J Bionic Eng 16 (2019) 442–454
DOI: https://doi.org/10.1007/s42235-019-0036-1

Journal of Bionic Engineering
http://www.springer.com/journal/42235

Robust RGB-D Camera and IMU Fusion-based Cooperative and Relative Close-range Localization for Multiple Turtle-inspired Amphibious Spherical Robots

Huiming Xing[1,2], Liwei Shi[1,2*], Kun Tang[1,2], Shuxiang Guo[1,2,3*], Xihuan Hou[1,2], Yu Liu[1,2], Huikang Liu[1,2], Yao Hu[1,2]

1. *Key Laboratory of Convergence Medical Engineering System and Healthcare Technology*, *the Ministry of Industry and Information Technology*, *School of Life Science*, *Beijing Institute of Technology*, *Beijing* 100081, *China*
2. *Key Laboratory of Biomimetic Robots and Systems*, *Ministry of Education*, *Beijing Institute of Technology*, *Beijing* 100081, *China*
3. *Faculty of Engineering*, *Kagawa University*, 2217-20 *Hayashicho*, *Takamatsu*, *Kagawa* 761-0396, *Japan*

Abstract

In the narrow, submarine, unstructured environment, the present localization approaches, such as GPS measurement, dead-reckoning, acoustic positioning, artificial landmarks-based method, are hard to be used for multiple small-scale underwater robots. Therefore, this paper proposes a novel RGB-D camera and Inertial Measurement Unit (IMU) fusion-based cooperative and relative close-range localization approach for special environments, such as underwater caves. Owing to the rotation movement with zero-radius, the cooperative localization of Multiple Turtle-inspired Amphibious Spherical Robot (MTASRs) is realized. Firstly, we present an efficient Histogram of Oriented Gradient (HOG) and Color Names (CNs) fusion feature extracted from color images of TASRs. Then, by training Support Vector Machine (SVM) classifier with this fusion feature, an automatic recognition method of TASRs is developed. Secondly, RGB-D camera-based measurement model is obtained by the depth map. In order to realize the cooperative and relative close-range localization of MTASRs, the MTASRs model is established with RGB-D camera and IMU. Finally, the depth measurement in water is corrected and the efficiency of RGB-D camera for underwater application is validated. Then experiments of our proposed localization method with three robots were conducted and the results verified the feasibility of the proposed method for MTASRs.

Keywords: vision localization, bio-inspired robots, RGB-D camera, histogram of oriented gradient and color names fusion feature, Cooperative and Relative Localization (CRL)

1 Introduction

In recent years, owing to the increasing demand of ocean, oceanographic exploration and research are still challenging activities. In the open ocean, many methods can realize the cooperative localization of multiple AUVs or underwater robots[1,2]. However, in the narrow, submarine, unstructured environments, it is hard to realize Cooperative and Relative Localization (CRL), especially for small-scale robots. While there are many mineral resources in these narrow environments that are inaccessible to human directly, it's essential to study the precise localization of multiple small-scale underwater robots in such environment.

The underwater localization mainly can be divided into two categories — localization in structured 3D environment and unstructured 3D environment. In structured 3D environment, many approaches, for example, artificial visual landmarks-based method are utilized. Kim *et al.* realized the vision-based localization techniques[3] with artificial landmarks in structured underwater environments. To mitigate these problems of low visibility, noise and large areas of featureless scene, Kim *et al.* designed artificial landmarks to be utilized with a camera for localization, and proposed a novel vision-based object detection technique and applied it to the Monte Carlo Localization (MCL) algorithm, a map-based localization technique. Carreras *et al.* proposed a vision-based localization approach[4,5] to estimate the position and orientation of an underwater robot in a structured environment. The down-looking camera attached to the robot was used to detect the landmarks in a

***Corresponding author:** Liwei Shi, Shuxiang Guo
E-mail: shiliwei@bit.edu.cn, guoshuxiang@bit.edu.cn

coded pattern placed on the bottom of a water tank. This method has lots of limitations that the experimental setup needs to be arranged elaborately, which is unable to be used in the unstructured 3D environment.

In unstructured 3D environment, the Remotely-Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) with the acoustic localization system, GPS measurement[6,7] above water or the dead-reckoning approach, especially in deep-water operation, are essential for application as varied environmental surveying, geology, archeology and cable inspection. Navinda *et al.* analyzed the disadvantage of radio communication between UUVs and introduced the method to synchronize data transmission[8] for Cooperative Localization (CL) with the use of Pulse Per Second (PPS) signal. The recent experiments[9] about the development of a synchronous-clock acoustic localization system which was suitable for CL of multiple UUVs were reported. A delayed extended Kalman filter which is designed to deal with the cooperative localization problem with acoustic communication delay[10] is presented. Matsuda *et al*. proposed the Alternating Landmark Navigation (ALN)[11] by multiple AUVs. Each AUV alternately became a landmark, which remains stationary on the seafloor, whereas the other AUVs can navigate based on acoustic positioning with the landmark AUV.

Localization of unmanned underwater vehicles typically makes use of an Inertial Measurement Unit (IMU) in combination with a Doppler Velocity Log (DVL). If the direction is given by the IMU, the position can be calculated with the velocity provided by DVL. Such a navigation system called dead-reckoning also exists for divers[12]. DVL requires sufficient beam measurements (at least three) to calculate the 3-D velocity. However, in cases that DVL only has limited (few than three) beam measurements in the underwater environment. Liu *et al*. proposed a method based on the tightly coupled approach for INS/DVL integrated navigation with limited DVL beam measurements[13]. But the devices of this method are bulky, heavy and expensive. The other shortcoming is that the error of positioning will increase with time, which cannot be avoided, and just reduced with advanced algorithms[14]. Therefore, the dead-reckoning localization cannot be suitable for small-scale underwater robots.

Although the acoustic positioning system[15] and dead-reckoning with IMU and DVL are quite precise, these localization approaches are not suitable for the close-range surveying in unstructured 3D environment, such as submarine caves and narrow passages. Recently, there are many researches on cooperative close-range localization, such as artificial visual landmarks-based method, light beacons-based method and so on. Josep *et al.* proposed a new close-range tracking system[16,17] for Autonomous Underwater Vehicles (AUVs) navigating in a close formation, using computer vision and active light beacons. This proposed system allows the estimation of the pose and location of a target vehicle at short ranges with an omnidirectional camera in an extended field of view. Matthias *et al.* proposed a new pose estimation system[18] that consists of multiple infrared LEDs and cameras with an infrared-pass filter. The observing ground robot was equipped with a camera to track a quadrotor attached with LEDs. The infrared LEDs can be detected by vision system, thus their position on the target object can be precisely determined. However, if the infrared LEDs were sheltered by the body of the robot or others, the pose of the robot is unable to be estimated.

In our previous work, we proposed an amphibious spherical robot[19–21] inspired by turtles. In order to carry more sensors and improve multi-locomotion performance for the exploration in amphibious environments, a novel turtle-inspired amphibious spherical robot[22] was redesigned. Based on multiple amphibious spherical robots, rather than the robotic fish[23] or other bio-inspired robots[24–29], this paper proposed a novel RGB-D camera and IMU fusion-based cooperative and relative close-range localization approach. Because the spherical shape can keep TASRs rotate flexibly along the plumb line with zero-radius[30], which causes less shift in submarine 3-D space. These capabilities of TASRs contribute to the proposed localization approach. The main contribution is that the approach is able to be used in the narrow, submarine, unstructured environments. Other methods, such as GPS measurement, dead-reckoning, acoustic positioning, artificial landmarks-based method, are hard to be used to these environments.

The reminder of this paper is organized as follows:

section **2** introduces an amphibious spherical robot, the waterproof structure of RGB-D camera and the advanced electric system. In section **3**, we present the procedure of an automatic detection approach with HOG and CNs fusion feature and SVM classifier for the TASR. The MTASRs model and cooperative and relative close-range localization approach for the MTASRs are proposed in section **4**. Experiments of cooperative and relative close-range localization approach with three robots are conducted in section **5** and the performance evaluation of this localization method is presented. Finally, section **6** concludes this paper.

2 The turtle-inspired amphibious spherical robot

In order to offer more space for sensors and improve the stability and velocity performance in amphibious environment, we proposed a multiple locomotion-based amphibious spherical robot. The robot was inspired by turtles. As shown in Fig. 1, the amphibious robot includes a hemispheroid waterproof hull, two quarter spherical hull, a middle plate, four mechanical legs, a lifting and supporting wheel mechanism, the electronic circuit, and five batteries. The size of our spherical robot is 350 mm. The actuator system of the robot is wheel-legged composite driving mechanism including the LSWM and mechanical legs. In water, the two quarter spherical hulls can close like a ball and the robot can move by four water-jet thrusters. On land, the two hulls open and the robot can realize the sliding locomotion mode and the walking locomotion mode with the wheel-legged composite driving mechanism. In order to make a steadily electrical system of the robot, a stable and powerful PSU with the electric isolation is essential. When the robot begins to stand up, eight servos are supplied electricity simultaneously, which triggers the dramatic surge current. The instantaneous and huge current is a huge challenge for the controller and sensors. In order to realize the electric isolation of control electricity and dynamic electricity, the control electricity is not common-grounded with the dynamic electricity, and the control part outputs the Pulse Width Modulation (PWM) signals to the motors in the dynamic part *via* optocouplers which can realize photoelectric signal transformation. As shown in Fig. 1, two main

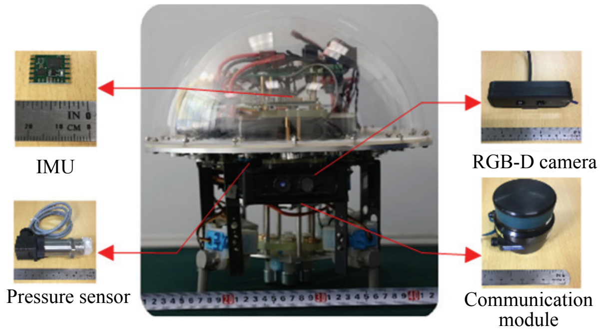

Fig. 1 Block of the novel turtle-inspired amphibious spherical robot (TASR).

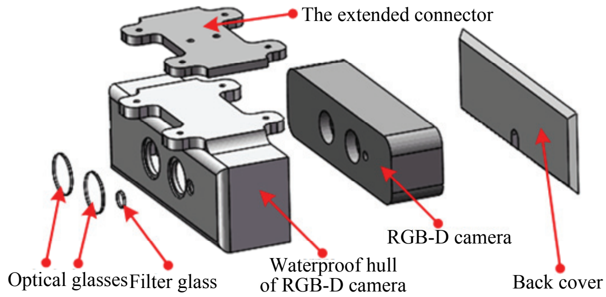

Fig. 2 The waterproof structure of the RGB-D camera.

sensors, RGB-D and IMU, are used. And for realizing the communication of multiple robots, a small-size communication modem is utilized in this paper.

As shown in Fig. 2, a Softkinetic RGB-D camera which can capture color images with 720p resolution and depth map images with 320 × 240 resolution at 25 fps – 60 fps is adopted to the robot to perceive the environment. The depth noise of the camera is under 1.4 cm at 1 m for indoor application, which is excellent in the prevalent RGB-D cameras. Recently, Song *et al.* proposed a robust vision-based relative-localization approach[31] for a moving target based on an RGB-D camera and sensor measurements from 2D light detection and ranging (LiDAR). Underwater application researches of the RGB-D camera are caused more attention, such as vision localization[32], visual tracking[33,34], point cloud imaging[35], 3D mapping and scene reconstruction[36,37]. In order to apply it into the submarine environment, a customized waterproof hull is made using 3D printing technology and mounted under the robot by the extended connector. Three pieces of optical glasses are fixed in the front of the RGB sensor, the depth sensor and the laser and diffuser. Considering the changes of the optical flying path, a calibration needs to be conducted, which will be described in section **4**.

3 The automatic recognition approach

In this section, an automatic recognition algorithm based on HOG-CNs fusion features and the SVM classifier is proposed. The framework of HOG and CNs detector is shown in Fig. 3. The features set used to detect the TASRs is the combination of two types of features (HOG and CNs). Firstly, the input images are normalized to be the size of 64 × 64. Secondly, for each image, we calculated HOG feature and CNs feature respectively. Thirdly, all HOG-CNs feature vectors are trained by SVM. Then a HOG-CNs detector can be achieved for testing. In the following part, an overview of the two features extraction is given.

3.1 HOG and CNs features fusion

For HOG feature extraction, an image is divided into $N \times N$ non-overlapping pixel regions, which are named as cells. For each cell, the histogram of the pixels is calculated. Through combining the histograms of individual cells, the local appearance of a patch is effectively represented. In our method, 31-dimensional histogram vector is used to describe each cell.

To embed the CNs feature into the HOG feature, the histogram of the CNs feature in each divided cell is calculated. Then, a new feature vector is constructed by extending the p-dimensional HOG vector with the 11-dimensional color names histogram vector. We extend the p-dimensional HOG vector with the 11-dimensional color names histogram vector to construct a new feature vector. For each detection window $\boldsymbol{W}_i$, the representation is obtained as:

$$\boldsymbol{W}_i = [\boldsymbol{HOG}_i, \boldsymbol{CNs}_i], \tag{1}$$

where $\boldsymbol{HOG}_i$ and $\boldsymbol{CNs}_i$ represent the ith detection window's HOG vector and CNs histogram vector. Thus, $(p + 11)$ dimension feature vector for each detection window is obtained. Through computing concatenated feature vectors of all detection window, a patch is represented.

3.2 Recognition method with SVM

Feature classification is the key step in the object recognition, and excellent feature classifier can increase the accuracy of the object recognition. Support Vector

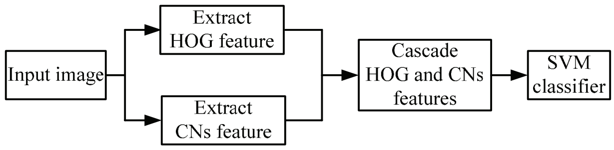

Fig. 3 The framework of HOG and CNs detector.

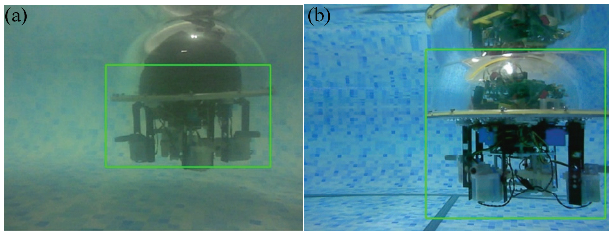

Fig. 4 The detection results in stained water and clear water. (a) Stained water; (b) clear water. Green boxes show our results.

Machine (SVM) classification is the most popular classification method in pattern recognition, especially in the binary classification. We trained a linear SVM using image patches of TASRs and non-TASRs. The basic definition of an SVM classifier is:

$$h(x) \leftarrow \operatorname{sgn}\left(\sum_{s_i} a_i \boldsymbol{y}_i \boldsymbol{s}_i \cdot \boldsymbol{x} + b\right), \ \operatorname{sgn} = \begin{cases} -1 & x < 0 \\ 0 & x = 0, \\ 1 & x > 1 \end{cases} \tag{2}$$

where, a_i is the Lagrange multiplier, $\boldsymbol{y}_i$ is the labels of the support vectors, $\boldsymbol{s}$ is the support vectors, $\boldsymbol{x}$ is our input feature vector, and b is the offset. In our robot detection, the feature space is non-separable, so we choose a soft margin SVM classifier. For saving the cost of the run time, a linear kernel is selected.

In detection of TASRs, one thousand TASRs color images were captured in the pool whose size is 3 m × 2 m × 1 m *via* RGB-D camera. Then five hundred images of the whole TASRs were chosen as positive samples, and five hundred images of underwater stone, coral and fish were selected as negative samples. In addition, two hundred images were collected as test samples which contain images of TASRs.

In order to show the effective of HOG-CNs fusion features, the SVM classier experiments based on HOG feature and HOG-CNs fusion features were carried out, respectively. Test samples contain 100 images of TASRs and 100 images of coral and fish. Images of TASRs were captured in multiple points of view and in different water quality environment (Fig. 4). As shown in Table 1, the recognition rate is 73.08% with HOG feature, and 89.26%

Table 1 The results of SVM classifier

Features	HOG	HOG-CNs
The recognition rate	73.08%	89.26%

Table 2 The intrinsic parameters table of the color camera

Camera	The intrinsic parameters			
	f_x	f_y	u_0	v_0
The color camera	652.495	654.334	339.072	242.607

with HOG-CNs fusion features. The recognition rate with HOG and CNs fusion features is improved 16.18% than the recognition rate with HOG feature, which proves that the recognition of TASRs with the proposed HOG-CNs fusion features has high accuracy, strong robustness and good results. The recognition results with HOG-CNs fusion features in stained water and clear water are shown in Figs. 4a and 4b, respectively.

4 Cooperative and relative close-range localization approach for MTASRs

In MTASRs, the Softkinetic RGB-D camera and IMU were adopted to realize the cooperative and relative localization for MTASRs in submarine environment. In the positioning of one robot (the master) to another (the slave), the relative RGB-D camera-based measurement model was established with the depth map captured by the RGB-D camera. Using the rotation movement[30] with zero-radius, the master is able to detect the slave and log self-pose estimates with IMU. With the RGB-D camera and IMU fusion, the MTASRs model is established. However, the camera is designed for on land application. Before the camera with the waterproof hull is used in water, the calibration of the color camera and the correction of the depth camera are carried out urgently.

4.1 The calibration of RGB-D camera in the submarine environment

In section **2**, we designed a 3-D printed waterproof hull which has two optical glasses to enclose the RGB-D camera, but optical glasses change the beam path. For the rigor of the scientific research, the RGB-D camera needs to be calibrated and the depth camera needs to be corrected. The color camera model can be obtained as Eq. (3) by the pinhole camera model.

$$Z_C\begin{bmatrix}u_C\\v_C\\1\end{bmatrix}=\begin{bmatrix}f_x^C&0&u_0^C&0\\0&f_y^C&v_0^C&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}X_C\\Y_C\\Z_C\\1\end{bmatrix}. \tag{3}$$

Using the waterproof hull with two optical glasses, the radial distortion and tangential distortion are very strong. The corrected equation for radial distortion and tangential distortion can be derived from the following Eqs. (4) and (5), respectively.

$$\begin{bmatrix}x_c\\y_c\end{bmatrix}=(1+k_1r^2+k_2r^4+k_3r^6)\cdot\begin{bmatrix}x\\y\end{bmatrix}, \tag{4}$$

$$\begin{bmatrix}x_c\\y_c\end{bmatrix}=\begin{bmatrix}x+2p_1y+p_2(r^2+2x^2)\\y+p_1(r^2+2y^2)+2p_2x\end{bmatrix}, \tag{5}$$

where $[x, y]^{\mathrm{T}}$ is the original location of the distorted point, $r^2 = x^2 + y^2$, and $[x_c, y_c]^{\mathrm{T}}$ is the corrected location, k_1, k_2, k_3, p_1 and p_2 are distortion coefficients.

In order to calculate the color camera, OpenCV software development kit is utilized. The intrinsic parameters (the focal length and the principal point) are shown in Table 2. Table 3 gave the five distortion parameters computed for the color camera.

The calibration of the color camera and the depth camera is the fundamental to extract metric information from 2D images. The RGB-D camera was calibrated by calculating the rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{T}$ from the right camera (depth camera) to the left one (color camera) as shown in Fig. 5. 3D point $\boldsymbol{P}_W$ in the world coordinate frame $\mathcal{F}^W$ ($[X_W, Y_W, Z_W, 1]^{\mathrm{T}}$) is represented to be $\boldsymbol{P}_C$ in the color camera coordinate frame ${}_i\mathcal{F}^C$ ($[X_C, Y_C, Z_C, 1]^{\mathrm{T}}$) and $\boldsymbol{P}_D$ in the depth camera coordinate frame ${}_i\mathcal{F}^D$ ($[X_D, Y_D, Z_D, 1]^{\mathrm{T}}$) of Robot i, respectively. Specifically, $\boldsymbol{R}$ and $\boldsymbol{T}$ can be derived by the following equation as described in Eq. (6).

$$\begin{bmatrix}\boldsymbol{P}_C\\1\end{bmatrix}=\begin{bmatrix}\boldsymbol{R}&\boldsymbol{T}\\0^{\mathrm{T}}&1\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_D\\1\end{bmatrix}. \tag{6}$$

The depth camera model can be given by Eq. (7).

$$Z_D\begin{bmatrix}u_D\\v_D\\1\end{bmatrix}=\begin{bmatrix}f_x^D&0&u_0^D&0\\0&f_y^D&v_0^D&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}X_D\\Y_D\\Z_D\\1\end{bmatrix}=\begin{bmatrix}\boldsymbol{M_D}&0\end{bmatrix}\begin{bmatrix}\boldsymbol{P_D}\\1\end{bmatrix}. \tag{7}$$

Table 3 Distortion parameters table of the color camera

Camera	Distortion parameters				
	k_1	k_2	k_3	p_1	p_2
The color camera	0.0989159	4.44508	−22.9803	−0.000654106	0.0217542

$$\begin{cases} \boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 0.999991 & 0.001237 & 0.003848 \\ 0.001284 & -0.999926 & -0.012077 \\ -0.003832 & -0.012082 & 0.999197 \end{bmatrix} \\ \boldsymbol{T} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} 0.026000 \\ -0.000508 \\ -0.000863 \end{bmatrix} \end{cases}, \tag{8}$$

$$(Z_D r_{33} + t_3)\begin{bmatrix} u_C \\ v_C \\ 1 \end{bmatrix} = \begin{bmatrix} f_x^C & 0 & u_0^C & 0 \\ 0 & f_y^C & v_0^C & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}_C & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_C \\ 1 \end{bmatrix}. \tag{10}$$

Similarly, based on the OpenCV library, the parameters of calibration can be given by Eq. (8).

Eq. (9) of the relation between Z_C and Z_D can be obtained by the Eqs. (6) and (8).

$$Z_C = Z_D r_{33} + t_3. \tag{9}$$

Eq. (3) can be *revised* to be Eq. (10).

The pixel coordinate (u_C, v_C) of the object in the color camera coordinate frame can be acquired by the method mentioned in the section 3. With Eq. (11), the pixel coordinate (u_D, v_D) of the object in the depth camera coordinate frame will be obtained:

$$\begin{bmatrix} u_D \\ v_D \\ 1 \end{bmatrix}^{\mathrm{T}} \boldsymbol{F} \begin{bmatrix} \upsilon_C \\ v_C \\ 1 \end{bmatrix} = 0, \tag{11}$$

where $\boldsymbol{F}$ is the fundamental matrix which can be calculated using the OpenCV library.

The value Z_D can be get by the depth camera using the pixel coordinate (u_D, v_D). Therefore, the coordinate $\boldsymbol{P}_C$ of the only point of the 3-D space relative to the color camera can be calculated using Eq. (10).

4.2 The correction of the depth camera in the submarine environment

The depth measurement principle of RGB-D camera is based on the flight time of the infrared ray emitted by infrared emitter in the medium of RGB-D camera

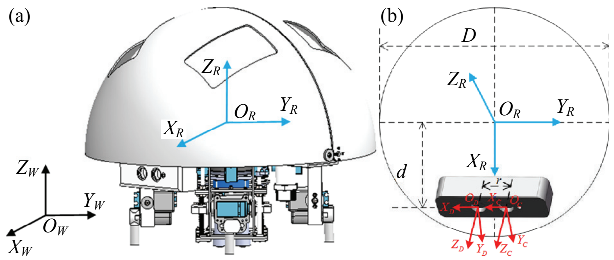

Fig. 5 The amphibious spherical robots in Cartesian coordinates representation.

which is used in air rather than in the submarine environment. If the RGB-D camera is used in water, the double attenuation between the lens and water will lead to the large offset of the depth data. To ensure the accuracy of depth measurement, the elimination of the offset is a crucial step. Therefore, the experiment needs to be conducted to correct the depth data by the experience compensation.

The depth measurement correction structure is shown in Fig. 6. The RGB-D camera with the waterproof structure is sunk into the water through a sliding bracket, which can slide with the top pulley and change the distance from the camera to the object. The distance measuring tools are equipped on the top pulley runner and the bottom of the pool parallel to the ruler. In this experiment, the target is a large plane plate, the plate is parallel to the plane of the camera and placed vertically at the bottom of the pool. The experimental environment

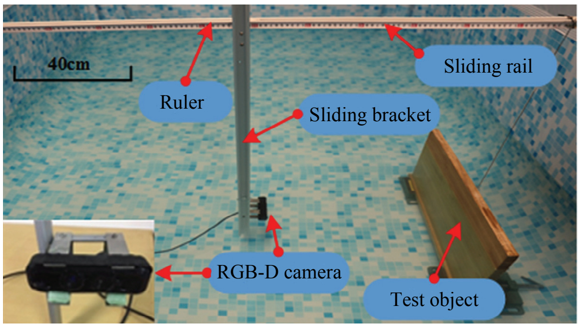

Fig. 6 The experiment setup of the depth camera correction.

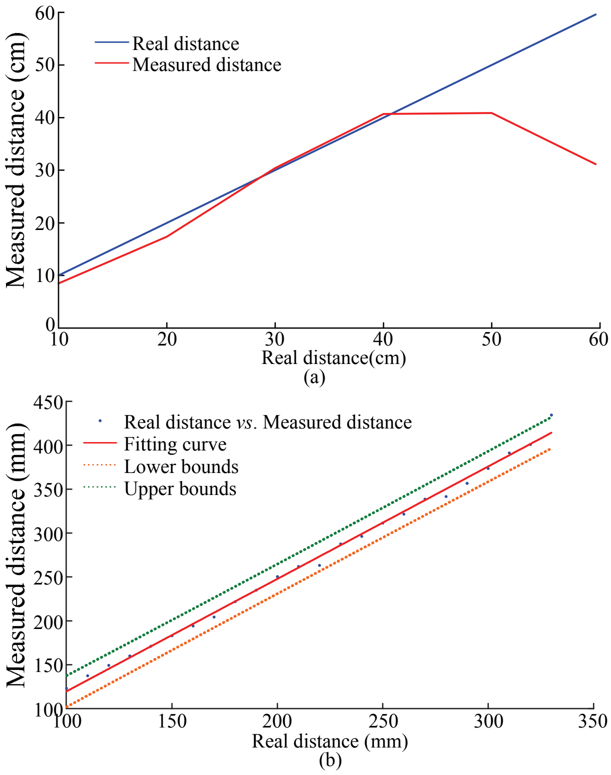

Fig. 7 The correction of the depth camera in water. (a) The data ranging 10 cm to 60 cm; (b) the data ranging 10 cm to 40 cm.

is an indoor pool whose size is 3.0 m × 2.0 m × 1.00 m, and the depth of water is 0.7 m.

In our experiment of distance calibration, initially, the camera starts from the minimum imaging distance (*i.e.*, the minimum effective depth data), and the sampling interval is 10 cm, 10 sets of data are recorded every time. The results are shown in Fig. 7a. We can find that, with the relative distance between the camera and the measured object increases from zero, the depth data appeared about 10 cm, and the depth data decay dramatically after 40 cm. The effective range of RGB-D camera in water is about 10 cm – 40 cm.

In order to increase the accuracy of the depth data, we change the sampling interval as 1 cm in the effective range. The result is shown in Fig. 7b. It can be seen that the attenuation caused by the lens and the water is obvious, but the response of the depth camera was appropriate linear. Therefore, we establish the composite experience compensation error correction data model. The empirical formula is following as Eq. (12):

$$D_{\text{corrected}} = 0.778 \times D_{\text{messured}} + 7.289. \tag{12}$$

4.3 The proposed MTASRs localization

In section **4.1**, the measured position of the object relative to the color camera is mentioned. However, we need to know the measured coordinate of one robot relative to the other. As shown in Fig. 8, the world coordinate frame $\mathcal{F}^W$ ($[X_W, Y_W, Z_W]^{\mathrm{T}}$), the body coordinate frame ${}_i\mathcal{F}^R$ (${}_i[X_R, Y_R, Z_R]^{\mathrm{T}}$), the color camera coordinate frame ${}_i\mathcal{F}^C$ (${}_i[X_C, Y_C, Z_C]^T$) and the depth camera coordinate frame ${}_i\mathcal{F}^D$ (${}_i[X_D, Y_D, Z_D]^T$) were given. In order to evaluate the extended localization method of MTASRs, the coordinate frame $\mathcal{F}^{W_i'}$ ($\left[X_{W_i'}, Y_{W_i'}, Z_{W_i'}\right]^{\mathrm{T}}$) was defined, and the axes are parallel to the axes of the world coordinate frame using a rotation matrix and a translation vector.

Considering the close measuring distance of RGB-D camera in the submarine, the size R of the robot, the distance d from the camera to the center of the robot and the distance r between the color camera and the depth camera shown in Fig. 5 cannot be ignored. The rigid transformation between the body coordinate frame and the color camera coordinate frame is described in Eq. (13).

$$\begin{bmatrix} {}_R\boldsymbol{P}_C \\ 1 \end{bmatrix} = \begin{bmatrix} {}_C^R\boldsymbol{R} & {}_C^R\boldsymbol{T} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} {}_C\boldsymbol{P}_C \\ 1 \end{bmatrix} = {}_C^R\boldsymbol{S} \begin{bmatrix} {}_C\boldsymbol{P}_C \\ 1 \end{bmatrix}, \tag{13}$$

where ${}_{C_i}^{R_i}\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and ${}_C^R\boldsymbol{T} = \begin{bmatrix} d \\ r/2 \\ 0 \end{bmatrix}$ are the rotation matrix and translation vector of Robot i from the color camera coordinate frame to the body coordinate frame, respectively. ${}_R\boldsymbol{P}_C$ presents the color camera position in the body coordinate frame.

As shown in Fig. 8, Robot j is in the observation scope of Robot i. Robot i and Robot j were named as the

$$\begin{bmatrix} {}_{R_j}\boldsymbol{P}_{C_i} \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{r_j}_{C_i}\boldsymbol{R} & {}^{r_j}_{C_i}\boldsymbol{T} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} {}_{C_i}\boldsymbol{P}_{C_i} \\ 1 \end{bmatrix} + \begin{bmatrix} {}^{R}_{r}\boldsymbol{T} \\ 1 \end{bmatrix} = {}^{r_j}_{C_i}\boldsymbol{S} \begin{bmatrix} {}_{C_i}\boldsymbol{P}_{C_i} \\ 1 \end{bmatrix} + \begin{bmatrix} {}^{R}_{r}\boldsymbol{T} \\ 1 \end{bmatrix}, \tag{14}$$

$$\begin{bmatrix} {}_{R_j}\boldsymbol{P}_{R_i} \\ 1 \end{bmatrix} = {}^{r_j}_{C_i}\boldsymbol{S}\,{}^{R_i}_{C_i}\boldsymbol{S}^{-1} \begin{bmatrix} {}_{R_i}\boldsymbol{P}_{C_i} \\ 1 \end{bmatrix} + \begin{bmatrix} {}^{R}_{r}\boldsymbol{T} \\ 1 \end{bmatrix} = {}^{r_j}_{R_i}\boldsymbol{S} \begin{bmatrix} {}_{R_i}\boldsymbol{P}_{C_i} \\ 1 \end{bmatrix} + \begin{bmatrix} {}^{R}_{r}\boldsymbol{T} \\ 1 \end{bmatrix}, \tag{15}$$

$${}^{W_i'}_{C_i}\boldsymbol{R} == \begin{bmatrix} \cos\varphi_i\cos\theta_i & -\sin\varphi_i\cos\phi_i + \cos\varphi_i\sin\theta_i\sin\phi_i & \sin\varphi_i\sin\phi_i + \cos\varphi_i\sin\theta_i\cos\phi_i \\ \sin\varphi_i\cos\theta_i & \cos\varphi_i\cos\phi_i + \sin\varphi_i\sin\theta_i\sin\phi_i & -\cos\varphi_i\sin\phi_i + \sin\varphi_i\sin\theta_i\cos\phi_i \\ -\sin\theta_i & \cos\theta_i\sin\phi_i & \cos\theta_i\cos\phi_i \end{bmatrix}. \tag{16}$$

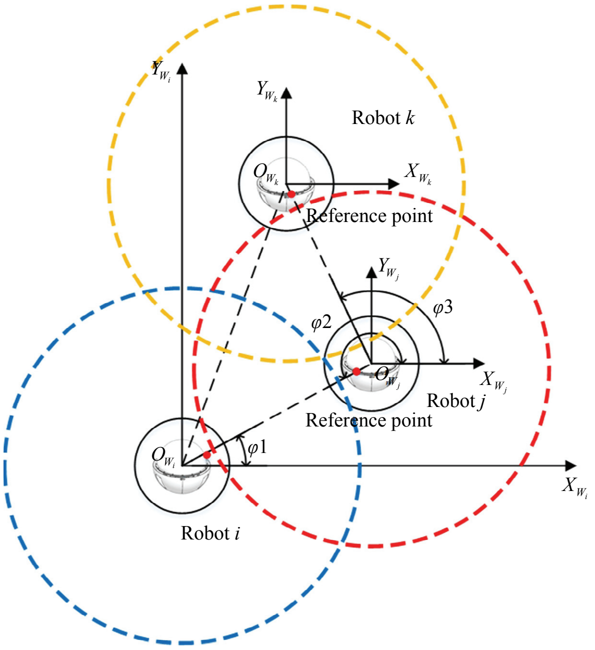

Fig. 8 Multiple turtle-inspired amphibious spherical robots in Cartesian coordinates representation. The red point indicates the reference point.

master robot and slave robot, respectively. In order to get the position of Robot *j*, a reference point is defined. As shown in Fig. 8, the red point indicates the reference point. Then, the position of Robot *j* can be obtained by Eq. (14) *via* the transformation from Robot *i* to Robot *j*. In Eq. (14), ${}^{r_j}_{C_i}S$ is the transformation from the color camera of robot *i* to the reference point of Robot *j*. ${}^{R}_{r}\boldsymbol{T} = [D/2\ 0\ 0]^{\mathrm{T}}$ is the translation vector from the reference point to the center of the body. In the localization of Robot *i* to Robot *j*, Robot *i* can rotate and detect Robot *j*. The pose of Robot *i* is adjusted until Robot *j* is stable and in the center of the color image. And then, the position of Robot *j* is calculated.

Combining Eqs. (13) and (14), the relationship between Robot *i* and Robot *j* can be described in Eq. (15). In Eq. (15), ${}^{r_j}_{R_i}\boldsymbol{S}$ is the transformation matrix between Robot *i* and the reference point of Robot *j*.

Define the parameters ϕ_i, φ_i and θ_i as the roll, yaw and pitch angles provided by MEMS IMU, respectively. Then, the body coordinate frame of Robot *i* can be translated into the world coordinate frame using Eq. (16).

Assumption *I* (planar localization): the cooperative and realative localization is such that:

(1) The robots in aquatic environment have the same depth, *i.e.*, the robots are in the horizontal planar.

(2) The motion is just around the *Z*-direction (yaw rotations), and the radius of yaw rotation is zero, *i.e.*, there is no translation in the horizontal plane.

(3) The roll and pitch angles are left as degree of freedom for motion, but in the steady-state rotation, they are equal to zero.

Therefore, Eq. (16) can be simplified as Eq. (17).

$${}^{W_i'}_{R_i}\boldsymbol{R} = \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{17}$$

Then, the coordinate of Robot *i* can be obtained by Eq. (18).

$$\begin{bmatrix} {}_{W_j'}\boldsymbol{P}_{R_i} \\ 1 \end{bmatrix} = {}^{W_j'}_{R_i}\boldsymbol{R} \begin{bmatrix} {}_{R_j}\boldsymbol{P}_{R_i} \\ 1 \end{bmatrix}. \tag{18}$$

5 Evaluating experiments of cooperative and relative close-range localization for the MTASRs

The localization experiments are conducted in the pool with the size of 3.0 m × 2.0 m × 1.0 m. Three rulers with the size of 1.5 m and 2.0 m in the bottom of the pool

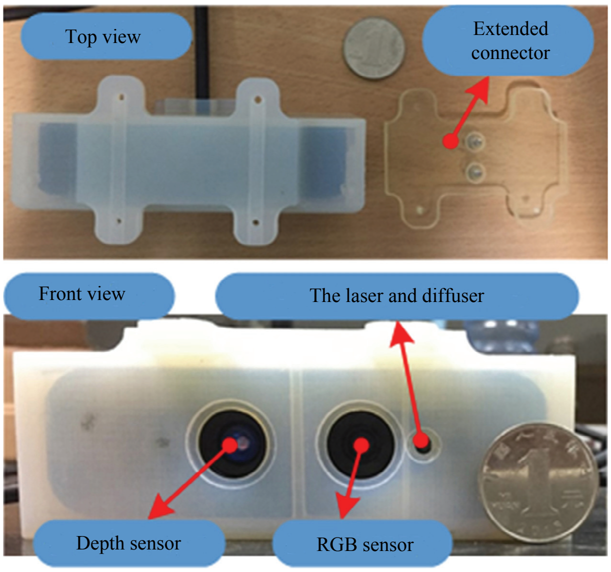

Fig. 9 The waterproof hull of the RGB-D camera.

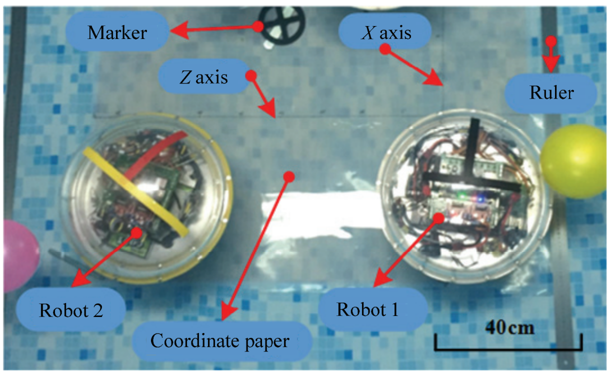

Fig. 10 The experiment setup of the localization experiment with two robots.

are used to keep the position of the amphibious spherical robot. In experiments, two or three amphibious spherical robots configured RGB-D cameras were adopted. As shown in Fig. 9, the waterproof hull of the RGB-D camera is produced using 3-D printing technology.

5.1 Experiments of RGB-D camera-based localization with two robots

As mentioned in the experiment of the depth camera correction (section **4.2**), we learned that the distance measurement of RGB-D camera is effective at a close-range distance. To evaluate the efficiency of positioning the slave robot, this experiment with two robots was conducted. In this experiment, two robots were arranged on a coordinate paper. As shown in Fig. 10, Robot 1 was placed on the original point marked with a

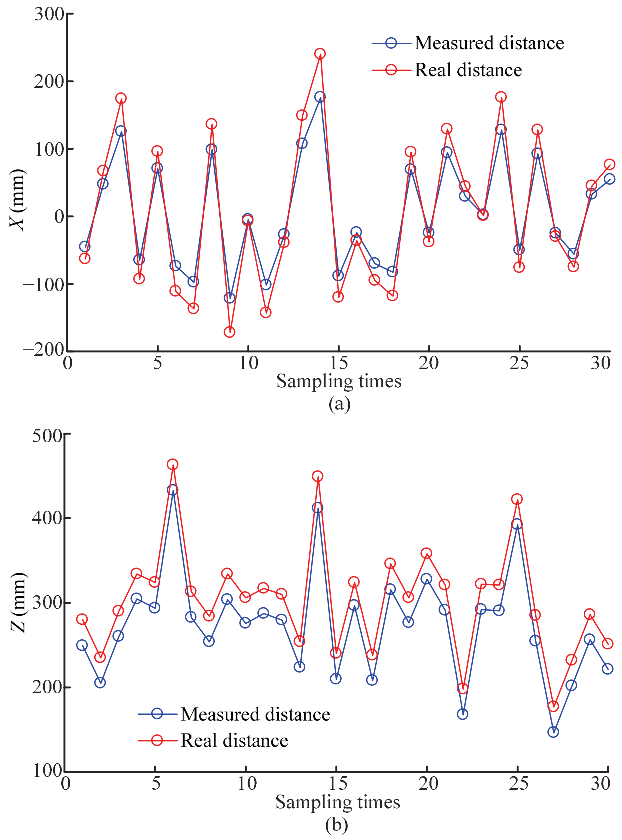

Fig. 11 Localization results with two robots. (a) The curve of *x* coordinate; (b) the curve of *z* coordinate.

Table 4 Underwater localization errors in *X* and *Y* axial directions

	X (Calibrated)	*Y* (Calibrated)	*Z* (Corrected)
Mean deviations (mm)	27.569	9.292	30.272
Maximum absolute deviations (mm)	63.852	30.395	37.151
Standard deviations	15.353	8.883	2.438

yellow ball on the coordinate paper. Robot 2 was put in the valid range of Robot 1, and the position of Robot 2 was marked with a red ball.

With the movement of Robot 2, we recorded the images with the current position for localization evaluation. We recorded 30 groups of position points. Compared the localization results using the original camera parameters and the calibrated parameters, Figs. 11a and 11b demonstrate 30 groups of data. The development trend between the measured data by Robot 1 and the reference has high regularity. Table 4 shows underwater localization errors including the average values, the maximum deviation and the standard deviation in *X* and *Z* axes. The maximum deviations in *X* and *Y* directions are 63.852 mm and 30.395 mm. Due to the shape of the TASRs and the experiment setup, the resistance of the rotation motion is smaller than that of the floating and sinking motion. Therefore, the fluctuation in the rotation motion is much easier, which caused that the maximum deviation in X is double than that in Y approximately. Besides, the recognition method and the measurement error also can result in the deviations. The maximum error and the standard deviation of depth measurement are 37.151 mm and 2.438 mm, respectively. However, compared with the size of TASR, it is insignificant. Obviously, the feasibility of the localization using RGB-D camera is validated.

5.2 Experiments of RGB-D camera and IMU-based localization with three robots

In order to realize the cooperative and relative localization of the MTASRs, this experiment with three robots validated the extension of vision localization *via* MEMS IMU. We know that MTASRs localization system can be realized with the normalization of coordinates. In our lab pool, initially, three robots were set to be a shape of an obtuse triangle shown in Fig. 12. The coordinate frame of Robot 1 was set in the original point of coordinate paper and set as the world coordinate frame $\mathcal{F}^W$. The distances from Robot 1 to Robot 3 and from Robot 1 to Robot 2, respectively, are all 30 cm and the distance between Robot 1 and Robot 3 is 50 cm. Robot 1 is in X axis of the body coordinate frame of Robot 2, and Robot 2 is also in X axis of the body coordinate frame of Robot 1 and Robot 3, respectively. Firstly, the relative localization between Robot 1 and Robot 2 was carried out simultaneously. After that, Robot 2 rotated automatically with vision-based PID control until Robot 3 was in X axis of the body coordinate frame of robot 2 steadily. Then the localization program between Robot 2 and Robot 3 was conducted. Finally, the cooperative localization of three robots was realized.

In the whole localization, three robots all perform self-pose estimates with Kalman Filter (KF), which is essential for the relative localization. Owing to none rotation of Robot 1 and Robot 3, Fig. 13 only shows self-pose estimates of Robot 2, *i.e.*, the roll (blue line), pitch (red line) and yaw (green line) angles. During the detection and positioning of Robot 2, the roll and pitch

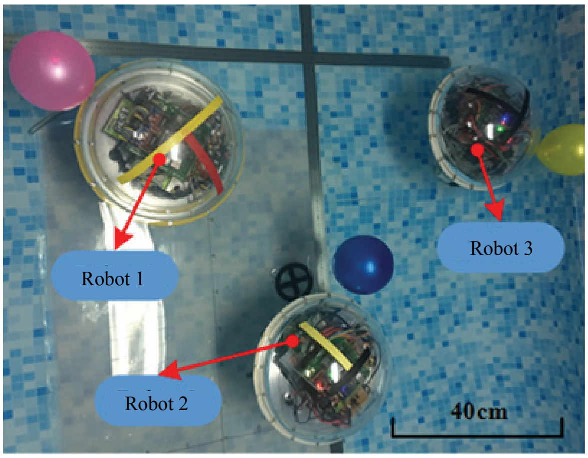

Fig. 12 The experiment setup of the localization experiment with three robots.

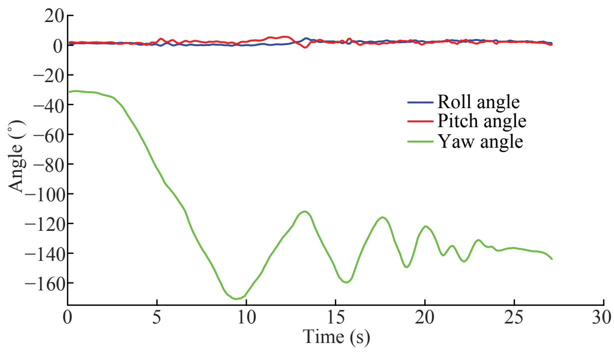

Fig. 13 The roll, pitch and yaw angle of Robot 2. The blue, red and green lines indicate roll, pitch and yaw angles, respectively.

angles vary slightly, and the maximum offsets of roll and pitch angles are up to 7.8° and 5.1°, respectively. These slight offsets justify Assumption *I*. The green line indicates that the initial yaw angle is about 31°. After 25 s, the robot keeps stable and the yaw angle is up to 138°. Therefore, the relative pose of Robot 3 to Robot 2 will be calculated using Eq. (16).

Fig. 14 shows experimental results that were fused into the body coordinate frame of Robot 1. The red "✲", blue "+" and green "○" makers demonstrated the localization results of Robot 2 in perspective of Robot 1, Robot 1 in perspective of Robot 2 and Robot 3 in perspective of Robot 2, respectively. Compared to the predefined position of three robots in Fig. 13a, mean deviations of measured horizontal positions were calculated. As described in Table 5, in X axial direction, the mean deviations of Robot 1, Robot 2 and Robot 3 were 3.0 cm, 4.0 cm and 4.7 cm. And in Y axial direction, mean

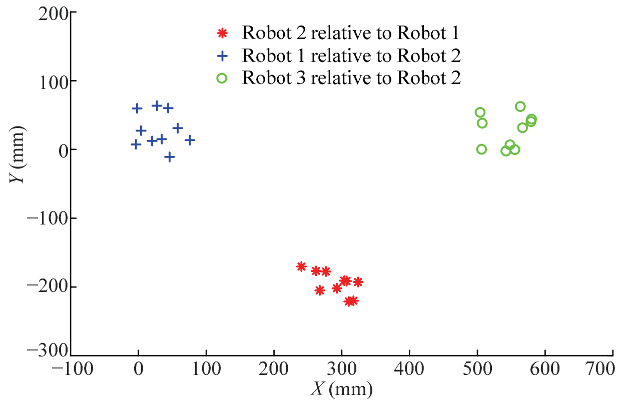

Fig. 14 Localization results with three robots.

Table 5 The localization error of three robots in X and Y axial direction

Robot	1		2		3	
Mean deviation	X	Y	X	Y	X	Y
	3.0 cm	2.8 cm	4.0 cm	2.9 cm	4.7 cm	3.6 cm

Table 6 The localization error table of three robots

Robot	1	2	3
Localization errors	4.1 cm (11.7%)	4.9 cm (14%)	5.9 cm (16.9%)

deviations of Robot 1, Robot 2 and Robot 3 were 2.8 cm, 2.9 cm and 3.6 cm. Then, localization errors of three robots shown in Table 6 were 4.1 cm, 4.9 cm and 5.9 cm, respectively. Therefore, the phenomenon of error accumulation existed. On the one hand, it may be caused by measurement errors in the vision-based localization. On the other hand, Robot 2 conducted rotation movement and produced the water wave, which led to the slight offset of other robots. From the results, we can get that the automatic localization errors are about 0.3 cm in X axis and 2 cm in Y axis larger than the errors in the robot-fixed environment. That's because the rotation motion of Robot 2 causes the drift of other robots. However, these localization errors of three robots only occupied 11.7%, 14% and 16.9% to the size (35 cm) of the robot, respectively, which can be accepted. In conclusion, experiments of the cooperative localization with three robots prove that the proposed localization approach has effectiveness in the submarine environment.

6 Conclusion

This paper presented a robust RGB-D camera and IMU fusion-based cooperative and relative close-range localization for MTASRs. It is able to be used in the narrow submarine environment, such as underwater caves, and environments that the acoustic localization method, GPS measurement method and dead-reckoning method cannot be applied to. In our proposed localization system, an SVM classifier with HOG and CNs features was trained to realize the automatic object detection firstly. The recognition rate with HOG and CNs features is up to 89.26% and 16.18% higher than with HOG features. Secondly, the measurement model with RGB-D camera was established. In this paper, owing to the underwater application, the RGB camera was calibrated and the depth camera was corrected. The effective range of the camera is about 10 cm to 40 cm. In order to realize the localization of MTASRs, the MTASRs model was established with depth map captured by RGB-D camera and self-pose estimates derived from IMU. In experiments of RGB-D camera-based localization, the maximum absolute deviation in X, Y and Z direction occupied 18.24%, 8.68% and 10.61%, respectively, which proved RGB-D camera-based localization method was feasible. Then experiments of RGB-D camera and IMU fusion-based localization approach were conducted with three robots. From experimental results, the phenomenon of error accumulation existed. However, compared to the size (35 cm) of the robot, these localization errors of three robots only occupied 11.7%, 14% and 16.9%, respectively, which can be accepted. Therefore, the cooperative and relative close-range localization approach with three robots has effectiveness in the narrow, submarine environment.

In the future work, considering the precision of the RGB-D camera, multiple high-power infrared LEDs will be added to the camera. Besides, an outdoor experiment will be conducted to verify the effectiveness and feasibility of this method, and this system will be used in the cooperation of MTASRs.

Acknowledgment

This research is partly supported by the National Natural Science Foundation of China (Nos. 61773064, 61503028), Graduate Technological Innovation Project of Beijing Institute of Technology (No. 2018CX10022), and National High Tech. Research and Development Program of China (No. 2015AA043202).

References

[1] Gu S X, Guo S X. Performance evaluation of a novel propulsion system for the spherical underwater robot (SURIII). *Applied Sciences*, 2017, **7**, 1196.

[2] Li Y X, Guo S X, Wang Y. Design and characteristics evaluation of a novel spherical underwater robot. *Robotics & Autonomous Systems*, 2017, **94**, 61–74.

[3] Kim D, Lee D, Myung H, Choi H T. Artificial landmark-based underwater localization for AUVs using weighted template matching. *Intelligent Service Robotics*, 2014, **7**, 175–184.

[4] Carreras M, Ridao P, Batlle J, Ribas D. High-accuracy localization of an underwater robot in a structured environment using computer vision. *Pattern Recognition and Image Analysis*, Springer Berlin Heidelberg, 2003, **2652**, 150–157.

[5] Carreras M, Ridao P, Garcia R, Nicosevici T. Vision-based localization of an underwater robot in a structured environment. In: Perales F J, Campilho A J C, de la Blanca N P, Sanfeliu A, eds., *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, 2003, **1**, 971–976.

[6] Kuch B, Buttazzo G, Azzopardi E, Sayer M, Sieber A. GPS diving computer for underwater tracking and mapping. *Underwater Technology*, 2012, **189**, 189–194.

[7] Baccou P, Jouvencel B, Creuze V, Rabaud C. Cooperative positioning and navigation for multiple AUV operations. *Proceedings of MTS/IEEE Oceans*, Honolulu, HI, USA, 2001, 1816–1821.

[8] Navinda K, Zimmer U R. Underwater acoustic localization for small submersibles. *Journal of Field Robotics*, 2011, **28**, 40–69.

[9] Campos R, Gracias N, Ridao P. Underwater multi-vehicle trajectory alignment and mapping using acoustic and optical constraints. *Sensors*, 2016, **16**, 387.

[10] Paull L, Seto M, Leonard J J. Decentralized cooperative trajectory estimation for autonomous underwater vehicles. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Chicago, IL, USA, 2014, 184–191.

[11] Matsuda T, Maki T, Sato Y, Sakamaki Y. Experimental evaluation of accuracy and efficiency of alternating landmark navigation by multiple AUVs. *IEEE OES Joint Symposium/Workshop on Autonomous Underwater Vehicles* (*AUV*), Tokyo, Japan, 2018, **43**, 288–310.

[12] Hartman R, Hawkinson W, Sweeney K. Tactical underwater navigation system (TUNS). *Proceedings of IEEE/ION Position*, *Location and Navigation Symposium*, Monterey, CA, USA, 2008, 898–911.

[13] Liu P J, Wang B, Deng Z H, Fu M Y. INS/DVL/PS tightly coupled underwater navigation method with limited DVL measurements. *IEEE Sensors Journal*, 2018, **18**, 2994–3002.

[14] Allotta B, Caiti A, Chisci L, Costanzi R, Corato F D, Fantacci C. An unscented Kalman filter based navigation algorithm for autonomous underwater vehicles. *Mechatronics*, 2016, **39**, 185–195.

[15] Gao W, Liu Y L, Xu B, Che Y T. An improved cooperative localization method for multiple autonomous underwater vehicles based on acoustic round-trip ranging. *Proceedings of IEEE/ION Position*, *Location and Navigation Symposium*, Monterey, CA, USA, 2014, 1420–1423.

[16] Josep B, Nuno G, Pere R, Klemen I, David R. Close-range tracking of underwater vehicles using light beacons. *Sensors*, 2016, **16**, 429.

[17] Nuno G, Josep B, Karim M E. Pose estimation for underwater vehicles using light beacons. *IFAC-PapersOnLine*, 2015, **48**, 70–75.

[18] Faessler M, Mueggler E, Schwabe K, Scaramuzza D. A monocular pose estimation system based on infrared LEDs. *Proceedings of the IEEE International Conference on Robotics and Automation*, Hong Kong, China, 2014, 907–913.

[19] Shi L W, Guo S X, Mao S, Yue C F, Li M X, Asaka K. Development of an amphibious turtle-inspired spherical mother robot. *Journal of Bionic Engineering*, 2013, **10**, 446–455.

[20] Guo S X, He Y L, Shi L W, Pan S W, Xiao R, Tang K, Guo P. Modeling and experimental evaluation of an improved amphibious robot with compact structure. *Robotics and Computer-Integrated Manufacturing*, 2018, **51**, 37–52.

[21] Hou X H, Guo S X, Shi L W, Xing H M, Liu Y, Liu H K, Hu Y, Xia D B, Li Z. Hydrodynamic analysis-based modeling and experimental verification of a new water-jet thruster for an amphibious spherical robot. *Sensors*, 2019, **19**, 259.

[22] Xing H M, Guo S X, Shi L W, He Y L, Su S X, Chen Z, Hou X H. Hybrid locomotion evaluation for a novel amphibious spherical robot. *Applied Sciences*, 2018, **8**, 156.

[23] Yu J Z, Ding R, Yang Q H, Tan M, Wang W B, Zhang J W. On a bio-inspired amphibious robot capable of multimodal motion. *IEEE/ASME Transactions on Mechatronics*, 2012, **17**, 847–856.

[24] Zhong B, Zhang S W, Xu M, Zhou Y C, Fang T, Li W H. On a CPG-based hexapod robot: AmphiHex-II with variable stiffness legs. *IEEE/ASME Transactions on Mechatronics*, 2018, **23**, 542–551.

[25] Crespi A, Karakasiliotis K, Guignard A, Ijspeert A J. Salamandra Robotica II: An amphibious robot to study salamander-like swimming and walking gaits. *IEEE Transactions on Robotics*, 2013, **29**, 308–320.

[26] Zhang S W, Liang X, Xu L C, Xu M. Initial development of a novel amphibious robot with transformable fin-leg composite propulsion mechanisms. *Journal of Bionic Engineering*, 2013, **10**, 434–445.

[27] Zhong B, Zhou Y C, Li X X, Xu M, Zhang S W. Locomotion performance of the amphibious robot on various terrains and underwater with flexible flipper legs. *Journal of Bionic Engineering*, 2016, **13**, 525–536.

[28] Gao Z H, Shi Q, Fukuda T, Li C, Huang Q. An overview of biomimetic robots with animal behaviors. *Neurocomputing*, 2019, **332**, 339–350.

[29] Shi Q, Li C, Li K, Huang Q, Ishii H, Takanishi A, Fukuda T. A modified robotic rat to study rat-like pitch and yaw movements. *IEEE/ASME Transactions on Mechatronics*, 2018, **23**, 2448–2458.

[30] Xing H M, Guo S X, Shi L W, Pan S W, He Y L, Tang K, Su S X, Chen Z. Kalman filter-based navigation system for the amphibious spherical robot. *Proceedings of the IEEE International Conference on Mechatronics and Automation*, Takamatsu, Japan, 2017, 638–643.

[31] Song H, Choi W, Kim H. Robust vision-based relative-localization approach using an RGB-depth camera and LiDAR sensor fusion. *IEEE Transactions on Industrial Electronics*, 2016, **63**, 3725–3736.

[32] Tang K, Shi L W, Guo S X, Pan S W, Xing H M, Su S X, Guo P, Chen Z, He Y L. Vision locating method based RGB-D camera for amphibious spherical robots. *Proceedings of the IEEE International Conference on Mechatronics and Automation*, Takamatsu, Japan, 2017, 1509–1514.

[33] Pan S W, Shi L W, Guo S X. A kinect-based real-time compressive tracking prototype system for amphibious spherical robots. *Sensors*, 2015, **15**, 8232–8252.

[34] Guo S X, Pan S W, Shi L W, Guo P, He Y L, Tang K. Visual detection and tracking system for a spherical amphibious robot. *Sensors*, 2017, **17**, 870.

[35] Lu H M, Zhang Y, Li Y J, Zhou Q, Tadoh R, Uemura T. Depth map reconstruction for underwater kinect camera using inpainting and local image mode filtering. *IEEE Access*, 2017, **5**, 7115–7122.

[36] Anwer A, Ali S S A, Mériaudeau F. Underwater online 3D mapping and scene reconstruction using low cost kinect RGB-D sensor. *Proceedings of the International Conference on Intelligent and Advanced Systems*, Kuala Lumpur, Malaysia, 2017.

[37] Karras G C, Kyriakopoulos K J. Localization of an underwater vehicle using an IMU and a laser-based vision system. *Proceedings of IEEE Conference on Control & Automation*, Athens, Greece, 2007, 817–822.