



Experimental verification of turbidity tolerance of stereo-vision-based 3D pose estimation system

Myo Myint¹ · Khin Nwe Lwin¹ · Naoki Mukada¹ · Daiki Yamada¹ · Takayuki Matsuno¹ · Yuuichirou Toda¹ · Saitou Kazuhiro² · Mamoru Minami¹

Received: 22 May 2018 / Accepted: 13 August 2018
© JASNAOE 2018

Abstract

This paper presents the verification of the turbidity tolerance of a stereo-vision-based 3D pose estimation system for underwater docking applications. To the best of the authors' knowledge, no studies have yet been conducted on 3D pose (position and orientation) estimation against turbidity for underwater vehicles. Therefore, the effect of turbidity on the 3D pose estimation performance of underwater vehicles and a method of operating under turbid conditions were studied in this work. A 3D pose estimation method using the real-time multi-step genetic algorithm (RM-GA) proposed by the authors in the previous works shows robust pose estimation performance against changing environmental conditions. This paper discusses how and why the RM-GA is well suited to effective 3D pose estimation, even when turbid conditions disturb visual servoing. The experimental results confirm the performance of the proposed 3D pose estimation system under different levels of turbidity. To demonstrate the practical usefulness of the RM-GA, docking experiments were conducted in a turbid pool and a real sea environment to verify the performance and tolerance of the proposed system under turbid conditions. The experimental results verify the robustness of the system against turbidity, presenting a possible solution to a major problem in the field of robotics.

Keywords Visual servoing · 3D pose estimation · Sea docking · Stereo-vision · Robustness against turbidity · Real-time multi-step genetic algorithm

1 Introduction

Autonomous underwater vehicles (AUVs) play an important role in many undersea operations, such as the inspection of underwater structures (e.g., dams [1] and bridges), ship-hull inspections [2], and deep-water archaeology [3]. Despite recent advancements to power storage technologies, the operation time of underwater vehicles is a limiting factor. A recharging unit with an underwater docking function can enable the extended operation of AUVs in the sea independent of a surface vehicle to which they must return for recharging. Docking is very useful not only for battery

recharging but also for other applications, such as sleeping under the mother ship and downloading new missions [5]. Therefore, many studies have been conducted on underwater docking [4–16]. However, a number of challenging issues hinder these applications, which require high accuracy and robustness against disturbances that occur in actual undersea environments. To overcome these problems in underwater vehicles, we have developed a vision-based docking system using stereo-vision [7]. Depending on the application, different homing sensors have been used for the docking of AUVs. However, each sensor has its own limitations. In [8], a novel approach to docking using an electromagnetic guidance system was proposed. However, the accuracy of this approach is limited by the presence of magnetic abnormalities near a station. For acoustic-based sensors used in AUVs [5], the effects of undesirable acoustic reflections and attenuation may reduce the accuracy of AUV navigation. Similarly, there remain some limitations in systems using vision sensors applied in AUVs. The limitations to vision-based underwater vehicles with lighting units are particularly challenging. The motion of such vehicles produces fluctuations

✉ Myo Myint
ecmyomyint@gmail.com

¹ Okayama University Graduate School of Natural Science and Technology Division of Mechanical and Systems Engineering 3-1-1 Tsushimanaka, Kita-ku, Okayama 700-8530, Japan

² Ushimado Marine Institute Okayama, 130-17 Ushimadocho Kashino, Setouchi, Japan

in the lighting direction, which cause the lighting conditions to change. This type of variable lighting environment presents challenges when the available light is dominated by the lighting system of the AUV, as in deep-sea missions or during nighttime operation. In addition, when a vision-based underwater vehicle approaches the sea bottom, a combination of water turbidity and fluctuations in the lighting direction produces artifacts in camera images and acts as a disturbance for visual servoing, as shown in Fig. 1.

To the best of the authors' knowledge, there have been no studies on the 3D pose estimation of underwater vehicles under turbid conditions. The detection of points of interest in turbid underwater images has been investigated using a collection of images acquired by a trinocular camera system under gradually increasing turbidity [17]. In [18], the robustness of local feature detection in underwater images was analyzed using a new data set called TURBID, which consists of real seabed images with different amounts of degradation. In addition, methods of underwater image quality assessment, visibility enhancement, and disparity computation under turbid conditions have been proposed in [19]. None of the above studies consider image recognition in real-time dynamic images, which is an indispensable technology for visual servoing in underwater vehicle docking. The papers concerning image processing of fog environment have been found. [20–22] are concerned with static images analyses in the field of car guidance and safety technology, then they are difficult to be utilized as real-time control feedback information. [23] has discussed analysis in dynamic images about how to distinguish fog, but this idea has not

yet been utilized for automatic driving as a visual feedback sensor.

Our research group has conducted a number of studies on our vision-based docking system, which uses the real-time multi-step genetic algorithm (RM-GA) method [24–28]. In [24–26], several experiments were conducted using two cameras and a known 3D marker to confirm the robustness of the docking system. The development system has been named “Three Dimensional Move on Sensing” (3D-MoS) [7]. Docking trials using a remotely operated vehicle (ROV) as a test bed were also successfully conducted in a real sea environment near the city of Wakayama in Japan [27]. In [28], docking experiments with the AUV Tuna-Sand 2 were conducted in a pool. The robustness of the 3D pose estimation system against air bubbles [24] and target occlusion [29] has been verified experimentally. The effectiveness of the proposed system in tracking the pose of a moving 3D object has been reported in [24]. In [25], visual servoing while a physical disturbance was applied in a specific direction by pushing the AUV with a stick and the docking performance of the AUV were tested. Through these experiments, the authors have demonstrated the robustness of the proposed RM-GA method against a number of disturbances.

However, we have yet to confirm the robustness of the proposed system against the effects of the water turbidity. The most important points of investigation regarding the RM-GA method are the reasons it shows robust against turbidity and how much turbidity it can tolerate. These aspects have not been examined despite being closely connected to the practicality of the method when applied in real undersea environments.

Based on the above motivation, some experiments were conducted to confirm the practicality of the proposed system against turbidity in the present paper. In this study, three experiments were conducted in three different environments, as shown in Fig. 2, to assess the system performance in turbid environments and demonstrate the potential of the proposed system for real undersea applications. The recognition performance against different turbidity levels was verified first in a small pool, as shown in Fig. 2a. After this assessment of the recognition performance, the turbidity tolerance was verified by conducting docking experiments under different turbidity levels in a larger pool, as shown

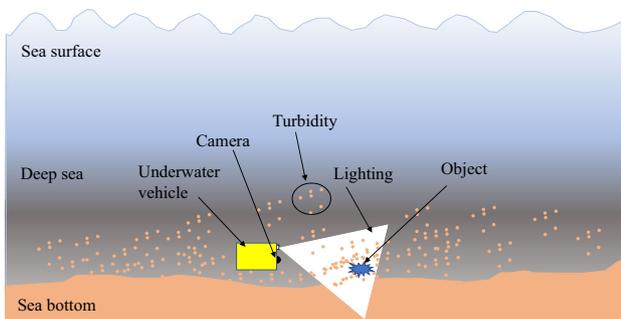
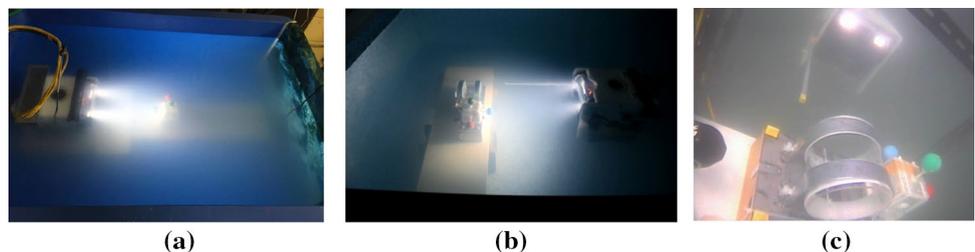


Fig. 1 Visual servoing in a deep-sea environment with turbidity

Fig. 2 ROV and 3D marker in turbid water. **a** Recognition experiment in a small pool. **b** Docking experiment in a large pool. **c** Continuous iterative docking experiment in the sea



in Fig. 2b. Finally, a continuous iterative docking experiment was conducted in a shallow sea region near the town of Ushimado, Japan, as shown in Fig. 2c. Because this study focused on turbidity tolerance, a turbid coastal environment was selected over clear oceanic water to conduct the docking experiments.

This paper is organized as follows: Sect. 2 introduces the problem statement. Section 3 explains the stereo-vision-based real-time 3D pose estimation method. Section 4 outlines the experiments for real-time 3D recognition under different turbidity levels. Section 5 describes the docking experiments under different turbidity levels in a pool. Section 6 presents the continuous iterative docking experiment in the sea. The final section contains the conclusions of this study and plans for future work.

2 Problem statement and contributions of this study

2.1 Social demand for deep-sea operations

Recently, social demand for deep-sea resources, such as food and energy, has increased rapidly with the aid of developing technologies. At present, almost all metal and industrial mineral extraction is conducted onshore. Deep-sea extraction would expand the available resource pool and provide a level of independence from onshore resource extraction. Among the available deep-sea resources, expensive rare metals and methane hydrate in the seabed are of particular economical importance. Japan has been considering the deep-sea mining of methane hydrate, which is expected to be a future energy resource. Furthermore, the information that can be gathered from deep-sea exploration is useful in the prediction of disasters such as earthquakes and tsunamis and can help us understand how we are affecting and being affected by changes in Earth's climate and atmosphere. Therefore, the advancement of deep-sea research technologies would be highly beneficial in a number of applications.

2.2 Energy supply for autonomous underwater vehicles

To meet the above-mentioned social demands for deep-sea operations, underwater robots have been developed worldwide. AUVs have become essential in deep-sea operations such as ocean bottom exploration and underwater surveying. However, the operation time of AUVs is limited because of their limited power capacity. Returning to the surface to recharge their batteries reduces their efficiency because of the time lost returning to the surface and the manpower required on the surface vessel. One possible solution to overcome this problem is the use of an underwater battery

recharging unit with a docking function that would allow the AUV to continue its operation in the seabed without returning to the surface station for recharging.

2.3 Turbidity

Although AUVs do not need to descend to the sea bottom for some tasks, such as bottom topology surveys, working near the sea floor is necessary for most ocean exploration operations, including oil pipe inspection and the detection and extraction of precious metals. At this time, the most challenging and unavoidable problem in deep-sea operations is turbidity, which deteriorates the visual capabilities of AUVs. Turbidity is defined as cloudiness in a liquid caused by the presence of suspended particles that scatter and absorb light. Therefore, the verification of the turbidity tolerance of an AUV and the development of a method to overcome disturbances caused by turbidity are important research questions not only for AUV development but also for the field of vision-based underwater systems. Because the intended application in this study is underwater battery recharging at the sea bottom to extend the operation time of AUVs, turbidity cannot be avoided by simply operating the AUVs in clean water.

Some studies on underwater image processing related to turbidity have been conducted [19, 30]. However, most are based on static images rather than real-time dynamic images. When visual servoing—in which only visual information is used as feedback in the real-time control of the underwater vehicle—is performed in the sea, where the environment dynamically changes with turbidity, the tolerance of the proposed visual-servoing approach against turbidity must be verified. The difficulty with this is that turbidity can vary from site to site based on the characteristics of the suspended materials, including their particle shape, refractive index, and color. Unfortunately, there are no universal techniques for analyzing the effects of turbidity on the performance of image-based approaches for different applications, even though the limits at which each approach achieves an acceptable level of accuracy must be identified. Therefore, it is critical to develop a systematic assessment method that can derive a meaningful link between the on-site conditions and the performance of image-based methods.

2.4 Contributions of the present study

As discussed above, turbidity is a practical problem for AUVs that are operated at the sea bottom. To the best of the authors' knowledge, the turbidity tolerance of real-time visual-servoing-based docking at the sea bottom has not yet been investigated through experiments conducted in an actual sea environment. In addition, the robustness of the proposed stereo-vision-based pose estimation method

against turbidity has not yet been discussed, and the performance of the proposed system in a turbid sea environment has not yet been experimentally validated. Therefore, the main contribution of this study is that it is the first experimental investigation of the practicality of undersea docking under turbid conditions in an actual undersea environment. Visual servoing in a turbid environment using a dual-eye camera for pose estimation was verified experimentally through the analysis of the performance of the proposed method against different turbidity levels. Although visual servoing is not the only solution for the docking of AUVs, it is practically important to evaluate its tolerance before combining visual-servoing technology with an integrated control system. Docking experiments were conducted under turbid conditions in pools and an actual undersea environment to verify the effect of turbidity on pose estimation and assess the real-time visual-servoing performance and turbidity tolerance of the proposed method. This study is expected to extend the applications and capabilities of underwater robotics research and technology.

3 Stereo-vision-based real-time 3D pose estimation

Apart from image-based visual servoing, position-based visual servoing has been developed for the vision-based docking approach proposed by our research group. A tutorial on the visual-servoing technology developed in field of terrestrial robotics can be found in [31]. We introduced and explained our approach in which the real-time relative pose of a 3D marker relative to the underwater vehicle is estimated using stereo-vision through a model-based matching method [7]. This section presents a brief explanation of the proposed 3D pose estimation method to provide background for the present study for the reader's convenience; a more detailed discussion can be found in the previous report [7].

The left and right cameras are fixed to the ROV with parallel lines of sight. The coordinate frames of the ROV and the 3D marker are shown in Fig. 3. As shown in this figure, the search space was set to be around the target on the assumption that there should exist a real 3D marker in the genetic algorithm (GA) search space. When the 3D marker is not found in the search space, the fitness value—a scalar that indicates how well the 3D model matches the real 3D marker—remains low at approximately 0.1 or 0; thus, the vehicle controller can determine the vehicle is not near the 3D marker. This judgment can cause the vehicle to globally search for the 3D marker. However, this recovery behavior is not discussed in the present report. The relative pose between the ROV and the 3D marker is determined from six parameters: x , y , z , e_1 , e_2 , and e_3 ; the first three are the Cartesian coordinates of the 3D marker in the ROV frame

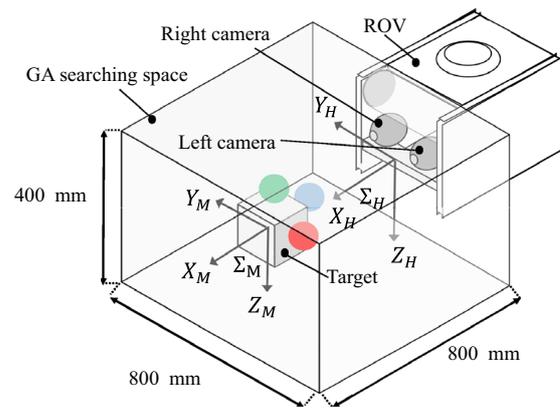


Fig. 3 Coordinate systems of the ROV and 3D marker. Σ_M and Σ_H denote the coordinate systems of the 3D marker and the ROV, respectively. The pose of the marker relative to the ROV, represented by the position and orientation of Σ_M with respect to Σ_H , is considered to be the unknown in the 3D pose estimation process

of reference, and the latter three are the orientation of the marker in the ROV frame represented by a unit quaternion avoiding singularities.

3.1 3D model-based matching method

A model-based matching method was used to recognize the 3D marker and estimate its pose in real time using stereo-vision. Figure 4 shows the 3D marker coordinate system Σ_M , the i th model coordinate system Σ_{M_i} , the left and right camera coordinate systems Σ_{CL} and Σ_{CR} , and the left and right image coordinate systems Σ_{IL} and Σ_{IR} . The target is a 3D marker that consists of three spheres of 40 mm in diameter colored red, green, and blue, and the origins of Σ_M and Σ_{M_i} are the intersections of the three lines perpendicular to the faces to which the spheres are attached.

In conventional approaches, object recognition including relative pose information is implemented by feature-based recognition using 2D-to-3D reconstruction calculations, in which the information of the target object is determined from a set of points in different images, generally using epipolar geometry. The main challenge in this type of approach is ensuring that points are correctly mapped. If a point in one image is incorrectly mapped to a point in another image, the pose of the reconstructed object does not represent that of the real 3D object. Figure 4 shows the incorrect reconstruction of a point through 2D-to-3D reconstruction resulting from incorrect mapping. To avoid incorrect mapping, which results from the original problem being ill posed, a model-based pose estimation approach based on 3D-to-2D projection was applied in this study, because the forward projection from 3D to 2D generates unique points in 2D images without any errors, meaning incorrect mapping is avoided. As shown in Fig. 4, the j th point on the i th model in

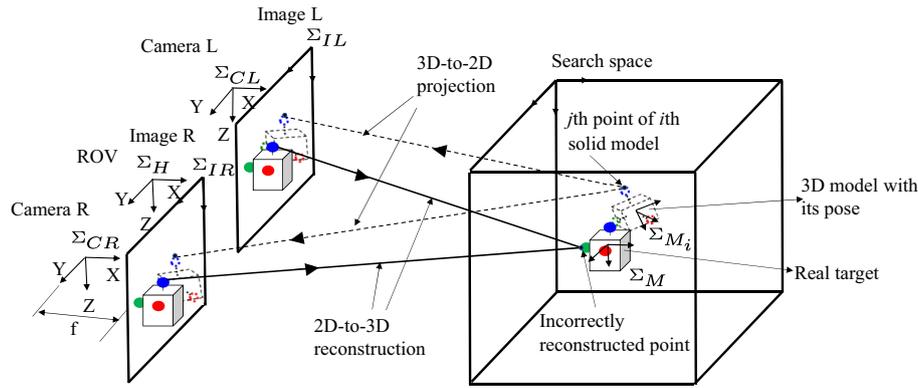


Fig. 4 Model-based pose estimation using the dual-eye vision system in the coordinate systems of the left and right cameras, the object (solid object), and the model (represented by a dotted box and dotted spheres). The j th point on the i th model in 3D space can be described

in each coordinate system using these coordinates and homogeneous transformations. Similarly, a 3D model with its pose defined as a group of points in 3D space is projected onto the left and right camera images through 3D-to-2D projection

3D space is projected onto the left and right camera images correctly. With this approach, 3D model-based recognition can be implemented.

A solid model of the target is defined in the computer system based on known information about the target, such as its shape, size, and color, and this model is projected onto 2D planes. The relative pose is then calculated by computing the correlation between the projected solid model image and the 2D images captured by the dual-eye cameras. In the pose estimation process, the main task is to plot the number of solid models with different poses within the search space, which is defined according to the field of view of the cameras. The next step is to search for models that match the target in the 2D images to within a defined fitness value. Finally, the pose of the model that most closely fits the target in the 2D images is selected as the estimated relative pose. Figure 4 shows how a solid model is defined in 3D space and projected onto 2D images to match the captured real target in 2D images. To measure the fitness value between the target in the image and the projected model with its pose, a correlation function called the fitness function in this study was

designed, as explained in the next section. Note that the pose calculation and convergence are executed in 3D space and the evaluation is performed using the 2D images.

3.2 Fitness function

The fitness function was designed for use as an evaluation parameter in the pose estimation process. It is defined the correlation between a projected model and a real target in the image. In Fig. 5a, the three solid circles and the circles outlined with broken lines (broken circles) represent the spheres on the real target and those on the j th model obtained from 3D-to-2D projection, respectively. The pose ϕ_M^j (j means j th point on a model and M stands for model's coordinates, Σ_M) of the 3D model is an unknown variable composed of six parameters ($x, y, z, \epsilon_1, \epsilon_2, \epsilon_3$, where the first three are position and the latter are orientation) and is determined in the pose estimation process by RM-GA. The 2D projection of each sphere in the model is divided into two regions, as shown by the dashed circles in Fig. 5b. Instead of evaluating the positions of all of the points in

Fig. 5 **a** Real target (solid circles with three colors, red, green, and blue) and projected 3D model (circles with broken lines) in a 2D image obtained by the right camera. **b** Projection of the green sphere of a model with selected sample points. There are a total of 60 points (36 and 24 points in the inner and outer regions, respectively) in the projection, and the diameter of the inner region is same as that of the actual sphere. (Color figure online)

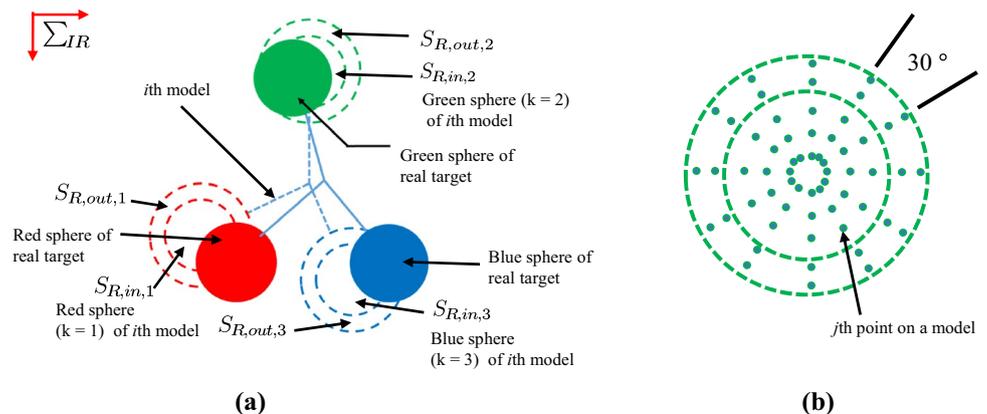
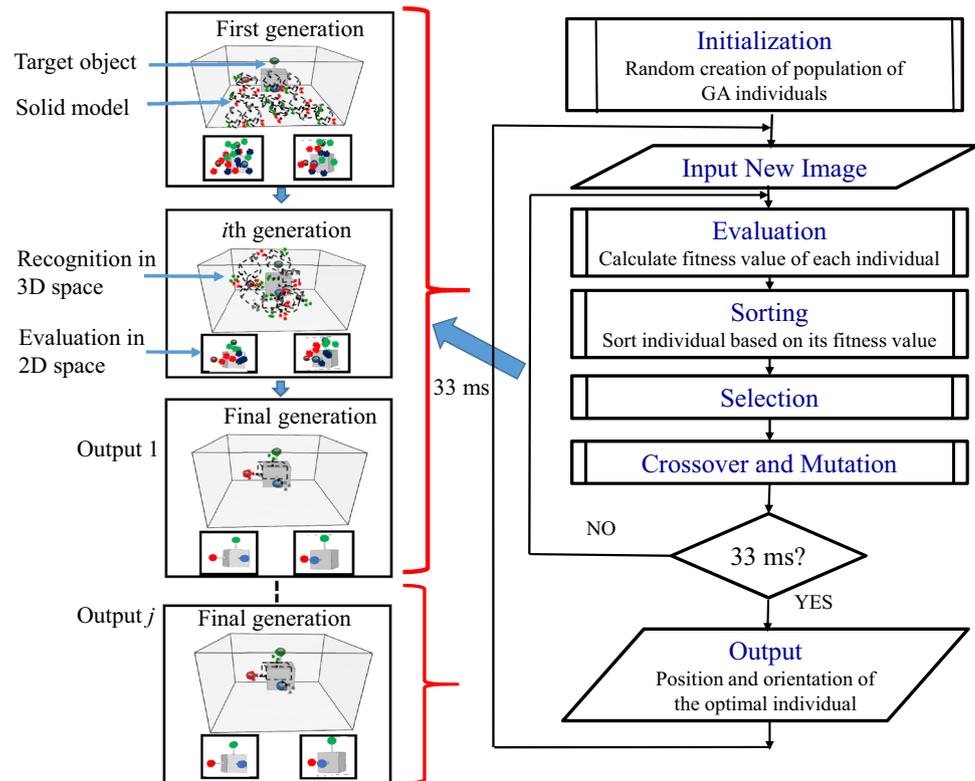


Fig. 6 Pose estimation using the RM-GA. In the initialization step, 60 models with different random poses are generated. The models with the best fit are selected based on their fitness values and evaluated using GA processes (selection, crossover, and mutation). The pose of the model with the highest fitness value in the final generation of the GA process within 33 ms is output as the estimated pose of the 3D marker relative to the vehicle



the model, only select points are considered, as shown in Fig. 5b. When the j th model is projected onto the 2D images of the left and right cameras, the value of fitness function for that model is calculated. Note that the value of fitness function is hereafter referred to as fitness value and defined as F . Portions of the target object that lie inside the inner and outer regions of each corresponding sphere of the projected model proportionally increase and decrease the fitness value, respectively. Therefore, the fitness function is maximized when the pose of the model fits that of the target object depicted in the images of the left and right cameras. Then, solving the pose estimation problem has been converted into optimization problem, then how much the height of the peak does not concern with the solving procedure of optimization. The evaluation parameters of the fitness function were designed to reduce the effect of noise, which is considered here as peaks in the fitness function that represent incorrect poses of the target. A detailed definition of the fitness function can be found in [7, 32], and [33]. The concept of the fitness function in this study can be said to be extension of the work in [33], in which different models, including a model with rectangular surface strips, were evaluated using images from a single camera.

3.3 Real-time multi-step genetic algorithm

In the process of 3D pose estimation, it is assumed that there are many models in the search area, as shown in the top left image in Fig. 6. To determine which model is closest to the actual target, the fitness function defined in the previous

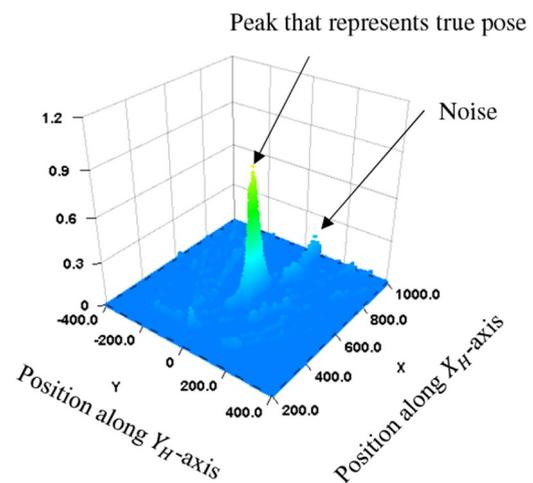


Fig. 7 Fitness distribution. The peak represents the true pose detected by the designed fitness function. The noise, which represents incorrect poses, is generated in the fitness distribution as a result of image deformation caused by environmental effects

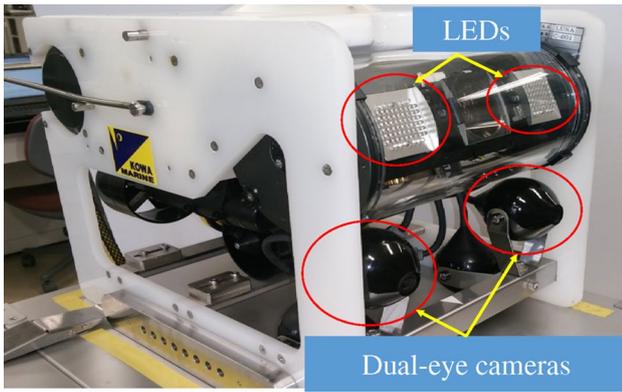


Fig. 8 Photograph of ROV

section was used to quantify the correlation between the models and the target. The main task of the pose estimation process is to search for the optimal model with the pose that is most strongly correlated with that of the real 3D marker. Therefore, the problem of finding the target object and detecting its pose can be converted to searching for the pose ϕ_M^j that maximizes the fitness value $F(\phi_M^j)$. To solve this optimization problem, the RM-GA was developed. The time convergence performance of the RM-GA as a dynamic evaluation function has been confirmed by Lyapunov analysis in [34]. Real-time 3D pose estimation using 3D-model-based recognition and the RM-GA has been presented in detail in the previous papers [7, 29]. Figure 6 shows the flowchart of the RM-GA and how the best model is obtained. The real-time pose can be estimated for every image with an image frame rate of 30 fps. We explained why and how RM-GA was developed for real-time 3D pose estimation in a previous study [7].

3.4 Turbidity tolerance

The fitness distribution with respect to a position in the XY plane based on Σ_H (see Fig. 3) is illustrated in Fig. 7. Because the pose of the target is composed of six parameters (three for position and three for orientation), the fitness distribution with a peak at the true pose can be seen in 3D space, including the fitness value and any pairs of dimensions of pose parameters, as shown in Fig. 7. In the plot in Fig. 7, there is a large peak that corresponds to the true pose, and some additional peaks that correspond to other incorrect poses are present. The proposed system can be considered robust as long as the highest peak of the fitness distribution represents the true pose and the effect of the noise that represents incorrect pose is significantly less than this peak. The shape of the fitness distribution will change in a dynamic image with a video rate of 30 fps.

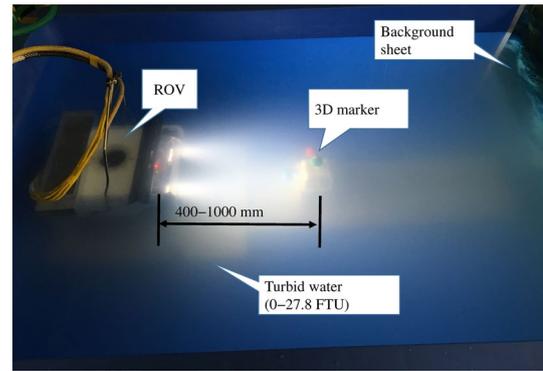


Fig. 9 ROV and 3D marker in turbid water. The turbidity range was from 0 to 27.8 FTU, as measured by the turbidity sensor, and the distance between the ROV and the 3D marker was varied from 400 to 1000 mm

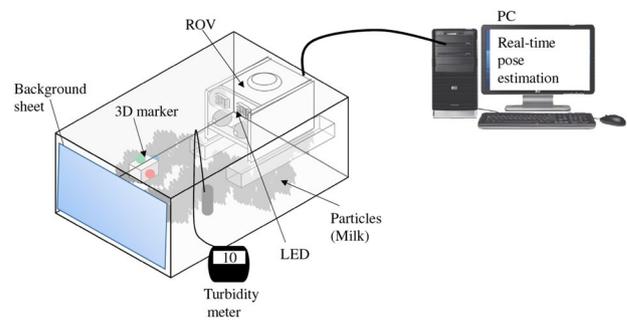


Fig. 10 Experimental layout for 3D pose estimation against turbidity. The system was implemented in a PC (Intel @ Core™ i7-3770 CPU 3.40 GHz, 8.00 GB RAM, 64 bits)

In the sea, turbidity is one of the greatest disturbances to visual servoing and a major source noise. When the turbidity level is high enough to render the designed fitness function ineffective, there will be no peak that represents the true pose of the target. The reason the proposed system can be considered robust against turbidity is that the problem

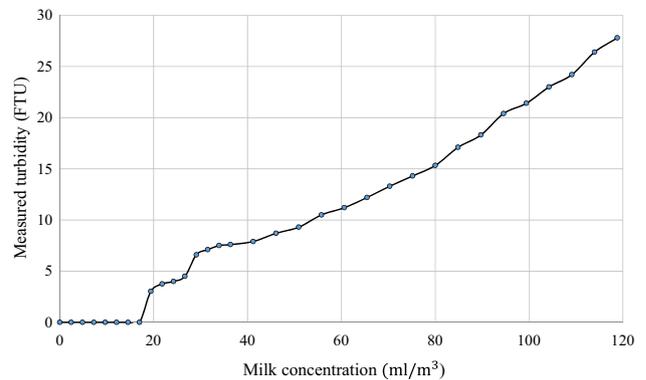
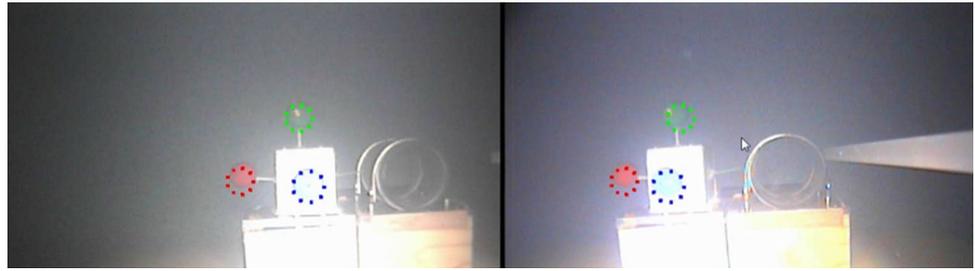


Fig. 11 Relationship between the measured turbidity and the milk concentration

Fig. 12 Projection of the recognized pose onto images taken by the left and right cameras with dotted spheres indicating the positions of the three spheres for user visualization during experiments and analysis



of finding the target object and detecting its pose has been converted to an optimization problem. Therefore, it is necessary to verify the turbidity tolerance of the proposed system. To do this, it is critical to simulate the turbidity levels using a suitable medium and assess the performance of the proposed system against different turbidity levels. The turbidity tolerance of the proposed system was verified in this study, and the experimental results provide an assessment of the system performance against turbidity and demonstrate the potential of the proposed approach for actual sea docking applications.

4 Real-time 3D pose estimation against turbidity

4.1 Underwater vehicle

The remotely controlled underwater robot (Kowa, maximum depth 50 m) used in this experiment is shown in Fig. 8. Two fixed forward-facing cameras with the same specifications (imaging element: CCD, pixel number: 640×480 , pixel focal length: 2.9 mm, signal system: NTSC, minimum illumination: 0.8 lx, no zoom) were mounted on the ROV. These two cameras were used for 3D object recognition. The thruster system of the ROV consists of one vertical, one lateral, and two horizontal thrusters with a maximum thrust of 4.9 N each. In this experiment, only recognition was conducted, and the thrusters of the ROV were not controlled. The ROV is equipped with two LED light units (5.8 W) as an illumination source.

4.2 Experimental layout

In this experiment, 3D pose recognition was conducted using the proposed system under different turbidity levels. In this experiment, the ROV was fixed at the same distance from the 3D marker, and illumination was kept constant by setting the two light-emitting diode (LED) units of the ROV to emit light aimed directly at the 3D marker, as shown in Fig. 9, with an illuminance of 200 lx. The illuminance was measured using a lux sensor (model: LX-1010B manufactured by Milwaukee) placed 600 mm in front of the LED of the

ROV. Figure 10 shows the experimental layout for 3D pose estimation under different turbidity levels. The experiments were conducted in a dark environment, where the LED of the ROV is the dominant light source.

Water turbidity was simulated by adding milk to the water in which the system was submerged. According to the

Table 1 Average fitness value distribution for different turbidity levels and distances between the ROV and 3D marker [the first column gives the turbidity level measured by the turbidity sensor (FTU) and the corresponding amount of milk (ml/m^3). The fitness values, represented by F , are given at each of the considered turbidity levels and with the distance between the ROV and 3D marker ranging from 400 to 1000 mm. Area I represents the controllable area ($F \geq 0.60$). Area II represents the recognition area ($0.22 \leq F < 0.60$). The remaining, Area III, is the loss of recognizable area ($F < 0.22$). The labels A–F represent the conditions in which docking experiments were conducted in another pool.]

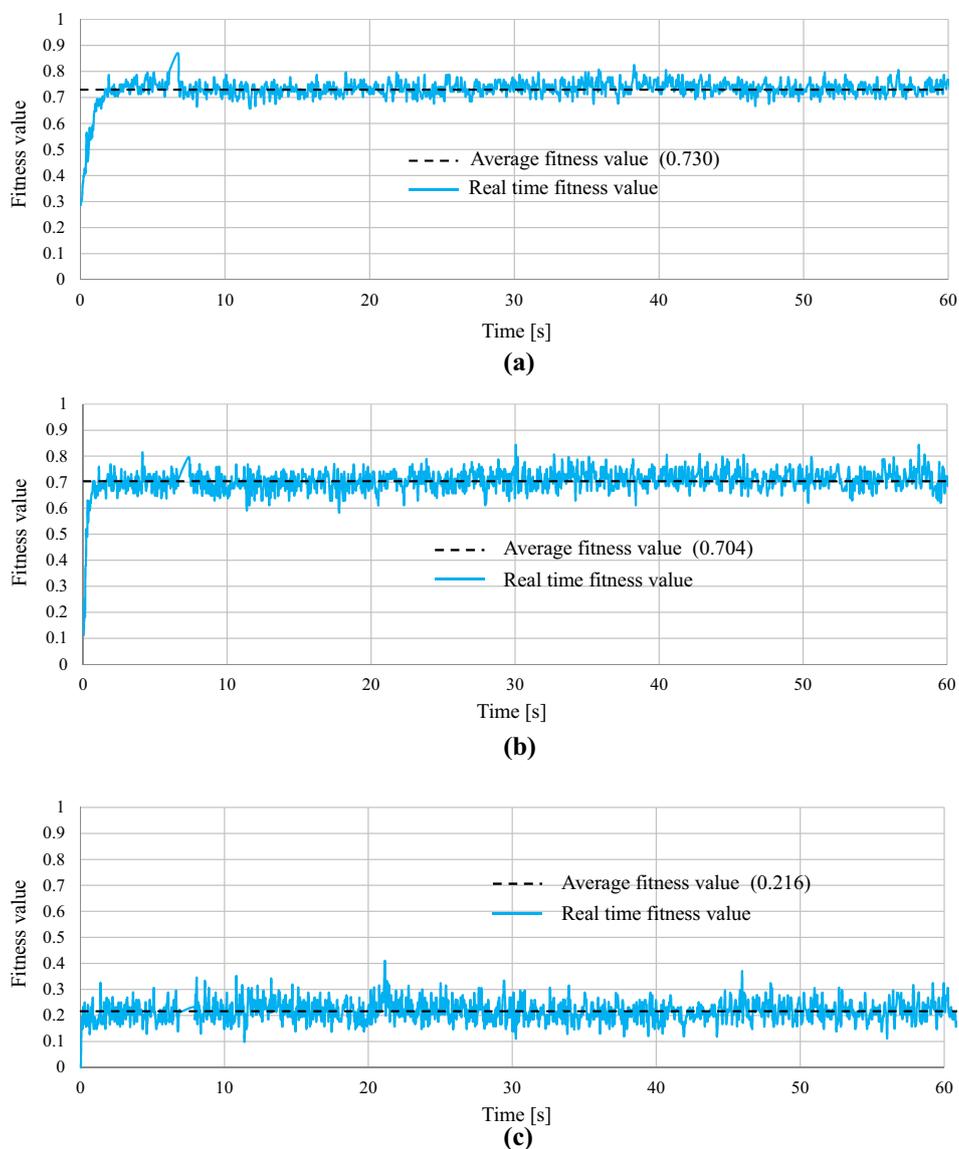
Turbidity	Distance	400 mm	600 mm	800 mm	1000 mm
0.00 FTU (0.00 ml/m^3)		0.680	A 0.730	0.775	0.748
0.00 FTU (2.43 ml/m^3)		0.873	0.886	0.759	0.750
0.00 FTU (4.85 ml/m^3)		0.766	0.834	0.721	0.589
0.00 FTU (7.28 ml/m^3)		0.921	0.817	0.660	0.555
0.00 FTU (9.70 ml/m^3)		0.849	0.746	0.573	0.420
0.00 FTU (12.13 ml/m^3)		0.764	B 0.744	0.498	0.490
0.00 FTU (14.55 ml/m^3)		0.739	0.700	0.672	0.455
0.00 FTU (16.98 ml/m^3)		0.708	0.654	0.528	0.301
3.03 FTU (19.40 ml/m^3)		0.697	0.644	0.535	0.283
3.75 FTU (21.83 ml/m^3)	I	0.671	0.750	0.327	0.0619
4.00 FTU (24.25 ml/m^3)		0.687	C 0.704	0.395	0.0575
4.50 FTU (26.68 ml/m^3)		0.673	0.710	0.325	0.133
6.60 FTU (29.10 ml/m^3)		0.682	0.654	II 0.380	0.0596
7.10 FTU (31.53 ml/m^3)		0.667	0.645	0.298	0.298
7.50 FTU (33.95 ml/m^3)		0.666	D 0.651	0.214	III 0.0556
7.60 FTU (36.38 ml/m^3)		0.646	0.589	0.257	0.0572
7.90 FTU (41.23 ml/m^3)		0.639	0.593	0.182	0.183
8.70 FTU (46.08 ml/m^3)		0.606	0.582	0.159	0.0579
9.30 FTU (50.93 ml/m^3)		0.618	E 0.527	0.157	0.0552
10.50 FTU (55.78 ml/m^3)		0.578	0.294	0.124	0.135
11.20 FTU (60.63 ml/m^3)		0.578	F 0.216	0.098	0.151
12.20 FTU (65.48 ml/m^3)		0.545	0.262	0.130	0.118
13.30 FTU (70.33 ml/m^3)		0.565	0.0876	0.137	0.167
14.30 FTU (75.18 ml/m^3)		0.446	0.217	0.0529	0.0599
15.30 FTU (80.03 ml/m^3)		0.549	0.150	0.0582	0.153
17.10 FTU (84.88 ml/m^3)		0.371	0.118	0.0604	0.166
18.30 FTU (89.73 ml/m^3)		0.485	0.144	0.149	0.152
20.40 FTU (94.58 ml/m^3)		0.444	0.120	0.189	0.223
21.40 FTU (99.43 ml/m^3)		0.422	0.155	0.148	0.145
23.00 FTU (104.28 ml/m^3)		0.396	0.149	0.167	0.0599
24.20 FTU (109.13 ml/m^3)		0.100	0.126	0.129	0.140
26.40 FTU (113.98 ml/m^3)		0.0892	0.159	0.166	0.169
27.80 FTU (118.83 ml/m^3)		0.172	0.0919	0.210	0.215

previous literature reviews [17, 18], the diameter of milk molecules ranges from 10 to 600 nm. Particles with a diameter of 10 nm scatter equal amounts of light forward and backward. The forward scattering begins to dominate for particles with diameters of approximately 100 nm, and close to 1000 nm, there is strong small-angle forward scattering and weak backscattering. Therefore, milk was selected to model the turbidity, because it can provide all types of scattering. On the basis of the maximum milk concentrations of 0.19 ml/l (190 ml milk in 1000 l water) in [17] and 1.5 ml/l in [18], the experiments in this study were conducted with milk concentrations of up to 0.12 ml/l (95 ml milk in 800 l water). Note that the light sources used in [17] and [18] are different from that used here. Two fluorescent light strips

were used in [17], and a halogen lamp was used in [18]. In the present experiment, two LED units installed on the ROV, as shown in Fig. 8, were used as a light source. The ROV was positioned in front of the 3D marker at a fixed distance ranging from 400 to 1000 mm. During an actual docking operation, the ROV approaches the 3D marker from a distance of approximately 1000 mm. It then performs visual servoing and the final docking stage from distances of approximately 600 mm and 400 mm, respectively. Details of this docking process are provided in Sect. 5.1.

To provide an experimental environment similar to a real undersea environment, a background sheet printed with an image similar to what would be observed in a real sea environment was placed behind the 3D marker, as shown

Fig. 13 Real-time and average fitness values under the conditions labeled **a** A, **b** C, and **c** F in Table 1



in Fig. 10. The pool size is 1580 mm × 1100 mm × 590 mm, and the pool was filled with 800 l of water. Milk was added to the water in increments of 2 g and 4 g up to a total of 30 g and 98 g, respectively, to run the experiment at different levels of turbidity. The turbidity of the water was measured using a turbidity sensor (model: TD-500 manufactured by OPTEX) with a measurement range of 0.0–500 FTU (Formazin Turbidity Unit). The relationship between the measured turbidity and the milk concentration is illustrated in Fig. 11.

4.3 Evaluation of 3D recognition

In this experiment, the fitness value was used to evaluate the performance of the proposed pose recognition method at different turbidity levels. The correlation function of the real target projected onto the camera images and the assumed model, which was represented by poses in the chromosomes, is used as the fitness function in the GA process, in which selection, crossover, and mutation are performed to reproduce the next generation via evaluation by the fitness

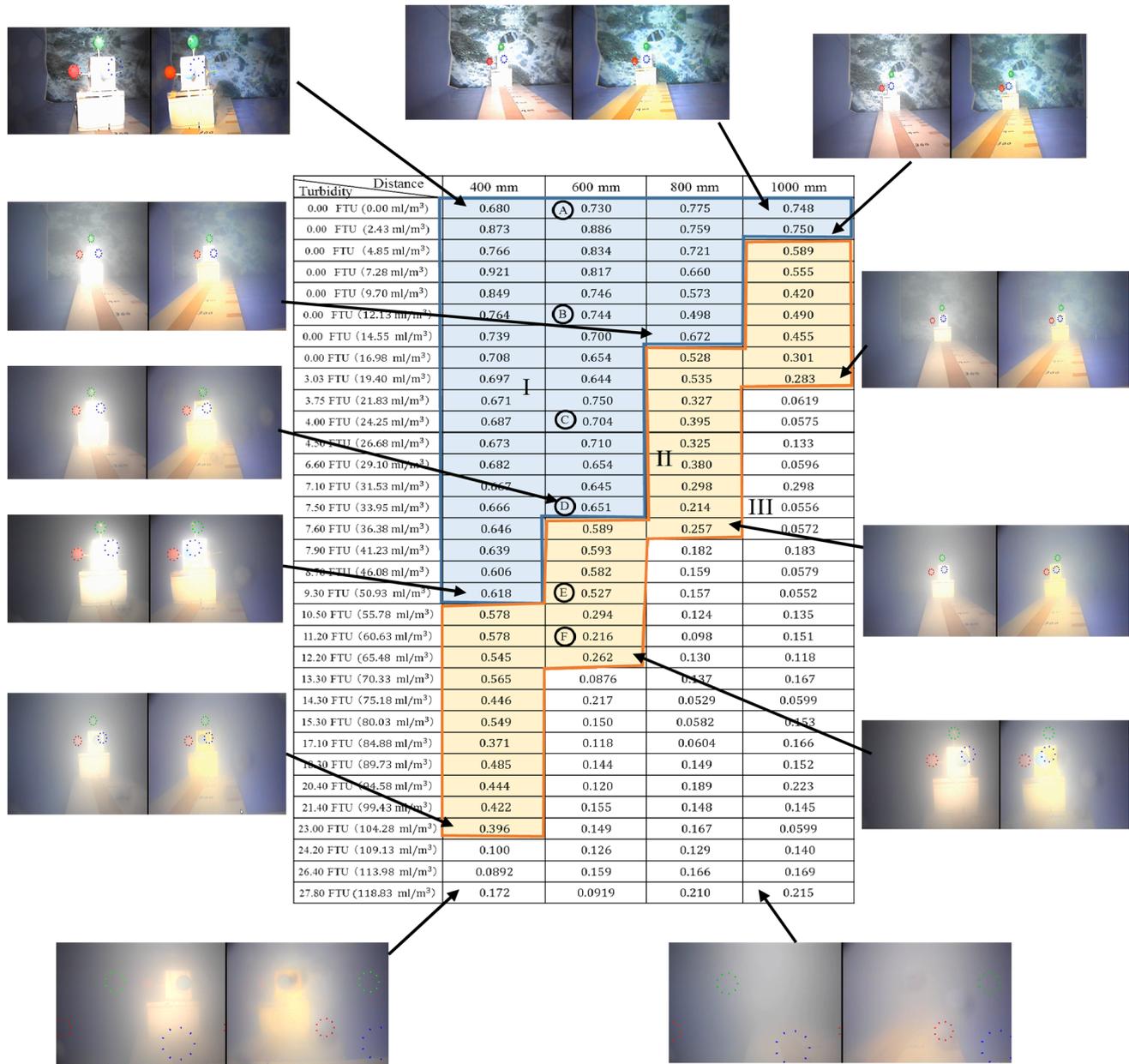


Fig. 14 Left and right camera images under the maximum turbidity conditions in the control and recognition areas at each considered distance. Images taken at the maximum and minimum distances in clean

water and at the maximum turbidity, in which the 3D marker is not observable, are also shown at the top and bottom, respectively

function. The fitness function was modified based on the voting performance and the characteristics of the target (color, size, and shape). In this study, the fitness value averaged over 60 s was used to verify the performance of the proposed system under different turbidity levels. In addition to using the fitness value to evaluate the recognition performance, the recognized pose was visually evaluated by projecting the spheres on the target onto the left and right camera images, as shown by the dotted circles in Fig. 12. Moreover, to evaluate the accuracy of RM-GA, a full-search method in which

the vehicle searches globally for the 3D marker is performed using left and right cameras images for offline analysis.

4.4 Pose recognition against turbidity

A total of 132 iterations of this experiment were conducted at different turbidity levels and distances between the ROV and the 3D marker. The turbidity tolerance in terms of the fitness value under each set of conditions was analyzed. The maximum amount of milk added to the 800 l of water in the pool was 98 g (118.825 ml/m³); this corresponds to a

Fig. 15 Left and right camera images with the pose recognized by the pose estimation system at different turbidity levels and a distance of 600 mm between the ROV and 3D marker. The recognized pose is indicated by dotted circles in each photograph. The water turbidity measured by the turbidity sensor is shown in units of FTU, and the amount of added milk is given in units of ml/m³



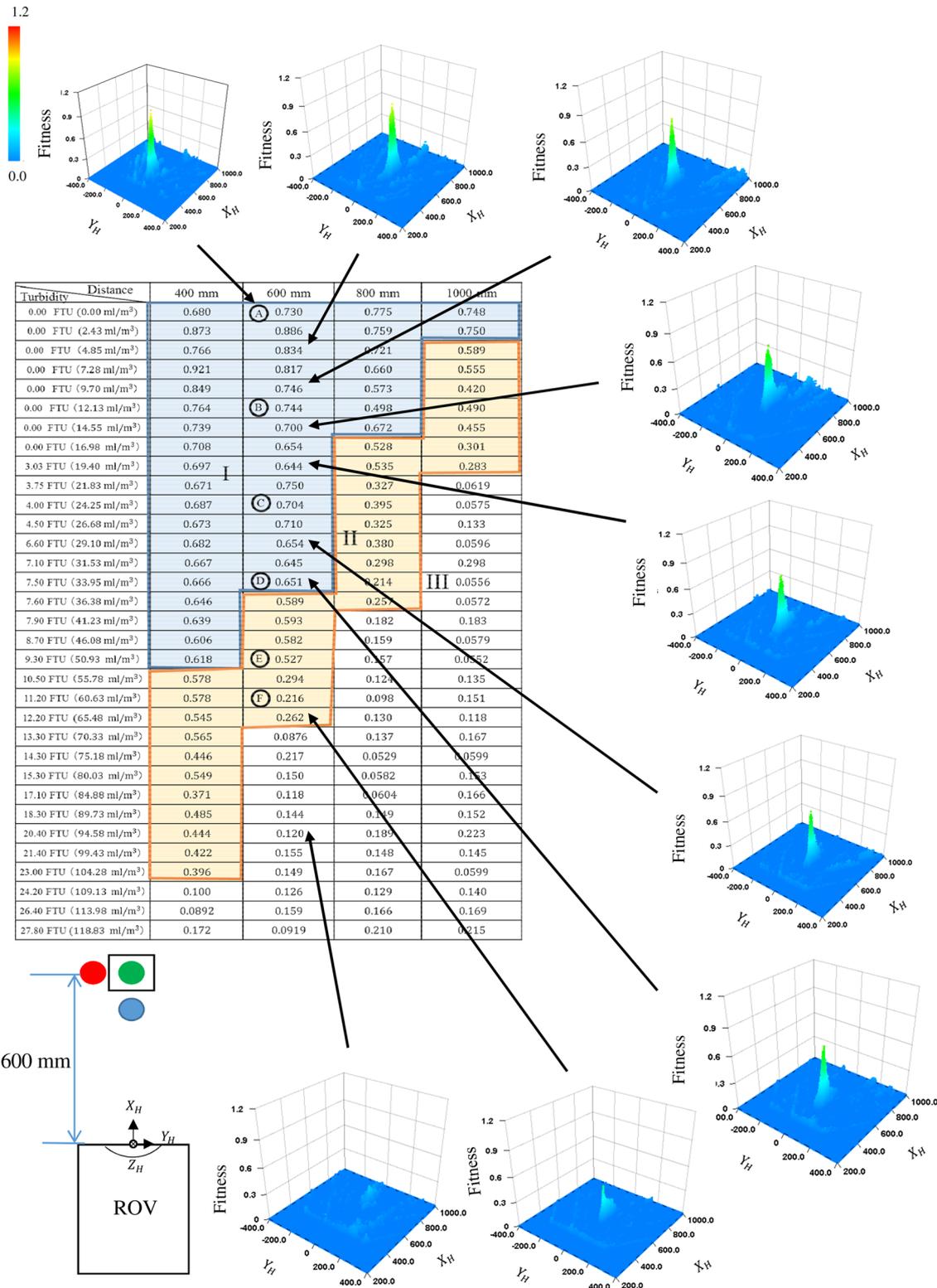


Fig. 16 Fitness value distributions confirming the robustness of the system at a distance of 600 mm. The position of the peak corresponding to the true pose of the marker was maintained even though the

height of the peak was reduced by increasing turbidity. The gradual reduction in the height of peak shows the effect of turbidity on image recognition

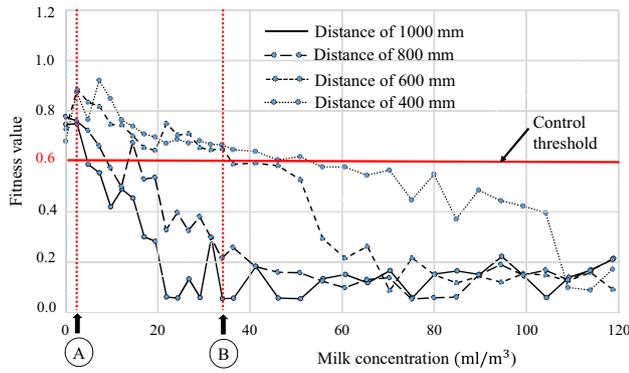


Fig. 17 Fitness value plotted against milk concentration at different distances between the ROV and 3D marker. The fitness value decreased with increasing milk concentration (increasing turbidity). The maximum turbidity levels for approaching from 1000 mm, visual servoing from 600 mm, and docking at 400 mm determined experimentally according to the defined control threshold are indicated by vertical lines. For example, given the defined control threshold of 0.6, approaching from 1000 mm, visual servoing and docking can be performed up to a milk concentration of 2.43 ml/m³ that is indicated by (A) in this figure, and visual servoing and docking can be performed from 600 mm up to a milk concentration of approximately 33.95 ml/m³ indicated by (B). (Refer to Table 1 for more accurate numerical values.)

maximum turbidity of 27.8 FTU, as measured by the turbidity sensor. The maximum amount of milk was selected as the value at which the system could not detect the 3D marker from a distance of 400 mm. Table 1 gives the fitness values at all considered turbidity levels and distances between the ROV and the 3D marker. The first column gives the turbidity level measured by the turbidity sensor in units of FTU and the corresponding amount of milk in units of milliliters per cubic meter. The remaining columns give the fitness values at each of the given turbidity levels with the distance

between the ROV and the 3D marker ranging from 400 to 1000 mm. The fitness values were calculated by averaging the fitness values over a real-time recognition period of 60 s in each case. Figure 13 shows the time profiles of the real-time and average fitness values for some illustrative cases.

There exists a fitness value threshold below which the ROV cannot reliably recognize the 3D marker. Thus, in an actual docking scenario, the ROV would not be controlled by visual servoing when the fitness value is less than the threshold. Area I, shown in Table 1, represents the controllable area ($F \geq 0.6$, F means fitness function value detected RM-GA), which is the area in which the ROV can be controlled by visual servoing. This upper fitness value threshold of 0.6, hereafter called the control threshold, was determined experimentally. Area II is the recognizable area ($0.22 \leq F < 0.6$), in which the ROV can still recognize the 3D marker but the ROV can no longer be reliably controlled using visual servoing. According to the experimental results, the system cannot recognize the 3D marker when the fitness value is less than this lower control threshold of 0.22. Area III in Table 1, the loss of recognition area, represents the cases in which the fitness value is below the recognition control threshold. Figure 14 shows examples of the left and right camera images with the recognized pose of the 3D marker under conditions near the control and recognition thresholds. Example images at the minimum and maximum considered distances and turbidity levels are also shown for comparison. Figure 15 shows examples of the left and right camera images under different turbidity levels at a distance of 600 mm between the ROV and 3D marker. According to the experimental results, the system can recognize the 3D marker up to a turbidity of 12.2 FTU at this distance, which corresponds to the recognition threshold (Table 1). Similarly, the maximum turbidity at which the marker can be recognized for each distance is given in Table 1.

Fig. 18 Docking operation steps, including approaching, visual servoing, and docking. Note that approaching is performed by manual control at the beginning of the docking operation in this study. The error allowance for docking is ± 20 mm in the y - and z -directions and $\pm 7^\circ$ in the rotation about the z -axis

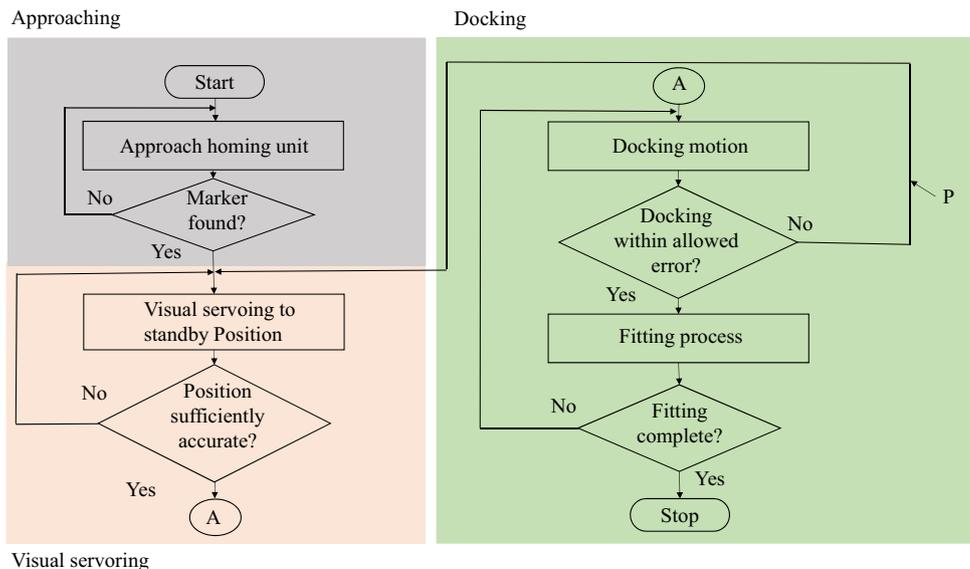


Fig. 19 Layout of the docking experiment showing the process of aligning the ROV with the 3D marker. **a** Desired pose in the visual-servoing step. **b** Desired pose at the completion of the docking step

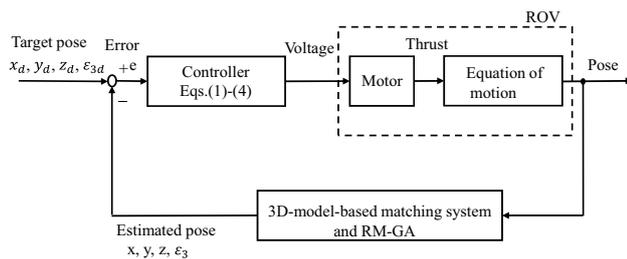
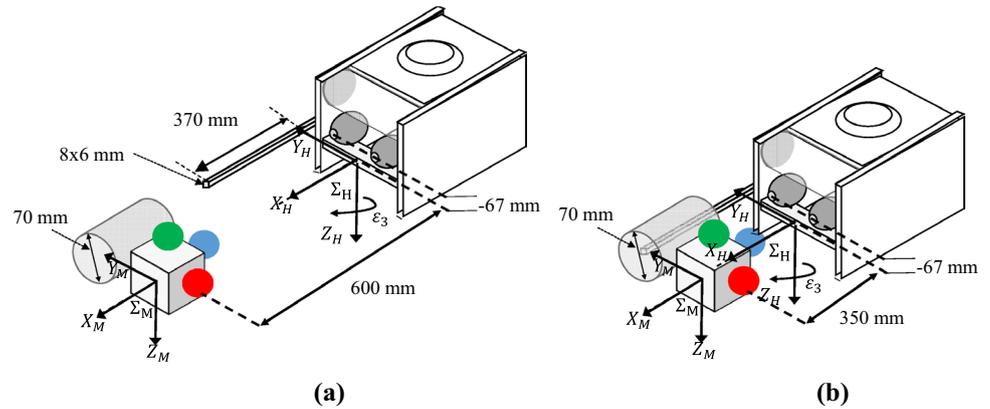


Fig. 20 Block diagram of 3D-MoS for visual-servoing-based underwater vehicle

Figure 16 shows the fitness distributions at different turbidity levels for a distance of 600 mm between the ROV and 3D marker. The height of the peak of the fitness distribution decreased with increasing turbidity. However, in areas I and II, the pose represented by the peak corresponded with the true pose even though the height of the peak was reduced by increasing turbidity.

These experimental results confirm that the proposed system is robust against turbidity up to 12.2 FTU at a distance of 600 mm. When the turbidity level reached 20.4 FTU, there was no peak at the true pose (Fig. 16), indicating that the 3D marker could not be recognized. In a real undersea environment, it is necessary for the ROV to be able to determine whether the turbidity level is too high to conduct the docking operation. To enable this independent determination, the control threshold could be defined based on the experimental results, as shown in Fig. 17. The recognition results given in terms of the fitness are useful as a turbidity indicator during docking operations. For example, the ROV can return to the sea surface or wait until the turbidity level is low enough to dock and recharge its battery. In Fig. 17, fitness value is plotted against milk concentration at different distances between the ROV and 3D marker. When the control threshold is determined experimentally as 0.6, approaching from 1000 mm, visual servoing and docking can be performed up to a milk concentration of 2.43 ml/m³, and visual servoing

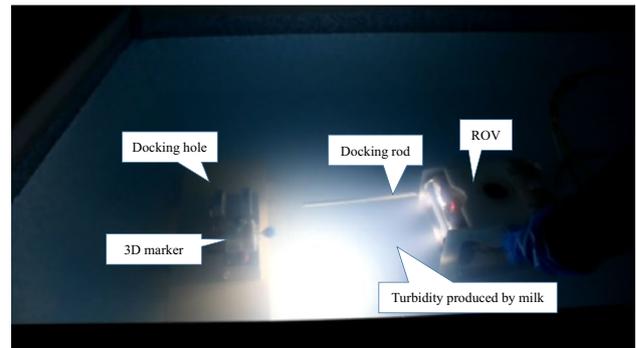


Fig. 21 Photograph of the docking experiment under turbid conditions in a dark environment

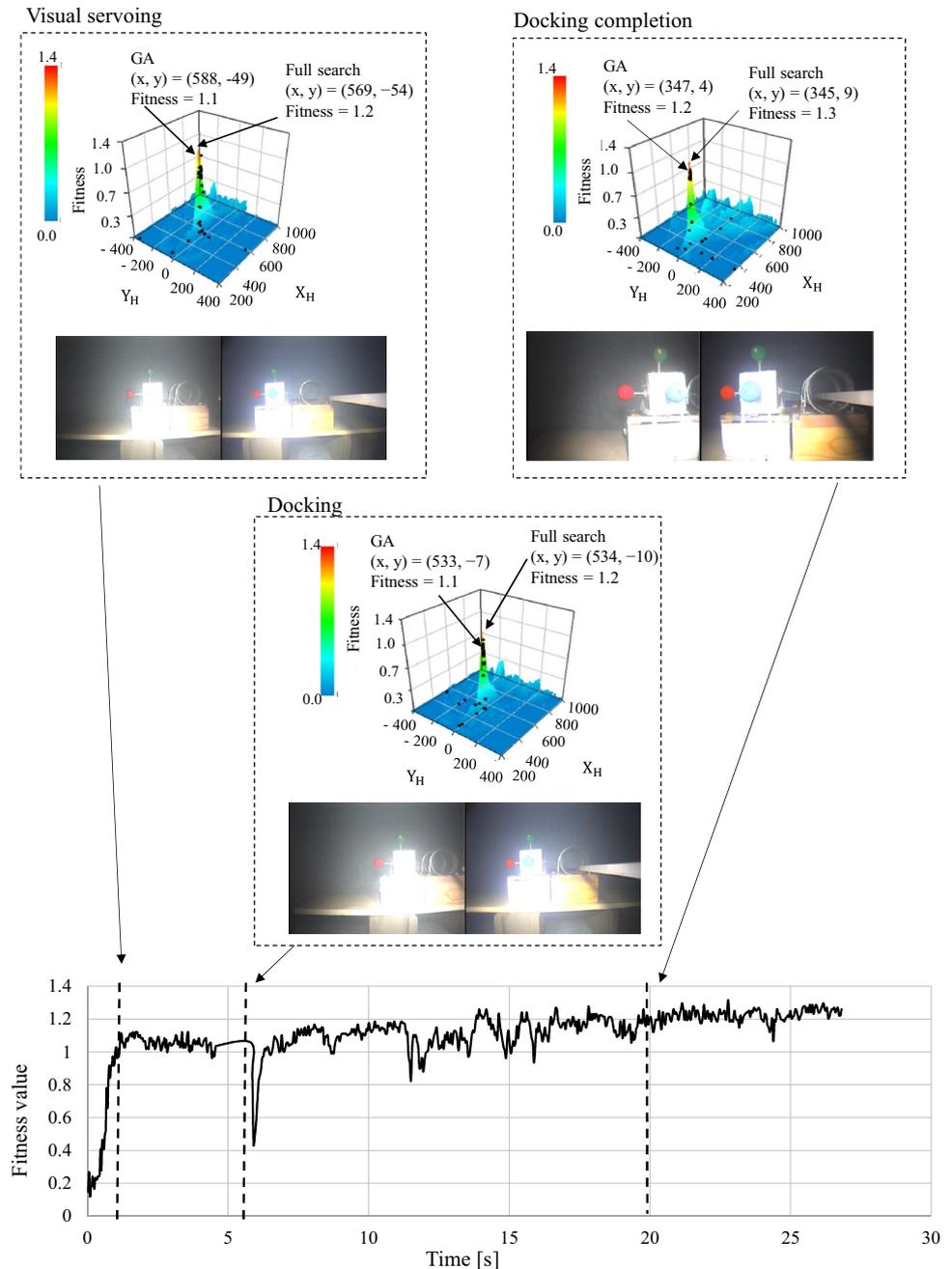
and docking can be performed from 600 mm up to a milk concentration of approximately 33.95 ml/m³ with the turbidity of 7.5 FTU. Based on this experimental result, the docking performance were verified under selected different turbidity levels and explained in the next section.

5 Docking experiment under turbidity in a pool

5.1 Docking method

In an underwater battery recharging operation, the AUV must navigate to a seafloor station and dock at the station for recharging. Normally, a cone-shaped docking station with light sources mounted around its entrance is used for a torpedo-shaped AUV. The hovering-type AUV studied in [15] docks by descending to the station. To perform docking experiments for underwater automatic charging using the proposed approach, a rod was installed on the right side of the underwater robot, and a matching cylindrical hole was attached to the left side of the target. When the ROV is in the correct pose relative to the object, it must move forward to insert the rod into the hole. A flowchart of the proposed docking method is shown in Fig. 18.

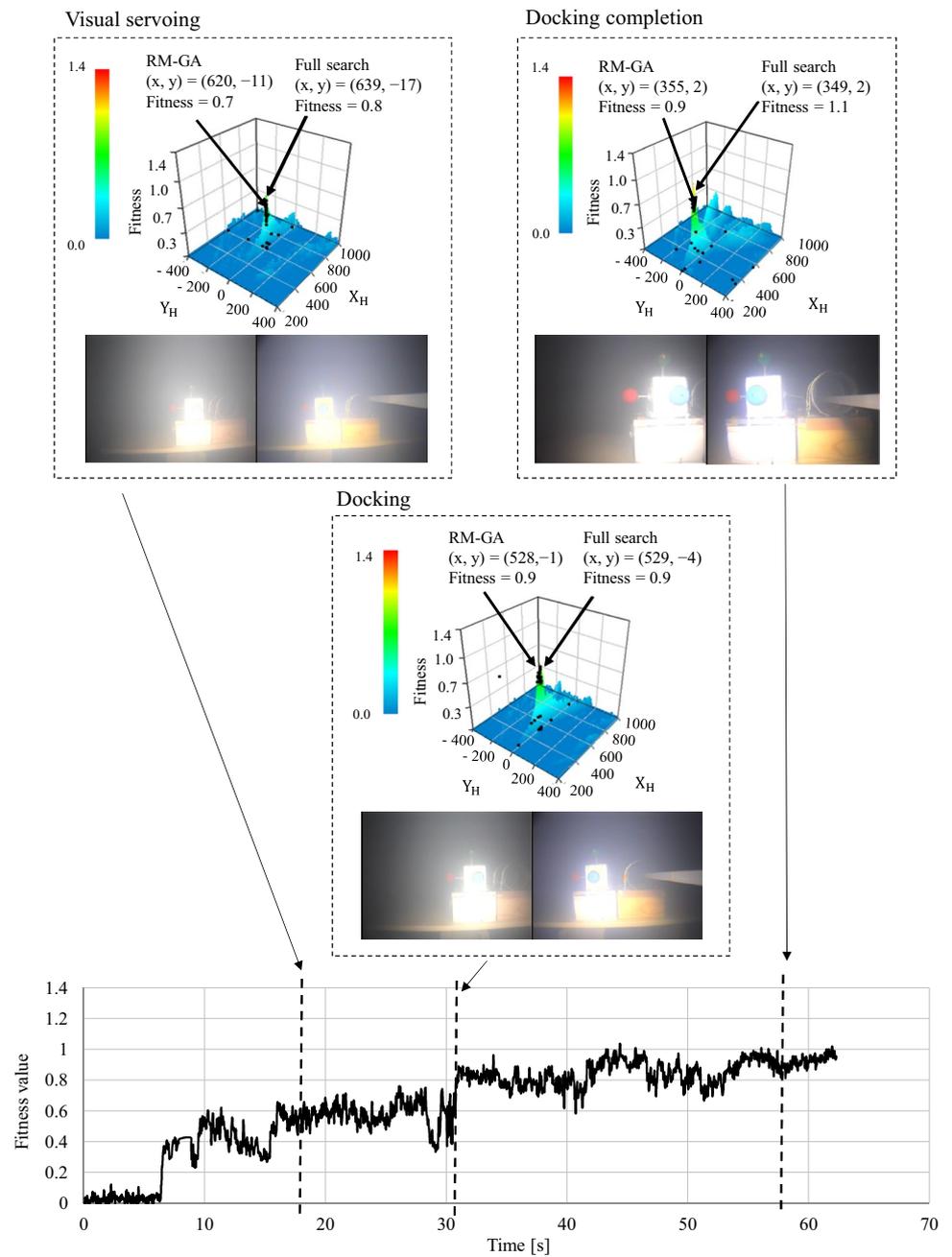
Fig. 22 Fitness value results for experiment B. The photographs show examples of the left and right camera images from which the pose was estimated using the RM-GA. From left to right, the photographs show selected images from the visual-servoing step, the docking step, and after the completion of the docking step. The poses estimated using the RM-GA and the full-search method are indicated in the fitness value distributions for each of these docking steps. The area around the peak of the fitness distribution was searched by scanning all planes of the images. The presence of a peak in the distribution indicates the robustness of the recognition method against turbidity, and the correspondence between the peak and the black points indicates the accuracy of the RM-GA results. The black point represents each gene of RM-GA. The pose yielded by the RM-GA is shown in Fig. 24



This method is divided into the three steps: approaching, visual servoing, and docking. In the approaching step, the AUV estimates its approximate position relative to the station using other methods, such as an acoustic method, and approaches the station. The approaching step was performed manually in this study. In the visual-servoing step, the AUV measures the precise relative pose using stereo-vision-based estimation and remains at the entry point for the final docking step, as shown in Fig. 19a. In the docking step, the AUV moves to the final desired pose, as shown in Fig. 19b, at a speed of 30 mm/s in the x -direction as long as the errors of the y - and z -components

of the position of the ROV and that of its orientation about the z -axis relative to the target remain within ± 20 mm and $\pm 7^\circ$, respectively, for a minimum of 165 ms (five times the length of the control loop). Whenever any of these components of the error of the relative pose exceeds this allowance, the process switches to visual servoing, as shown by the path labeled “P” in Fig. 18. This process of switching between the visual servoing and docking steps was implemented to avoid any physical damage that could result from contact between the rod and the hole in the target.

Fig. 23 Same as Fig. 22 for experiment E. The pose yielded by the RM-GA is shown in Fig. 25



5.2 Motion controller using proportional controller

The proportional controller is used to control the vehicle. The four thrusters that are mounted on the underwater robot are controlled by sending the command voltage based on the feedback relative pose between the underwater robot and the object. The block diagram of the control system is shown in Fig. 20. The control voltage of the four thrusters is given by the following equations:

$$\text{Surge direction} : v_1 = k_{p1}(x_d - x) + 2.5 \tag{1}$$

$$\text{Vertical axis rotation} : v_2 = k_{p2}(\epsilon_{3d} - \epsilon_3) + 2.5 \tag{2}$$

$$\text{Heave direction} : v_3 = k_{p3}(z_d - z) + 2.5 \tag{3}$$

$$\text{Sway direction} : v_4 = k_{p4}(y_d - y) + 2.5, \tag{4}$$

where $v_1, v_3,$ and v_4 are the control voltage of the four thrusters of $x, z,$ and y directions, respectively, shown by Σ_H in Fig. 3. x_d, y_d, z_d are the desired relative pose between the vehicle and the target. ϵ_{3d} is the rotation direction around the z -axis and its command voltage is expressed as the value of v_2 . According to the experimental results, the gain coefficients have been adjusted to perform the best condition for visual servoing.

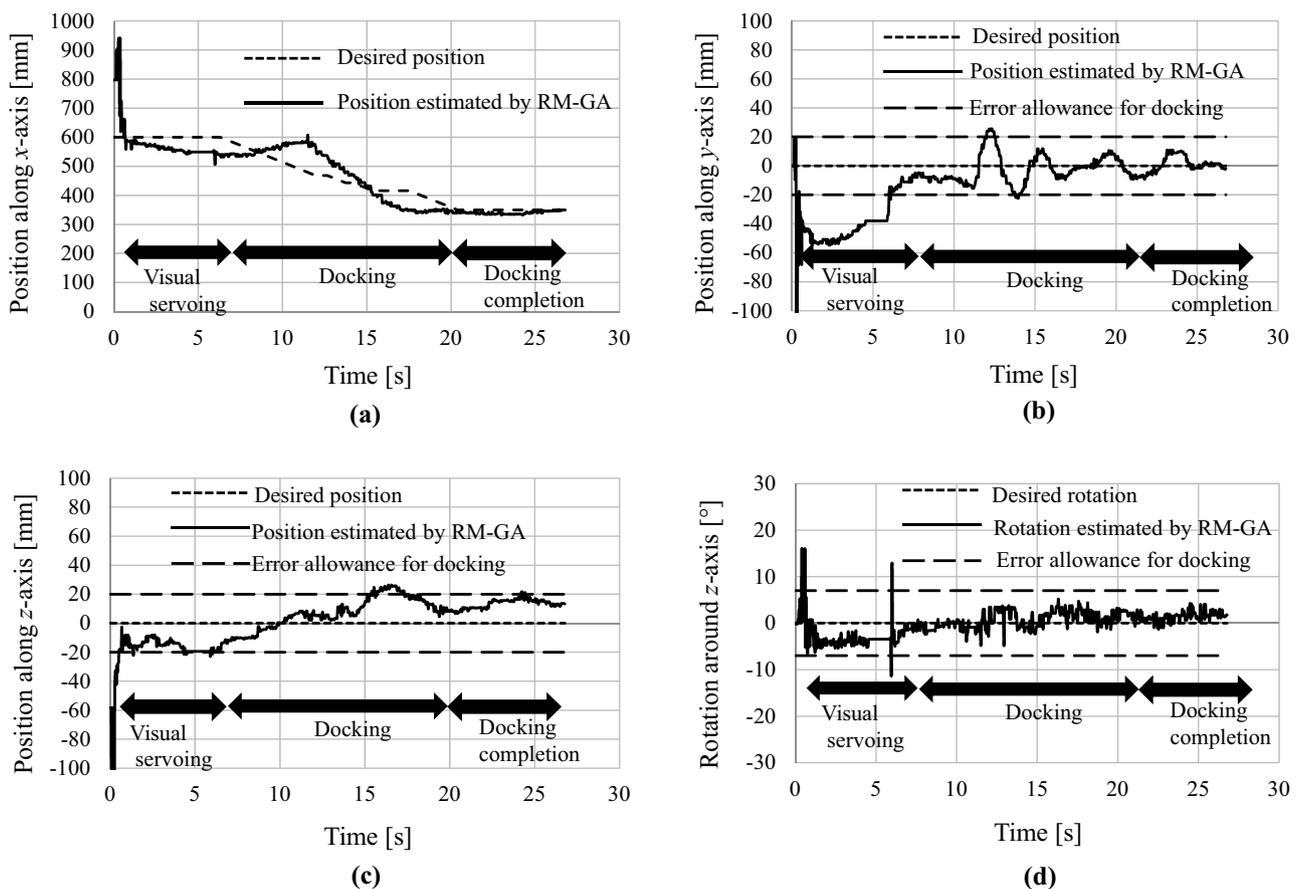


Fig. 24 Position along the **a** x -, **b** y -, and **c** z -axes and **d** rotation about the z -axis estimated using the RM-GA in the docking experiment for experiment B. In this case, the control threshold is 0.6

5.3 Docking task

The docking experiment was conducted under turbid conditions in a large pool (length 2870 mm \times width 2000 mm \times height 1000 mm) filled with 4000 l of clear water. Experiments were conducted in a dark environment, and turbidity was simulated by adding milk in amounts ranging from 0 to 250 g in increments of 50 g, corresponding to the concentration conditions labeled (A)–(F) in Table 1; these docking experiments are hereafter referred to as experiments A–F, respectively. A photograph of this experiment is shown in Fig. 21. This experiment was conducted to verify whether each of the considered conditions were in the control area. In this system, the images acquired from the dual-eye camera are sent to the PC. The real-time recognition of the 3D pose of the target is then executed using the model-based matching method and the RM-GA in the PC software. Finally, based on the error between the actual and recognized poses, command signals generated by a position controller for the thrusters are input into the ROV to ensure it maintains the target pose. In the docking experiments, the ROV is placed at a distance of approximately 600 mm in front of the 3D

marker with an arbitrary initial orientation. The docking alignment process is shown in Fig. 19 along with the marker and ROV coordinate systems.

Among the six docking experiments conducted in the large pool, experiments B and E (milk concentrations of 12.125 ml/m³ and 48.5 ml/m³, respectively) are discussed in detail. Figures 22 and 23 show the fitness values for experiments B and E, respectively. The results obtained using the full-search and RM-GA methods were compared for some sample points during the visual servoing and docking steps and after docking completion. The poses estimated by all of the genes in the RM-GA are represented by black points, and the peak fitness value represents the estimated pose. The fitness distribution for each pose was searched using the full-search method, which involves scanning all planes to find the true pose. A comparison of the fitness values obtained in experiments B and E, especially those obtained during visual servoing, reveals that increasing the turbidity reduces the fitness value.

Figures 24 and 25 show the desired pose of the 3D marker and the pose estimated by the RM-GA. Note that the ROV is controlled manually until the 3D marker is

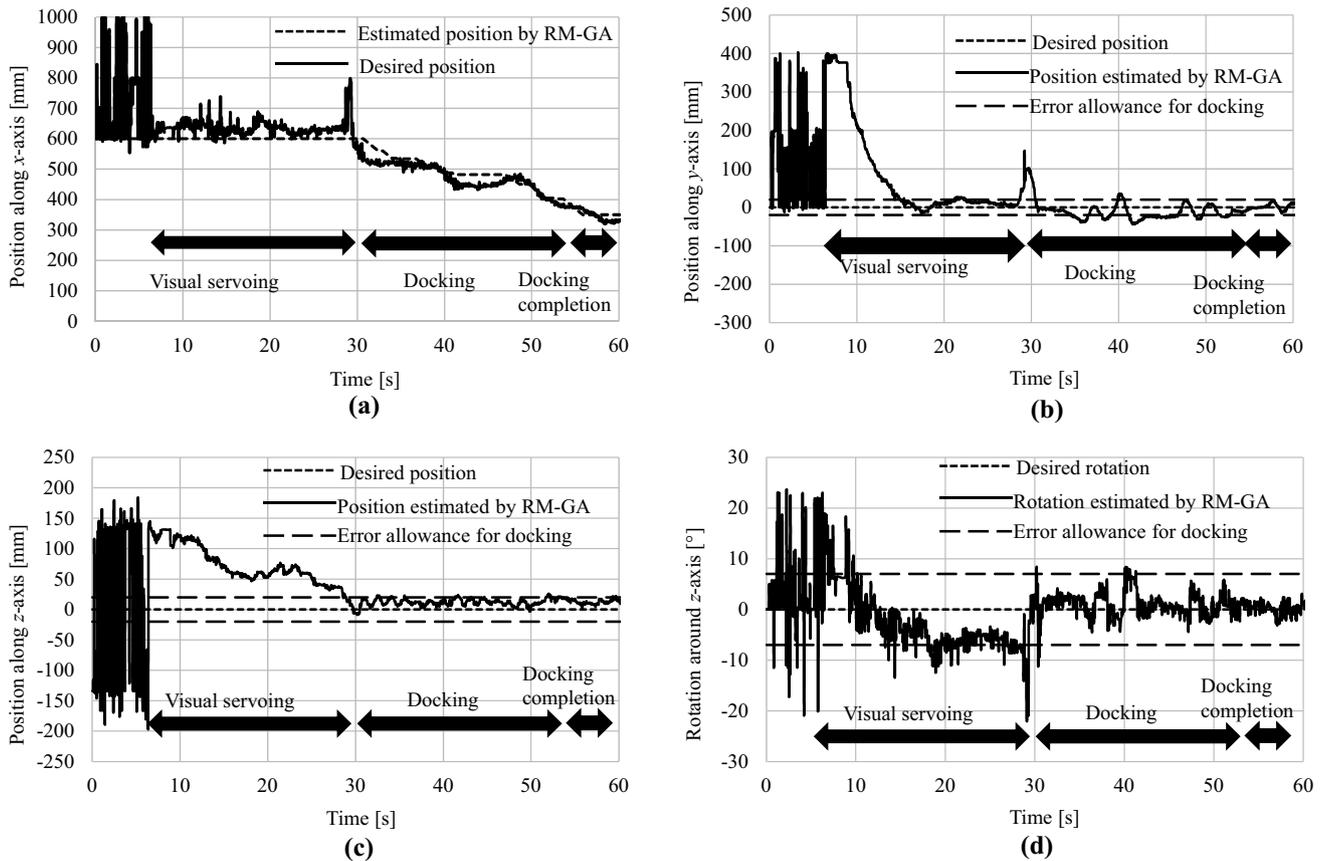
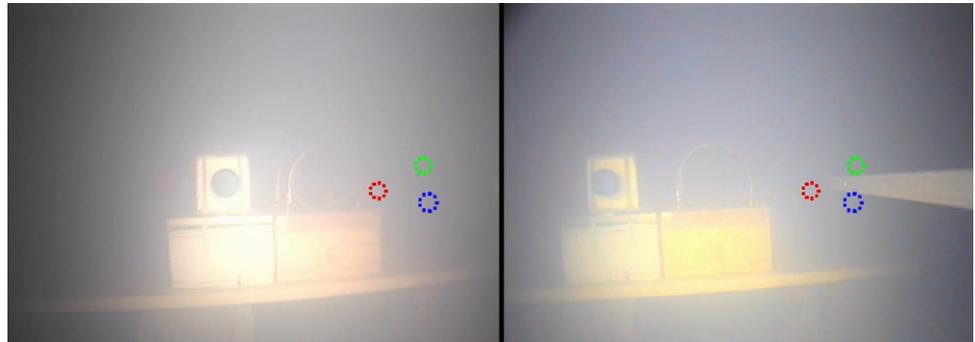


Fig. 25 Same as Fig. 24 for experiment E. In this case, the control threshold is 0.4

Fig. 26 Left and right camera images in experiment F. The dashed circles, which are not aligned with the target, represent the system's failure to recognize the target



recognized by the system. For example, the recognized poses during the first 10 s in Fig. 23 are not meaningful, because the fitness value is less than 0.22. Therefore, visual servoing started at approximately 8 s in this case as a result of the system switching from manual mode to visual-servoing mode. Visual servoing continued until the y - and z -components of the estimated position were within the error allowances of the desired pose. When these docking criteria were satisfied, the docking step was initiated. During the docking process, the rod is inserted into the

docking hole by gradually decreasing the desired value of the position along the x -axis. Note that the desired position along the x -axis reduces until it reaches a distance of approximately 350 mm from the 3D marker, at which point docking is complete. In experiments A–E, the docking operation was completed successfully within 60 s after the start of visual servoing. The longest time required for completion among the five successful docking experiments was 60 s in experiment E. In experiment F, in which the turbidity level was 11.2 FTU (60.6 ml/m³ of milk), the 3D

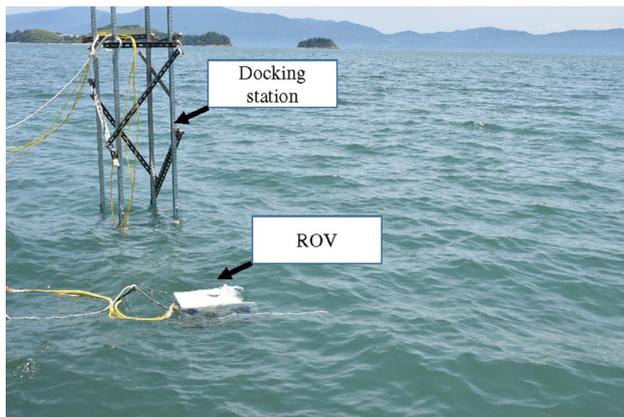


Fig. 27 ROV and a frame in which docking station has been installed. The docking station is shown in Fig. 28

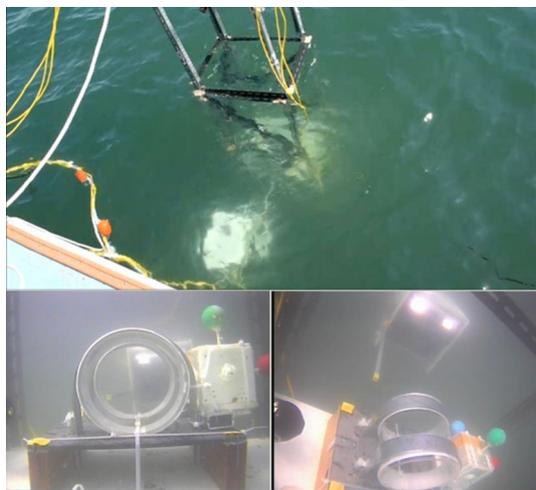


Fig. 28 Continuous iterative docking experiments in the sea. These photographs were taken by two underwater cameras installed in the docking station and from a pier

marker could not be recognized by the system, as shown in Fig. 26, because the fitness value was less than 0.2. Thus, docking was successfully executed under the turbidity levels in experiments A–E, and the system failed to recognize the target under the conditions of experiment F.

6 Continuous iterative docking experiment in the sea

Continuous iterative at-sea docking trials were conducted near Ushimado, Japan, as shown in Fig. 27. The docking station (length 600 mm \times width 450 mm \times height 3000 mm) was oriented with its long sides perpendicular to the pier. Underwater cameras were installed in the docking station to observe the performance of the ROV during operation,

as shown in Fig. 28. Docking tests began with the vehicle at a distance of 1.5 m in front of the dock. A shallow sea area was selected as the docking area, because the high turbidity in a shallow region would allow the verification of the robustness of the proposed system against turbidity. The turbidity level measured by the turbidity sensor during the experiment was 7.7 FTU; as indicated by Table 1, a turbidity of approximately 7 FTU is within the control area. The turbidity was measured at the position of 600 mm in front of the 3D marker in the sea. The depth of the sea floor in the docking area is 2.1 m. Natural waves in the sea continued, while the experiments were conducted. The ROV was tethered to an onshore platform with a cable of 200 mm in length. To demonstrate the underwater battery recharging operation, a docking rod was attached to vehicle, and a docking hole affixed with a 3D marker was designed. The main task for the vehicle was to automatically insert the docking rod into the docking hole under visual-servoing control. First, the vehicle was guided to the dock by manual control until the 3D marker was in the field of view (at a distance of approximately 600 mm from the target). In the visual-servoing step, the vehicle took the desired pose for docking. When the vehicle stably achieved the position within an error of ± 30 mm in the image plane (y , z) for 165 ms, it began to insert the docking rod by gradually decreasing the distance between the vehicle and target in the x -direction until it reached 350 mm, as shown in Fig. 18. After the docking operation was complete, the vehicle returned to a distance of 600 mm from the target in the x -direction for the next docking iteration.

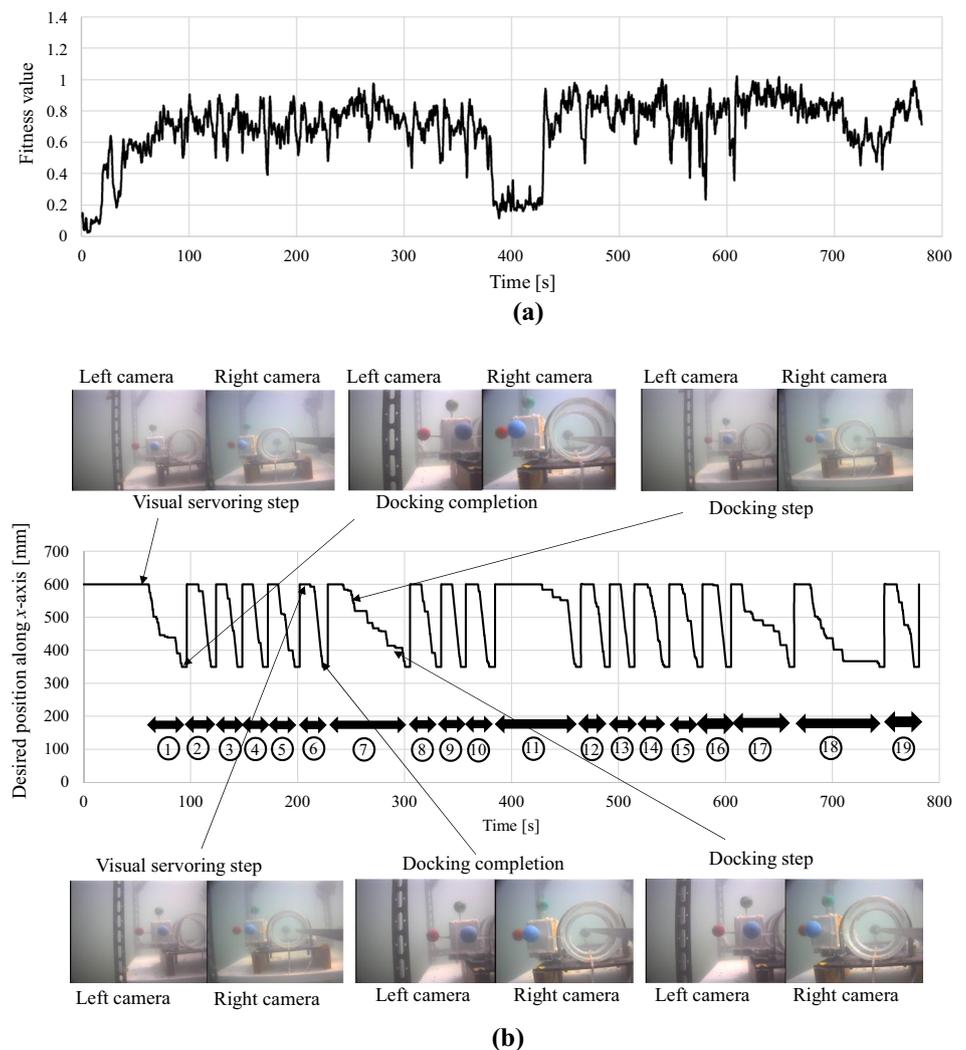
Continuous iterative docking was conducted successfully for 19 iterations. The fitness function and desired position in the x -direction in this experiment are shown in Fig. 29. Among the 19 iterations, docking iteration 3, which was one of the shortest docking operations, and docking iteration 7, which was one of the longest, were analyzed in detail; the results of these two iterations are shown in Figs. 30 and 31, respectively. Figure 30a, b, c–f shows the fitness function, the vehicle trajectory in 3D space, and the components of the recognized and desired poses, respectively, for docking iteration 3. The same results are shown in Fig. 31 for docking iteration 7. Docking iteration 3 was completed successfully within 30 s. In contrast, the completion of docking iteration 7 took more than 60 s. The position along the y -axis and the rotation about the z -axis fluctuated significantly, which delayed docking completion. This fluctuation seems to have been an effect of the waves. Therefore, the vehicle trajectory in docking iteration 7 (Fig. 31b) shows much larger variations than that of docking iteration 3 (Fig. 31b). As shown in Fig. 31c, there was a gap between the desired and estimated positions along the x -axis, because the error allowance for the docking operation is defined for only the

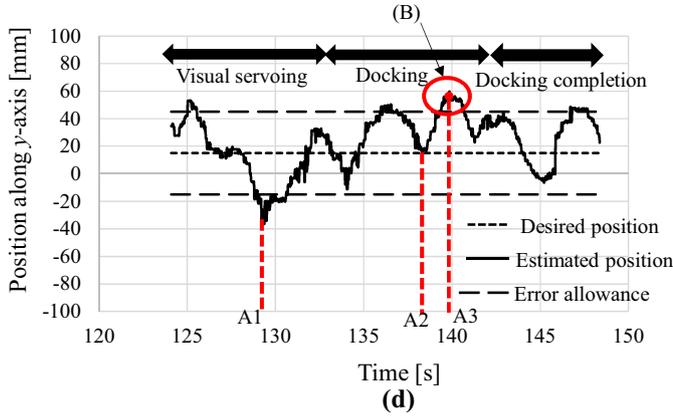
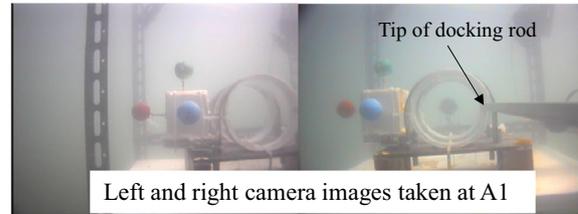
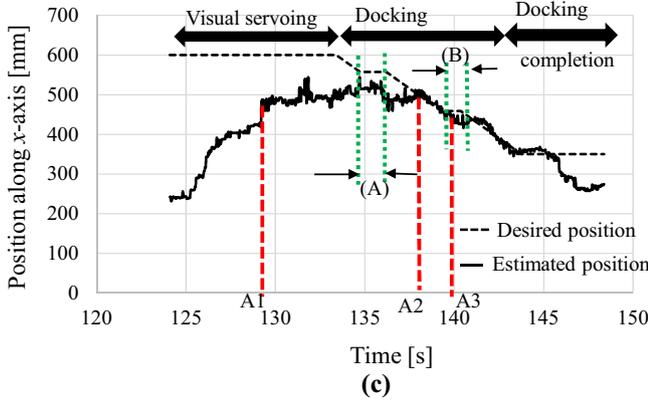
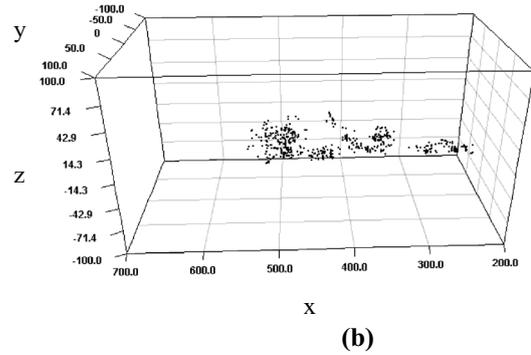
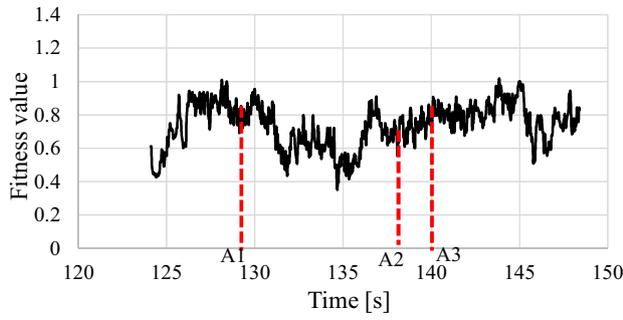
positions along the y - and z -axes and the rotation about the z -axis. Additionally, the desired position along the x -axis remained constant for some periods during the docking step because of some fluctuations in the position along the y -axis and especially the rotation about the z -axis that exceeded the error allowance, as shown in Fig. 31d, f. This condition triggers a switch from the docking step to the visual-servoing step, as shown by the path labeled “P” in Fig. 18.

During the undersea docking experiments, all data were stored for offline analysis. However, the left and right camera images were stored only up to docking iteration 7 because of limitations to the memory of the PC. As shown by the experimental results of the docking iterations, the docking operations conducted in the sea at turbidity levels below 7.7 FTU were executed successfully with good agreement between the analysis of the recognition accuracy in the pool under turbid conditions and the experimental docking results; the turbidity limit of 7.7 FTU agrees well with the set of

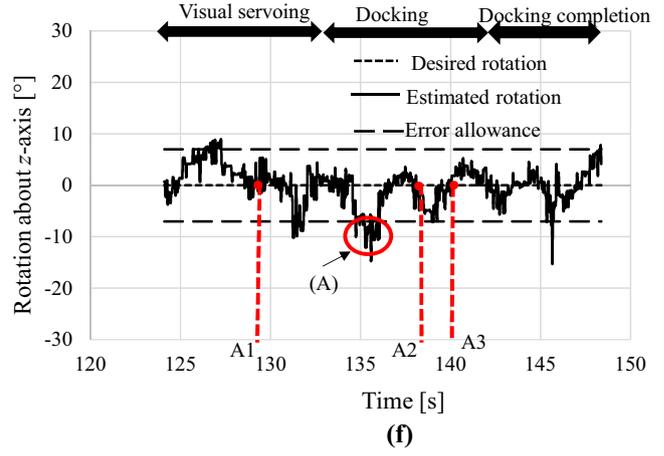
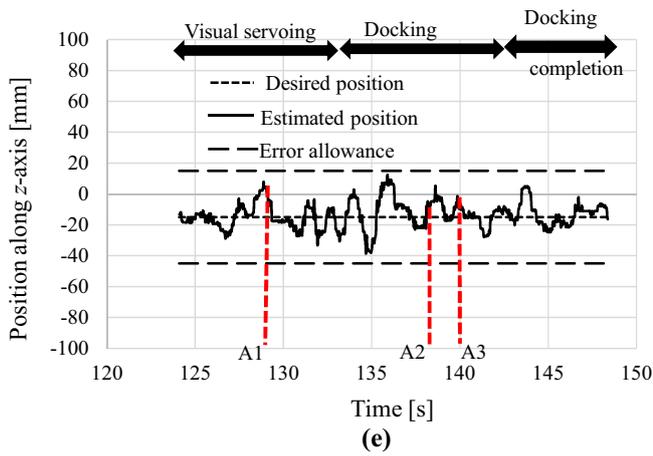
conditions labeled E in Table I. A comparison of the docking performance in the sea in docking iteration 7 with that in the pool in experiment E reveals that the docking period in the sea docking experiment was nearly twice that in the pool docking experiment and the fluctuation in the pose in the sea docking experiment, especially regarding the position along the y -axis and the rotation about the z -axis (Fig. 31d, f, was larger than that in the pool docking experiment (Fig. 25).

Fig. 29 Results of continuous iterative docking experiment. **a** Fitness value plotted against time. **b** Desired position in the x -direction during 19 docking iterations in the sea. The numbers along the bottom of the plot represent the docking iteration number, and the duration of each docking iteration is represented by the length of the corresponding arrow. Examples of the left and right camera images taken during the visual-servoing and docking steps and after docking completion are shown above and below the plot. Detailed results for docking iterations 3 and 7 are presented in Figs. 30 and 31, respectively





(g)



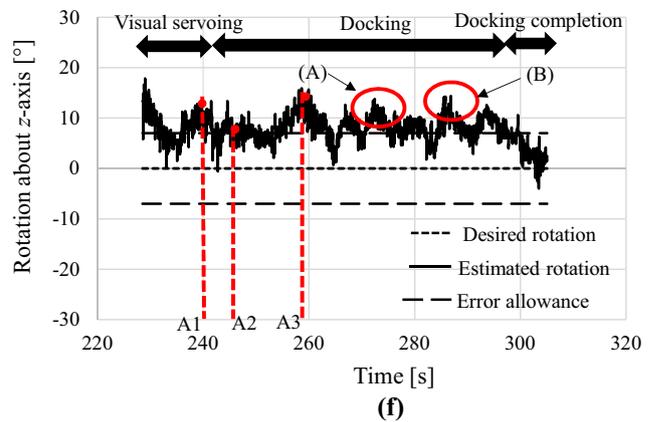
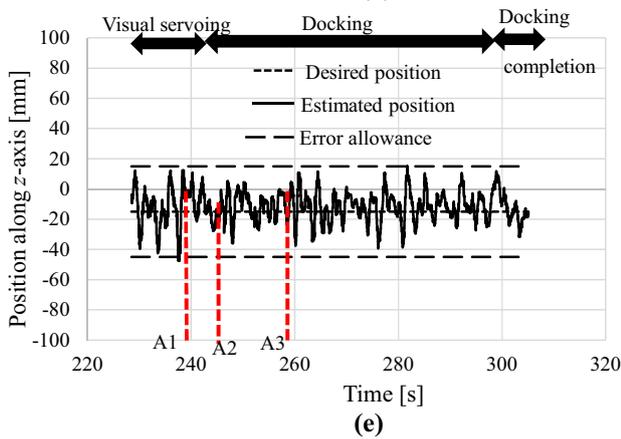
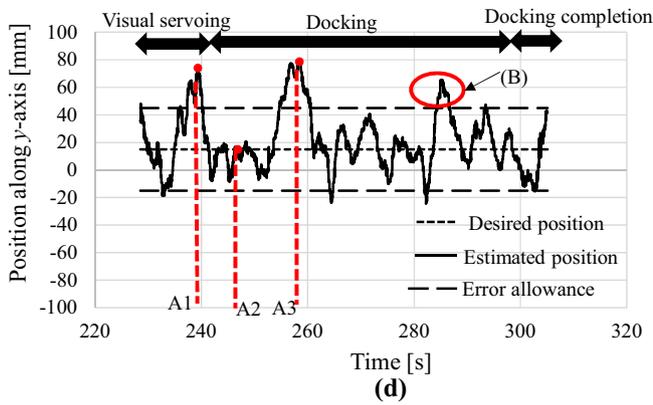
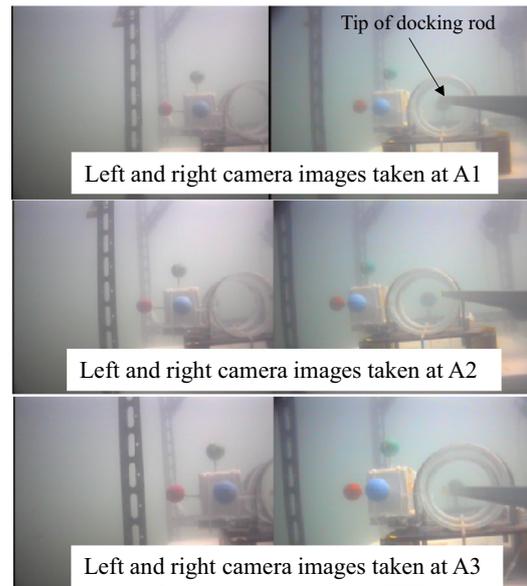
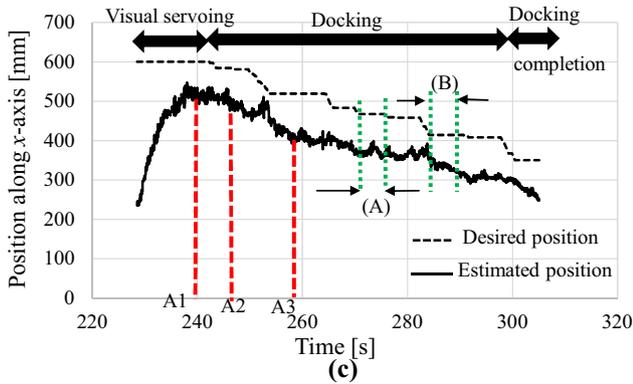
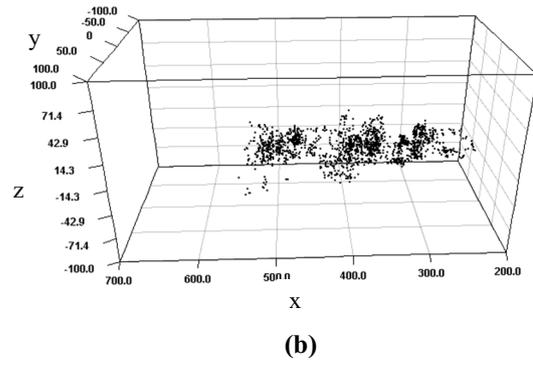
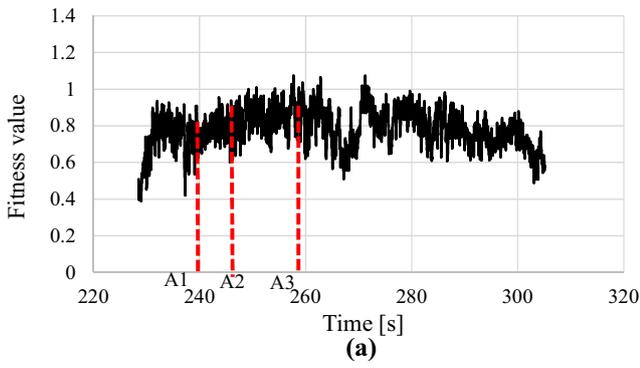


Fig. 31 Same as Fig. 30 for docking iteration 7. **a** Fitness value plotted against time. **b** Vehicle trajectory in 3D space. **c–f** Recognized position along the x -, y -, and z -axes and rotation about the z -axis obtained by the RM-GA. The desired position along the x -axis remained constant for the periods labeled (A) and (B) in **c**, because the rotation error about the z -axis labeled (A) in **e** and both the position error along the y -axis and the rotation error about the z -axis labeled (B) in **d** and **e**, respectively, surpassed the error allowance. At the time labeled A2 in the time profiles, the ROV is at the desired position along the y -axis, and the rotation angle about the z -axis is within the error allowance. At A1 and A3, the position along the y -axis and the rotation about the z -axis both surpassed the error allowance. This indicates that the rotation about the z -axis and the position along the y -axis are coupled. Therefore, the tip of docking rod appears to be within the allowed area in the images taken at A1 and A3 even though there are some deviations in the position along the y -axis and the rotation about the z -axis

Therefore, the turbidity tolerance described in Table 1 for the proposed system in a pool environment was verified experimentally in a real sea environment. The control and recognition areas (areas I and II in Table 1) can be expanded by improving the system in future.

7 Conclusion

In this work, the turbidity tolerance of a real-time stereo-vision-based pose estimation system using RM-GA was verified. The critical problem of determining how and why the RM-GA is suited to estimate the pose in real time in a changing environment, especially in terms of turbidity, was discussed in this paper. Experiments were conducted in different environments including the sea. The performance of the 3D pose estimation system under different turbidity levels was analyzed, and the turbidity tolerance of the system was examined experimentally. In future work, the system will be developed to be able to work under higher turbidity levels exceeding 10 FTU and in day- and nighttime operations.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number JP16K06183. The authors would like to thank Monbukagakusho; Mitsui Engineering and Shipbuilding Co., Ltd.; and Kowa Corporation for their collaboration and support for this study. The authors would also like to express their thanks to Dr. Yuya Nishida, Prof. Toshihiko Maki, and Prof. Tamaki Ura for their help and support.

References

- Ridao P, Carreras M, Ribas D, Garcia R (2010) Visual inspection of hydroelectric dams using an autonomous underwater vehicle. *J Field Robot* 27(6):759–778
- Negahdaripour S, Firoozfam P (2006) An ROV stereo vision system for ship-hull inspection. *IEEE J Ocean Eng* 31(3):551–564
- Bingham B, Foley B, Singh H, Camilli R, Delaporta K, Eustice R, Mallios A, Mindell D, Roman C, Sakellariou D (2010) Robotic tools for deep water archaeology: surveying an ancient shipwreck with an autonomous underwater vehicle. *J Field Robot* 27(6):702–717
- Negre A, Pradalier C, Dunbabin M (2008) Robust vision-based underwater homing using self-similar landmarks. *J Field Robot* 25(6–7):360–77
- McEwen RS, Hobson BW, McBride L, Bellingham JG (2008) Docking control system for a 54-cm-diameter (21-in) AUV. *IEEE J Ocean Eng* 33(4):550–62
- Cowen S, Briest S, Dombrowski J (1997) Underwater docking of autonomous undersea vehicles using optical terminal guidance. *OCEANS'97 MTS/IEEE Conf Proc* 2:1143–1147
- Myint M, Yonemori K, Lwin KN, Yanou A, Minami M (2017) Dual-eyes vision-based Docking system for autonomous underwater vehicle: an approach and experiments. *J Intell Robot Syst*. <https://doi.org/10.1007/s10846-017-0703-6>
- Feezor MD, Sorrell FY, Blankinship PR, Bellingham JG (2001) Autonomous underwater vehicle homing/docking via electromagnetic guidance. *IEEE J Ocean Eng* 26(4):515–21
- Teo K, Goh B, Chai OK (2015) Fuzzy docking guidance using augmented navigation system on an AUV. *IEEE J Ocean Eng* 40(2):349–61
- Palomeras N, Penalver A, Massot-Campos M, Vallicrosa G, Negre PL, Fernandez JJ, Ridao P, Sanz PJ, Oliver-Codina G, Palomer A (2014) I-AUV docking and intervention in a subsea panel. In: *Intelligent robots and systems (IROS 2014)*, 2014 IEEE/RSJ international conference, pp 2279–2285
- Batista P, Silvestre C, Oliveira P, (2012) A two-step control strategy for docking of autonomous underwater vehicles. In: *American control conference (ACC)*, pp 5395–5400
- White KA, Smith SM, Ganesan K, Kronen D, Rae GJ, Langenbach RM (1996) Performance results of a fuzzy behavioral altitude flight controller and rendezvous and docking of an autonomous underwater vehicles with fuzzy control. In: *Autonomous underwater vehicle technology, 1996 AUV'96*, proceedings of the 1996 symposium, pp 17–124
- Park JY, Jun BH, Lee PM, Lee FY, Oh JH (2007) Experiment on underwater docking of an autonomous underwater vehicle ISiMI' using optical terminal guidance. In: *OCEANS 2007, Europe*, pp 1–6
- Park JY, Jun BH, Lee PM, Oh J (2009) Experiments on vision guided docking of an autonomous underwater vehicle using one camera. *Ocean Eng* 36(1):48–61
- Maki T, Shiroku R, Sato Y, Matsuda T, Sakamaki T, Ura T (2013) Docking method for hovering type AUVs by acoustic and visual positioning. In: *Underwater technology symposium (UT)*, 2013 IEEE international, pp 1–6
- Palomeras N, Ridao P, Ribas D, Vallicrosa G (2014) Autonomous I-AUV docking for fixed-base manipulation. *IFAC Proc* 47(3):12160–5
- Garcia R, Gracias N (2011) Detection of interest points in turbid underwater images. In: *OCEANS, 2011 IEEE-Spain*, pp 1–9
- Codevilla F, Gaya JD, Duarte N, Botelho S (2004) Achieving turbidity robustness on underwater images local feature detection. *Int J Comput Vis* 60(2):91–110
- Roser M, Dunbabin M, Geiger A (2014) Simultaneous underwater visibility assessment, enhancement and improved stereo. In: *Robotics and automation (ICRA)*, 2014 IEEE international conference, pp 3840–3847
- Toka V, Sankaramurthy NH, Kini RP, Avanigadda PK, Kar S (2016) A fast method of fog and haze removal. In: *Acoustics, speech and signal processing (ICASSP)*, 2016 IEEE international conference on 20 Mar 2016, pp 1224–1228

21. Negru M, Nedeveschi S (2013) Image based fog detection and visibility estimation for driving assistance systems. In: Intelligent computer communication and processing (ICCP), 2013 IEEE international conference on 5 Sept 2013, pp 163–168
22. Tan RT (2008) Visibility in bad weather from a single image. In: Computer vision and pattern recognition. CVPR 2008. IEEE conference on 23 June 2008, pp 1–8
23. Satat G, Tancik M, Raskar R (2018) Towards photography through realistic fog. In: Computational photography (ICCP), 2018 IEEE international conference on 4 May 2018, pp 1–10
24. Myint M, Yonemori K, Yanou A, Ishiyama S, Minami M, (2015) Robustness of visual-servo against air bubble disturbance of underwater vehicle system using three-dimensional marker and dual-eye cameras. In: OCEANS' 15 MTS/IEEE Washington, pp 1–8
25. Myint M, Yonemori K, Yanou A, Minami M, Ishiyama S (2015) Visual-servo-based autonomous docking system for underwater vehicle using dual-eyes camera 3D-pose tracking. In: System integration (SII), 2015 IEEE/SICE international symposium, pp 989–994
26. Myint M, Yonemori K, Yanou A, Lwin KN, Minami M, Ishiyama S (2016) Visual-based deep sea docking simulation of underwater vehicle using dual-eyes cameras with lighting adaptation. In: OCEANS 2016-Shanghai, pp 1–8
27. Myint M, Yonemori K, Yanou A, Lwin KN, Mukada N, Minami M, (2016) Dual-eyes visual-based sea docking for sea bottom battery recharging. In: OCEANS 2016 MTS/IEEE Monterey, pp 1–7
28. Li X, Nishida Y, Myint M, Yonemori K, Mukada N, Lwin KN, Takayuki M, Minami M (2017) Dual-eyes vision-based docking experiment of AUV for sea bottom battery recharging. In: OCEANS 2017 Aberdeen, pp 1–5
29. Myint M, Yonemori K, Yanou A, Lwin KN, Minami M, Ishiyama S (2016) Visual servoing for underwater vehicle using dual-eyes evolutionary real-time pose tracking. JRM 28(4):543–58
30. Chen HH, Wu CM (2002) Use of a numerical technique for reducing effects of environmental luminance and turbidity on underwater imaging. OCEANS'02 MTS/IEEE 4:2383–2389
31. Hutchinson S, Hager GD, Corke PI (1996) A tutorial on visual servo control. IEEE Trans Robot Autom 12(5):651–70
32. Song W, Minami M, Aoyagi S (2008) On-line stable evolutionary recognition based on unit quaternion representation by motion-feedforward compensation. Int J Intell Comput Med Sci Image Process 2(2):127–39
33. Minami M, Agbanhan J, Asakura T (2003) Evolutionary scene recognition and simultaneous position/orientation detection. Soft Comput Meas Inf Acquis. Springer, Berlin, Heidelberg, pp 178–207
34. Song W, Fujia Y, Minami M (2014) 3D visual servoing by feed-forward evolutionary recognition. J Adv Mech Des Syst Manuf 4(4):739–55