A Real-time 3D Pose Estimation Method towards Arbitrary Target with Stereo Vision

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

[1]Graduate School of Natural Science and Technology, Okayama University, Japan
(Tel: 81-86-251-8233, Fax: 81-86-251-8233)
[1]ptlg9dvi@s.okayama-u.ac.jp

Abstract: Visual servoing methods have been utilized as a control method to make robots conduct various tasks. In conventional visual servoing methods, some prerequisite conditions, such as feature points or predefined models, are needed to complete the estimation. This demand limited the robots to track the 3D pose of the arbitrary targets because the visual servoing methods can only recognize the assigned targets. On the other hand, humans can comprehend 3D perception of any surrounding situations instantly. This ability relies on the parallax error in the stereo vision of humankind. Aroused by this feature, a new 3D pose estimation method named "projection-based 3D perception (Pb3DP)" is proposed in this paper. In Pb3DP, a stereo vision is completed by dual-eyes camera configuration. Contrary to other methods, the Pb3DP needs no priori-knowledge of target objects. Therefore, the premeditated definition of models of target objects can be omitted, which increased the versatility and the application area of robots. In this paper, the methodology of Pb3DP is introduced in detail. To examine the adaptability of Pb3DP, the real-time 3D pose tracking results of different unknown target objects are shown.

Keywords: Visual servoing, 3D pose estimation, Real-time, Unknown/ arbitrary target object, Pb3DP

1 INTRODUCTION

Recently, the automation of robots has been expected as a solution to the lack of labour. In order to realize the automation of robots, the visual servoing [1], also known as vision-based control approach is one of the most popular methods. Visual servoing has been applied to many problems as a conventional solution, such as guiding the robots in public areas; conveying, processing, and assembling works in factories. Therefore, there are many researches have been conducted about visual servoing. In order to measure the pose of the target object, the position-based visual servoing (PBVS) [2] uses a precisely pre-defined 3D geometric model to retrieve the three-dimensional information about the scene. Meanwhile, the image-based visual servoing (IBVS) [3] proposed by Chaumette, F, provided another method to realize visual servoing with simplified model. In IBVS, the feature points (such as apexes of the shapes of a target object) in camera's image are used to designate the desired movement of robot. Furthermore, in order to improve PBVS and IBVS, another hybrid visual servoing has been proposed, which is known as "2 $\frac{1}{2}$D visual servoing." [4]. "2 $\frac{1}{2}$D" method ensures the convergence of the control law in the whole task space. More importantly, contrary to PBVS and IBVS, the "2 $\frac{1}{2}$D" visual servoing provides a controlling solution in the camera's depth direction. The above method is used as a general method in various fields to achieve visual servoing. However, one of the most fundamental problems in the above methods is that the pre-defined model or assigned model are indispensable as a prerequisite to perceive the environments, means the recognizable target by the methods above are limited in a small range. Because the unknown targets can hardly be recognized. To overcome this problem, in this paper, we proposed a new approach named "projection-based 3D perception (Pb3DP)" to achieve the visual servoing towards the arbitrary target objects without priori knowledge, in which 3D perception enabled by stereo vision.

Most of the conventional methods of visual servoing are based on the monocular vision, they utilize single camera to minimize the process time of vision information extraction. However, the difficulty in camera depth direction estimation limits the adaptive servoing operation abilities of robots. Comparing to monocular vision, the binocular vision, which provides a stereo-vision [5], [6], is superior because it can perceive 3D perception of the environment using parallax. This feature in binocular vision ensures high possibility of 3D pose estimation, especially in the position estimation of camera's depth direction. But at the same time, a fundamental problem in stereo-vision approaches, called "Corresponding Points Identification Problem" that is how to make a point in one camera image correctly correspond to a point in another camera image—to confirm whether both two points in dual cameras' images represent a point on the 3D target objects—, is considered hard to be solved and time-consuming [7], [8]. However, the Pb3DP exploits the 2D point-cloud model to conduct visual servoing, the relative position of each sampling points in 2D point-cloud model is destined and the projection of each points is completed at the same time. Therefore, the corresponding points identification problem can be avoided in Pb3DP. This feature of 2D point-cloud model has been discussed in [9].

The main purpose of this paper is to propose a new pose tracking approach to estimate the 3D pose of arbitrary target, in which 3D perception enabled by stereo vision. The schematic is that (1) target object is selected in the scene in one of the stereo cameras, (2) the selected 2D target is inversely projected in 3D space with assumed pose, (3) the target in 3D space is projected again into the other camera scene, (4) if the target projected through assumed pose happens to be matched with the real target in the camera scene, then the assumed pose represents real target's pose in 3D space. In addition, Real-time Multi-step GA (RM-GA) [10] is exploited as the optimization method to process the dynamic image. This combination of Pb3DP and RM-GA with stereo vision is the main contribution of proposed paper.

The remainder of this paper is organized as follows: In section 2, a detailed explication of the projection-based method is presented. Section 3 describes the optimization method. Experiment results are shown in Section 4. In the final, section 5 concludes this paper.

2 PROJECTION-BASED METHOD

2.1 Projection-based 3D Pose Estimation Method Using Stereo Vision

In the proposed projection-based 3D perception method, the main purpose is to use the image of the arbitrary target's image to estimate it's pose. The schematic is shown as Fig.1, (1) target object is selected in the scene in one of the stereo cameras, (2) the selected 2D target is inversely projected in 3D space with assumed pose, (3) the target in 3D space is projected again into the other camera scene, (4) if the target projected through assumed pose happens to be matched with the real target in the camera scene, then the assumed pose represents real target's pose in 3D space. In this section, methodology of the Pb3DP method is explained.

2.1.1 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.

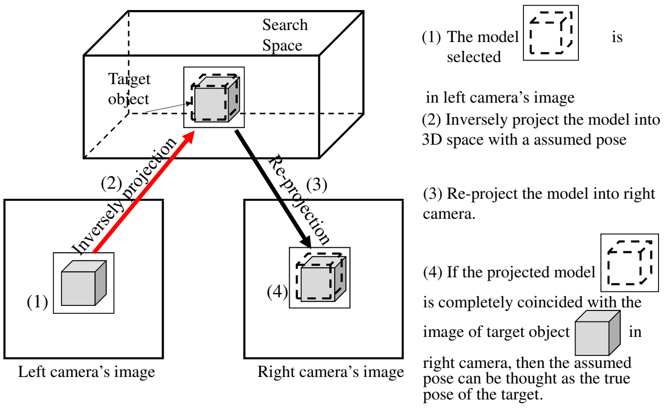
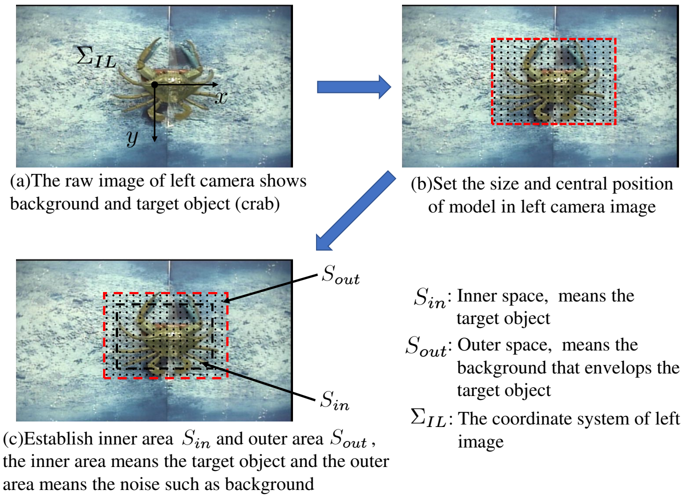

Search
Space
Target
object
(2)
Inversely projection
Re-projection
(3)
(1)
(4)
Left camera's image
Right camera's image
(1) The model is selected
in left camera's image
(2) Inversely project the model into 3D space with a assumed pose
(3) Re-project the model into right camera.
(4) If the projected model is completely coincided with the image of target object in right camera, then the assumed pose can be thought as the true pose of the target.

Fig. 1. The schematic of Pb3DP, (1) target is selected in the left camera image as the model, (2) the selected target's image is inversely projected into 3D space with assumed pose, (3) re-project the image of target into right camera image, (4) if the target projected through assumed pose is matched with the real target in right camera, then the assumed pose can be thought as the real pose of target object in 3D space

Figure 2 shows the procedure of a model's establishment. In this figure, a mock-up of crab is set as the target object. 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. In Fig. 2(a), a raw image from the left camera is read as the basement to generate the model and the coordinate of left camera image is defined as Σ_{IL}. The origin of Σ_{IL} is at the center of the left camera image. The size and position to generate the model are selected manually. Then the sampling points are generated in the model area at regular intervals. The arbitrary position of the point in the created model in the left camera coordinate system is given as ${}^{IL}\boldsymbol{r}^{j}_{Mi}$. As shown in Fig.2(b), the plastic model is fully included in the model area. However, due to the model's shape is set as a rectangle and the targets' shape are usually irregular, it's inevitable that some background is included in the selected area. Therefore, it is necessary to distinguish the background from the model. For this reason, the model consists of the inner area (S_{in}) and the outer area (S_{out}), where the S_{in} means the target object and the S_{out} means the background. As shown in Fig.2(c), the outer area envelops the inner area as a subtraction to exact accurate recognition result. The outer area is generate around the inner area with the same regular intervals.

2.2 The kinematics of stereo-vision

The coordinates of the this system is shown as Fig. 3. The proposed system utilized eye-in-hand configuration and two cameras to complete stereo vision. The coordinate systems of two cameras and target object are consisted of world coordinate system Σ_W, i-th model coordinate system Σ_{M_i}, hand position coordinate system Σ_H, left and right camera coordinate system Σ_{CL} and Σ_{CR}, left and right image coordinate system Σ_{IL} and Σ_{IR}, coordinate system of target object Σ_M, and they are shown in Fig. 3. The position vectors of an arbitrary j-th point in the i-th 3D model coordinate Σ_{Mi} based


Σ_{IL}
x
y
S_{out}
S_{in}
S_{in}: Inner space, means the target object
S_{out}: Outer space, means the background that envelops the target object
Σ_{IL}: The coordinate system of left image

(a)The raw image of left camera shows background and target object (crab)

(b)Set the size and central position of model in left camera image

(c)Establish inner area S_{in} and outer area S_{out}, the inner area means the target object and the outer area means the noise such as background

Fig. 2. Model generation process are described as (a)∼(c): (a) shows the raw image in left camera, (b) represents the model area set by assigned central position and size, (c) represents a inner area S_{in} and outer area S_{out} envelops S_{in}

on each coordinate system are as following:

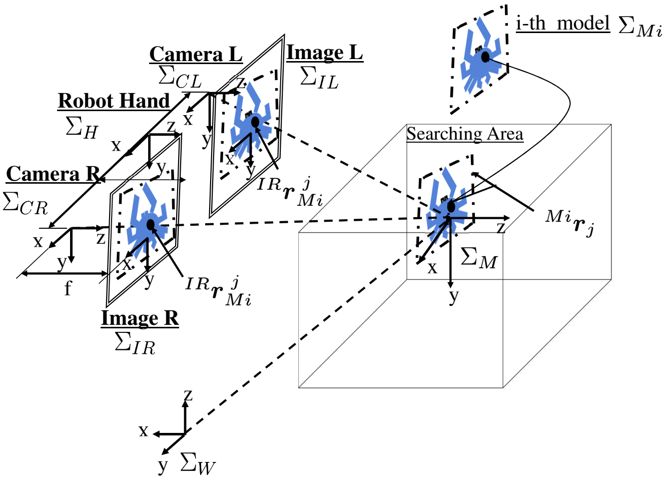


Fig. 3. The coordinate systems of the projection-based method

- ${}^{W}\boldsymbol{r}^j_{Mi}$: position of an arbitrary j-th point on i-th 3D model based on Σ_W.
- ${}^{Mi}\boldsymbol{r}_j$: position of an arbitrary j-th point on i-th 3D model in Σ_{Mi}, where ${}^{Mi}\boldsymbol{r}_j$ is a constant vector.
- ${}^{CL}\boldsymbol{r}^j_{Mi}$ and ${}^{CR}\boldsymbol{r}^j_{Mi}$: position of an arbitrary j-th point on i-th 3D model based on Σ_{CL} and Σ_{CR}.
- ${}^{IL}\boldsymbol{r}^j_{Mi}$: the position of j-th point of i-th model in left image coordinate system Σ_{IL}
- ${}^{IR}\boldsymbol{r}^j_{Mi}$: the position of j-th point of i-th model in left image coordinate system Σ_{IR}

Meanwhile, the projective transformation matrix is given as following

$$\boldsymbol{P}({}^{C}z_j) = \frac{1}{{}^{C}z_j}\begin{bmatrix} f/\eta_x & 0 & {}^{I}x_0 & 0 \\ 0 & f/\eta_y & {}^{I}y_0 & 0 \end{bmatrix}. \qquad (1)$$

Therefore, the arbitrary point of target object naturally projected result in Σ_{IL} and Σ_{IR} can be given as,

$$\begin{aligned} {}^{IL}\boldsymbol{r}_M &= \boldsymbol{P}({}^{CL}z_j){}^{CL}\boldsymbol{r}_M \\ &= \boldsymbol{P}({}^{CL}z_j){}^{CL}\boldsymbol{T}_H{}^{H}\boldsymbol{T}_M(\boldsymbol{\phi}_M, \boldsymbol{q}){}^{M}\boldsymbol{r} \qquad (2) \end{aligned}$$

$$\begin{aligned} {}^{IR}\boldsymbol{r}_M &= \boldsymbol{P}({}^{CR}z_j){}^{CR}\boldsymbol{r}_M \\ &= \boldsymbol{P}({}^{CR}z_j){}^{CR}\boldsymbol{T}_H{}^{H}\boldsymbol{T}_M(\boldsymbol{\phi}_M, \boldsymbol{q}){}^{M}\boldsymbol{r} \qquad (3) \end{aligned}$$

On the other hand, the inverse projection transformation matrix $\boldsymbol{P}^{+}$ can be achieve based on Eq.(1) as

$$\boldsymbol{P}^{+}({}^{C}z_j) = {}^{C}z_j \begin{bmatrix} \frac{\eta_x}{f} & 0 & 0 & 0 \\ 0 & \frac{\eta_y}{f} & 0 & 0 \end{bmatrix}^{T} \qquad (4)$$

where, the ${}^{C}z_j$ is the distance from the coordinate of Σ_{Mi} to Σ_{CL}, which is assumed by RM-GA.

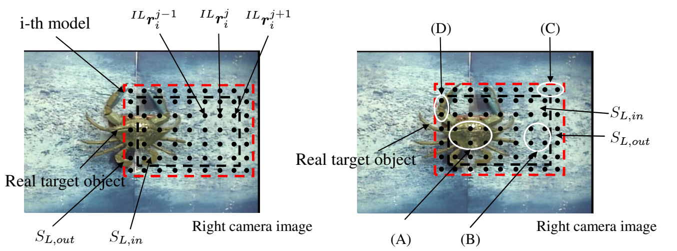
To achieve the pose of 3D searching model in space

$$\begin{aligned} {}^{Mi}\boldsymbol{r}_j &= {}^{Mi}\boldsymbol{T}_{CL}{}^{CL}\boldsymbol{r}^j_{Mi} \qquad (5) \\ &= {}^{Mi}\boldsymbol{T}_{CL}\left[\boldsymbol{P}^{+}({}^{CL}z^j_{Mi}){}^{IL}\boldsymbol{r}^j_{Mi} + (\boldsymbol{I}_4 - \boldsymbol{P}^{+}\boldsymbol{P})\boldsymbol{l})\right] \end{aligned}$$

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 coincidence 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. 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



(a) $^{IL}\boldsymbol{r}_i^j$, that is the j-th point in i-th model, which is projected on left image whose pose is given by RM-GA

(b) Classification of evaluation points (A)~(D) on the projected model. (A) represents the points that satisfy the first case of Eq.(18), $|H_M(^{IL}\boldsymbol{r}_i) - H_I(^{IL}\boldsymbol{r}_i)| \leq 20$, represents that inner model $S_{L,\mathrm{in}}$ overlaps with the real target. (B) does $|H_M(^{IL}\boldsymbol{r}_i) - H_I(^{IL}\boldsymbol{r}_i)| > 20$, represents the inner area $S_{L,out}$ overlaps with the background. (C) does $|H_M(^{IL}\boldsymbol{r}_i) - H_I(^{IL}\boldsymbol{r}_i)| \leq 20$, meaning that the outer area overlaps with the background, and (D) shows $S_{L,out}$ overlaps with the real target object.

Fig. 4. The calculation of the coincidence degree of each point in right camera.

of each points in model that forward projected into right camera image is $IR\boldsymbol{r}_i^j$, and evaluation value of each point in inner portion of the model ($^{IR}\boldsymbol{r}_i \in S_{R,in}(\boldsymbol{\phi})$) is $\boldsymbol{P}_{R,in}(^{IR}\boldsymbol{r}_i^j)$ calculated by Eq.(6). The outer portion ($^{IR}\boldsymbol{r}_i \in S_{R,out}(\boldsymbol{\phi})$) is $\boldsymbol{P}_{R,out}(^{IR}\boldsymbol{r}_i^j)$ calculated by Eq.(7).

$$\boldsymbol{P}_{R,in}(^{IR}\boldsymbol{r}_i^j) = \begin{cases} 2, if(|H_M(^{IR}\boldsymbol{r}_i^j) - H_I(^{IR}\boldsymbol{r}_i)| \leq 20) \\ -1, if(|H_M(^{IR}\boldsymbol{r}_i^j) - H_I(^{IR}\boldsymbol{r}_i)| > 20) \end{cases} \tag{6}$$

$$\boldsymbol{P}_{R,out}(^{IR}\boldsymbol{r}_i^j) = \begin{cases} 0.1, if(|H_M(^{IR}\boldsymbol{r}_i^j) - H_I(^{IR}\boldsymbol{r}_i)| \leq 20) \\ -2, if(|H_M(^{IR}\boldsymbol{r}_i^j) - H_I(^{IR}\boldsymbol{r}_i)| > 20) \end{cases} \tag{7}$$

where

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

The fitness function can be given by the following equation:

$$F_R(\boldsymbol{\phi}) = \left\{ \sum_{^{IR}\boldsymbol{r}_i \in S_{R,in}(\boldsymbol{\phi})} p(^{IR}\boldsymbol{r}_i) + \sum_{^{IR}\boldsymbol{r}_i \in S_{R,out}(\boldsymbol{\phi})} p(^{IR}\boldsymbol{r}_i) \right\} \Big/ (2 \times N_{R,in} + 0.1 \times N_{R,out}) \tag{8}$$

If the projected 2D model is completely coincide with the captured target object in the left and right images, the fitness value that calculated by Eq.(8) 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 [12], in which different models including a rectangular shape surface-strips model were evaluated using images from a single camera.

$$\overbrace{\underbrace{01\cdots01}_{10bit}}^{{}^{H}z_M}\overbrace{\underbrace{00\cdots01}_{7bit}}^{{}^{H}z_{\theta_x}}\overbrace{\underbrace{11\cdots01}_{7bit}}^{{}^{H}z_{\theta_y}}$$

Fig. 5. Gene information

Figure 4 shows the principle to calculate the fitness value. In Fig.4 (a), the sampling points are indicated by white dots as $^{IL}\boldsymbol{r}_i^{j-1}, ^{IL}\boldsymbol{r}_i^{j}, ^{IL}\boldsymbol{r}_i^{j+1}$. In Fig.4 (b), it shows another situation that the model overlapped more area than Fig.4 (a). In Fig.4 (b), the (A) shows the situation that the sampling points in inner area overlaps with the target object. In this situation, the hue value of the sampling point ($H_M(^{IL}\boldsymbol{r}_i^j)$) is close to the hue value of target object ($H_I(^{IL}\boldsymbol{r}_i)$), if the difference of them is less than 20, that is $|H_M(^{IR}\boldsymbol{r}_i^j) - H_I(^{IR}\boldsymbol{r}_i)| \leq 20$). In this case, the the fitness value would be increased with the voting value "+2." Else, as the situation (B) in Fig.4 (b), the sampling points overlaps with the background area, in which case the difference of the $H_I(^{IL}\boldsymbol{r}_i)$ and $H_M(^{IL}\boldsymbol{r}_i)$ is larger than 20, the fitness value will decreased with voting value "-1." Similarly, to the sampling points in outer area, the (C) means the sampling points overlaps with the background, whose hue value $H_I(^{IL}\boldsymbol{r}_i)$ is close to the sampling points $H_I(^{IL}\boldsymbol{r}_i)$. In this case, it meet the rules as ($|H_M(^{IR}\boldsymbol{r}_i^j) - H_I(^{IR}\boldsymbol{r}_i)| \leq 20$), therefore the fitness value of situation (C) will increased with voting value "+0.1." Otherwise, in the case (D), if the sampling points in outer area overlap with the target object, the fitness value will decreased with voting value "-2"

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 real-time 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 real-time recognition in 30 FPS. The reason why we choose RM-GA has been discussed in [13].

In proposed RM-GA, each chromosome includes 24 bits for searching three parameters: ten for position and fourteen for orientation as shown in Fig.5. The GA operation are conducted in the sequence as evaluation, sorting, obsolete, crossover and mutation. These operations are repeated several times in 33[ms] to generate the best individual.

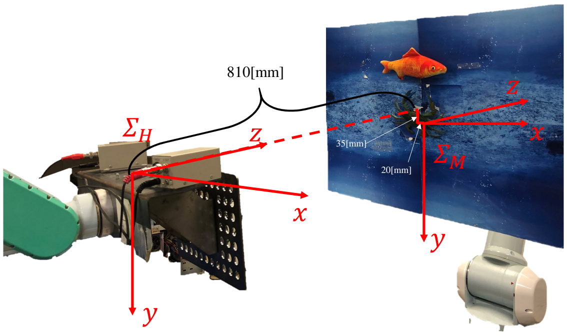


Fig. 6. The experiment environment to conduct pose estimation.

4 EXPERIMENT

Experiment environment

In order to confirm the effectiveness of Pb3DP in estimating 3D pose of target objects, this section is consisted of two parts. In the first part, the experiment is conducted to examine the position estimation result. In the second part, the experiment are focused on the orientation recognition. The experimental environment in two experiments is same, as shown in Fig.6. The target objects are set on the background, which is handled by a manipulator. The coordinate system of target object is defined as Σ_M and the coordinate system of robot hand is Σ_H. The Eq.(9) shows the position relationship of Σ_H and Σ_M.

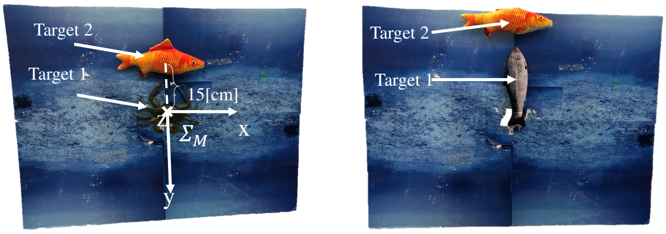
$$^{H}\boldsymbol{T}_M = \begin{bmatrix} 1 & 0 & 0 & 20[mm] \\ 0 & 1 & 0 & 35[mm] \\ 0 & 0 & 1 & 810[mm] \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

In two experiments, the target objects are different and they are shown in Fig.7. The Fig.7(a) shows the targets that used in the position recognition experiment, and the Fig.7(b) shows the ones used in orientation recognition experiment. During the experiment, the manipulator that holds the background moved in a set trajectory which changed the target objects' position and orientation as shown in Fig.8. The upper part of Fig.8 shows the trajectory change in position recognition and the lower part is the trajectory change in orientation recognition experiment.

4.1 Results of experiment

4.1.1 (a). The estimation towards position in x, y, z axes

The experiment is conducted in 60 [sec], the results can be referred to Fig.9 (a), the blue lines mean the estimated position of target objects and the red lines mean true position of target objects. The movement of target objects are shown in the upper part of Fig.8. From the result, in x, y axes, the estimated position is very close to the true position of target objects, in recognition of z axis, due to the objects themselves have a certain thickness, the result of recognition is slightly different from the true value. After considering the thickness of the objects, the recognition results are also basically coincides with the true value.



(a) The target objects used in position recognition experiment

(b)The target objects used in orientation recognition experiment

Fig. 7. The target objects used during position and orientation recognition experiments

4.1.2 (b). The estimation towards orientations around x, y, z axes

The experiment is conducted in 70 [sec], the results can be referred to Fig.9 (b), the black lines mean the estimated position of target objects while the red lines mean true orientation of target objects. The movement of target objects are shown in the lower part of Fig.8. From results, the estimation towards orientation around x, y axes has a obviously gap from true value of orientation of target objects. In the first half of estimation result of target 1, the recognition result in x axis is better than y axis as the change around x axis was estimated obviously. It may because the target 1 is placed vertically and the rotation around the x axis is more pronounced for objects placed vertically. Similarly, for objects placed horizontally, the rotation around the y-axis will produce more obvious parallax changes. Therefore, for the Target 2, which was placed horizontally, the estimation result of orientation change around y axis is better than x axis.

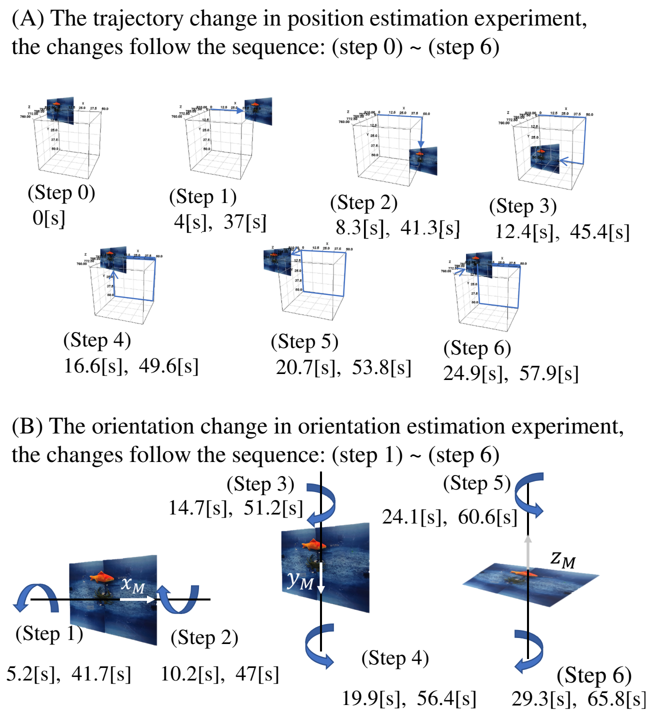
However, the estimation of orientation around z axis has improved significant improvement as the parallax changes more obviously in left and right camera.

5 CONCLUSION

This paper introduced the methodology of Pb3DP, and conducted the experiment to examine if the Pb3DP method could recognize the target's pose in real-time. From the results, the position recognition result can basically coincides with the truth value while the orientation recognition results are depend on the whether there was a obvious parallax change.

From this paper, the following conclusions can be confirmed:

- As the combination of projection-based method and


(A) The trajectory change in position estimation experiment, the changes follow the sequence: (step 0) ~ (step 6)

(B) The orientation change in orientation estimation experiment, the changes follow the sequence: (step 1) ~ (step 6)

Fig. 8. The pose change during the experiment. In the upper part, (A) shows the trajectory change to examine the position estimation by Pb3DP. The lower part (B) shows the orientation change during experiment to examine the effectiveness of orientation recognition.

RTMS-GA, the Pb3DP method that utilizes stereo-vision can recognize target's pose in real-time.

- The priori knowledge is unnecessary for Pb3DP as different targets can be recognize in the same time, and the target objects are switched at any time.

REFERENCES

[1] Hutchinson, S., Hager, G. D., & Corke, P. I. (1996), A tutorial on visual servo control. IEEE transactions on robotics and automation, 12(5): 651-670

[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

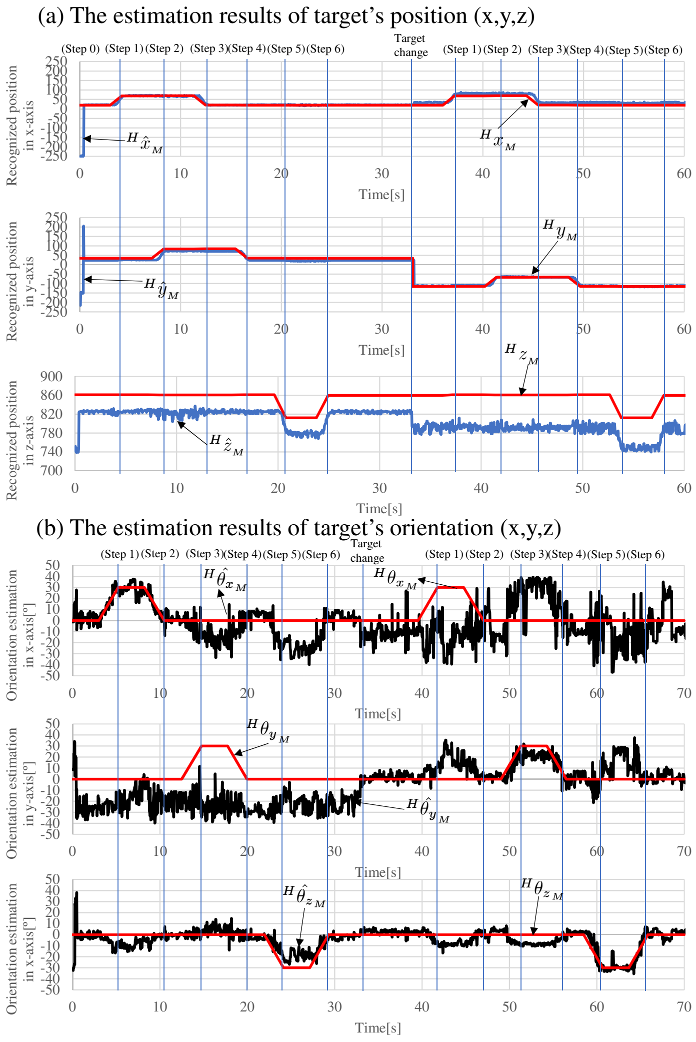
[3] Chaumette, F. (1998), Potential problems of stability and convergence in image-based and position-based visual servoing. In The confluence of vision and control, pp. 66-78

[4] 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

[5] Marr, D., & Poggio, T. (1976), Cooperative computation of stereo disparity. Science, 194(4262), 283-287

[6] Barnard, S. T., & Fischler, M. A. (1982), Computational stereo. SRI INTERNATIONAL MENLO PARK CA ARTIFICIAL INTELLIGENCE CENTER

[7] Poggio, T., & Edelman, S. (1990), A network that learns to recognize three-dimensional objects. Nature, 343(6255), 263


(a) The estimation results of target's position (x,y,z)

(b) The estimation results of target's orientation (x,y,z)

Fig. 9. The result of pose experiment, (a) shows the results in position estimation which conduct in x, y, z axes, (b) shows the orientation estimation of two target objects, in which the target objects changed their orientation around x, y, z axes

[8] Ullman, S., & Basri, R. (1989), Recognition by linear combination of models (No. AI-M-1152). MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB.

[9] Lwin, K. N., Mukada, N., Myint, M., Yamada, D., Yanou, A., Matsuno, T., & Minami, M. (2018), Visual Docking Against Bubble Noise With 3-D Perception Using Dual-Eye Cameras. IEEE Journal of Oceanic Engineering

[10] Lwin, K. N., Myint, M., Mukada, N., Yamada, D., Matsuno, T., Saitou, K., & Minami, M. (2019), Sea Docking by Dual-eye Pose Estimation with Optimized Genetic Algorithm Parameters. Journal of Intelligent & Robotic Systems, 1-22

[11] Yoshikawa, T. (1990), Foundations of robotics: analysis and control. MIT press

[12] Song, W., Minami, M., Mae, Y., & Aoyagi, S. (2007), On-line evolutionary head pose measurement by feedforward stereo model matching. In Proceedings 2007 IEEE International Conference on Robotics and Automation, pp. 4394-4400

[13] Myint, M., Lwin, K.N., Mukada, N., Yamada, D., Matsuno, T., Toda, Y., Kazuhiro, S. and Minami, M., 2019. Experimental verification of turbidity tolerance of stereo-vision-based 3D pose estimation system. Journal of Marine Science and Technology, 24(3), pp.756-779.