

Expanding the recognition distance using the Model-based Matching method and the 2D model by zoom cameras

Siyu Pan¹, Renya Takahashi¹, Jincheng Li¹, Yuichiro Toda¹, and Mamoru Minami¹

¹Okayama University, Japan
(Tel: 81-86-251-8233, Fax: 81-86-251-8233)
¹pl2540ow@s.okayama-u.ac.jp

Abstract: In recent years, Autonomous Underwater Vehicles (AUVs) are essential to explore the ocean that humans are unable to go directly. AUVs could be operated automatically and required to be recharged after the long-term work. In our previous research, we have conducted the docking experiments with the distance of 600 [mm] from docking station to Remotely Operated Vehicle (ROV) in real sea using dual-eye cameras. However, it can be difficult to recognize the 3D marker from further distance owing to the target getting smaller on the screen and the turbid water. In the next step, we need an approaching ability that can guide AUV to reach the docking range from a greater distance. For this purpose, we proposed a system using 2D model and zoom cameras in the approaching step. In this paper, the effectiveness of pose estimation that using 2D model with zoom cameras is confirmed by experiments.

Keywords: Dual-eyes based visual servoing, Model-based Matching method, 3D pose estimation

1 INTRODUCTION

AUVs are applied to seabed exploration, pipeline maintenance, oilfield exploration and other fields [1][2]. AUVs usually have limited working time and moving distance because of the limited battery capacity. They require recycling to get recharged, maintain support and transfer data. It is essential to provide an underwater battery recharging system for AUVs to return to the docking station. There are three stages in a docking operation: long-distance navigation, approach and short-distance docking. We have made the short-distance docking system using visual servoing based ROV in docking step with the distance of 600 [mm] [3][4]. We used to perform the approaching step by manual. In the next step, we need a system that can guide the underwater vehicle from longer distance automatically in the approaching step.

In the real sea environment, it is hard to observe objects at a long distance because of the water current disturbance, turbidity and refraction effect, etc. The main disturbance for the vision-based underwater vehicle is the turbidity. As the distance increases, the influence of the turbidity will also increase. That is the reason why we required the system to have the ability to against the water turbidity. To solve the issue mentioned above we newly designed a 2D model and a fitness function. In our previous research, only hue value was used in fitness function. However, as the level of turbidity increase, it is hard to get the hue information of the 3D marker from camera images. Therefore, we use a new fitness function determined by hue and brightness value to improve the recognition in high turbidity environment. In addition, the target will be a small proportion in the camera images at a long distance and hard to be recognized. We may zoom

out the target to make true there is enough proportion of 3D marker in camera images. Therefore, we also use zoom cameras to check out if the 2D model is useful for pose estimation at long distance with different focal length.

In this paper, section 2 describes the method of real-time 3D pose estimation. The author will introduce the Model-based Matching method, 2D model and the fitness function we used. In section 3, the 3D pose estimation experiment using 2D model with different fitness function in turbid water and the approaching experiments using 2D model with different focal length of zoom cameras in air environment are presented.

2 REAL-TIME 3D POSE ESTIMATION METHOD

2.1 Model-based Matching Method

In this section, we discuss the 3D-perception based move on sensing (3D-MoS) based on dual-eye cameras for the underwater vehicle and the 3D pose estimation method. In this system, the model-based matching method is used to estimate the matching degree between the projected model and the captured images. The pose estimation based on 2D-to-3D reconstruction using feature-based recognition is applied in other conventional methods. The set of image points in different images to determine the information of the target object is implemented in that approach. The main drawback of this approach is the complexity of searching for the corresponding point and time taken. Apart from this, we use the model-based pose estimation approach based on 3D-to-2D projection to avoid the influence of the wrong mapping

points in images using the dual-eye cameras.

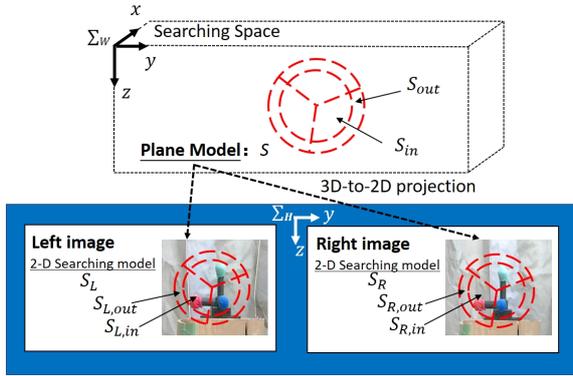


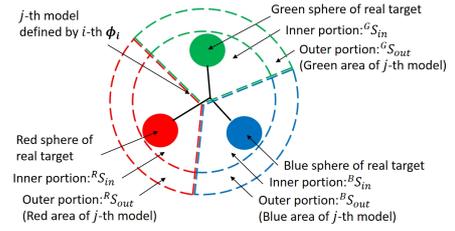
Fig. 1. Real target and projected 2D model in 2D images obtained by the right and left camera

Figure 1 shows the model-based matching method using dual-eye cameras for 3D pose estimation. Σ_H is the reference coordinate frame of the right camera image and the left camera image. The 2D 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 the gene of Genetic Algorithms (GA) is projected from 3D-to-2D [5]. 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. 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-eye camera [6][7], which are used as fundamental knowledge for this research.

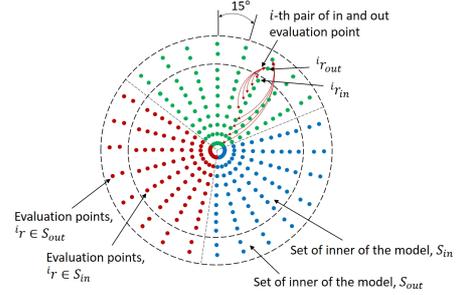
2.2 2D Model and Fitness Function

The 2D model to evaluate object is composed of points. Figure 2 shows the consistence of the model. There are three kinds of colors (red, green and blue) in the model to match the corresponding colors of three spherical balls in the 3D marker. The inner portion is the same size as the real target object (3D marker) and the outer portion is the background area. The points in inner portion are paired with the points in the outer portion for the calculation of the fitness value. The points in each portion mean points to calculate the correlation degree on how much the inner portion overlaps the target object and the outer portion does not overlap the target object.

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 peak at the true position of the target. The construction of the fitness function affects the optimum search performance.



(a) 3D marker made of 3-spheres with red, green and blue colors



(b) Evaluation points defined in 2D model projected in camera image

Fig. 2. Real target and projected 2D model

The calculation of conventional fitness function is based on hue value. However, as the distance and turbidity level increase. It is hard to recognize the whole marker and the color in the background also influences the result of recognition. In that case, the brightness of the marker can still be recognized by system. So, we propose a new fitness function that adds the evaluation of brightness value to the conventional fitness function. At first we describe the conventional function as Eq. (1)~(3).

In equation (1), N is the number of points in inner portion. $I r_j$ represents the I -th point of the model. $I r_{j,in}(\phi_i)$ is the point that satisfies $I r_j(\phi_i) \in S_{in}$. ϕ_i defines the position of the model as $\phi_i = [x, y, z]$. j means the j -th model given by GA. When the ball of the 3D marker situated in the corresponding inner portions (red, green, blue) of 2D model, $p_{H,1}(I r_{j,in}(\phi_i)) = +1$, otherwise $p_{H,1}(I r_{j,in}(\phi_i)) = -1$. If the ball is situated in outer portion, $p_{H,2}(I r_{j,out}(\phi_i)) = -1$, otherwise $p_{H,2}(I r_{j,out}(\phi_i)) = +1$. When the three balls situated in corresponding inner portions at the same time. $p_{H,3}(I r_{j,in}(\phi_i), I r_{j,out}(\phi_i)) = +2$ otherwise $p_{H,3}(I r_{j,in}(\phi_i), I r_{j,out}(\phi_i)) = -1$.

$$f'_H = \frac{1}{N} \begin{pmatrix} \sum_{I r_j(\phi_i) \in S_{in}} p_{H,1}(I r_{j,in}(\phi_i)) + \\ \sum_{I r_j(\phi_i) \in S_{out}} p_{H,2}(I r_{j,out}(\phi_i)) + \\ \sum_{\substack{I r_j(\phi_i) \in S_{in} \\ I r_j(\phi_i) \in S_{out}}} p_{H,3}(I r_{j,in}(\phi_i), I r_{j,out}(\phi_i)) \end{pmatrix} \quad (1)$$

To make sure the peak will be situated in the true position of the marker and the value is in the range of 0 to 1, we use

Eq. (2) as follows. $H_{u,sum}(u = R, G, B)$ represents the number of points overlap the marker in the corresponding inner portion (red, green, blue) of 2D model.

$$f_H = \begin{cases} f'_H + 0.5 & (4 < H_{R,sum} < 20) \cap \\ & (4 < H_{G,sum} < 20) \cap \\ & (4 < H_{B,sum} < 20) \\ f'_H + 0.3 & \{(4 < H_{R,sum} < 20) \cap (4 < H_{G,sum} < 20)\} \cup \\ & \{(4 < H_{R,sum} < 20) \cap (4 < H_{B,sum} < 20)\} \cup \\ & \{(4 < H_{G,sum} < 20) \cap (4 < H_{B,sum} < 20)\} \\ f'_H & (otherwise) \end{cases} \quad (2)$$

Equation (3) is used to take the average of the fitness value calculated in left camera images ${}^L f$ and right camera images ${}^R f$.

$$F = \frac{1}{2}({}^L f_H + {}^R f_H) \quad (3)$$

Newly designed fitness function is based on hue and brightness value, the calculation of the hue fitness is the same as Eq. (1),(2). In the calculation of brightness fitness, we take the difference between the paired points in inner portion and outer portion. Eq. (4),(5) are used to calculate the brightness fitness. If the difference between the brightness value of the paired points is over 30, $p_{Br,1}({}^I r_{j,in}(\phi_i), {}^I r_{j,out}(\phi_i)) = +1$, otherwise $p_{Br,1}({}^I r_{j,in}(\phi_i), {}^I r_{j,out}(\phi_i)) = -1$. When the three balls situated in the corresponding inner portion at the same time, $p_{Br,2}({}^I r_{j,in}(\phi_i), {}^I r_{j,out}(\phi_i)) = +2$, otherwise $p_{Br,2}({}^I r_{j,in}(\phi_i), {}^I r_{j,out}(\phi_i)) = -1$.

$$f'_{Br} = \frac{1}{N} \left(\begin{array}{l} \sum_{\substack{{}^I r_j(\phi_i) \in S_{in} \\ {}^I r_j(\phi_i) \in S_{out}}} p_{Br,1}({}^I r_{j,in}(\phi_i), {}^I r_{j,out}(\phi_i)) \\ + p_{Br,2}({}^I r_{j,in}(\phi_i), {}^I r_{j,out}(\phi_i)) \end{array} \right) \quad (4)$$

To make sure the peak will situated in the true position of the marker and the value is in the range of 0 to 1, we use the Eq. (5) as follows. $Br_{u,sum}(u = R, G, B)$ represents the number of paired points that meet the difference between the paired points situated in the corresponding areas (red, green, blue).

$$f_{Br} = \begin{cases} f'_{Br} + 0.8 & (1 < Br_{R,sum} < 20) \cap \\ & (1 < Br_{G,sum} < 20) \cap \\ & (1 < Br_{B,sum} < 20) \\ f'_{Br} + 0.5 & \{(1 < Br_{R,sum} < 20) \cap (1 < Br_{G,sum} < 20)\} \cup \\ & \{(1 < Br_{R,sum} < 20) \cap (1 < Br_{B,sum} < 20)\} \cup \\ & \{(1 < Br_{G,sum} < 20) \cap (1 < Br_{B,sum} < 20)\} \\ f'_{Br} & (otherwise) \end{cases} \quad (5)$$

To make the comprehensive evaluation of the hue and brightness fitness value and enhance the performance of

brightness fitness value, we use the Eq. (6) as follows. The equation also makes sure that the result will be in the range of 0 to 1.

$$f = f_H + f_{Br} - f_H \cdot f_{Br} \quad (6)$$

Equation (7) is used to take the average value of the fitness value calculated in left camera images ${}^L f$ and right camera images ${}^R f$.

$$F = \frac{1}{2}({}^L f + {}^R f) \quad (7)$$

3 EXPERIMENTS USING 2D MODEL

Two experiments are presented in this section to prove the effectiveness of the 2D model in long distance recognition. To prove the ability of 2D model to against the turbidity, we conducted the experiment of 3D pose estimation in turbid water environment with different fitness functions.

To prove the target tracking ability using 2D model with zoom cameras. We conducted the approaching experiments in air environment with different focal length. The zoom cameras can change the focal length from 3.3 [mm] to 7.2 [mm].

3.1 First Experiment: 3D Pose Estimation using 2D Model in Turbidity Water

The experimental layout is in an indoor pool which was filled with fresh water. The cameras and 3D marker were fixed in position. The camera coordinate system Σ_H and the model coordinate system Σ_M are situated as Fig. 3. The amount of turbidity was controlled by adding milk. The milk was chosen because milk is a highly scattering liquid than the other [8]. The turbidity level (Formazin Turbidity Unit, FTU) was measured by using portable turbidity monitoring sensor TD-M500 (manufactured by OPTEx). The turbidity level is gradually increased until the system cannot recognize the 3D marker. ROV starts the approaching operation after it recognizes the target object before the docking operation at the distance further than 600 [mm]. Therefore, we choose $L = 2000$ [mm] as the distance for recognition performance in this section.

Table 1 shows the maximum of the fitness value between y, z plane against different turbidity levels at 2000 [mm] and the images taken by left and right cameras. Fig. 6 shows the details of the fitness distribution between y, z plane against different turbidity levels at 2000 [mm] and (I) left and right cameras images and level of turbidity, (II) graph of fitness distribution using fitness function based on hue and (III) graph of fitness distribution using fitness function based on hue and brightness are described. The pose estimated using

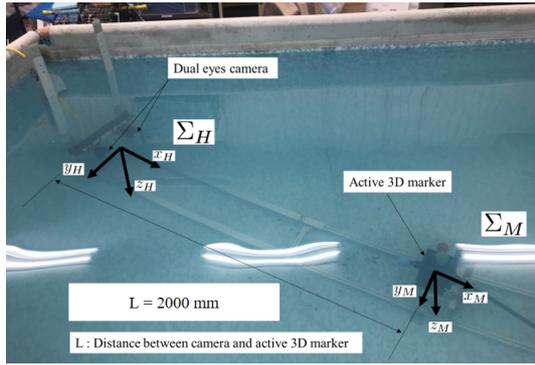


Fig. 3. First Experiment: Experimental environment using dual-eye camera and active 3D marker

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 y, z planes of the images and the fitness value of every points which are 1 [mm] apart in the entire searching area were calculated. The position of the target object is represented by the highest peak as shown in Fig. 6.

Table 1. Maximum fitness value of different fitness functions at 2000 [mm] against different turbidity levels.

Fitness function	hue	hue + brightness
Milk ml/m ³ (FTU)		
0(0)	0.618	0.793
4.85(0.56)	0.562	0.799
7.28(0.76)	0.530	0.744
9.70(1.03)	0.458	0.718
1.21×10(1.2)	0.486	0.758
1.46×10(1.6)	0.432	0.745
1.70×10(1.76)	0.462	0.686
1.94×10(2.0)	0.448	0.633
2.18×10(2.46)	0.295	0.485
2.43×10(2.6)	0.307	0.367
2.67×10(3.03)	0.402	0.335
2.91×10(3.5)	0.272	0.251
3.15×10(3.56)	0.404	0.238
3.40×10(3.9)	0.268	0.184
3.64×10(4.06)	0.277	0.196

According to the results. It can be confirmed that the fitness value decreases as the turbidity increase. In Table 1, Area(I) means that there is a peak around the true position of the 3D marker in the fitness distribution shown as Fig. 6(a)(b) with both fitness functions. However, in Area(II), it is difficult to observe the whole 3D marker and the peak changes to the position of the red ball in y, z plane using the fitness function based on hue and brightness as shown in Fig. 6(c)(III). The peak shows the possibility of ROV getting close to 3D marker at the long distance in turbid environment and the three balls of the 3D marker can be recognized again as the

distance decrease between ROV and 3D marker. In Area(III), the highest peak is not situated in the position around the 3D marker with both fitness function which means 3D marker can not be recognized. This experiment shows the 2D model with the fitness function based on hue and brightness has the ability to against the turbidity, even the 3D marker can not be recognized completely as the turbidity increase.

3.2 Second Experiment: Approaching Experiment using 2D model with Zoom Cameras in Air Environment

The experiment is conducted in air environment. The camera coordinate system Σ_H and the model coordinate system Σ_M are determined as Fig. 4.

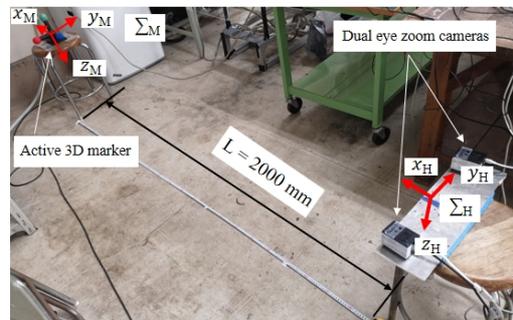


Fig. 4. Second Experiment: Experimental environment using zoom cameras and active 3D marker

To simulate the approaching operation, we set the 3D marker on the chair at the position $x = 2000$ [mm], $z = 20$ [mm], $y = 0$ [mm] in coordinate system Σ_H and push the chair from $x = 2000$ [mm] to 600 [mm] then back to 2000 [mm] for one approach. To make sure the marker can be completely photographed by the cameras all the time, the focus length $f = 7.2$ [mm] is used further than 1500 [mm] and switches to $f = 3.3$ [mm] less than 1500 [mm] by system.

Figure 5 shows the experimental results. Red lines represents the timing of switching the focal length and blue lines shows the images of left and right cameras before and after changing the focal length. According to the results, the pose estimation is effective when using different focal length. In x-axis, the distance from 3D marker to zoom cameras changes from 2000 [mm] to 600 [mm] is the same as the distance we moved the 3D marker. The pose estimation in z-axis is always around 20 [mm] and y-axis is always around 0 [mm] within an error less than 50 [mm]. This error is considered acceptable in the approaching step. The fitness value is around 0.8 which means the effectiveness of recognition during the experiments. However pose estimation is unstable when changing the focal length and recovers soon after that. This problem may influence the control of ROV and

should be discussed in our future research. This experiment shows the effectiveness of pose estimation using 2D model with different focal length.

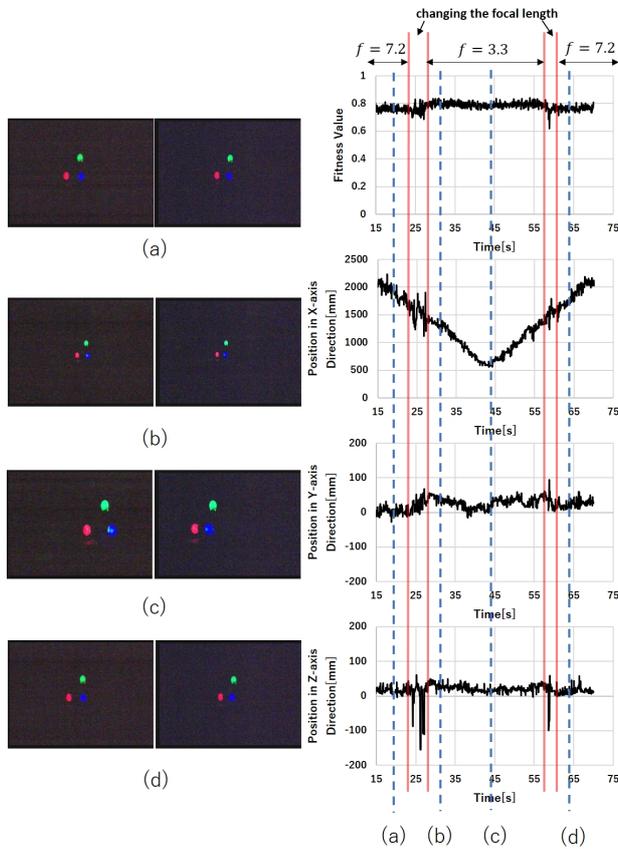


Fig. 5. Approaching experiment using 2D model changing focal length from 7.2 [mm] to 3.3 [mm]

4 CONCLUSION

In the present paper, pose estimation using 2D model and Model-based Matching method was proposed for long distance recognition. The effectiveness of pose estimation that using 2D model was confirmed experimentally. 2D model with fitness function based on hue and brightness can against the turbidity at a long distance. Results show that the proposed system can recognize the relative pose of a 3D marker robustly even though the turbid environment. In addition, pose estimation that using 2D model with zoom cameras was also confirmed in the approaching experiment. It proves that the proposed system can recognize the position of the target at a long distance with different focal length. These experiments demonstrate the possibility of 2D model being applied on ROV for long distance recognition in our future research.

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] Krupar 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] Lwin K N, Mukada N, Myint M, Yamada D, Minami M, Matsuno T, Saitou K, Godou W, "Docking at Pool and Sea by Using Active Marker in Turbid and Day/Night Environment," *Artificial Life and Robotics*, DOI 10.1007/s 10015-018-0442-1, 2018.
- [4] Yamada D, Mukada N, Myint M, Lwin K N, Matsuno T, Minami M, "Docking Experiment in Dark Environment Using Active/Lighting Marker and HSV Correlation," *Proceedings of IEEE/MTS Techno-Ocean Conference 2018*, Paper No.171129-098, Kobe, Japan, May 29-31, 2018.
- [5] Lwin K N, Myint M, Mukada N, Yamada D, Matsuno T, Saitou K, Godou W, Sakamoto T, Minami M, "Sea Docking by Dual-eye Pose Estimation with Optimized Genetic Algorithm Parameters," *Journal of Intelligent Robotic Systems*, Vol.92, Issue 1, pp.159-186, DOI (identifier) 10.1007/s10846-018-0970-x, 2018.
- [6] 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.
- [7] 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.
- [8] 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.

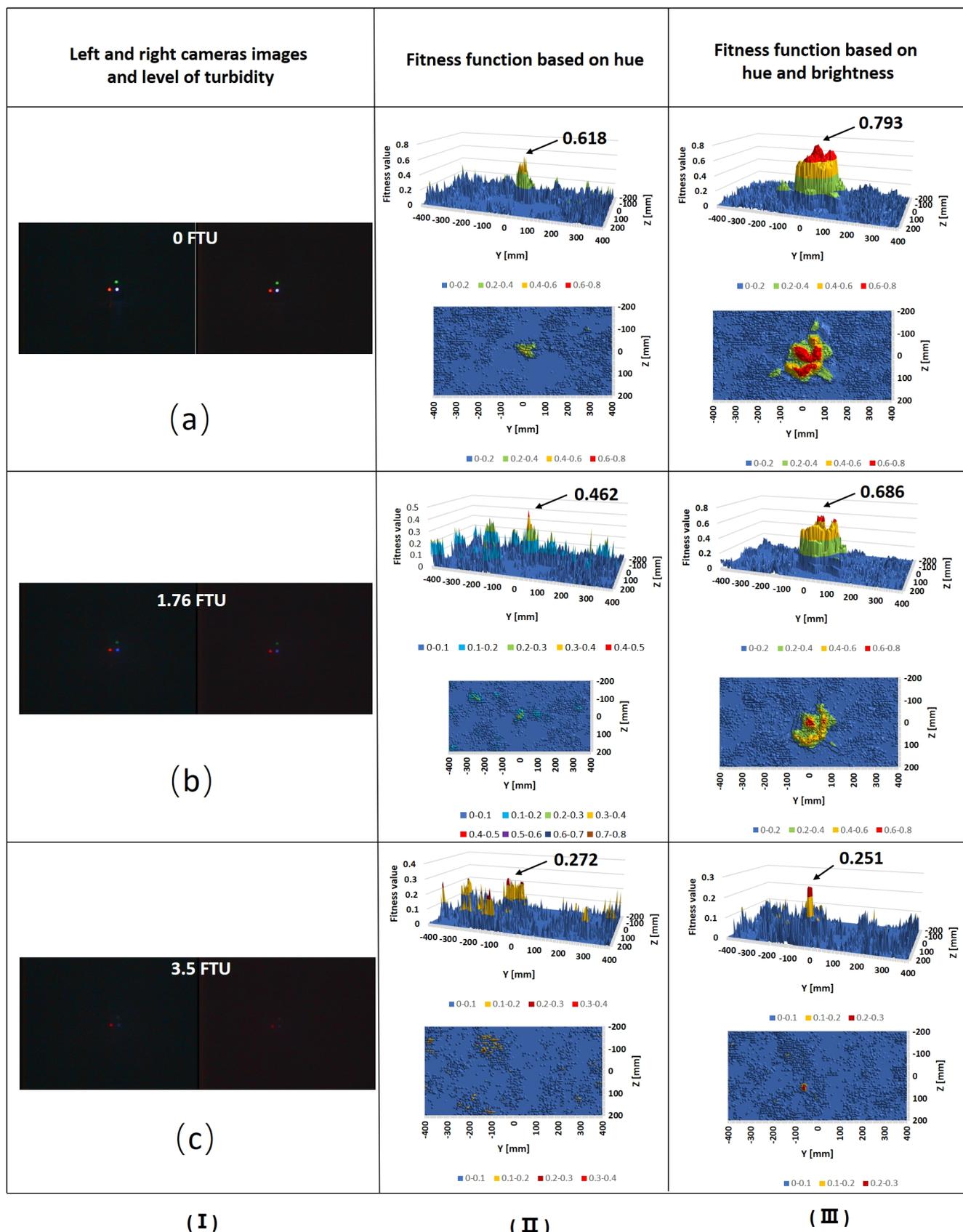


Fig. 6. Fitness distribution in y-z plane confirming the robustness of the proposed system against different turbidity levels at 2000 [mm].(I) left and right cameras images, (II) graph of fitness distribution using fitness function based on hue and (III) graph of fitness distribution using fitness function based on hue and brightness