axis.



(e) Angle on yaw.



(f) Maximum SNRs between AUV and buoys.



(g) Time and energy comparison.

Fig. 18. Experimental results when the SNR is considered in obstacle-dense environment.

a more close distance in obstacle-dense environment. Particularly, the demo for these results is provided in the third part of [44].

Along with this, we incorporate the SNR quality into the motion planning of AUV. Particularly, the DDPG-based solution in this paper is compared with Q learning method [19] and DQN method [20]. By employing the parallax-based measurement, the motion trajectories of AUV with the above three methods are shown in Fig. 18(a). Correspondingly, the

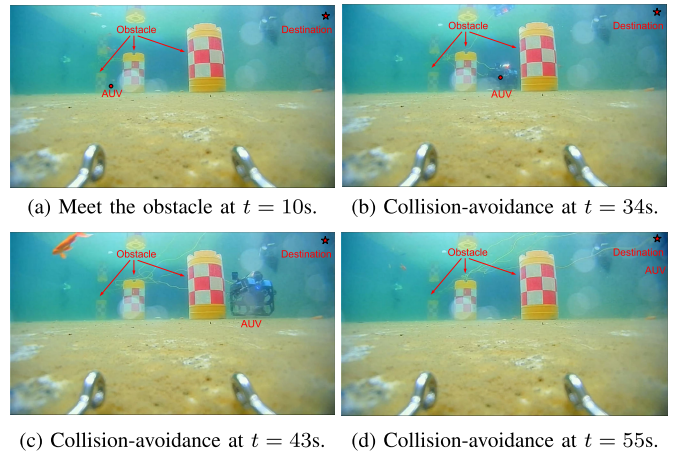


Fig. 19. Video screenshots of DDPG-based solution when the SNR is considered in obstacle-dense environment.

motion states of AUV on X axis, Y axis, Z axis and yaw axis are shown in Figs. 18(b)-(e), respectively. Clearly, the DDPG-based solution in this paper has the smallest path length (i.e., 8.61m), as compared with the DQN and Q learning based solutions. Besides that, the maximum SNR between AUV and buoys is depicted by Fig. 18(f). We find that the maximum SNR is larger than the threshold, which means AUV can maintain a minimum required SNR with at least one buoy. Finally, the total time cost and energy consumption of tracking controller are shown in Fig. 18(g). It can be seen that the DDPG-based solution in this paper has the minimum time step and energy consumption, as compared with the other two solutions. The video screenshots of DDPG-based solution are shown in Figs. 19(a)-(d). These results reflect the merits of our solution over the DQN and Q learning solutions. The demo for these results is provided in the fourth part of [44].

V. CONCLUSION AND FUTURE WORKS

This paper presents a communication-aware motion planning solution for AUV in obstacle-dense environment. By integrating our previously proposed channel estimation approaches, a binocular vision-based co-optimization framework is constructed for AUV to balance the communication quality and motion stability. Based on this, the vision measurements are employed to design a DDPG-based obstacle avoidance algorithm for AUV, which can move AUV away from obstacles and toward to destination. Finally, simulation and experimental results are both performed to validate the effectiveness.

In the future, we will adopt the binocular vision-based deep learning method to solve the swarm control of multiple AUVs in more complex maritime transportation systems. Besides that, the co-design of underwater sensing, transmission and control is also our future work.

REFERENCES

[1] J. Zhang, G. Han, J. Sha, Y. Qian, and J. Liu, "AUV-assisted subsea exploration method in 6G enabled deep ocean based on a cooperative Pac-Men mechanism," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1649–1660, Feb. 2022.

- [2] C. Lin, G. Han, T. Zhang, S. B. H. Shah, and Y. Peng, "Smart underwater pollution detection based on graph-based multi-agent reinforcement learning towards AUV-based network ITS," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7494–7505, Jul. 2023, doi: 10.1109/TITS.2022.3162850.
- [3] J. Ahn, S. Yasukawa, T. Sonoda, Y. Nishida, K. Ishii, and T. Ura, "An optical image transmission system for deep sea creature sampling missions using autonomous underwater vehicle," *IEEE J. Ocean. Eng.*, vol. 45, no. 2, pp. 350–361, Apr. 2020.
- [4] J. Guo and B. He, "Improved iSAM based on flexible re-linearization threshold and error learning model for AUV in large scale areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7678–7687, Dec. 2021.
- [5] R. Cui, Y. Li, and W. Yan, "Mutual information-based multi-AUV path planning for scalar field sampling using multidimensional RRT," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 993–1004, Jul. 2016.
- [6] J. Zhang, M. Liu, S. Zhang, R. Zheng, and S. Dong, "Multi-AUV adaptive path planning and cooperative sampling for ocean scalar field estimation," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.
- [7] X. Yu et al., "Path planning in multiple-AUV systems for difficult target traveling missions: A hybrid metaheuristic approach," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 3, pp. 561–574, Sep. 2020.
- [8] P. Yao, L. Qiu, J. Qi, and R. Yang, "AUV path planning for coverage search of static target in ocean environment," *Ocean Eng.*, vol. 241, Dec. 2021, Art. no. 110050.
- [9] R. A. Khalil, N. Saeed, M. I. Babar, and T. Jan, "Toward the Internet of Underwater Things: Recent developments and future challenges," *IEEE Consum. Electron. Mag.*, vol. 10, no. 6, pp. 32–37, Nov. 2021.
- [10] S. Heshmati-Alamdari, A. Nikou, and D. V. Dimarogonas, "Robust trajectory tracking control for underactuated autonomous underwater vehicles in uncertain environments," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 1288–1301, Jul. 2021.
- [11] J. Wang, Z. Wu, M. Tan, and J. Yu, "3-D path planning with multiple motions for a gliding robotic dolphin," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 5, pp. 2904–2915, May 2021.
- [12] K. Xue, J. Liu, N. Xiao, X. Ji, and H. Qian, "A bio-inspired simultaneous surface and underwater risk assessment method based on stereo vision for USVs in nearshore clean waters," *IEEE Robot. Autom. Lett.*, vol. 8, no. 1, pp. 360–367, Jan. 2023.
- [13] M. Zhang, W. Cai, Q. Xie, and S. Xu, "Binocular-vision-based obstacle avoidance design and experiments verification for underwater quadcopter vehicle," *J. Mar. Sci. Eng.*, vol. 10, no. 8, p. 1050, Jul. 2022.
- [14] Y. Jin, B. Ku, J. Ahn, S. Kim, and H. Ko, "Nonhomogeneous noise removal from side-scan sonar images using structural sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1215–1219, Aug. 2019.
- [15] J. Abbott, "Acoustic speckle: Theory and experimental analysis," *Ultrason. Imag.*, vol. 1, no. 4, pp. 303–324, Oct. 1979.
- [16] E. Ochoa, G. Nuno, I. Klemen, B. Josep, C. Patryk, and G. Rafael, "Collision detection and avoidance for underwater vehicles using omnidirectional vision," *Sensors*, vol. 22, no. 14, p. 5354, Jul. 2022.
- [17] D. Youakim, P. Cieslak, A. Dornbush, A. Palomer, P. Ridao, and M. Likhachev, "Multirepresentation, multiheuristic A* search-based motion planning for a free-floating underwater vehicle-manipulator system in unknown environment," *J. Field Robot.*, vol. 37, no. 6, pp. 925–950, Sep. 2020.
- [18] P. Zhu, S. Liu, T. Jiang, Y. Liu, X. Zhuang, and Z. Zhang, "Autonomous reinforcement control of visual underwater vehicles: Real-time experiments using computer vision," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8237–8250, Aug. 2022.
- [19] P. Zhu, S. Yao, Y. Liu, S. Liu, and X. Liang, "Autonomous reinforcement control of underwater vehicles based on monocular depth vision," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 9201–9206, 2020.
- [20] X. Yang, Y. Shi, W. Liu, H. Ye, W. Zhong, and Z. Xiang, "Global path planning algorithm based on double DQN for multi-tasks amphibious unmanned surface vehicle," *Ocean Eng.*, vol. 266, Dec. 2022, Art. no. 112809.
- [21] Y. Yang, Y. Xiao, and T. Li, "A survey of autonomous underwater vehicle formation: Performance, formation control, and communication capability," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 815–841, 2nd Quart., 2021.
- [22] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–14.
- [23] T. Manderson, J. C. G. Higuera, R. Cheng, and G. Dudek, "Vision-based autonomous underwater swimming in dense coral for combined collision avoidance and target selection," in *Proc. IEEE/RJS Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1885–1891.
- [24] Y.-C. Chou, H.-H. Chen, C.-C. Wang, H.-M. Chou, and C.-C. Wang, "An AI AUV enabling vision-based diver-following and obstacle avoidance with 3D-modeling dataset," in *Proc. IEEE 3rd Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, Jun. 2021, pp. 1–4.
- [25] G. Han, X. Qi, Y. Peng, C. Lin, Y. Zhang, and Q. Lu, "Early warning obstacle avoidance-enabled path planning for multi-AUV-based maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 2, pp. 2656–2667, Feb. 2023, doi: 10.1109/TITS.2022.3157436.
- [26] Z. He, L. Dong, C. Sun, and J. Wang, "Asynchronous multithreading reinforcement-learning-based path planning and tracking for unmanned underwater vehicle," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 5, pp. 2757–2769, May 2022.
- [27] M. C. Domingo, "Magnetic induction for underwater wireless communication networks," *IEEE Trans. Antennas Propag.*, vol. 60, no. 6, pp. 2929–2939, Jun. 2012.
- [28] Y. Yan and Y. Mostofi, "Robotic router formation in realistic communication environments," *IEEE Trans. Robot.*, vol. 28, no. 4, pp. 810–827, Aug. 2012.
- [29] D. B. Licea, M. Bonilla, M. Ghogho, S. Lasaulce, and V. S. Varma, "Communication-aware energy efficient trajectory planning with limited channel knowledge," *IEEE Trans. Robot.*, vol. 36, no. 2, pp. 431–442, Apr. 2020.
- [30] U. Ali, H. Cai, Y. Mostofi, and Y. Wardi, "Motion-communication co-optimization with cooperative load transfer in mobile robotics: An optimal control perspective," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 2, pp. 621–632, Jun. 2019.
- [31] J. Yan, W. Cao, X. Yang, C. Chen, and X. Guan, "Communication-efficient and collision-free motion planning of underwater vehicles via integral reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 13, 2022, doi: 10.1109/TNNLS.2022.3226776.
- [32] W. Cao, J. Yan, X. Yang, X. Luo, and X. Guan, "Communication-aware formation control of AUVs with model uncertainty and fading channel via integral reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 1, pp. 159–176, Jan. 2023.
- [33] USBL. Accessed: Dec. 2022. [Online]. Available: <https://www.waterlinked.com/underwater-gps>
- [34] J. Yan, X. Li, X. Yang, X. Luo, C. Hua, and X. Guan, "Integrated localization and tracking for AUV with model uncertainties via scalable sampling-based reinforcement learning approach," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 11, pp. 6952–6967, Nov. 2022.
- [35] R. Cui, C. Yang, Y. Li, and S. Sharma, "Adaptive neural network control of AUVs with control input nonlinearities using reinforcement learning," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 6, pp. 1019–1029, Jun. 2017.
- [36] Y. Zhang, Z. Zhang, L. Chen, and X. Wang, "Reinforcement learning-based opportunistic routing protocol for underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2756–2770, Mar. 2021.
- [37] J. Yan, Z. Zhang, X. Yang, X. Luo, and X. Guan, "Target detection in underwater sensor networks by fusion of active and passive measurements," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 4, pp. 2319–2333, Jul./Aug. 2023, doi: 10.1109/TNSE.2023.3246098.
- [38] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 1993–2016, Aug. 2017.
- [39] H. Barrett and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. Cambridge, MA, USA: Academic Press, 1996.
- [40] X. Yu, A. Khani, J. Chen, H. Xu, and H. Mao, "Real-time holding control for transfer synchronization via robust multiagent reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 23993–24007, Dec. 2022.
- [41] J. Peng, S. Zhang, Y. Zhou, and Z. Li, "An integrated model for autonomous speed and lane change decision-making based on deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21848–21860, Nov. 2022.
- [42] H. Li, H. Cheng, Z. Wang, and G.-C. Wu, "Distributed Nesterov gradient and heavy-ball double accelerated asynchronous optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5723–5737, Dec. 2021.
- [43] Q. Hu, B. Jiang, and M. I. Friswell, "Robust saturated finite time output feedback attitude stabilization for rigid spacecraft," *J. Guid., Control, Dyn.*, vol. 37, no. 6, pp. 1914–1929, Nov. 2014.

[44] *Video of the Experimental Results*. Accessed: Jun. 2023. [Online]. Available: <https://www.youtube.com/watch?v=s1WQRfjfvGg> and https://v.youku.com/v_show/id_XNTk3MTIxODY2OA===.html



Jing Yan (Senior Member, IEEE) received the B.Eng. degree in automation from Henan University, Kaifeng, China, in 2008, and the Ph.D. degree in control theory and control engineering from Yanshan University, Qinhuangdao, China, in 2014. He is currently a Full Professor with Yanshan University. He has authored over 60 refereed international journals and conference papers. Meanwhile, he has published two books and he is also the inventor of 17 patents. His research interests include underwater acoustic sensor networks and underwater

cyberphysical systems.

He received the Excellent Youth Project for NSF of China in 2022, the Distinguished Youth Project for NSF of Hebei Province in 2022, the Youth Top Talent Project for Hebei Education Department in 2018, and the Excellent Youth Project for NSF of Hebei Province in 2020. Currently, he serves as an Associate Editor for *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *IEEE SYSTEMS JOURNAL*, *Ocean Engineering*, *IET Control Theory and Applications*, and *Wireless Networks* (Springer). He also serves as the Career Associate Editor for *IEEE/CAA JOURNAL OF AUTOMATICA SINICA*.



Liang Zhang received the B.S. degree in automation from the Shandong University of Science and Technology, Qingdao, China, in 2021. He is currently pursuing the master's degree in control engineering with Yanshan University, Qinhuangdao, China. His research interests include motion planning control and deep learning. He won the First Prize in Hebei University Robot Competition and the Second Prize in the World University Underwater Robot Competition.



Xian Yang received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from Yanshan University, Qinhuangdao, China, in 2010 and 2016, respectively. She is currently an Associate Professor with Yanshan University. Her research interests include networked teleoperation systems, underwater cyberphysical systems, and nonlinear control.



Cailian Chen (Member, IEEE) received the B.Eng. and M.Eng. degrees in automatic control from Yanshan University, Qinhuangdao, China, in 2000 and 2002, respectively, and the Ph.D. degree in control and systems from the City University of Hong Kong, Hong Kong, in 2006. In 2008, she joined the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, as an Associate Professor, where she is currently a Full Professor. She has authored and/or coauthored two research monographs and over 100 refereed international journals and conference papers. She is the inventor of more than 20 patents. Her research interests include wireless sensor and actuator network and application in industrial automation, vehicular networks and application in intelligent transportation systems, and estimation and control for multiagent systems. She received the *IEEE TRANSACTIONS ON FUZZY SYSTEMS* Outstanding Paper Award in 2008. She was one of the First Prize Winners of the Natural Science Award from the Ministry of Education of China in 2006 and 2016. She was honored as a Changjiang Young Scholar by the Ministry of Education of China in 2015 and the Excellent Young Researcher by NSF of China in 2016.



Xinping Guan (Fellow, IEEE) is currently a Chair Professor with Shanghai Jiao Tong University, China, where he is the Deputy Director of the University Research Management Office and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. Before that, he was a Professor and the Dean of electrical engineering, Yanshan University, China. He is a Leader of the prestigious Innovative Research Team of the National Natural Science Foundation of China (NSFC). His current research

interests include industrial cyber-physical systems, wireless networking and applications in smart city and smart factory, and underwater sensor networks. He is an Executive Committee Member of the Chinese Automation Association Council and the Chinese Artificial Intelligence Association Council. He received the First Prize of Natural Science Award from the Ministry of Education of China in 2006 and 2016 and the Second Prize of the National Natural Science Award of China in 2008. He was a recipient of the *IEEE TRANSACTIONS ON FUZZY SYSTEMS* Outstanding Paper Award in 2008. He is a National Outstanding Youth honored by NSF of China, the Changjiang Scholar by the Ministry of Education of China, and the State-Level Scholar of New Century Bai Qianwan Talent Program of China.