Underwater Formation System Design and Implement for Small Spherical Robots

Liwei Shi, Member, IEEE, Pengxiao Bao, Student Member, IEEE, Shuxiang Guo, Fellow, IEEE, Zhan Chen, Student Member, IEEE, and Zhongyin Zhang, Student Member, IEEE

Abstract—Aiming at the difficulty of communication and positioning of the existing underwater robot system, this article combines the vision system with a closed leader–follower formation structure to realize the underwater formation of three underwater robots. On account of the long-standing tracking requirements during vision-based formation process, an antiloss and redetecting strategy is fused in the proposed detecting and tracking algorithm and the binocular vision system can achieve real-time positioning underwater and calculate the 3-D coordinate so as to complete the follower’s follow-up to the leader. An artificial submarine detection experiment and an antiloss redetection experiment are applied to demonstrate the robustness of the algorithm. And the underwater formation experiment is implemented so that the effectiveness of the proposed vision-based formation strategy is proved, which also provides a new solution for expanded formation system and multilevel cooperation.

Index Terms—Antiloss and redetection, binocular vision, target detection and tracking, underwater formation.

I. INTRODUCTION

The ocean, which accounts for 71% of the earth’s terrestrial area, is a rich and far-undeveloped treasury. With the expansion of population in the 21st century, it is the shortage of land resources and the deterioration of living environment that accelerate the process of utilizing oceans, which contains fishing industry, submarine oil extraction, and other underwater operations. However, submarine operations pose significant risks for mankind. To diminish the potential risks, with the help of fast-developing robot technology, many countries in the world are conducting research on underwater robots [1], [2], [3], [4], [5], [6].

Since small robots have shown advantages in working under conditions such as pipelines, narrow valley landform, and other complex environments, where large-volume robots are not available to accomplish exploring tasks, the miniaturization of robots become an important research interest [7], [8]. However, the operating time, load capacity, and scope of subtle robots are strictly limited by their volume. To compensate for the lack of various functions, the research of multirobots cooperative formation movements is on the rise nowadays. Although the strategy of cooperative formation movements is a complicated and abstract concept to pursue, the researchers have made significant breakthroughs. Strategies that are rooted in the coordinated movement of biomemes rather than following the man-made principle is increasingly embraced in the search for cooperative formation movements of robots. It can combine the advantages of small and large robots, which can extend the reach of just one robot to accomplish complex tasks.

More specifically, a number of countries are conducting research on underwater multirobots systems. Several small underwater robots, Ranger, have been used to implement formation and collaboration among robots and could be used for tasks such as monitoring, searching, and formation [9]. A number of unmanned aerial vehicles (UAVs), have been used by Edward Fiorelli for collaborative multirobots underwater experiments in Monterey Bay [10]. In the Gulf of Buzzaz, Massachusetts, DA Paley developed and tested a Glider Co-ordinated Control System (GCCS) using a water glider [11]. Yu Junzhi of the Shenyang Automation Institute of the Chinese Academy of Sciences performed the coordinated control of robotic fishes in the underwater environment in 2016 [12]. Researchers in the Mcgill University and University of Minnesota used the visual detection and tracking algorithm to realize the multirobots convoying and carrying out the experiment under the water, based on six-foot amphibious robot platform [13]. Liu from Dexas Alington worked on a robust distributed formation control protocol based on graph theory and robust compensation theory for UAVs [14]. He of Jilin used a fusion control strategy with a redistribution mechanism (RM) based on virtual structure and leader–follower formation characteristics to improve speed, stability, and accuracy of the multi-UAVs [15].

However, it is a prerequisite for autonomous multirobots to carry out the task that transformation of information between members can be achieved [16]. At present, the accuracy of the unmanned aerial vehicle (UAV) formation system is capable of reaching millimetre level [17]. And it relies on electromagnetic waves for communication. Because of the peculiarity of communication in the underwater environment, electromagnetic waves...
are shielded, which leads to a situation that the acoustic wave is more efficient during the underwater communication process. But, the speed of acoustic wave propagates five orders of magnitude lower than electromagnetic wave in water media, which will result in low information transmission speed, severe latency, packet loss, and fast attenuation [18], [19]. Moreover, the limited volume of small robots makes the use of large sonar systems impossible. As a solution, only a few small hydrophone devices are available, which are not only expensive, but also inefficient and prolonged [20]. Therefore, using visual system to obtain position and stance information is more reasonable in robots’ visual range.

In this article, a combined detection and tracking method is proposed and resolves the problem of loss and recovery of underwater sight, which have been proved the robustness and effectiveness through the artificial submarine detection experiment and the occluder experiment. The proposed method is fused with a cascade leader–follower formation structure to form a vision-based formation system. By carrying out formation experiments on the platform of three small spherical robots, we demonstrate the feasibility of the vision-based method and put forward a new solution for the underwater multirobots formation cooperation of small robots.

The framework of this article is expanded as below. First, in Section II, the experimental platform is set up. Second, in Section III, a detection and tracking algorithm is proposed in combination with antiloss and redetection strategy. Meanwhile, the artificial marine environment detection experiment is implemented to verify the robustness of the detection algorithm and the occlusion experiment is carried out to test the redetection method. Third, in Section IV, a vision-based formation framework is purposed and a 3-D positioning system is discussed. Then, underwater formation experiments are constructed in Section V. Finally, Section VI concludes this article.

II. UNDERWATER FORMATION EXPERIMENTAL PLATFORM

To adapt the robot to the shallow sea water environment, bionic principles were applied to the design of the spherical robot by referring to amphibious and jellyfish turtles. The spherical robot is covered by a spherical hull and uses four legs to move. Every leg has a two-degree freedom. It can crawl on the ground and use the water jet motor of each leg to move underwater [5], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. As shown in Fig. 1, these are the results that represent the linear motion experiment on the glossy ground of the spherical robot and the red dot points toward the end.

The original amphibious robot has weaknesses in power system and water resistance. In addition, the main frequency of the control board is low, so the processing of information cannot be performed effectively. To complete the formation experiment, we modified the team’s original spherical amphibious robot and set up a new robot. The new robot can control the headangle with a control precision of ±5° and its maximum floor speed is 8.4 cm/s, while the maximum underwater speed is 16.2 cm/s [31]. In this article, we focus only on the underwater movement.

A. Mechanical Structure of the Underwater Spherical Platform

The diameter of the new robot is 350 mm. Fig. 2 illustrates the design of the experimental platform. The underwater formation platform still has the mode of coverage opening and the mode of coverage closure of the lower spherical hull, using four water jet motors for underwater operation like the previous module.

The platform enhanced the waterproof housing, which consists of an O-ring and a single-layer ABS hemispherical shell. The new housing can not only improve the use of space, but also improve its water-resistant performance. And a new type of high-thrust water jet motor is adopted, which can significantly improve the underwater movement performance of the robot. The robot using the new water jet motor has a better power system and good controllability, which can provide a stable platform for the vision-based navigation and leader–follower formation system.

B. Electronic System of the Underwater Spherical Platform

The robot’s electrical structure is the basis for controlling the robot’s movement and performing the functions of the vision system. The electrical structure mainly includes: the main processing module, the coprocessing module, the perception module, the communication module, the drive execution module, and the power module. Fig. 3 shows the electrical structure of the underwater formation platform. And Table I shows the details of the main equipment. The communication module of the new robot, which is not designed for common communications but for emergency situations, is composed of a small hydrophone and a fiber transceiver. The fiber transceiver is primarily used for debugging and underwater testing, such as tuning proportion integration differentiation (PID) parameters and debugging robot formations. During the debugging process, visual feedback can be observed through the VNC remote desktop.
The improvement of the electronic system for spherical robots is mainly reflected in the upgrade of the electronic equipment. The upgraded underwater formation platform uses NVIDIA Jetson TK1 Development Board, which has a quad-core ARM Cortex-A15, as the robot’s main data processor. The coprocessing module is mainly composed of STM32F103 controller and the servo control board. The main function of the STM32 controller is to control the depth gauge and create a waterproof alarm, while the servo control board connects the controller and the steering engine. The sensor module is mainly composed of the vision system, a MEMS sensor and depth sensors. In this article, binocular camera is used as the visual sensing module, which has a direct output resolution ratio of $640 \times 480$ and a $105^\circ$ wide-angle camera with a size of $75 \text{ mm} \times 15 \text{ mm}$. The drive execution module is mainly composed of water jet motors, steering engines, and electronic speed controllers (ESCs). The power module that accounts for power supplement is mainly composed of lithium batteries, voltage boost modules, and voltage drop modules.

### TABLE I

<table>
<thead>
<tr>
<th>Device</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development board</td>
<td>NVIDIA Jetson TK1</td>
</tr>
<tr>
<td>Single-chip microcomputer</td>
<td>STM32F103</td>
</tr>
<tr>
<td>Servo control board</td>
<td>Arduino-USB-1</td>
</tr>
<tr>
<td>MEMS</td>
<td>JY901</td>
</tr>
<tr>
<td>Battery</td>
<td>VB6000</td>
</tr>
<tr>
<td>Binocular camera</td>
<td>XC3440</td>
</tr>
<tr>
<td>Buck module</td>
<td>LM2596-ADJ</td>
</tr>
<tr>
<td>Small underwater acoustic device</td>
<td>Micron Sonar</td>
</tr>
<tr>
<td>Fiber transceiver</td>
<td>Wpd-smolIG</td>
</tr>
<tr>
<td>Steering engine</td>
<td>HS-3086WP</td>
</tr>
<tr>
<td>Electronic speed control</td>
<td>HOBBYWING(2-3S)(40A)</td>
</tr>
<tr>
<td>Depth sensor</td>
<td>MSS803-01BA</td>
</tr>
</tbody>
</table>

The underwater binocular vision system

Since the communication strategy is to use visual system to estimate the position and stance information, it is necessary for the followers to track the leader. In this way, the working process can be decomposed into two steps, leader follow-up and redetect after loss.

**A. Underwater Target Detecting Algorithm**

Currently, there are many detection algorithms, which can be divided into classical algorithms and deep learning algorithms. Deep learning algorithms have developed quickly in recent years, and numerous algorithms have been applied to underwater detection such as R-CNN [32], YOLO [33], and SSD [34]. However, when it is applied to fish posture, the detection effect is not good, because multiangles and distortion will limit the accuracy of its detection and it is not suitable for embedded low-power robot platform, either. For classical algorithms, Viola-Jones detectors [35], [36], HOG pedestrian detectors [37], Deformable PartModel algorithm (DPM) [38], [39], [40], [41], [42] are widely used.

The major difference between underwater image and land image is that the light of underwater image is weak, and the object is easy to blur and deform and may have low contrast, but it has less interference information and less obstacles. As a result, an algorithm, which can withstand deformation is more reasonable.

The DPM model is highly intuitive, splitting the overall detection of the target into the detection of each component of the...
model. The final detection results are obtained by aggregating the detection results of each component. And DPM is a pattern based on star structure, which is composed of rootfilter and partfilters, where rootfilter is unique while partfilters are various. The spatial resolution of the image elements depicted in the rootfilter is half of that of the partfilter.

The initial score principle of partfilters and rootfilter are the dot product of the parameters and the feature vector of the window in the feature pyramid. For the partfilter, the degree of deviation from the ideal position is expressed by the deformation cost function. And the difference between the initial score of the partfilter at a position and the deformation cost of the position is the score of the partfilter at that position and the maximum value of the score at all spatial positions is the finite score of the partfilter. The scores summed by each partfilter and rootfilter at a given position are the finite scores of the star model at a given position and scale.

Although a single star structure may adapt to numerous changes in the shape of the target, in practice, a single deformable model cannot fully represent the target because of the different viewing angles and multiple forms of the target, particularly in the underwater environment. Because our spherical robot has good symmetry properties, only two star structures are used, where two components are formulated in the front view and the side view of the robot.

Because of the real-time requirement of robot, we combine ARM’s NEON acceleration engine to accelerate the DPM detection algorithm and apply it to our platform and the robot was trained offline by latent support vector machine method in advance.

B. Underwater Target Tracking Algorithm

The second step is to achieve target tracking. For the vision system of underwater robots, there is a relationship between the front and back frames. Therefore, it is not necessary to carry out global target detection and matching on the image of each frame, but to know the initial location of the target. Currently, there are two broad strategies to initialize target tracking. The first involves the position information of the tracking target in advance, and the other is to box select the position of the target subjectively. Yet neither of them is suitable for multirobots formation system. Moreover, the target tracking may be lost. Although the antiloss mechanism is added to some tracker algorithms, most of which can only recover follow-up where loss time is short and is limited for long-term retracking.

At present, the development direction of the tracking algorithm is mainly based on correlation filtering. Considering that in the leader–follower formation, the follower robot needs to track the leader robot in real time and the resource occupation should be finite, multiscale kernel correlation filtering algorithm (KCF) based on online learning [43], [44], [45], [46] was chosen as our algorithm framework. A scale filter to estimate the target scale in the current frame was added to the algorithm and make it similar to discriminative scale space tracker (DSST) [47]. The kernel correlation matrix of the training sample in the KCF algorithm is

$$k_{MM} = k_{s,t+1}^{x,y+1}$$

and the parameter vector is $$\alpha = \frac{S}{x+y}$$, which is updated with the following formula:

$$\begin{align*}
\alpha_t &= (1-\eta)\alpha_t + x\eta \\
\alpha_{t+1} &= (1-\eta)\alpha_{t+1} + \alpha_t
\end{align*}$$

In the target position obtained by the filter, with the target position as the center, 33 samples of different scales are collected, for example, the size of the current frame target is $$M \times N$$, and the features of the size $$\alpha^m M \times \alpha^m N$$ image are extracted around it, where $$n \in \left\{ \left[ \frac{(S-1)}{2} \right], \ldots, \left[ \frac{(S+1)}{2} \right] \right\}$$, $$S$$ is the size of the scale filter and $$\alpha$$ is the scale factor between the feature layers. The characteristic samples collected from various scales are used to select the appropriate scale of the current frame target using the scale filter. The scaling filter update is similar to the position filter update strategy. Then, a new image is input, in a continuous loop, and the predicted result is continuously obtained in each new image, thus completing the tracking. The KCF target tracking algorithm constructs training samples through cyclic shift, and uses cyclic matrix to convert the classifier into frequency domain correlation filtering operation, which avoids the inversion of matrix and reduces the complexity of operation. However, KCF algorithm uses fixed factors to update the model for the weight vector of each training setting, which cannot solve issues like target loss. This results in a combined detection and tracking method.

C. Combined Detecting and Tracking Algorithm

In this article, we combine the DPM detection algorithm and the KCF tracking algorithm to achieve detecting, tracking, and antiloss mechanism. When the binocular camera separately tracks the leader, the threshold value of the maximum response score and the distance change rate detected by the follower are used as criteria for whether the leader robot is lost or not. Specifically, test the distribution of the largest response score and the distance change rate detected by the follower tracks the leader, the threshold value of the maximum response in the confidence graph when the target is not lost, then test the transformation of the maximum response value in the confidence graph when the target is lost. When the response value is less than a certain value and the target is considered to be lost.

Because of the characteristic of the binocular camera, two cameras simultaneously follow the target of the leading robot. And if both cameras are lost, both cameras carry out a new detection. More precisely, depending on the characteristics of the periods of binocular cameras, the rate of distance change is fixed as another criterion for the loss of the robot target. The target distance measured by the past-time frame is expressed as $$d_t$$, the target distance measured by the current frame is expressed as $$d_{t+1}$$, and the distance change rate is defined as $$\xi = \frac{d_{t+1} - d_t}{d_t}$$. When $$|\xi|$$ is above the set threshold, it is considered that the leader robot target is lost and redetection should be started. Moreover, the advantage of the DPM redetection mechanism is that there is no need to worry about whether the training sample is contaminated during the tracking process or if the training sample is changed after losing.
D. Visual Experiment

To verify that our algorithm can be applied in an underwater environment, we have set up an underwater landscape scene and implemented a detection experiment.

In order to verify the effectiveness of the combined detection and tracking algorithm, we set up a redetecting experiment with a submarine model and a occluder.

1) Robust Detecting Experiment Based on an Artificial Submarine Setting: In order to create an artificial submarine settings, we used a transparent fish tank whose size is $50 \times 30 \times 35$ cm and a coral moss cave whose size is $31 \times 16 \times 17$ cm, as well as three colorful sea anemones, three fish whose length is 10 cm, 10 different-sized shells, several clusters of aquatic plants, a floating jellyfish, and a yellow submarine marker. The water in the experiment is semiturbid, which can better simulate the real situation underwater.

The setting diagram of the experiment is shown in Fig. 4(a), and the detecting result is presented in Fig. 4(b). Experimental results show that the DPM algorithm can achieve target detection in the presence of obstructions and water turbidity and demonstrate the robustness of our detection algorithm.

2) Antiloss Tracking and Redetection Experiment: Fig. 5 is the schematic diagram and the results from the antiloss and redetection experiment. Fig. 5(a) shows the three cases. In case 1, the target is moving rapidly and then binocular camera is obstructed for a long time, which is similar with the fluctuating distance changing rate from frame 70th to frame 92th. And the 70th frame to the 84th frame represents the rapid movement process and the 84th frame to the 92th frame represents the obstruction and redetection process. Case 2 represents a short-term shelter from frame 159th to frame 165th. Case 3 shows a long-term shelter from frame 196th to frame 208th.

And Fig. 5(b) shows the changes in the maximum response values of the left camera and the right camera, the distance change rate of binocular ranging, and the redetection performance during the entire experiment. In Fig. 5(b), when the target is lost or blocked, the maximum response values of the left and right cameras become smaller and the rate of distance change suddenly becomes large sharply, and the redetection process is started at the same time.

The experiment verifies the effectiveness of redetection and tracking algorithm combined with target antiloss strategy within a short range, which means that followers of the underwater formation system can track and detect the leader within a short distance.

IV. UNDERWATER VISION-BASED FORMATION STRATEGY

The principle concept in formation control is the use of small, simple, and relatively cheap agents instead of a costly specialized device, which executes compound and complicated commands cooperatively [48]. Generally, the formation of underwater vehicles may be categorized as path planning, path following, and path tracking [49], [50], [51].

Because the resistance of water far exceeds that of air, the friction between deepwater diving and shipping is different, which means that the speed of vessels, underwater vehicles, and submarines is much slower than that of vehicles traveling on land. The speed of the world’s fastest submarine, the papa class submarine of the former Soviet Union, is 47 knots (24.18 m/s), while the speed of the conventional cargo ship is usually 15–20 knots (7.72–10.29 m/s). The power and volume of small underwater vehicles are much lower than that of these vehicles, so the speed of navigation is far lower than that of these vehicles. In general, the speed of small underwater robots is no more than 4 knots (2.05 m/s). Due to the frame rate of the underwater binocular camera is 15FPS and higher, the detection and tracking period will not be influenced by the motion of the robot. Therefore, a vision-based formation strategy can be designed.

A. Others Work

Karras presented a vision-based leader–follower cooperative scheme, which consists of two underwater vehicles in the absence of explicit communications and direct information interchange. It calculates its distance and orientation with respect to the surface using its own laser pointers, which are equipped on
each robot platform [52]. Zhang discussed about the formation control of soft robotic fish swarm based on the visual images acquired by the overhead camera, which is able to track the position and direction of the fish and the target [53]. Shahab’s work has designed and demonstrated an effective visual position servo approach for autonomous underwater vehicles using nonlinear model predictive control. And in his proposed scheme, a mechanism is incorporated within the vision-based controller that determines when the visual tracking algorithm should be activated and new control inputs should be calculated [54]. Hamamatsu purposed a short distances tracking method by using a camera and an LED array mounted on another moving vehicle to estimate self-position and velocity with regard to the target vehicle [55]. Liu proposed a vision framework for the automatic retrieval of an AUV by another AUV in shallow water, where the framework is mainly consist of a Laplacian-of-Gaussian-based coarse-to-fine blockwise method for the detection of underwater landmarks to overcome ambient light and nonuniform spreading, which are the two main problems in shallow water [56]. Ozay has integrated a vision-based framework with an ultra-short baseline acoustic sensor to achieve short-range underwater docking [57].

However, the use of laser and rectangular markers for positioning does not suit all small robots. Because many robots have no such equipment. And tracking algorithm that relies on the overall camera is not suitable for outdoor and has many limitations in the application. Many underwater vision frameworks for submarine formation systems place emphasis on measuring attitude and distance rather than on the loss prevention function, which contain defects in long-term tracking and may present latent dangers in underwater formation tasks. Therefore, our vision-based formation approach is a priority.

B. Cascade Leader–Follower Formation Structure

Based on the proposed combined detection and tracking algorithm, we use the leader–follower structure to control the formation.

The classical leader–follower formation control is mainly used in a 2-D plane and two methods are usually used to realize formation control. One is based on the angle-distance method, while the other is based on the distance–distance method. The first method involves controlling the movement of the following robot using the deflection angle and the offset between the leader and the follower, while the second method uses the distance deviation between the follower and two leaders.

For the following robot, information about the leader’s position and stance can be obtained directly. Therefore, a cascade leader–follower structure is designed. Two robots are regarded as a small formation system and there may be multiple small formation systems in the entire formation system. In the interior of a small formation system, each follower has a leader, whereas the entire system has a single leader.

The follower realized the 3-D positioning of the leader under its coordinate system through the vision system and drift angles and coordinates are calculated, serving as input into the motion control to maintain a fixed leading position from the follower’s perspective. Fig. 6 is a schematic diagram of the cascade leader–follower formation structure.

C. Multirobot Relative Underwater Positioning System

In order to achieve the position and stance information of the leader, a relative underwater positioning system is discussed. Binocular camera is used to obtain the 3-D information.

The principle of binocular vision is parallax and triangulation, which obtains the target information through two cameras and performs a matching solution to calculate the 3-D information. And the matching process is to match points of image features to reduce the amount of computation by reducing the number of corresponding pixels. And the 3-D position of the robot can be reached simply by matching the pixels in the center.

As shown in Fig. 7, the left camera and the right camera form a binocular camera. The coordinate of point $P$ under the earth coordinate system $O_w = (X_w, Y_w, Z_w)$ is $(X_{w1}, Y_{w1}, Z_{w1})$ and the coordinate under the left camera coordinate system $O_{c1} = (X_{c1}, Y_{c1}, Z_{c1})$ is $(X_{c1}, Y_{c1}, Z_{c1})$, while under the right camera coordinate system $O_{c2} = (X_{c2}, Y_{c2}, Z_{c2})$ is $(X_{c2}, Y_{c2}, Z_{c2})$. The corresponding mapping points under
(a) Global perspective when $t = 0s$. (b) Global perspective when $t = 5s$. (c) Global perspective when $t = 10s$. (d) Global perspective when $t = 15s$. (e) Follower perspective when $t = 0s$. (f) Follower perspective when $t = 5s$. (g) Follower perspective when $t = 10s$. (h) Follower perspective when $t = 15s$.

Fig. 8. Images of global perspective and follower perspective of two robots steer-following formation experiment. (a) Global perspective when $t = 0s$. (b) Global perspective when $t = 5s$. (c) Global perspective when $t = 10s$. (d) Global perspective when $t = 15s$. (e) Follower perspective when $t = 0s$. (f) Follower perspective when $t = 5s$. (g) Follower perspective when $t = 10s$. (h) Follower perspective when $t = 15s$.

\[
\begin{bmatrix}
X_{c2} \\
Y_{c2} \\
Z_{c2}
\end{bmatrix} = R_2 \begin{bmatrix}
X_w \\
Y_w \\
Z_w
\end{bmatrix} + t_2
\] 

where $R_1$, $t_1$ represents the rotation matrix and translation vector of the left camera, respectively. And $R_2$, $t_2$ represents the rotation matrix and translation vector of the right camera, respectively. Compare (2) and (3) to obtain the relationship between the left camera and the right camera and use similar expression to obtain (4).

Based on the imaging feature, the central position of the light in the camera’s optical system is the far point $O_C$. The far point of the left camera and the right camera is $O_{C1}$ and $O_{C2}$, respectively, and the component on the optical axis is $Z_{c1}$ and $Z_{c2}$. The relationship between the corresponding mapping point $p_1(u_1, v_1), p_2(u_2, v_2)$ and the coordinate point $P(X_w, Y_w, Z_w)$ is as (5) and (6), where $R^{12}, t^{12}$ is the rotation matrix and translation vector between the left camera and the right camera, respectively. And $R^{12} = R_1 R_2^{-1}, t^{12} = t_1 - R_1 R_2^{-1} t_2$.

\[
\begin{bmatrix}
X_{c1} \\
Y_{c1} \\
Z_{c1}
\end{bmatrix} = R_1 R_2^{-1} \begin{bmatrix}
X_{c2} \\
Y_{c2} \\
Z_{c2}
\end{bmatrix} + t_1 - t_2 R_1 R_2^{-1}
\]

\[
\begin{bmatrix}
X_{c1} \\
Y_{c1} \\
Z_{c1}
\end{bmatrix} = R^{12} \begin{bmatrix}
X_{c2} \\
Y_{c2} \\
Z_{c2}
\end{bmatrix} + t^{12}
\] 

Fig. 9. Heading angle measurements of visual system and MEMS sensor over time. (a) Heading angle measured by the follower. (b) Heading angle measured by the leaders MEMS sensor.

Fig. 10. Three dimensional coordinates and distance between the leader robot and the follower robot measured by the follower visual system. (a) Three-dimensional coordinates. (b) Distance on each axis.

\[
\begin{bmatrix}
X_w \\
Y_w \\
Z_w
\end{bmatrix} = M_1 \begin{bmatrix}
X_{c1} \\
Y_{c1} \\
Z_{c1}
\end{bmatrix} = \begin{bmatrix}
m_{11}^1 & m_{12}^1 & m_{13}^1 & m_{14}^1 \\
m_{21}^1 & m_{22}^1 & m_{23}^1 & m_{24}^1 \\
m_{31}^1 & m_{32}^1 & m_{33}^1 & m_{34}^1
\end{bmatrix} \begin{bmatrix}
X_{c1} \\
Y_{c1} \\
Z_{c1}
\end{bmatrix}
\]

the image coordinate system of the left camera $O_1 - u_1 v_1$ and the right camera $O_2 - u_2 v_2$ is $p_1(u_1, v_1)$ and $p_2(u_2, v_2)$, respectively.

According to the imaging principle of the camera, the process of converting the world coordinate system to the camera coordinate system of the point $P$ in the left camera and the right camera can be expressed as as follows:

\[
\begin{bmatrix}
X_{c1} \\
Y_{c1} \\
Z_{c1}
\end{bmatrix} = R_1 \begin{bmatrix}
X_w \\
Y_w \\
Z_w
\end{bmatrix} + t_1
\]

\[
\begin{bmatrix}
X_{c2} \\
Y_{c2} \\
Z_{c2}
\end{bmatrix} = R_2 \begin{bmatrix}
X_w \\
Y_w \\
Z_w
\end{bmatrix} + t_2
\]
First-level follower perspective when $t = 0s.$ (g) First-level follower perspective when $t = 3s.$ (h) First-level follower perspective when $t = 6s.$ (i) Second-level follower perspective when $t = 0s.$ (j) Second-level follower perspective when $t = 3s.$ (k) Second-level follower perspective when $t = 6s.$ (l) Second-level follower perspective when $t = 9s.$

Fig. 11. Images of global perspective and follower perspective of three robots straight line following formation experiment. (a) Global perspective when $t = 0s.$ (b) Global perspective when $t = 3s.$ (c) Global perspective when $t = 6s.$ (d) Global perspective when $t = 9s.$ (e) First-level follower perspective when $t = 0s.$ (f) First-level follower perspective when $t = 3s.$ (g) First-level follower perspective when $t = 6s.$ (h) First-level follower perspective when $t = 9s.$ (i) Second-level follower perspective when $t = 0s.$ (j) Second-level follower perspective when $t = 3s.$ (k) Second-level follower perspective when $t = 6s.$ (l) Second-level follower perspective when $t = 9s.$

Fig. 12. Heading angle measurements of visual system and MEMS sensors by two follower robots over time. (a) First-level follower. (b) Second-level follower.

$Z_{c2} = \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} m_{11}^2 \\ m_{21}^2 \\ m_{31}^2 \\ m_{12}^3 \\ m_{22}^3 \\ m_{32}^3 \\ m_{14}^2 \\ m_{24}^2 \\ m_{34}^2 \end{bmatrix}$

$$Z_{c2} = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} u_1 m_{11} - m_{11} \\ u_1 m_{12} - m_{12} \\ u_1 m_{13} - m_{13} \\ v_1 m_{21} - m_{21} \\ v_1 m_{22} - m_{22} \\ v_1 m_{23} - m_{23} \\ u_2 m_{31} - m_{31} \\ u_2 m_{32} - m_{32} \\ u_2 m_{33} - m_{33} \\ v_2 m_{34} - m_{34} \\ v_2 m_{33} - m_{33} \\ v_2 m_{32} - m_{32} \end{bmatrix}$$

where $M_1$ and $M_2$ are the projection matrix of the left camera and the right under the world coordinate system $O_w - X_w Y_w Z_w.$ And $\alpha_x, \alpha_y, u_1, v_1$ are the internal parameters of the left camera and $\alpha_x, \alpha_y, u_0, v_0$ are the internal parameters of the right camera. The internal parameters can be obtained from the binocular camera parameter manual.

Assume that the coordinate system $O_{c1} - X_{c1} Y_{c1} Z_{c1}$ of the left camera coincides with the earth coordinate system $O_w - X_w Y_w Z_w.$ At this time, the rotation matrix $R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ of the left camera has a translation vector of $t_2 = (0 \ 0 \ 0)^T.$ The parameters of the camera are ideal parameters, which are different from the actual camera parameters. Consequently, we have to calibrate the camera to obtain the actual parameters. Thus, the rotation matrix of the right camera $R_2 = R_1^{12} = R_{1\text{Calib}} R_{2\text{Calib}}^{-1},$ $t_2 = t_{1\text{Calib}} = R_{1\text{Calib}} R_{2\text{Calib}}^{-1} t_{2\text{Calib}}.$

Compare (5) and (6) and eliminate $Z_{c1}, Z_{c2},$ we can obtain (7). Suppose that $K$ is the parameter matrix and $U$ is nonhomogeneous term, (7) can be simplified into the following:

$$K \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = U.$$  (8)

The least squares solution of (8) is

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = (K^T K)^{-1} K^T U.$$  (9)

Thus, the target’s 3-D coordinate $P(X_w, Y_w, Z_w)$ can be calculated. Therefore, a positioning system based on the visual system is achieved.
V. UNDERWATER FORMATION EXPERIMENTS

In order to study the effect of the proposed method, we implemented two experiments in a swimming pool, whose size is $3 \text{m} \times 2 \text{m} \times 1 \text{m}$. One is the steer-following formation experiment and the other is the straight line following formation experiment.

The first experiment includes two robots and the leader robot moves approximately linearly in a certain direction while the follower robot steers its heading angle to align while following. The second experiment involves three robots, where the leader robot remains stationary and the first-level follower keeps up with the leader and the second-level follower keeps up with the first-level follower.

A. Steer-Following Formation Experiment

The improved previous-generation spherical robot serves as the leader and the improved underwater robot serves as the follower with different heading angle. And the leader robot moves forward along a straight ruler positioned at the bottom of the pool, while follower robot steers and moves under the lead of the leader robot.

Fig. 8(a)–(d) shows the global perspective in the experiment. Fig. 8(e)–(h) shows the state of the leader robot. Fig. 9 compares the heading angle of the sensors with the visual system of both robots. Moreover, since the MEMS sensors of the two robots show different results from the same heading angles, the results are assessed by observing the trend of the MEMS sensor. Fig. 9(a) shows the heading angle of the leader robot measured by the follower’s binocular vision system and MEMS sensor. The MEMS sensor shows that the follower’s heading angle gradually increases to ensure that the leader is in sight of the follower and approaches the leader robot’s. Fig. 9(b) is the leader robot’s heading angle measured by the MEMS sensor. When $t = 8 \text{s}$, the MEMS sensor itself is affected by the magnetic field and sets off a steep cliff. Fig. 10 is the 3-D coordinates and the distance between the leader and the follower measured by the follower’s visual system. Fig. 10(b) shows the distance between the follower robot and the leader robot in the Z-axis, which becomes stable gradually in the following motion.

During the whole movement, it can be seen that the leader robot can lead the follower’s linear movement and steering movement.

B. Straight Line Following Formation Experiment

In the experiment, the previous-generation spherical robots is set as the leader robot and is static. The improved underwater robot works as the first-level follower and also the second-level leader. The improved previous-generation spherical robots is the second-level follower.

Fig. 11(a)–(d) shows the state of the three robots captured by the global camera. The first-level follower approaches the leader and the second-level follower follows the second-level leader, which is also the first-level follower. Fig. 11(e)–(h) shows the state of the leader robot during the tracking process viewed by the first-level follower and Fig. 11(i)–(l) shows the state of the second-level leader during the tracking process viewed by the second-level follower robot. Fig. 12(a) and (b) shows the heading angle of the leader robot measured by the first-level follower and the heading angle of the second-level leader robot measured by the second-level follower robot, respectively. And the measurement is compared with their MEMS sensors.

In the experiment, the first-level follower robot, which is also the second-level leader robot moves to the leader robot and guide the second-level follower robot. It can be concluded that this method is feasible to track both static targets and moving targets.

VI. CONCLUSION

In this article, a combined detecting and tracking algorithm is designed for an underwater vision-based formation method aiming at the difficulty of communication and positioning of the existing underwater formation strategy.

At present, most underwater tracking algorithms are short-time tracking algorithms that may lead to the failure of vision-based formation due to target loss. Therefore, an antiloss and redetection strategy is fused with the proposed algorithm. The detection algorithm is tested in the environment of an artificial submarine environment, which demonstrates the robustness of the detection period. And the effect of the redetection strategy is tested by an occluded experiment that shows that the targeted algorithm can achieve long-time tracking.

In addition, a cascade leader–follower formation structure is adopted and the 3-D coordinate and distance is calculated through the follower’s visual system and can achieve underwater positioning without microwave communication. Two formation experiments are implemented and prove that the formation strategy can be expanded from two robots to three robots and can track a static leader and also a dynamic leader. The formation experiments also show it is feasible and effective to use vision-based method to complete formation tasks and the cascade leader follower formation structure can expand to more robots because every single can be a leader of the downstream robot.

REFERENCES


**Pengxiao Bao** (Student Member, IEEE) received bachelor’s degree in automation from the Beijing Institute of Technology, Beijing, China, in 2020. Her research interests include underwater bionic-inspired robots, biomems, and dynamics.

**Shuxiang Guo** (Fellow, IEEE) received the Ph.D. degree in mechanical information system from Nagoya University, Nagoya, Japan, in 1995. He is currently a Professor in biomedical engineering with the Beijing Institute of Technology, Beijing, China, and Kagawa University, Takamatsu, Japan. His research interests include surgical robotics, rehabilitation robotics, endoscopic capsule robotics, and underwater robotics. Prof. Guo is the leader of Key Laboratory of Convergence Biomedical Engineering System and Healthcare Technology and the Editor in Chief of *IJMA (International Journal of Mechatronics and Automation)*.

**Zhan Chen** (Student Member, IEEE) received the M.S. degree in biomedical engineering from the Beijing Institute of Technology, Beijing, China, in 2018. His research interests include bio-inspired robots and underwater formation structure.

**Zhongyin Zhang** (Student Member, IEEE) received the B.S. degree in special type energy from the Beijing Institute of Technology, Beijing, China, in 2019. His research interests include bio-inspired robots, bionic fish, and mechanical structure design.