J Intell Robot Syst
https://doi.org/10.1007/s10846-017-0703-6

Dual-eyes Vision-based Docking System for Autonomous Underwater Vehicle: an Approach and Experiments

Myo Myint[1] · Kenta Yonemori[1] · Khin Nwe Lwin[1] · Akira Yanou[2] · Mamoru Minami[1]

Received: 21 February 2017 / Accepted: 29 August 2017

Abstract A critical challenge for autonomous underwater vehicles (AUVs) is the docking operation for applications such as sleeping under the mother ship, recharging batteries, transferring data, and new mission downloading. The final stage of docking at a unidirectional docking station requires the AUV to approach while keeping the pose (position and orientation) of the vehicle within an allowable range. The appropriate pose therefore demands a sensor unit and a control system that have high accuracy and robustness against disturbances existing in a real-world underwater environment. This paper presents a vision-based AUV docking system consisting of a 3D model-based matching method and Real-time Multi-step Genetic Algorithm (GA) for real-time estimation of the robot's relative pose. Experiments using a remotely operated vehicle (ROV) with dual-eye cameras and a separate 3D marker were conducted in a small indoor pool. The experimental results confirmed that the proposed system is able to provide high homing accuracy and robustness against disturbances that influence not only the captured camera images but also the movement of the vehicle. A successful docking operation using stereo vision that is new and novel to the underwater vehicle environment was achieved and thus proved the effectiveness of the proposed system for AUV.

Keywords Visual servoing · Genetic algorithm · Underwater vehicle · Underwater docking · Stereo vision

1 Introduction

Nowadays, autonomous underwater vehicles (AUVs) are expected be one of the essential tools in applications such as inspection of underwater structures (e.g., dams and bridges) and underwater cable tracking [1]. Even though many studies have been conducted and published worldwide, researchers face rapidly increasing demands to expand the roles of AUVs. In spite of developing technologies related to power storage devices, underwater vehicles are especially limited in operations that take longer than the duration supported by the vehicle's power capacity. A recharging unit with an underwater docking function can enable AUVs to operate for extended periods in the sea independently of a surface vessel, making the docking operation important not only for battery recharging applications but also other applications such as sleeping under the mother ship or new mission downloading. Moreover, the docking capacity can be extended to provide navigation for other underwater vehicles on the way to their own missions [2]. However, a number of challenging issues hinder these applications, which require high accuracy and robustness against disturbances that occur in the underwater environment. To achieve these tasks in underwater vehicles, we have developed a vision-based docking system using stereo vision.

Research on docking operations using various homing sensors and techniques for the underwater robot has been

✉ Myo Myint
puqs1ci8@s.okayama-u.ac.jp

Akira Yanou
yanou-a@mw.kawasaki-m.ac.jp

Mamoru Minami
minami-m@cc.okayama-u.ac.jp

[1] Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan

[2] Faculty of Health Science and Technology, Kawasaki University of Medical Welfare, 288 Matsushima, Kurashiki, Okayama, 701-0193, Japan

Published online: 13 October 2017

conducted worldwide [3–10, 15–23]. The optical terminal guidance technique was introduced in [3]. A docking guidance system was designed and implemented by the Sugeno fuzzy inference system (FIS) in [4]. In [5], an electromagnetic homing (EM) system for docking was proposed and tested. In [6], the AUV homed to a docking station by using an ultrashort base line (USBL) sonar transceiver mounted in the vehicle nose. The work in [7] proposed a robust AUV docking guidance and navigation approach that can handle unknown current disturbances without a velocity sensor. The work in [8] addressed robust vision-based target recognition by presenting a scaled and rotationally invariant target design and a recognition routine based on self-similar landmarks (SSL).

Recently, due to the progress in computer vision, a vision-based system has been highlighted as a promising navigation system. As in land and space systems, numerous studies on underwater vehicles using visual servoing have recently been conducted worldwide. Each study has different merits and limitations depending on the intended application. Most research is based on monocular vision [9, 11]. In [10], features in a man-made plate were extracted and the relative pose was estimated from camera images. In [12], a vision system using two cameras and artificial underwater landmarks for autonomous operation was reported. Even though two cameras were installed in the vehicle in [10, 12], both cameras did not see the same object at the same time to estimate the relative pose. One camera detected the target and the second camera performed other tasks.

In contrast, we have developed a 3D pose tracking system using stereo vision to provide high homing accuracy [13]-[17]. Stereo vision can use parallactic displacement, which is effective for camera depth estimation in 3D pose detection. On the other hand, the monovision estimation method cannot use parallactic displacement. We think the benefit of stereo vision over a single camera exists in the inherited merit that dual-eye has in a parallactic looking way by the proposed pose estimation, as shown in [13]. In our previous papers, the robustness of the 3D pose estimation against disturbances (air bubbles [13] and in the case of target occlusion [14]) are discussed with the experimental results of visual servoing, which is one of docking steps in this paper. In [13], a visual servoing experiment in which a 3D marker moved in a back and forth direction was conducted. That experiment showed the effectiveness of the proposed system for pose tracking of a moving 3D object. Then, the robustness of regulating performance under the condition of noise due to air bubbles in front of the 3D marker was verified. The air bubble disturbances added not only random noise to the captured images but also physical disturbances to the movement of the vehicle. In [14], the robustness of visual servoing was discussed, especially in the case when the target is partially occluded. In [15], visual servoing against a physical disturbance in a specific direction and briefly the docking performance was reported. As a continuation of the work reported in [15], this paper proposes a docking strategy with a detailed explanation of how 3D real-time pose estimation has been achieved. The experimental docking results when the vehicle is in arbitrary underwater positions are also discussed. As a follow-up work reported in [16], the proposed system was developed to perform docking simulations to examine the illumination variation, especially when the LED light installed on the vehicle is dominant. The difference between the previous paper [17] and the present paper is that a detailed explanation of a proposed approach (3D pose estimation and motion controller by visual servoing using two cameras in real-time for underwater vehicle) is presented in the present paper but a detailed methodology discussion is not included in the previous paper [17]. The topic of the previous work [17] is exemplified on the practicality of the proposed system by showing experimental results of docking in a real sea. The present paper is intended to be the first paper that explains about the proposed methodology of 3D-pose real-time estimation in details with experimental results. Moreover, Not only docking performances for three different starting positions but also performance against physical disturbances simulating water currents is discussed in the present paper, which has not been included in [17]. To the best knowledge of the authors, our proposed system is the first initiated research using two cameras as standalone stereo vision for 3D pose tracking in underwater docking.

In this paper, an efficient and robust real-time autonomous docking system using visual servoing with stereo vision is presented. The system, named Three-Dimensional Move-on Sensing (3D-MoS), is shown in Fig. 1, in which 3D perception enabled by dual-eyes visual pose tracking by using known 3D marker is used for controlling the vehicle's relative pose to the desired one. The 3D-MoS system recognizes a relative pose between a remotely operated vehicle (ROV) and a target object by utilizing 3D model-based recognition using dual-eye cameras images with a video frame rate of 30 fps. In the proposed approach, visual information

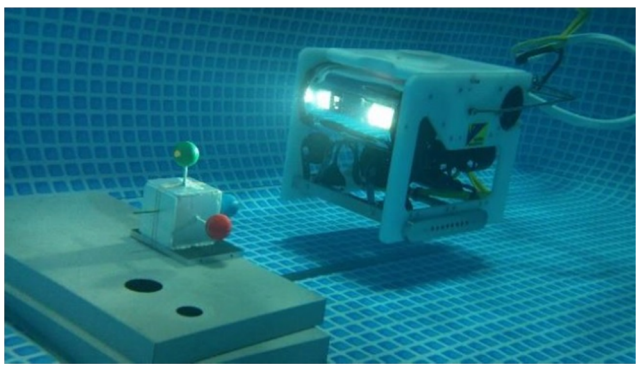


Fig. 1 Underwater vehicle and 3D marker

is directly used in feedback control in real-time. Additionally, developed optimization method named Real-time Multi-step GA is implemented in accordance with the concept of optimization of dynamic images for real-time target tracking. This combination of 3D-MoS and Real-time Multi-step GA in feedback loop using two cameras images is the main novelty of this study. There are three experiments in this paper to evaluate the effectiveness of proposed approach. The first one is regulating performance experiment in which the defined relative pose between the ROV and the target was maintained by the proposed system. Secondly, the robustness of visual servoing was evaluated by exerting an abrupt disturbance force while the ROV was controlled by visual servoing. Finally, a docking experiment for battery recharging was conducted in an indoor pool, which simulated an underwater sea environment. Implementation of the docking system for underwater automatic battery charging using the proposed 3D-MoS system was examined experimentally, particularly with respect to robustness against disturbances. This paper further details the docking strategy using 3D-MoS through visual feedback of Real-time Multi-step GA model-based matching.

The remainder of this paper is organized as follows: Section 2 presents the problem statement of a visual servoing based docking system for underwater vehicles with discussions on related works. Section 3 describes the novelties of the proposed approach. In Section 4, a detailed explanation of 3MoS with Real-time Multi-step GA is described. Section 5 describes the proposed docking system. Experimental results are described in Section 6. The final section concludes the paper.

2 Problem Statement

Since an intended application is a sea docking operation, this section presents the problem statements of the whole proposed system with discussions of the background and related work. The specific research questions are discussed in the view of novelty in next Section 3.

2.1 Docking Station

The two common configurations of docking stations are omnidirectional [18], where the docking hole can rotate to allow a vehicle to approach and dock from any azimuth, and unidirectional [19], where the docking hole is oriented in a specific direction. The unidirectional station has been selected in many studies because of its robustness and simplicity. However, the final approach of unidirectional docking is a difficult task, even though expensive navigation sensor suites and large-scale dead-reckoning sensors are able to provide position data. In this work, a simulated

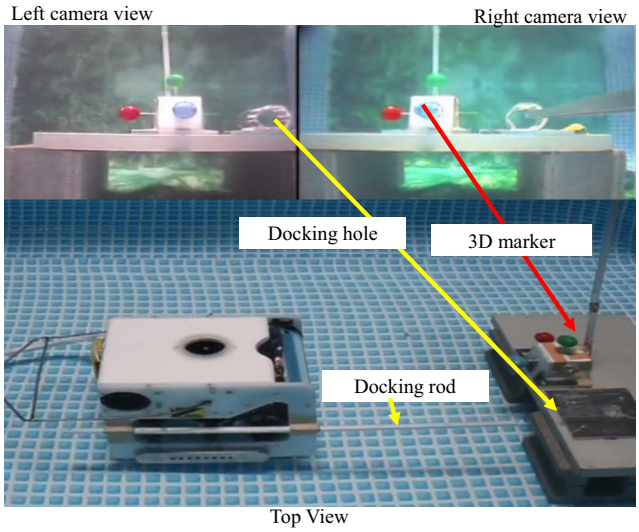


Fig. 2 ROV and simulated docking station with left camera view and right camera view

docking station with unidirectional entry, as shown in Fig. 2, is designed for underwater battery recharging. Therefore, the initial objective for successful docking is that the proposed system provide high homing accuracy and robustness against disturbances.

With regard to homing mechanisms, a docking system [5] has a mechanical capturing function that can physically damage the AUV and the dock. In research [6] using a torpedo-type vehicle, the lack of a hovering function reduced the docking efficiency, even though the speed of the vehicle was suitable. To avoid these issues, hovering-type ROVs and switching ability between the visual servoing mode and the docking mode are considered in the proposed system.

2.2 Sensor Configuration

Few studies have provided reasonable homing accuracy from meter level to centimeter level, because the sensors are expensive in money and space due to the increased numbers of sensor units and the computational load. Some studies such as [23] focused on vision-based localization using data fusion through integration of a global navigation satellite system (GNSS), inertial navigation system (INS), and vision system. In [24], Doppler Velocity Log (DVL) and a Ultra-Short Base Line position system (USBL) were used for underwater vehicle localization. However, in the final stage of unidirectional docking, when precise homing accuracy is necessary, the vision system needs to be dominant. Therefore, instead of integration with other sensors, a standalone dual-eye vision system was added in this work. Nevertheless, the proposed system can be extended by being integrated with other navigation systems.

2.3 3D-MoS Using Landmarks

In some docking experiments in other works, the relative pose is estimated by one camera and a known target or landmark [2, 20–23]. The pose estimation techniques reported in these works employ feature-based recognition. The work in [20] provided the relative position and distance from a geometric arrangement of lights set at docking station. Especially, the calculation of relative orientation was more complicated and difficult than detection of the position. In [23], the feature-based algorithm ORB was used. ORB is a combination of oriented features from the accelerated segment test (FAST) and rotated binary robust independent elementary features (BRIEF). ORB was applied in [25] for pose estimation of a man-made plate by using camera images to support the navigation system when position data from other sensors were no longer available. Even though the applied vision-based docking detector algorithm utilized the ORB feature extractor, the derived relative information was input into a localization filter for information synthesis rather than a standalone vision sensor. In [2], a signboard system including four colored balls was used as a passive target object. Based on the known information about the signboard system, the distance and orientation (heading angle only) of a vehicle to the target were calculated for position and heading error correction. The vehicle was assumed to be in a horizontal plane at the same level as three of the four balls, and so the accuracy was very dependent on other sensors, such as the altimeter, as well as stability control of the vehicle.

In spite of the advantages of the above works, feature extraction and matching of feature points to estimate the relative position and especially the orientation of a landmark or a target have some degree of uncertainty in terms of the reliability of mapping features. On the other hand, 3D recognition through learning [26] and 3D reconstruction [27, 28] require expensive computational loads. To avoid the limitations while increasing the effectiveness of the estimation of the relative pose in real time, we propose model-based recognition utilizing a GA on dynamic image input with a video frame rate based on a projection of target models from 3D to 2D. The 3D passive target is simple but has full information about the size, color, and shape. Moreover, the projection of a solid object as a whole rather than individual pixels and the evaluation with a fitness function using voting performance make the proposed system more robust.

2.4 Optimization Problem

In the proposed 3D-MoS system with the Real-time Multi-step GA, the best chromosome that represents the most trustful pose is the chromosome with the highest fitness function value for correlation between the model, whose pose is defined by the chromosome, and real target in the input images. The best chromosome has to be evaluated by an optimization technique instead of evaluating all possible chromosomes. Many kinds of powerful advanced optimization methods are available. However, almost all focus on accuracy rather than real-time application merit. In contrast, with the goal of constructing a pose feedback control system for docking, the two criteria of accuracy and real-time performance are indispensable to extract pose estimation in dynamic images input by video rate. Therefore, instead of comparing different optimization methods, GAs, which have a long history of usefulness, are selected and utilized in the form of the Real-time Multi-step GA for the proposed system. In other words, our strategy is utilizing the Real-time Multi-step GA, which has simplified optimizing calculations with reasonable performance in one loop and increasing accuracy with repeatability within a real-time video frame rate, that is 30 fps.

2.5 Robustness Against Disturbances

As in the space environment, the underwater world gives complexity to underwater vehicle operation due to disturbances [7]. Because the proposed system is a vision-based system, not only the physical disturbances of ocean currents but also the noise in recognized images should be considered in the experiments. By completing the experimental tasks while including these considerations, the proposed docking system demonstrates its effectiveness against different disturbances.

Regarding very low light conditions, it is intended to use a LED lighting unit that is attached to the vehicle to be able to operate in a deep sea or night time. Some experiments in the dark environment in which only the LED lighting unit installed on the ROV was used and the effect of the illumination variation on visual servoing was discussed in a previous work [16]. According to experimental result in [16], we confirmed that the proposed approach works in low light conditions even though the LED installed on the ROV addresses an illumination variation with a certain degree caused by the dynamic movement of the vehicle while docking operation has been pursued. One of the reasons why the proposed system works under illumination variation is a choice of hue value representation of color in HSV that is less sensitive to light changing.

Concerning strong water turbidity, the turbidity tolerance of the proposed system have not been reported even though docking experiments have been conducted in a real sea where there was a certain level of water turbidity. Instead of study on performance against turbidity, a previous study [29] in which an image distortion is provided by air bubbles that appear in front of the cameras have been reported. We believe that there is a similar characteristic between turbidity and

air bubbles disturbance that can affect optimization (which is used directly in the feedback in the proposed system). According to the experimental results in [29], 3D pose estimation performance is maintained even though air bubbles exist. This image-disturbance-tolerable performance can be thought that the pose estimation problem is translated into an optimization problem in our approach, which means that the height of a maximum point - that is influenced by lighting condition and turbidity - is not affected by the problem structure of optimization. Apart from above disturbances, robustness against physical disturbance simulating water current was verified in this paper.

3 Novelties of Proposed Approach

This section highlights the novelties of the proposed approach solving the problem statement discussed in Section 2 for this docking application.

3.1 3D Pose Tracking Using Two Cameras and 3D Marker

The main task in this docking experiment is 3D pose tracking in time by following dynamic images with a video rate of 30 frames per second. There is no study that has achieved real-time 3D pose estimation by using dual-eye cameras for AUV in docking operation in which only visual information is directly used in a feedback loop. Therefore, as a main contribution of this paper, a new method of real-time 3D pose estimation in successively input dynamic images from two cameras as shown in Fig. 3 using 3D model-based recognition method utilizing Real-time Multi-step GA is proposed. Moreover, instead of point markers, spheres as 3D marker are used because they have quantifiable diameters, meaning the spheres positioned nearer to the cameras appear as bigger circles in images. Consequently, the spheres reinforce the camera depth estimation. Additionally, spheres are more resilient than points, because an obstacle could more easily hide point markers but not spheres.

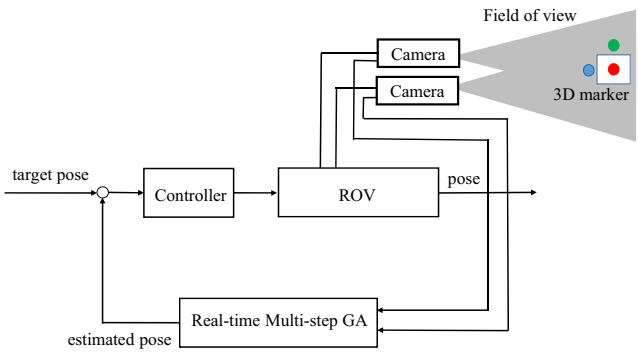


Fig. 3 Block diagram of the proposed 3D-MoS system with Real-time Multi-step GA

3.2 3D Pose Recognition by 3D to 2D Projection

Most of other approaches for object recognition are based on 2D to 3D reconstruction that can address some issues such as expensive computation [27, 28]. In contrast, real-time 3D pose estimation by 3D to 2D projection using known information of a target 3D marker is proposed, avoiding the effects of incorrect pair points, whose problem arises in 2D to 3D reconstruction process used as usual image recognition processes. Contrarily in the proposed 3D model-based pose estimation, 3D models are used in recognition and evaluating is performed using 2D images from left and right cameras. Since 3D to 2D method does not pose incorrect pair problem, pose estimation can be more accurate comparing to other model-based approaches that used 2D models and 2D images. This feature is derived from the simple nature that all points on 3D marker are predefined and corresponding points on 2D images can be projected correctly. Particularly, 3D models in this study are designed to include two areas (inner and outer) for each sphere (red, green and blue) with the concept that was extended from the work in [36]. In [36], different models including surface-strips model were evaluated in recognition in terms of sensitivity, reliability and speed. However, rectangular shape models and images from single camera were used in [36].

3.3 Practical Approach for Docking Application

The overall target of this study is to check the functionality and practicality of our proposed algorithm for an intended docking application. Docking strategy was designed and implemented experimentally. Experiments were conducted in a pool using an ROV to confirm that the proposed approach is able to guide an ROV to insert a rod attached on the ROV into a docking hole with a radius of 35 mm attached with a 3D marker as shown in Fig. 2.

4 3D-MoS with Real-Time Multi-Step GA

Instead of localizing the vehicle and target in an absolute pose in world coordinates to address the requirement of measurements using sensors such as GNSS and INS, localizing the vehicle relative to the target through recognition with a known target's information is implemented in feedback control using standalone dual cameras and a 3D marker.

4.1 Model-Based Pose Estimation Using Dual-Eye Cameras and 3D Marker

In conventional approaches, object recognition including relative pose information is implemented by feature-based recognition based on 2D to 3D reconstruction. The information

of the target object is determined by a set of image points in different images, and the process entails a time-consuming complex search of the corresponding points. A model-based pose estimation approach based on 3D to 2D projection is applied in this work to avoid the effects of incorrect mapping points in images. Both 2D to 3D reconstruction and 3D to 2D projection are shown in Fig. 4. Points B in image 1 and C in image 2 are mapped incorrectly as a pair of points during 2D to 3D point-to-point reconstruction as shown in Fig. 4a. Consequently, the reconstructed 3D point A does not represent a true 3D object. In contrast, points including A and B are correctly projected in group from object in 3D in both images to 2D projection as shown in Fig. 4b. This is possible because the forward projection from 3D to 2D generates unique points in 2D images without errors. Based on this way of thinking, 3D model-based recognition is implemented. Other model-based approaches that are mostly based on template matching have used a 2D model and evaluated 2D images. These kinds of techniques cannot be extended to 3D pose estimation. However, our method is based on the idea of recognition using a 3D model and evaluating 2D images from left and right cameras. The method evaluates how much the 3D model's pose overlaps the actual 3D target, that is, the correlation of the 3D model and the 3D target in 3D space through left and right projected 2D images. This paper experimentally shows the functionality and effectiveness of the chosen strategy, in which simple optimization with short calculation time is applied to dual-eye 3D pose tracking in dynamic images.

Figure 5 shows model-based pose estimation using the dual-eye vision system. The coordinate systems of left and right cameras, an object (solid object) and a model (represented by a dotted box and dotted spheres) in Fig. 5 consist of object coordinate system as Σ_M, i-th model coordinate system as Σ_{M_i}, left and right cameras coordinate systems as Σ_{CL} and Σ_{CR}, and left and right images coordinate systems as Σ_{IL} and Σ_{IR}. Please note that the origins of Σ_M and Σ_{M_i} are the intersections of three axes to which each sphere is attached. A j-th point on a model in 3D space can be described based on each coordinate system using these coordinates and homogeneous transformations. Then, that point can be projected onto the left and right cameras images through a projection matrix. As the information of the 3D marker (color, size, shape) is known, a group of points on the 3D model are defined in a computer system and projected onto the left and right cameras images. In other words, the relation between the arbitrary points on the model and projected points on the left and right images including the 3D marker is evaluated on the 2D image plane by the fitness function with six variables of the pose. In this way, a solid model with its pose (that pose is defined as individual in GA evolution process) is supported to exit in 3D space and projected onto 2D images to match the captured real target in 2D images. By comparing the projected solid model image with the captured 2D images by the dual-eye cameras, the relative pose difference is calculated. The target object is a 3D marker that consists of three spheres (40 mm in diameter) whose colors are red, green, and blue.

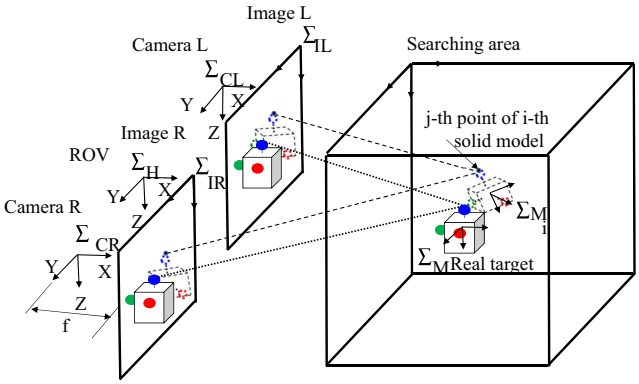


Fig. 5 Model-based pose estimation using dual-eye vision system

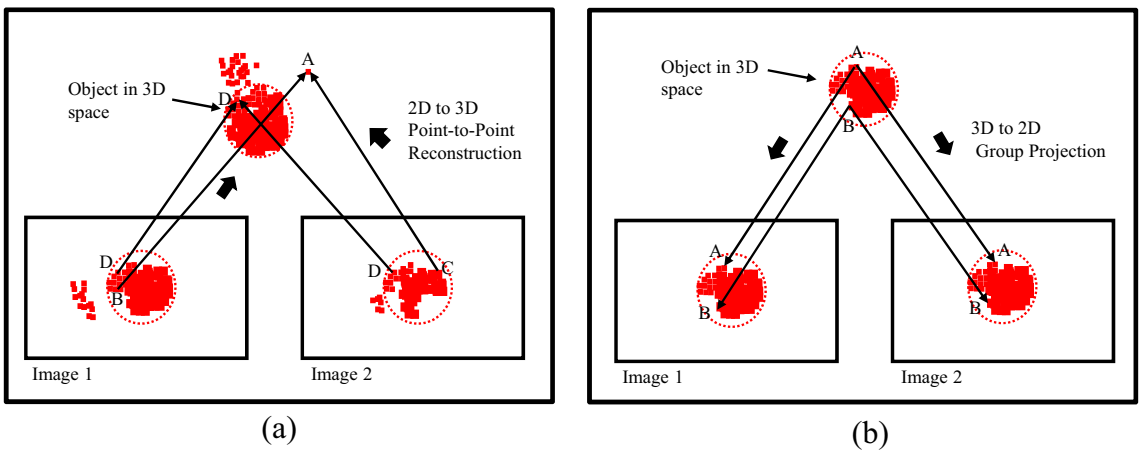


Fig. 4 **a** Mis-mapping in 2D to 3D reconstruction, and **b** Pairing of points in 3D to 2D projection

In the pose estimation process, the main task is to define the number of solid models with different poses within the search space that is defined according to the field of view of the cameras. Then, models that match the target in 2D images to a defined degree are searched for. Finally, the pose of the model that has the highest degree of matching with the target in 2D images is selected as the estimated relative pose.

Regarding the performance of the proposed system in terms of occlusions, we would like to refer to a previous work [14]. In [14], the red and green spheres are occluded individually during a recognition process. According to experimental results in [14], the system can recognize the 3D marker even though one of the three spheres in the 3D marker is occluded. Additionally, some experiments in the case that the 3D marker is occluded by air bubbles that appear randomly in front of the dual-eye cameras were conducted and discussed in [13].

4.2 Real-Time Multi-Step GA

4.2.1 What is Real-Time Multi-Step GA

In the proposed 3D model-based recognition method, searching for all possible models is time consuming for real-time recognition. Therefore, the problem of finding/recognizing the 3D marker and detecting its pose is converted into an optimization problem with a multi-peak distribution, which can be confirmed directly by calculating the distribution of a fitness function against the 3D pose [31, 37]. The highest peak of the fitness distribution is at the value equal to the true pose of the 3D marker. This can be stated in another way: the correlation function used for the fitness function and the target 3D model should be designed to have a dominant peak at the true pose of the target.

The optimization should be completed in the changing condition, for example, dynamic images input by a video rate when visual feedback is required. In this situation, we have two options: (1) A simple optimization method that needs a short time to complete with appropriate accuracy and repeats the optimizing procedure, and (2) Sophisticated optimization methods that might provide better accuracy but require large calculation time. Here the question is which of these options is better for pose tracking in dynamic images. The first one, simple optimization, was chosen for the underwater docking experiment based on its simple logic and effectiveness. The following is one of the reasons for choosing (1). Fast evolution due to shorter life spans, such as mouse evolution, can enable an animal to adapt itself to a changing environment faster than can slower evolution of animals such as an elephant, which can live for several decades with fewer chances to adapt. For example, support that the life span of a mouse is 1 year and that of an elephant is about 80 years. Therefore, the mouse has 80 times more chances in the time domain than does the elephant to adapt to a changing environment.

The discussions on optimization performance in other studies such as [38–40] are based on the speed measured by iteration times. We think that performance evaluation of different methods based on iteration times is unfair because operation time for one iteration of each method in different systems may not be the same. For example, one method may take one hour for one iteration while another method may finish one iteration within one minute. It is difficult to find performance comparison in time domain. According to authors' knowledge, our work is the first one to measure the optimization performance in time domain. On the other hand, other discussions in [41, 42] are concentrated on finding out valuables' numbers to give a maximum function number that are used for the controller's parameters for tuning up the system's performances. In contrast, optimization is directly used in a feedback of the control system in the present paper.

We have developed Real time Multi-step GA, formerly known as the 1-step GA, which can deal with non-differential distribution with a multi-peak, for this underwater docking experiment, although it may not be the best GA in comparison to other optimization methods. We did not compare GA with other optimization methods in this study. Real-time Multi-step GA evolves the chromosomes with as many generations as possible within the video frame rate for each image; in our experimental system, nine generations are possible. The practical performance of the Real-time Multi-step GA was confirmed in a previous work [31]. In [31], the Real-time Multi-step GA was used to estimate the pose of a fish in real time.

4.2.2 How Real-Time Multi-Step GA Works

In the proposed Real-time Multi-step GA, each chromosome encodes 12 bits for each of six parameters: three for position and three for orientation. Figure 6 shows the 3D model-based recognition process in 3D space that evolves through the evaluations of chromosomes by forward projection from the 3D marker onto 2D images. The defined number of chromosomes that represent the different relative poses of the 3D model to the ROV in back-projection is initiated randomly, as shown in the first generation in Fig. 6. Fitter chromosomes as evaluated by a fitness function have a higher chance to be selected for the reproduction of offspring by using designed operators (selection, crossover, and mutation). After reaching the predefined number of generations, the chromosome that has the best fitness value is selected to represent the actual pose of the object. Because the main objective is real-time performance, termination of GA evolution is determined by the video frame rate, which is 30 fps in this work.

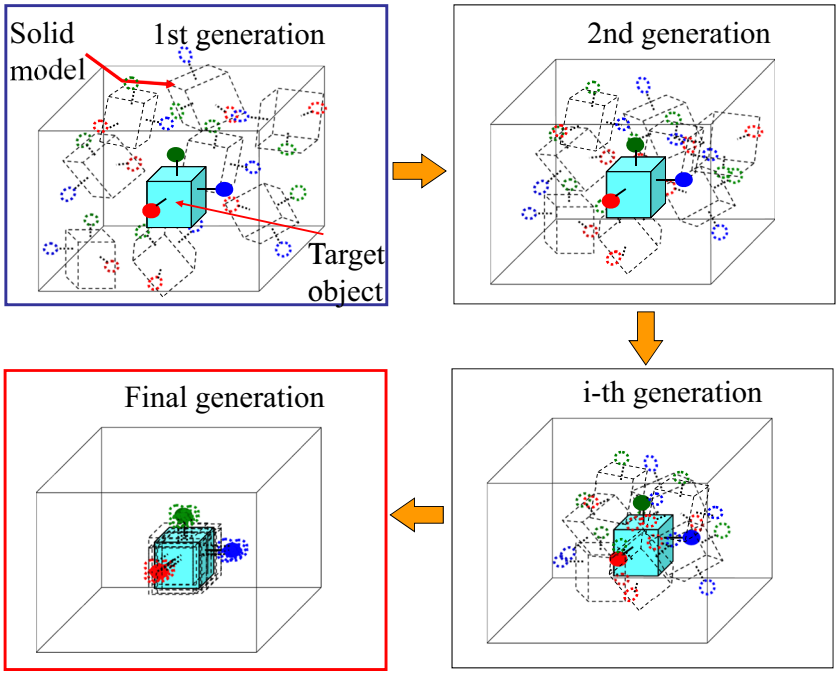


Fig. 6 3D model-based recognition process

A correlation function of the real target projected in camera images with the assumed model, represented by poses in the chromosomes, is used as the fitness function in the GA process. We modified the fitness function based on the voting performance and the target's structure (color, size, and shape). As shown in Fig. 7, two spaces in the model object can have a scored fitness value: the inner space that is the same size as the target sphere and the other space that is the background area. The portion of the captured target that lies inside the inner area of the model will score a higher fitness value and the portion that lies inside the background area will score a lower value. Therefore, the fitness value is maximum when the poses of the target and the model are coincident. Note that the evaluation of models for comparing the real target is in 3D space and the matching target model in terms of the fitness function is done in 2D images. The color information in hue space is used to search for the 3D marker in images, because hue space is less sensitive to the lighting condition [16]. The effectiveness of this method was confirmed in our previous research [33–35]. The time-convergence performance of the Real-time Multi-step GA as a dynamic evaluation function was approved mathematically by a Lyapunov analysis in [30].

Figure 8 shows the flowchart of the Real-time Multi-step GA (right sub figure) and recognition process in 3D space (left sub figure). Please note that a solid model in 3D space

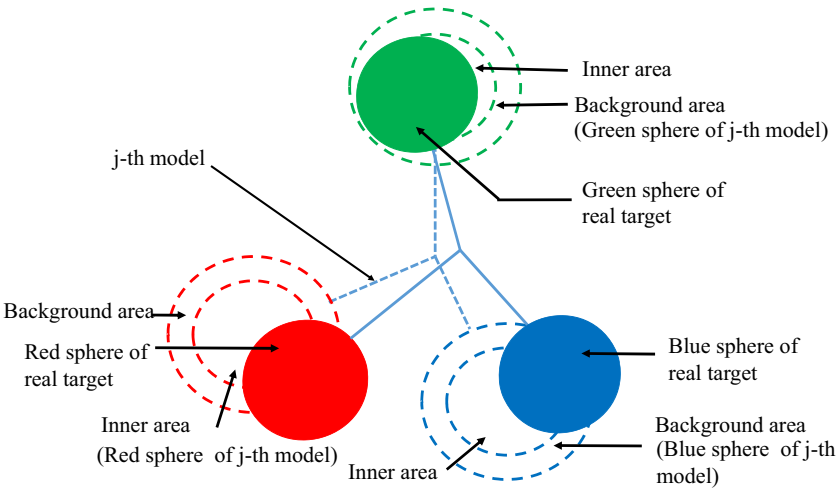


Fig. 7 Target and model object in 2D image from right camera

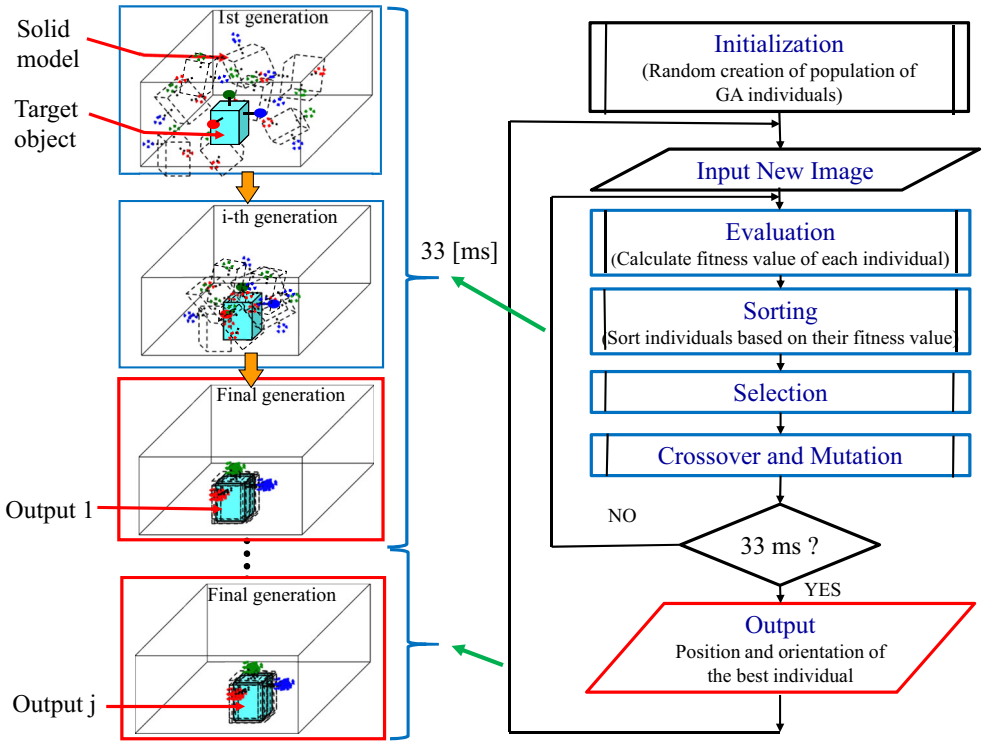


Fig. 8 Flowchart of Real-time Multi-step GA

represents a GA individual. GA operations such as Selection, Crossover and Mutation are performed to reproduce the next generation through evaluation by a fitness function (explained in next section). Several solid models that represent different relative poses converge to the target object through GA evolution process within 33 ms as shown in Fig. 8 (left sub figure). The solid model (Output j in Fig. 8) that represents the true pose with the highest fitness value is searched for every 33 ms. Then, these fit models are forwarded to the next step as the initial models for the next new images in real time.

4.2.3 Fitness Function

A fitness function, which is a shape-based integration/differentiation calculation, is modeled to calculate the correlation between a model and images captured by two cameras using hue value of images. In other words, the intention of the designed fitness function is to have a dominant peak at the true pose of the target. Here is the brief explanation on why designed fitness function is a shape-based integration/differentiation calculation. Since a model has spheres with quantitative diameters rather than a point,

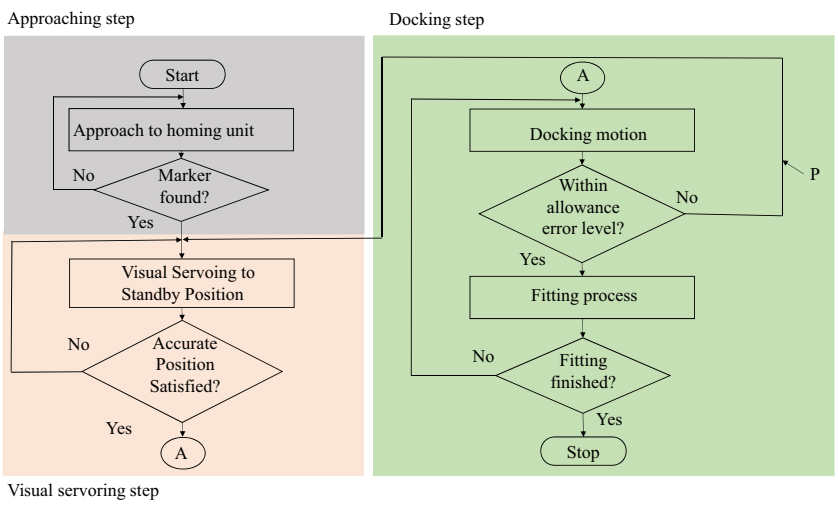


Fig. 9 Flowchart of docking strategy

shape information is used when calculating the correlation between the model and the target object. Therefore, it is said to be a shape-based approach. As a group of image points that lie inside the inner area and outer area (see Fig. 7) of the projected model are evaluated together and added all together in the area, it is said to be in terms of integration. Integration operation can reduce the noise that appears in images like spike noise. To increase the sensitivity, differentiation operation is also considered in the construction of fitness function. The evaluation value is calculated by subtracting the values for points that lie in the outer area from the ones that are overlapped with the inner area of the model. It is therefore said to be in terms of differentiation. The total fitness value is calculated from averaging two fitness functions of the left and right camera. Please note that there is no individual evaluation of left and right images. Finally, the fitness function will have a maximum value when the pose of the searching model fits the one of the target object being imaged in the right and left cameras' images. The evaluation parameters of the objective function (that is fitness function in this study) are designed to reduce the noise (noise in here means some peak points that represent incorrect poses of the target). Please refer to [32, 36] for a detailed definition of the fitness function. The concept of fitness function in this study can be said to be an extension of the work in [36] in which different models including a rectangular shape surface-strips model were evaluated using images from a single camera.

5 Proposed Docking System

5.1 Docking Strategy

The proposed docking strategy consists of three steps. First, the ROV has to approach the 3D target until the target is in its field of view. Second, detecting the object and regulating the vehicle to the defined relative pose of the target is performed in the visual servoing step. Third, the docking operation is completed. The flowchart of the docking strategy is shown in Fig. 9. The originality of this work is concentrated on the dual-eye visual servoing as a possible new docking strategy rather than conventional docking methods. Therefore, the main contribution in the present paper is focused on the second and the third steps of Fig. 9 to demonstrate the effectiveness of the proposed docking system.

The first step can be extended for real-world application by using a long-distance navigation sensor to guide the vehicle into the field of view of the cameras. In [23], a state machine was proposed to generate a waypoint around the estimated target position and inside the vehicle's field of view, but that discussion was limited to the approaching step

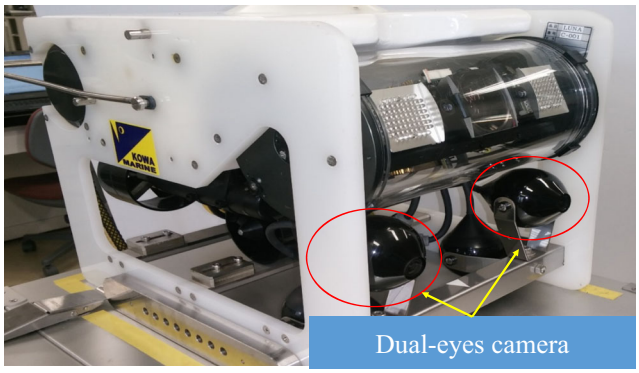


Fig. 10 Photograph of ROV

in Fig. 9. In this study, after approaching with constant speed and a constant proceeding direction while trying to detect the 3D marker, the vehicle is stabilized in the visual servoing step and controlled to keep the ROV with a defined pose relative to the target. In the docking step, when the vehicle is stable within the tolerance of the position error for the defined time period, the forward thrust that enables the docking pole attached to the ROV to fit into the dock is generated by gradually decreasing the distance between the vehicle and the target object. Switching between the visual servoing mode and the docking mode by using the continuous pose feedback in the docking strategy makes the system robust with little surfacing of the dock and minimizes the mechanical aspect as well.

5.2 Underwater Vehicle

The remotely controlled underwater robot (Kowa, maximum depth 50 m) used in this experiment is shown in Fig. 10. Two fixed forward cameras with the same specifications (imaging element CCD, pixel number 640 × 480, pixel focal length 2.9 mm, signal system NTSC, minimum illumination 0.8 lx, no zoom) are mounted on the ROV. These two fixed cameras are used for 3D object recognition in visual servoing. The thruster system of the

Table 1 Specification of ROV

Items	Specification
Max. operating depth [m]	50
Dimensions [mm]	280 (W) × 380 (L) × 310 (H)
Dry weight [kg]	15
Number of thrusters	2 (horizontal), 1 (vertical), 1 (traverse)
Number of cameras	2 (front, fixed)
Number of LED lights	2 (5.8 W)
Tether cable [m]	200
Structural materials	Aluminum alloy and acrylate resin
Maximum thrust force [N]	9.8 (horizontal), 4.9 (vertical, traverse)

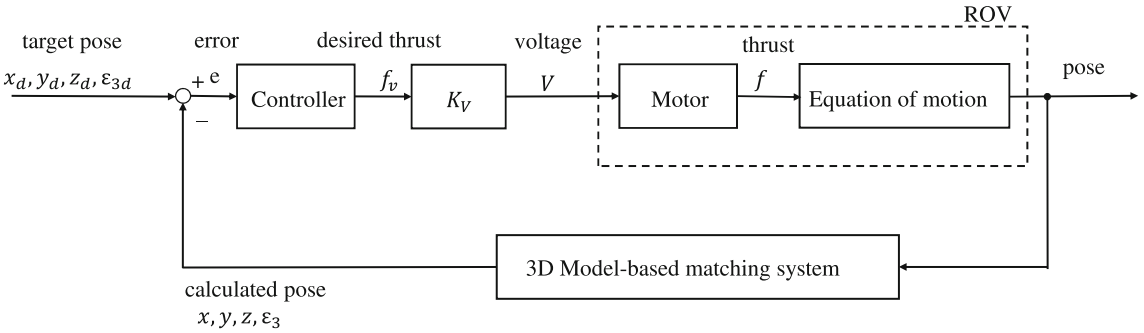
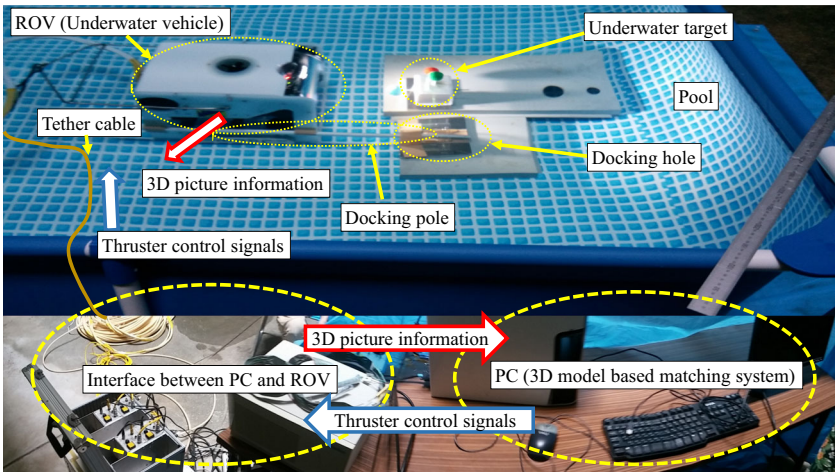


Fig. 11 Block diagram of proposed system for automatic docking

Fig. 12 Layout of underwater experimental devices



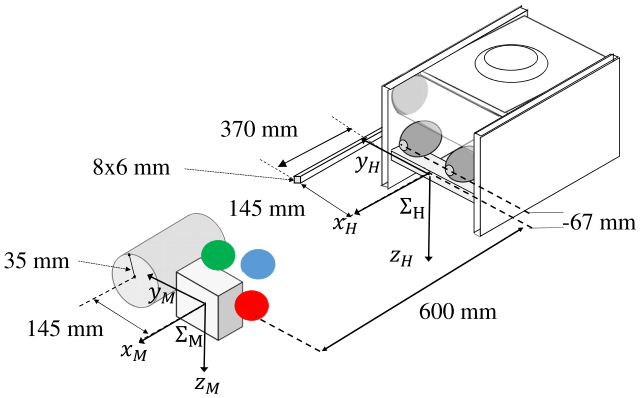


Fig. 13 Coordinate system in docking layout

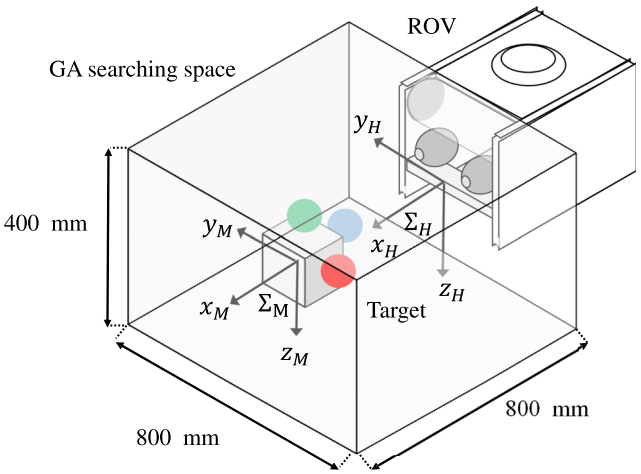


Fig. 14 Underwater target and GA searching space

ROV consists of two horizontal thrusters with a maximum thrust of 4.9 N each, and one vertical thruster and one lateral thruster with a maximum thrust of 4.9 N each. In addition, the ROV is equipped with two units of LED lights (5.8 W) as the illumination source. The specifications of the main hardware components are summarized in Table 1.

5.3 Experimental System

The block diagram of the proposed control system is shown in Fig. 11. The images acquired from the dual-eye cameras are sent to the PC. Then, the real-time recognition of the 3D pose of the target object by 3D-MoS and the Real-time Multi-step GA is executed by software installed in the PC. Finally, based on the error between the target value and the recognized value, command signals generated from calculating the voltage value gained by the P controller for the thrusters are input into the ROV. Rolling and pitching orientation controls are neglected in this experiment due to their self-stabilization characteristic, and the rolling and pitching stabilizing transition motions were confirmed to not influence to the other controlled axes. Therefore, ϵ_1 and ϵ_2, which represent the roll angle and the pitch angle, respectively, are neglected in the control of the vehicle. The 3D model-based matching process is executed within 33 ms, as shown in the Real-time Multi-step GA in Fig. 8, by synchronizing the video frame rate of the dual-eye cameras. The specifications of the PC are Intel Core(TM) i7-3770 CPU @ 3.40 GHz, RAM 8.00 GHz, system type 64 bits. Two interfacing boards, PCI 5523, are used in the PC to receive images from the two cameras. To output the control voltage from the PC to the ROV, PCI 3343A is installed in the PC as a digital-to-analog converter in this experiment.

A pool (2 m (L) × 3 m (W) × 0.75 m (H)) filled with tap water was used as an experimental tank for the underwater vehicle experiments. The ROV was tethered by a cable 200 m in length to receive image information and output control signals, as shown in Fig. 12. To perform experiments that simulate underwater automatic charging, a rod on the right side of the underwater robot and a cylinder hole on the left side of the target were designed as shown in Figs. 12 and 13. Figure 13 shows the alignment process between the rod installed on the ROV and the cylinder of the dock station when the ROV is stable enough to insert the rod into the docking hole for the experiments of this study. When the robot is in the correct relative pose to the 3D marker, it has to move ahead to insert the rod into the cylinder hole. To verify the stability against disturbances, abrupt external forces were applied to the vehicle by pushing the vehicle in different directions with a rod while visual servoing.

5.4 Experimental Conditions

5.4.1 Real-Time Multi-Step GA

The relative pose estimation by 3D model-based recognition is assumed to be executed in the GA search area set in front of the underwater robot, as shown in Fig. 14. We considered the visibility range in real seawater as about 1 m. Therefore, the searching space is defined as shown in Table 2. The searching space depends on the camera lens specification, which has a focal length of 2.9 [mm]. Table 2 shows the conditions of the GA.

Table 2 Parameters of real-time multi-step GA

Items	Specification
Number of genes	60
Evolved pose (position and orientation)	$(x, y, z, \epsilon_1, \epsilon_2, \epsilon_3)$, all genes are coded by binary 12 bits
	$(\epsilon_1, \epsilon_2, \epsilon_3)$ are represented by quaternion
Pose used for controlling	Position (x mm, y mm, z mm)
(Position, Orientation)	Orientation (ϵ_3) around z-axis of Σ_H in Fig. 13
Searching space defined by Σ_H in Fig. 13	{x,y,z,ϵ_3}={±400, ±400, ±200, ±0.15 (equal to $\pm 17.3 degree$)}
Control period [ms]	33
Number of gene evolution [times/33ms]	9
Selection rate [%]	60
Mutation rate [%]	10
Crossover	Two-point
Evolution strategy	Elitism preservation

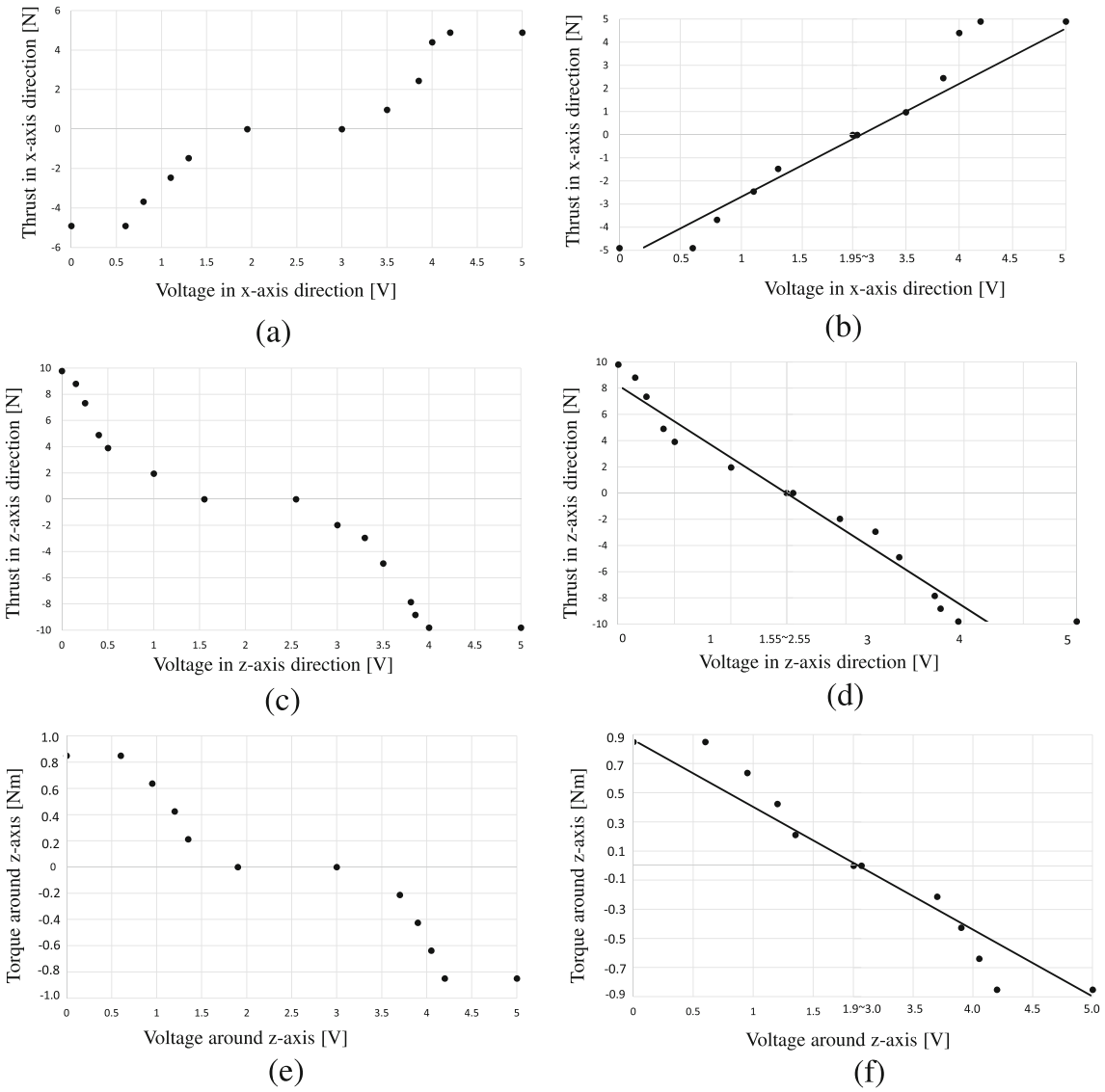


Fig. 15 Initial characteristics of thrust and torque control voltage and those adjusted by removing the dead band and linearization: **a** initial characteristics in x-axis direction, **b** characteristics after removing dead band (black dots) and adjusting (solid line) in x-axis direction, **c**, **d** characteristics in z-axis direction, and **e**, **f** characteristics around z-axis

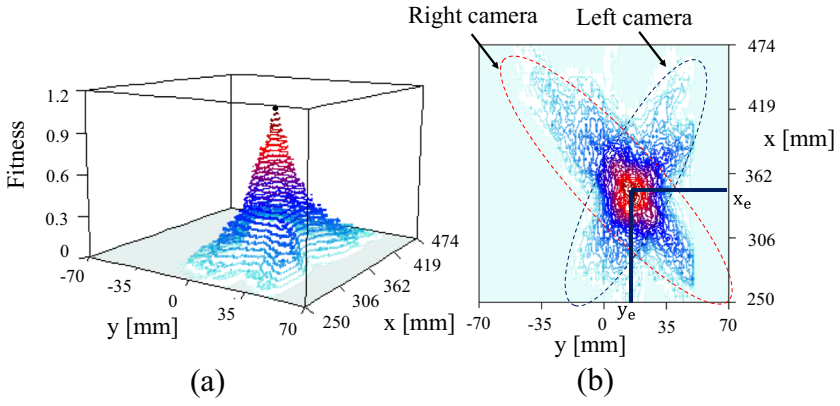


Fig. 16 Fitness value of GA candidates in x-y position: **a** side view **b** top view

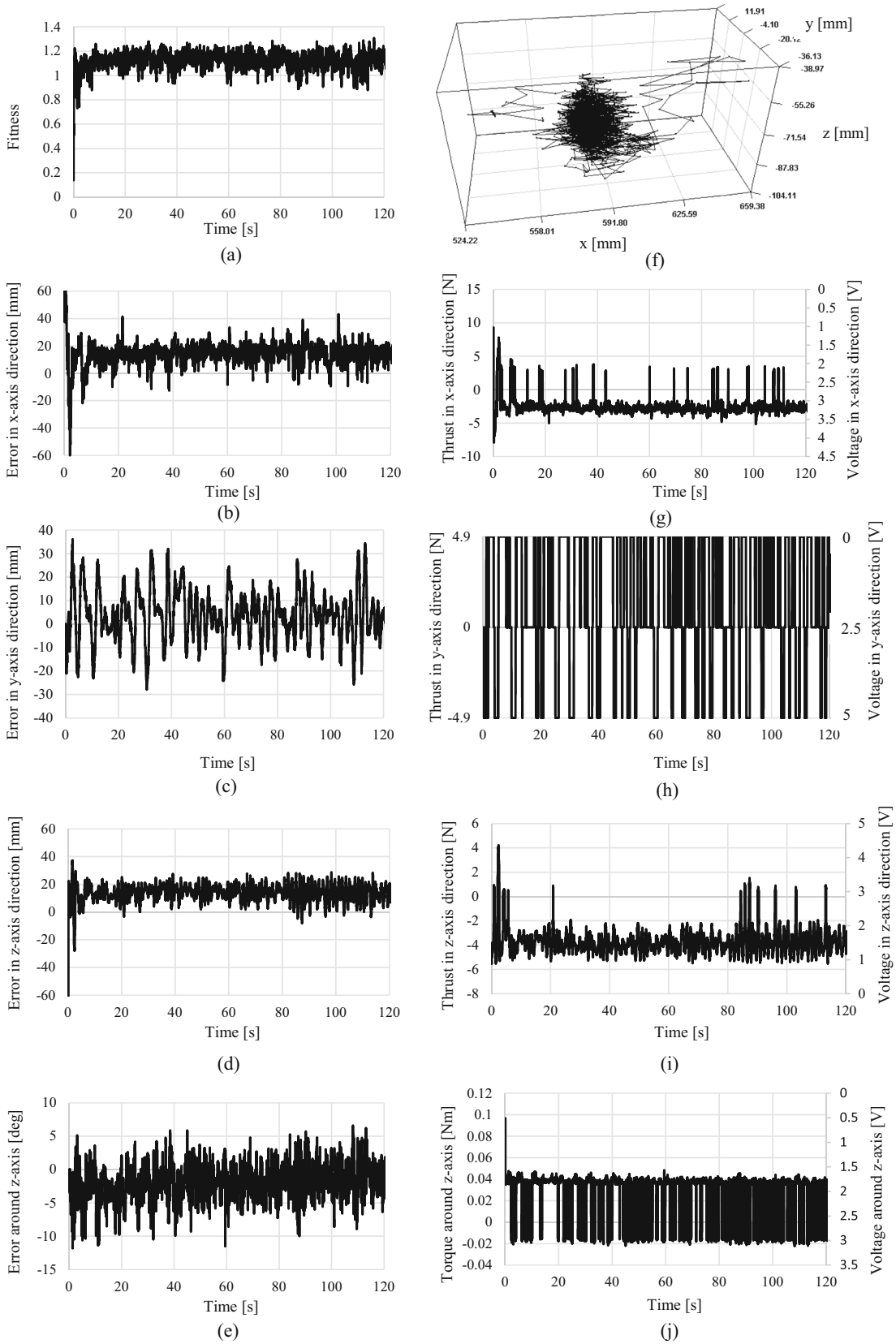


Fig. 17 Regulating performance without additional disturbance: **a** fitness value, **b** error in x-axis direction, **c** error in y-axis direction, **d** error in z-axis direction, **e** error around z-axis, **f** 3D trajectory of underwater vehicle, **g** thrust in x-axis direction, **h** thrust in y-axis direction, **i** thrust in z-axis direction, and **j** torque around z-axis

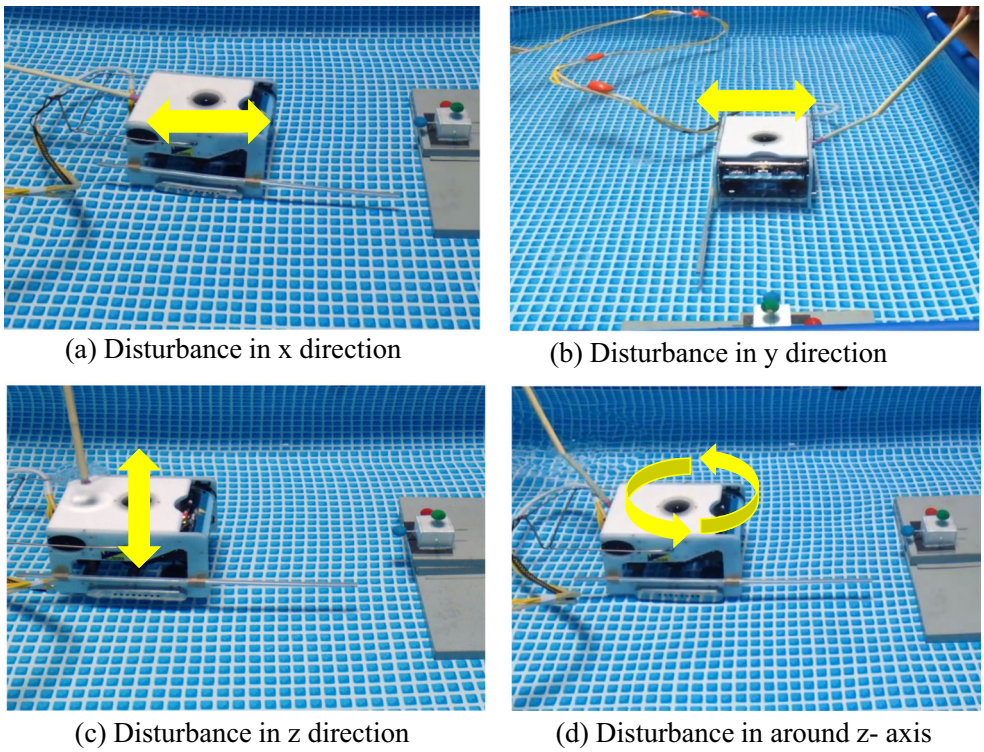


Fig. 18 Additional disturbance directions for regulating performance test: **a** disturbance in x-axis direction, **b** disturbance in y-axis direction, **c** disturbance in z-axis direction, and **d** disturbance around z-axis

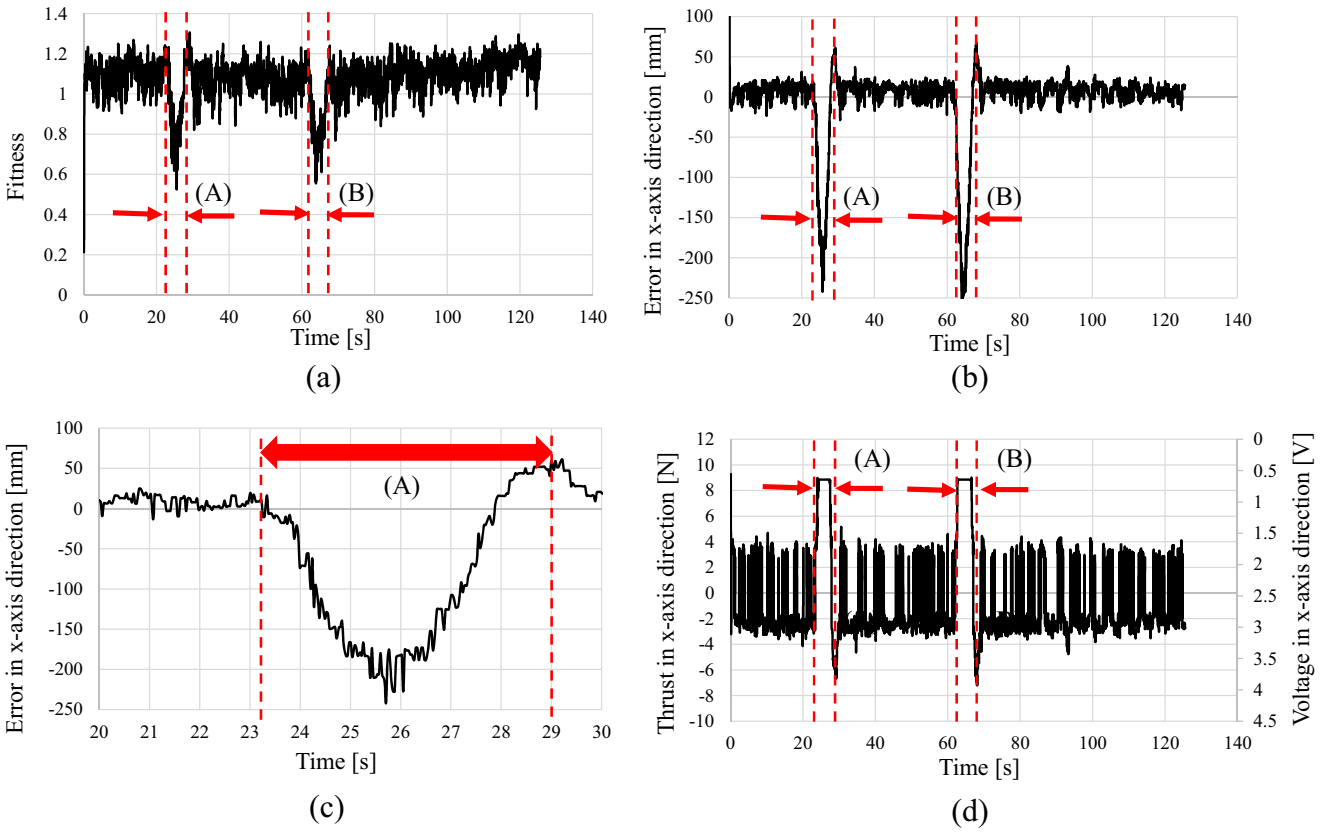


Fig. 19 Regulating performance with disturbance in x-axis direction: **a** fitness value, **b** error in x-axis direction, **c** error in x-axis direction (enlarged view from 20 s to 30 s), and **d** thrust in x-axis direction

5.4.2 Desired Pose

The following relative pose between the ROV and the 3D marker (x_d [mm], y_d [mm], z_d [mm], ϵ_{3d} [deg]) is controlled according to the visual servoing step in Fig. 9.

$$x_d = {}^{H}x_M = 600\ (350)\,mm,\quad y_d = {}^{H}y_M = 0\ (0)\,mm,$$
$$z_d = {}^{H}z_M = -67\ (-67)\,mm,\ \epsilon_{3d} = 0\ (0)\,deg$$

Each number in the above formulas is a target value for regulating the underwater robot immediately after recognizing the object in the visual servoing step in Fig. 8. ${}^{H}x_M$ represents the x position of the origin of Σ_M in reference to Σ_H, where Σ_H and Σ_M are defined in Fig. 14. It should be noted that the numbers in parentheses are the defined target values at the time of completion of the fitting in the docking experiment.

5.5 Controller

To regulate the underwater robot with the relative pose to the target, the following command voltage values v_1 to v_4 calculated by proportional control were fed to the respective thrusters. Even though the PD controller looks more efficient when the damping characteristic of water is not sufficient for stabilizing control, it was found experimentally that the damping characteristic was sufficient when using the P controller. Therefore, a simple P controller was used with appropriate gains instead of comparing the gains of different controllers such as PD and PID.

$$\begin{array}{ll} Back and Forth: & v_1 = k_{p1}(x_d - x) + 2.5 \\ Direction & (v_1 = 0\,V\ for\ thrust\ 9.8\,N\ in\ x_H \\ (x_H\ axis\ in\ Fig.13) & of\ \Sigma_H,\ v_1 = 5\,V\ for\ -9.8\,N) \end{array} \quad (1)$$

$$Left and right: \quad Direction \quad (y_H\ axis\ in\ Fig.13) \qquad v_2 = \begin{cases} 5\,V & ((y_d - y < -5\ \text{mm}) \\ & \text{for thrust in } y_H \text{ of} \\ & \Sigma_H = -4.9\text{N}) \\ 2.5\,V & ((-5 \le y_d - y \le 5) \\ & \text{meaning thrust} \\ & \text{is equal to zero}) \\ 0\,V & ((y_d - y > 5\ \text{mm}) \\ & \text{for thrust in } y_H \text{ of} \\ & \Sigma_H = 4.9\text{N}) \end{cases} \quad (2)$$

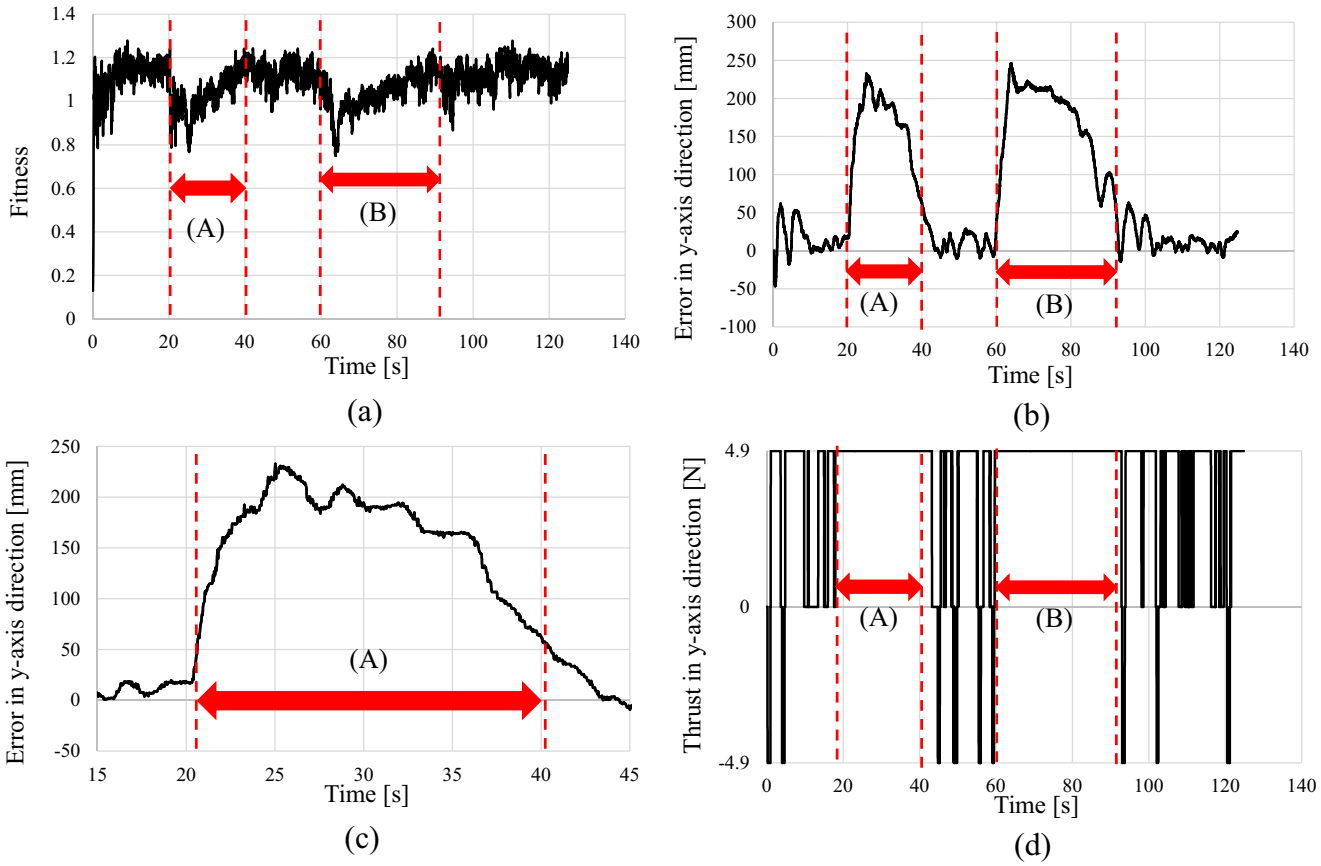


Fig. 20 Regulating performance with disturbance in y-axis direction: **a** fitness value, **b** error in y-axis direction, **c** error in y-axis direction (enlarged view from 15 s to 45 s), and **d** thrust in y-axis direction

$$\text{Rotation } (around\ z_H\ axis\ in\ Fig.\ 13): \quad v_3 = k_{p2}(\epsilon_{3d} - \epsilon_3) + 2.5 \quad (3)$$
$$(v_3 = 0\ \text{V for}\ \Sigma_H = 0.882\ \text{Nm in}\ z_H,\ v_3 = 5\ \text{V for}\ \Sigma_H = -0.882\ \text{Nm in}\ z_H)$$

$$\text{Vertical direction } (z_H\ axis\ in\ Fig.\ 13): \quad v_4 = k_{p3}(z_d - z) + 2.5 \quad (4)$$
$$(v_4 = 0\ \text{V for}\ \Sigma_H = -4.9\ \text{N in}\ z_H,\ v_4 = 5\ \text{V for}\ \Sigma_H = 4.9\ \text{N in}\ z_H)$$

where v_1 is the input voltage for each of the two horizontal thrusters for movement of the ROV in the back and forth direction (x_H in Fig. 14); v_2 is the input voltage for the traverse thruster for movement of the ROV in the left and right direction (y_H in Fig. 14); v_3 is the input voltage for thrusters for rotation movement of the ROV around z_H in Fig. 14); and v_4 is the input voltage for the vertical thruster for movement of the ROV in the vertical direction (z_H in Fig. 14). Note that the rotation of the vehicle is controlled by the two horizontal thrusters that rotate in opposite directions. In future work, a P or PD controller will be considered for all thrusters.

6 Results and Discussion

6.1 Thruster Output Control

For manual operation by a joystick in the remote-operated mode, the joystick operation has to have a certain amount of dead zone in order to prevent a malfunction due to the operating motion of the human finger. In this study, when approaching the object by thruster propulsion, the realization of the movement as well as the orientation control performance of accuracy should be at the millimeter level. The dead band characteristics might cause oscillation of the vehicle when they are used directly in the control system for visual servoing. Therefore, the dead band characteristics of the ROV that were confirmed in preliminary experiments in Fig. 15a, c, e are eliminated, as shown by solid lines in Fig. 15b, d, f. However, the saturations are not compensated for by assuming that the fringe areas of 0 V and 5 V, as shown in Fig. 15b, d, f, are not important during stabilizing and docking. The settings shown in Fig. 15b, d, f for this experiment were adopted because for the sake of experimental effectiveness and simplification.

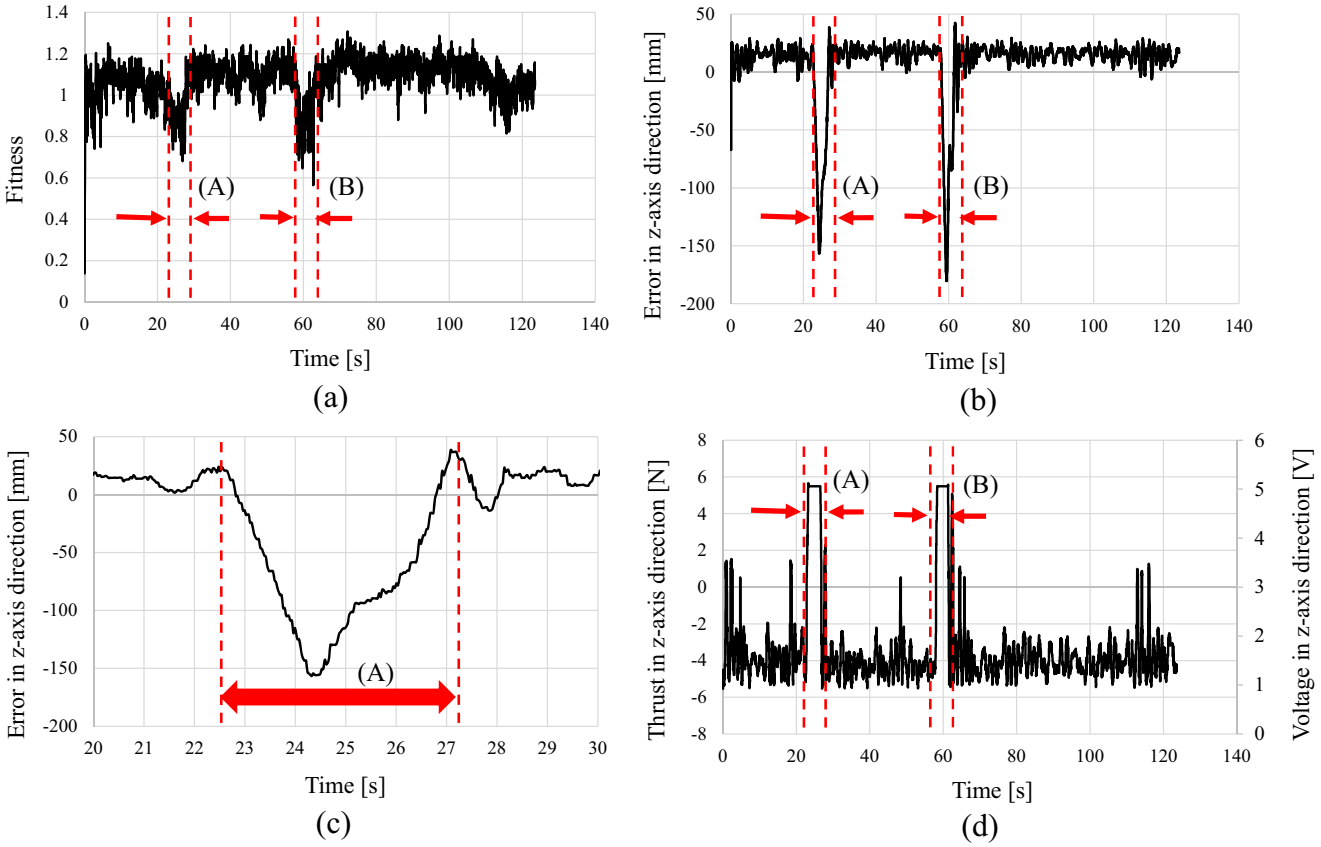


Fig. 21 Regulating performance with disturbance in z-axis direction: **a** fitness value, **b** error in z-axis direction, **c** error in z-axis direction (enlarged view from 20 s to 30 s), and **d** thrust in z-axis direction

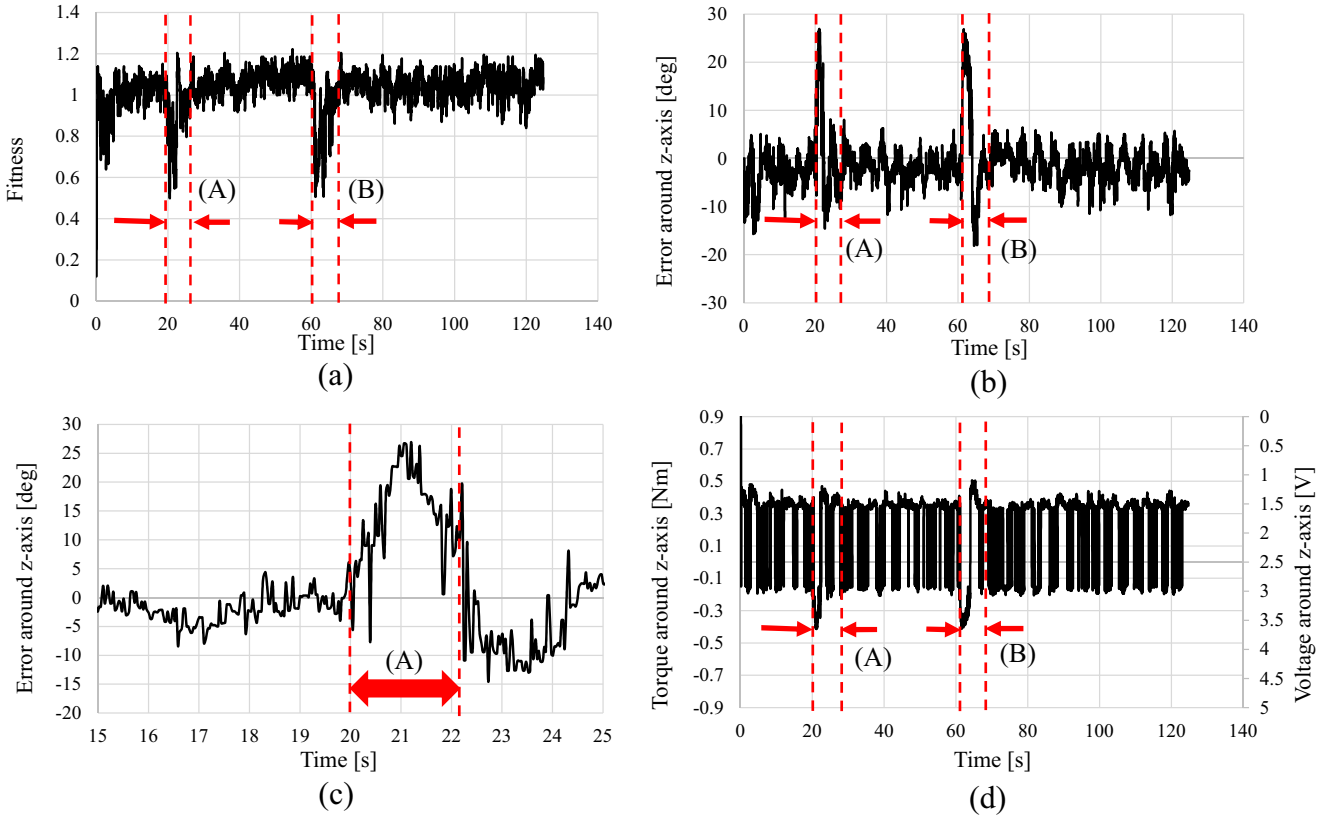


Fig. 22 Regulating performance with disturbance around z-axis: **a** fitness value, **b** error around y-axis, **c** error around y-axis (enlarged view from 15 s to 25 s), and **d** torque around z-axis

6.2 Recognition Accuracy of Real-Time Multi-Step GA

Figure 16 shows the fitness value distribution of all candidates that represent different poses at the sampled time (here, only the x-y position is shown, where x_e and y_e represent the estimated x and y positions, respectively). This distribution of the fitness function indicates that the pose estimation problem of the 3D marker has been converted into an optimization problem to find the values giving the highest peak with the constraint of real-time

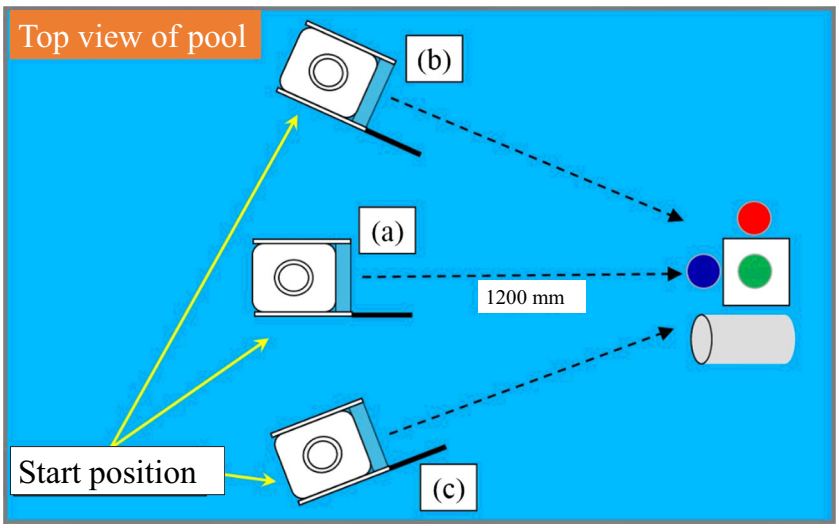


Fig. 23 Start position of underwater vehicle: **a** in front of 3D marker, **b** on the left side of pool relative to 3D marker, and **c** on the right side of pool relative to 3D marker

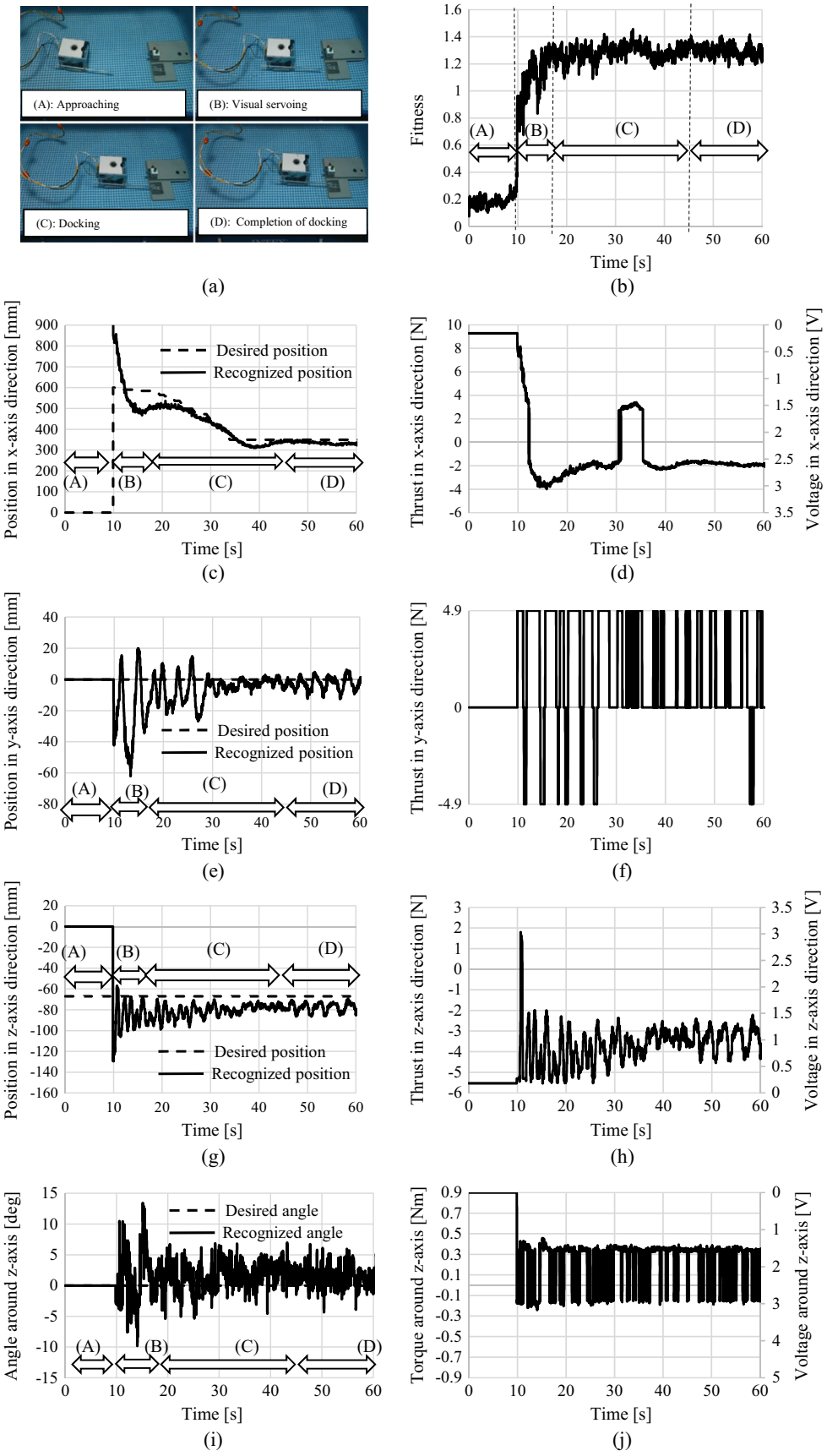


Fig. 24 Docking experimental results for start position of underwater vehicle in front of 3D marker, position (**a**) in Fig. 23: **a** photo of docking experiment, **b** fitness value, **c** position in x-axis direction, **d** thrust in x-axis direction, **e** position in y-axis direction, **f** thrust in y-axis direction, **g** position in z-axis direction, **h** thrust in z-axis direction, **i** angle around z-axis, and **j** torque around z-axis

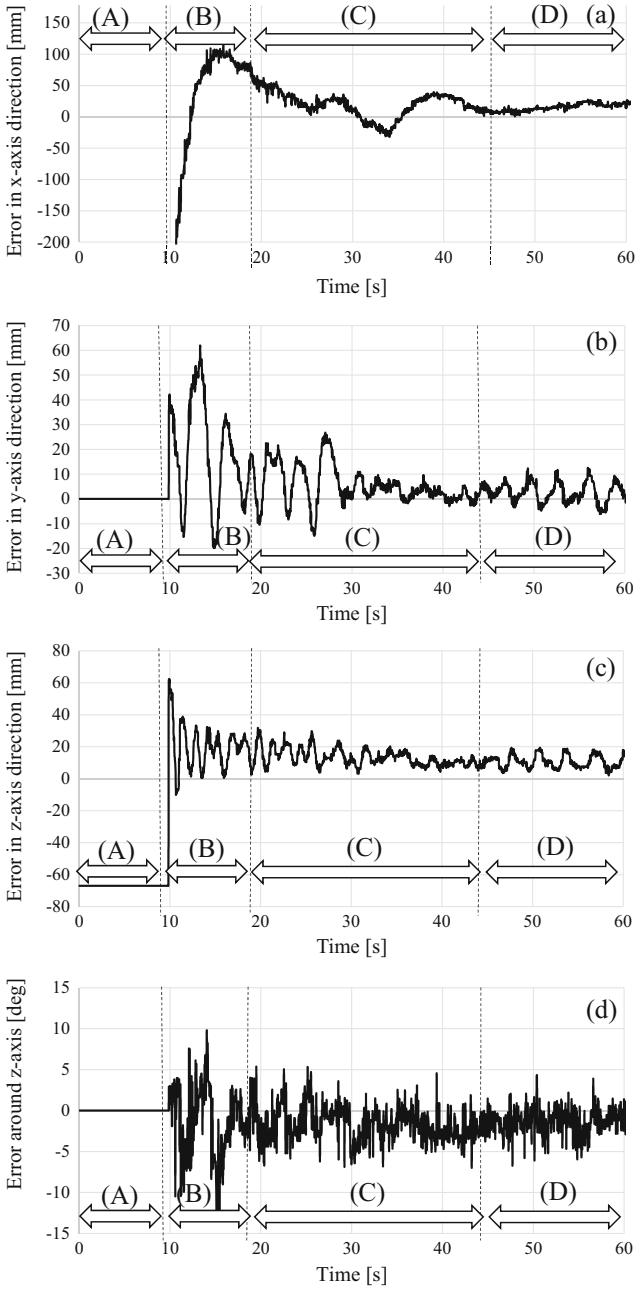


Fig. 25 Further docking experimental results for start position of underwater vehicle as in front of 3D marker, position (**a**) in Fig. 23: **a** error in x-axis direction, **b** error in y-axis direction, **c** error in z-axis direction, and **d** error around z-axis

convergence within the video frame rate. The peak value of the fitness values from the intersection of two images from the dual-eye cameras, as shown in Fig. 16b, provides the recognition accuracy. According to the experimental result, the error of pose estimation is less than 5 mm. This experimental result highlights the main benefit, which is the accuracy, of stereo vision in 3D pose estimation.

Figure 17a shows the time variation of the fitness value at the time of Real-time Multi-step GA recognition of the underwater robot that was regulated at $x_d = 600\,mm$, $y_d = 0\,mm$, $z_d = -67\,mm$, $\epsilon_{3d} = 0\,deg$. It can be seen that the fitness value is maintained above 0.8 within a few seconds from the recognition start. According to the experimental results, the minimum fitness value for good recognition performance is 0.5 when performing visual servoing with the proposed system.

6.3 Regulating Performance

The regulating performance without physical disturbances is shown in Fig. 17. Figure 17a is the fitness value recognized by the Real-time Multi-step GA, where the fitness value is kept above 1.0. This result means that the model and the real 3D marker match well, as shown in Figs. 6 and 8. Figure 17f shows the position of the ROV in the regulating performance measured in the Real-time Multi-step GA. Figure 17b to e represents the errors between the desired and estimated relative pose of the target position. Figure 17g to j shows the torques and control voltages to restore the ROV to be regulated in the desired pose. According to the experimental results, the ROV can be regulated to be in the desired pose by using the proposed system within a ±20 mm position error.

6.4 Stability Against Disturbance

To verify the stabilizing ability of the proposed system against disturbances caused by collision or ocean currents, for example, the regulating performance was evaluated to determine whether the proposed system can restore the relative pose to the target by applying external forces in the x direction (Fig. 18a), y direction (Fig. 18b), z direction (Fig. 18c), and around the z-axis (Fig. 18d) manually by using the rod (wood, full length 2 m) from outside the experimental pool. It should be noted that the disturbance here pushed the robot to move 150 to 200 mm between 1.5 and 2.0 s in the x, y, and z directions and to rotate 15 deg per 1 s around the z-axis.

The regulating performance with a disturbance in each direction is shown in Figs. 19, 20, 21 and 22. The term "stability" in this paper means the property in which the underwater robot can be restored to the target pose relative to the 3D marker, even when a disturbance is given to the ROV. Figures 19 to 22 show (a) the fitness of GA recognition, (b) the error between the relative pose between the target object and the underwater robot recognized for each variable, and (c) the same results of (b) enlarged during a disturbance. Furthermore, (d) represents the thrust force (torque) applied to the thruster. The disturbance was applied twice around 20 s and 60 s from the beginning of

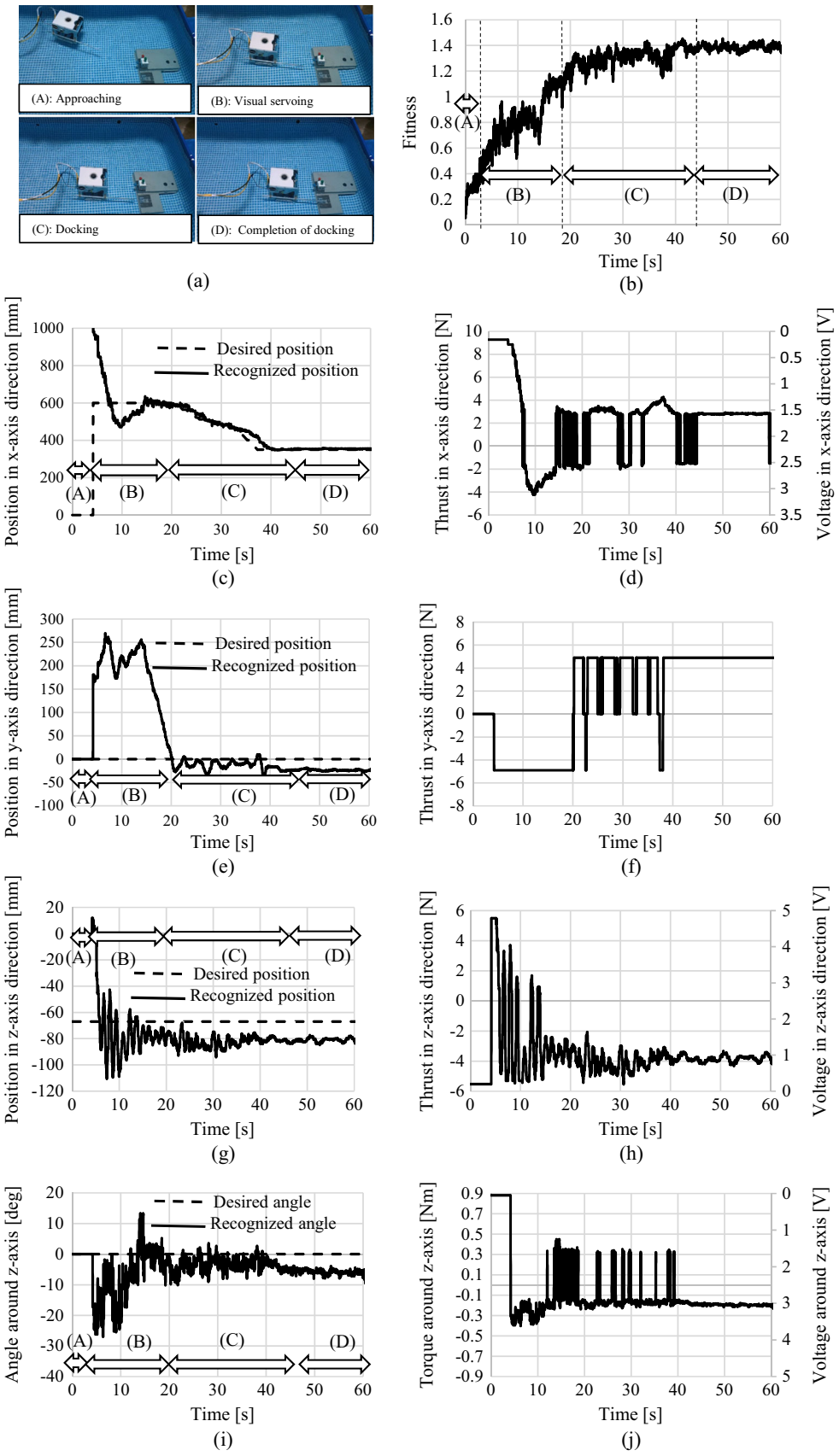


Fig. 26 Docking experimental results for start position of underwater vehicle on the left side of pool relative to 3D marker, position (**b**) in Fig. 23: **a** photo of docking experiment, **b** fitness value, **c** position in x-axis direction, **d** thrust in x-axis direction, **e** position in y-axis direction, **f** thrust in y-axis direction, **g** position in z-axis direction, **h** thrust in z-axis direction, **i** angle around z-axis, and **j** torque around z-axis

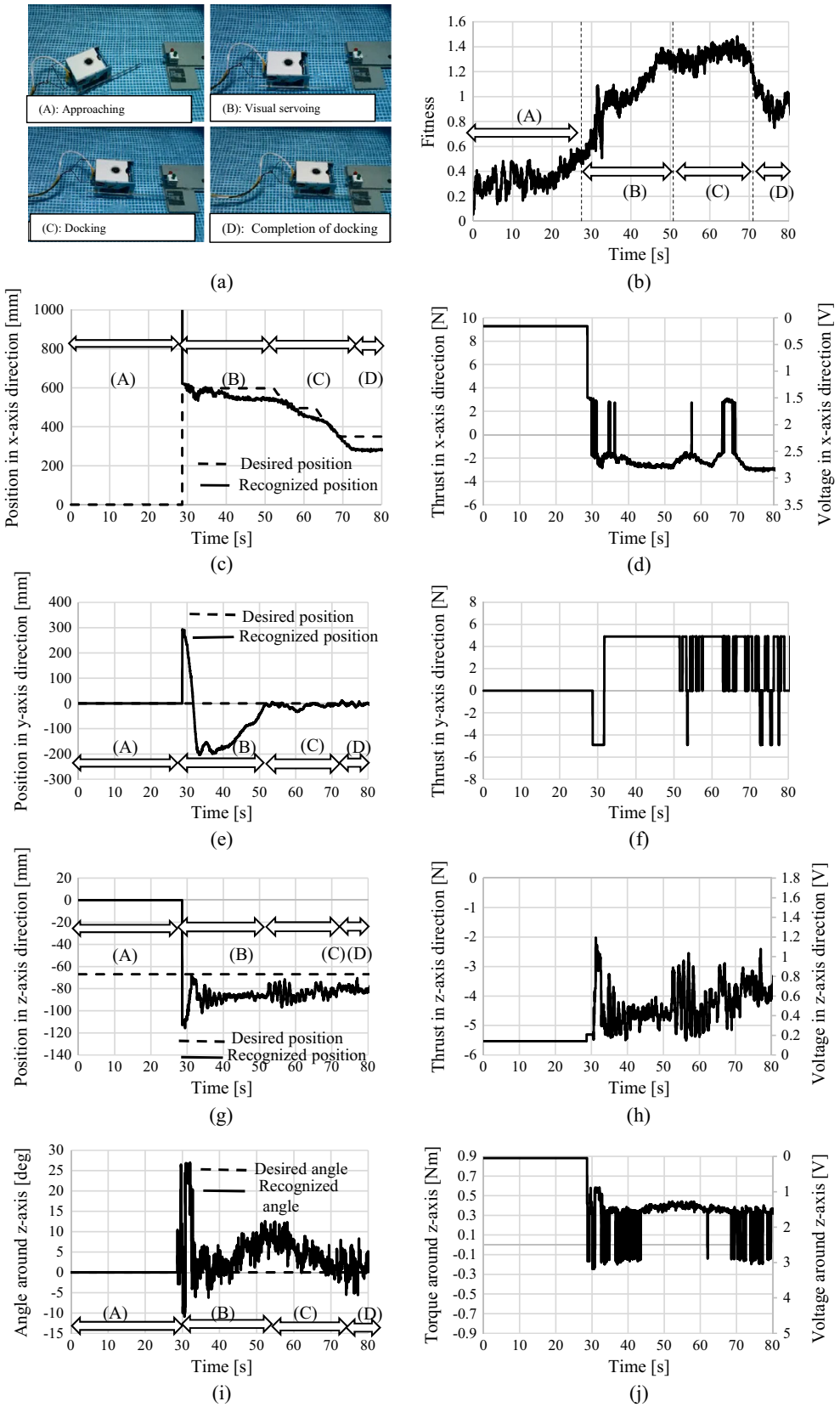


Fig. 27 Docking experimental results for start position of underwater vehicle on the right side of pool relative to 3D marker, position (**c**) in Fig. 23: **a** photo of docking experiment, **b** fitness value, **c** position in x-axis direction, **d** thrust in x-axis direction, **e** position in y-axis direction, **f** thrust in y-axis direction, **g** position in z-axis direction, **h** thrust in z-axis direction, **i** angle around z-axis, and **j** torque around z-axis

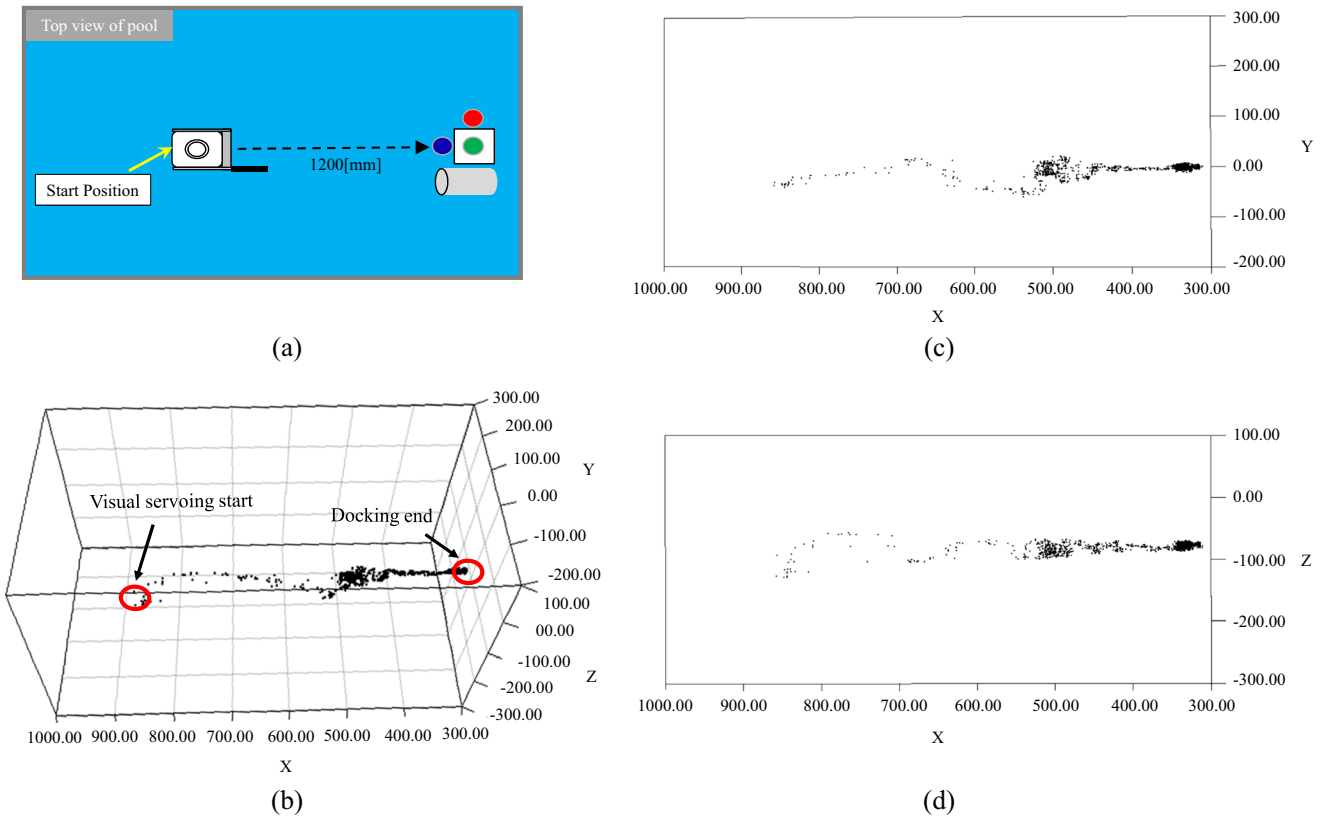


Fig. 28 Recognized trajectory for start position of underwater vehicle in front of 3D marker, position (**a**) in Fig. 23: **a** Start position of underwater vehicle, **b** recognized position in 3D by Real-time Multi-step GA, **c** recognized position in x-axis and y-axis, and **d** recognized position in x-axis and z-axis

the experiment. In the section shown by (A) and (B) in (a), (b), (c), it is found that varying the thrust (torque) applied to the thrusters in response to an error of the relative target pose maintains the relative pose during visual servoing, although the fitness is temporarily lowered when a disturbance is applied. In other words, it is possible to confirm that an operation for correcting the error has occurred and consequently will change the pose of the ROV to restore the relative target pose. From the above results, the proposed system could restore the robot to the original position within a few to several tens of seconds for all of these disturbances.

6.5 Docking Experiments

Experiments were carried out with different start positions: (a) in front of the 3D marker, (b) on the left side of the pool relative to the 3D marker, and (c) on the right side of the pool relative to the 3D marker, as shown in Fig. 23. The docking experiments were carried out as shown in Figs. 24, 26, and 27 following the four states (A) approaching the object (approach), (B) visual servoing to keep the relative pose to the object (visual servoing), (C) fitting to the fixed homing unit (docking), and (D) fully fitting into the homing unit (completion of docking) as shown in Fig. 24a.

In the approaching step (A) in Figs. 24, 26, and 27, the robot's speed is low. This is the state until the underwater robot finds the 3D marker (recognition) under the assumption that the object is presented in front of the underwater robot. In other words, the underwater robot does not know the relative pose to the object in the initial condition, and then goes forward and transits to the state of visual servoing after discovering the object (judged by the fitness function rising to 0.5). Then, visual servoing is the state in which the underwater robot is regulated in the desired pose. As described above, the control process is performed to maintain the robot in the relative pose to the 3D marker. After transition to this state, the underwater robot moves forward. The relative target position x_d decreases by 30 mm/s in the x_H-axis direction when the error of the relative position of the robot (y_d and z_d) with respect to the object is stable within ± 20 mm for the minimum period of 165 ms (control loop $\times$ 5). Then the docking state performs the fitting to the homing unit.

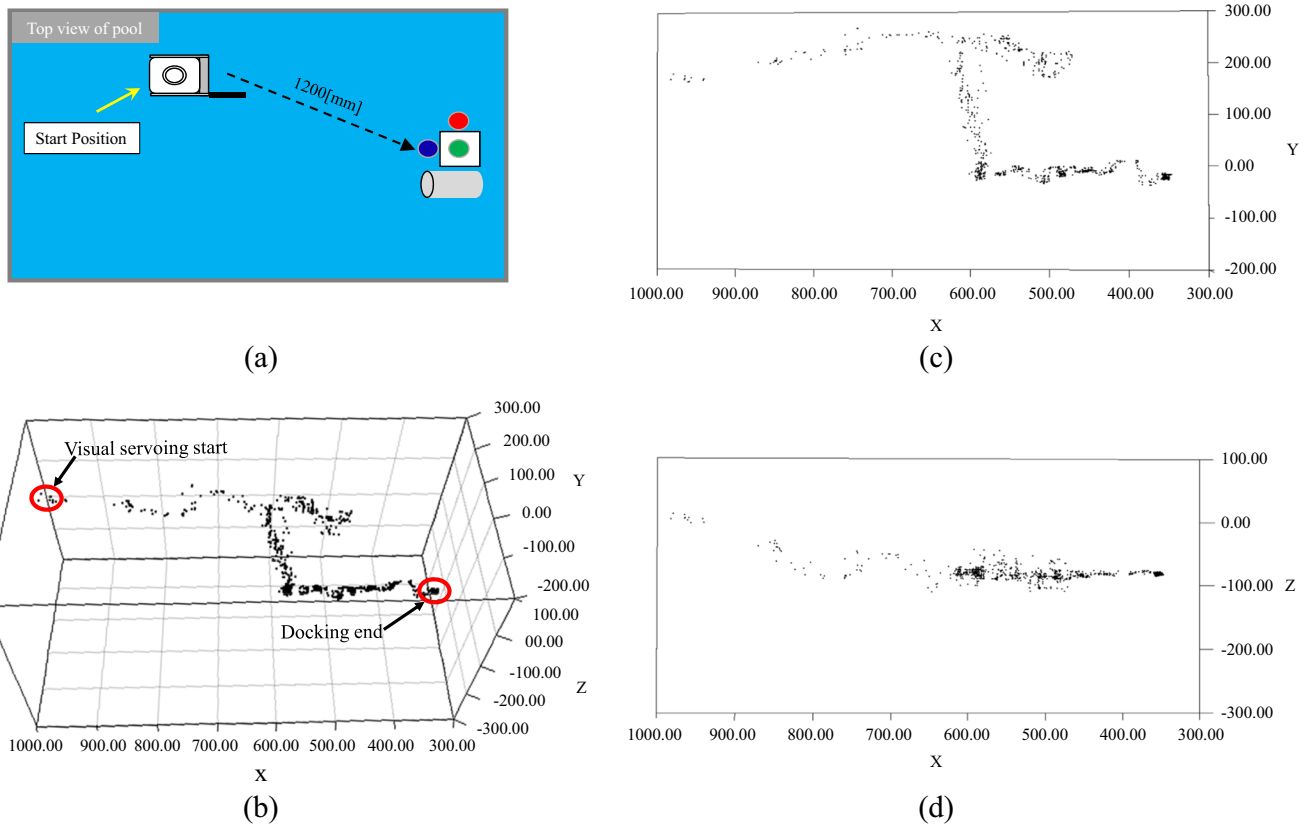


Fig. 29 Recognized trajectory for start position of underwater vehicle on the left side of pool relative to 3D marker, position (**b**) in Fig. 23: **a** Start position of underwater vehicle, **b** recognized position in 3D by Real-time Multi-step GA, **c** recognized position in the xy plane, and **d** recognized position in the xz plane

However, when the range of the abovementioned errors exceeds a defined value in the docking process, the underwater robot suspends the docking process and goes back to the visual servoing state to execute the docking process again, as shown by the arrow "P" in Fig. 9. In the state of visual servoing, when the error of the relative target position and the posture between the object and the underwater robot reaches below the abovementioned threshold, the process transits to docking, in which the robot is expected to fit into the homing unit while recognizing the target object. The docking process is done by performing visual servoing until the robot moves to ($x_d = 350$ mm, $y_d = 0$ mm, $z_d = -67$ mm, $\epsilon_{3d} = 0$ deg). In completion of the docking state, the underwater robot is expected to keep the relative target pose of the object in the connected state by visual servoing. Figures 24, 26, and 27b to j show the result of docking experiments with different start positions: (b) is fitness, (c) (e) (g) (i) are the position and orientation of the underwater robot, and (d) (f) (h) (j) are the thrust and torque in each axis. Figure 25 represents the error of the relative pose with respect to Fig. 24. From each of the figures, the transition of the state to (A) approach, (B) visual servoing, (C) docking, and (D) completion of docking can be seen. In other words, the experiments confirm that the underwater robot can achieve docking by using the proposed system. Figures 28, 29, and 30 show the trajectories of the ROV recognized by the Real-time Multi-step GA when the ROV starts from different positions. Please note that these trajectories start by using visual servoing and end at the point where docking is completed. From the trajectories in Figs. 29c and 30c, surge motion appeared faster than sway motion.

6.6 Evaluation with Full Search for 3D Pose Estimation Performance and Comparison with Other Systems

To evaluate 3D pose recognition performance of Real-time Multi-step GA, we compare a pose estimated by Real-time Multi-step GA with one estimated by full search method at the sampled operation time. In a full search method, all possible poses in the searching area (see Table 2) are evaluated by the fitness function. Therefore, the pose estimated by the full search method that has the highest peak means a true pose of the 3D marker at the time of the image being input. Figure 31 shows fitness value of (a) all genes (models) (b) top 60 % genes and (c) bottom 40 % genes searched by Real-time multi-step GA. It can be seen that

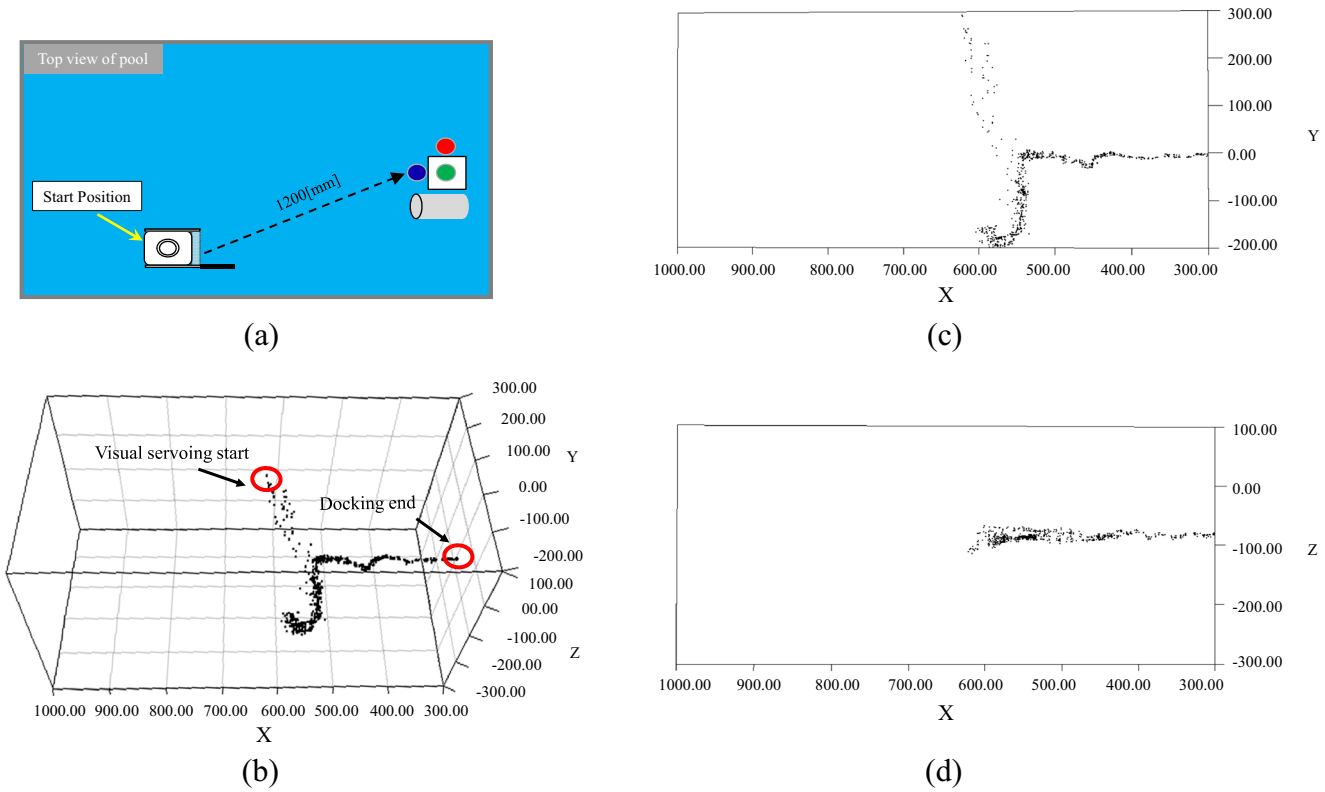
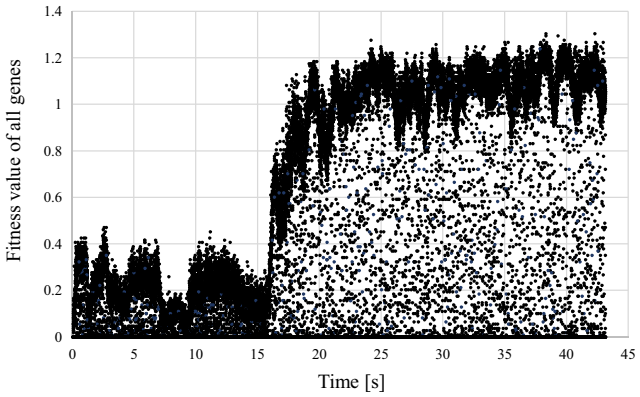


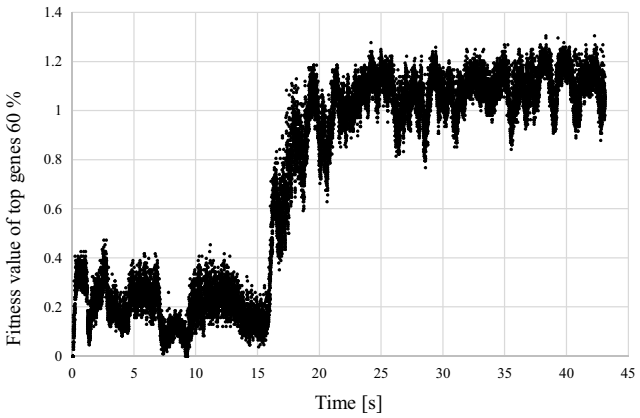
Fig. 30 Recognized trajectory for start position of underwater vehicle on the right side of pool relative to 3D marker, position (**c**) in Fig. 23: **a** Start position of underwater vehicle, **b** recognized position in 3D by Real-time Multi-step GA, **c** recognized position in the xy plane, and **d** recognized position in the xz plane

the top 60 % models converge to the top gene that represents the estimated pose. Other 40 % genes are distributed (seems to be uniform) as shown in Fig. 31c for preparing to detect the sudden appearance of growing peak that may represent a true 3D marker. Figure 31d-f show the comparisons between the fitness distributions of full search and Real-time Multi-step GA at the operation time of 16 s, 20 s, and 40 s respectively. The estimated poses represented by the fittest genes of both Real-time Multi-step GA and full search method are also expressed to analyze the 3D pose recognition performance. In Fig. 31d-f, positions in Y-Z plane are estimated by both Real-time Multi-step GA and full search method. Mountain-shaped fitness value distribution represents full search and black dots represent the poses of all genes searched by Real-time Multi-step GA. By comparing the position represented by the peak of each method, the error between the pose estimated by Real-time Multi-step GA and one by full search method can be calculated. For example, the position in Y-Z plane estimated by Real-time Multi-step GA at 16 s (See Fig. 31d) is (Y= 91 mm, Z=-108 mm) and the position estimated by full search is (Y= 90 mm, Z=-103 mm). Therefore, the error in Y-axis is -1 mm and the error in Z-axis is 5 mm. According to the results in Fig. 31d-f, it can be seen that the maximum error in 3D pose estimation of Real-time Multi-step GA is 5 mm.

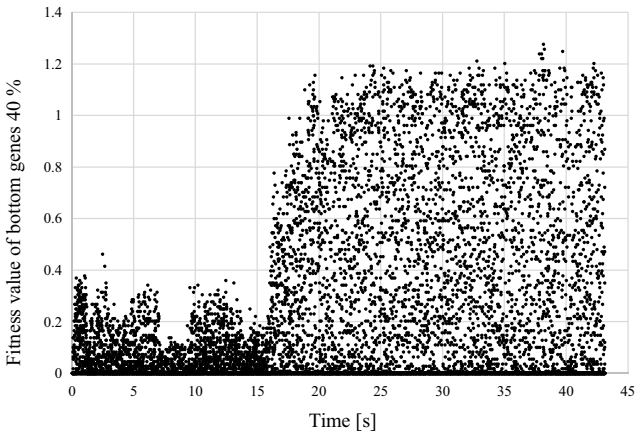
Even though the experimental result of the proposed system is not evaluated with other methods rather than full search (discussed above) in this paper for a square comparison, the functionality and practicality of the proposed approach were verified experimentally showing some advantages over other studies. A conical shape docking station with a large diameter is used to capture the vehicle in the following studies [5–7, 9], in which different sensors including vision sensor are used. In [9], vision guided docking of an AUV was conducted using the docking device with five lights around the rim of the entrance. This kind of approach may address some physical damages to the vehicle as well as the docking station. In contrast, precise docking is performed using our proposed system without hard touching of the vehicle to the docking station. In [6], a docking station is large enough for a torpedo-typed AUV with a diameter of 54-cm to enter. In contrast, a 7-cm diameter of docking hole is used for docking operation in our docking method. Therefore, we would like to highlight our performance comparing to other studies in terms of accuracy that can be said centimeter level. Regarding operational duration



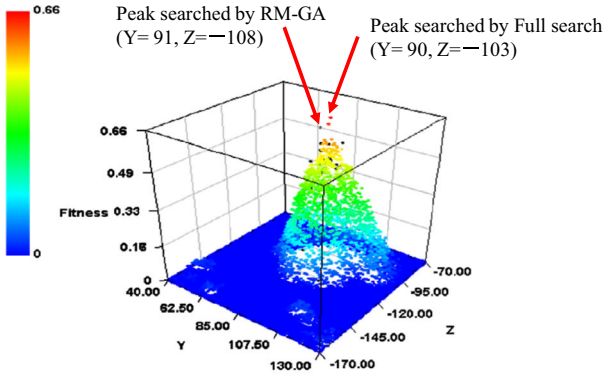
(a)



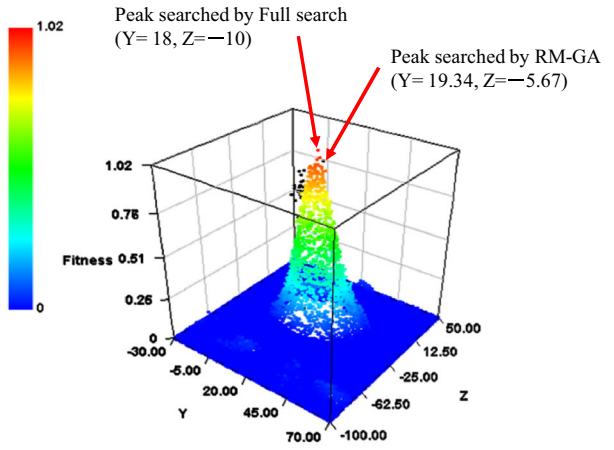
(b)



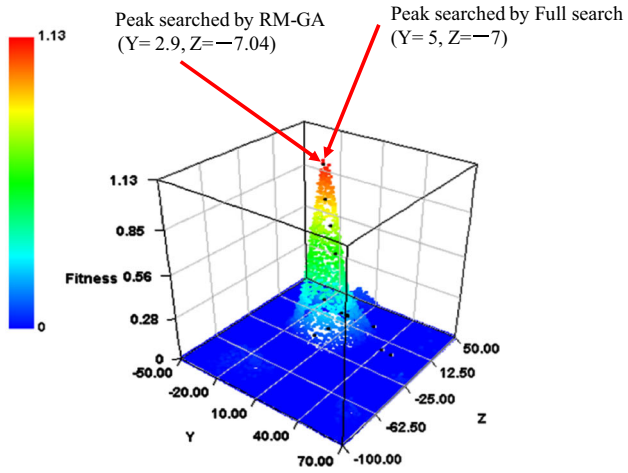
(c)



(d)



(e)



(f)

Fig. 31 Fitness values of **a** all genes, **b** top 60 % genes, **c** bottom 40 % genes in Real-time Multi-step GA (RM-GA), and comparison between pose estimated by full search —meaning fitness function distribution against y,z (y,z: y_H and z_H position in Σ_H in Fig. 13) position of 3D model have been calculated in full space of y-z plane —and Real-time Multi-step GA search at the operation time of **d** 16 s, **e** 20 s, and **f** 40 s. Mountain shaped fitness values in (**d**) (**e**) (**f**) represent for full search and black dots for Real-time Multi-step GA. (position unit is in mm)

time, it took less than 80 seconds for docking operation by visual servoing in the proposed system while it took 99 seconds in the study of [23] that is one of the novel works on precise docking operation using visual information.

7 Conclusion

In this work, visual servoing by using two cameras for an underwater vehicle was designed and implemented successfully by using the system called 3D-MoS and a 3D marker on an ROV. Model-based pose estimation using two cameras was validated for real-time applications. Experiments of regulating the performance with and without physical disturbances were conducted to verify the robustness of the proposed system. Then, docking experiments in a simulated pool were carried out to demonstrate underwater automatic battery charging. A unidirectional docking station that requires high homing accuracy was designed to evaluate the vehicle docking capability. The experimental results showed the proposed system could successfully carry out docking operations. In future work on the development of the underwater automatic charging system, we will verify the effectiveness of the proposed system in an actual sea environment.

Acknowledgments The authors would like to thank the Kowa cooperation for their collaboration in the experiments.

References

1. Balasuriya, B.A.A.P., Takai, M., Lam, W.C., Ura, T., Kuroda, Y.: Vision based autonomous underwater vehicle navigation: underwater cable tracking. Proc. MTS/IEEE OCEANS Conf. **2**, 1418–1424 (1997)
2. Son-Cheol, Y., Ura, T., Fujii, T., Kondo, H.: Navigation of autonomous underwater vehicles based on artificial underwater landmarks. Proc. MTS/IEEE OCEANS Conf. **1**, 409–416 (2001)
3. Cowen, S., Briest, S., Dombrowski, J.: Underwater docking of autonomous undersea vehicle using optical terminal guidance. Proc. MTS/IEEE OCEANS Conf. **2**, 1143–1147 (1997)
4. Teo, K., Goh, B., Chai, O.K.: Fuzzy docking guidance using augmented navigation system on an AUV. IEEE J. Ocean. Eng. **40**(2), 349–361 (2015)
5. Feezor, M.D., Yates Sorrell, F., Blankinship, P.R., Bellingham, J.G.: Autonomous underwater vehicle Homing/Docking via electromagnetic guidance. IEEE J. Ocean. Eng. **26**(4), 515–521 (2001)
6. McEwen, R.S., Hobson, B.W., McBride, L., Bellingham, J.G.: Docking control system for a 54-cm-diameter (21-in) AUV. IEEE J. Ocean. Eng. **33**(4), 550–562 (2008)
7. Teo, K., An, E., Beaujean, P.-P.J.: A robust fuzzy autonomous underwater vehicle (AUV) docking approach for unknown current disturbances. IEEE J. Ocean. Eng. **37**(2), 143–155 (2012)
8. Negre, A., Pradalier, C., Dunbabin, M.: Robust vision-based underwater homing using self similar landmarks. J Field Robot., Wiley-Blackwell, Special Issue on Field and Service Robot. **25**(6-7), 360–377 (2008)
9. Park, J.-Y., Jun, B.-H., Lee, P.-M., Oh, J.: Experiments on vision guided docking of an autonomous underwater vehicle using one camera. IEEE J. Ocean. Eng. **36**(1), 48–61 (2009)
10. Palomeras, N., Ridao, P., Ribas, D., Vallicrosa, G.: Autonomous i-AUV docking for fixed-base manipulation. Proc. Int. Federation Autom. Control **47**(3), 12160–12165 (2014)
11. Foresti, G.L., Gentili, S., Zampato, M.: A vision-based system for autonomous underwater vehicle navigation. Proc. MTS/IEEE OCEANS Conf. **1**, 195–199 (1998)
12. Ura, T., Kurimoto, Y., Kondo, H., Nose, Y., Sakamaki, T., Kuroda, Y.: Observation behavior of an AUV for ship wreck investigation. Proc. MTS/IEEE OCEANS Conf. **3**, 2686–2691 (2005)
13. Myint, M., Yonemori, K., Yanou, A., Ishiyama, S., Minami, M.: Robustness of visual-servo against air bubble disturbance of underwater vehicle system using three-dimensional marker and dual-eye cameras. In: Proceedings of the MTS/IEEE OCEANS Conference, pp. 1–8 (2015)
14. Myint, M., Yonemori, K., Yanou, A., Lwin, K.N., Minami, M., Ishiyama, S.: Visual servoing for underwater vehicle using dual-eyes evolutionary real-time pose tracking. J. Robot. Mechatronics **28**(4), 543–558 (2016)
15. Myint, M., Yonemori, K., Yanou, A., Minami, M., Ishiyama, S.: Visual-servo-based autonomous docking system for underwater vehicle using dual-eyes camera 3D-pose tracking. In: Proceedings of the 2015 IEEE/SICE International Symposium on System Integration (SII), pp. 989–994 (2015)
16. Myint, M., Yonemori, K., Yanou, A., Lwin, K.N., Minami, M., Ishiyama, S.: Visual-based deep sea docking simulation of underwater vehicle using dual-eyes cameras with lighting adaptation. In: Proceedings of the MTS/IEEE OCEANS Conference, pp. 1–8 (2016)
17. Myint, M., Yonemori, K., Yanou, A., Lwin, K.N., Mukada, N., Minami, M.: Dual-eyes visual-based sea docking for sea bottom battery recharging. In: Proceedings of the MTS/IEEE OCEANS Conference Monterey, pp. 1–7 (2016)
18. Stokey, R., Purcell, M., Forrester, N., Austin, T., Goldsborough, R., Allen, B., von Alt, C.: A docking system for REMUS, an autonomous underwater vehicle. Proc. MTS/IEEE OCEANS Conf. **2**, 1132–1136 (1997)
19. Park, J.Y., Jun, B.H., Lee, P.M., Oh, J.H., Lim, Y.K.: Underwater docking approach of an under-actuated AUV in the presence of constant ocean current. IFAC Control Appl. Marine Syst. **43**(20), 5–10 (2010)
20. Park, J.-Y., Jun, B.-H., Lee, P.-M., Lee, F.-Y., Oh, J.-h.: Experiment on underwater docking of an autonomous underwater vehicle ISimI using optical terminal guidance. In: OCEANS'2007-Europe, pp. 1–6 (2007)
21. Dunbabin, M., Lang, B., Wood, B.: Vision-based docking using an autonomous surface vehicle. In: IEEE International Conference on Robotics and Automation, pp. 26–32. Pasadena (2008)
22. Batista, P., Silvestre, C., Oliveira, P.: A two-step control strategy for docking of autonomous underwater vehicles. In: American Control Conference, pp. 5395–5400. Fairmont QueenElizabeth, Montrreal, Canada (2012)
23. Palomeras, N., Penalver, A., Massot-Campos, M., Vallicrosa, G.N., Gre, P.L., Fernandez, J.J., Ridao, P., Sanz, P.J., Oliver-Codina, G., Palomer, A.: I-AUV docking and intervention in a Subsea panel. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2279–2285. Chicago (2014)
24. Vasilijevic, A., Borovic, B., Vukic, Z.: Underwater vehicle localization with complementary filter: performance analysis in the shallow water environment. J. Intell. Robot. Syst. **68**(3-4), 373–386 (2012)

25. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (2011)
26. Kasaei, S.H., Oliveira, M., Lim, G.H., Lopes, L.S., Tomé, A.M.: Interactive open-ended learning for 3D object recognition: an approach and experiments. J. Intell. Robot. Syst. **80**(3-4), 537–553 (2015)
27. Wu, P., Liu, Y., Ye, M., Li, J., Du, S.: Fast and adaptive 3D reconstruction with extensively high completeness. In: IEEE Transactions on Multimedia (2016)
28. Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M., Szeliski, R.: Building rome in a day. Commun. ACM **54**(10), 105–112 (2011)
29. Lwin, K.N., Yonemori, K., Myint, M., Naoki, M., Minami, M., Yanou, A., Matsuno, T.: Performance analyses and optimization of real-time multi-step ga for visual-servoing based underwater vehicle. In: Techno-oceans (2016)
30. Song, W., Fujia, Y., Minami, M.: 3-D visual servoing by feedforward evolutionary recognition. J. Adv. Mech. Des., Syst. Manuf. **4**(4), 739–755 (2010)
31. Suzuki, H., Minami, M.: Visual servoing to catch fish using global/local GA search. IEEE/ASME Trans. Mechatronics **10**(3), 352–357 (2005)
32. Song, W., Minami, M., Aoyagi, S.: On-line stable evolutionary recognition based on unit quaternion representation by Motion-Feedforward compensation. Int. J. Intell. Comput. Medical Sci. Image Process. (IC-MED) **2**(2), 127–139 (2008)
33. Song, W., Minami, M.: On-line motion-feedforward pose recognition invariant for dynamic hand-eye motion. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1047–1052 (2008)
34. Song, W., Minami, M.: Stability / precision improvement of 6-DoF visual servoing by motion feedforward compensation and experimental evaluation. In: IEEE International Conference on Robotics and Automation, pp. 722–729 (2009)
35. Song, W., Minami, M.: Hand and eye-vergence dual visual servoing to enhance observability and stability. In: IEEE International Conference on Robotics and Automation, pp. 714–721 (2009)
36. Minami, M., Agbanhan, J., Asakura, T.: Evolutionary scene recognition and simultaneous position/orientation detection. In: Soft Computing in Measurement and Information Acquisition, pp. 178–207. Springer, Berlin (2003)
37. Suzuki, H., Minami, M.: Real-time recognition of multiple pedestrians using car-mouted camera. Electron. Commun. Part 3 **89**(4), 21–33 (2006)
38. Trelea, I.C.: The particle swarm optimization algorithm: convergence analysis and parameter selection. Infor. Process. Lett. **85**(6), 317–325 (2003)
39. Lin, H.I.: A fast and unified method to find a minimum-jerk robot joint trajectory using particle swarm optimization. J. Intell. Robot. Syst. **75**(3-4), 379 (2014)
40. Polden, J., Pan, Z., Larkin, N., van Duin, S.: Adaptive partial shortcuts: path optimization for industrial robotics. J. Intell. Robot. Syst **86**(1), 35–47 (2017)
41. Mousavian, S.H., Koofigar, H.R.: Identification-based robust motion control of an AUV: optimized by particle swarm optimization algorithm. J. Intell. Robot. Syst. **85**(2), 331–352 (2017)
42. Ma, X.M.: Application of ant colony algorithm in PID parameter optimization for mining hoist direct torque control system. In: ICACC'09 International Conference of Advanced Computer Control, pp. 632–636 (2009)

Myo Myint received his M.E. degree of electronic engineering from Yangon Technological University, Myanmar in 2006. He has obtained his doctorate degree in electronic engineering in 2009 from Mandalay Technological University, Myanmar. He worked as an associate professor in Yangon Technological University from 2011 to 2015. He is now with the Department of Intelligent Mechanical System, Okayama University, Japan. Currently, he is working on vision-based control for autonomous underwater vehicles. His research interests include control of autonomous underwater vehicles, robotics, electronics and renewable energy.

Kenta Yonemori received his B.E. degree from Okayama University in 2014. He obtained his M.E. degree in robotics from Okayama University in 2016. His research interests include 3D pose estimation, robotics and optimization.

Khin Nwe Lwin received her B.E. in Mechatronic engineering from the Technological University (Kyaukse), Myanmar, in 2006, and graduated in M.E. of Mechatronic Engineering from Mandalay Technological University, Myanmar, in 2011. She has worked as a Lecturer on Faculty of Precision Engineering, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar. Currently, she is working on intelligent robotics and control of underwater vehicle, at the Intelligent Mechanical System department, Graduated School of Natural Science and Technology, Okayama University, Japan. Her research interests include visual servoing, control of autonomous underwater vehicle, robotics and genetic algorithm.

Akira Yanou received his Ph.D. in Engineering from Okayama University, Japan in 2001. He worked with School of Engineering, Kinki University from 2002 to 2008, and Graduate School of Natural Science and Technology, Okayama University from 2009 to 2016. He joined Department of Radiological Technology, Kawasaki College of Allied Health Professions in 2016, and Faculty of Health Science and Technology, Kawasaki University of Medical Welfare in 2017, as an Associate Professor. His research interests include adaptive control, strong stability systems and estimation.

Mamoru Minami obtained his Ph.D. degree from Kanazawa University in 1990. He worked for University of Fukui from 1994 to 2002. He has worked as a Professor in the Department of Intelligent Mechanical System, Okayama University, Japan since 2010. His research interests include adaptive control, visual servoing, robotics, autonomous underwater vehicle, cloth recognition and underwater docking system.