

Visual Servoing to Arbitrary Target by Using Photo-Model Definition

- Hongzhi TIAN, Member, Okayama University, psnc8ytd@s.okayama-u.ac.jp
 Yejun KOU, Student Member, Okayama University
 Khaing Win PHYU, Okayama University
 Ryuki FUNAKUBO, Okayama University
 Mamoru MINAMI, Member, Okayama University

Visual servoing is one of the methods to control robots. By incorporating visual information obtained from the installed visual sensor into the feedback loop, it is expected to enable robots to work in changing environments or unknown environment. For a visual servoing system, it is difficult to quickly switch the tracking object for different tasks. Using a dual hand-eye cameras manipulator, authors have proposed photo-model-based recognition method for handling different objects with static pictures. In this paper, the recognition method is used for the pose tracking of different objects with dynamic pictures. The visual servoing experimental results demonstrate the proposed recognition method are feasible, flexible and effective.

Key Words: Visual servoing, Photo-model-based matching, Arbitrary target, Dual hand-eye manipulator

1 Introduction

From the instant of birth, human beings are talented at identifying and distinguishing different objects. However, human have to go through a period of education in order to understand objects' names. For example, as shown in 1, let a student pick up a pencil. Firstly, the student must already know what the pencil is. Similarly, if a human want a robot to instead of him to pick up a pencil, a book or a phone, firstly, he needs to teach the robot what they are. Only pure image recognition, there are many great researches. However, it is still difficult to identify the distance and posture of an object only by using cameras without other sensors.

In the previous works [1] and [2], a photo-model-based matching method has been proposed. With this method, the robot will know what the object is and how much it is. In [3], 3D recognition accuracy has been confirmed experimentally by using 12 different samples cloths. Their pictures have been prepared in advance.

Furthermore, about real time recognition and tracking, a robot control technology called visual servoing is playing an important role in many of applications. The developed visual servoing system is shown in Fig. 2. 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. The proposed photo-model-based recognition system is intended to get better performance and higher accuracy than fisherman. In this paper, the authors design experiments with arbitrary aquatic creature toys. Take pictures of them to teach robot what they are. And conduct visual servoing tracking experiments for verifying the ability of tracking with the photo-model-based recognition method.

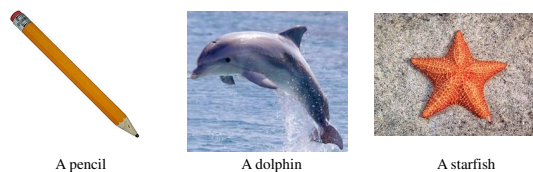


Fig.1 A language is a tool. Every creature, object, or concept is named. If people haven't learned these names they cannot communicate and collaborate.

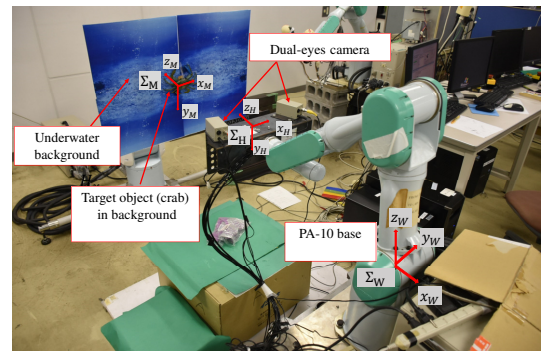


Fig.2 Visual servoing system with dual eye-in-hand cameras. World coordinate system (Σ_W) and target coordinate system (Σ_M)

2 Photo-model-based recognition

This section is to explain the photo-model-based crab recognition method. The model-based matching method [4] adopt a set-point-model-thinking. That is all points of the solid model in the 3D searching space as a group are projected onto the left and right camera image planes (2D image) without the miss-comparing problem. So, all projections for each point are correct. Here is the description of the kinematics of stereo-vision before an explanation of the proposed system in details.

2.1 Kinematics of stereo-vision

Figure 3 shows a perspective projection of the dual-eyes vision system. The coordinate systems of dual-eyes cameras and the target object in Fig. 3 consist of world coordinate system Σ_W , j-th model coordinate system Σ_{M_j} , hand coordinate system Σ_H , camera coordinate systems as Σ_{CL} and Σ_{CR} , and image coordinate systems as Σ_{IL} and Σ_{IR} . In Fig. 3, the position vectors of an arbitrary i-th point of the j-th 3D model Σ_{M_j} based on each coordinate system are as follows:

- ${}^M \mathbf{r}_i^j$: position of an arbitrary i-th point on j-th 3D model

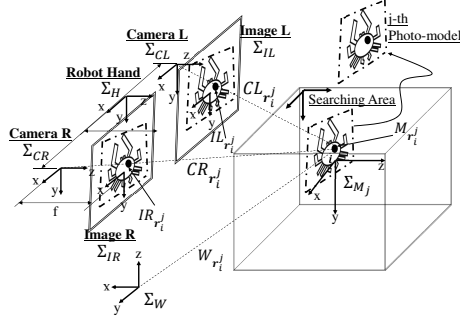


Fig.3 Perspective projection of dual-eye vision-system: In the searching area, a 3D solid model is represented by the picture of crab with black point (j-th photo model)

in Σ_{M_j} , where ${}^M\mathbf{r}_i^j$ is a constant vector

- ${}^{CR}\mathbf{r}_i^j$ and ${}^{CL}\mathbf{r}_i^j$: position of an arbitrary i-th point on j-th 3D model based on Σ_{CR} and Σ_{CL}
- ${}^{IL}\mathbf{r}_i^j$ and ${}^{IR}\mathbf{r}_i^j$: projected position on Σ_{IL} and Σ_{IR} of an arbitrary i-th point on j-th 3D model

The homogeneous transformation matrix from the right camera coordinate system Σ_{CR} to the target object coordinate system Σ_M is defined as ${}^{CR}T_M(\phi_M^j, \mathbf{q})$, where ϕ_M^j is j-th model's pose and $\mathbf{q}$ means robot's joint angle vector. Then, ${}^{CR}\mathbf{r}_i^j$ can be calculated by using Eq. (1),

$${}^{CR}\mathbf{r}_i^j = {}^{CR}T_M(\phi_M^j, \mathbf{q}) {}^M\mathbf{r}_i^j. \quad (1)$$

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

$${}^{IR}\mathbf{r}_i^j = \mathbf{P}_k {}^{CR}\mathbf{r}_i^j = \mathbf{P}_k {}^{CR}T_M(\phi_M^j, \mathbf{q}) {}^M\mathbf{r}_i^j \quad (2)$$

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

$$\begin{cases} {}^{IR}\mathbf{r}_i^j(\phi_M^j) = \mathbf{f}_R(\phi_M^j, {}^M\mathbf{r}_i^j) \\ {}^{IL}\mathbf{r}_i^j(\phi_M^j) = \mathbf{f}_L(\phi_M^j, {}^M\mathbf{r}_i^j) \end{cases} \quad (3)$$

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

2.2 Model generation

The model generation process is represented as Figure 4. 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. 4 (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. 4 (b). As shown in Fig. 4 (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. 4 (d). It is gotten that the size of the tangential plane of the object in the real 3D space.

2.3 3D model-based matching

In Fig. 5, 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. 5 shows a generated 3D solid model with

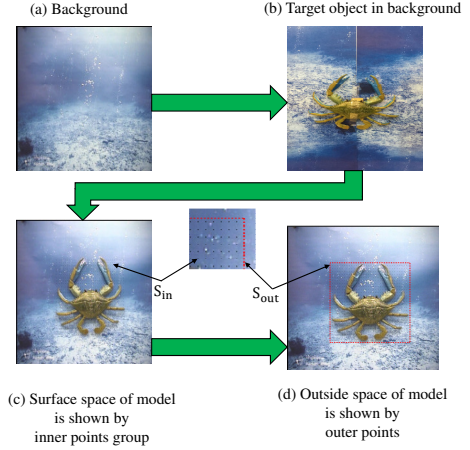


Fig.4 Model generation process described in chapter 2.2. Fig. 4 (b) is captured as a 640×480 pixels picture at a distance of 400[mm] from a camera

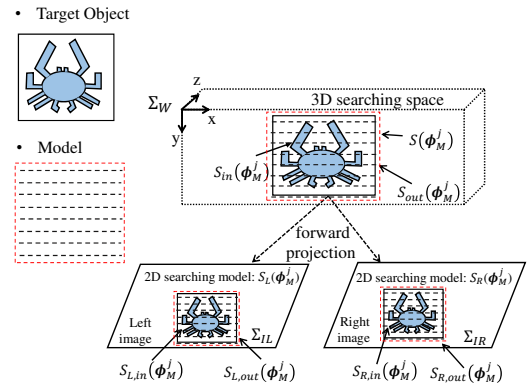


Fig.5 Through projection transformation, a 3D solid model in the 3D searching space is projected to 2D left and right images. Searching models represented as $S_L(\phi_M^j)$ and $S_R(\phi_M^j)$

its pose $S_{in}(\phi_M^j)$ (inner dotted points) and the outside space enveloping $S_{in}(\phi_M^j)$ denoted as outer dotted line ($S_{out}(\phi_M^j)$). The sub figure on the left/right bottom of Fig. 5 show the left/right 2D searching models $S_L(\phi_M^j)$ and $S_R(\phi_M^j)$ respectively. Both $S_L(\phi_M^j)$ and $S_R(\phi_M^j)$ consist of $S_{L,in}(\phi_M^j)$ and $S_{L,out}(\phi_M^j)$ and $S_{R,in}(\phi_M^j)$ and $S_{R,out}(\phi_M^j)$. 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 [4] 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 equations

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

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

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 $IL \mathbf{r}_i^j \in S_{L,in}(\phi_M^j)$ and $IL \mathbf{r}_i^j \in S_{L,out}(\phi_M^j)$ respectively. Eqs. (5) and (6) is used for calculating $p_{L,in}(IL \mathbf{r}_i^j)$ and $p_{L,out}(IL \mathbf{r}_i^j)$.

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

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

where

- $H_{IL}(IL \mathbf{r}_i^j)$: the hue value of the left camera image at the point $IL \mathbf{r}_i^j$ (i-th point in $S_{L,in}$),
- $H_{ML}(IL \mathbf{r}_i^j)$: the hue value of the point $IL \mathbf{r}_i^j$ (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 $p_{R,in}(IR \mathbf{r}_i^j)$ and $p_{R,out}(IR \mathbf{r}_i^j)$ are represented for the right camera image.

2.5 GA (Genetic Algorithm)

GA evaluation process is applied to find the maximum value as an optimal solution. The 20 individuals of GA are used in this experiment. Each individual chromosome consists of six variables. Each variable is coded by 12bits that can easily implement 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.

$$\underbrace{01 \dots 01}_{12bits} \underbrace{00 \dots 01}_{12bits} \underbrace{11 \dots 01}_{12bits} \underbrace{01 \dots 01}_{12bits} \underbrace{01 \dots 11}_{12bits} \underbrace{01 \dots 10}_{12bits}$$

Readers can refer to [3], which has a more detailed explanation.

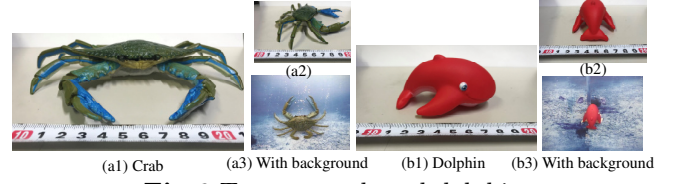


Fig.6 Targets: crab and dolphin

3 Experimental environment

In order to verify that different objects can be tracked with photo-based recognition method, two visual servoing experiments with crab and dolphin toy are performed. As shown in Fig. 2, during the experiments the desired position and orientation is ${}^H \psi_M = (0, -80[mm], 500[mm], 0, 0, 0)$. The target trajectories of two experiments are the same. Both of them are sine curves with amplitude of 100 [mm], period of 20 [s] in the x and y axis directions and amplitude of 100 [mm], period of 60 [s] in the z axis direction.

4 Experimental results and discussion

The world coordinate system (Σ_W), the hand coordinate system (Σ_H) and the crab coordinate system (Σ_M) that are used in the experiments respectively. In Fig.7 and Fig.8, the actual positions of end-effector ${}^W x_H, {}^W y_H, {}^W z_H$ are the tracking results of the end-effector. The desired positions of end-effector ${}^W x_{Hd}, {}^W y_{Hd}, {}^W z_{Hd}$ are ideal positions during the experiments.

Through the tracking results as shown in Fig.7 and Fig.8, it can be seen that even though the tracking curves are somewhat delayed in phase, the visual servoing system with photo-model-based recognition method can track the object in time. It can also be seen that 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 future, we would like to discuss this problem.

References

- [1] 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, pp. 742-747, 2016.
- [2] 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, September 20-23 2016.
- [3] Phyu, K.W., Funakubo, R., Fumiya, I., Shinichiro, Y. and Minami, M., 2017, August. Verification of recognition performance of cloth handling robot with photo-model-based matching. In Mechatronics and Automation (ICMA), 2017 IEEE International Conference on (pp. 1750-1756). IEEE.
- [4] 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, pp. 178-207, 2003.

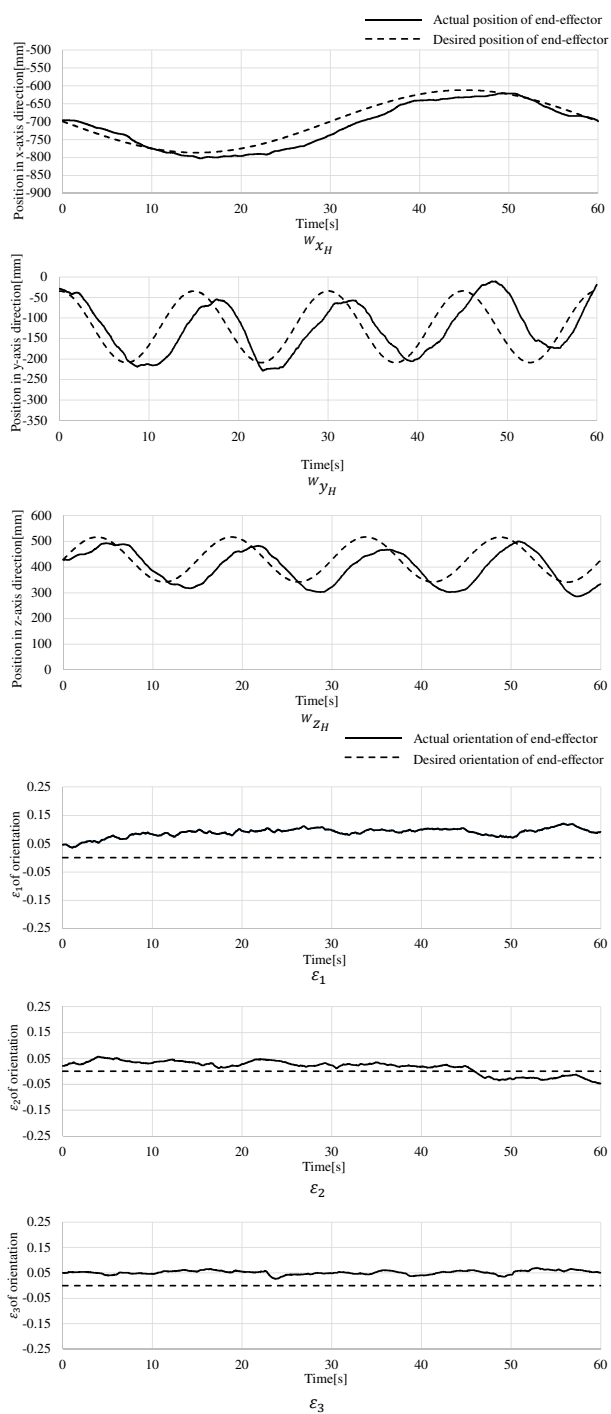


Fig.7 Position and orientation tracking results. Target: crab

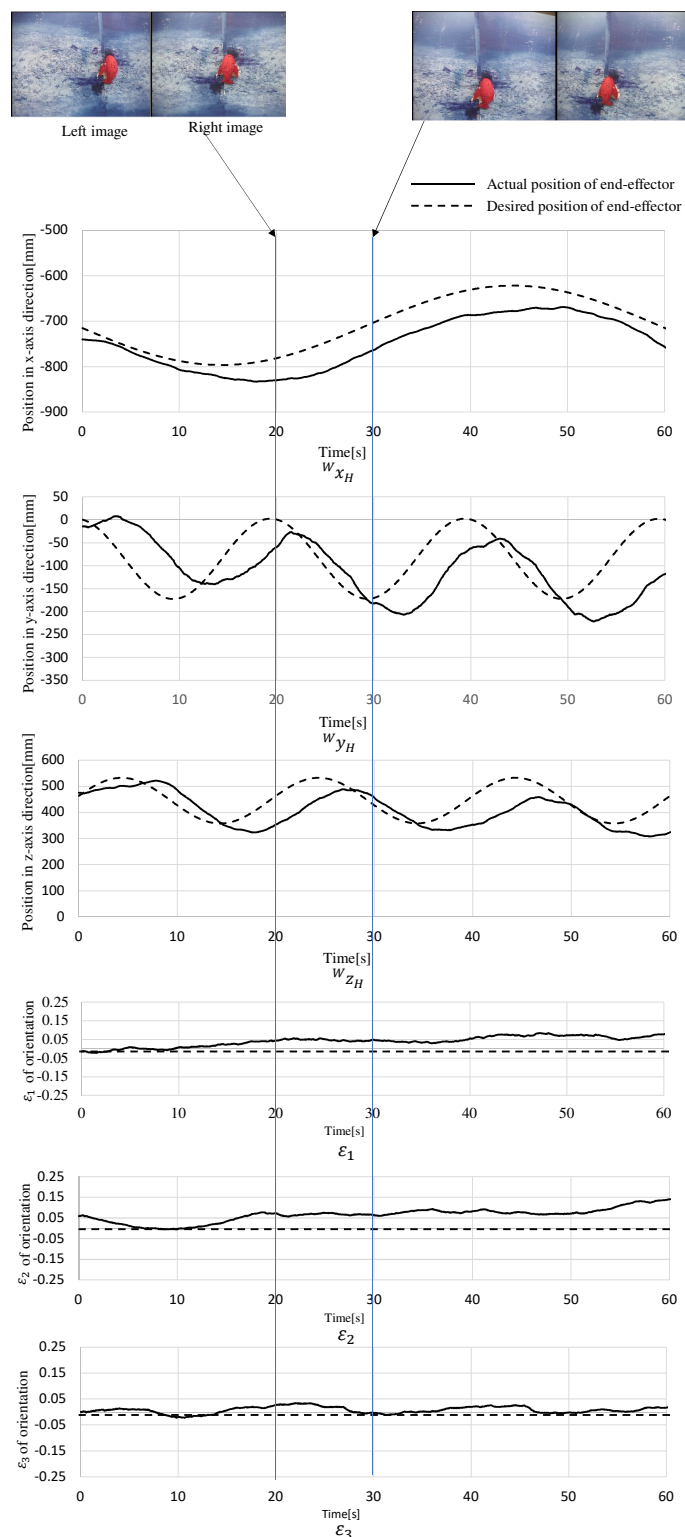


Fig.8 Position and orientation tracking results. Target: dolphin