The Twenty-Fourth International Symposium on Artificial Life and Robotics 2019 (AROB 24th 2019),
The Fourth International Symposium on BioComplexity 2019 (ISBC 4th 2019),
B-Con Plaza, Beppu, Japan, January 23-25, 2019

Visual servoing to arbitrary target with photo-model-based recognition method

Hongzhi Tian[1], Yejun Kou[1], and Mamoru Minami[1]

[1]Okayama University, Japan
(Tel: 81-86-251-8233, Fax: 81-86-251-8233)
[1]psnc8ytd@s.okayama-u.ac.jp

Abstract: Visual servoing is one of the methods to control robots. By incorporating visual information obtained from the installed vision sensor into the feedback loop, it is desirable for the robot to be able to operate in a changing environment or an unknown environment. For visual servo systems, the authors proposed a photo-model-based recognition method for processing an arbitrary object with a still picture. In the previous work, a flat shape clothes handling robotic system has been proposed to handle deformable and unique clothes. In this paper, we extend the photo-model-based robot handling system (pick and place) to a real-time pose tracking system. And the recognition method is used for attitude tracking of different objects of dynamic pictures. Furthermore, to verify the ability to track with the photo-model-based recognition method. The authors design some frequency response experiments with arbitrary aquatic creature toys to keep the relative pose between a sea animal and hand-eye. The results of visual servoing experiments show that the proposed identification method is feasible, flexible and effective.

Keywords: visual servoing, photo-model-based recognition, genetic algorithm(GA)

1 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.

However, until now, robots cannot entirely replace humans. While human beings can conduct intended tasks in pending circumstances, an automated robot is not adept at being similarly adaptable. Therefore, the researchers have tried to improve the abilities of automated robots.

About automated robots, the visual servoing, a robot control technology using visual information obtained from a vision sensor (camera) in the feedback loop, is expected to be able to allow the robot to adapt to changing or unknown environments [1, 2, 3].

In the previous works [4] and [5], a photo-model-based matching method has been proposed. With this method, we developed a clothes-handling robot. In [6], 3D recognition accuracy has been confirmed experimentally by using 12 different samples cloths. Except for static clothes handling, there is a need for tracking arbitrary moving target object. With photo-model-based recognition method, we developed a visual servoing system. As shown in Fig. 1. The dual-eye cameras that are fixed at the end-effector of a PA-10 robot perform the object recognition and pose estimation process based on the photograph model.

For verifying the ability of tracking an arbitrary object, in this paper, the authors conduct some frequency response experiments with arbitrary aquatic creature toys to keep the relative pose between a sea animal and hand-eye. The results of visual servoing experiments show that the proposed identification method is feasible, flexible and effective.

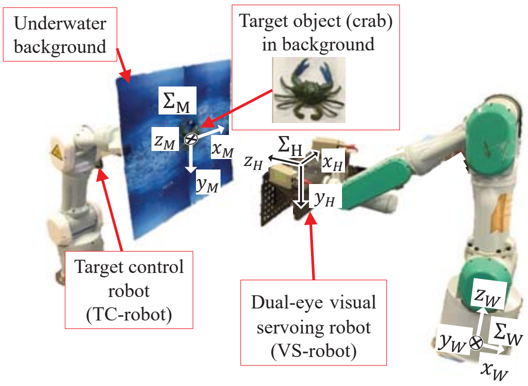

Underwater background
Target object (crab) in background
Σ_M
z_M
x_M
y_M
Σ_H
z_H
x_H
y_H
Target control robot (TC-robot)
Dual-eye visual servoing robot (VS-robot)
z_W
Σ_W
y_W
x_W

Fig. 1. The visual servoing system with dual eye-in-hand cameras. World coordinate system (Σ_W) and target coordinate system (Σ_M)

2 PHOTO-MODEL-BASED RECOGNITION

To make it easier to understand the photo-model-based recognition method, we will describe the kinematics of stereo-vision before an explanation of the proposed system in details.

2.1 Kinematics of stereo-vision

Figure 2 shows a perspective projection of the dual-eyes vision system. Each coordinate system are as follows:

- Σ_W: world coordinate system,
- Σ_H: end-effector (hand) coordinate system, as
- Σ_{CL}, Σ_{CR}: left and right camera coordinate systems,
- Σ_{IL}, Σ_{IR}: left and right image coordinate systems,

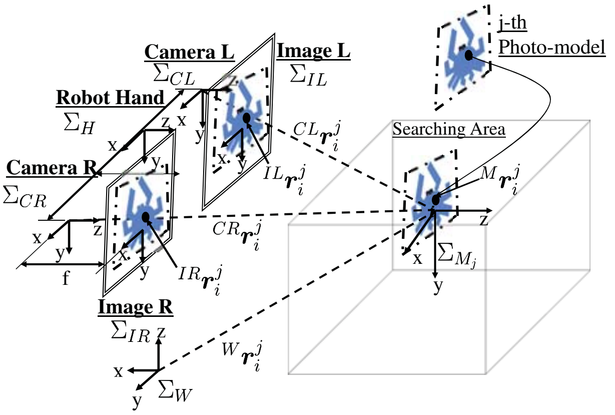


Fig. 2. Perspective projection of dual-eye vision-system: In the searching area, a 3D solid model is represented by the picture of crab with black point (j-th photo model)

- Σ_{M_j}: j-th model coordinate system,
- Σ_M: object coordinate system.
- $^{M}\boldsymbol{r}_i^j$: position of an arbitrary i-th point on j-th 3D model in Σ_{M_j}, where $^{M}\boldsymbol{r}_i^j$ is a constant vector
- $^{CR}\boldsymbol{r}_i^j$ and $^{CL}\boldsymbol{r}_i^j$: position of an arbitrary i-th point on j-th 3D model based on Σ_{CR} and Σ_{CL}
- $^{IL}\boldsymbol{r}_i^j$ and $^{IR}\boldsymbol{r}_i^j$: projected position on Σ_{IL} and Σ_{IR} of an arbitrary i-th point on j-th 3D model

The homogeneous transformation matrix from Σ_{CR} to Σ_M is defined as $^{CR}\boldsymbol{T}_M(\boldsymbol{\phi}_M^j, \boldsymbol{q})$, where $\boldsymbol{\phi}_M^j$ is j-th model's pose and $\boldsymbol{q}$ means robot's joint angle vector. Then, $^{CR}\boldsymbol{r}_i^j$ can be calculated by using Eq. (1),

$$^{CR}\boldsymbol{r}_i^j = {}^{CR}\boldsymbol{T}_M(\boldsymbol{\phi}_M^j, \boldsymbol{q})\ {}^{M}\boldsymbol{r}_i^j. \tag{1}$$

The position vector of the i-th point in the right and left camera image coordinates $^{IR}\boldsymbol{r}_i^j$ can be described by using projective transformation matrix $\boldsymbol{P}_k$ as,

$$^{IR}\boldsymbol{r}_i^j = \boldsymbol{P}_k\ {}^{CR}\boldsymbol{r}_i^j = \boldsymbol{P}_k {}^{CR}\boldsymbol{T}_M(\boldsymbol{\phi}_M^j) {}^{M}\boldsymbol{r}_i^j \tag{2}$$

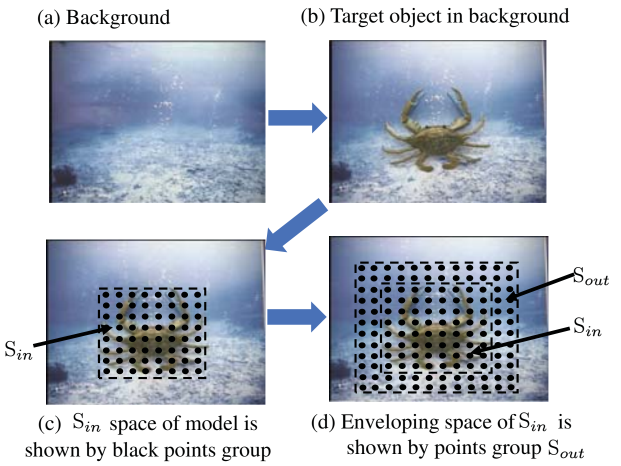
Then, $^{IR}\boldsymbol{r}_i^j$ can be described as,

$$\begin{cases} {}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j) = \boldsymbol{f}_R(\boldsymbol{\phi}_M^j, {}^{M}\boldsymbol{r}_i^j) \\ {}^{IL}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j) = \boldsymbol{f}_L(\boldsymbol{\phi}_M^j, {}^{M}\boldsymbol{r}_i^j) \end{cases} \tag{3}$$

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

2.2 Model generation

The model generation process is represented as Figure 3. It should be noted that the model is only part of a picture and the picture is not the model. Firstly, a background image is captured by the first camera and the averaged hue value of the background image is calculated as shown in Fig. 3 (a). Then, the crab is put on the background. Take a 640 × 480 pixels picture at a distance of 400[mm] from the object as shown in Fig. 3 (b). As shown in Fig. 3 (c), scan from the four corners at the same time in the arrow direction compare with the averaged hue value of the background, and generate the surface space S_{in} of the model. Finally, the outside space S_{out} of the model is generated by enveloping S_{in} as shown in Fig. 3 (d). It is gotten that the size of the tangential plane of the object in the real 3D space.



(c) S_{in} space of model is shown by black points group

(d) Enveloping space of S_{in} is shown by points group S_{out}

Fig. 3. (a) shows a photograph of background image, (b) shows a photograph of the target object (the blue crab) in background, (c) represents a photograph of surface space model S_{in} by inner points group and (d) represents a photograph of outside space of model S_{out} that enveloping S_{in}.

2.3 3D photo-model-based matching

In Fig. 4, a generated solid model is projected from the 3D space onto the left and right 2D searching planes. The sub figure on the top of Fig. 4 shows a generated 3D solid model with its pose $S_{in}(\boldsymbol{\phi}_M^j)$ (inner dotted points) and the outside space enveloping $S_{in}(\phi_M)$ denoted as outer dotted line ($S_{out}(\boldsymbol{\phi}_M^j)$). The sub figure on the left/right bottom of Fig. 4 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 from the dual-eye cameras attached at the end-effector defined as a fitness function.

2.4 Definition of the fitness function

The concept of the fitness function in this study can be said to be an extension of the work in [8] in which different models including a rectangular shape surface-strips model was evaluated using images from a single camera. The correlation between the projected model ϕ and captured images on the left and right 2D searching areas is calculated by the

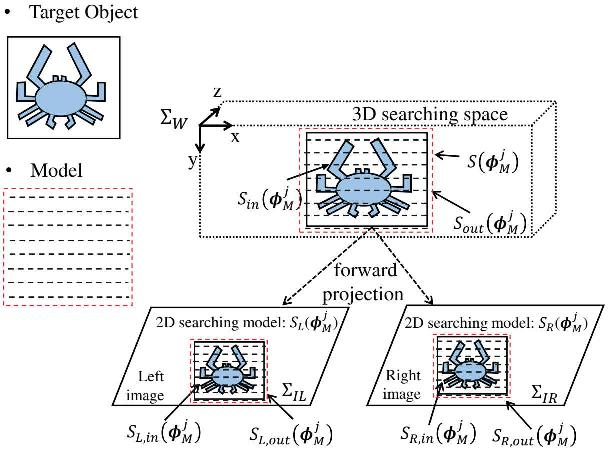

• Target Object
• Model
Σ_W
z
x
y
3D searching space
$S(\boldsymbol{\phi}_M^j)$
$S_{in}(\boldsymbol{\phi}_M^j)$
$S_{out}(\boldsymbol{\phi}_M^j)$
forward
projection
2D searching model: $S_L(\boldsymbol{\phi}_M^j)$
2D searching model: $S_R(\boldsymbol{\phi}_M^j)$
Left
image
Right
image
Σ_{IL}
Σ_{IR}
$S_{L,in}(\boldsymbol{\phi}_M^j)$
$S_{L,out}(\boldsymbol{\phi}_M^j)$
$S_{R,in}(\boldsymbol{\phi}_M^j)$
$S_{R,out}(\boldsymbol{\phi}_M^j)$

Fig. 4. 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 $S_R(\boldsymbol{\phi}_M^j)$

equations Eq. (4) to Eq. (6).

$$F(\boldsymbol{\phi}_M^j) = \left\{ \left(\sum_{\substack{{}^{IR}\boldsymbol{r}_i^j \in \\ S_{R,in}({}^{CR}\boldsymbol{\phi}_M^j)}} p({}^{IR}\boldsymbol{r}_i^j) + \sum_{\substack{{}^{IR}\boldsymbol{r}_i^j \in \\ S_{R,out}({}^{CR}\boldsymbol{\phi}_M^j)}} p({}^{IR}\boldsymbol{r}_i^j) \right) \right.$$
$$\left. + \left(\sum_{\substack{{}^{IL}\boldsymbol{r}_i^j \in \\ S_{L,in}({}^{CL}\boldsymbol{\phi}_M^j)}} p({}^{IL}\boldsymbol{r}_i^j) + \sum_{\substack{{}^{IL}\boldsymbol{r}_i^j \in \\ S_{L,out}({}^{CL}\boldsymbol{\phi}_M^j)}} p({}^{IL}\boldsymbol{r}_i^j) \right) \right\} / (2N) \tag{4}$$

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) is used for calculating $p_{L,in}({}^{IL}\boldsymbol{r}_i^j)$ and $p_{L,out}({}^{IL}\boldsymbol{r}_i^j)$.

$$p_{L,in}({}^{IL}\boldsymbol{r}_i^j) = \begin{cases} 2, \text{ if}(|H_{IL}({}^{IL}\boldsymbol{r}_i^j) - H_{ML}({}^{IL}\boldsymbol{r}_i^j)| \leq 30); \\ -1, \text{if}(|H_{IL}({}^{IL}\boldsymbol{r}_i^j) - H_{ML}({}^{IL}\boldsymbol{r}_i^j)| \geq 50); \\ -0.005, \quad \text{if}(|\bar{H}_B - H_{ML}({}^{IL}\boldsymbol{r}_i^j)| \leq 30); \\ 0, \quad \text{otherwise.} \end{cases} \tag{5}$$

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

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

The evaluation values are tuned experimentally. In Eq. (5), if the hue value of each point of captured images, which lies inside the surface model frame $S_{L,in}$, is same 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 that 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 same 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.

2.5 Fitness distribution of position

To evaluate the feasibility of fitness function Eq. (4), a good way is a brute-force search or an exhaustive search. For still pictures of a moment, the fitness values of all candidates that represent different poses of models are calculated. We call it "fitness distribution."

In fact, for the measurement of position or orientation, it is impossible to exhaust all possibilities. For fitness distribution, the accuracy of position ${}^{H}x_M -{}^{H} y_M$ is 5[mm]. We prepared 3D toys of 6 marine creatures as shown in Fig. 5 (a). The four labels corresponding to each model are number, English name, size and Japanese name. To recognize, all the objects were taken pictures and saved in a database as shown in Fig. 5 (b). The true value of all objects is

$${}^{H}\boldsymbol{\psi}_M = [0, 0, 500, 0, 0, 0][mm]. \tag{7}$$

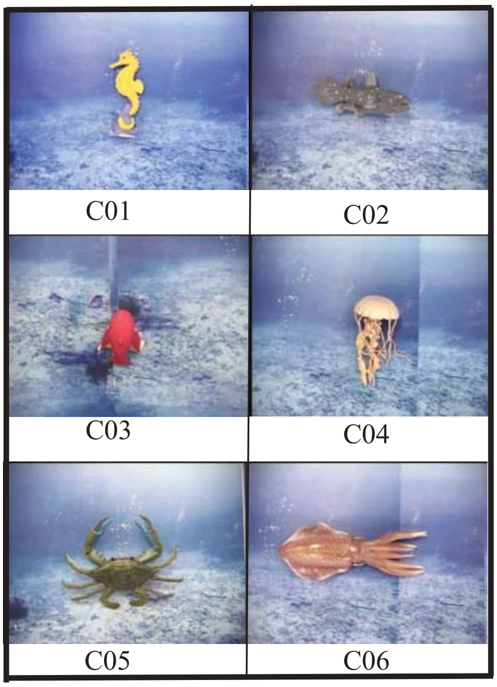
Fig. 6 are experimental results. In each frame, the lower right two pictures are the images captured by the left and right camera. And target object is explained on the left side. From the results, each peak appears near the true value. The error is small. The last experiment searches C02 from several objects. Since its peak is near the true value, it can be said that this method can search for an object from a complex background.

2.6 Real-time Multi-step Genetic Algorithm (RM-GA)

As shown in Fig. 8, searching for all possible models is time-consuming for real-time recognition. Therefore, we convert the problem of finding/recognizing the target object's pose into an optimization problem with a multi-peak distribution. GA is a simple and effective searching method.For real-time recognition through one frame image in 33[ms], we have proposed a Real-time Multi-step Genetic Algorithm

C01	Seahorse	C02	Coelacanth
13×4.5 ×2.7	タツノオトシゴ	7.0×14.5 ×6.0	シーラカンス
C03	Dolphin	C04	Jellyfish
8×6×4.5	イルカ	9×9×11	クラゲ
C05	Crab	C06	Bigfin reef squid
17.5×14×4	カニ	21×8.25	アオリイカ

(a) Details of 6 target toys (Unit: cm)



(b) 12 target images in database

Fig. 5. (a) 6 marine biological models. Code name is from C01 to C06. The size of each 3D toy is shown in left down of each frame. The four labels corresponding to each model are number, English name, size and Japanese name. (b) pictures of marine biological models with blue sea background. The size of each picture is 640 × 480 pixels. It should be noted that the model is only part of a picture and the picture is not the model.

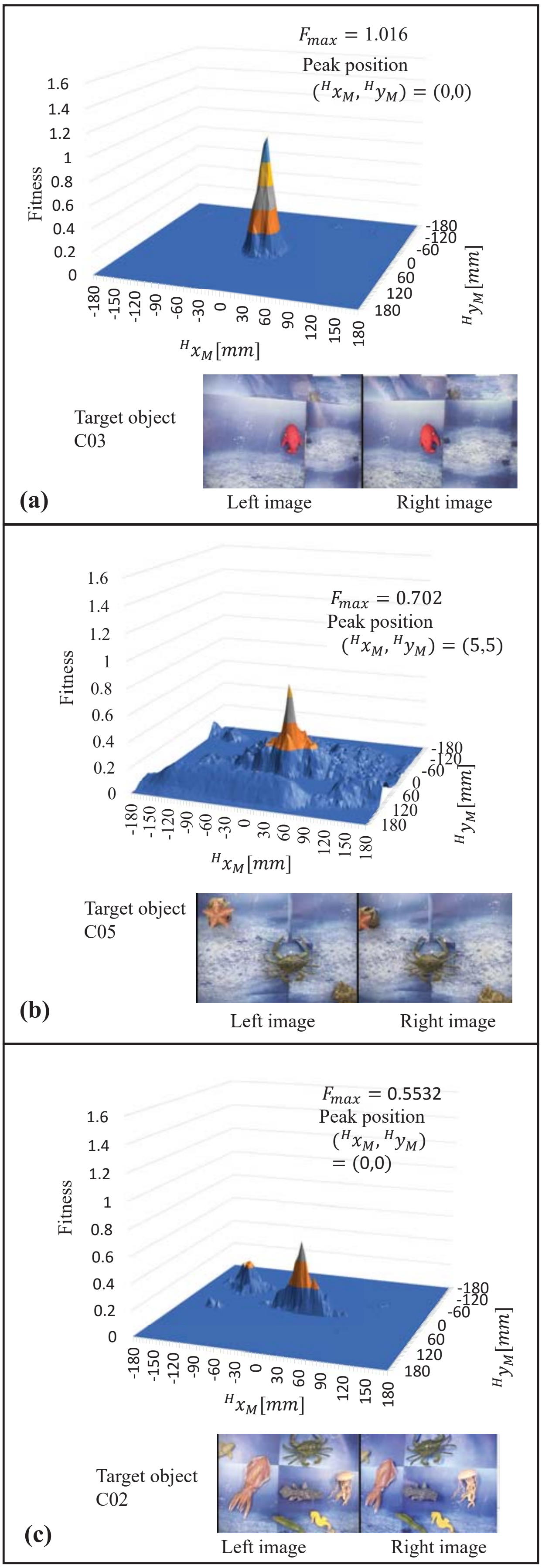


Fig. 6. Due to the limited space, this figure shows fitness distribution results of object C02, C03, and C05 on ${}^{H}x_M - {}^{H}y_M$ plane. C03 and C05 are chosen randemly for later visual servoing experiment. In frame (c), target object is C02. We verify the robustness of this recognition method to the background with (c) experiment.

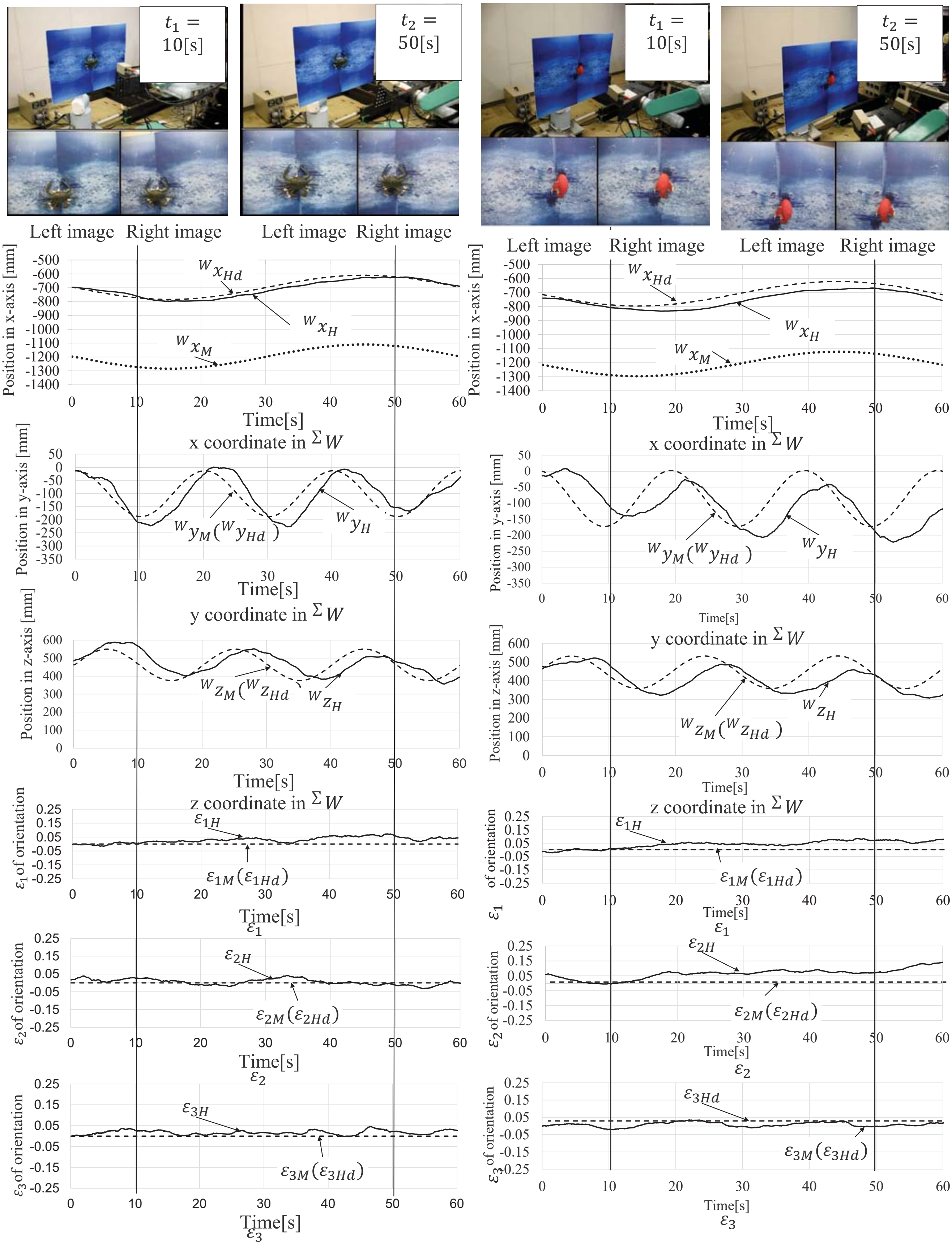
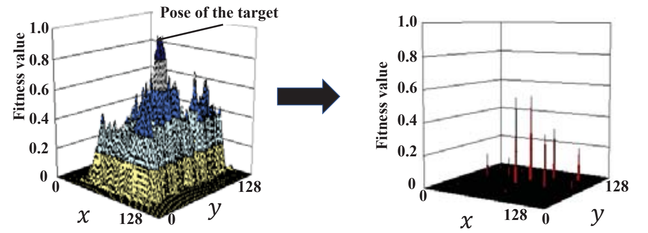


Fig. 7. Pose estimation and tracking performance experiments of the 3D solid target object. Positions of crab and dolphin change in sine wave. Object's orientation does not change.



Fitness distribution: Exploring every possible location takes a lot of time.

RM-GA：Calculation of some selected poses to save time.

Fig. 8. we have proposed Real-time Multi-step Genetic Algorithm (RM-GA) for searching the pose of target object in real-time.

(RM-GA)[7] 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. With this algorithm, Each chromosome consists of six variables. Each variable is coded by 12bits that can provide sufficient accuracy to get the optimal solution. The first three variables of a model in 3D space (t_x, t_y, t_z) are represented as the position and the last three variables $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are represented as the orientation.

$$\underbrace{\overbrace{01\cdots01}^{t_x}}_{12bits}\underbrace{\overbrace{00\cdots01}^{t_y}}_{12bits}\underbrace{\overbrace{11\cdots01}^{t_z}}_{12bits}\underbrace{\overbrace{01\cdots01}^{\varepsilon_1}}_{12bits}\underbrace{\overbrace{01\cdots11}^{\varepsilon_2}}_{12bits}\underbrace{\overbrace{01\cdots10}^{\varepsilon_3}}_{12bits}.$$

Readers can refer to [7], which has a more detailed explanation.

3 EXPERIMENTAL ENVIRONMENT

Two frequency response experiments with crab and dolphin toy are conducted to confirm the tracking ability of the developed photo-model-based visual servoing system. As shown in Fig. 1, during the experiments the desired position and orientation is the same as Eq. (7). The target trajectories of the two experiments are the same. Both of them are sine curves with an amplitude of 100 [mm], a period of 20 [s] in the x and y-axis directions and amplitude of 100 [mm], a period of 60 [s] in the z-axis direction.

4 EXPERIMENTAL RESULTS AND DISCUSSION

In Fig.7, ${}^{W}\boldsymbol{r}_H$ and ${}^{W}\boldsymbol{\varepsilon}_H$ are the tracking results of the end-effector. ${}^{W}\boldsymbol{r}_M$ and ${}^{W}\boldsymbol{\varepsilon}_M$ are the motion of target object. And the desired positions of end-effector ${}^{W}\boldsymbol{r}_{Hd}$ and ${}^{W}\boldsymbol{\varepsilon}_{Hd}$ are ideal positions during the experiments. It can be seen that even though the tracking curves delay somewhat in phase, the visual servoing system with photo-model-based recognition method can track the object in time. And different objects have different effects on the tracking results. On this point, we will conduct further research and discussion.

5 CONCLUSION

The visual servoing experiments were conducted to confirm the performance of the photo-model-based recognition method. According to the experimental results, this system can recognize and track the 3D crab and dolphin toys with the prepared pictures.

We conclude that if we have prepared the pictures of objects the system can recognize them and track them. However, different objects seem to have different effects on the tracking performance of the system. In the future, we would like to discuss this problem.

REFERENCES

[1] S.Hutchinson, G.Hager, and P.Corke, A Tutorial on Visual Servo Control, IEEE Trans. on Robotics and Automation, vol. 12, no. 5, 1996, pp. 651-670.

[2] P.Y.Oh, and P.K.Allen, Visual Servoing by Partitioning Degrees of Freedom, IEEE Trans. on Robotics and Automation, vol. 17, no. 1, 2001, pp. 1-17.

[3] P.K.Allen, A.Timchenko, B.Yoshimi, and P.Michelman, Automated Tracking and Grasping of a Moving object with a Robotic Hand-Eye System, IEEE Trans. on Robotics and Automation, vol. 9, no. 2, pp. 152-165, 1993.

[4] Funakubo R, Phyu KW, Tian H, Minami M, Recognition and handling of clothes with different pattern by dual hand-eyes robotic system, IEEE/SICE International Symposium, 2016, pp. 742-747.

[5] Phyu KW, Cui Y, Tian H, Hagiwara R, Funakubo R, Yanou A, Minami M, Accuracy on Photo-Model-Based Clothes Recognition, SICE Annual Conference, Tsukuba, Japan, 2016, September 20-23.

[6] Phyu, K.W., Funakubo, R., Fumiya, I., Shinichiro, Y. and Minami, M., Verification of recognition performance of cloth handling robot with photo-model-based matching, IEEE International Conference on Mechatronics and Automation (ICMA), 2017, pp. 1750-1756.

[7] 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, 92(1), 2018, pp.159-186.

[8] Minami M, Agbanhan J, Asakura T, Evolutionary scene recognition and simultaneous position/orientation detection, in Soft Computing in Measurement and Information Acquisition, Springer Berlin Heidelberg, 2003, pp. 178-207.