Photo-Model-Based Stereo-Vision 3D Perception for Marine Creatures Catching by ROV

1st Hongzhi Tian
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
psnc8ytd@s.okayama-u.ac.jp

2nd Yejun Kou
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
ptlg9dvi@s.okayama-u.ac.jp

3rd Takuro Kawakami
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
p4zi5w9s@s.okayama-u.ac.jp

4th Renya Takahashi
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
pcz88iou@s.okayama-u.ac.jp

5th Mamoru Minami
Graduate School of Natural Science and Technology, Okayama University
Okayama, Japan
minami-m@cc.okayama-u.ac.jp

***Abstract*—By incorporating visual information obtained from the installed vision sensor into the feedback loop, visual servoing enables the robot to be able to operate in a changing environment or an unknown environment. In the past, the authors have proposed a real-time monocular fish-catching robot. However, it is insufficient for new demands, especially fish-catching in 3D space. To overcome these problems, a binocular stereo-vision system with photo-model-based recognition method has been proposed for picking and placing clothes. Then, this previous research has been the base for extending the stereo-vision pick-and-place system to a three-dimensional (3D) object visual servoing system, which used the genetic algorithm (GA) of the fishing-catching robot. This paper verifies the visual servoing ability of the system for a moving 3D object through visual servoing experiment. In the experiment, marine creature toy floated on the water in the pool without pose constraints. It is confirmed from the experimental results that the proposed system can be used to capture a moving marine creature target and is not susceptible to partial occlusion conditions.**

***Index Terms*—Photo-model-based recognition method, Stereo-vision, Pose estimation, Visual servoing**

I. Introduction

Since robots have higher reliability and accuracy than humans, they have been used extensively in production factories to perform a wide variety of tasks instead of human workers. Moreover, robots equipped with hand-eye cameras have been utilized in factories for handling metal parts. Obviously, controlling robots with visual information has been studied mainly for ground-based robots. For example, target object detection and recognition for mobile robots using a monocular camera were proposed in [1], [2]. On the other hand, some vision-based autonomous underwater vehicles (AUVs) have been researched, e.g., for eliminating underwater invasive species with a monocular camera [3]. In contrast to the current state of robot vision research for use on land, applying robot vision in water is at a lower stage.

Object recognizing based on the input image from a monocular camera has also been studied in our laboratory [5]. As shown in Fig. 1, a fish-catching robot has been developed to deal with problems of the target's three degrees-of-freedom (3DOF) recognition. The 3DOF means the fish's x,y positions and heading concerning the $x - y$ plane. A model-based recognition method is utilized. The search models identify the fish directly from continuously imported dynamic raw images. Through a fitness function which evaluates the correlation degree between a model and the target, the recognition and simultaneous detection of the position/orientation (pose) can be converted to optimization problem. Then, GA is applied to visual recognition in dynamic scenes because of its high performance of optimization. The robot can successfully catch a fish by a net and a camera attached at the hand of the manipulator based on the real-time visual servoing. Figure 2 shows a state of the capturing by visual servoing. It has shown the effectiveness of the monocular model-based recognition method for manipulator visual servoing [5].

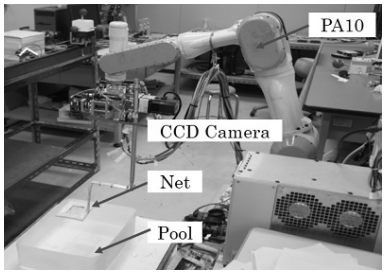


Fig. 1. Fish catching robot and experiment environment

In addition, the intelligence degree between fish and the robot has been evaluated by continuous catching and releasing operation. It is considered that the antagonistic relationship can be very meaningful as one way to discuss robotic intelligence [6]. The merit of the monocular camera is that the configuration is simple, and processing time seems to be less than that of the multi-cameras unit.

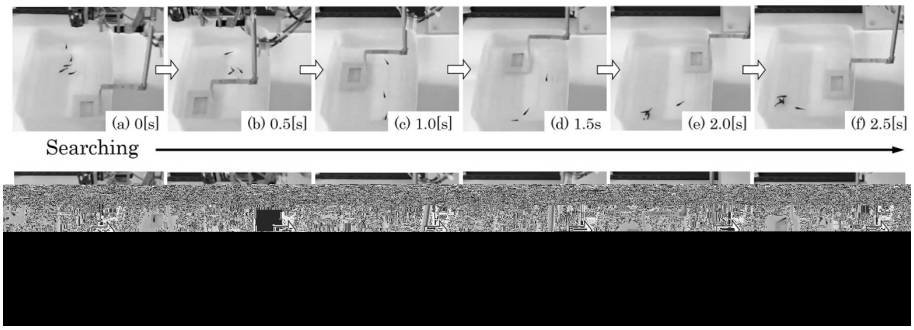


Fig. 2. Catching motion by visual servoing

On the other hand, single-camera 3D pose estimation has been studied thoroughly, but the estimated position accuracy in the camera depth of field has been proven insufficient. How the inherent complexities faced by a monocular system are resolved with binocular vision has been experimentally verified in [8]. Since the space recognition ability of a stereo vision is superior to the monocular camera, it is expected to be able to allow the robot to adapt to 3D estimation of the pose of a target object [7], [8].

However, there still have been difficulties for robots with vision cameras to accurately detect and handle the unique object under different kinds of environmental conditions. The approach in [7] to obtain visual information, by actively rocking the cameras, has not been used for real-time operations. Despite the fact that our previous study [6] can work for real-time visual servoing, it just detects the fish in a plane by a monocular camera. Even though the stereo vision was used aiming for an underwater vehicle, the methodologies in [7] were not applied in the underwater animal visual servoing test to prove the functionality and practicality of their proposed methods.

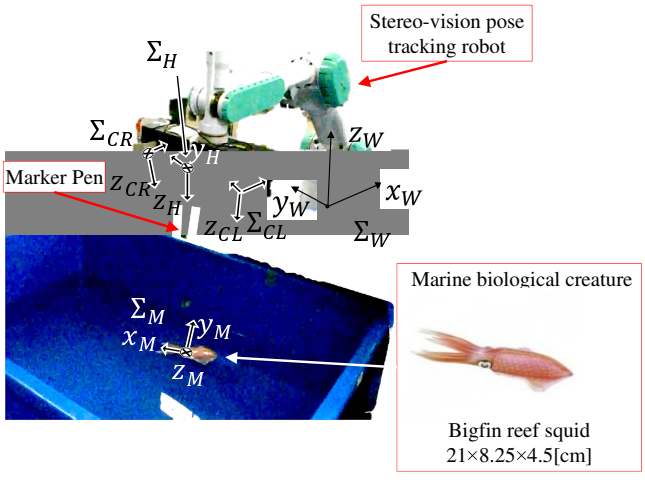


Fig. 3. Photo-model-based stereo-vision system

To overcome monocular vision's disadvantages and response new demands, binocular stereo-vision photo-model-based clothes handling robot was developed and the target clothes were still [9]. Although multi-view stereo with three or more cameras can give more details, it makes a system too complex and time consuming [10]. Therefore, this paper only talks about binocular stereo-vision.

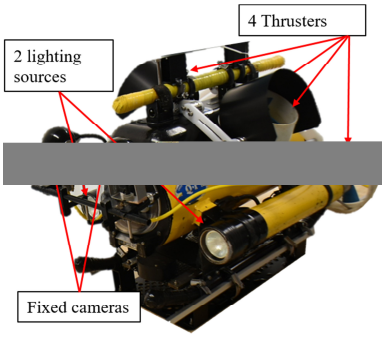


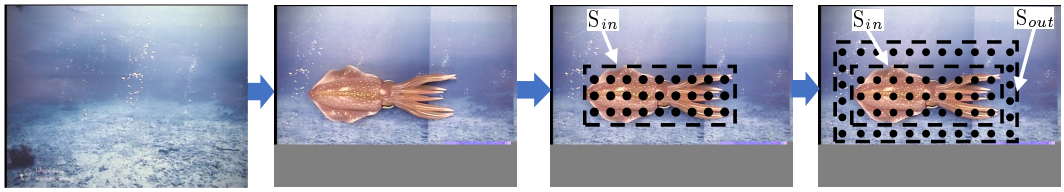
Fig. 4. Photograph of New ROV (DELTA-150).

Our previous study [11] expands the clothes handling robot to a real-time visual servoing system for moving underwater animal tracking with GA, which was used in the fishing-catching robot. As an optimization method, the GA has been improved to a "Real-time Multi-step Genetic Algorithm (RM-GA)" which has simplicity, repeatable ability and especially effectiveness in the real-time recognition performance [12]. Therefore, this paper applies the photo-model-based pose estimation method and RM-GA to video-rate stereo-vision pose visual servoing. The developed photo-model-based visual servoing system is shown in Fig. 3.

Even though frequency response visual servoing experiments were conducted in [11], the motion trajectories of the marine creature toys are sine curves which did not exist in nature. Moreover, the orientation of each object was fixed. In this paper, referring to the fish-catching research [6], position visual servoing has been conducted to moving marine creature toy and then the toy has been caught by a spear that was released from the robot. This experiment aims to verify the availability whether the proposed photo-model-based visual servoing system has an ability to catch aquatic creatures. In the test, the pose of the target object is not fixed.

More precisely, the contributions of this paper are as follows.

- A stereo-vision 3D perception method is proposed to estimate the pose of a 3D solid shape target by using



(a) Background (b) Target object in background (c) S_{in} model is set by black points group on the target (d) Strip area of S_{out} is set as it envelops S_{in}

Fig. 5. (a) shows a photograph of background image, (b) shows a photograph of the target object, the crab, in background, (c) represents a photograph of surface space model S_{in} by inner points group and (d) represents an outer points group of outside space of model S_{out} that enveloping S_{in}.

binocular stereo-vision and 2D photo-model.

- With the proposed method a photo-model-based stereo-vision visual servoing system (Fig. 3) is developed which can conduct visual servoing tasks.
- The developed system's visual servoing abilities will be verified through a position visual servoing experiment with a 3D moving object floating on a pool.

As shown in Fig. 4, in the future, the proposed photo-model-based stereo-vision system will be utilized on a ROV (DELTA-150). Now it is used for docking research [12].

II. Photo-model-based recognition

This section discusses the methodology of the proposed photo-model-based recognition method. Firstly, the kinematics of a binocular stereo-vision system will be described to make it easy to understand the recognition method. Secondly, the generation and matching of a photo-model are introduced. Then, an evaluation function is designed to convert the object recognition problem into an optimization problem. In the end, a genetic algorithm is chosen as a solution to the optimization problem to ensure that the recognition method can detect an object in real-time.

A. Photo-model generation

The model generation process is represented as Fig. 5. It should be noted that the photo-model is only part of a picture including target shape. Firstly, a background image is captured by the camera and the averaged hue value of the background image is calculated as shown in Fig. 5 (a). Then, the solid crab toy is put on the background. Take a 640×480 pixels picture at a distance of 400[mm] from the object as shown in Fig. 5 (b). As shown in Fig. 5 (c), a photo-model composed of dots with color information of hue is set as S_{in}. Finally, the outside space S_{out} of the model is generated by enveloping S_{in} as shown in Fig. 5 (d).

B. 3D photo-model-based matching

The developed photo-model-based visual servoing system as shown in Fig. 3. Each coordinate system is as follows:

- Σ_W: world coordinate system,
- Σ_H: end-effector (hand) coordinate system,
- Σ_M: target object coordinate system,
- Σ_{CL}, Σ_{CR}: left and right camera coordinate systems.

In Fig. 6, a generated photo-model is projected from the 3D space onto the left and right 2D searching planes. Each coordinate system is as follows:

- Σ_{Mj}: j-th model coordinate system,
- ${}^{Mj}\boldsymbol{r}_i$: position of an arbitrary i-th point on j-th 3D model in Σ_{Mj}, where ${}^{Mj}\boldsymbol{r}_i$ is a constant vector,
- Σ_{IL}, Σ_{IR}: left and right image coordinate systems,
- ${}^{IL}\boldsymbol{r}_i^j$, ${}^{IR}\boldsymbol{r}_i^j$: projected position on Σ_{IL} and Σ_{IR} of an arbitrary i-th point on j-th 3D model.

Based on Σ_W, the positions of Σ_H and Σ_M are ${}^W\boldsymbol{r}_H = [{}^Wx_H, {}^Wy_H, {}^Wz_H]^T$ and $[{}^Wx_M, {}^Wy_M, {}^Wz_M]^T$ respectively. The pose of Σ_M based on Σ_H, including three position variables and three orientation variables in quaternion, is

$$\begin{aligned} {}^H\boldsymbol{\phi}_M &= [{}^H\boldsymbol{r}_M^T, {}^H\boldsymbol{\varepsilon}_M^T]^T \\ &= [{}^Hx_M, {}^Hy_M, {}^Hz_M, {}^H\varepsilon_{1M}, {}^H\varepsilon_{2M}, {}^H\varepsilon_{3M}]^T. \end{aligned} \quad (1)$$

The pose of j-th 3D model based on Σ_H is represented as

$${}^H\boldsymbol{\phi}_M^j = [{}^Hx_M^j, {}^Hy_M^j, {}^Hz_M^j, {}^H\varepsilon_{1M}^j, {}^H\varepsilon_{2M}^j, {}^H\varepsilon_{3M}^j]^T. \quad (2)$$

For simplicity, the ${}^H\boldsymbol{\phi}_M^j$ is written as $\boldsymbol{\phi}_M^j$ hereafter. The homogeneous transformation matrix from Σ_H to Σ_{Mj} is defined as ${}^H\boldsymbol{T}_M^j(\boldsymbol{\phi}_M^j)$. The position vector of the i-th point in the left camera image coordinates ${}^{IL}\boldsymbol{r}_i^j$ can be described by using projective transformation matrix $\boldsymbol{P}_{CL}$ as,

$${}^{IL}\boldsymbol{r}_i^j = \boldsymbol{P}_{CL}{}^{CL}\boldsymbol{r}_i^j = \boldsymbol{P}_{CL}{}^{CL}\boldsymbol{T}_H{}^H\boldsymbol{T}_{Mj}(\boldsymbol{\phi}_M^j)^{Mj}\boldsymbol{r}_i. \quad (3)$$

And ${}^{IR}\boldsymbol{r}_i^j$ can also be described as the same manner like ${}^{IL}\boldsymbol{r}_i^j$.

The sub figure on the top of Fig. 6 shows a generated 3D solid model with its pose $S_{in}(\boldsymbol{\phi}_M^j)$ (inner dotted points) and the outside space enveloping $S_{in}(\boldsymbol{\phi}_M^j)$ denoted as outer dotted line $S_{out}(\boldsymbol{\phi}_M^j)$. The sub figure on the left/right bottom of Fig. 6 show the left/right 2D searching models $S_L(\boldsymbol{\phi}_M^j)$ and S_R $(\boldsymbol{\phi}_M^j)$ respectively. Both $S_L(\boldsymbol{\phi}_M^j)$ and $S_R(\boldsymbol{\phi}_M^j)$ consist of $S_{L,in}(\boldsymbol{\phi}_M^j)$ and $S_{L,out}(\boldsymbol{\phi}_M^j)$ and $S_{R,in}(\boldsymbol{\phi}_M^j)$ and $S_{R,out}(\boldsymbol{\phi}_M^j)$. The evaluation of the correlation between the projected model and the images including real target object that are input from the binocular cameras is defined as a fitness function.

(a) Target Object

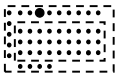


Fig. 6. Through projection transformation, a 3D solid model in the 3D searching space is projected to 2D left and right images. Searching models represented as $S_L(\boldsymbol{\phi}_M^j)$ and $S_R(\boldsymbol{\phi}_M^j)$

C. Definition of the fitness function

The correlation between the projected model and captured images on the left and right 2D searching areas is calculated by Eqs. (4)-(6).

$$
\begin{aligned}
F(\boldsymbol{\phi}_M^j) \quad = \quad & \Big\{ \Big[\sum_{{}^{IR}\boldsymbol{r}_i^j \in S_{R,in}(\boldsymbol{\phi}_M^j)} p_{R,in}({}^{IR}\boldsymbol{r}_i^j) + \sum_{{}^{IR}\boldsymbol{r}_i^j \in S_{R,out}(\boldsymbol{\phi}_M^j)} p_{R,out}({}^{IR}\boldsymbol{r}_i^j) \Big] \\
+ \quad & \Big[\sum_{{}^{IL}\boldsymbol{r}_i^j \in S_{L,in}(\boldsymbol{\phi}_M^j)} p_{L,in}({}^{IL}\boldsymbol{r}_i^j) + \sum_{{}^{IL}\boldsymbol{r}_i^j \in S_{L,out}(\boldsymbol{\phi}_M^j)} p_{L,out}({}^{IL}\boldsymbol{r}_i^j) \Big] \Big\} / (2N)
\end{aligned}
\tag{4}
$$

The evaluation of every point in the input image that lie inside the surface model frame and outside area of the model frame is represented as ${}^{IL}\boldsymbol{r}_i^j \in S_{L,in}(\boldsymbol{\phi}_M^j)$ and ${}^{IL}\boldsymbol{r}_i^j \in S_{L,out}$ $(\boldsymbol{\phi}_M)$ respectively. N is the total number of sampling points. Eqs. (5) and (6) are used for calculating $p_{L,in}({}^{IL}\boldsymbol{r}_i^j)$ and $p_{L,out}({}^{IL}\boldsymbol{r}_i^j)$ that are included in Eq. (4).

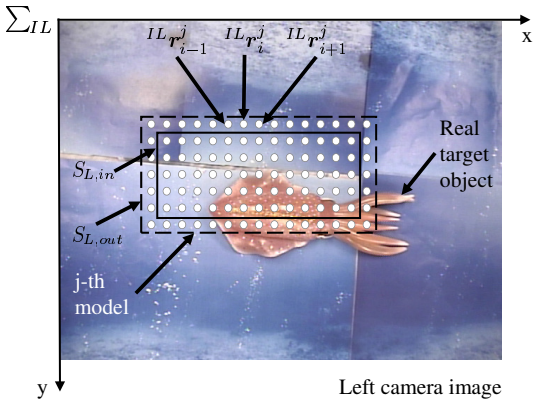
$$
p_{L,in}({}^{IL}\boldsymbol{r}_i^j) = \begin{cases} 2, \text{if}(|H_{IL}({}^{IL}\boldsymbol{r}_i^j) - H_{ML}({}^{IL}\boldsymbol{r}_i^j)| \leq 30); \\ -0.005, \text{else if}(|\bar{H}_B - H_{ML}({}^{IL}\boldsymbol{r}_i^j)| \leq 30); \\ 0, \text{otherwise.} \end{cases}
\tag{5}
$$

$$
p_{L,out}({}^{IL}\boldsymbol{r}_i^j) = \begin{cases} 0.1, & \text{if}(|\bar{H}_B - H_{IL}({}^{IL}\boldsymbol{r}_i^j)| \leq 20); \\ -0.5, & \text{otherwise.} \end{cases}
\tag{6}
$$

where

- $H_{IL}({}^{IL}\boldsymbol{r}_i^j)$: the hue value of the left camera image at the point ${}^{IL}\boldsymbol{r}_i^j$ (i-th point in $S_{L,in}$),
- $H_{ML}({}^{IL}\boldsymbol{r}_i^j)$: the hue value of the point ${}^{IL}\boldsymbol{r}_i^j$ (i-th point in $S_{L,in}$) on the model ,
- $\bar{H}_B$: the average hue value of the background image, i.e., Fig. 5 (a).

The evaluation values are tuned experimentally. In Eq. (5), if the hue value of each point of captured images, which lies



(a) Evaluation position ${}^{IL}\boldsymbol{r}_i^j$, that is i-th point of j-th model, which is projected on left image whose pose $\boldsymbol{\phi}_M^j$ is given by evolutionary process of GA.

(b) Classification of evaluation points (A)∼(D) on the photo model is explained. (A) represents points that satisfy the first case of Eq. (5), $|H_{IL}({}^{IL}\boldsymbol{r}_i^j) - H_{ML}({}^{IL}\boldsymbol{r}_i^j)| \leq 30$, representing that inner model $S_{L,in}$ overlaps with the real target. (B) does $|\bar{H}_B - H_{IL}({}^{IL}\boldsymbol{r}_i^j)| \leq 30$, representing that inner model $S_{L,in}$ overlaps with background. (C) does $|\bar{H}_B - H_{IL}({}^{IL}\boldsymbol{r}_i^j)| \leq 20$, meaning that the outer model $S_{L,out}$ overlaps with background, and (D) shows $S_{L,out}$ overlaps with the real target.

Fig. 7. Calculation of the matched degree of each point in model space ($S_{L,in}$ and $S_{L,out}$).

inside the surface model frame $S_{L,in}$, is similar to the hue value of each point in a model, the fitness value will increase with the voting value of "+2." The fitness value will decrease with the value of "−0.005" for every point of crabs in the left camera image when hue values of $S_{L,in}$ are similar to the average hue value of the background. Similarly, in Eq. (6), if the hue value of each point in the left camera image, which are in $S_{L,out}$, is near to the hue value of the background, with the tolerance of 20, the fitness value will increase with the value of "0.1." Otherwise, the fitness value will be decreased with the value of "−0.5." Similarly, a function $p_{R,in}({}^{IR}\boldsymbol{r}_i^j)$ and $p_{R,out}({}^{IR}\boldsymbol{r}_i^j)$ are represented for the right camera image.

D. Real-time Multi-step Genetic Algorithm (RM-GA)

Previous experiments have confirmed that the problem of recognizing the pose of the target object has been converted

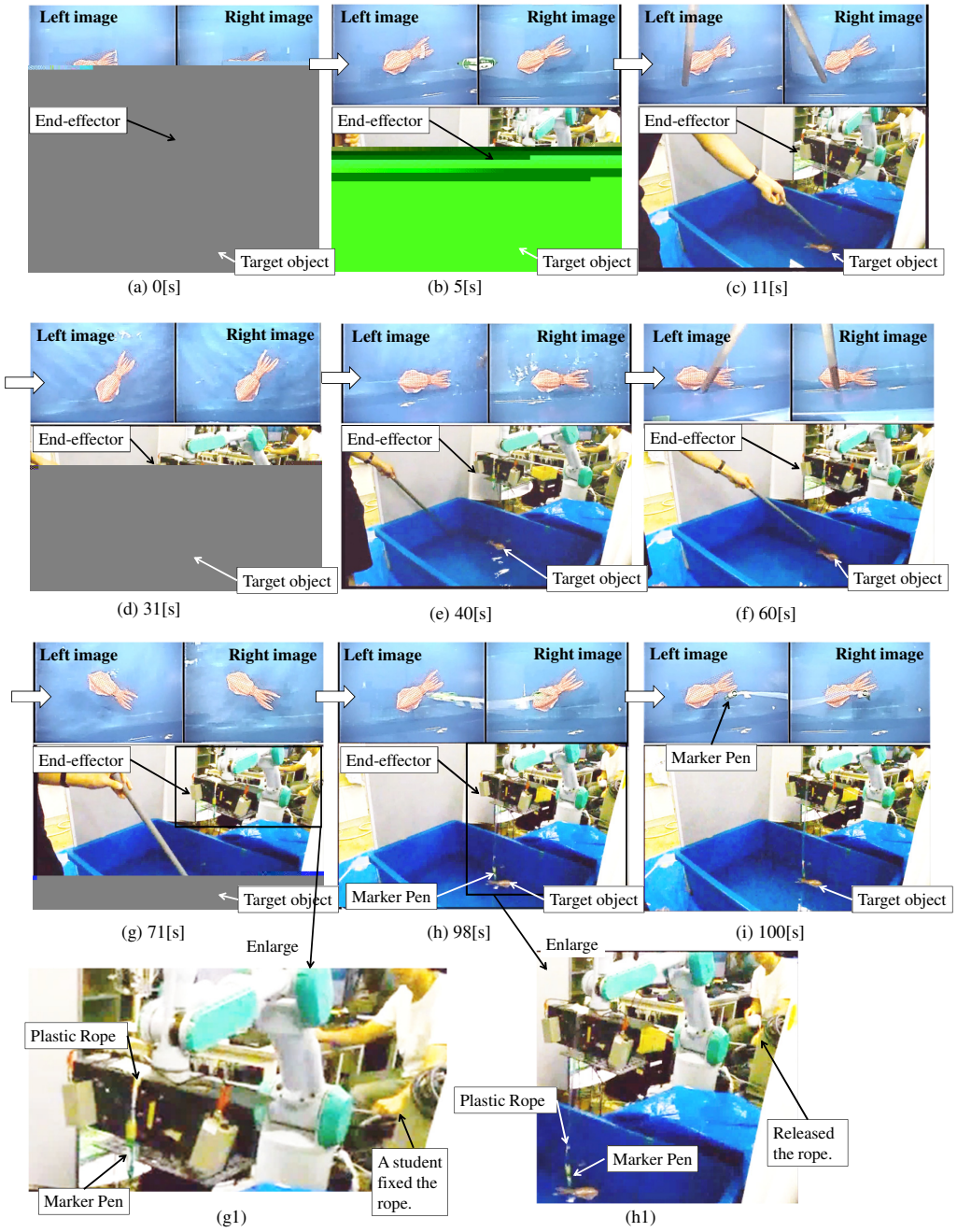


Fig. 8. Position 3DOF visual servoing experiment with pose (position/orientation) 6DOF estimation. The marker pen was tied on a rope and hung near the end-effector. From (a) to (g), the rope was fixed by a student. At (h), the rope is released, and the marker pen hit the squid. At the end (i), the squid drifted away due to the impact. (g1) and (h1) are enlarged views of a part of (g) and (h) respectively.

into an optimization problem of fitness. Moreover, the pose of the model with the highest fitness value can indicate the actual pose of the target object [11]. However, to measure the pose of the target object, trying all the possible pose with the model is very time consuming and cannot be used for real-time identification. On the real-time optimization problem, fishing research proves that GA is a simple and effective real-time recognition search method [6]. It has been improved to a Real-Time Multi-Step Genetic Algorithm (RM-GA). Within 33 [ms] (30[fps]), it can detect the pose of a target object with two images captured by the left and right cameras [12].

With this algorithm, as shown in Eq. (7), each chromosome consists of six variables, coded by 12[bit]. The first three variables of a model in 3D space $({}^{H}x_{M}^{j}, {}^{H}y_{M}^{j}, {}^{H}z_{M}^{j})$ are the position and the last three variables $({}^{H}\varepsilon_{1M}^{j}, {}^{H}\varepsilon_{2M}^{j}, {}^{H}\varepsilon_{3M}^{j})$ are the orientation based on Σ_H.

$$\overbrace{01\cdots01}^{{}^{H}x_{M}^{j}\ 12bits}\overbrace{00\cdots01}^{{}^{H}y_{M}^{j}\ 12bits}\overbrace{11\cdots01}^{{}^{H}z_{M}^{j}\ 12bits}\overbrace{01\cdots01}^{{}^{H}\varepsilon_{1M}^{j}\ 12bits}\overbrace{01\cdots11}^{{}^{H}\varepsilon_{2M}^{j}\ 12bits}\overbrace{01\cdots10}^{{}^{H}\varepsilon_{3M}^{j}\ 12bits}. \quad (7)$$

Readers can refer to [12], which has a more detailed explanation.

III. position 3DOF VISUAL SERVOING WITH POSE 6DOF ESTIMATION

A. Experimental environment and content

The pose visual servoing experimental environment is shown in Fig. 3. The utilized manipulator in the system is a PA-10 robot arm manufactured by Mitsubishi Heavy Industries. And two CCD cameras mounted on the end-effector are FCB-1X11A manufactured by Sony Corporation. The frame frequency of stereo cameras is set as 30[fps]. The image processing board, CT-3001, receiving the image from the CCD camera is connected to the host computer (CPU: Intel Core i7-3770, 3.40 GHz).

The purpose of the experiment is to verify the availability of the proposed photo-model-based visual servoing system for catching a marine creature.

As shown in Fig. 3, a squid toy is a target object. And a marker pen is hung on the end-effector and along the Z_H direction. In the experiment, as shown in Eq. 1, the stereo-vision detects the pose six parameters ${}^H\boldsymbol{\phi}_M$ of the object. With the detection results, the end-effector moves to the top of the squid object. The desired target position relationship $\boldsymbol{r}_d$ between the end-effector Σ_H and the object Σ_M is set as:

$$ {}^H\boldsymbol{r}_{dM} = [0, 0, 600][\text{mm}]. \tag{8} $$

The squid object floats on the water in the pool without pose constraints. In the end, a marker pen is released and falls off to spear the squid to confirm the position visual servoing ability.

B. Results and discussion of the experiment

Figure 8 shows the experimental states. At the beginning, in (a), the distance between Σ_H and Σ_M at the vertical direction was ${}^H z_M = {}^W z_H - {}^W z_M = 680$[mm]. In other directions, ${}^H x_M$ and ${}^H y_M$ were unknown. Figure 8 (b) shows that at 5[s], the visual servoing system has controlled the end-effector to move to the target position with a height about ${}^H z_M = 600$[mm]. Comparing with (a), the height in (b) had a significant drop. In two camera images of (b), the squid became bigger than that in (a). As shown in Fig. 8 (c) and (f), during the experiment, the squid target object was moved through a stick randomly and then by inertia. And in (d), (e) and (g), the wave was made by the stick to mimic the natural situation. In the end, in (h), the marker pen was released and hit the squid.

According to the results, even though there were waves and light reflection on the water, it can be seen that the visual servoing system can detect the object's pose and track it in time. And the system is not susceptible to partial occlusion conditions. It is confirmed from the experimental results that the system has an ability to conduct a visual servoing task for a moving target and has certain robustness against external disturbances.

IV. CONCLUSION

In this paper, photo-model-based stereo-vision 3D perception method has been presented. Based on it, a photo-model-based stereo-vision visual servoing system has been developed. In order to evaluate the adaptability of the system, a pool visual servoing experiment has been conducted. According to the experimental results, the merits of the proposed stereo-vision perception method has been confirmed as follows.

- The photo-model-based stereo-vision method is able to estimate the pose of a 3D solid shape target in real-time by using stereo-vision and 2D photo-model.
- The proposed binocular stereo-vision system can be utilized on a manipulator for visual servoing tasks.
- The developed stereo-vision visual servoing system can work for a moving marine creature toy visual servoing in a pool.

In the future, the proposed photo-model-based stereo-vision system will be utilized on a ROV (DELTA-150) for marine creature catching.

REFERENCES

[1] Aoyagi, S., Hattori, N., Kohama, A., Komai, S., Suzuki, M., Takano, M. and Fukui, E., Object detection and recognition using template matching with SIFT features assisted by invisible floor marks. Journal of Robotics and Mechatronics, Vol.21, No. 6, pp. 689-697, 2009.

[2] Tomono, M., 3D object modeling and segmentation using image edge points in cluttered environments. Journal of Robotics and Mechatronics, Vol. 21, No. 6, pp.672-679, 2009.

[3] Dayoub, F., Dunbabin, M. and Corke, P., Robotic detection and tracking of crown-of-thorns starfish. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1921-1928, 2015.

[4] Stavnitzky, J. and Capson, D., Multiple camera model-based 3-D visual servo. IEEE transactions on robotics and automation, Vol. 16, No. 6, pp. 732-739, 2000.

[5] Suzuki, H. and Minami, M., Visual servoing to catch fish using global/local GA search. IEEE/ASME Transactions on Mechatronics, Vol. 10, No. 3, pp. 352-357, 2005.

[6] Inukai, H., Minami, M. and Yanou, A., Validity analysis of chaos generated with Neural-Network-Differential-Equation for robot to reduce fish's learning speed. International Journal of Applied Electromagnetics and Mechanics, Vol. 52, No. 3-4, pp. 883-889, 2016.

[7] Sagara, S., Ambar, R.B. and Takemura, F., A Stereo Vision System for Underwater Vehicle-Manipulator Systems-Proposal of a Novel Concept Using Pan-Tilt-Slide Cameras-. JRM, Vol. 25, No. 5, pp. 785-794, 2013.

[8] Negahdaripour, S. and Firoozfam, P., An ROV stereovision system for ship-hull inspection. IEEE Journal of oceanic engineering, Vol. 31, No. 3, pp. 551-564, 2006.

[9] Phyu, K.W., Funakubo, R., Hagiwara, R., Tian, H. and Minami, M., Verification of photo-model-based pose estimation and handling of unique clothes under illumination varieties. Journal of Advanced Mechanical Design, Systems, and Manufacturing, Vol. 12, No. 2, pp. JAMDSM0047-JAMDSM0047, 2018.

[10] Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006, June. A comparison and evaluation of multi-view stereo reconstruction algorithms. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 1, pp. 519-528). IEEE.

[11] Tian, H., Kou, Y. and Minami, M., Visual servoing to arbitrary target with photo-model-based recognition method, 24th International Symposium on Artificial Life and Robotics, B-Con PLAZA, Beppu, JAPAN, pp. 950-955, 2019.

[12] Myint, M., Yonemori, K., Lwin, K.N., Yanou, A. and Minami, M., Dual-eyes vision-based docking system for autonomous underwater vehicle: an approach and experiments. Journal of Intelligent & Robotic Systems, Vol. 92, No. 1, pp. 159-186, 2018.