The Twenty-Seventh International Symposium on Artificial Life and Robotics 2022 (AROB 27th 2022),
The Seventh International Symposium on BioComplexity 2022 (ISBC 7th 2022),
The Fifth International Symposium on Swarm Behavior and Bio-Inspired Robotics 2022 (SWARM 5th 2022)
ONLINE, January 25-27, 2022

Space sensing system using photo/projection combined method

Yejun Kou[1†], Lujie Wang [2], Yuichiro Toda[3] and Mamoru Minami[4]

[1]Graduate School of Natural Science and Technology Okayama University, Okayama, Japan
(Tel: +81-090-6834-0079; E-mail: ptlg9dvi@s.okayama-u.ac.jp)
[2]Hiroshima University, Hiroshima, Japan
(Tel: +81-070-7564-0169; E-mail: wanglujie2017@yahoo.co.jp)
[3]Graduate School of Natural Science and Technology Okayama University, Okayama, Japan
(E-mail: ytoda@okayama-u.ac.jp)
[4]Graduate School of Natural Science and Technology Okayama University, Okayama, Japan
(E-mail: minami-m@cc.okayama-u.ac.jp)

Abstract: The technologies of space sensing have been advancing nowadays. Many researches have resorted to making a recognition model in advance to estimate the target's pose and detect the target object. Although it can provide pose recognition results, preparing a correct recognition model is time-consuming and complicated. For this problem, our research group has proposed a 3D space sensing method named Projection-based 3D perception (Pb3DP), which can estimate an arbitrary target's pose without prior knowledge if the target is designed by some procedure. However, without the designation of a specific target, the Pb3DP method is functionless because the Pb3DP cannot understand which target's pose should be measured. To this problem, our research group combined the Pb3DP method with the photo-based method, which is a method that can distinguish a specific target from the background and estimate its pose through a 2D photo-model. Then the combined method can specify the target whose pose is required to measure, then Pb3DP can output the pose of the target.

Keywords: Pose sensing, 3D perception, projection-based 3D perception, photo-based model method

1. INTRODUCTION

It is not a simple task to make robots obtain intelligence and perceive abilities like humans. Many researchers have made progress in the field of robot vision. Most of them use predefined object models or depth cameras to obtain the orientation or posture of the object[1-5], but if these conditions such as predefined models and infrared light distance measurement are not given, it seems that there is still a long way to go for the robot to keep working in an unknown environment.

The 3D perception ability of robots is essential in its mobility in an unknown environment. Take the underwater robot as an example. An intelligent underwater robot should be able to cruise itself in the unknown ocean without colliding with the underwater environment and recognize the target and track it when the underwater robot finds pests such as the devil starfish. In order to achieve this function, the author proposes a new system that combines the previous research: a combination of projection-based 3D perception method (Pb3DP)[6] and photo-based model methods[7]. Pb3DP is a new 3D space recognition method. It does not require any prior knowledge of the object. Thus the Pb3DP method can estimate an arbitrary target's pose as long as it can be seen in the camera image. Through the 2D image of the object projected by the camera, it can infer the pose of the object relative to the robot in space based on the parallax.

The Pb3DP only uses visual information to function in environments where infrared rays cannot be used effectively, such as underwater and outdoors. The photo-based model method uses a pre-photographed 2D photo of the object to recognize the position and orientation of the specified object in the 3D space. Then the photo-model based method can distinguish the specific target object from the other everything. These two methods complement each other and cope with the demand mentioned above for the robot's 3D perception abilities.

In this paper, the authors explained the system of the combination of the Pb3DP method and the photo-based model method, and proved its effectiveness through experiments: tracking the object when it appears in the field of view, and maintaining itself relative to the object when it leaves the field of view The relative position of the environment.

† Yejun Kou is the presenter of this paper.

2. PHOTO/PROJECTION COMBINED SYSTEM

As mentioned in the Introduction, this paper proposes a new system that combines the model method and the projective method. Due to the characteristics of these two methods, the photo-model based method can use the 2D photo model to recognize the specified target; the Pb3DP method can measure the distance of any target from the camera. Therefore, this system can make the robot own the ability to work in an unknown envrionment without colliding with the environment.

In this system, the Pb3DP method and the photo-model based method run simultaneously, so we need a way to distinguish whether the object specified by the model method appears in the field of view. To this problem, the author designed that if a specified target's fitness value is greater than a threshold f_{pbm}, then switch the robot into a mode of tracking the specified object. In other cases, the Pb3DP method is used to avoid collisions by measuring the distance between

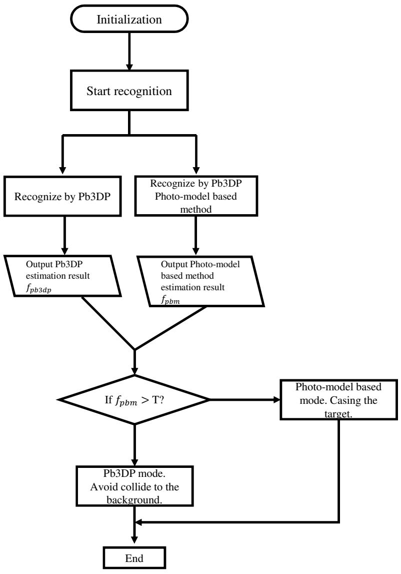

Fig. 1. The schematic of combined system

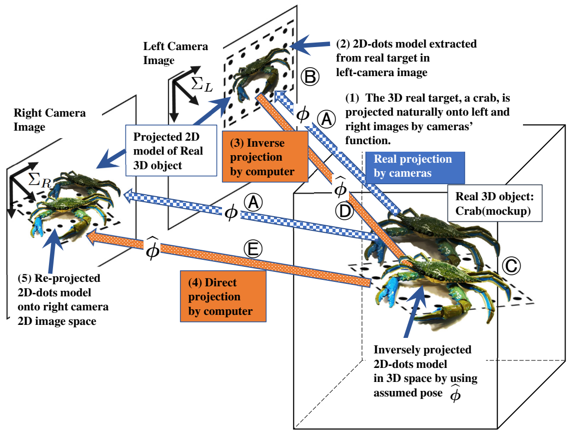

Fig. 2. The schematic of Pb3DP method

the environment and the robot itself. The system can be depicted as a flow-chat, figure. 1.

3. METHODOLOGY

For the paper proposed a new system that combined the Pb3DP and the photo-based model method, this section will explain the methodologies of the this two methods separately.

3.1. 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.

3.2. The conception of Pb3DP

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 Ⓐ. The natural projection is thought to be completed based on true pose, $\boldsymbol{\phi}$, but it is unknown. (2): Target object is selected in the left camera scene as a model depicted by Ⓑ. (3): The selected 2D target model is inversely projected into the 3D space with an assumed pose $\hat{\boldsymbol{\phi}}$, which is 2D and flat, shown by Ⓒ and the inverse projection is indicated by Ⓓ. 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 $\hat{\boldsymbol{\phi}}$, Ⓔ. (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, $\hat{\boldsymbol{\phi}} = \boldsymbol{\phi}$.

3.3. 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:

- ${}^W\boldsymbol{r}_{Mi}$: 3D position vector in Σ_W of an i-th point on a 2D model defined by Σ_M.
- ${}^M\boldsymbol{r}_i$: 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 ${}^M\boldsymbol{r}_i$ is a constant vector since Σ_M is attached to the model.
- ${}^L\boldsymbol{r}_{Mi}$: 3D position vector in Σ_L, projection coordinates defined in Fig.**??**, of an i-th point on a 2D model in space defined by Σ_M.
- ${}^{IL}\boldsymbol{r}_{Mi}$: 2D position in left image coordinate system Σ_{IL}, left-camera image coordinates defined in Fig.**??**, 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.**??** and Σ_L as ${}^L\boldsymbol{T}_M$, the relation between ${}^L\boldsymbol{r}_{Mi} = \left[{}^L x_{Mi}, {}^L y_{Mi}, {}^L z_{Mi}, 1\right]^{\mathbf{T}}$ and ${}^M\boldsymbol{r}_i = \left[{}^M x_i, {}^M y_i, {}^M z_i, 1\right]^{\mathbf{T}}$ is represented by

$$
{}^L\boldsymbol{r}_{Mi} = {}^L\boldsymbol{T}_M \; {}^M\boldsymbol{r}_i. \tag{1}
$$

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

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

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

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

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.

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

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

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

The Eq.(2) can be modified into

$$^{L}\boldsymbol{T}_{M}(\boldsymbol{\phi})^{M}\boldsymbol{r}_{i} = \boldsymbol{P}^{+}(\boldsymbol{\phi})^{IL}\boldsymbol{r}_{Mi} + (I_4 - \boldsymbol{P}^{+}\boldsymbol{P})\boldsymbol{l}. \tag{5}$$

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

$$^{W}\boldsymbol{r}_{Mi}(\hat{\boldsymbol{\psi}}) = {}^{W}\boldsymbol{T}_{M}(\hat{\boldsymbol{\psi}})^{L}\boldsymbol{T}_{M}^{-1}(\hat{\boldsymbol{\psi}}) \Big[\boldsymbol{P}^{+}(\hat{\boldsymbol{\psi}})^{IL}\boldsymbol{r}_{Mi} + (\mathrm{I}_4 - \boldsymbol{P}^{+}(\hat{\boldsymbol{\psi}})\boldsymbol{P}(\hat{\boldsymbol{\psi}}))\boldsymbol{l} \Big]. \tag{6}$$

Here, $\boldsymbol{\psi}$ is defined as $\boldsymbol{\psi} = [^{L}z_{M}, ^{L}\theta_x, ^{L}\theta_y]$. The three components $^{L}z_{M}, ^{L}\theta_x, ^{L}\theta_y$ of $\boldsymbol{\psi}$ 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 $\boldsymbol{\psi}$ be chosen and fixed, shall we describe the fixed valuable as $\hat{\boldsymbol{\psi}}$.

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

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

3.5. The conception of photo-based method

The photo-based mdoel method has been utilized with the adoption of a set-point-model-thinking.That is all points of the solid 3D model in the 3D searching space as a group are projected onto the left and right camera image planes (2D images) without the Corresponding Points Identification Problem that has been pointed out as the difficulty existing in pose estimation by using plural cameras. Since all points on the 3D model are projected into 2D camera images in our method, all projections for each point are correct. This means that forward projection of the 3D object has been used, which does not raise up Corresponding Points Identification Problem. The pose of 2D photo model $\boldsymbol{\phi} = (x, y, z, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ is decided by GA parameter. By using information on the pose of the 3D model, it can be easily understood that how 3D model would be reflected in an image. In order to find out how 3D model appears in the image from the relationship between the camera and the pose of the 3D solid model, it is necessary to project the 3D model to the image plane using the center of projection to generate a plane model. In this section, the projective transformation is described in detail.

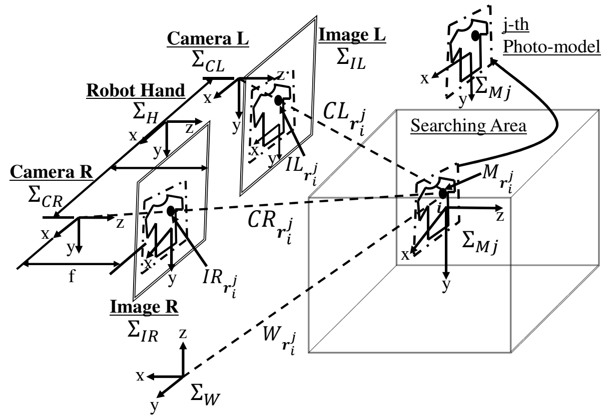

Fig. 3. Perspective projection of dual-eye vision-system: In the searching area, a 3D solid model is represented by the picture of cloth with black point (j-th photo-model). The coordinate systems of photo-model, camera and image are represented by Σ_{Mj}, Σ_{CL}, Σ_{CR}, Σ_{IL} and Σ_{IR} respectively. A 3D solid model that is assumed to be in the searching area is projected from 3D space to 2D left and right camera images.

3.6. Projection transformation in photo-based model method

Figure 3 shows a perspective projection of the dual-eyes vision system. The coordinate systems of dual-eyes cameras and the target object (cloth) in Fig. 3 consist of world coordinate system Σ_W, j-th model coordinate system Σ_{Mj}, 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_{Mj} based on each coordinate system are as follows:

- $^{W}\boldsymbol{r}_i^j$: position of an arbitrary i-th point on j-th 3D model based on Σ_W
- $^{M}\boldsymbol{r}_i^j$: position of an arbitrary i-th point on j-th 3D model in Σ_{Mj}, where $^{M}\boldsymbol{r}_i^j$ is 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 the right camera coordinate system Σ_{CR} to the target object coordinate system 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. (8),

$$ {}^{CR}\boldsymbol{r}_i^j = {}^{CR}\boldsymbol{T}_M(\boldsymbol{\phi}_M^j, \boldsymbol{q})\ {}^{M}\boldsymbol{r}_i^j. \tag{8} $$

where ${}^{M}\boldsymbol{r}_i^j$ is predetermined as fixed vectors since Σ_{Mj} is fixed on the j-th model. ${}^{CL}\boldsymbol{r}_i^j$ that represents the same i-th point on j-th model based on Σ_{CL} is also calculated by using ${}^{CL}\boldsymbol{T}_M(\boldsymbol{\phi}_M^j, \boldsymbol{q})$. Since $\boldsymbol{q}$ can be measured by robot's joint sensors, it could be thought to have been known, then $\boldsymbol{q}$ is omitted hereafter. Eq. (9) represents the projective transformation matrix $\boldsymbol{P}_k$.

$$ \boldsymbol{P}_k = \frac{1}{{}^{k}z_i}\begin{bmatrix} \frac{f}{\eta_x} & 0 & {}^{I}x_0 & 0 \\ 0 & \frac{f}{\eta_y} & {}^{I}y_0 & 0 \end{bmatrix}. \tag{9} $$

where, k = CL, CR, ${}^{k}z_i$; position of the i-th point in the camera sight direction in Σ_{CR} and Σ_{CL}, f ; focal length, $\eta_x; [mm/pixel]$ in x-axis, $\eta_y; [mm/pixel]$ in y-axis. Eq. (10), ${}^{IR}\boldsymbol{r}_i^j$ is the position vector of the i-th point on the j-th model in the right camera image coordinates.

$$ {}^{IR}\boldsymbol{r}_i^j = \begin{bmatrix} {}^{IR}x_i^j \\ {}^{IR}y_i^j \end{bmatrix} \tag{10} $$

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

$$ {}^{IR}\boldsymbol{r}_i^j = \boldsymbol{P}_k\ {}^{CR}\boldsymbol{r}_i^j = \boldsymbol{P}_k {}^{CR}\boldsymbol{T}_M(\boldsymbol{\phi}_M^j) {}^{M}\boldsymbol{r}_i^j \tag{11} $$

$$ \begin{bmatrix} {}^{IR}x_i^j \\ {}^{IR}y_i^j \end{bmatrix} = \frac{1}{{}^{k}z_i}\begin{bmatrix} \frac{f}{\eta_x} & 0 & {}^{I}x_0 & 0 \\ 0 & \frac{f}{\eta_y} & {}^{I}y_0 & 0 \end{bmatrix} \tag{12} $$

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

$$ \begin{cases} {}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j) = \boldsymbol{f}_R(\boldsymbol{\phi}_M^j, {}^{M}\boldsymbol{r}_i^j) \\ {}^{IL}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j) = \boldsymbol{f}_L(\boldsymbol{\phi}_M^j, {}^{M}\boldsymbol{r}_i^j) \end{cases} \tag{13} $$

where ${}^{IL}\boldsymbol{r}_i^j$ can also be described as the same manner like ${}^{IR}\boldsymbol{r}_i^j$. The measurement of the pose $\boldsymbol{\phi}_M^j$, $\boldsymbol{\phi}_M^j = [x_M, y_M, z_M, \theta_M]^T$ will be explained in the following sections.

3.7. 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 $\boldsymbol{\psi}$ in Pb3DP and $\boldsymbol{\phi}_M^j$to the optimization problem. Here we take the true pose $\boldsymbol{\psi}$ in Pb3DP as an example. Please assume that $\hat{\boldsymbol{\psi}}$ means estimated value of $\boldsymbol{\psi}$. If a scalar function $F(\hat{\boldsymbol{\psi}})$ should satisfy that the distribution of $F(\hat{\boldsymbol{\psi}})$ has a single maximum peak F_{max} at true pose $\boldsymbol{\psi}$, and that also satisfy $F(\hat{\boldsymbol{\psi}}) = F_{max}$, then $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}$. This could be rewritten as,

$F(\hat{\boldsymbol{\psi}}) = F_{max}$ if and only if $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi} \in L$ (Single peak assumption, where L means parameter space of $\hat{\boldsymbol{\psi}}$), then the problem to estimate the true pose $\boldsymbol{\psi}$ can be converted to an another problem as,
Find $\hat{\boldsymbol{\psi}}$ to maximize $F(\hat{\boldsymbol{\psi}})$ subject to $\hat{\boldsymbol{\psi}} \in L$.
It means that the estimation of true pose can be completed by optimizing $F(\hat{\boldsymbol{\psi}})$ with parameters $\hat{\boldsymbol{\psi}}$. Then how to constitute a scalar function $F(\hat{\boldsymbol{\psi}})$ satisfying the single peak assumption above appears to be a next problem.

4. REAL-TIME MULTI-STEP GA

4.1. Evaluation Method

In proposed Pb3DP method and photo-based model, 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 [8]. 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 ${}^{IR}\boldsymbol{r}_i^j$, 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. (14). 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. (15).

$$ \boldsymbol{P}_{R,in}({}^{IR}\boldsymbol{r}_i^j) = \begin{cases} 2, if(|H_M({}^{IR}\boldsymbol{r}_i^j) - H_I({}^{IR}\boldsymbol{r}_i)| \leq 20) \\ -1, if(|H_M({}^{IR}\boldsymbol{r}_i^j) - H_I({}^{IR}\boldsymbol{r}_i)| > 20) \end{cases} \tag{14} $$

$$ \boldsymbol{P}_{R,out}({}^{IR}\boldsymbol{r}_i^j) = \begin{cases} 0.1, if(|H_M({}^{IR}\boldsymbol{r}_i^j) - H_I({}^{IR}\boldsymbol{r}_i)| \leq 20) \\ -2, if(|H_M({}^{IR}\boldsymbol{r}_i^j) - H_I({}^{IR}\boldsymbol{r}_i)| > 20) \end{cases} \tag{15} $$

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:

$$ F_R(\boldsymbol{\phi}) = \left\{ \sum_{{}^{IR}\boldsymbol{r}_i \in S_{R,in}(\boldsymbol{\phi})} p({}^{IR}\boldsymbol{r}_i) + \sum_{{}^{IR}\boldsymbol{r}_i \in S_{R,out}(\boldsymbol{\phi})} p({}^{IR}\boldsymbol{r}_i) \right\} \Big/ (2 \times N_{R,in} + 0.1 \times N_{R,out}) \tag{16} $$

4.2. Real-time Multi-step GA (RM-GA)

Searching all possible pose of target object through calculating the fitness value is time-consuming for realtime pose

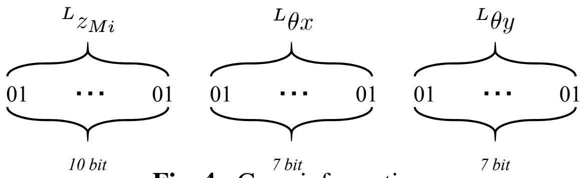

Fig. 4. Gene information

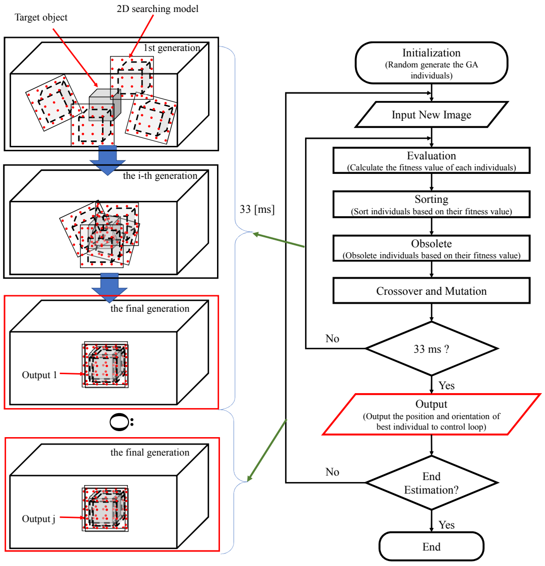

Fig. 5. Flowchart of the RM-GA

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 this paper, the RM-GA is same as the previous researches, we employed Real-time Multi-step GA (RM-GA) to satisfy the realtime recognition in 30 FPS. The and detail and the reason why we choose RM-GA has been discussed in [9].

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 seraching model (Output j in Fig. 5) that represents the true pose with the highest fitness value that calculated by Eq.(16) 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.

5. EXPERIMENT

As the example mentioned in the introduction, an intelligent underwater robot should be able to cruise itself in the unknown ocean without colliding with the underwater environment, which means the underwater robot should be able to maintain the relative pose to the background and it can track the target object when the target object shows up in the field of view. Therefore, in order to meet this demand, we did the following experiments. In the experiment, we did not capture the object but completed the real-time recognition and tracking of the object.

5.1. Experimental contents

The experimental environment of this verification experiment is the same as the experiment performed in the previous chapter. In the experiment, as the Fig.6 shows, the authors placed 6 objects on the rotating table, two of which were the objects to be tracked in this experiment. One of the objects is a model of a redfish, and the other is a model of a sea horse.

In the experiment, the rotating table will be rotated, and the robot will look for the object and track the object when it appears in the camera. However, due to the limitation of the length of the robot in the experiment, when the tracking distance reaches the maximum distance, we assume that the object has been captured or has escaped. After the object disappears from the camera, the robot will move to find the next object. When another object appears in the camera, the robot will track the object that appears in the camera. To accomplish the above goals, I performed the following three experiments.

- Recognition and tracking experiment of rotating table under clockwise rotation.

The experimental results will be explained in the next chapter.

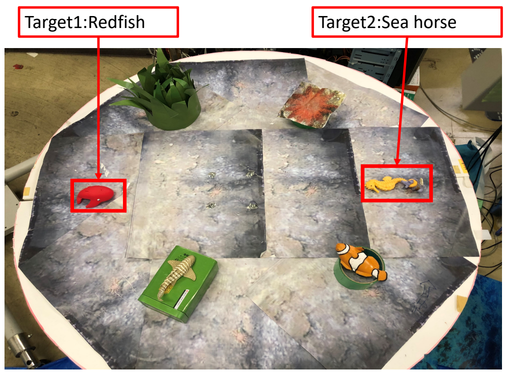

Fig. 6. Placement of the object on the rotating table and two of which are the target of the experiment.

5.2. Recognition and tracking experiment of rotating table under clockwise rotation

The vertical lines in Fig.7 divide the data into three cases. The first case is that the target one is being recognized. In the second case, after the target disappears, the robot's hand moves to the initial position and explores the next target. In this process, we use the project-based method(this method uses the left camera to extract the middle part as the object, and the right eye measures the distance from the object to the robot hand) to measure the height of other objects. The third case is that the target two is being recognized. In search of the target, when the fitness value of the target 1 is greater than

0.6, I consider that the target 1 appears in the camera, and transmit the position and posture information of the target in the camera to the robot. Similarly, when the fitness value of the target 2 is greater than 0.33, I transmit the information of the current target in the camera to the robot.

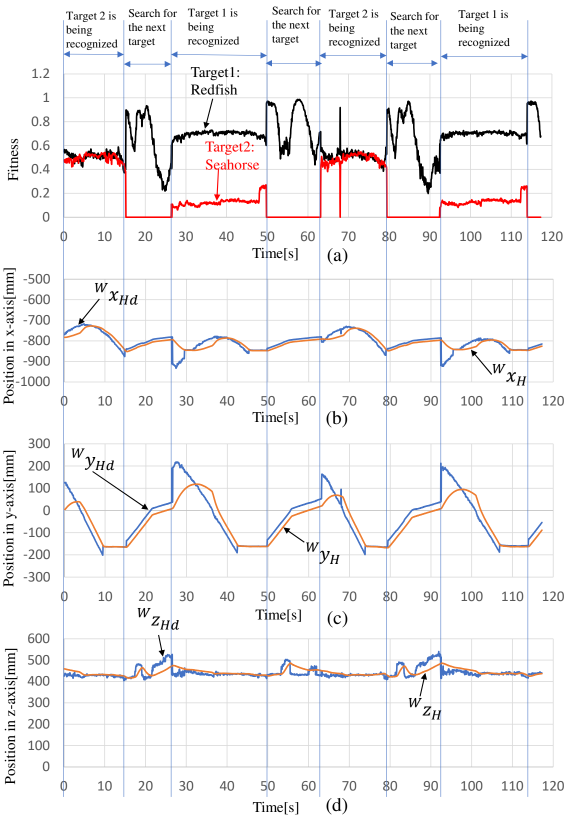

Fig. 7. Experimental results when rotating table rotated clockwise. The red line in (a) is the fitness of the target 2(Seahorse), and the black line in (a) is the target 1(Redfish), But in the process of finding the object, the black line is the fitness of right camera. ${}^{W}x_H$,${}^{W}y_H$,${}^{W}z_H$in (b),(c),(d) are the position tracking results of the end-effector. The desired position of the end-effector is $({}^{W}x_{Hd}, {}^{W}y_{Hd}, {}^{W}z_{Hd})$.

Through the results in (b)∼(d), on the overall trend, it can be seen that the robot hand position varies with the changing of the desired position varies. During the tracking of Object 1, we can see that for a short period of time, the target was not tracked in time, but will immediately track the target immediately after a while. In the process of target exploration, we set a target value to make the robot hand look for the next target. In the target 2 tracking,(b)∼(d) indicates that the real position ${}^{W}r_H = [{}^{W}x_H, {}^{W}y_H, {}^{W}z_H]^T$ of the robot hand is also near to the desired one ${}^{W}r_{Hd} = [{}^{W}x_{Hd}, {}^{W}y_{Hd}, {}^{W}z_{Hd}]^T$. Therefore, the results verified that the visual servoing system with the photo-model-based recognition method could track the target's position in time.

6. CONCLUSION

This paper proposes a new system combining Pb3DP method and photo-based model method, which aims to enable the robot to obtain the movement ability in the unknown environment and the ability to pursue the target objects. Through experiments, the author confirmed that it can recognize and track the specified object, and maintain the distance to the background when the object disappears from the field of view, so that it can ensure the safety of its operation to a certain extent.

REFERENCES

[1] Barbosa G B, Da Silva E C. and Leite A C (2021) Robust Image-based Visual Servoing for Autonomous Row Crop Following with Wheeled Mobile Robots. In 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE): pp 1047-1053

[2] Sharma R S, Shukla S, Beher L, Subramanian V K (2020) Position-based visual servoing of a mobile robot with an automatic extrinsic calibration scheme. Robotica 38(5): pp 831-844

[3] He Z, Wu C, Zhang S, Zhao X (2018) Moment-based 2.5-D visual servoing for textureless planar part grasping. IEEE Transactions on Industrial Electronics 66(10): pp 7821-7830.

[4] Zeng X, Gao Y, Hou S, Peng S (2015) Real-Time Multiscale Tracking via Online RGB-D Multiple Instance Learning. J. Softw: 10(11), pp 1235-1244.

[5] Susperregi L, Martínez-Otzeta J M, Ansuategui A, Ibarguren A, Sierra B (2013) RGB-D, laser and thermal sensor fusion for people following in a mobile robot. International Journal of Advanced Robotic Systems: 10(6), p.271.

[6] Yejun K, Hongzhi T, Mamoru M (2020) A Realtime 3D Pose Estimation Method towards Arbitrary Target with Stereo Vision. 25th International Symposium on Artificial Life and Robotics: 831-836

[7] Tian H, Kou Y, Li X, & Minami M (2020) Real-time pose tracking of 3D targets by photo-model-based stereo-vision. Journal of Advanced Mechanical Design, Systems, and Manufacturing, 14(4): JAMDSM0057-JAMDSM0057.

[8] Minami M, Zhu J, Miura M (2003) Real-time Evolutionary Recognition of Human with Adaptation to Environmental Condition, The Second International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS), Proceeding CD-ROM PS010103

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