

A New Concept of Pose Estimation of Arbitrary 3D Object without Prerequisite Knowledge: Projection-based 3D Perception

Yejun Kou¹, Yuichiro Toda¹, Mamoru Minami¹

¹Graduate School of Natural Science and Technology, Okayama University, Japan

(Tel: 81-86-251-8233, Fax: 81-86-251-8233)

¹ptlg9dvi@s.okayama-u.ac.jp

Abstract: Generally speaking, the tasks that utilize robot vision such as visual servoing or pose estimation towards a solid object require a prerequisite condition. Because the system can only check out objects with predefined features or models, the requirement of a priori knowledge of the target object seems to have been a hindrance for tasks applied for the arbitrary target. For this reason, a new approach named: Pb3DP (Projection-based 3D perception) is proposed in this paper. The concept of Pb3DP employs stereo-vision effectively, that is (1) use the 2D image of the target object in the left camera as the 2D model, then (2) inversely project the model into 3D space with an assumed pose, and then (3) the projected 2D model in space onto right camera's image again. The projection (2) and (3) are calculated by computer. Here, (4) if the 2D model reprojected by the computer coincides with the real 3D target object projected naturally by camera's function, then the assumed pose could be thought to present the real target's pose. In this paper, we apply Pb3DP to estimate a solid target object's pose in real-time, by showing the experimental result, the effectiveness of the proposed method will be discussed.

Keywords: 3D pose estimation, Pb3DP, Projection-based

1 INTRODUCTION

It is a common sense that most living being in nature possess two eyes inherently, which may come out of the need for survival of animals because it is essential for animals to judge the distance reaching the prey. Inspired by this fact, some approaches have been proposed with binocular vision for robots to acquire a 3D perception. On the other hand, derived from its simplicity, in the field of visual servoing, some classic methods such as image-based method [1], position-based method [2] and 2- $\frac{1}{2}$ methods [3], have been devised using monocular vision. However, the difficulty in position estimation along to the camera depth direction limits the adaptive operation abilities of vision-based robots. Comparing to monocular vision, a binocular vision is inherently superior to monocular vision in depth recognition because it can provide more vision information such as parallax and another complete image for reference.

However, the configuration of binocular raises a fundamental problem in stereo-vision, called "Correspondence Point Identification Problem,"[4], which means the difficulty to make a point on a camera image correctly matches with another point in the other camera, and the two points are different expressions of the same feature on the target object in different camera image. In stereo-vision it is difficult to judge whether both two points in dual cameras' images represent the same point on the 3D target object or not. This difficulty is derived from the ill-posed problem setting[5]. For the conventional methods employing stereo-vision, the 3D perception is usually be measured through feature-based recog-

niton and 2D-to-3D reconstruction calculation. The 2D-to-3D reconstruction usually exploits triangulation and epipolar geometries. For artifacts, feature points such as vertices and edges are easier to be extracted through image processing techniques. However, the natural objects such as rocks and plants are usually in irregular shapes, the corresponding problem may become more complex because it is hard to specify the corresponding point correctly. Fig. 1 shows such a situation that a potato is selected as the target object. An incorrect mapping of points in the left and right cameras through 2D-to-3D reconstruction, results in wrong point reconstruction in 3D space as shown in B'' in Fig. 1(a). This problem occurs because the expanded 3D information in 3D space is to be reconstructed from the degenerated image information of the 2D images, this intractable feature is derived from ill-posed problem setting.

Contrarily to the 2D-3D reconstruction, the direct projection from 3D-to-2D generates unique points in the 2D image without any errors in corresponding points matching. Therefore, the pose estimation method based on 3D-to-2D model projection [6] is applied in this study. Figure 1 (b) means that the 3D-to-2D projections can resolve the problem of wrong mapping in 2D-to-3D. The point cloud defined on surface on 3D target object can be projected correctly onto 2D camera images by 3D-to-2D projection that is not annoyed by ill-posed problem. Providing that the positions of the point cloud are assumed to be defined on 3D potato surface, then projected on to 2D image planes of left and right cameras as a whole.

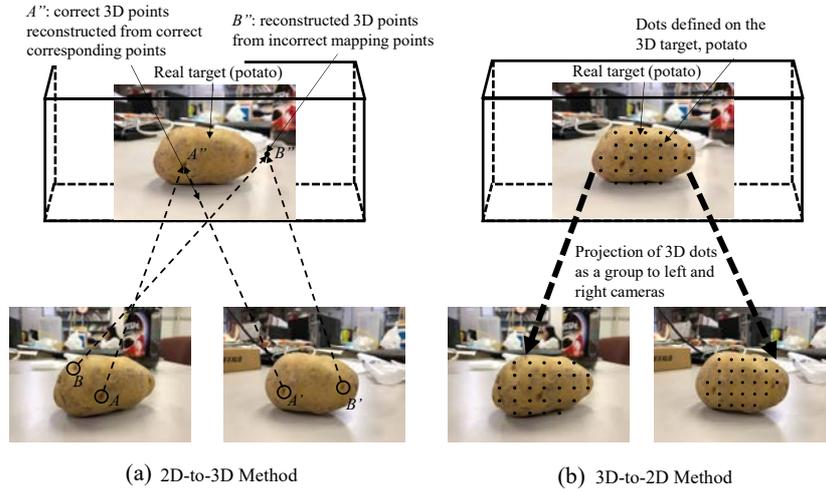


Fig. 1. (a) Miss-mapping in 2D-to-3D reconstruction, and (b) Paring of points in 3D-to-2D projection

On the other hand, there are also some new method for measuring distance. For example, the ToF (Time of Fly) principle used in Microsoft Kinect v2 and Intel-SoftKinect DepthSense, and the triangular ranging principle using structured light, used in Microsoft v1 and Intel Realsense. As the character of infra-red camera depth distance measurement, the ToF method has a featured ability about measurement suitable for situations when long-distance estimations are required. However, it also caused some problem such as the performance of estimation under sunlight and estimation towards glasses, specular surface could be degraded [7] and [8]. But in the proposed Pb3DP method, the problem talked above can be avoided because the Pb3DP only exploited the natural vision. Therefore, recognition can be conducted without considering the environment situation and the appearance of the object.

This paper is going to explain how Pb3DP works and examine the feasibility of estimation towards the arbitrary target object with Pb3DP. The experimental result of real-time recognition will be introduced. The remainder of this paper is organized as following: In section 2, the detailed explication of Pb3DP is extended. Section 3 explained the combination of Pb3DP and RM-GA (Real-time Multi-step Genetic Algorithm). The experiment results are shown in section 4. The final section concludes this paper.

2 PROJECTION-BASED 3D PERCEPTION

The Pb3DP is a combination of projection-based method and RM-GA. To the best knowledge of the authors, there are no studies achieved the real-time 3D pose tracking towards an unknown target, meaning that any information about the target object is unknown. The Pb3DP utilized the stereo-vision with RM-GA to estimation the pose of arbitrary target object without prerequisite model. The methodology of Pb3DP is explained in following content.

2.1 The conception of Pb3DP

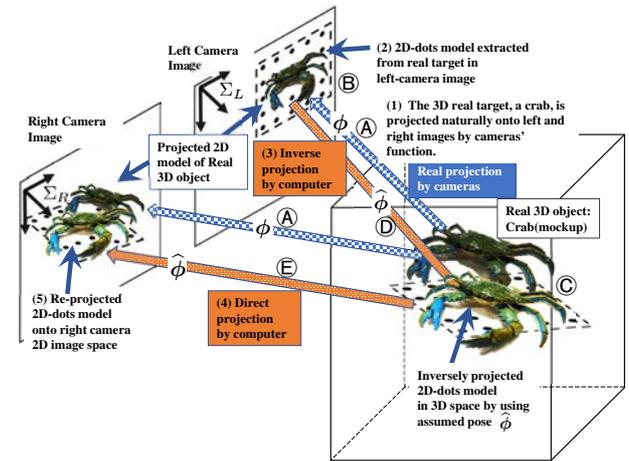


Fig. 2. The schematic of Pb3DP, the detail is explained in section 2.1. But in this figure, since this figure is depicted on the assumption of $\hat{\phi} \neq \phi$, the two images, naturally projected crab by right camera and the computer-projected crab with light color, do not coincide with each other in right camera image. If $\hat{\phi} = \phi$, they would coincide.

In this section, a methodology of the Projection-based 3D Perception (Pb3DP) is explained. The primary purpose of Pb3DP is to estimate a 3D arbitrary target's pose using stereo images. The concept is shown in Fig. 2. In the figure, the procedure of how to estimate the pose of the 3D target by Pb3DP is explained schematically in the order from (1) to (5). (1): A real target, the mock-up of a crab is depicted with dark color, is projected onto left and right cameras images — that is natural light projections into left and right cameras' images — are indicated by (A). The natural projection is thought to be completed based on true pose, ϕ , but it is unknown. (2): Target object is selected in the left camera scene

as a model depicted by (B). (3): The selected 2D target model is inversely projected into the 3D space with an assumed pose $\hat{\phi}$, which is 2D and flat, shown by (C) and the inverse projection is indicated by (D). In the figure the inversely projected 2D model in the 3D space is depicted with light color comparing with actual target in 3D space. (4): The 2D flat model in 3D space is re-projected back to the right camera scene with the same assumed pose $\hat{\phi}$, (E). (5): If the re-projected model with an assumed pose coincides with the real target in the right camera scene, then the assumed pose represents the real target's pose in 3D space, that is, $\hat{\phi} = \phi$.

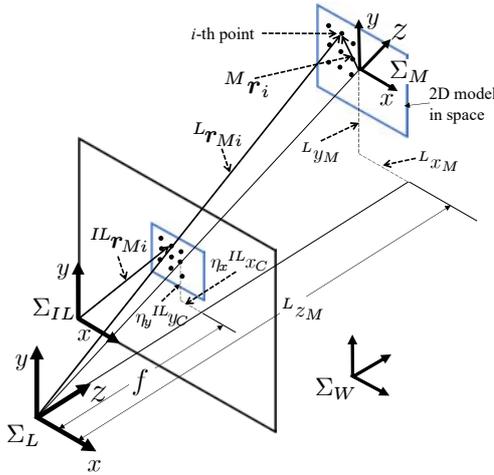


Fig. 3. Projection schematic diagram: The i -th point ${}^M \mathbf{r}_i = [{}^M x_i, {}^M y_i, {}^M z_i]^T$ on a model shown by Σ_M in space is converted into the point represented by Σ_L as ${}^L \mathbf{r}_i = [{}^L x_i, {}^L y_i, {}^L z_i]^T$, and the point ${}^L \mathbf{r}_i$ is projected to ${}^{IL} \mathbf{r}_i = [{}^{IL} x_i, {}^{IL} y_i]^T$ on the left camera image defined by Σ_{IL} .

2.2 Kinematical relations of camera projections

The position vectors in world coordinates Σ_W of an arbitrary i -th point on a 2D model placed in space on which the model coordinate Σ_M is set, are defined as following:

- ${}^W \mathbf{r}_{Mi}$: 3D position vector in Σ_W of an i -th point on a 2D model defined by Σ_M .
- ${}^M \mathbf{r}_i$: 2D position vector on x-y plane in Σ_M of an i -th point on a 2D model whose x-y plane coincides with the 2D model plane, where ${}^M \mathbf{r}_i$ is a constant vector since Σ_M is attached to the model.
- ${}^L \mathbf{r}_{Mi}$: 3D position vector in Σ_L , projection coordinates defined in Fig.3, of an i -th point on a 2D model in space defined by Σ_M .
- ${}^{IL} \mathbf{r}_{Mi}$: 2D position in left image coordinate system Σ_{IL} , left-camera image coordinates defined in Fig.3, of an i -th point of a model.

Given the homogeneous matrix connecting Σ_M that is fixed at a 2D model shown in Fig.3 and Σ_L as ${}^L \mathbf{T}_M$, the relation between ${}^L \mathbf{r}_{Mi} = [{}^L x_{Mi}, {}^L y_{Mi}, {}^L z_{Mi}, 1]^T$ and ${}^M \mathbf{r}_i = [{}^M x_i, {}^M y_i, {}^M z_i, 1]^T$ is represented by

$${}^L \mathbf{r}_{Mi} = {}^L \mathbf{T}_M {}^M \mathbf{r}_i. \quad (1)$$

The i -th point ${}^L \mathbf{r}_{Mi}$ on a model defined by Σ_L in space is projected to ${}^{IL} \mathbf{r}_{Mi} = [{}^{IL} x_{Mi}, {}^{IL} y_{Mi}]^T$ on the left camera image defined by Σ_{IL} as follows, using $\phi = [{}^L \mathbf{r}_M^T, {}^L \theta_M^T]^T$,

$$\begin{aligned} {}^{IL} \mathbf{r}_{Mi} &= \mathbf{P}({}^L z_{Mi}) {}^L \mathbf{r}_{Mi} \\ &= \mathbf{P}(\phi) {}^L \mathbf{T}_M (\phi) {}^M \mathbf{r}_i. \end{aligned} \quad (2)$$

The projective transformation matrix $\mathbf{P}({}^L z_{Mi})$ is given as:

$$\mathbf{P}({}^L z_{Mi}) = \frac{1}{{}^L z_{Mi}} \begin{bmatrix} f/\eta_x & 0 & 0 & 0 \\ 0 & f/\eta_y & 0 & 0 \end{bmatrix}, \quad (3)$$

where

- ${}^L z_{Mi}$: z-axis position of the i -th point in Σ_L on the model Σ_M ,
- f : focal length,
- η_x, η_y : coefficients [mm/pixel] in x-axis and y-axis of image frame.

The projection of right camera can be discussed in the same manner.

2.3 Inverse projection from left camera image to 3D space and re-projection to right camera image

For preparation of inverse projection of ${}^{IL} \mathbf{r}_{Mi}$ to 3D space, the pseudo-inverse projection matrix $\mathbf{P}^+({}^L z_{Mi})$ of $\mathbf{P}({}^L z_{Mi})$ defined by Eq. (3) is needed,

$$\mathbf{P}^+({}^L z_{Mi}) = {}^L z_{Mi} \begin{bmatrix} \eta_x/f & 0 & 0 & 0 \\ 0 & \eta_y/f & 0 & 0 \end{bmatrix}^T. \quad (4)$$

The Eq.(2) can be modified into

$${}^L \mathbf{T}_M (\phi) {}^M \mathbf{r}_i = \mathbf{P}^+(\phi) {}^{IL} \mathbf{r}_{Mi} + (\mathbf{I}_4 - \mathbf{P}^+ \mathbf{P}) \mathbf{l}. \quad (5)$$

The inversely projected flat model that is determined dependently by $\hat{\psi}$ in Σ_W is derived as

$$\begin{aligned} {}^W \mathbf{r}_{Mi}(\hat{\psi}) &= {}^W \mathbf{T}_M(\hat{\psi}) {}^L \mathbf{T}_M^{-1}(\hat{\psi}) \left[\mathbf{P}^+(\hat{\psi}) {}^{IL} \mathbf{r}_{Mi} + (\mathbf{I}_4 \right. \\ &\quad \left. - \mathbf{P}^+(\hat{\psi}) \mathbf{P}(\hat{\psi})) \mathbf{l} \right]. \end{aligned} \quad (6)$$

Here, ψ is defined as $\psi = [{}^L z_M, {}^L \theta_x, {}^L \theta_y]$. The three components ${}^L z_M, {}^L \theta_x, {}^L \theta_y$ of ψ are independent valuables for inversely projecting the flat model in left camera 2D image into space. Providing a set of valuables in the variety of

ψ be chosen and fixed, shall we describe the fixed valuable as $\hat{\psi}$.

Then the image projected to right camera plane of flat target model, ${}^{IR}\mathbf{r}_{Mi}$, is calculated by using assumed $\hat{\psi}$ as

$${}^{IR}\mathbf{r}_{Mi}(\hat{\psi}) = \mathbf{P}(\hat{\psi})^R \mathbf{T}_W(\hat{\psi})^W \mathbf{r}_{Mi}(\hat{\psi}). \quad (7)$$

2.4 Problem conversion from pose estimation to optimization

The purpose of this subsection is to show how to convert the estimation problem of a 3D target's true pose ψ to the optimization problem. Please assume that $\hat{\psi}$ means estimated value of ψ . If a scalar function $F(\hat{\psi})$ should satisfy that the distribution of $F(\hat{\psi})$ has a single maximum peak F_{max} at true pose ψ , and that also satisfy $F(\hat{\psi}) = F_{max}$, then $\hat{\psi} = \psi$. This could be rewritten as,

$F(\hat{\psi}) = F_{max}$ if and only if $\hat{\psi} = \psi \in L$ (Single peak assumption, where L means parameter space of $\hat{\psi}$), then the problem to estimate the true pose ψ can be converted to an another problem as,

Find $\hat{\psi}$ to maximize $F(\hat{\psi})$ subject to $\hat{\psi} \in L$.

It means that the estimation of true pose can be completed by optimizing $F(\hat{\psi})$ with parameters $\hat{\psi}$. Then how to constitute a scalar function $F(\hat{\psi})$ satisfying the single peak assumption above appears to be a next problem.

2.5 The Establishment of a Model

In the conventional visual servoing method, the model that created beforehand limits the visual servoing system because they can only recognize the assigned target objects. In order to realize the recognition of the arbitrary objects, the models in Pb3DP are designed to be created at any time. In this section, the establishment of the model will be described.

The models used in this method consist of 2-D point cloud, each of the sampling points contains the color information of the image at the location of the point. The color information is used to evaluate the recognition result. The model is consisted by two portions, the inner area that mainly presented the target object and the outer area, which is presenting the noise around the target object to make the result accurate. The model can be made in real-time by using the target's image shown in camera, this feature of model establishment ensured that no prerequisite information of target is needed, and an arbitrary target object can be recognized. The components of the model is same as the previous research of the authors, the detail of inner area and outer area can be referred in [9].

3 REAL-TIME MULTI-STEP GA

3.1 Evaluation Method

In proposed Pb3DP method, the models with assumed pose are utilized to infer the true pose of target object. A co-

incidence degree, between the projected model and the target in right camera captured by dual-eye cameras can be thought as a method to evaluate the recognition result (Minami et al., 2003). In this evaluation method, the fitness is used as a numerical value to evaluate the coincidence degree. Therefore, the problem of finding the true pose of target object can be converted into finding the maximum value of fitness.

A model is consisted of two portions, the inner area and outer area, which are composed of sampling points. The number of sampling points in inner area and outer area are N_{in} and N_{out} . The coordinate of each points in model that direct projected into right camera image is ${}^{IR}\mathbf{r}_i^j$, and evaluation value of each point in inner portion of the model (${}^{IR}\mathbf{r}_i \in S_{R,in}(\phi)$) is $\mathbf{P}_{R,in}({}^{IR}\mathbf{r}_i^j)$ calculated by Eq. (8). The outer portion (${}^{IR}\mathbf{r}_i \in S_{R,out}(\phi)$) is $\mathbf{P}_{R,out}({}^{IR}\mathbf{r}_i^j)$ calculated by Eq. (9).

$$\mathbf{P}_{R,in}({}^{IR}\mathbf{r}_i^j) = \begin{cases} 2, & \text{if } (|H_M({}^{IR}\mathbf{r}_i^j) - H_I({}^{IR}\mathbf{r}_i)| \leq 20) \\ -1, & \text{if } (|H_M({}^{IR}\mathbf{r}_i^j) - H_I({}^{IR}\mathbf{r}_i)| > 20) \end{cases} \quad (8)$$

$$\mathbf{P}_{R,out}({}^{IR}\mathbf{r}_i^j) = \begin{cases} 0.1, & \text{if } (|H_M({}^{IR}\mathbf{r}_i^j) - H_I({}^{IR}\mathbf{r}_i)| \leq 20) \\ -2, & \text{if } (|H_M({}^{IR}\mathbf{r}_i^j) - H_I({}^{IR}\mathbf{r}_i)| > 20) \end{cases} \quad (9)$$

where

- $H_M({}^{IR}\mathbf{r}_i^j)$: the hue value of the model in right camera image at the point ${}^{IR}\mathbf{r}_i^j$ (j -th point in i -th model, lying in $S_{R,in}$).
- $H_I({}^{IR}\mathbf{r}_i)$: the hue value of right camera image at the point ${}^{IR}\mathbf{r}_i$.

The fitness function can be given by the following equation:

$$F_R(\phi) = \left\{ \frac{\sum_{{}^{IR}\mathbf{r}_i \in S_{R,in}(\phi)} p({}^{IR}\mathbf{r}_i) + \sum_{{}^{IR}\mathbf{r}_i \in S_{R,out}(\phi)} p({}^{IR}\mathbf{r}_i)}{(2 \times N_{R,in} + 0.1 \times N_{R,out})} \right\} / \quad (10)$$

If the projected 2D model entirely coincides with the captured target object in the left and right images, the fitness value calculated by Eq. (10) is designed to have a maximum value. Therefore, the fitness value distribution for all models will shaped with a peak that represented the real pose of the target object. The concept of the fitness function in this method can be said as an extension of the work in (Song et al., 2007) in which different models, including a rectangular shape surface-strips model, were evaluated using images from a single camera.

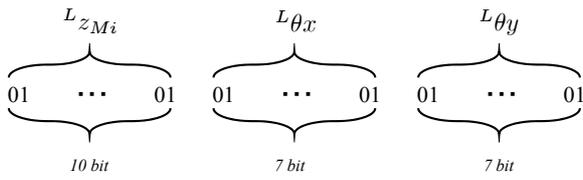


Fig. 4. Gene information

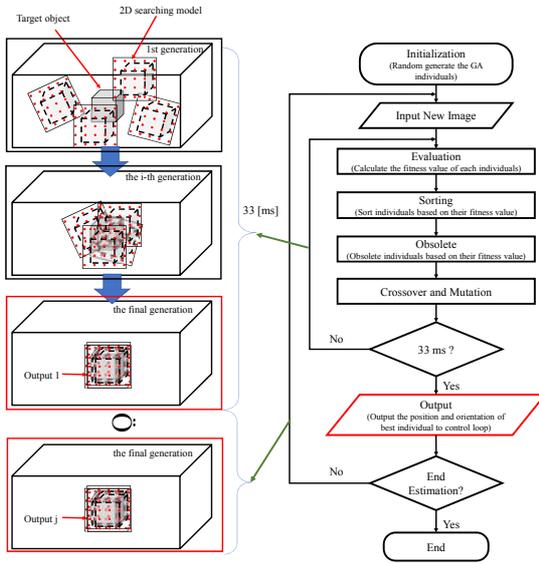


Fig. 5. Flowchart of the RM-GA

3.2 Real-time Multi-step GA (RM-GA)

In Pb3DP method, searching all possible pose of target object through calculating the fitness value is time-consuming for realtime pose estimation. Therefore, the problem of recognizing the target object’s pose can be transformed into a optimization to find the maximum value of fitness. In Pb3DP, we employed Real-time Multi-step GA (RM-GA) to satisfy the realtime recognition in 30 FPS. The reason why we choose RM-GA has been discussed in (Myint et al., 2018).

In proposed RM-GA, each chromosome includes 24 bits for searching three parameters: ten for the position and fourteen for orientation, as shown in Fig. 4. Figure 5 shows the flowchart of the Real-time Multi-step GA, and the recognition process in 3D space is presented in the left. Here, a 2D searching model in 3D space represents a GA individual. The GA operation is conducted in the sequence as evaluation, sorting, obsolete, crossover, and mutation. Several 2D searching models that represent different relative poses converge to the target object through GA evolution process within 33 [ms]. The 2D searching model (Output j in Fig. 5) that represents the true pose with the highest fitness value that calculated by Eq.(10) is searched for every 33 ms. Then, these fit models are directly projected to the next step as the initial models for the next new images in real time.

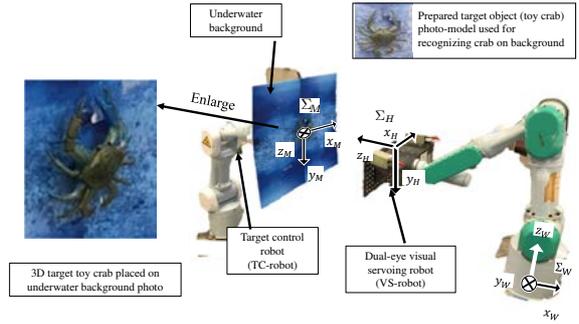


Fig. 6. Experimental layout. The motion of the target animal, crab, is given by TC-robot, and the VS-robot moves to keep desired relative pose of the VS-robot against the crab attached on a panel with sea bottom backdrop whose motion is given by TC-robot. World coordinate system Σ_W , hand coordinate system Σ_H , and target coordinate system Σ_M are depicted in the figure.

4 EXPERIMENTS

The experiment in this paper presented the visual servoing result by Pb3DP method. During the experiment, the target object changed its position and orientation with time. The robot’s hand recognize the target object’s pose in real time and kept a assigned pose relationship to track the trajectory of the target object. The following parts of this section is: experimental environment and results of the experiment. The last part is the discussion of the experimental results.

4.1 Experimental Environment

The experimental environment can be referred to Fig. ???. Two manipulator are employed in this experiment, which are PA-10 robot arm manufactured by Mitsubishi Heavy Industries. Two cameras are mounted on the end-effector and connected to a host computer (CPU: Intel i7-3770, 3.40[GHz]). The layout of these two cameras formed a binocular vision configuration. The resolution of dynamic images is 640480 [pixel]. The frame frequency of stereo cameras is set as 30[fps].

During the experiment, the target object was move in a defined trajectory, which is shown in the Fig. 7(a)~(f) as black lines. The movement range of position change is 100[mm] in each direction and the degree of orientation change is set as 20[deg], which can be transformed into quaternion as “0.173.” The time of each step is marked as (1)~(12) in the top of Fig. 7. The target object changed its position and orientation with time. Meanwhile, a pose relationship between robot hand and the target object is kept as Eqn. (11).

$${}^M T_{Hd} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -500[\text{mm}] \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (11)$$

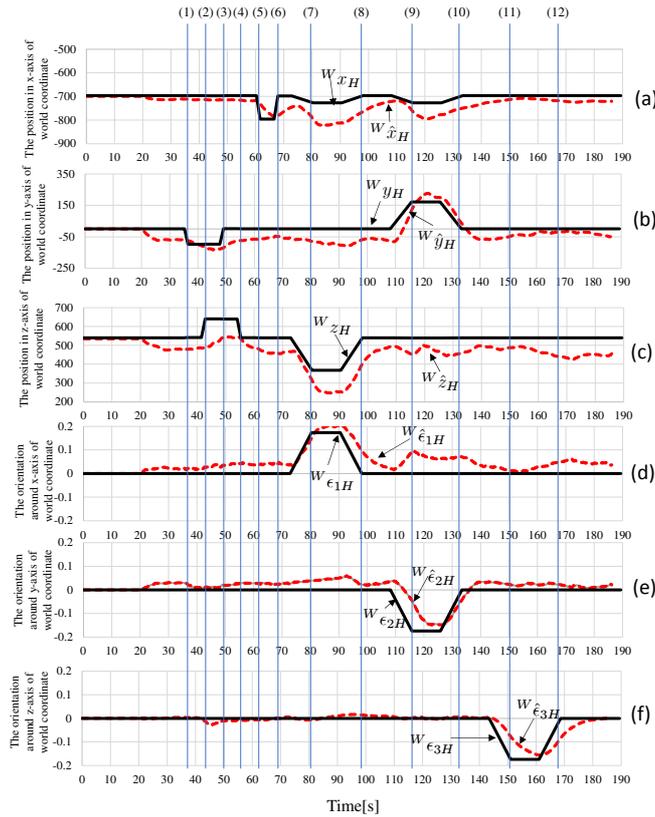


Fig. 7. The experiment result. The trajectory of target object’s movement is shown as black lines in (a)~(f), and the movement of robot’s hand is shown as red lines.

4.2 Experimental Result

As the result, the Pb3DP method can recognize the pose of target object in real time, and lead the robot’s hand to track the movement of target object in time. The position visual servoing results are shown as Fig. 7(a)~(d), and the orientation visual servoing results are shown as Fig. 7(e)~(f). To the position tracking result, the movement of robot’s hand can follow the trajectory of target object, but there are some error between the desired hand position ${}^W k_H(k = x, y, z)$ and the real hand position ${}^W \hat{k}_H(k = x, y, z)$. The reason is considered as the hand received the influence of the real-time orientation recognition result, to maintain the the relationship defined in Eqn. (11), the error between real and desired hand position is occurred.

To the orientation tracking results as shown in Fig. 7, the robot hand can track the trajectory of target object with some time delay. Meanwhile, there were some error during the experiment because the recognition of orientation recognition result is always changing due to the pose recognition is completed by RM-GA. However, according to the experimental results, we can see the PA-10 can recognize and track the target object in real time by employing the Pb3DP method.

5 CONCLUSION

In this paper, the Pb3DP stereo-vision system is introduced. It is verified that the proposed method can make the robot’s hand-eye track the 3D target object without predefined model. It means that the 3D target’s pose can be estimated in real-time to arbitrary target object by its 2D photo-model. Meanwhile, the RM-GA ensured the feasibility and robustness of recognition towards target objects. The RM-GA can converge to the true pose of target object after some generations of evolution. And it can converge to the target object again in a short time after the image of the target object changed in the left and right camera.

REFERENCES

- [1] Chaumette, F. (1998), Potential problems of stability and convergence in image-based and position-based visual servoing. In *The confluence of vision and control*, pp. 66-78
- [2] Allen, P. K., Timcenko, A., Yoshimi, B., & Michelman, P. (1993), Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Transactions on Robotics and Automation*, 9(2), 152-165
- [3] Chaumette, F., & Malis, E. (2000), 2 1/2 D visual servoing: a possible solution to improve image-based and position-based visual servoings. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings*, 1: pp. 630-635
- [4] Phyu, K.W., Funakubo, R., Hagiwara, R., Tian, H. and Minami, M., Verification of illumination tolerance for photo-model-based cloth recognition. *Artificial Life and Robotics*, Vol.23, No.1, (2018), pp.118-130.
- [5] Poggio, T. and Edelman, S., A network that learns to recognize three-dimensional objects. *Nature*, Vol.343, No.6255,(1989), p.263.
- [6] 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 (2018), pp.159-186.
- [7] Bajard, A., Aubretton, O., Bokhabrine, Y., Verney, B., Truchetet, F., Eren, G. and Eiril, A., 2012. Three-dimensional scanning of specular and diffuse metallic surfaces using an infrared technique. *Optical Engineering*, 51(6), p.063603.
- [8] Gupta, M., Yin, Q. and Nayar, S.K., 2013. Structured light in sunlight. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 545-552).
- [9] Hongzhi TIAN, Yejun KOU, Xiang LI, Mamoru MINAMI, "Real-time pose tracking of 3D targets by photo-model-based stereo-vision", *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, Vol.14, No.4, 2020.