Robustness of 3D Pose Estimation against Turbidity Using Dual-eye Cameras And Active/Lighting 3D Marker for Visual-servoing Based AUV

1st Khin Nwe Lwin
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
pdoj8yez@s.okayama-u.ac.jp

2nd Myo Myint
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
puqs1ci8@s.okayama-u.ac.jp

3rdNaoki Mukada
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
pola9zud@s.okayama-u.ac.jp

4th Daiki Yamada
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
p7kw2h27@s.okayama-u.ac.jp

5th Takayuki Matsuno
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
matsuno@cc.okayama-u.ac.jp

6th Mamoru Minami
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
minami-m@cc.okayama-u.ac.jp

***Abstract*—This paper presents the 3D pose estimation against turbidity under dark environment by using dual-eye cameras and an active —meaning light emitting— 3D marker for visual-servoing based underwater vehicle. The authors have proposed a 3D-perception based move on sensing (3D-MoS) system using a 3D position and orientation (pose) estimation method with dual-eye cameras that exploits the parallactic nature that enables reliable 3D pose estimation in real-time, named as "Real-time Multi-step Genetic Algorithm (RM-GA)." The active/lighting 3D marker was designed and constructed to improve the 3D pose estimation especially in turbidity and low illumination. In real-time pose estimation, not only recognition but also robustness are important. This paper focus on the robustness of 3D pose recognition performance using dual-eye cameras and 3D marker against turbidity. The experimental results have confirmed that the effectiveness and robustness of the proposed system for the real-time 3D pose estimation under turbidity and night condition.**

***Index Terms*—Turbidity, Real-time Multi-step GA, Active 3D marker, 3D pose estimation, Dual-eye cameras**

I. Introduction

Visual servoing based underwater vehicle has become essential for sea exploration and exploitation such as inspection, repair oil and gas, pipeline tracking, docking task, scientific studies of the deep ocean, etc [1], [2]. Some references are based on single camera to estimate the pose of the target object [3]- [5]. A binocular vision was used in some of these studies in order to estimate the relative pose of the target object in [6], [7]. Even though two cameras were used in [6], one was facing downward for shooting the sea-floor images and the second camera was pointed forward for the purpose of obstacle avoidance. In [7], Girona 500 AVU was developed for inspection and intervention tasks for the seafloor survey. In that work, several sensors and two video cameras are used to localize an object. But the two cameras look at difference targets, then these approaches do not materialize parallactic nature.

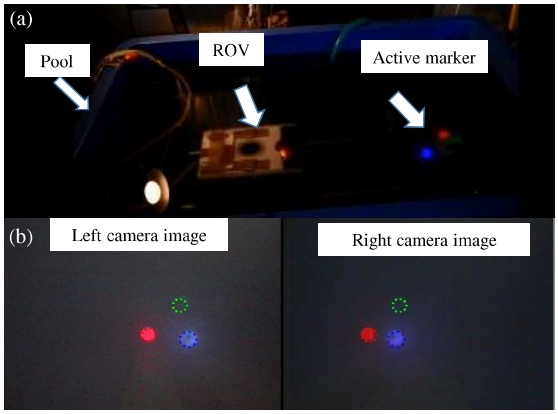

Fig. 1. experimental setup of the ROV and an active marker under dark environment: (a) an underwater vehicle and an active 3D Marker in a simulated pool. (b) left and right camera images of the ROV. The dotted circle in left and right cameras images is the estimated pose by real-time multi-step genetic algorithm (RM-GA).

In the actual sea environment, it is difficult to observe underwater environment using the cameras because of the disturbances such as lighting effect, water current disturbance, turbidity and refraction effect, etc. The common disturbances for the vision-based underwater vehicle is the turbidity and low illumination when the vehicle are used to explore the deep sea environment. The role of turbidity should be considered in the underwater environment because it can degrade the visual quality of the camera images. Especially how to recognize the target objects by the camera in high turbidity condition, it is an important fact to considered to expand the possibility

978-1-5386-1654-3/18/$31.00 ©2018 IEEE

of utilization of underwater vehicle in the exploration of the underwater environment. In [8], [9], two cameras and three cameras are used to increase the image recognition robustness by analysing degradation factors in turbid water. But the discussion about turbidity is not enough in these references. They evaluated the multiple features of detectors and compared the performance of detectors on images degraded by turbidity. They did not perform the pose estimation.

The authors have proposed a new 3D pose estimation method with dual-eye cameras that exploits the parallactic nature that enables reliable 3D pose estimation in real-time as shown in Fig. 1. Visual servoing using stereo vision and parallactic character for the underwater vehicle utilizing 3D model-based recognition and Realtime Multi-step Genetic Algorithm (RM-GA) has been initiated by our research group. The dual-eye cameras based on perception means solving the corresponding points problem. If the corresponding points in the real object are not connected with the corresponding points in images during 2D-to-3D reconstruction, the true 3D object cannot be represented because of the wrong reconstructed 3D points. Based on this point, 3D model-based recognition is implemented. Furthermore, solid object that is represented a group of points on the sphere of the 3D marker is projected rather than the individual pixel.

To the best of authors' knowledge, there are no related works in which the real-time pose estimation using stereo-vision based real-time performance against turbidity and night condition. In previous works [10]- [12], the robustness of the proposed system was confirmed by conducting the different experiments. However, the robustness of the proposed system against turbidity had not been done previously. This present paper is intended to explain how the proposed system performs real-time pose estimation robustly against turbidity on the image, aiming at confirming the system can keep recognizing the pose of 3D marker despite in the dark environment.

The remainder of the paper is organized as follows: Section II describes the method of real-time 3D pose estimation. Experiment results are reported in section III with discussion and conclude in section IV.

II. Real-time 3D Pose Estimation Method

A. Model-based Matching Method

In this section, dual-eye based 3D MoS for underwater docking especially 3D pose estimation method is briefly discussed for reader convenience. In proposed system, the model-based matching method is used to estimate the matching degree between the projected model and the captured images. In other conventional methods, the pose estimation method is implemented using feature-based recognition based on 2D-to-3D reconstruction. In that approach, the set of image points in different images is used to determine the information of the target object. The main drawback is complex for searching the corresponding points and time taken. Apart from this, the model-based pose estimation approach based on 3D to 2D projection is applied in this work avoiding the effects of wrong mapping points in images using dual-eye cameras.

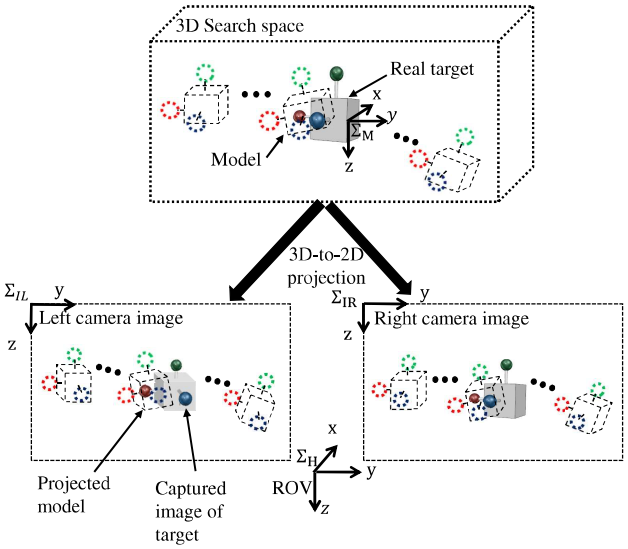

Fig. 2. Model-based matching method using dual-eye cameras and 3D marker

Figure 2 shows the model-based matching method using dual-eye cameras for 3D pose estimation. In Fig. 2, Σ_{IR} and Σ_{IL} are the reference coordinate frames of the right camera image and the left camera image. Σ_H is the reference frame of the ROV. Σ_M is the reference frame of the real target object. The solid model of the real target object in space is projected naturally to the dual-eyes cameras images and the dotted 3D marker model where pose is given by one of GA's genes is projected from 3D-to -2D. The different relative pose is calculated by comparing the projected model and the captured images by the dual-eye cameras. Finally, the best model of the target object that represents the true pose can be obtained based on its highest fitness value. Please refer to [13] for detailed explanation. There are some works done on visual-servoing experiments concerning hand eye manipulator in the air using 3D model-based matching method utilizing genetic algorithms and dual-eyes camera [14], [15], which are used as fundamental knowledge for this research.

B. Fitness Function

The fitness function is constructed to evaluate the matching degree between the projected model and the captured image. 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 directly influences the RM-GA's convergence speed [16]. Figure 3(a) shows the real 3D active marker and Fig. 3(b) shows model with enlarged view of the blue ball model, where the inner area is the same size as the real target object (blue ball) and the outer area is the background area. Each model consists of three spherical ball (red, green and blue). Each spherical ball consists of two portions, where the inner portion is the same size as the real target object and the outer area is the background area. The dots in each ball mean points to calculate the correlation degree on how much the inner area

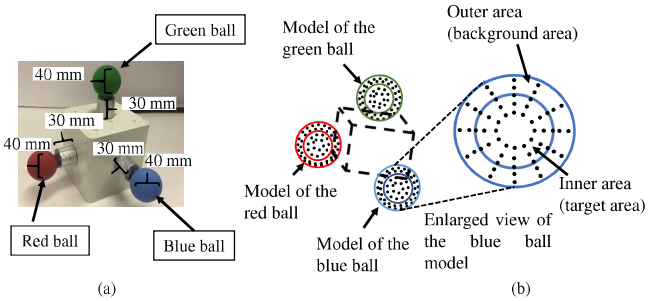

Fig. 3. Real 3D marker and model: (a) active/lighting marker, (b) model with enlarged view of the blue ball model, where the inner area is the same size as the real target object (blue ball) and the outer area is the background area. The dots in enlarged view of blue ball mean points to calculate the correlation degree on how much the inner area overlaps the blue ball and the outer area does not overlap the blue ball.

overlaps the target object and the outer area does not overlap the target object. Only hue value is used for recognition of 3D marker because of less sensitive to the environment. The captured image (pixel) is detected in 2D image as (green or blue or red) in hue space. If the captured image (pixel) situated in inner portion, the fitness value increases and the capture image (pixel) is situated in outer portion, the fitness value decreases. Therefore, the fitness value will be maximum when the model and the real target are identical. Finally, the pose of the model with the maximum fitness value is to represent the pose of the real target 3D marker. Detailed explanation about the fitness function is referred to our previous paper [17].

C. Real-time Multi-step GA (RM-GA)

The genetic algorithm is used as a search and an optimization method to estimate the relative pose between the ROV and 3D marker. Even though there are many powerful optimization methods, GA was selected and modified as Real-time Multi-step GA (RM-GA) because of its simplicity and especially effectiveness in real-time performance. Real-time multi-step GA is means the capable of real time recognition of the true pose of the target object within 33 ms. Figure 4 shows an individual of GA population. Position and orientation of the three-dimensional model is represented as 72 bits string of length as shown in Fig. 4. The former 36 bits represent the position of the 3D marker and the later 36 bits describe the orientation defined by a quaternion. Figure 5(a) shows the flowchart of the RM-GA and Fig. 5(b) illustrates the behavior of GA convergence from the first generation to the final generation. Please not that although the pose of the target object is evaluated in 2D, convergence occurs in 3D.

Firstly, a random population of model is generated. A new pair of left and right images that was captured by ROV's cameras is input. The fitness value of each model was evaluated by using the fitness function. Each model is sorted based on the fitness value and selected the better model from the current generation to reproduce the new generation. Then, again new generations are formed from the two-point

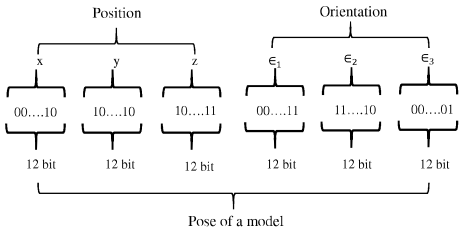

Fig. 4. An individual of GA population: 12 bits for each x, y, z represents the position coordinate of the three dimensional model of the gene and 12 bits for each $\epsilon_1, \epsilon_2, \epsilon_3$ describes the orientation defined by a quaternion.

crossover and mutation operation of GA. The real-time multi-step GA evolves the models with as many generations as possible within the video frame rate for each image. In the present study, the number of evolution times of the RM-GA was chosen to be nine, which is a maximum that the computer used in the present study could calculate within 33 ms (determined by the video frame rate) during the GA evolution process. The RM-GA find repeatedly the solutions to get the optimum value that indicates the best pose of the target object. For the next input, a new video image is used. The convergence performance to an optimum value of the GA's evolution function used as fitness function has been proved mathematically by a Lyapunov analysis in a previous work [18]. The effectiveness of the GA was demonstrated in a previous study on visual servoing for catching fish using a GA search [19].

D. Active/Lighting Marker

In our previous works, the passive marker was used to conduct the experiments. In the present study, the active marker was designed and newly constructed to improve the pose estimation at high turbidity and low illumination. Figure 6(a) shows the appearance of the active marker and Fig. 6(b) shows the internal circuit diagram. The circuit was created by combining the variable resistors, resistors, and light emitting diodes such as red, green, and blue. The resistance value of the variable resistors, and the number of resistors determined by trial and error. The value of each variable resistors and resistors are shown in Fig. 6. The 3D marker was constructed from a water proof fiber box (100 mm $\times$ 100 mm $\times$ 100 mm) and white spheres (diameter: 40 mm). The red, green and blue LED were installed into the white spherical ball and covered by color balloon as shown in Fig. 6(a). The pose estimation has improved by emitting the light of LED, especially in the night time. The effectiveness of the active marker will be discussed in the next section.

III. Experimental Results and Discussion

A. Experiment Environment

Figure 1 shows the experimental layout in an indoor pool (length 750 mm$\times$ width 570 mm $\times$ height 490 mm) which was filled 800 litres with fresh water. The ROV manufactured by KOWA cooperation is used as a test bed in this study. As

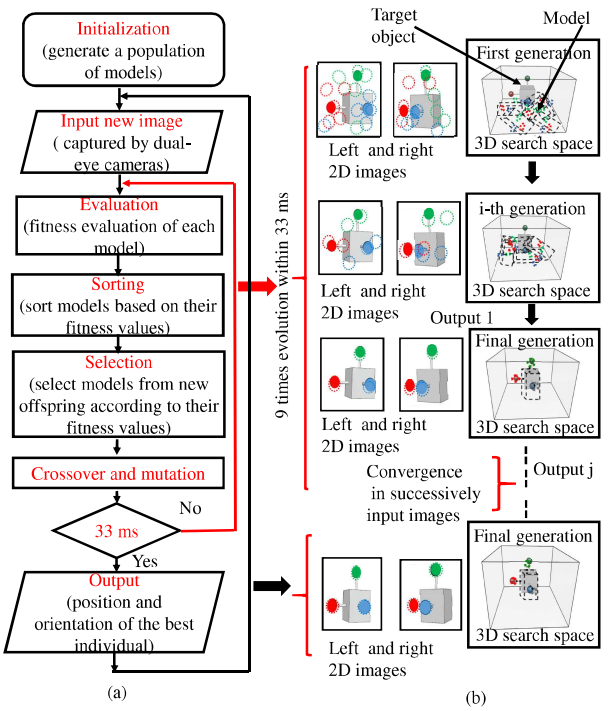

Fig. 5. Flowchart of the real-time multi-step GA: (a) the flowchart of the RM-GA, the true pose of the target object is evaluated within 33 ms through the GA process (b) the convergence behavior of RM-GA from the first generation to the final generation to represent the true pose of the target object in which evaluation is performed in 2D and convergence occurs in 3D.

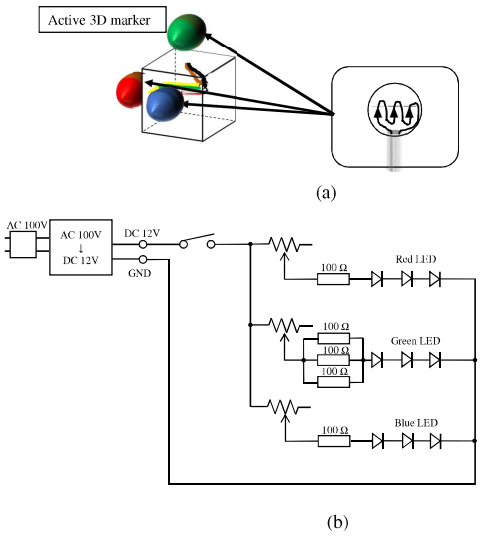

Fig. 6. Active 3D marker and its internal circuit.

an image captured device, an imaging element CCD, signal system NTSC with a resolution 380,000 pixels was used. The ROV and active 3D marker were fixed in position. The amount of turbidity was controlled by adding whole milk and mud in water in the tank. The milk was chosen because milk is a highly scattering liquid than the other [20]. The mud was chosen to simulate as the natural condition. It was carried from near the sea environment. The turbidity level(Formazin Turbidity Unit, FTU) was measured by using portable turbidity monitoring sensor TD-M500 (manufactured by OPTEX). The illumination condition was 9 lx that was measured by using using a lux sensor (model: LX-1010B, manufactured by Milwaukee) when the experiment was conducted. The turbidity level is gradually increased until the ROV cannot recognize the 3D marker.

B. 3D Pose Estimation in Turbid Water Under Dark Environment

The ROV performed the visual servoing at about 600 mm in docking operation [12], meaning waiting and stabilizing for docking operation to recognize the target object. Therefore, we choose 600 mm distance for recognition performance in this section. Figures 7 and 8 show the fitness value in different turbidity levels in the case of milk and mud at the distance 600 mm between the ROV and 3D marker under dark environment. The horizontal axis is described by milk and mud amount in ml/m^3, the left vertical axis is expressed in terms of fitness value and the right vertical axis is described in terms of FTU by using turbidity sensor.

According to the graphical results in the case of milk, the fitness value decreased from 0.6 to 0.1 when the milk amount is gradually increased from 0 ml/m^3 (0 FTU) to 169.75 ml/m^3 (20.9 FTU). The FTU value linearly increased from 0 FTU to 20.9 FTU when the milk amount was increased. The control threshold fitness value can be adjusted depending on FTU level. For example, the control threshold fitness value can be adjusted 0.4 when the FTU level was 11 FTU. But, when the FTU level was above 20 FTU, the ROV cannot recognize the 3D marker well. At that time, the fitness value was about 0.01.

In the case of mud, the fitness value decreased from 0.6 to 0.1 when the turbidity was gradually increased from 0 ml/m^3 (0 FTU) to 375.875 ml/m^3 (50.2 FTU) . In the case of mud, the FTU level rapidly increased in some amount of mud when we added. The ROV cannot recognize the 3D marker when the FTU level reached to 375.875 ml/m^3 (50.2 FTU). At that time, the fitness value was 0.02. The performance of 3D pose estimation under different turbidity levels are analyzed and the maximum turbidity level can be determined according to the defined threshold of fitness value. The fitness distribution was analysed at the points of "A"-"E" that was choose among different turbidity levels as shown in Figs. 7 and 8. The fitness distribution of these points are discussed in the next section.

C. Robustness in 3D Pose Estimation

Figures 9 and 10 show the fitness distribution between x, y plane against different turbidity levels for the case of milk and mud in the night condition. Figures 9 and 10 "(A)"-"(E)" correspond to the points denoted by dotted lines "A"-"E" in Figs. 7 and 8. In figs. 9 and 10, (I) 3D graph of fitness distribution, (II) 2D graph of fitness distribution and

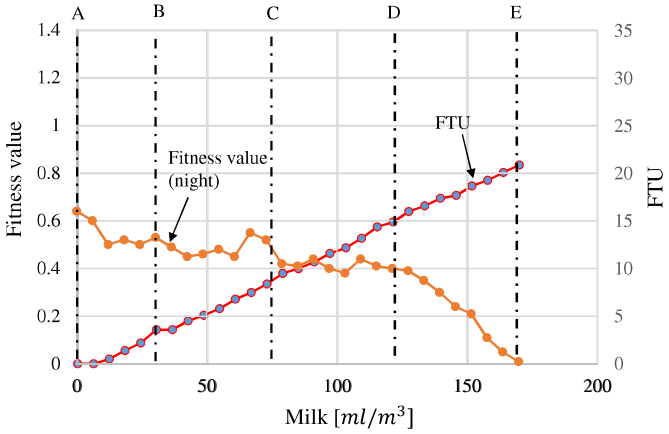

Fig. 7. Turbidity tolerance under night condition in the case of milk at the distance 600 mm between the ROV and the active 3D marker.

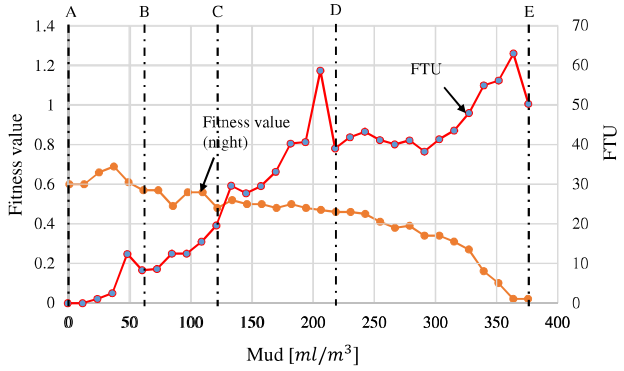

Fig. 8. Turbidity tolerance under night condition in the case of mud at the distance 600 mm between the ROV and the active 3D marker.

(III) left and right cameras images are described, respectively. The pose estimated using full search method indicated in the fitness value distribution for each of the turbidity levels. The full search method is the scanning of points in all planes (x,y, and z) of the images and the fitness value of every points which are 1 [mm] apart in the entire searching area were calculated. The true pose of the target object is represented by the peak of the mountain shape as shown in Figs. 9 and 10(A)-(E). The presence of a peak of the fitness distribution indicates the robustness of the recognition method against turbidity.

According to the experimental results, the position of the peak corresponding to the true pose of the target was maintained even though the height of the mountain shape of fitness values gradually reduces when the turbidity level is increased, but the estimated pose represented by the peak is maintained as shown in Figs. 9 and 10(A)-(E)(I). In figs. 9 and 10(A)-(E)(II), it can be seen the intersection shape of "X" type in x-y plane, the peak of the mountain shape occurs at the intersection point. This is because of the dual-eye cameras that exploits the parallactic nature that enables reliable 3D pose estimation in real-time. The appearance of turbidity can be seen in related two camera images Figs. 9 and 10(A)-(E)(III).

In two camera images, the dotted circle means the pose of the 3D marker recognized by RM-GA. It can be clearly seen that the dotted circle and the real target exactly coincide at 0 FTU. At that time, the fitness value is maximum. When the turbidity level is gradually increased, the fitness value of the true pose of the mountain shape is also reduced but the position of the mountain shape still remained at different turbidity levels in both cases of milk and mud. The problem conversion from pose estimation into optimization problem has merit to make the visual servoing feedback system to be robust against turbidity. This is brought from the fact that the problem conversion does not care about the height of fitness distribution as shown in Figs. 9 and 10. In other words, the robustness of the proposed system using the RM-GA can recognize the position and orientation of the 3D marker in real time.

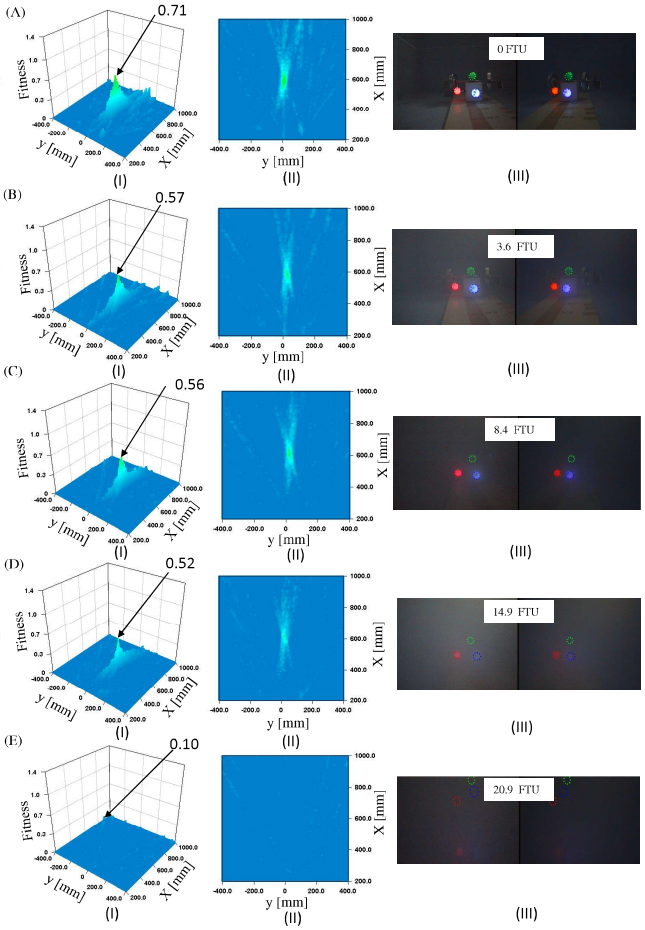

Fig. 9. Fitness value distributions confirming the robustness of the proposed system at a distance 600 mm between the ROV and the active 3D marker against different turbidity levels under night condition in the case of milk: (I) fitness distribution in 3D graph, (II) fitness distribution in 2D graph, and (III) left and right cameras images.

IV. Conclusion

In the present paper, visual-servo-based 3D pose estimation system against turbidity for underwater vehicle in the dark environment was proposed. A real-time pose detection scheme

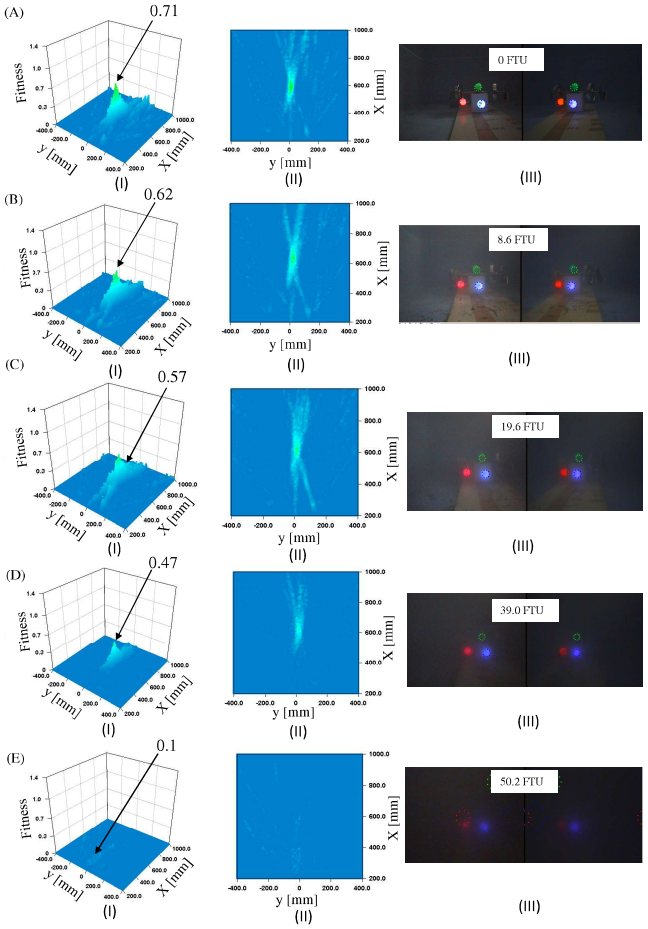

Fig. 10. Fitness value distributions confirming the robustness of the proposed system at a distance 600 mm between the ROV and the active 3D marker against different turbidity levels under night condition in the case of mud: (I) fitness distribution in 3D graph, (II) fitness distribution in 2D graph, and (III) left and right cameras images.

was implemented by means of 3D model-based recognition and real-time multi-step GA (RM-GA) using stereo vision and 3D marker as the active target. The robustness of the proposed system against turbidity under dark environment was confirmed experimentally in the simulated pool. Results show that the proposed system can recognize the relative pose of a 3D marker robustly even though the turbid environment that can degrade the images.

REFERENCES

[1] Agbakwuru J, "Oil/Gas Pipeline Leak Inspection and Repair in Underwater Poor Visibility Conditions: Challenges and Perspectives," *Journal of Environmental Protection*, 2012.

[2] Krupi r ki S, Allibert G, Hua M D, Hamel T, "Pipeline tracking for fully-actuated autonomous underwater vehicle using visual servo control," *In American Control Conference (ACC)*, pp. 6196-6202. IEEE, 2012.

[3] Eustice R M, Pizarro O, Singh H, "Visually augmented navigation for autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, 33(2), pp. 103-122, 2008.

[4] Jung J, Choi S, Choi H T, Myung H, "Localization of AUVs using depth information of underwater structures from a monocular camera," *In Ubiquitous Robots and Ambient Intelligence (URAI), 2016 13th International Conference*, pp. 444-446, IEEE, 2016.

[5] Ghosh S, Ray R, Vadali SR, Shome SN, Nandy S, "Reliable pose estimation of underwater dock using single camera: a scene invariant approach," *Machine Vision and Applications*, 27(2), pp. 221-36, 2016.

[6] Kume A, Maki T, Sakamaki T, Ura T, "A method for obtaining high-coverage 3D images of rough seafloor using AUV-realtime quality evaluation and path-planning," *Journal of Robotics and Mechatronics*, 25(2), pp. 364-374, 2013.

[7] Ribas D, Palomeras N, Ridao P, Carreras M, Mallios A, "Girona 500 auv: From survey to intervention," *IEEE/ASME Transactions on Mechatronics*, 17(1), pp. 46-53, 2012.

[8] Codevilla F, Gaya J D O, Duarte N, Botelho S, "Achieving turbidity robustness on underwater images local feature detection," *International journal of computer vision*, 60(2), pp. 91-110, 2004.

[9] Garcia R, Gracias N, "Detection of interest points in turbid underwater images,"*In OCEANS, 2011 IEEE-Spain*, pp. 1-9. IEEE, 2011.

[10] 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," *Proceedings of the International Conference on OCEANS (Washington DC,USA)(MTS/IEEE)*, 150521-002, 2015.

[11] 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," *Journal of Robotics and Mechatronics*, 28 (4):543-558, 2016.

[12] 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," *Proceedings of OCEANS 2016-Shanghai*, pp. 1-8, 2016.

[13] Myint M, Yonemori K, Lwin K N, Yanou A, Minami M, "Dual-eyes Vision-based Docking System for Autonomous Underwater Vehicle: an Approach and Experiments," *Journal of Intelligent and Robot Systems*, DOI 10.1007/s10846-017-0703-6, 2017.

[14] Yu F, Minami M, Song W, Zhu J, Yanou A, "On-line head pose estimation with binocular hand-eye robot based on evolutionary model-based matching," *Journal of Computer and Information Technology*, 2(1), pp. 43-54, 2012.

[15] Song W, Minami M, Aoyagi S, "Feedforward on-line pose evolutionary recognition based on quaternion," *Journal of the Robot Society of Japan*, 28(1), pp. 55-64, 2010.

[16] Lwin K N, Yonemori K, Myint M, Mukada N, Minami M, Yanou A, Matsuno T, "Performance analyses and optimization of real-time multi-step GA for visual-servoing based underwater vehicle," *In Proceedings 2016 IEEE International Conference on Techno-Ocean 2016, IEEE*, 2016.

[17] Song W, Minami M, Aoyagi S, "On-line stable evolutionary recognition based on unit quaternion representation by motion-feedforward compensation," *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, 2(2), pp. 127-139, 2008.

[18] Song W, Fujia Y, Minami M, "3D visual servoing by feedforward evolutionary recognition," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, 4(4), pp. 739-755, 2010.

[19] Suzuki H, Minami M, "Visual servoing to catch fish using global/local GA search," *IEEE/ASME Transactions on Mechatronics*, 10(3), pp. 352-357, 2005.

[20] Narasimhan, S.G, Gupta. M, Donner. C, Ramamoorthi. R, Nayar. S.K, and Jensen. H. W, "Acquiring scattering properties of participating media by dilution," In ACM Transactions on Graphics (TOG), Vol. 25, No. 3, pp. 1003-1012, ACM, 2006.