Visual Docking Against Bubble Noise With 3-D Perception Using Dual-Eye Cameras

Khin Nwe Lwin, Naoki Mukada, Myo Myint, Daiki Yamada, Akira Yanou, Takayuki Matsuno, Kazuhiro Saitou, Waichiro Godou, Tatsuya Sakamoto, and Mamoru Minami

***Abstract*—Recently, many studies have been performed worldwide to extend the persistence of underwater operations by autonomous underwater vehicles. Underwater battery recharging technology is one of the solutions even though challenges still remain. The docking function plays an important role not only in battery recharging but also in other advanced applications, such as intervention. Visual servoing in undersea environments inevitably encounters difficulties in recognizing the environment when captured images are disturbed by noise. This study describes the effective recognition performance and robustness against air bubble disturbances in images captured by a real-time position and orientation (pose) tracking and servoing system using stereo vision for a visual-servoing-type underwater vehicle. The recognition of the vehicle pose based on dynamic images captured by dual video cameras was performed by a real-time multistep genetic algorithm (RM-GA). In previous studies, the docking performance was investigated under the condition that there were no disturbances in the captured images that address image degradation. In this paper, the robustness of the RM-GA against air bubble disturbances was verified through visual servoing and docking experiments in a pool test to confirm that the system can continue to recognize the pose of the 3-D marker and can maintain the desired pose by visual servoing. Then, the effectiveness of the proposed system against real disturbances such as turbidity that may degrade the visibility of the system in the sea was confirmed by conducting the docking experiment in a real sea, having verified the practicality of the proposed method.**

***Index Terms*—Air bubble noises, dual-eye cameras, genetic algorithm (GA), underwater vehicle, visual servoing.**

I. Introduction

A NUMBER of studies have examined control and guidance systems for underwater vehicles [1]–[5]. Robust vision-based target identification and homing using self-similar landmarks, which enables robust target pose estimation using a single camera, has been proposed [1]. The pose of an autonomous underwater vehicle (AUV) measured using an electromagnetic homing system has also been investigated [2]. In [3], Girona 500 AUV was developed for inspection and intervention tasks for the seafloor survey. In that work, several sensors, such as sonar, GPS, pressure sensor, velocity, fiber optic gyro, and plural video cameras, are used to localize an object. But the cameras look at different targets, so these approaches do not realize parallactic nature. In the approach of [4], an autonomous docking system for intervention using sonar and video camera system was developed. After performing the docking operation, the manipulator turns valves. The autonomous underwater manipulator for intervention mission in the oceanic environment was developed by SAUVIM [5].

Recently, vision-based systems for underwater vehicles have been studied for different applications [6]–[17]. Even though the vision-based approach is restricted to short-range measurement, and it degrades in the presence of turbidity and bubbles made by waves, it has the potential for the detection of 3-D poses, which is essentially necessary for docking a vehicle precisely. Most studies on guidance techniques using cameras for homing or docking operations are based on monocular vision to acquire the distance between a target and a vehicle [6], [7]. The reliable pose estimation of underwater docking using a single camera through a scene invariant approach was proposed in [8]. However, the methodology in [8] is thought to be difficult for real-time pose estimation, and it has not been applied to docking. An autonomous underwater torpedo-type vehicle was optically guided by lights mounted around the entrance of a docking station using one camera in [9]. A docking method for hovering-type AUVs using one camera based on both acoustic and visual positioning was proposed in [10]. The merit of the monocular camera is that the configuration is simple, and processing time seems to be less than that of the multicameras unit. The disadvantage is that the precision of distance measurement

Manuscript received November 1, 2017; revised May 13, 2018; accepted September 14, 2018. This work was supported by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid Scientific Research (KAKENHI) under Grant JP16K06183. This work was presented in part at the IEEE OCEANS Conference, Shanghai, China, April 10-13, 2016 and at the IEEE OCEANS Conference, Monterey, CA, USA, September 19-22, 2. *(Corresponding author: Khin Nwe Lwin.)*

Associate Editor: N. Cruz.

K. N. Lwin is with the Department of Intelligent Mechanical System, Okayama University, Okayama 700-8530, Japan (e-mail: knlwin85@gmail.com).

N. Mukada is with the Disco Corporation, Tokyo 143-8580, Japan (e-mail: pola9zud@s.okayama-u.ac.jp).

M. Myint is with the Technological University, Thanlyin 11292, Myanmar (e-mail: ecmyomyint@gmail.com).

D. Yamada, T. Matsuno, and M. Minami are with the Graduate School of Natural Science and Technology, Division of Mechanical and Systems Engineering, Okayama University, Okayama 700-8530, Japan (e-mail: p7kw2h27@s.okayama-u.ac.jp; matsuno@cc.okayama-u.ac.jp; minami-m@cc.okayama-u.ac.jp).

A. Yanou is with the Department of Radiological Technology, Kawasaki College of Allied Health Professions, Kurashiki 701-0194, Japan, and also with the Faculty of Health Science and Technology, Kawasaki University of Medical Welfare, Kurashiki 701-0193, Japan (e-mail: yanou-a@mw.kawasaki-m.ac.jp).

K. Saitou, W. Godou, and T. Sakamoto are with the Ushimado Marine Institute, Faculty of Science, Okayama University, Okayama 701-4303, Japan (e-mail: sc19412@s.okayama-u.ac.jp; godo@uml.okayama-u.ac.jp; ryu@uml.okayama-u.ac.jp).

Digital Object Identifier 10.1109/JOE.2018.2871651

of the camera's depth direction is not enough for applications in which high homing accuracy is required.

To overcome the limitation of single-camera-based systems, there are some studies using two cameras [11]–[17]. Binocular vision was used in some of these studies to detect the position of the vehicle and dock with a station [11]–[14]. In [11], AUV docks autonomously into a funnel-shaped docking station and then a vehicle-mounted manipulator performs several given tasks. In spite of that, two cameras have been mounted on the vehicle, one looking forward to estimate the panel pose and the other pointing down to help the mounted manipulator conduct some tasks. In [12], even though two cameras were used, one was facing downward and the other was looking forward for the purpose of obstacle avoidance and docking. A vision system for autonomous guidance using two cameras to track separate artificial underwater landmarks was proposed in [13]. In another study, the vehicle position was estimated using two cameras and a sonar system [14]. In that approach, the position of the vehicle was calculated by combining data from Doppler sonar and charge-coupled device (CCD) cameras, which does not deal with the orientation for controlling the vehicle. In two of the studies mentioned above [11], [12], "stereo vision" was not used even though two cameras were used. On the other hand, a concept of "stereo vision" that authors conceive creates a method that has a 3-D space perception by having plural images of one target object, which enables the pose estimation to exploit parallactic nature.

There are some studies using stereo vision for underwater vehicle [15]–[17]. The merit of the stereo camera is that the space recognition is superior to the monocular camera. A stereo vision system for an underwater vehicle-manipulator system consisting of two mobile cameras that can pan, tilt, and slide independently to observe and measure the position of the target object was proposed in [15]. The approach in [15] to obtain visual information, by actively rocking the cameras, has not been used for real-time operations, such as docking. A vision system for automated ship-hull inspection, based on computing the necessary information for positioning, navigation, and mapping of the hull from stereo images was proposed in [16] and [17]. How the inherent complexities faced by a monocular system are resolved with binocular vision has been experimentally proved in [17]. Even though stereo vision was used aiming for underwater vehicle, the methodologies in [15]–[17] were not applied in the docking test to prove the functionality and practicality of their proposed methods.

Some studies have used a stereo vision technique for pose estimation, although they have not been applied to underwater vehicles [18]–[21]. In [18], the pose estimation problem in real time was solved using multiple cameras and the extended Kalman filter for a moving robot. The method in [18] was based on epipolar geometry obtained from multiple views projected to multiple cameras. A cloth recognition and handling system for unique cloths using stereo vision that has tolerance against different light conditions was proposed in [19]. In [20], 3-D evolutionary pose tracking using an eye-vergence function was verified through frequency response experiments. In [21], the 3-D pose of the object was estimated and tracked in real time using a mono and stereo camera. The methodology in [21] was not yet applied in underwater environment but applied to robot set on the ground. Since the above-mentioned studies have been applied on the ground, the robustness against unstructured environment could not be concluded for a practical docking test in the sea environment.

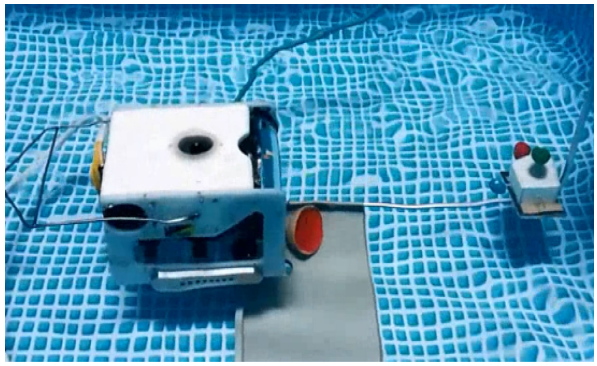

Fig. 1. Dual-eye visual servoing using 3-D marker and dual-eye cameras.

When a vehicle is used to explore an underwater environment using visual information, a number of disturbances in images, such as swirling mud from the bottom of the sea, fish, and plankton, can cause difficulties in pose estimation. Therefore, techniques are required for accurate recognition in water when disturbances occur in vision. There are some studies on image recognition with consideration of image degradation [22]–[25]. In [22] and [23], two cameras and three cameras, respectively, are used to increase the image recognition robustness by analyzing degradation factors in turbid water conditions. Several image recognition approaches for underwater robots to recognize a target object have been proposed in [24] and [25], where a ray tracing method was used to measure the 3-D shape of objects in consideration of refractive effects. In [25], noise was eliminated from a moving image to divide objects in the image into still backgrounds using an image processing technique. In that approach, epipolar constraints were used to search for corresponding points from a pair of cameras to measure the pose of the 3-D objects, which uses 2-D-to-3-D inverse reconstruction. The primary drawback of the inverse reconstruction is that 1) it is a problem to reconstruct 3-D pose from reduced information in 2-D spaces, an ill-posed problem; and 2) it is the effect of incorrectly mapping points in images, especially when images contain disturbances. The incorrect mapping causes pose estimation errors.

To extend the persistence time of an underwater operation, the dual-eyes vision-based docking approach avoiding the limitations of the above-mentioned studies was proposed in previous works [26]–[31], which have utilized a hovering-type remotely operated vehicle (ROV), as shown in Fig. 1. A 3-D pose-tracking system for docking using stereo vision was developed in [26] and was verified as follows. First, the stability of the proposed visual servoing regulation system was evaluated using abrupt disturbances. Second, underwater docking can be completed by switching between visual servoing and docking modes based on the error threshold. In a previous study, a deep-sea docking simulation experiment was conducted under a variable lighting environment [27]. Moreover, docking experiments were conducted

for undersea battery recharging application [28]. The 3-D-pose-tracking problem for estimating the pose of a 3-D marker was converted into an optimization problem, where the genetic algorithm (GA) was selected and utilized as a real-time multistep GA (RM-GA) for real-time pose estimation, since simple GA helps shorten optimizing calculation time for real-time image perception. In [29], the RM-GA performance was analyzed by choosing the best parameters by an evolutionary process, and the effectiveness of the RM-GA was evaluated by conducting an automatic docking experiment in an indoor pool. Experiments were conducted to verify the robustness of the proposed system against physical disturbances in different situations [30]. The turbidity tolerance of the proposed system was reported in [31], where the performance of the 3-D pose estimation was conducted under different turbidity levels in the pool to confirm the effectiveness of the proposed system. However, simulation of sea conditions in a pool is not enough to confirm the effectiveness and practicality of the proposed system against different disturbances. Therefore, in this study, the performance of the RM-GA was confirmed by conducting docking experiments in an environment disturbed by bubbles and in the sea with natural turbidity.

The rest of this paper is organized as follows. Section II presents the problem statement and contributions of this paper. Section III describes the proposed system for an underwater vehicle. The experimental results used to investigate the performance of the proposed system are presented, analyzed, and discussed in Section IV. Conclusions and areas for future research are presented in Section V.

II. Problem Statement and Contribution

A. Problem Statement

The deformation of images may appear in the water column at the bottom of the sea and also appear near the water surface when inspecting the hull of a ship, so the deformation caused by bubbles is a real obstacle for visual recognition in the sea. An additional effect of the bubbles is that the bubble disturbance induces upward fluid flow that affects the dynamical stability of the vehicle. As a result, bubble disturbance has the dual influences of degrading both image acquisition and dynamical stability. Therefore, construction of stable pose tracking and the stable control system in the presence of air bubble is important, and it may help improve the practical abilities of the ROV or AUV that could be useful in an actual sea environment. Then, the effectiveness of the proposed dual-eye control system should be evaluated from the viewpoints of practicality and functionality in natural environments by conducting docking experiments in the presence of turbidity in an actual sea.

B. Contribution

In previous works, the performance of the proposed system to estimate the vehicle's pose in real time was examined under the condition that images were not disturbed. Image deformation and occlusion are noise that hinders visual feedback control. Therefore, it is worth examining the performance of visual servoing in the presence of image degradation and occlusion. The stability of a visual-servo-type underwater vehicle system using a 3-D marker in the presence of air bubbles was investigated in an indoor pool in [32]. However, the effectiveness of the RM-GA when air bubbles disturbed pose estimation was not discussed in detail. In [33], the regulation performance in which the ROV maintains a desired pose with respect to a 3-D marker, even in the presence of air bubbles in front of the cameras, was evaluated. The 3-D pose-tracking performance was then checked while the 3-D marker was moving with a periodic surge motion having a sinusoidal position profile. Even though the proposed system was verified under air bubble disturbances, the docking performance against air bubble disturbances was not tested. The effectiveness and practicality of the real-time 3-D pose estimation system was not confirmed by conducting a sea docking experiment in previous studies. The main contributions of this paper are as follows.

1) Fitness distributions that RM-GA are needed to be optimized were scrutinized to clarify why the proposed RM-GA is tolerable against noise in images distorted by air bubbles, which is described in Section IV-A. The system is tolerable against noise because the pose-tracking problem is converted into an optimization problem, which enables the pose tracking to be independent of the height of the correlation function peak, as long as the true pose has the highest peak.
2) The tracking motion of the closed-loop dynamics of an ROV with real-time pose feedback when the deformation on the cameras images imposed by air bubbles was explored by analyzing the RM-GA's behavior with and without air bubble noise under different background sheet. The stability of the proposed system was confirmed, presented in Sections IV-B and IV-C.
3) The effectiveness of the proposed system was checked by conducting the docking experiment in a pool and an actual sea, described in Sections IV-D and IV-E.

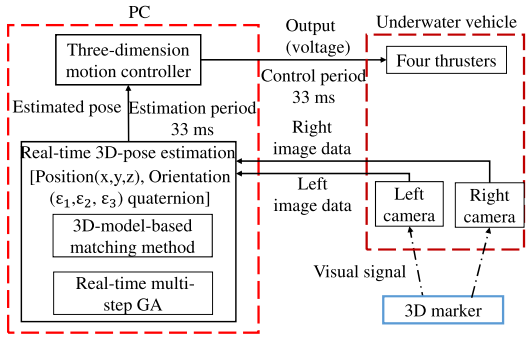

Fig. 2. Block diagram of dual-eye visual servo control system.

III. Proposed System

Fig. 2 shows a block diagram of dual-eye visual servoing for an underwater vehicle (see right-hand side of Fig. 2). A series of images are captured by dual-eye cameras mounted to the underwater vehicle and are then sent to a PC (see left-hand side of Fig. 2) through a cable. The relative pose of the vehicle is

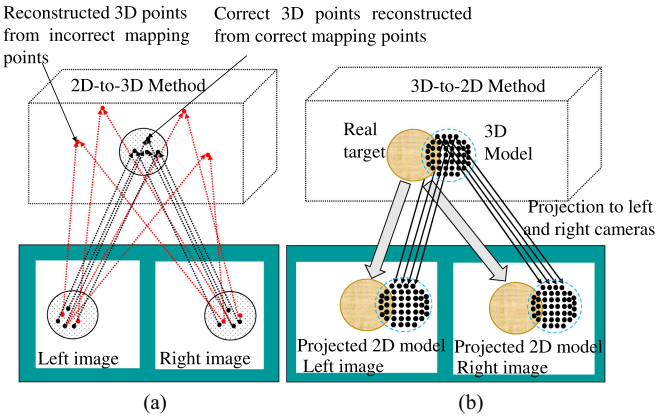

Fig. 3. (a) Correct and incorrect mapping in 2-D-to-3-D space and (b) pairing of points in 3-D-to-2-D projection.

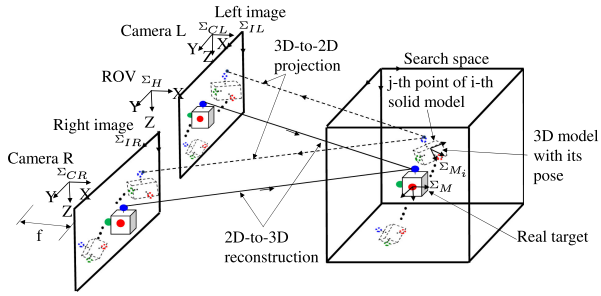

Fig. 4. Three-dimensional-model-based matching system with dual-eye cameras, including concepts of 3-D-to-2-D projection and 2-D-to-3-D reconstruction.

estimated in real time by the RM-GA, which is implemented in the PC. The estimated pose of the current condition of the underwater vehicle is entered into the 3-D motion controller as an input signal to adjust the thrust force of the robot. By eliminating the error between the estimated pose and the desired pose, the underwater vehicle is controlled to achieve a desired pose with respect to a 3-D marker.

A. Three-Dimensional-Model-Based Matching Method Using Dual-Eye Cameras and a 3-D Marker

The 3-D pose estimation method using dual-eye cameras and a 3-D marker was introduced and explained in [36]. In this section, the method is briefly discussed for the reader's convenience.

In conventional approaches for 3-D perception using plural cameras, recognition of objects and their relative pose is implemented by feature-based recognition using 2-D-to-3-D reconstruction calculations, in which the reconstructed shape of the 3-D target object is determined from a set of points in different images that are assumed to represent the same point on the target object surface in 3-D space, generally using triangulation and epipolar geometry. The main challenge in this type of approach is to ensure that points are correctly assigned to the same point on a 3-D object. If a point in one image is incorrectly mapped to a point in another image, the reconstructed point in 3-D space does not represent the real 3-D object. Fig. 3(a) shows the correct and incorrect reconstruction of a point through 2-D-to-3-D reconstruction, which has been recognized as an ill-posed problem, since expanded information in 3-D space needs to be reconstructed from degenerated image information in 2-D space.

A pose estimation approach [36] based on 3-D-to-2-D model projection was applied in this study because the forward projection from 3-D-to-2-D generates unique points in 2-D images without any errors, as shown in Fig. 3(b), meaning that the issue of incorrect mapping is avoided. With the 3-D-to-2-D approach, 3-D-model-based recognition was implemented in this study.

A 3-D-model-based matching method that uses the 3-D marker to estimate its pose in real time using stereo vision was first proposed in [34] and extended in [35]. Fig. 4 shows the 3-D marker coordinate system Σ_M, the ith model coordinate system Σ_{Mi}, 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 origins of Σ_M and Σ_{Mi} are the intersections of the three lines perpendicular to the faces to which the spheres are attached. The target is a 3-D marker that consists of three spheres 40 mm in diameter and colored red, green, and blue. The jth point on the ith model in 3-D space is projected onto the left and right camera images correctly, and these positions are calculated by computer using a camera projection geometry. The solid model of the real target object in space is projected naturally to the dual-eye camera images, and the 3-D models—the poses of which are given by an RM-GA gene—are projected from 3-D to 2-D. The correlation between the projected real 3-D marker and the projected model is calculated in 2-D space. A correlation function that considers the shape and color of the 3-D marker is used as a fitness function in the RM-GA evolution process (as described in detail in Section III-B). Multiple models that have the same 3-D marker information (color, shape, and size) with different poses are initially allocated randomly within the search space and are projected from 3-D space to a 2-D image plane. The projected models are then matched with the image captured by dual-eye cameras, including the projected real target in 2-D space. Finally, the best model with the maximum fitness value is selected to represent the measured relative pose.

B. Fitness Function

After the models whose poses are defined by genes in the RM-GA process are scattered in the 3-D search space, as shown in Figs. 4 and 7(b), the degree to which the pose of the individual 3-D models overlaps with the real 3-D target in 3-D space is evaluated in the 2-D space based on the correlation function used as a fitness function in the RM-GA. Refer to [34] and [35] for detailed descriptions of the derivation of a fitness function from a correlation function. The fitness function evaluates the correlation between an ith search model with its assumed pose $\boldsymbol{\phi}_i$ and the real 3-D marker in the left and right images. Note that the gene representing the pose of the ith model is expressed by $\boldsymbol{\phi}_i$. The intention of the designed fitness function is to have a dominant peak at the true pose of the target. The construction of the fitness function affects the optimum search performance and influences the RM-GA's convergence speed [29]. Fig. 5

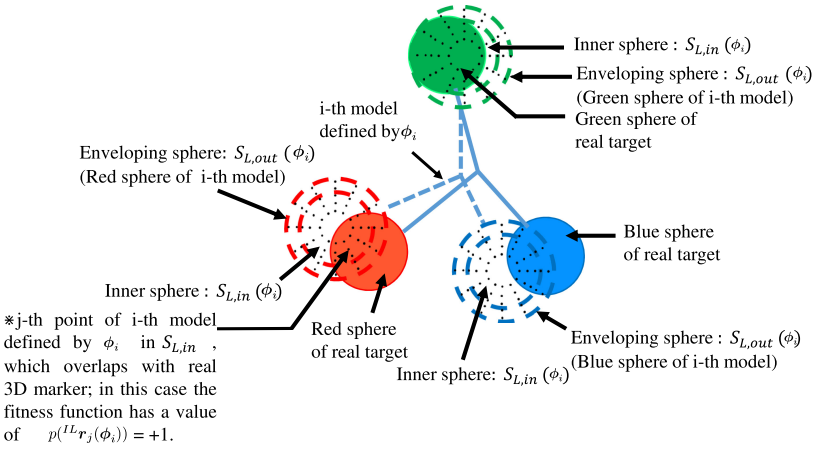

Fig. 5. Real 3-D marker and *i*th 3-D model projected to left camera's 2-D space.

shows the real 3-D marker projected onto the left image plane and the dotted circle that is the *i*th model obtained from the 3-D-to-2-D projection to the same plane. Each model consists of three spherical balls (red, green, and blue). Each ball of the model comprises an inner sphere $S_{L,\text{in}}$ and an enveloping sphere $S_{L,\text{out}}$, as shown in Fig. 5. The inner sphere of the red model $S_{L,\text{in}}$ is intended to evaluate the red ball area of the real target, and the enveloping sphere $S_{L,\text{out}}$ is for the background area. For example, if the *j*th point ${}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)$ on the red ball of an *i*th 3-D marker model defined by $\boldsymbol{\phi}_i$ in the left image is located in the inner sphere of the red ball of the real 3-D marker, as denoted by the explanation marked with "※" in the figure, then $p\left({}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\right) = +1$ in (1) and the fitness value of $F_{\text{red}}(\boldsymbol{\phi}_i)$ increases, whereas if the point in $S_{L,\text{in}}$ is not located in the ball area, $p\left({}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\right) = -1$, then the fitness value decreases. Concerning a point in $S_{L,\text{out}}$, if it does not overlap with the ball area, the fitness function increases by +1, and if it overlaps with the ball area, it decreases by -1. Therefore, when the real target and the model have identical poses, the fitness value reaches a maximum. In other words, the fitness function has been designed to have the highest peak only if the model's pose is identical to the pose of the real 3-D marker. Then, the pose of the model with the maximum fitness value represents the pose of the real target 3-D marker.

The fitness function in (1) was used in this study, as explained below. The fitness function for the red ball, $F_{\text{red}}(\boldsymbol{\phi}_i)$ with assumed pose $\boldsymbol{\phi}_i$ is calculated by averaging the fitness functions of both the left camera image $F_L(\boldsymbol{\phi}_i)$ and right camera image $F_R(\boldsymbol{\phi}_i)$. The summation in (1) is concerning the *j*th point ${}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)$, defined on the *i*th red 3-D model, whose pose is $\boldsymbol{\phi}_i$. In (1), *N* represents the number of *j*th points, where j = 1, 2, . . ., *N*

$$
\begin{aligned}
F_{\text{red}}(\boldsymbol{\phi}_i) &= \frac{1}{N}\left\{F_L(\boldsymbol{\phi}_i) + F_R(\boldsymbol{\phi}_i)\right\}/2 \\
&= \frac{1}{N}\left\{\left(\sum_{{}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\in S_{L,\text{in}}(\boldsymbol{\phi}_i)} p\left({}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\right)\right.\right. \\
&\quad \left.\left. + \sum_{{}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\in S_{L,\text{out}}(\boldsymbol{\phi}_i)} p\left({}^{\text{IL}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\right)\right)\right. \\
&\quad \left. + \left(\sum_{{}^{\text{IR}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\in S_{R,\text{in}}(\boldsymbol{\phi}_i)} p\left({}^{\text{IR}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\right)\right.\right. \\
&\quad \left.\left. + \sum_{{}^{\text{IR}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\in S_{R,\text{out}}(\boldsymbol{\phi}_i)} p\left({}^{\text{IR}}\boldsymbol{r}_j(\boldsymbol{\phi}_i)\right)\right)\right\} \Big/ 2. \qquad (1)
\end{aligned}
$$

The overall fitness function that evaluates all three colors is defined as $F(\boldsymbol{\phi}_i) = (F_{\text{red}}(\boldsymbol{\phi}_i) + F_{\text{green}}(\boldsymbol{\phi}_i) + F_{\text{blue}}(\boldsymbol{\phi}_i))/3$.

C. Real-Time Multistep GA (RM-GA)

Relative pose estimation using dynamic images should be optimized against a time-dependent and multipeak fitness distribution as fast as possible for making the closed-loop of visual servoing stable. Since the speed of the estimation is related to the time needed for the optimization calculation, not the iteration number, the performance should be evaluated using the convergence response measured in the time domain. However, most optimization methodologies have focused on accuracy and iteration number rather than calculation time. Instead of using other methods that could provide powerful accuracy but also increase the computational time, a simple optimization method (RM-GA) was selected for this study based on its simple logic and short calculation time, although it may not be the best in comparison to other optimization methods. A more detailed explanation of why RM-GA was used and how it works is provided in our previous work [36].

The pose of the model is expressed as a 72-b string to represent six pose parameters (x, y, z, ε_1, ε_2, and ε_3), as shown in Fig. 6. The first 36 b (12 b each for x, y, and z) represent the position coordinates of the position of the 3-D marker. The last 36 b (12 b each for ε_1, ε_2, and ε_3) describe the orientation defined by a quaternion.

Fig. 7(a) is a flowchart of the RM-GA process during convergence from the first to final generation using dynamic images input successively from video cameras. Fig. 7(b) is a graphical representation of chromosome evaluation and evolution during each 33-ms control period, and it illustrates how the RM-GA makes genes converge in successively input dynamic images.

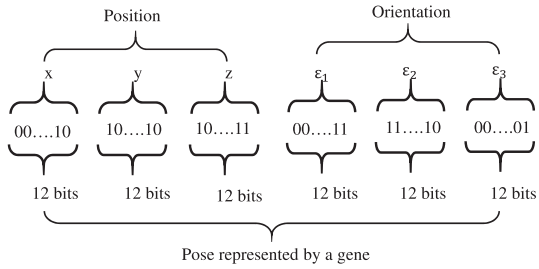

Fig. 6. Structure of RM-GA chromosome. Twelve bits each for x, y, and z represent the position coordinates of the 3-D model, and 12 b each for $\varepsilon_1, \varepsilon_2$, and ε_3 describe the orientation defined by a quaternion. The resolution for x and y are 0.20 mm, z is 0.10 mm, and ε_3 is 0.013°.

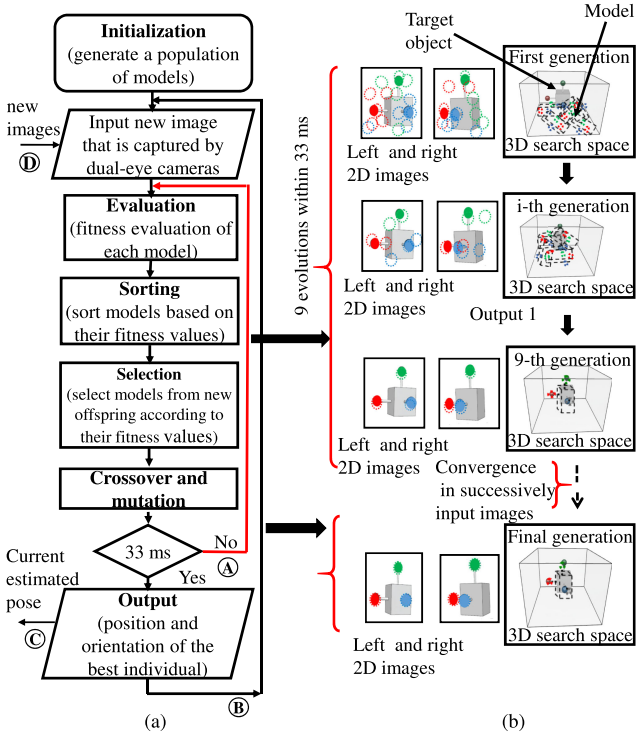

Fig. 7. Process flow in RM-GA. (a) Flowchart of process steps while the genes converge to the target 3-D marker from the first to the final generation and (b) graphical representation of solution evaluation and chromosome evolution during each 33-ms control period. The RM-GA converges to the solution in successively input dynamic images.

A random population is first generated, and a new pair of left and right images captured by the ROV are input. The appearance of the real target and the first generation of models in 3-D search space and the corresponding 2-D images of the target and models are shown at the top of Fig. 7(b). Each gene in the population defined by ϕ_i is evaluated using the fitness function $F(\phi_i)$. After the genes are ranked, the selection operator selects superior genes based on a ranking of the current generation, which are used for reproduction of new genes in the next generation. The selection operator uses an elitist preservation strategy to identify the best gene and carry it into the next generation as it is. That is, the model with the best pose is intended to be retained for the next evolving generation. A two-point crossover operator creates the next generation of models by randomly exchanging data between chromosomes at two bit locations. Then, the mutation operator randomly changes additional bits in the crossover

TABLE I
PARAMETERS OF THE REAL-TIME MULTISTEP GA

Parameter	Value
Number of genes	60
Evolved pose (position and orientation)	$(x, y, z, \varepsilon_1, \varepsilon_2, \varepsilon_3)$, where each gene is coded with 12 bits and $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are represented by a quaternion
Pose used for controlling (position, orientation)	Position (x, y, z) in millimeters Orientation (ε_3) around z-axis of Σ_H in Fig. 8
Search space defined by Σ_H in Fig. 8	{x,y,z,ε_3}={±400, ±400, ±200, ±0.15 (equal to ±17.3°)}
Control period	33 ms
Number of gene evolutions	9 times/33 ms
Selection rate	60%
Mutation rate	10%
Crossover	Two points
Evolution strategy	Elite preservation

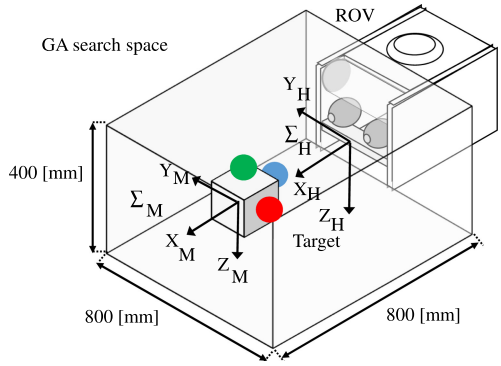

Fig. 8. GA search space and the vehicle coordinates Σ_H representing the pose of ROV.

operation. In the next generation, possibly superior genes with higher fitness values and better pose can be generated that more closely match the real target's pose than did the previous generation. Convergence of the genes to the maximum peak of the fitness distribution, which moves in the time domain as the input image changes, can be achieved through the RM-GA procedure. These genes evolve through successive input video images, and the fittest genes are transferred to the next generation to preserve the pose as a possible solution for evaluation in the next input images. This means that the fitness distribution in space ϕ_i is time-varying, and it becomes multipeak by image noise and other factors. Again, the convergence of the genes to a better solution with higher fitness value than the previous generation is continued within 33 ms, as shown by Ⓐ in Fig. 7(a). By performing this procedure repeatedly with a period of 33 ms, RM-GA searches the optimum solution that represents the real target's true pose, as shown at the bottom of Fig. 7(b), and the pose is output as shown by Ⓒ for visual feedback control; it is also shown by Ⓐ in the controller block diagram in Fig. 9. Although the pose of the target object is evaluated in 2-D space, convergence occurs in 3-D space. After 33 ms has passed, as shown by Ⓑ in Fig. 7(a), a new video image is input, as shown by Ⓓ. Table III-C presents the RM-GA parameters, and the RM-GA search space is shown in Fig. 8.

D. Three-Dimensional Motion Controller

The block diagram describing the control system is shown in Fig. 9. The four thrusters mounted on the ROV are controlled

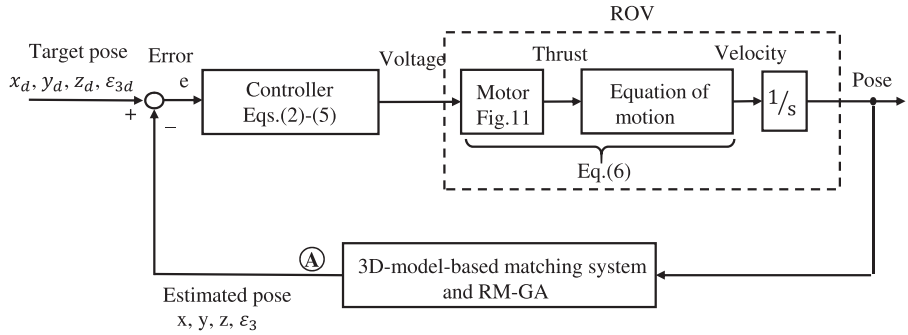

Fig. 9. Block diagram of visual servo control. Ⓐ in this figure represents output of pose estimated by RM-GA, which corresponds to Ⓒ in Fig. 7.

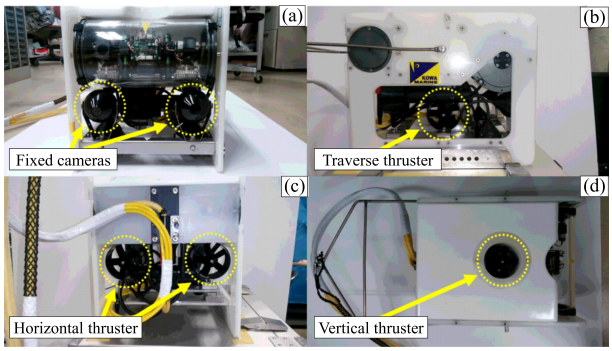

Fig. 10. Photographs of ROV. (a) Front view showing two cameras. (b) Side view showing traverse thruster. (c) Back view showing horizontal thrusters. (d) Top view showing vertical thruster. The height, width, and length were 310, 280, and 380 mm, respectively.

by sending the command voltage based on the feedback relative pose between the current pose (x, y, z, ε_3) of the ROV represented by Σ_H in Fig. 8 and the desired pose (x_d, y_d, z_d, ε_{3d}). In this study, the control of rotation around the x-axis and y-axis in Σ_H was ignored because the pose of the underwater robot (center of buoyancy to be located above the center of gravity) is passively restored to a stable state; that is, both rotation around x- and y-axes are reduced passively to zero.

$$\begin{aligned}\text{Surge direction}: v_x &= k_{p_1}(x_d - x) + 2.5 \\ &(v_x = 0\ [\text{V}] \text{ for thrust } 9.8\ [\text{N}] \\ &\text{in } X_H \text{ of } \Sigma_H \text{ in Fig. } 8 \\ &v_x = 5\ [\text{V}] \text{ for } -9.8\ [\text{N}]) \end{aligned} \tag{2}$$

$$\text{Sway direction}: \mathrm{v_y} = \begin{cases} 5\ [\text{V}]\ (y_d - y < -5\ [\text{mm}]) \\ \quad \text{for thrust } -4.9\ [\text{N}] \\ \quad \text{in } Y_H \text{ of } \Sigma_H \text{ in Fig. } 8 \\ 2.5\ [\text{V}] (-5 \le y_d - y \le 5) \\ \quad \text{thrust equals to zero} \\ 0\ [\text{V}]\ (y_d - y > 5\ [\text{mm}]) \\ \quad \text{for thrust } 4.9\ [\text{N}] \\ \quad \text{in } Y_H \text{ of } \Sigma_H \text{ in Fig. } 8 \end{cases} \tag{3}$$

$$\begin{aligned}\text{Heave direction}: v_z &= k_{p3}(z_d - z) + 2.5 \\ &(v_z = 0\ [\text{V}] \text{ for } -4.9\ [\text{N}] \\ &\text{in } Z_H \text{ of } \Sigma_H \text{ in Fig. } 8 \\ &v_z = 5\ [\text{V}] \text{ for } 4.9\ [\text{N}]) \end{aligned} \tag{4}$$

$$\begin{aligned}\text{Orientation around } Z_H \text{ axis}: v_{\varepsilon 3} &= k_{p2}(\varepsilon_{3d} - \varepsilon_3) + 2.5 \\ &(v_{\varepsilon 3} = 0\ [\text{V}] \text{ for } 0.882\ [\text{N}\cdot\text{m}] \\ &\text{in } Z_H \text{ of } \Sigma_H \text{ in Fig. } 8 \\ &v_{\varepsilon 3} = 5\ [\text{V}] \text{ for } -0.882\ [\text{N}\cdot\text{m}]). \end{aligned} \tag{5}$$

In (2)–(5), v_x, v_y, and v_z are the control voltages of the thrusters for the movement of the ROV in the x-, y-, and z-directions, respectively. Here, x_d, y_d, z_d, and ε_{3d} specify the desired relative pose between the vehicle and the target. Moreover, ε_{3d} is the orientation around the z-axis and is controlled by the value of $v_{\varepsilon 3}$. Since the horizontal thrusters shown in Fig. 10 are used not only for surge motion but also rotation around the z-axis of Σ_H depicted in Fig. 8, the voltage commands to the horizontal thrusters were determined so as to satisfy the requirements for surge and yaw motion. This thruster output calculation for decoupled motion was conducted by hardware logic in the ROV used in this experiment. This seems to be incorporated into the ROV's system because it was designed for human's manual operation, and the ROV has been modified into device enable computer control.

In the regulation performance experiment, the control voltage v_y for the y-direction was obtained by ON–OFF control, which was a given condition of the design of the ROV used in this study, and regulation of the other thrusters was obtained by the p controller. The gain coefficients were adjusted to achieve the best condition for visual servoing in the docking process based on preliminary experimental results. In this study, when approaching the docking station by thruster propulsion, the movement and yaw control need to be accurate, since the radius of the docking hole is 35 mm, as shown in Fig. 12. The dead band characteristics might cause oscillation of the vehicle when it is operated without linearization in the control system for visual servoing. Therefore, the dead band characteristics of

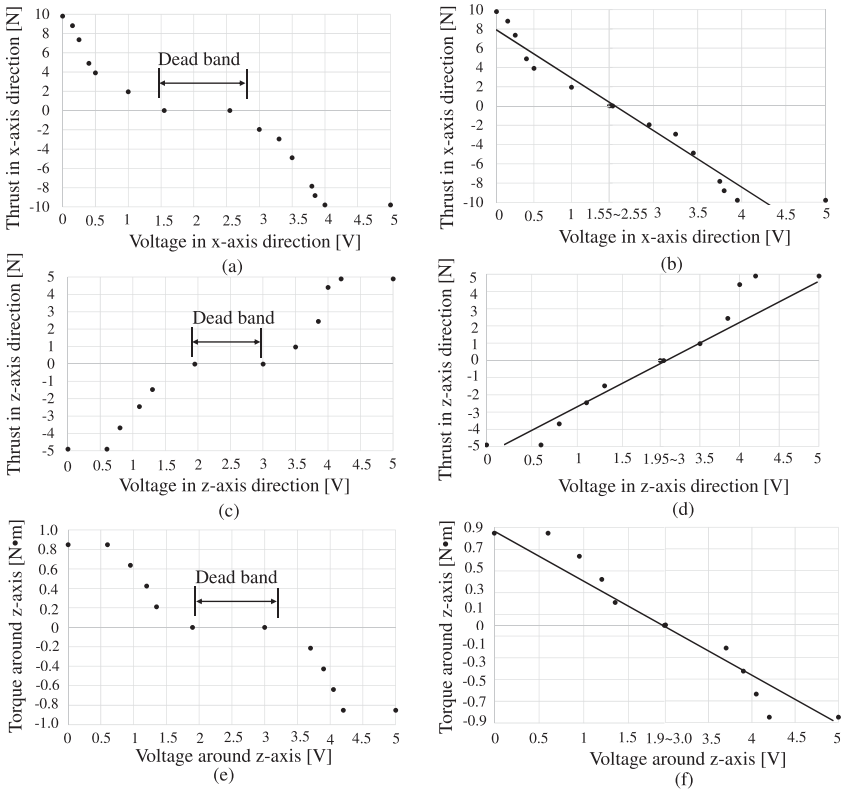

Fig. 11. Left column shows initial characteristics of thrust or torque against control voltage and the right column shows adjusted characteristics by removing the dead band and linearization. (a) Initial characteristics in x-axis direction. (b) Characteristics after removing dead band (black dots) and adjusting (solid line) in x-axis direction. (c) and (d) Initial and adjusted characteristics in z-axis direction. (e) and (f) Initial and adjusted characteristics around z-axis.

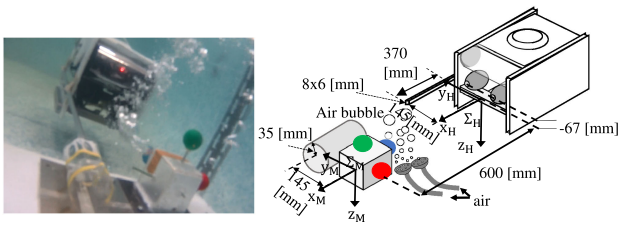

Fig. 12. Visual servoing experiment with bubble disturbances between the ROV and the 3-D marker, and layout of the docking experiment. The ROV is with a rod (8 mm × 6 mm) on its right-hand side, and the 3-D marker is with a docking hole having a radius of 35 mm on its left-hand side. The center distance between the docking hole and the 3-D marker was 145 mm. Two instruments were designed to intermittently generate bubbles in the water using air pumps.

the ROV that were confirmed in preliminary experiments in Fig. 11(a), (c), and (e) were eliminated, as shown by solid lines in Fig. 11(b), (d), and (f), respectively. However, the saturations were ignored in the linearized approximation, since pose of the visual servoing was confirmed to be stable by preliminary visual servoing experiments in a pool, in which the output voltages of v_x, v_y, v_z, and v_{ε_3} did not reach the fringe areas of 0 and 5 V during stabilizing and docking. For successful docking, the area of saturation should be avoided during visual servoing and docking, since the saturation character may increase oscillation of the vehicle, possibly leading to an unstable and uncontrollable state. The relation between voltage and induced velocity was derived experimentally using (6) as shown at the bottom of this page, including coupling effects. In (6), V_x, V_y, and V_z are the velocities of the x-, y-, and z-axis directions, respectively, in meter per second; $\omega_{\varepsilon 3}$ is the angular velocity of the orientation around the z-axis in radian per second; and v_x, v_y, v_z, and $v_{\varepsilon 3}$ are the output voltages in each direction that are calculated by the controller using (2)–(5).

E. Remotely Operated Vehicle and 3-D Marker

In this experiment, an ROV manufactured by KOWA Corporation, Fuji, Japan, is used, as shown in Figs. 10 and 12. The ROV was equipped with a compound eye visual sensor (imaging element: CCD, 380 000 pixels, signal system: NTSC,

$$\begin{bmatrix} V_x \\ V_y \\ V_z \\ \omega_{\varepsilon 3} \end{bmatrix} = \begin{bmatrix} 0.1483(\mathrm{m/s})/\mathrm{V} & 0.0030(\mathrm{m/s})/\mathrm{V} & 0 & 0 \\ 0.0148(\mathrm{m/s})/\mathrm{V} & 0.0303(\mathrm{m/s})/\mathrm{V} & 0 & 0.0442(\mathrm{m/s})/\mathrm{V} \\ 0 & 0 & -0.0690(\mathrm{m/s})/\mathrm{V} & 0 \\ 0 & 0 & 0.00020(\mathrm{rad/s})/\mathrm{V} & 0.0020(\mathrm{rad/s})/\mathrm{V} \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \\ v_{\varepsilon 3} \end{bmatrix}. \quad (6)$$

minimum illumination: 1.5 lx). In this experiment, the two front cameras were used simultaneously to perform 3-D object recognition. The distance between the two front cameras is 178 mm. Two horizontal thrusters, one vertical thruster, and one traverse thruster were mounted on the ROV, as shown in Fig. 10. The ROV height, width, and length were 310, 280, and 380 mm, respectively, and each thruster had a maximum thrust of 4.9 N. In addition, the ROV had two LED lights (5.8 W) for illumination. The ROV could operate at a maximum water depth of 50 m [36]. The 3-D marker is constructed from a box (100 mm $\times$ 100 mm $\times$ 100 mm) and red, green, and blue spheres (diameter: 40 mm). The coordinate of the 3-D marker and the vehicle is shown at the right-hand side of Fig. 12. A rod was attached to the right-hand side of the ROV and a docking hole was set on the left-hand side of the target, as shown in Fig. 12. The detailed explanations of the experiment layout are given in Section IV-D.

IV. Results and Discussion

A. Recognition Performance Against Air Bubble Disturbances

In this section, the recognition performance against air bubble disturbances with and without background sheet, and the RM-GA's behaviors while optimizing the fitness function are analyzed. The desired relative pose between the vehicle and the target is predefined. In this experiment, the population size of RM-GA is 60, and the input period of dynamical images 30 times a second is 33 ms. Then, the possible maximum number of evolutions is nine. The 3-D marker was fixed in the water with a relative pose of $x_t = 341$ mm, $y_t = 0$ mm, $z_t = -67$ mm, and $\varepsilon_{3t} = 0°$ in Σ_H depicted in Fig. 12. After 30 s from the beginning of the pose estimation experiment, air bubbles were emitted between the underwater robot camera and the 3-D marker. Moreover, recognition experiments of the 3-D marker with and without a background sheet of color patterns of a sea environment behind the 3-D marker were conducted.

Figs. 13 and 15 show the results for the case without the background sheet, and Figs. 14 and 16 show the results for the case with the background sheet. Figs. 13 and 14 show the distribution of the selected 60% (36 genes) genes for positions x, y, and z and orientation ε_3 around the z-axis in the evolution procedure, for the cases with and without the background sheet. In Fig. 13(a)–(d), although the recognition results of the position and orientation converge immediately after starting the experiment, bubble disturbances were confirmed to expand the gene recognition distribution, as compared to the previous generations before 30 s without the air bubble disturbance. The pose of the 3-D marker is difficult to recognize because the reflection of air bubble disturbance in the camera image. The variation of the gene recognition results in Fig. 13(a), as compared with Fig. 13(b) and (c), reveals the difficulty in estimating the position in the depth x-direction in case of bubble disturbance being imposed. The left and right camera images of the ROV are shown in Fig. 13(e) and (f) and Fig. 14(e) and (f), where Fig. 14(e) shows images 10 s from the start of the experiment and Fig. 14(f) shows images 40 s from the start of the experiment in the presence of air bubble disturbance.

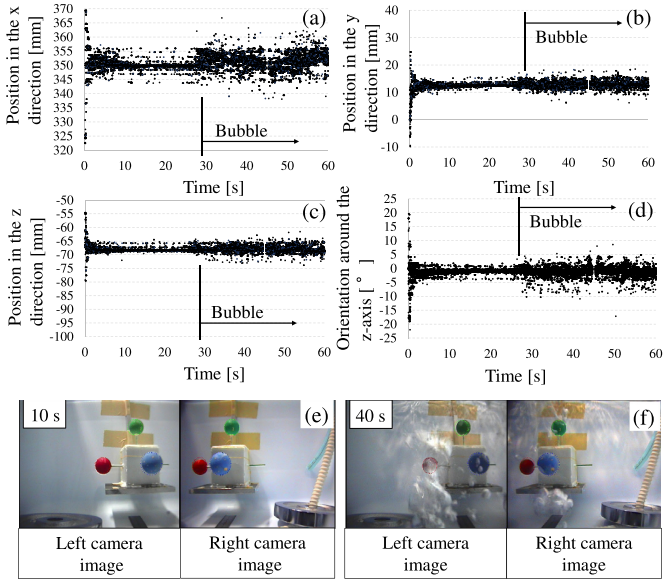

Fig. 13. Distribution of the top 36 genes in the case of a plain background. (a) Position in the x-direction. (b) Position in the y-direction. (c) Position in the z-direction. (d) Orientation around the z-axis. (e) Left and right camera images at 10 s from the beginning of the experiment. (f) Left and right camera images at 40 s from the beginning of the experiment in the presence of air bubble disturbance.

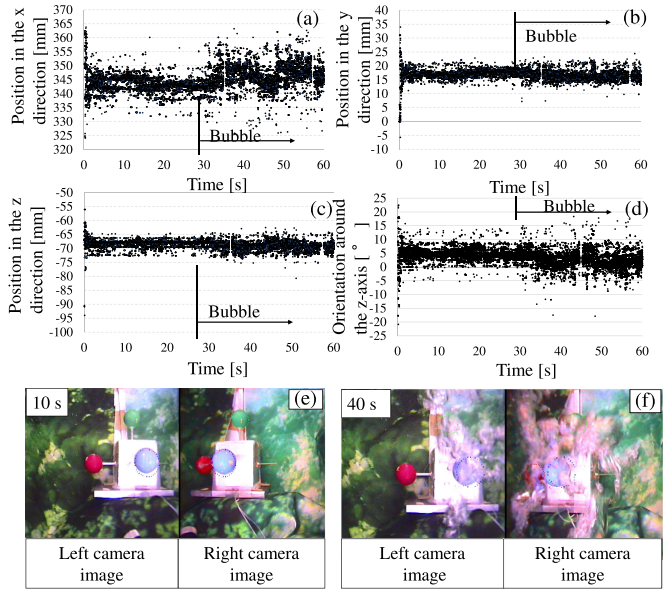

Fig. 14. Distribution of the top 36 genes in case of a simulated ocean background. (a) Position in the x-direction. (b) Position in the y-direction. (c) Position in the z-direction. (d) Orientation around the z-axis. (e) Left and right camera images at 10 s after the beginning of the experiment. (f) Left and right camera images at 40 s after the beginning of the experiment in the presence of air bubble disturbance.

Additional results for the case in which there is a background sheet simulating a real environment are shown in Fig. 14(a)–(d). These results were compared with those obtained for the case in which there was no background to confirm that the pose of the 3-D marker can be tracked in images input at the video frame rate. The experimental results shown in Figs. 13 and 14 reveal that the background sheet expands the variety of the distribution of the

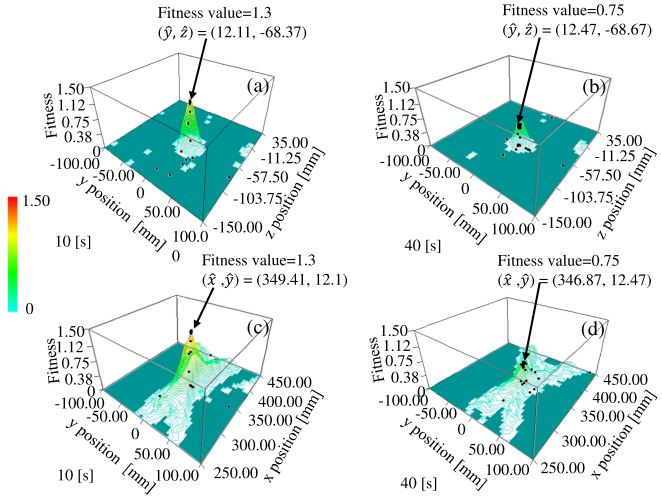

Fig. 15. Fitness distributions generated by scanning the assumed pose value in y–z plane and x–y plane with plain background at 10 and 40 s after the start of the experiment in the presence of air bubble disturbance. (a) Fitness distribution between the y and z positions at 10 s. (b) Fitness distribution between the y and z positions at 40 s. (c) Fitness distribution between the y and x positions at 10 s. (d) Fitness distribution between the y and x positions at 40 s.

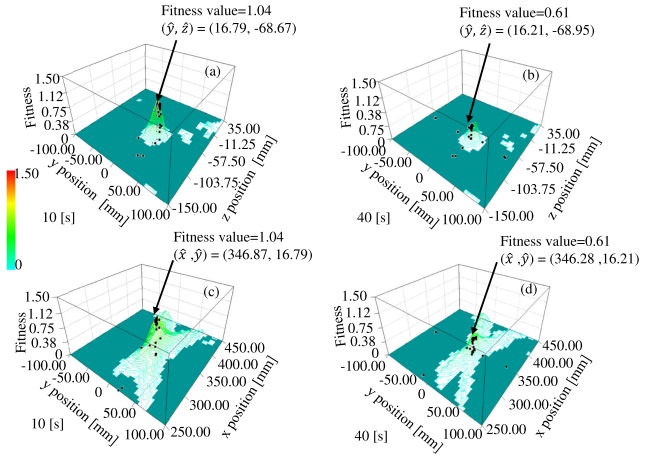

Fig. 16. Fitness distributions generated by scanning the assumed pose value in $y-z$ plane and $x-y$ plane with simulated ocean background sheet at 10 and 40 s after the beginning of the experiment in the presence of air bubbles. (a) Fitness distribution between the y and z positions at 10 s. (b) Fitness distribution between the y and z positions at 40 s. (c) Fitness distribution between the y and x positions at 10 s. (d) Fitness distribution between the y and x positions at 40 s.

genes, which causes the RM-GA to have difficulty optimizing the fitness function in real time.

The detected errors for x_e, y_e, z_e, and ε_{3e} that are the results of subtracting the pose estimated by the top gene for $\hat{x}$, $\hat{y}$, $\hat{z}$, and $\hat{\varepsilon_3}$ from the ground-truth measurement at a sample time of 10 s for the case of no air bubbles or background present are $x_e = x_t - \hat{x} = 341 - 350.20 = -9.20$ mm, $y_e = 0 - 12.11 = -12.11$ mm, $z_e = -67 - (-68.37) = 1.37$ mm, and $\varepsilon_{3e} = 0 - (-0.607) = 0.607°$.

Since the mean values of the x-, y-, and z-axes and orientation ε_3 among 36 genes are $\bar{x} = 349.90$, $\bar{y} = 12.44$, $\bar{z} = -68.38$, and $\bar{\varepsilon_3} = -0.653°$, then the mean errors are $x_e = x_t - \bar{x} = 341 - 349.90 = -8.90$ mm, $y_e = 0 - 12.44 = -12.44$ mm, $z_e = -67 - (-68.38) = 1.38$ mm, and $\varepsilon_{3e} = 0 - (-0.653) = 0.653°$. These values are reported in Table II, which also reports all statistics for the presence and absence of air bubbles and with and without the background sheet for sample times of 10 and 40 s. In this table, the maximum error was given by the value that maximizing the above subtracted results at the sample times, 10 and 40 s. The minimum error is given by minimizing the above subtracted results as a same manner. The standard deviations that are calculated from top genes at the sample times, 10 and 40 s in the case of air bubbles presence or absence for the case of with and without background were also analyzed.

TABLE II
POSITION RECOGNITION ACCURACY OF TOP GENES IN THE PRESENCE OR ABSENCE OF AIR BUBBLE DISTURBANCES FOR THE CASE OF WITH AND WITHOUT BACKGROUND SHEET

		Without background		With background	
		Without bubble (t=10s)	With bubble (t=40s)	Without bubble (t=10s)	With bubble (t=40s)
x-axis [mm]	Detected error, $x_t - \hat{x}$	-9.20	-12.32	-5.88	-5.27
	Mean error , $x_t - \bar{x}$	-8.90	-10.71	-4.42	-6.27
	Maximum error of $x_t - \hat{x}_i$	-8.22	-6.27	-0.99	-2.36
	Minimum error of $x_t - \hat{x}_i$	-9.39	-13.88	-6.66	-9.59
	Standard deviation	0.33	1.68	0.81	2.01
y-axis [mm]	Detected error, $y_t - \hat{y}$	-12.11	-12.30	-16.80	-16.21
	Mean error, $y_t - \bar{y}$	-12.44	12.69	-16.62	-16.63
	Maximum error of $y_t - \hat{y}_i$	-11.72	14.65	-14.65	-14.45
	Minimum error of $y_t - \hat{y}_i$	-13.28	10.94	-16.99	-18.95
	Standard deviation	0.24	0.86	0.41	0.94
z-axis [mm]	Detected error, $z_t - \hat{z}$	1.37	1.46	2.05	2.54
	Mean error, $z_t - \bar{z}$	1.38	1.11	1.06	2.23
	Maximum error of $z_t - \hat{z}_i$	1.46	1.86	2.44	3.51
	Minimum error of $z_t - \hat{z}_i$	1.27	-1.46	0.59	1.17
	Standard deviation	0.05	0.69	0.53	0.58
around z-axis [°]	Detected error, $\varepsilon_{3t} - \hat{\varepsilon}_3$	0.607	0.527	-2.911	0.046
	Mean error, $\varepsilon_{3t} - \bar{\varepsilon}_3$	0.653	0.787	-2.898	-0.813
	Maximum error of $\varepsilon_{3t} - \hat{\varepsilon}_{3i}$	0.894		-2.510	2.705
	Minimum error of $\varepsilon_{3t} - \hat{\varepsilon}_{3i}$	0.344	-0.264	-3.450	-4.459
	Standard deviation	0.163	0.589	0.256	1.296

1) $\hat{p}$ $(p = x, y, z, \varepsilon_3)$ represents p-position detected by top gene, and which is used for visual servoing feedback control.
2) $\bar{p} = (\sum_{i=1}^{36} p_i)/36 \quad (p = x, y, z, \varepsilon_3)$.
3) Maximum error is given by the value that maximizes $p_t - p_i (i = 1, 2, \ldots, 36)$, ($p_t$ represents ground-truth measured value, x_t, y_t, z_t, ε_{3t}).
4) Minimum error was given by the value that minimizes $p_t - p_i (i = 1, 2, \ldots, 36)$, ($p_t$ represents ground-truth measured value, x_t, y_t, z_t, ε_{3t}).
5) Unit of position value is [mm] and orientation is [°].

In the case of without background sheet, the standard deviation in x-axis in the case of without bubbles is 0.33 at the operation time of 10 s and the value of standard deviation in the case of with bubbles is 1.68 at the operation time of 40 s. By comparing these two standard deviation values, it can be confirmed that the air bubble disturbances have the variation of gene distributions expanded. Similarly, the position accuracy of top gene in other directions was analyzed and it was confirmed that the air bubble disturbances effected on the gene distribution, as given in Table II.

In the case of background sheet, the standard deviation in x-axis in the absence of air bubbles is 0.81 at the operation time of 10 s and the value of standard deviation in the presence of air bubbles is 2.01 at the operation time of 40 s. By comparing the standard deviation values in the case of with and without background sheet, the standard deviation value of with background is larger than without background. It means that the background sheet increases the expansion of

the genes. Similarly, the position recognition accuracy of top genes in other axes was analyzed as given in Table II and it was confirmed that the air bubbles and the background sheet make the RM-GA have difficulty in recognition of 3-D marker. However, the RM-GA could track the pose of the 3-D marker in real time even in the presence of air bubbles.

Why the system is tolerable against air bubble can be seen through Figs. 15 and 16. Figs. 15 and 16 represent the fitness distribution generated by each gene in the case of a plain background and a simulated ocean background. The fitness distributions between the y and z positions were evaluated 10 s after the start of the experiment and 40 s after the start of the experiment in the presence of air bubbles, as shown in Fig. 15(a) and (b). In Fig. 15(c) and (d), the fitness distributions between the y and x positions are also evaluated under the same conditions. When the fitness value is compared between the conditions with and without air bubble disturbance in Fig. 15(a)–(d), the fitness value of recognition without simulating air bubble disturbance was higher than that in under air bubble disturbance. According to the experimental results shown in Fig. 15, the fitness value is 1.3 and the recognized position is $y = 12.11$ mm and $z = -68.37$ mm for the condition of no background or bubbles at a sample time of 10 s. In the case without background and with bubbles at a sample time 40 s, the fitness value is 0.75 and the recognized position is $y = 12.47$ mm and $z = -68.67$ mm.

The results for the fitness distribution when the ocean background is placed behind the 3-D marker are shown in Fig. 16(a)–(d). In the case with the simulated ocean background and without bubbles at a sample time of 10 s, the fitness value is given by the top gene with the best fitness value of 1.04 and the recognized positions are $y = 16.79$ mm and $z = -67.87$ mm. The fitness value is 0.61 and the recognized positions are $y = 16.21$ mm and $z = -68.95$ mm in the case with both bubbles and background sheet, as shown in Fig. 16(b). The recognition accuracy of the highest gene with the best fitness value was found to be almost unchanged, regardless of the presence or absence of the background. Therefore, even though there was a disturbance of air bubbles and background sheet, only the height of the peak changed, but the estimated pose represented by the peak was maintained, as shown in Figs. 15 and 16. This is because the problem of pose estimation of 3-D marker is converted into an optimization problem. This confirmed that validity of finding the pose by optimization of the fitness distribution is preserved irrelevantly of background or bubble existence. In other words, the proposed RM-GA system has been confirmed to tolerate disturbance by air bubbles to some extent. These experimental results led to the sea docking trials that are presented in Section IV-E.

B. Regulation Performance Against Air Bubble Disturbances

The ROV is controlled to keep the desired relative pose ($x_d = 600$ mm, $y_d = 0$ mm, $z_d = -67$ mm, and $\varepsilon_{3d} = 0^\circ$) between the target and the ROV, predefined so that the ROV performs the relative pose regulation through visual servoing. The ROV is placed at the arbitrary start position in front of the 3-D marker about 600 mm in the x-direction. The ROV was kept to the desired pose by visual servoing after the 3-D marker had been recognized. The vehicle detects the target object, and then relative pose of the vehicle is estimated in real time through the model-based matching method and GA, i.e., RM-GA. The regulation performance without and with air bubble disturbances by visual servoing is shown in Figs. 17 and 18, respectively. The relative pose between the desired pose and estimated pose, that is, x, y, z, and orientation around the z-axis in Figs. 17 and 18, were measured by 3-D marker and RM-GA. From this point forward, all the relative poses presented in this paper were measured in the same way.

The regulation performance without air bubble disturbances, as shown in Fig. 17, was discussed in the previous work [36]. The previous work results were used to compare with the regulation performance with air bubble disturbances in this paper. According to Fig. 17(a), the fitness value exceeds 1 for a few seconds during recognition process and then remains above the value of 0.8. Fig. 18 shows the results in the presence of air bubbles in front of the 3-D marker. The fitness value shown in Fig. 18(a) exhibits greater variation than that shown in Fig. 17(a). In general, the minimum value of fitness function to be able to judge that RM-GA recognizes the 3-D marker well and estimates the relative pose between the ROV and 3-D marker was defined experimentally to be larger than 0.6.

Figs. 17(b)–(e) and 18(b)–(e) show the error between the current pose of the 3-D marker recognized by the RM-GA and the desired pose, and Figs. 17(g)–(j) and 18(g)–(j) show the thrust of each thruster required to restore the error. Figs. 17(f) and 18(f) show the trajectories of underwater robot based on Σ_H in Fig. 12 during the regulation process.

According to the experimental results in Figs. 17 and 18, the fitness function fluctuates and the thrust values oscillate. There seems to be three reasons why the fitness function has fluctuation. The first reason is the behavior of RM-GA, in which the top gene of RM-GA represents the pose that is close to the true pose. The gene that has the highest fitness value does not necessarily keep a constant height of the peak, then RM-GA's evolving behavior for optimization generates fluctuations. The second reason is that even though static environment is given for the recognition, the successively input dynamic images are not constant and not same, it seems to come from voltage fluctuation of CCD device unit. This is one additional reason while the solved pose by RM-GA fluctuates. The third reason is that the true pose is not constant in the dynamic images in the real case. The appearance of the dynamic images is changing due to the dynamic motion of vehicle and the environment changing. With these above discussions, the RM-GA fluctuating behaviors cannot be thought harmful thing but rather preferable nature of RM-GA to rapidly adapt for dynamic optimization in successively input dynamic images. Even though the desired pose is constant in regulating experiments, position error especially in y-direction in Figs. 17 and 18 oscillate with some amount. Consequently, the thrust value oscillates. This means that the closed loop dynamics of visual servoing may expand the ROV's oscillation. Therefore, the controller needs to be improved to expand leeway for keeping the closed loop dynamics within stable sphere enough away from unstable. While the authors have

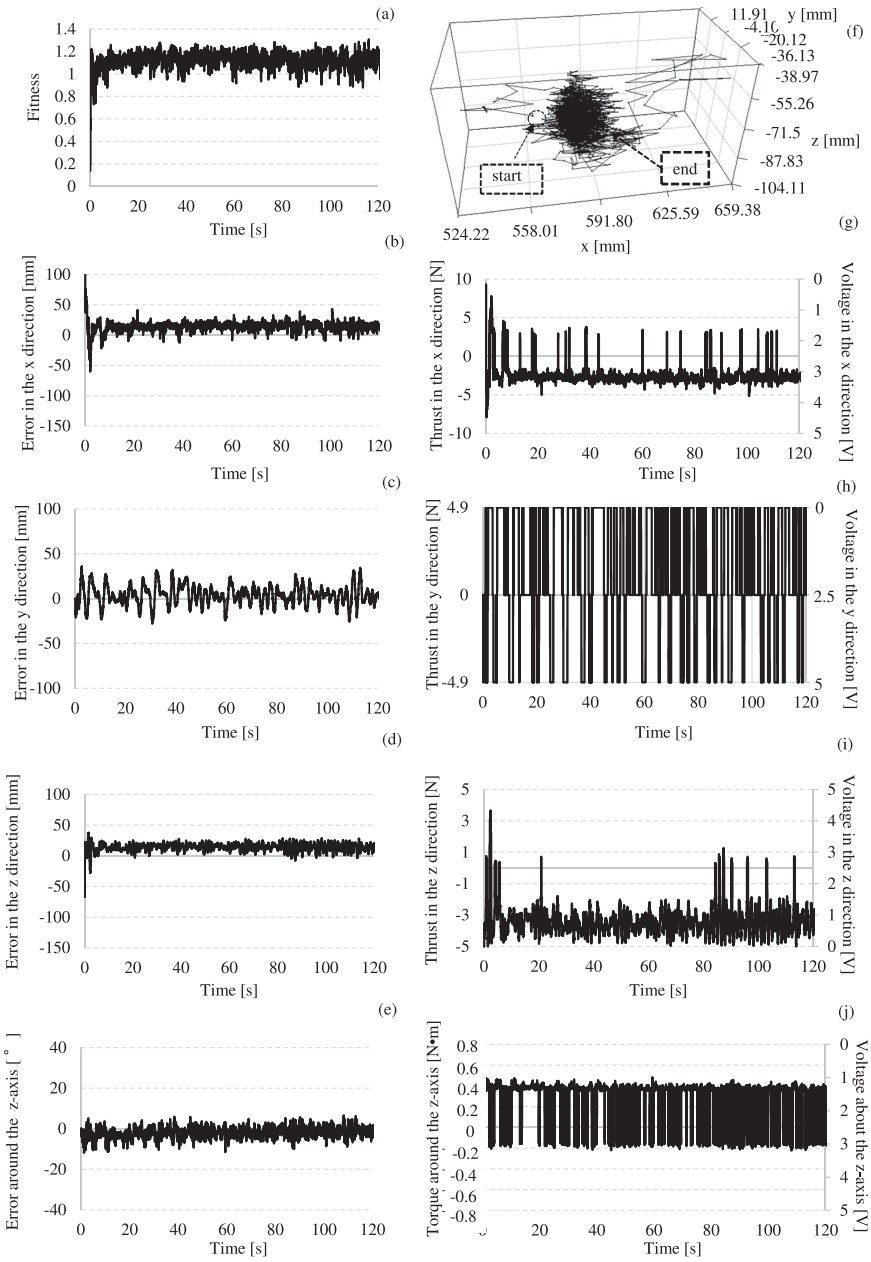

Fig. 17. Regulation performance of visual servoing by dual-eye image recognition with plain background without air bubble disturbance. (a) Fitness value. (b) Error in the x-direction. (c) Error in the y-direction. (d) Error in the z-direction. (e) Error around the z-axis. (f) 3-D trajectory of underwater vehicle. (g) Thrust in the x-direction. (h) Thrust in the y-direction. (i) Thrust in the z-direction. (j) Thrust around the z-axis.

proposed "RM-GA" as a 3-D perception method and verified its performance, what kind of real-time 3-D perception for visual servoing is best needs to be discussed in the future researches.

C. Regulation Performance for Periodic Motion of the 3-D Marker

To confirm the performance of the proposed system, the regulation motion of the underwater vehicle when 3-D marker is moving back and forth in front of the ROV was evaluated. The desired relative target pose was set to the same value as in Section IV-A. Fig. 19 shows a setup of visual servoing experiment with periodical motion of 3-D marker. Figs. 20–22 show the experimental results for the tracking performance without additional disturbances in front of the 3-D marker, whose motions had periods of 20, 15, and 10 s and an amplitude of 280 mm.

Figs. 20(a)–(d), 21(a)–(d), and 22(a)–(d) show the fitness value, the position in the x-direction, the tracking error in the x-direction, and the thrust in x-direction, respectively. The arrow labeled "(A)" in the figures indicates the motion of the 3D marker starting at 20 s and it also represents the duration of experiment in which the 3-D marker was in cyclic motion. The visual servoing has been started at 0 s with the 3-D marker being

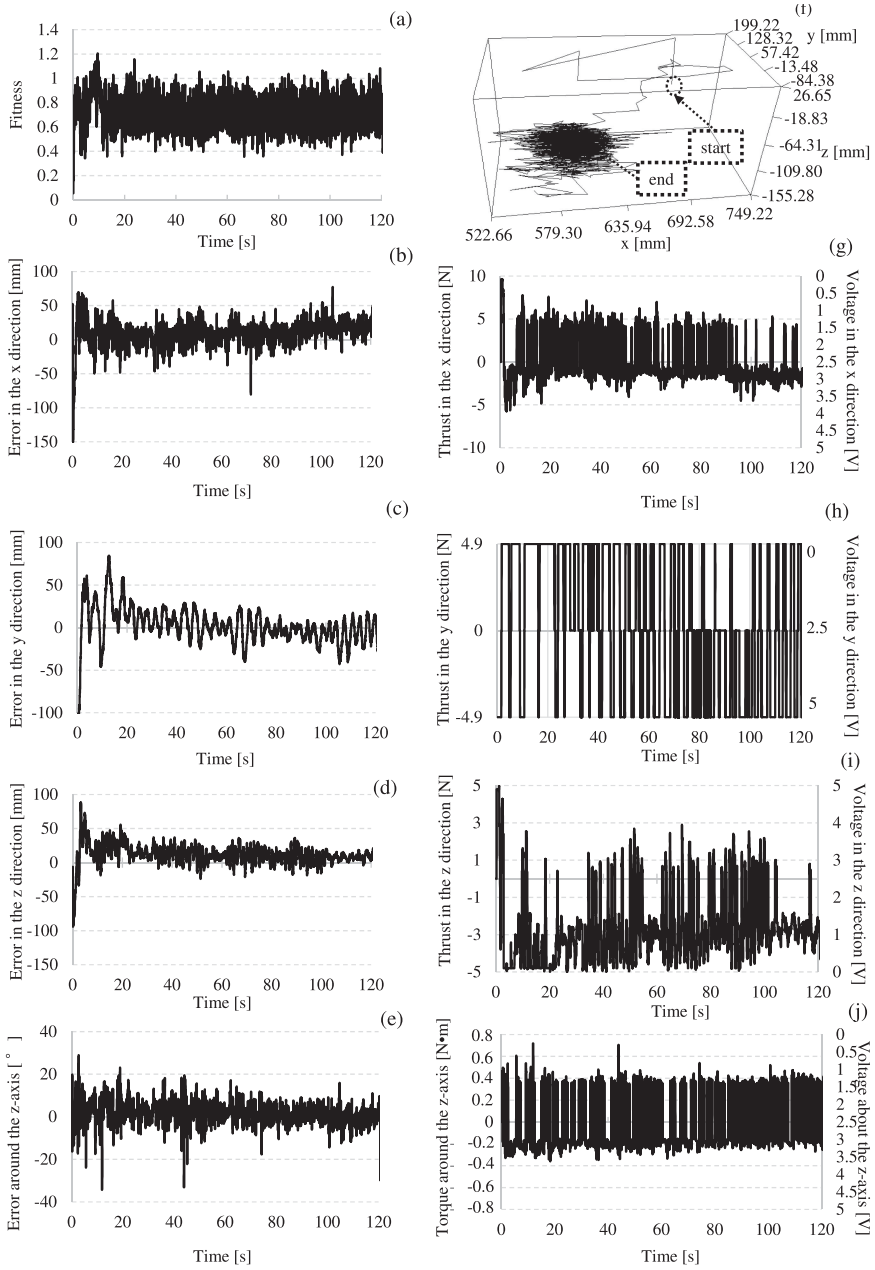

Fig. 18. Regulation performance of visual servoing by dual-eye image recognition with plain background with air bubble disturbance. (a) Fitness value. (b) Error in the x-direction. (c) Error in the y-direction. (d) Error in the z-direction. (e) Error about the z-axis. (f) 3-D trajectory of the underwater vehicle. (g) Thrust in the x-direction. (h) Thrust in the y-direction. (i) Thrust in the z-direction. (j) Torque about the z-axis.

stationary, and besides oscillatory motions of 3-D marker has begun at 20 s. The same conditions are given for Figs. 20–26. The tracking performance without air bubble disturbances for periods of 20, 15 and 10 s have the tracking errors, as shown in Figs. 20–22. In Fig. 20, the starting point of the ROV was set at the arbitrary point and the vehicle detects the target object and relative pose of the vehicle is estimated through the model-based matching method and RM-GA. Therefore, the adjusting behavior makes the fluctuation before the target starts moving at the 20 s in Fig. 20(a). But the fluctuation is not getting worse even though 3-D marker is moving. Since the tracking error in the x-direction for the period of 10 s in Fig. 22(c) is larger than that in Fig. 20(c), the thruster output was saturated in Fig. 22(d). The sign of the thruster output value in (d) is opposite the sign of the error value in (c). Since if $x_e = x_d - x < 0$, meaning the ROV position error is larger than x_d, the ROV is delayed and then the thruster for the x-axis needs to output with a positive sign. The motion period given to the marker is different, such as 20 s in Fig. 20, 15 s in Fig. 21, and 10 s in Fig. 22. The large tracking errors can be seen in Figs. 21 and 22 than Fig. 20. The motion delay may exist at the turning point of forward and backward moving of the 3-D marker.

Photographs taken every 10 s during the experiment in the case of a period of 20 s are shown in Fig. 23(a)–(e). The positional relationship between the ROV and the 3-D marker without additional disturbances is shown in Fig. 23 (left-hand column).

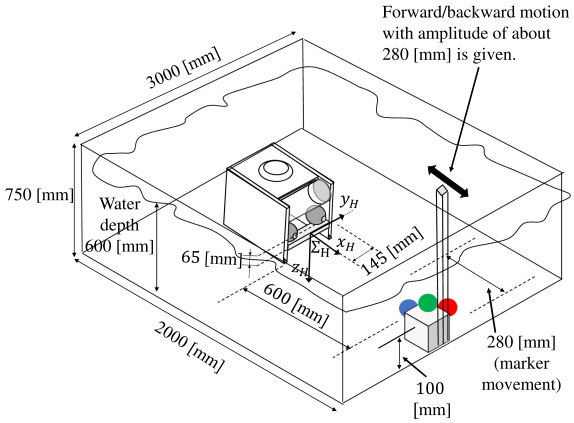

Fig. 19. Setup of visual servoing experiment with periodic motion of 3-D marker.

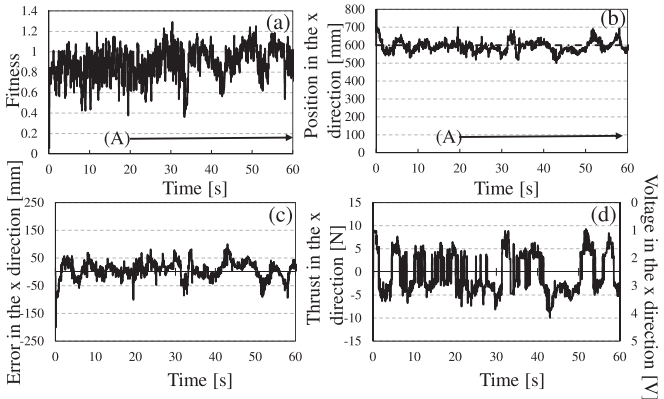

Fig. 20. Tracking performance of dual-eye image recognition with plain background without air bubble disturbance for the case in which "(A)" the 3-D marker moves in the x-direction with an amplitude of 280 mm and a period of 20 s after 20 s have passed from the start. (a) Fitness value. (b) Position of the underwater vehicle in the x-direction (dashed line is the desired position). (c) Tracking error in the x-direction. (d) Thrust in the x-direction. Photographs were taken every 10 s during this experiment and are shown in Fig. 23.

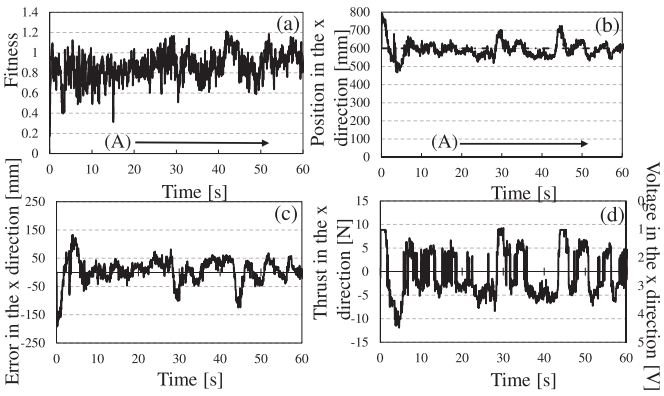

Fig. 21. Tracking performance of dual-eye image recognition with plain background without air bubble disturbance for the case in which "(A)" the 3-D marker moves in the x-direction with an amplitude of 280 mm and a period of 15 s after 20 s have passed from the start. (a) Fitness value. (b) Position of the underwater vehicle in the x-direction (dashed line is the desired position). (c) Tracking error in the x-direction. (d) Thrust in the x-direction.

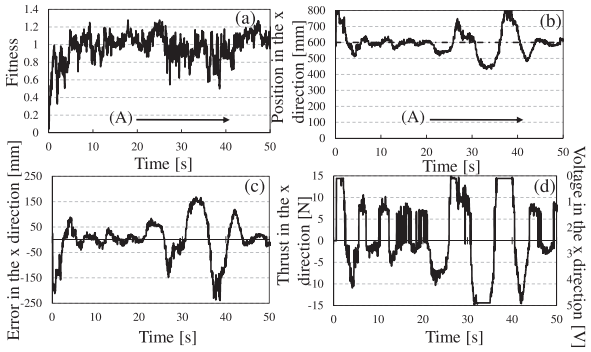

Fig. 22. Tracking performance of dual-eye image recognition with plain background without air bubble disturbance for the case in which "(A)" the 3-D marker moves in the x-direction with an amplitude of 280 mm and a period of 10 s after 20 s have passed from the start. (a) Fitness value. (b) Position of the underwater vehicle in the x-direction (dashed line is the desired position). (c) Tracking error in the x-direction. (d) Thrust in the x-direction.

In each image in the right-hand column of Fig. 23(a)–(e), the dotted circle represents the position and orientation of the 3-D marker as recognized by the RM-GA in real time. If each sphere indicated by the dotted circles and the 3-D marker matches, then the system can accurately recognize the position and orientation of the 3-D marker.

The experimental results for the case of an air bubble disturbance in front of the 3-D marker are shown in Figs. 24–26 when the 3-D marker is undergoing periodic motion in the x-direction with periods of 20, 15, and 10 s and an amplitude of 280 mm. Figs. 24(a)–(d), 25(a)–(d), and 26(a)–(d) show the fitness value, position of the underwater vehicle in the x-direction, the error of relative target position, and the thrust force to regulate the vehicle for periods of 20, 15, and 10 s, respectively. In each figure, arrow "(A)" indicates that the movement of the 3-D marker in the x-direction with an amplitude of 280 mm and periods of 20, 15, and 10 s started at 20 s. Arrow "(B)" indicates the started time of the air bubble disturbance and also duration time. The dashed lines in Figs. 20(b), 21(b), 22(b), 24(b), 25(b), and 26(b) represent the desired position along the x-axis of Σ_H in Fig. 12.

Although the fitness value sometimes decreases to approximately 0.4 in the presence of air bubbles in front of the dual-eye cameras, as shown in Figs. 24–26(a), the regulation performance can be confirmed to be maintained, as shown in Figs. 24(b)–(d), 25(b)–(d), and 26(b)–(d). According to the experimental results, the vehicle can track the 3-D marker with some tracking errors due to the GA performance, vehicle dynamics, dynamics of the motor, and the control gains. However, it can be verified that the ROV could keep tracking the moving object without losing visual feedback even in the presence of air bubble disturbances.

The experimental results shown in Fig. 24 for a period of 20 s with air bubble disturbance correspond to the photographs taken every 10 s, as shown in Fig. 27(a)–(e). The left-hand column of Fig. 27 shows the positional relationship between the ROV and the 3-D marker, and the position of the 3-D marker as viewed from the ROV is shown in the right-hand column of Fig. 27.

In Fig. 27(d), it looks like that detection of the 3-D marker has been lost and this photo corresponds to 40 s of Fig. 24.

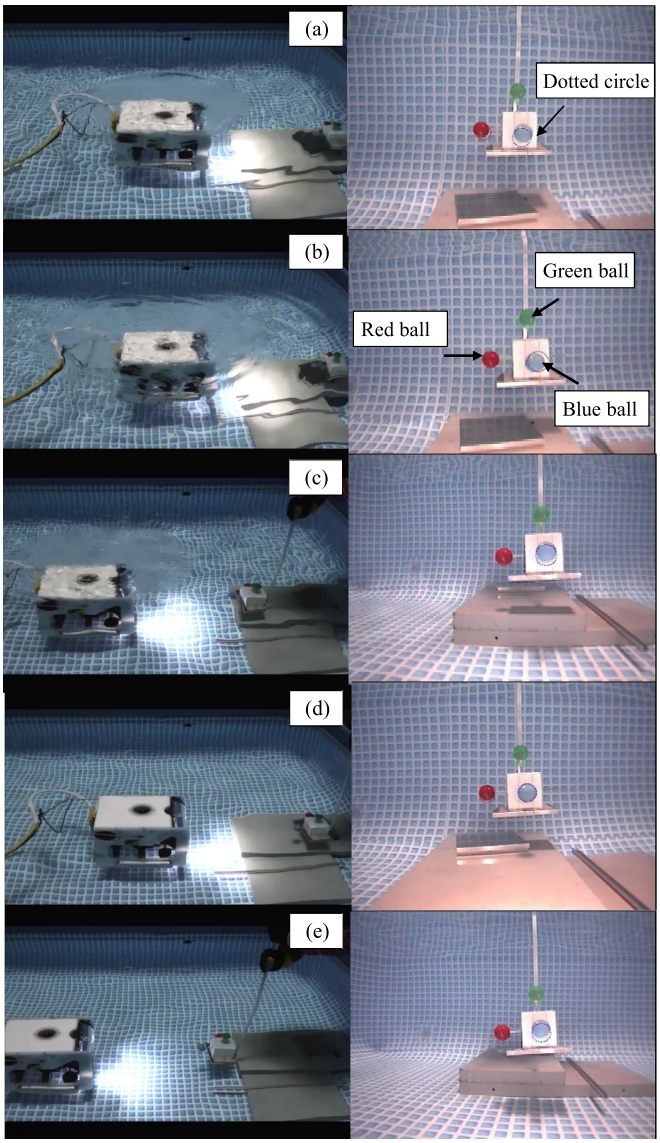

Fig. 23. Photographs of the underwater vehicle (left column) and captured images from the ROV (right column) without disturbances. These photographs correspond to the results shown in Fig. 20. (a) 10 s have passed (without making the 3-D marker move after starting experiment). (b) 20 s have passed (backward moving 3-D marker). (c) 30 s have passed (forward moving 3-D marker). (d) 40 s have passed (backward moving 3-D marker). (e) 50 s have passed (forward moving 3-D marker).

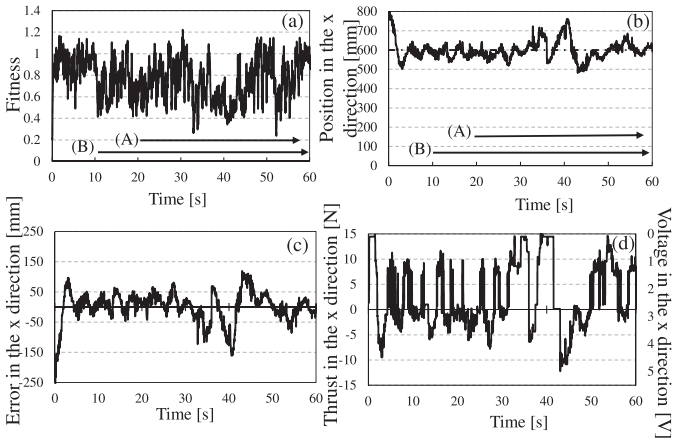

Fig. 24. Tracking performance of dual-eye image recognition with plain background with air bubble disturbance for the case in which "(A)" the 3-D marker moves in the x-direction with an amplitude of 280 mm and a period of 20 s after 20 s have passed from the start and "(B)" disturbances are generated after 10 s. (a) Fitness value. (b) Position of the underwater vehicle in the x-direction (dashed line is the desired position). (c) Tracking error in the x-direction. (d) Thrust in the x-direction. Photographs were taken every 10 [s] during this experiment and are shown in Fig. 27.

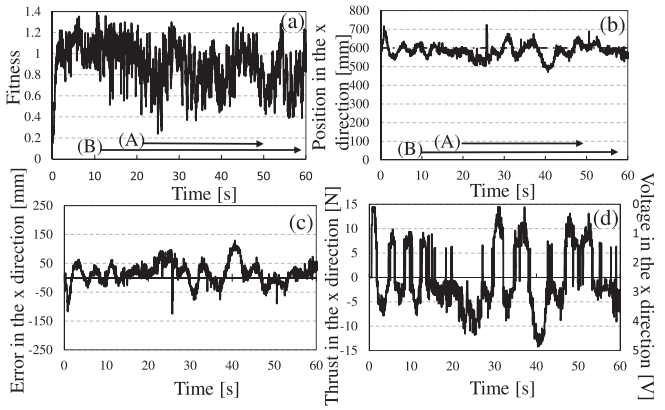

Fig. 25. Tracking performance of dual-eye image recognition with plain background with air bubble disturbance for the case in which "(A)" the 3-D marker moves in the x-direction with an amplitude of 280 mm and a period of 15 s after 20 s has passed from the start and "(B)" disturbances are generated after 10 s. (a) Fitness value. (b) Position of the underwater vehicle in the x-direction (dashed line is the desired position). (c) Tracking error in the x-direction. (d) Thrust in the x-direction.

The fitness function arises after 40 s in Fig. 24(a). And then, the desired position of x is kept as shown in Fig. 24(b). It can be seen in Figs. 22 and 24 that the large deviation in position error occurs due to thruster saturation between 30 and 40 s. Figs. 20–22 and 24–26 indicate that air bubbles have deteriorated the visual servoing closed loop stability, and it also has been worsened by the quick motion of 3-D marker with the short period. Although the air bubbles disturb the recognition of the 3-D marker while the marker is undergoing periodic motion, the ROV can keep tracking the moving 3-D marker without losing visual feedback information. The reader is referred to videos 4 to 7 in [37], which depict all the experiments performed in this study.

D. Docking Experiment Against Air Bubble Disturbances in Pool

The performance of the proposed system against air bubble disturbances was confirmed by conducting the docking experiment in the pool in this section. The docking experiment layout is shown in Fig. 12. To perform a docking experiment, a rod was attached to the right-hand side of the ROV and a round hole was set in the left-hand side of the 3-D marker as shown in Fig. 12. The ROV is with a rod (8 mm $\times$6 mm) on its right-hand side, and the 3-D marker is with a docking hole having a radius of 35 mm on its left-hand side. The center distance between the docking hole and the 3-D marker was 145 mm. The air bubbles were ejected from the chimney in front of the 3-D marker to verify the tolerance of the system against disturbances to the image, as shown in left-hand side of Fig. 12. The main task is to

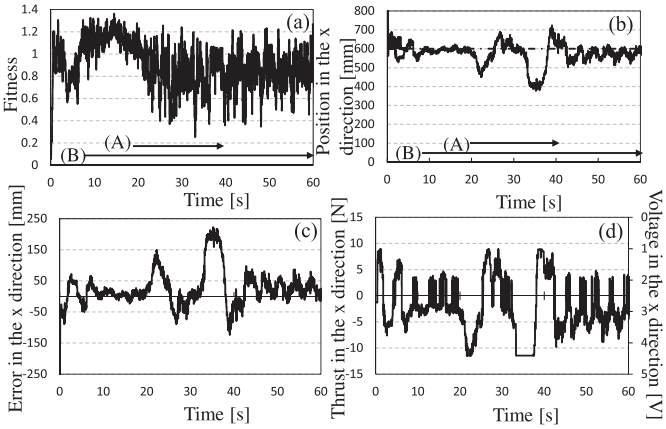

Fig. 26. Tracking performance of dual-eye image recognition with plain background with air bubble disturbance for the case in which "(A)" the 3-D marker moves in the x-direction with an amplitude of 280 mm and a period of 10 s after 20 s has passed from the start and "(B)" disturbances are generated after 10 s. (a) Fitness value. (b) Position of the underwater vehicle in the x-direction (dashed line is the desired position). (c) Tracking error in the x-direction. (d) Thrust in the x-direction.

confirm the restoration capacity to the desired pose, even in the presence of air bubble noise. Two instruments were designed to intermittently generate bubbles in the water using air pumps, as shown in Fig. 12. The experiment was conducted in a pool [2 m (length) $\times$3 m (width) $\times$0.75 m (height)] filled with tap water. Power was supplied and the control signal was transmitted from the PC through a tether cable (200 m). A proportional controller was used to control all thrusters in the docking experiment after confirming that the p controller was sufficient for this experiment. The integration and derivative terms could have been added into the controller if such adjustments were necessary, but preliminary experiments confirmed that adjustments were unnecessary.

The ROV moved forward to insert the rod into the docking hole. When the robot was within the error allowance range of pose relative to the 3-D marker, the range could be explained as follows. When the vehicle is stable within the tolerance of the position error, that is, $|e_i| < 20$ mm (where $i = x, y, z$) and $|e_{\varepsilon 3}| < 0.06$ (where $\varepsilon_{3d} = 0$ and is represented by a quaternion), which corresponds to 7°, where $\mathbf{e} = [e_x, e_y, e_z]^T = \mathbf{p}_d - \mathbf{p}, \mathbf{p}_d = [x_d, y_d, z_d]^T$, and $\mathbf{p} = [x, y, z]^T$, forward thrust causes insertion of the rod into the docking hole by gradually decreasing the desired distance between the vehicle and target object, where $x_d(t) = 600 - 30t$, with t being the time in seconds measured from the start of docking. The goal of this experiment was to complete the docking stage. Some conditions were applied in the pool experiment. There was no current and no wave. The water was clean.

The air bubble disturbances have been exerted to the ROV on the images while the vehicle is performing the docking experiments, aiming at confirming the tolerance of the proposed visual servoing system. The desired pose ($x_d = 600$ mm, $y_d = 12$ mm, $z_d = -10$ mm, and $\varepsilon_{3d} = 0°$) between the target and the ROV is predefined so that the ROV performs stationary hovering through visual servoing before docking. The air bubble disturbances were generated at the start of the experiment.

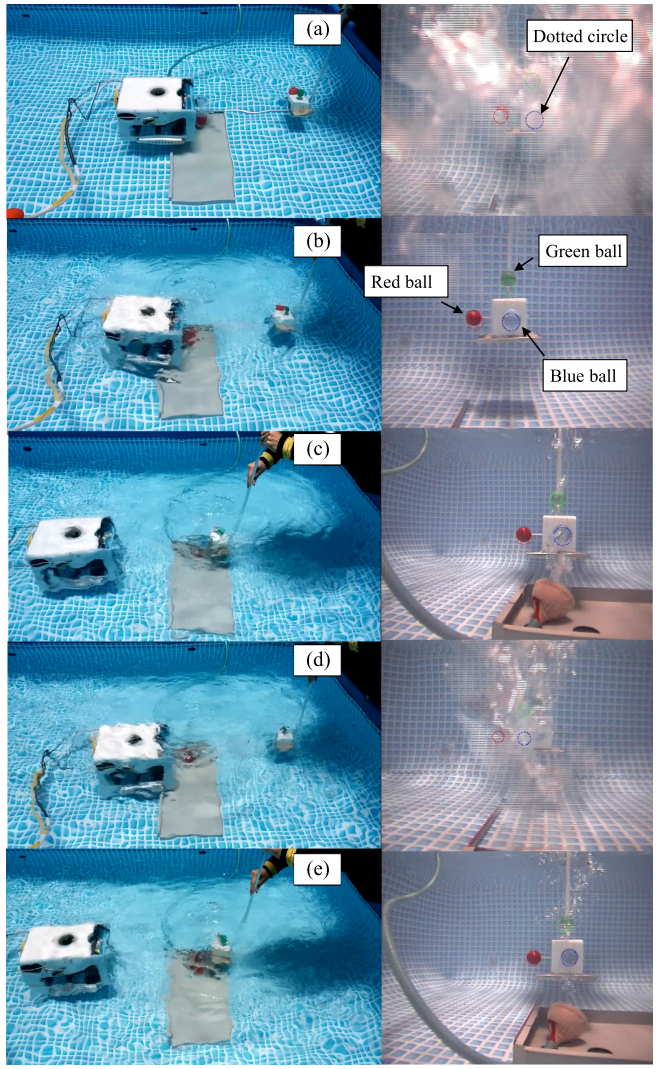

Fig. 27. Photographs of the underwater vehicle (left column) and captured images from the ROV (right column) with disturbances. These photographs correspond to the results shown in Fig. 24. (a) 10 s have passed (without making the 3-D marker move or air bubble disturbances after starting the experiment). (b) 20 s have passed (backward moving 3-D marker). (c) 30 s have passed (forward moving 3-D marker). (d) 40 s have passed (backward moving 3-D marker). (e) 50 s have passed (forward moving 3-D marker).

The results for the docking performance in the presence of air bubble disturbances are shown in Fig. 28. The dotted lines denoted as "A," "B," and "C" indicate the start of visual servoing, the start of docking, and the completion of docking, respectively. The photographs taken at these times are shown in Fig. 29. In Fig. 28(a), the fitness value increases to exceed 1 for the first few seconds of the recognition process and then maintains the allowable value of 0.6 for docking, which means that the system could recognize the 3-D pose of the target marker well. The fitness value 0.6 was set experimentally. Theoretically, the maximum fitness value is about 1.7. Experimentally, the maximum fitness value is about 1.2 in this study. Fig. 28(b)–(e) represents the relative pose between the desired pose and the estimated pose of the 3-D marker recognized by the RM-GA.

The docking strategy is explained in detail in [26], in which docking without air bubbles was conducted. In the docking

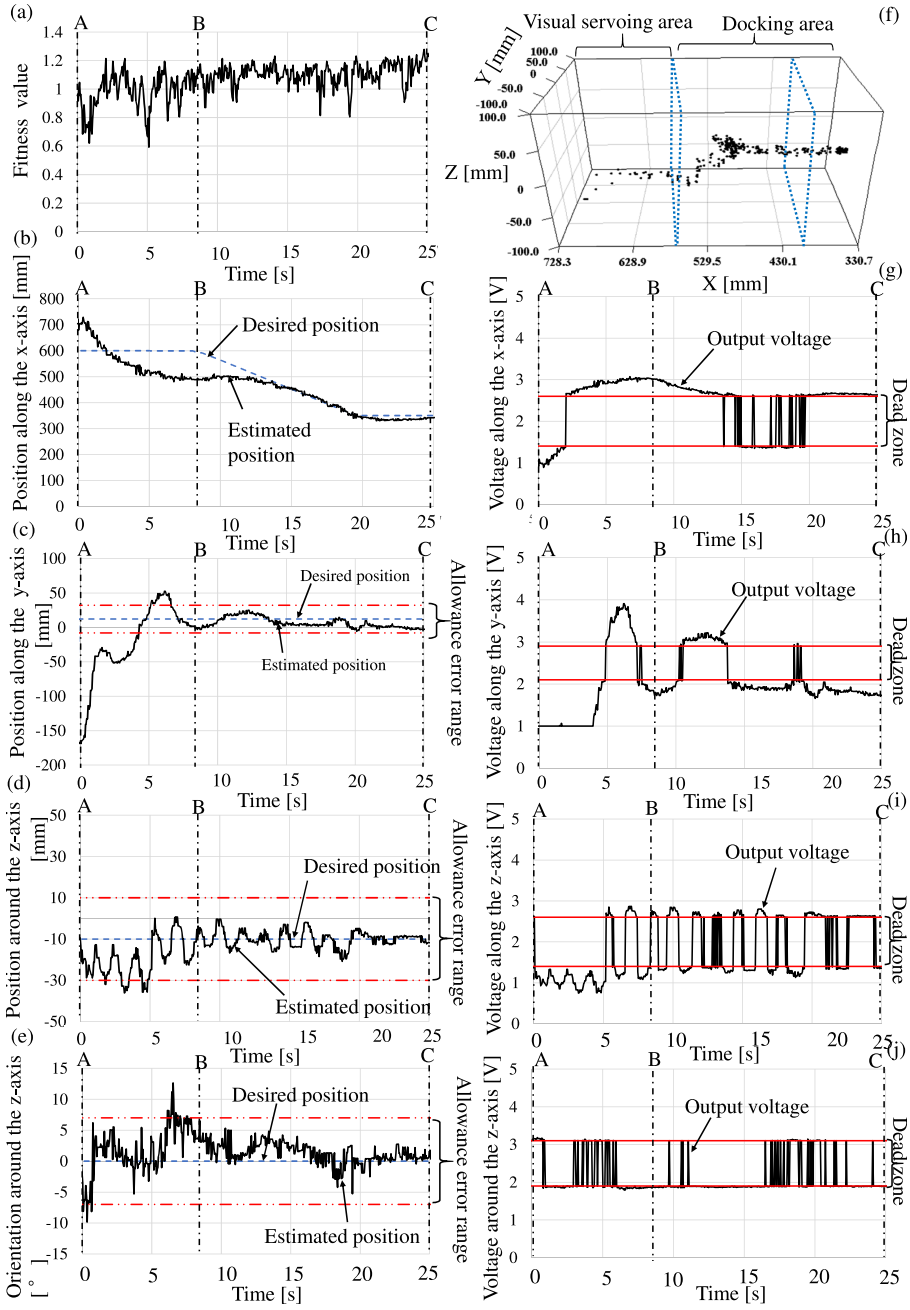

Fig. 28. Docking performance in pool with air bubble disturbance against dual-eye image recognition. (a) Fitness value. (b) Position along the x-axis. (c) Position along the y-axis. (d) Position along the z-axis. (e) Orientation around the z-axis. (f) detected 3-D pose of the underwater vehicle by RM-GA. (g) Voltage along the x-axis. (h) Voltage along the y-axis. (i) Voltage along the z-axis. (j) Voltage around the z-axis. The dotted lines denoted by "A," "B," and "C" in each of the subfigures indicate the start of visual servoing, the start of docking, and the completion of docking, respectively. Photographs taken at these times are shown in Fig. 29.

strategy, visual servoing starts when the 3-D marker is detected, which means the fitness value is above a defined threshold, 0.6. When the pose of the vehicle is within the allowable error range of ± 20 mm of the desired pose, as shown in Fig. 28(c) and (d), and the orientation around the z-axis is controlled to within 7°, as shown in Fig. 28(e), for the predefined period (165 ms, which is equal to five times the control loop period that is 33 ms) in this experiment, docking starts by decreasing the distance between the ROV and the 3-D marker from 600 to 350 mm, as shown in Fig. 28(b). The period between the dotted line labeled "A" and "B" in Fig. 28 indicates the visual servoing state, where the desired position along the x-axis is 600 mm. During the visual servoing state, the estimated position along the y-axis is out of error range, as shown in Fig. 28(c). A photograph taken at time in visual servoing period is shown in Fig. 29(a) and (b).

Therefore, visual servoing continues until the desired pose is within the error range for the y- and z-directions and the orientation around the z-axis, as shown in Fig. 28(c)–(e). At

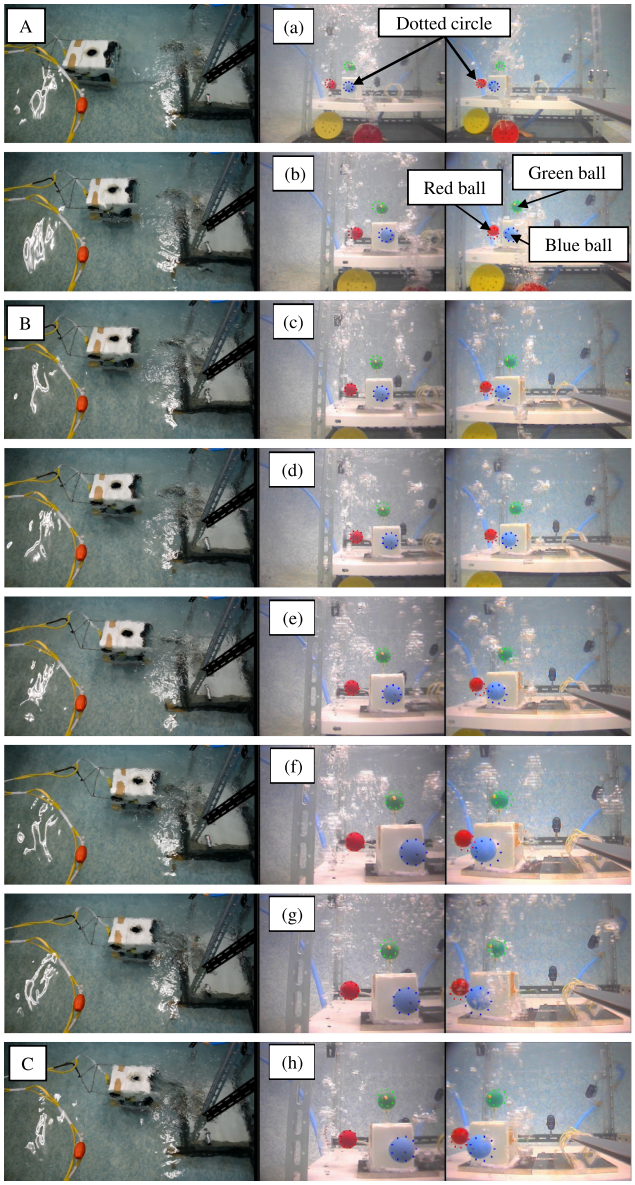

Fig. 29. Photographs of docking (left column) and dual-eye camera images from the ROV (right column) with air bubble disturbance. (a) and (b) 0–4 s visual servoing state. (c)–(e) 8–16 s docking process. (f)–(h) 20–25 s completion of docking. Each figure corresponds to the experimental results shown in Fig. 28, where the corresponding times are indicated by dotted lines "A," "B," and "C," respectively.

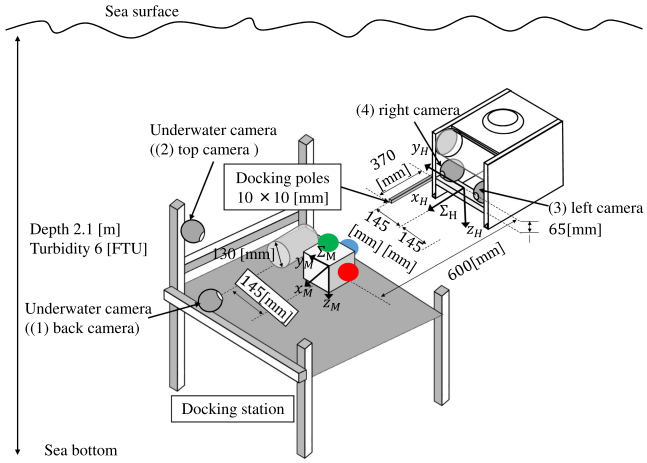

Fig. 30. Experimental setup of sea docking. The two underwater cameras, (1) back camera and (2) top camera, were attached to the docking station to monitor the behavior of the ROV.

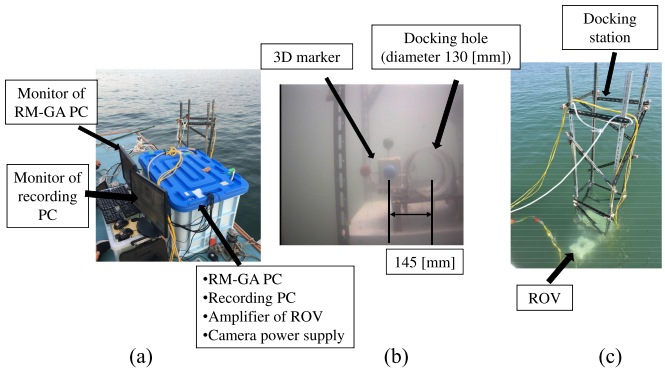

Fig. 31. Layout of the sea docking experiment. (a) ROV is controlled by the RM-GA PC while its behavior is monitored by the recording PC. (b) Left camera image of the ROV showing the 3-D marker and docking hole located in the docking station. The center distance between the docking hole and 3-D marker was 145 mm. (c) Position of ROV and docking station, which was a rectangular cuboid with length of 600 mm, width of 450 mm, and height of 3000 mm. The photograph (c) represents the completion of docking.

time B, as shown in Fig. 28(c)–(e), the docking criteria are satisfied and docking operation starts. Note that the position in the x-direction at point B is approximately 500 mm because only the positions in the y-and z-directions and the orientation around the z-axis are considered in the docking criteria. The docking operation started approximately 8 s after starting the experiment, and the image results for this condition are shown in the left- and right-hand columns of Fig. 29(c)–(e). The period between the dotted line labeled "B" and "C" in Fig. 28 indicates the docking process. In the docking process, the rod is fitted into the docking hole by decreasing the desired value of x_d, and this condition is expressed in Fig. 29(c)–(e). Finally, the docking operation was completed approximately at 25 s after starting the experiment, as shown in Fig. 29(f)–(h). The dotted line labeled "C" in Fig. 28 indicates the state whereby the docking is completed. During the docking process, if the docking state no longer meets the conditions of the desired allowance error range and the stable condition, the ROV returns to the visual servoing process and attempts to perform the docking process again. However, this condition did not occur in this experiment. Fig. 28(g)–(j) represents the thrust of each thruster required to restore the desired pose. The dotted line in Fig. 28(g)–(j) indicates the output voltage of each thruster and the solid lines express the dead zone range. The characteristic of thrust of each thruster changes with respect to the dead zone in the control voltage. Note that the dead zones shown in Fig. 28(g)–(j) are deleted, and the output values are bounced from upper limit to lower limit or *vice versa* to restore the desired pose for each thruster. This sudden transition means that the dead zones were compensated according to the configuration of the ROV to avoid fluctuations of the ROV's movement under manual control. The dots in Fig. 28(f) indicate the detected 3-D pose of the

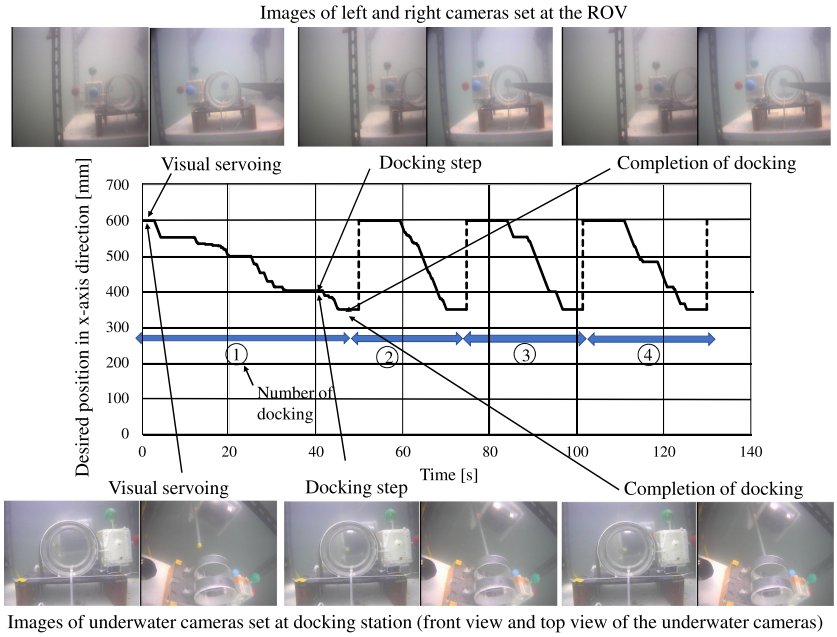

Fig. 32. Results of repeated docking experiment. Middle graph shows the desired position in the x-axis direction during four docking iterations in the sea. The upper photos represent the left and right cameras images set at the ROV taken at visual servoing stage, docking stage, and completion of docking stage. The lower photos are images taken by the underwater cameras set at the docking station at the same stages of docking process. Detailed results for the docking experiment ① is presented in Fig. 33 and the photographs of the docking experiment ① is shown in Fig. 34. The dotted line shows the returning procedure from docking completion position to restarting position of visual servoing, the ROV's returning motion took about 6 s.

underwater robot by RM-GA based on Σ_H (see Fig. 12) during the docking process. One black dot in Fig. 28(f) means the relative pose that the RM-GA calculated at every 33 ms. As shown in Fig. 28(f), when the pose distribution in visual servoing state and the docking state were compared, the pose distribution of the docking state is more compact than the visual servoing state because the vehicle is stable within the allowance error range in the docking area. The underwater robot was confirmed to have the ability to regulate the ROV to the desired pose while docking was performed, as shown in Fig. 28(b)–(e). Matching between the dotted circles and 3-D spheres of real 3-D marker in Fig. 29 (right-hand column) indicates how much the position of three spheres of 3-D marker model calculated by using the relative pose detected by RM-GA coincides with the real 3-D marker, which can show the degree of pose recognition of the 3-D marker. In particular, using the RM-GA, the proposed system was confirmed to sequentially recognize the relative pose of the 3-D marker and perform the docking process in real time, even in the presence of air bubbles.

E. Docking Experiment in Actual Sea Environment

In Section IV-D, the effectiveness of the proposed system in docking against air bubble disturbances in an artificial environment was discussed. However, real sea environments may degrade the visibility of the system more than the simulated artificial pool environment due to the other disturbances, such as turbidity, sun light, and waves. Since the RM-GA and control system have been confirmed to be effective and tolerable against image degradation caused by air bubbles, the author inferred that this system is also effective against turbidity. Therefore, the validity of the proposed system was confirmed in the real sea with the turbidity. Based on this motivation, the docking experiment was conducted in the coastal environment rather than clear water.

The sea docking experiments were conducted to verify the effectiveness of the proposed system against turbidity. The sea docking experiment was conducted within Okayama Prefecture, in Japan. The experimental setup and the layout of the sea docking experiment are shown in Figs. 30 and 31. The docking station was a rectangle of length (600 mm) $\times$width (450 mm) $\times$height (3000 mm), oriented with the long sides to the pier. Two underwater cameras are attached to the docking station for monitoring the behavior of the ROV during the experiments and recording for further analyses. The environmental conditions while conducting sea docking experiment are as follows: the time when the docking experiment was conducted is about 13:36 PM, the water depth from the surface to the sea bottom is 2.1 m, the turbidity level is 6 FTU [Formazin Turbidity Unit measured by TD-M500 (produced by OPTEX)], the illumination at the sea surface is 98 000 lx [measured by (model: LX-1010B, manufactured by Milwaukee, WI, USA)], the illumination in water at the depth of 1 m is 4800 lx and there were some waves in the sea during the docking experiment. In the sea docking experiment, the docking strategy is the same strategy as the pool test in the above explanation. The p controller was used to control the vehicle movement. But the diameter of the docking hole was 130 mm. The ROV was tethered and connected by a cable with 200 m length to the GA-PC (controller and 3-D pose estimation) on the pier.

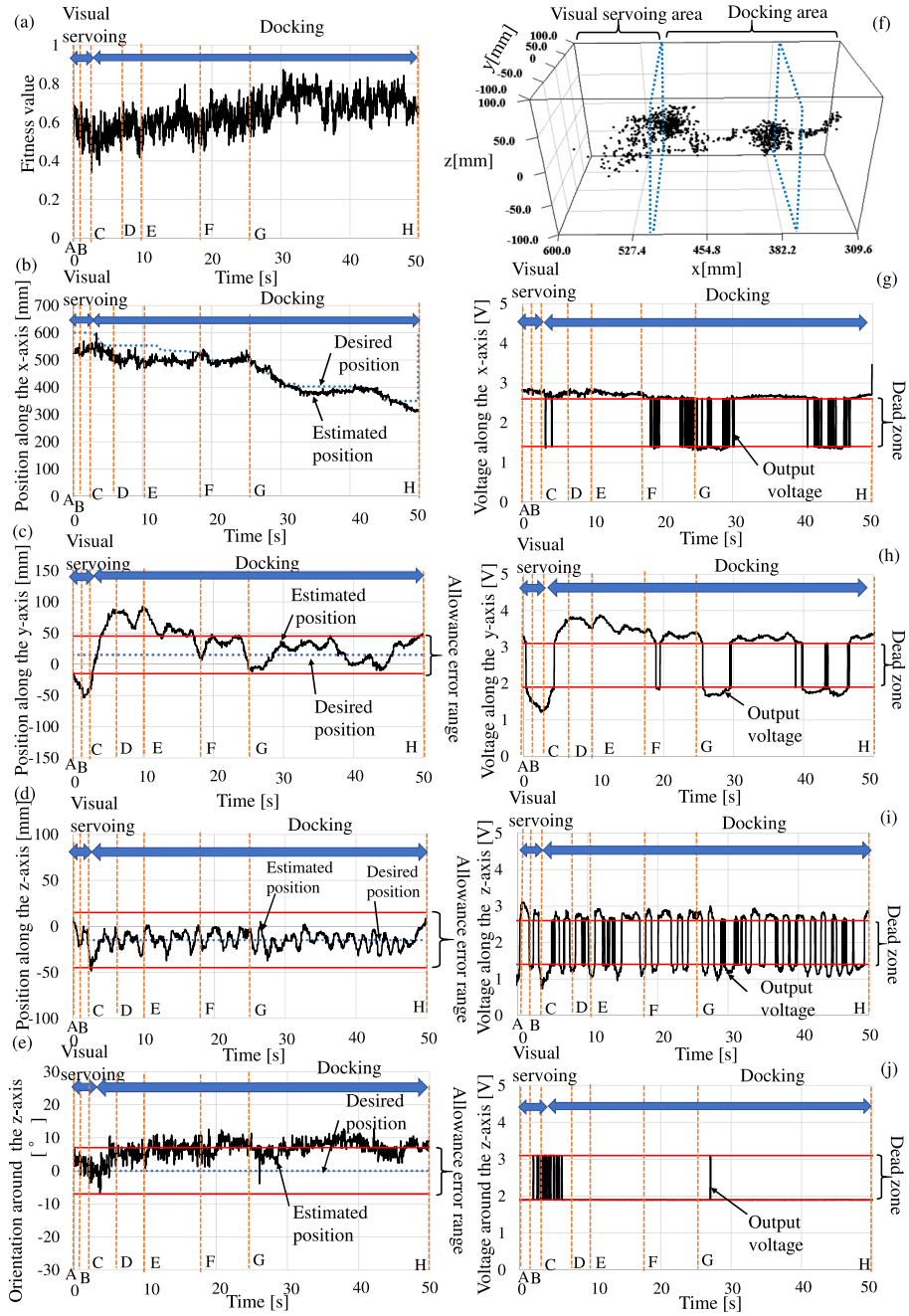

Fig. 33. Docking performance in an actual sea for the first time docking indicated by ① of Fig. 32. (a) Fitness value. (b) Position along the x-axis. (c) Position along the y-axis. (d) Position along the z-axis. (e) Orientation around the z-axis. (f) Detected 3-D pose of the underwater vehicle by RM-GA. (g) Voltage along the x-axis. (h) Voltage along the y-axis. (i) Voltage along the z-axis. (j) Voltage around the z-axis. The dotted lines denoted by "A"–"H" in each of the subfigures indicate the time during the first time docking of Fig. 32. Photographs taken at these times are shown in Fig. 34.

The docking experiment was performed successfully four times. After each docking completion, the ROV returned to its initial position automatically and repeated the docking/returning maneuvers four times. Desired position in the x-axis direction during four docking iterations in the sea with the example of photos that indicate the state of the visual servoing, the docking, and the completion of docking are shown in Fig. 32. Upon the completion of each docking operation, the desired position changed from 350 to 600 mm, as shown by the dotted lines in Fig. 32. As confirmed by these repeated dockings and shown in video 5 on our laboratory's home page [37], the return from the docking completion position to the initial position of visual servoing took about 6 s.

Among the four docking, the first docking ① took the longest to perform. This means that some disturbance is thought to have acted upon the ROV during this docking operation. Even though the pose of the vehicle sometimes exceeded the error allowance range, the vehicle could maintain the desired position within

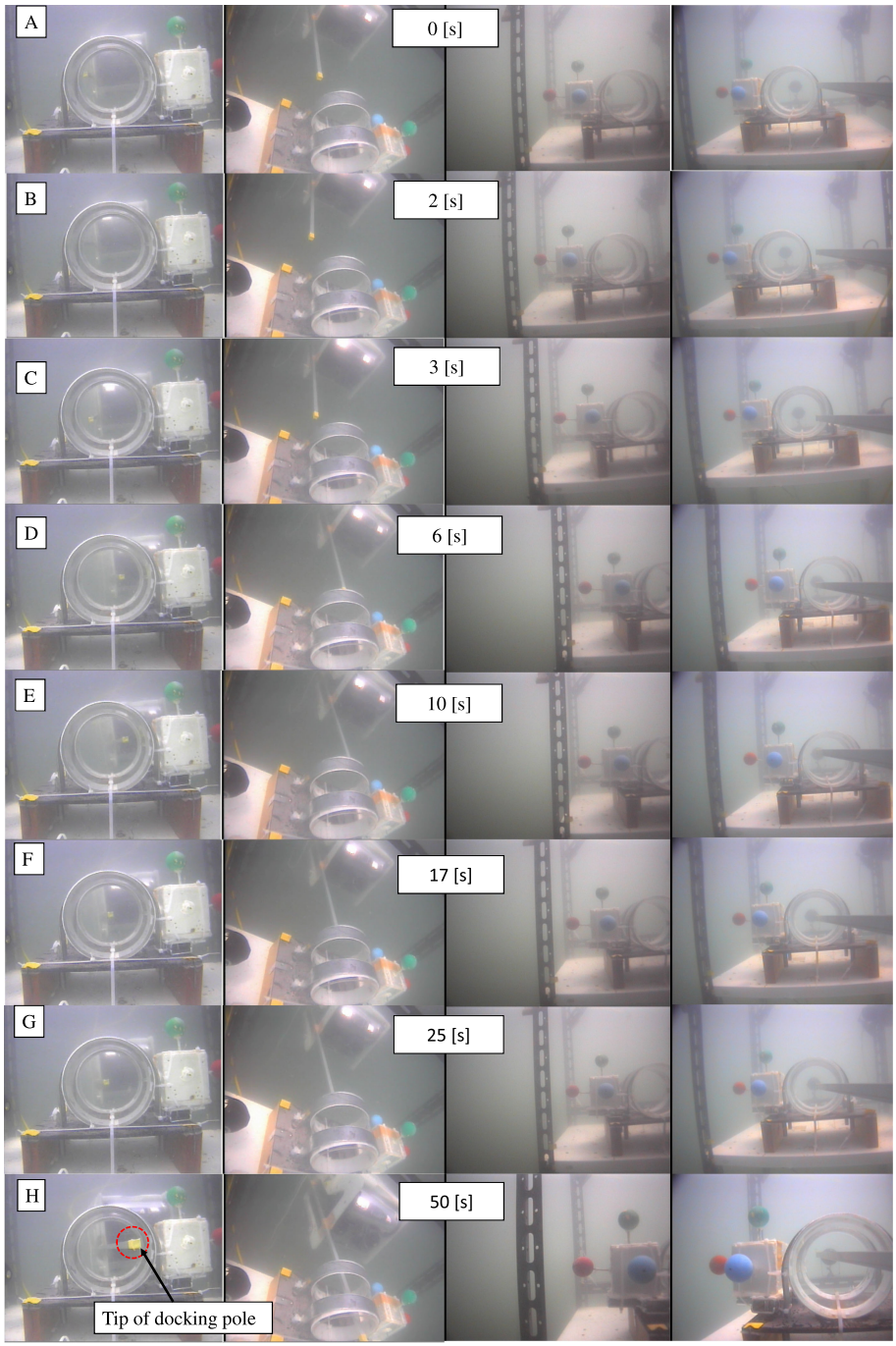

Fig. 34. First column shows the photographs of docking monitor by station back camera. Second column represents the photographs of docking monitor by station top camera. Third and fourth columns describe the ROV's left and right cameras images: "A"–"C" (0–3 s) visual servoing state, "C"–"H" (8–50 s) docking operation. Each figure corresponds to the experimental results shown in Fig. 33, where the corresponding times are indicated by dotted lines "A"–"H," respectively. Especially, it can be confirmed from image that the end of docking pole was inserted in docking hole at time "H".

the allowance range. The longest docking operation seems to be most useful for analysis of the docking motion. Therefore, the first time docking ① was chosen to examine in detail in Fig. 33.

The desired pose for sea docking ($x_d = 600$ mm, $y_d = 15$ mm, $z_d = -15$ mm, and $\varepsilon_{3d} = 0°$). When the ROV is stable in the above desired pose relative to the 3-D marker, the ROV moves forward to insert the docking pole into the docking hole as explained in the pool test.

Fig. 33(a) shows the fitness value in time domain. The marker detection is defined as the fitness value is 0.2 or above in the sea docking experiment. Therefore, when the fitness value is above 0.2, the visual servoing step started. Fig. 33(b)–(e), shows the positions along the x, y, z and orientation around the z-axis, respectively. The dotted lines labeled "A"–"H" in each subfigures indicate the time during the first time docking operation, and which are used as time indicator used in photographs at

"A"–"H" in Fig. 34. The two horizontal lines in Fig. 33(c)–(e) are the allowable error ranges that enable docking procedure. Fig. 33(f) shows the detect 3-D pose of the underwater vehicle by RM-GA. Fig. 33(g)–(j) described the thrust of each thruster required to restore the desired pose and continue docking. The horizontal lines in Fig. 33(g)–(j) express the dead zone range.

Fig. 33 shows that the first docking operation was completed approximately at 50 s after starting the experiment. At the start of the first docking experiment, the position along the y-axis is out of error range until the time "C" as shown in Fig. 33(c). Then, the controller tries to adjust the ROV within the allowable range with the output voltage in y-axis direction [see Fig. 33(h)]. The period between the dotted line "A"–"C" is the visual servoing state, as shown in Fig. 33. The corresponding photographs of these time are shown in Fig. 34 "A"–"C." When the vehicle pose is stable within the tolerance error range of the relative pose by adjusting in x-, y-, z-axes direction within ± 30 mm and the orientation within $\pm 7^{\circ}$ (see Fig. 12), the operation switched from the visual servoing step to the docking step.

According to the results, the positions along the z-axis is always within the allowable error range. After starting the docking step, the positions in y-axis are out of range again at the time of "D" and "E." At these times, the desired positions along the x-axis remained constant for some periods during the docking step because some fluctuations in the position along the y-axis and especially the rotation about the z-axis exceeded the error allowance. Consequently that, the controller tries to restore the vehicle to desired position can be seen the dotted line labeled "D" and "E" in Fig. 33(c) and (h). Photographs taken at these time are shown in Fig. 34 "D" and "E." After 17 s (see dotted line F), the position in the y-axis remains within the allowance error range until the end of the docking operation. At the time of 25 s (see dotted line G in Fig. 33), the orientation around the z-axis exceeds the allowance error range again. The desired position in the x-axis keep constant the position at that time. The photograph taken at this time can be seen in Fig. 34 "G." There are some fluctuations in the orientation around the z-axis during the docking process. This fluctuation seems to have occurred because of the effect of the waves. Therefore, the first time docking is the one with the longest operation time among the four repeated docking trials. However, the ROV could perform the docking operation by adjusting the error in orientation around the z-axis. Fig. 34 "H" is the end of docking, and the position of the docking pole can be seen through images of front view of underwater camera and top view of underwater camera. The part surrounded by the dotted circle is the docking pole tip and it can be confirmed that the docking pole can be fitted in the correct position in the docking hole. Therefore, the vehicle stopped visual servoing for a few seconds after docking completion to store the experimental data. After finished to store the docking data, the desired position in the x-axis direction changed to 600 mm, as shown in Fig. 32. The vehicle went back to the desired position about 600 mm in the x-axis direction and the second time docking started by visual servoing.

V. Conclusion

In this paper, visual servoing of an underwater vehicle using dual-eye cameras and a 3-D marker was conducted in the presence of air bubbles. Experiments were conducted to clarify the following:

1) the recognition performance in the presence of air bubbles with and without a background sheet;
2) the regulation performance when the 3-D marker was fixed;
3) the regulation performance when the 3-D marker is in periodic motion;
4) the docking performance in the presence of air bubbles in the pool;
5) the docking performance in the actual sea environments.

Since a hovering type ROV and a 3-D marker were used in our docking approach, and the vehicle could dock to a station from a distance of about 1 m, our results can be compared to those in [10]. Recognition accuracy of our proposed system is close to that from related works in [15] and [21], though those studies did not include docking operations. Comparing to other systems having each merit, the experimental results in this study provide an assessment of the system performance under air bubble disturbances and demonstrate the potential in practical environment in the sea. However, there are some limitations in our system. The controller needs to be improved for more stabilization. One of the limitations of our system that the vehicle should be guided to the field of view of camera about 1 m distance should be improved using other positioning systems in future. This means that our proposed visual servoing system needs to be combined with some functional unit to guide the vehicle to the space that the vehicle can find a 3-D marker, to make the combined system to be more usable and practical in real underwater world.

Acknowledgment

The authors would like to thank Mitsui E&S Shipbuilding Company, Ltd., and Kowa Corporation for their collaboration and support for this study. The authors would also like to express their thanks to Dr. Y. Nishida, Prof. K. Ishii, Prof. T. Maki, and Prof. T. Ura for their help and support.

References

[1] A. Negre, C. Pradalier, and M. Dunbabin, "Robust vision-based underwater target identification and homing using self-similar landmarks," in *Field and Service Robotics*. Berlin, Germany: Springer, 2008, pp. 51–60.
[2] M. D. Feezor, F. Y. Sorrell, P. R. Blankinship, and J. G. Bellingham, "Autonomous underwater vehicle homing/docking via electromagnetic guidance," *IEEE J. Ocean. Eng.*, vol. 26, no. 4, pp. 515–521, Oct. 2001.
[3] D. Ribas, N. Palomeras, P. Ridao, M. Carreras, and A. Mallios, "Girona 500 AUV: From survey to intervention," *IEEE/ASME Trans. Mechatronics*, vol. 17, no. 1, pp. 46–53, Feb. 2012.
[4] J. Evans, P. Redmond, C. Plakas, K. Hamilton, and D. Lane, "Autonomous docking for intervention-AUVs using sonar and video-based real-time 3D pose estimation," in *Proc. OCEANS Conf.*, 2003, vol. 4, pp. 2201–2210.
[5] G. Marani, S. K. Choi, and J. Yuh, "Underwater autonomous manipulation for intervention missions AUVs," *Ocean Eng.*, vol. 36, no. 1, pp. 15–23, 2009.

[6] A. Ortiz, M. Sim, and G. Oliver, "A vision system for an underwater cable tracker," *Mach. Vis. Appl.*, vol. 13, no. 3, pp. 129–140, 2002.
[7] J. Gao, P. Wu, B. Yang, and F. Xia, "Adaptive neural network control for visual servoing of underwater vehicles with pose estimation," *J. Mar. Sci. Technol.*, vol. 22, no. 3, pp. 470–478, 2017.
[8] S. Ghosh, R. Ray, S. R. K. Vadali, S. N. Shome, and S. Nandy, "Reliable pose estimation of underwater dock using single camera: A scene invariant approach," *Mach. Vis. Appl.*, vol. 27, no. 2, pp. 221–236, 2016.
[9] J. Y. Park, B. H. Jun, P. M. Lee, and J. Oh, "Experiments on vision guided docking of an autonomous underwater vehicle using one camera," *Ocean Eng.*, vol. 36, no. 1, pp. 48–61, 2009.
[10] T. Maki, R. Shiroku, Y. Sato, T. Matsuda, T. Sakamaki, and T. Ura, "Docking method for hovering type AUVs by acoustic and visual positioning," in *Proc. IEEE Int. Underwater Technol. Symp.*, 2013, pp. 1–6.
[11] N. Palomeras *et al.*, "I-AUV Docking and panel intervention at sea," *Sensors*, vol. 16, no. 10, pp. 1673, 2016.
[12] F. D. Maire, D. Prasser, M. Dunbabin, and M. Dawson, "A vision based target detection system for docking of an autonomous underwater vehicle," in *Proc. Australas. Conf. Robot. Automat.*, 2009, pp. 1–7.
[13] T. Ura, Y. Kurimoto, H. Kondo, Y. Nose, T. Sakamaki, and Y. Kuroda, "Observation behavior of an AUV for ship wreck investigation," in *Proc. MTS/IEEE OCEANS Conf.*, 2005, pp. 2686–2691.
[14] J. Guo, "Mooring cable tracking using active vision for a biomimetic autonomous underwater vehicle," *J. Mar. Sci. Technol.*, vol. 13, no. 2, pp. 147–153, 2008.
[15] S. Sagara, R. B. Ambar, and F. Takemura, "A stereo vision system for underwater vehicle-manipulator systems–proposal of a novel concept using pan-tilt-slide cameras–," *J. Robot. Mechatronics*, vol. 25, no. 5, pp. 785–794, 2013.
[16] S. Negahdaripour, X. Xu, and A. Khamene, "A vision system for real-time positioning, navigation, and video mosaicing of sea floor imagery in the application of ROVs/AUVs," in *Proc. 4th IEEE Workshop Appl. Comput. Vis.*, 1998, pp. 248–249.
[17] S. Negahdaripour and P. Firoozfam, "An ROV stereovision system for ship-hull inspection," *IEEE J. Ocean. Eng.*, vol. 31, no. 3, pp. 551–564, Jul. 2006.
[18] M. E. Ragab and K. H. Wong, "Multiple nonoverlapping camera pose estimation," in *Proc. 17th IEEE Int. Conf. Image Process.*, 2010, pp. 3253–3256.
[19] K. W. Phyu, R. Funakubo, R. Hagiwara, H. Tian, and M. Minami, "Verification of illumination tolerance for photo-model-based cloth recognition," *Artif. Life Robot.*, vol. 23, no. 1, pp. 118–130, 2018.
[20] H. Tian, Y. Cui, M. Minami, and A. Yanou, "Frequency response experiments of eye-vergence visual servoing in lateral motion with 3D evolutionary pose tracking," *Artif. Life Robot.*, vol. 22, no. 1, pp. 36–43, 2017.
[21] C. Choi, S. M. Baek, and S. Lee, "Real-time 3D object pose estimation and tracking for natural landmark based visual servo," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 3983–3989.
[22] R. Garcia and N. Gracias, "Detection of interest points in turbid underwater images," in *Proc. IEEE OCEANS Conf.*, Spain, 2011, pp. 1–9.
[23] F. Codevilla, J. D. O. Gaya, N. Duarte, and S. Botelho, "Achieving turbidity robustness on underwater images local feature detection," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[24] A. Yamashita, A. Fujii, and T. Kaneko, "Three dimensional measurement of objects in liquid and estimation of refractive index of liquid by using images of water surface with a stereo vision system," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2008, pp. 974–979.
[25] A. Yamashita, S. Kato, and T. Kaneko, "Robust sensing against bubble noises in aquatic environments with a stereo vision system," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2006, pp. 928–933.
[26] M. Myint, K. Yonemori, A. Yanou, M. Minami, and S. Ishiyama, "Visual-servo-based autonomous docking system for underwater vehicle using dual-eyes camera 3D-pose tracking," in *Proc. Int. Symp. Syst. Integr.*, 2015, pp. 989–994.
[27] M. Myint, K. Yonemori, A. Yanou, K. N. Lwin, M. Minami, and S. Ishiyama, "Visual-based deep sea docking simulation of underwater vehicle using dual-eyes cameras with lighting adaptation," in *Proc. Int. OCEANS Conf.*, Shanghai, China, 2016, pp. 1–8.
[28] M. Myint, K. Yonemori, A. Yanou, K. N. Lwin, N. Mukada, and M. Minami, "Dual-eyes visual-based sea docking for sea bottom battery recharging," in *Proc. MTS/IEEE OCEANS Conf.*, Monterey, CA, USA, 2016, pp. 1–7.
[29] K. N. Lwin *et al.*, "Performance analyses and optimization of real-time multi-step GA for visual-servoing based underwater vehicle," in *Proc. TECHNO-OCEAN Conf.*, 2016, pp. 519–526.
[30] M. Myint, K. Yonemori, A. Yanou, K. N. Lwin, M. Minami, and S. Ishiyama, "Visual servoing for underwater vehicle using dual-eyes evolutionary real-time pose tracking," *J. Robot. Mechatronics*, vol. 28, no. 4, pp. 543–558, 2016.
[31] M. Myint, K. N. Lwin, N. Mukada, T. Matsuno, and M. Minami, "Stereo vision-based 3D pose estimation under turbid water for underwater vehicles," in *Proc. Robot. Soc. Japan*, 2017, pp. 1–4, Paper RSJ2017AC1C3-02.
[32] M. Myint, K. Yonemori, A. Yanou, S. Ishiyama, and M. Minami, "Robustness of visual-servo against air bubble disturbance of underwater vehicle system using three-dimensional marker and dual-eye cameras," in *Proc. MTS/IEEE OCEANS Conf.*, Washington, DC, USA, 2015, pp. 1–8.
[33] A. Yanou, K. Yonemori, S. Ishiyama, M. Minami, and T. Matsuno, "Control characteristics of visual-servo type underwater vehicle system using three-dimensional marker for air bubble disturbance," *Trans. Soc. Instrum. Control Eng.*, vol. 52, no. 5, pp. 284–291, 2016.
[34] M. Minami, J. Agbanhan, and T. Asakura, "Evolutionary Scene Recognition and Simultaneous Position/Orientation Detection," in *Soft Computing in Measurement and Information Acquisition*, Berlin, Germany: Springer, 2003, pp. 178–207.
[35] W. Song, M. Minami, and S. Aoyagi, "On-line stable evolutionary recognition based on unit quaternion representation by motion-feedforward compensation," *Int. J. Intell. Comput. Med. Sci. Image Process.*, vol. 2, no. 2, pp. 127–139, 2008.
[36] M. Myint, K. Yonemori, K. N. Lwin, A. Yanou, and M. Minami, "Dual-eyes vision-based docking system for autonomous underwater vehicle: An approach and experiments," *J. Intell. Robot Syst.*, vol. 92, pp. 1–28, 2017, doi: 10.1007/s10846-017-0703-6.
[37] [Online]. Avaliable: http://www.suri.sys.okayama-u.ac.jp/underwater/index.html

Khin Nwe Lwin received the B.E. degree in mechatronic engineering from Technological University, Kyaukse, Myanmar, in 2006, the M.E. degree in mechatronic engineering from Mandalay Technological University, Mandalay, Myanmar, in 2011, and the Ph.D. degree from the Division of Industrial Innovation Sciences, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan, in 2018, working on visual servoing control and docking of ROV with real-time 3-D pose estimation by stereo-vision.

She was a Lecturer with the Faculty of Precision Engineering, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar. Her research interests include visual servoing, control of autonomous underwater vehicle, robotics, and GA.

Naoki Mukada received the B.E. degree from Okayama University, Okayama, Japan, in 2016, and the M.E. degree from the Division of Industrial Innovation Sciences, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan, in 2018, working on robust visual pose estimation for docking of ROV.

He is currently with Disco Corporation, Tokyo, Japan. His research interests include 3-D pose estimation, robotics, and optimization.

Myo Myint received the M.E. degree from Yangon Technological University Yangon, Myanmar, in 2006, and the Ph.D. degree from Mandalay Technological University, Mandalay, Myanmar, in 2009, both in electronic engineering. He received another Ph.D. degree in intelligent control and robotics from Division of Industrial Innovation Sciences, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan, in 2018.

From 2011 to 2015, he was an Associate Professor at Yangon Technological University. He is currently a Professor of Electronic Engineering at Technological University, Thanlyin, Myanmar. His research interests include autonomous underwater vehicles, robotics, electronics, and renewable energy.

Daiki Yamada received the B.E. degree from the Division of Industrial Innovation Sciences, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan, in 2017, working on robust visual pose estimation for docking of ROV.

His research interests include 3-D pose estimation, robotics, and optimization.

Akira Yanou received the Ph.D. degree in engineering from Okayama University, Okayama, Japan, in 2001.

From 2002 to 2008, he was with the School of Engineering, Kinki University, Osaka, Japan. From 2009 to 2016, he was with the Graduate School of Natural Science and Technology, Okayama University. He joined the Department of Radiological Technology, Kawasaki College of Allied Health Professions, Kurashiki, Okayama, Japan, in 2016, and Faculty of Health Science and Technology, Kawasaki University of Medical Welfare, Kurashiki, Okayama, Japan, in 2017, as an Associate Professor. His research interests include adaptive control and strong stability systems and estimation.

Takayuki Matsuno received the M.E. and Ph.D. degrees from Nagoya University, Nagoya, Japan, in 2000 and 2005, respectively.

He was with Nagoya University in 2004. He was with Toyama University, Japan, in 2006. In 2011, he was with the Okayama University, Okayama, Japan. His research interest focuses on manipulation of deformable linear objects.

Kazuhiro Saitou is currently member of the Technical Staff at the Ushimado Marine Institute, Faculty of Science, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan.

Waichiro Godou is currently member of the Technical Staff at the Ushimado Marine Institute, Faculty of Science, Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan.

Tatsuya Sakamoto received the Ph.D. degree from The University of Tokyo, Tokyo, Japan, in 1992.

Since 2003, he has been a Professor and the Director at the Ushimado Marine Institute, Faculty of Science and a Professor in the Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan. Since 2014, he has been the Chairperson of the Director-Council of the Japanese National Marine and Inland Water Biological Stations.

Mamoru Minami received the B.S. and M.S. degrees in aeronautical engineering from the Osaka Prefecture University, Osaka, Japan, in 1978 and 1980, respectively, and the Ph.D. degree from the Faculty of Engineering, Kanazawa University, Kanazawa, Japan, in 1992, working on visual servoing control of a mobile manipulator utilizing hand-eye camera image perception.

Since 1994, he has been an Associate Professor of Mechanical Engineering at the University of Fukui, Fukui, Japan, and moved to a newly established Department of Human and Artificial Intelligence Systems in 1999. He has been a Professor there since 2002. He was a Professor with the Faculty of Engineering, Okayama University, Okayama, Japan, in April 2010. His current research interests include robotics and automation in dynamics, kinematics and control of robots including mobile manipulators, and intelligent motion control. To research intelligent robot, he has been engaged in machine vision for real-time adaptive operation, stereo vision, eye-vergence visual servoing, autonomous underwater vehicle, and underwater docking system.