The Twenty-Third International Symposium on Artificial Life and Robotics 2018 (AROB 23rd 2018),
The Third International Symposium on BioComplexity 2018 (ISBC 3rd 2018),
B-Con Plaza, Beppu, Japan, January 18-20, 2018

# Improvement of 3D-Pose Real-time Estimation by Active Marker and HSV-evaluated Function

Daiki Yamada[1],Naoki Mukada[1], Myo Myint[1], Khin Nwe Lwin[1], Takayuki Matsuno[1] and 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]p7kw2h27@s.okayama-u.ac.jp

**Abstract:** Nowadays, AUV is playing an important role for human society in different applications such as inspection of underwater structures (dams, bridges). Underwater recharging function is one of the solution to enable the AUV to operate for extended periods independently on a surface vehicle for recharging. we have developed dual-eyes vision-based docking system especially for final docking step. The most challenging and unavoidable problems in sea operations are turbidity and light changing. In this study, we newly designed an active -light emitting- 3D marker and a fitness function determined by HSV color components to improve the performance of the system especially in a more turbid environment. We conducted docking operation using this system in a pool with dark and turbid water (8.0 FTU) for verifying utility of methods we proposed. The experimental results have confirmed the robustness of the docking system using the improved method against turbidity.

**Keywords:** Visual Servoing, Remotely Operated Vehicle(ROV), Genetic Algorithm, Docking, Turbid Environment.

## 1 INTRODUCTION

The most challenging and unavoidable problems in sea operations are, we think, turbidity and light changing. Since underwater battery recharging units are supposed to be installed in deep sea to save the time consuming and work done from human beings in the case of returning surface vehicle for recharging, the deep-sea docking experiments cannot avoid turbidity and low light environment. In previous studies, we conducted sea docking experiments using a passive marker and fitness function based on only hue information. However, the performance of our system is limited in lower turbid sea when we conducted in coastal area. Therefore, we improved our system to expend tolerance of our system against turbidity.

In this study, to improve our system removing above defects, we newly designed an active -light emitting- 3D marker and a fitness function determined by HSV color components to improve the performance of the system especially in a more turbid environment. The advantage of using an active 3D marker is to suppress the diffusion of light more than the previous method in recognizing. In this way, AUV can detect 3D marker more easily because there is not clipped whites on the camera images caused by turbidity. However, the above problem could not be solved completely only using the active marker. More turbidity becomes high, more a border between lighting parts in the marker and non-lighting parts become blurred by scattering of light. Therefore, to overcome this problem, we designed the new fitness function with Hue in which "Saturation" and "Value" information which are components of HSV color representation are used additionally as evaluation indexes for discriminating between lighting and non-lighting parts.

## 2 3D RECOGNITION AND CONTROL BY 3D-MOS

The system, named 3D Move-on Sensing (3D-MoS) in which 3D perception enabled by dual-eyes visual pose tracking by using known 3D marker is used for controlling the vehicle's relative pose to the desired one. The 3D-MoS system recognizes a relative pose between a robotic system (remotely operated vehicle (ROV) in this study) and a target object by utilizing 3D model-based recognition using dual-eye cameras images with a video frame rate of 30 fps. In the proposed approach, visual information is directly used in feedback control in real-time. Additionally, developed optimization method named Real-time Multi-step GA is implemented in accordance with the concept of optimization of dynamic images for real-time target tracking. This combination of 3D-MoS and Real-time Multi-step GA in feedback loop using two cameras images is the main novelty of this study.

In the proposed 3D-MoS, we first create a shape attribute model that incorporates attributes into the 3D known shape model of the recognition object. Then we define a scalar function which means the correlation between multiple models projected onto images of dual eyes and actual objects appearing in an image of dual eyes. It is possible to set the model shape, light environment or correlation function so as to take the maximum fitness value when the position and orientation of the 3D model overlaps those of the object. In such a state, the problem of finding the 3D position and orientation

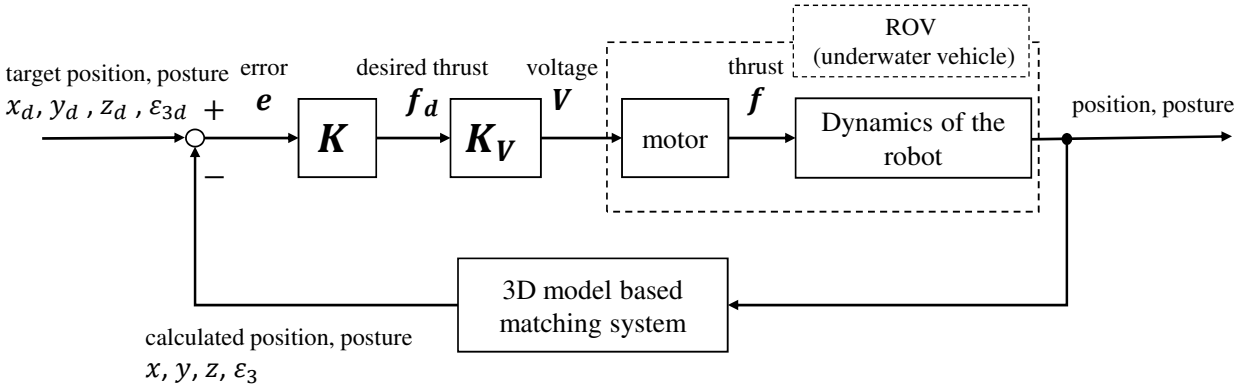



Fig. 1. Control logic for the proposed system

of the object is an optimization problem for finding a variable that maximizes the correlation function. Furthermore, by conducting an optimal solution search using the correlation function as a fitness of the genetic algorithm (GA), it is possible to effectively search for the optimal solution even when the distribution of the correlation function with respect to the position and orientation becomes a multi modal function due to noise.

### 2.1 3D Model-based Matching Method

Knowing the information of the target and predefined relative pose to the AUV, the solid model of the target is predefined and projected to 2D images. Comparing the projected solid model image with the captured 2D images by dual cameras, the relative pose difference can be calculated. A control system using this method is shown in Fig. 1.

### 2.2 3D Model-based Recognition using Real-time Multi-step GA

Fitness value which is correlation function of projected model against the real target in the image is used as the evaluation parameter of recognition process. Even through there are classical computer vision algorithms to obtain the relative pose estimation, GA provides recognition performance in terms of effectiveness, simplicity and repeatable evaluation for real-time performance. Therefore, GA named as Real-time Multi-step GA (Detail explanation can be seen in [1][2][3]) in this experimental system is capable of real-time recognition of the moving image effectively and confirmed in our previous works [4][5]. In this method, the genes which represent the different relative poses of 3D model to the ROV are initiated randomly. According to defined fitness function, the gene with the highest fitness function value represents the pose of the real target. Therefore, the searching problem of real target pose addresses the optimization problem. Through the steps of GA (Selection, Cross over and Mutation), a number of genes that represent different poses are evaluated by the defined fitness function to get the best gene with the most truthful estimated pose. This 3D model-based matching process is executed within 33 ms synchronizing with the video rate of dual-eyes camera. Fig. 2 shows searching area of RM-GA.

### 2.3 Underwater robot system

Remotely controlled underwater robot used in this experiment (manufactured by KOWA) is shown in Fig. 3.

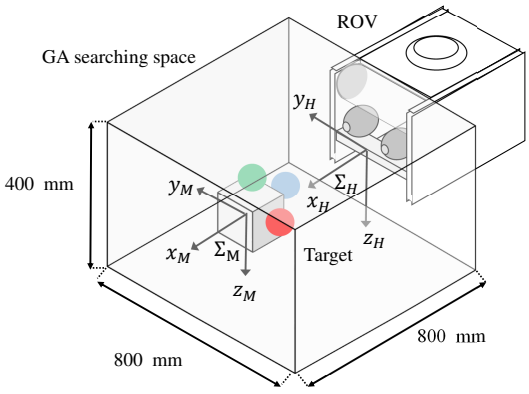



Fig. 2. GA searching area and coordinate systems of the robot and the real target.

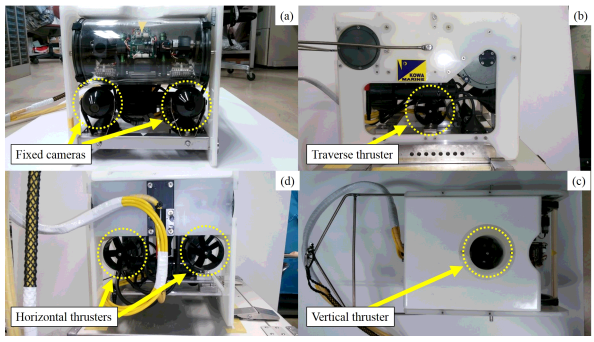


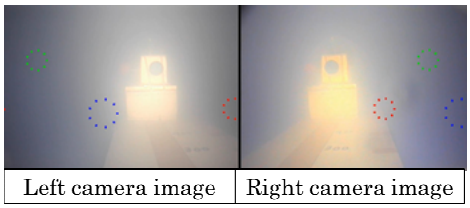

Fig. 3. Underwater vehicle which we used in the experiment.

As the main sensor that is visual sensor for this robot , the two fixed forward cameras are used for 3D object recognition in visual servo. In the thruster system of ROV, 2 horizontal thrusters with maximum thrust of 9.8 N, 1 vertical thruster with maximum thrust of 4.9 N and 1 lateral thruster with 4.9 N are installed.

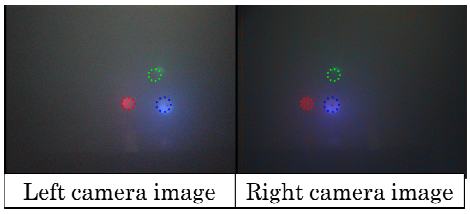
## 3 NEW PROPOSED METHODS

### 3.1 Lighting 3D Marker

We call the target object used in this study "3D marker". A conventional 3D marker did not emit light (passive); thus, in recognition under dark and turbid environments, the robot had to use the lighting mounted on itself. However, in that case, as shown in Fig. 4(A), it is has been confirmed that the recognition has failed because the camera image is entirely blurred white. This is thought to be caused by diffused reflection of light on the particles in the turbid water. This is a factor that makes recognition in the dark and turbid environment difficult, which can be a fatal obstacle to use the system in the actual sea area. From this point of view, it can be said that it is necessary to minimize the amount of light reaching the camera in order to prevent diffused reflection and accurately recognize the object. Therefore, we newly propose "lighting 3D marker" which emits light. By doing so, it can take in only the light necessary for recognition, which can be recognized even in higher turbid environment (Fig. 4(B)).




(*A*) Non-lighting marker in 15.3 FTU

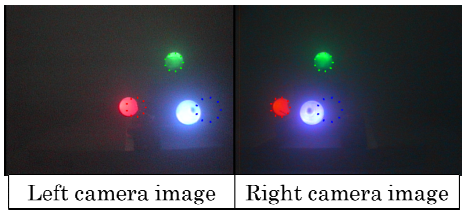



(*B*) Lighting marker in 15.0 FTU

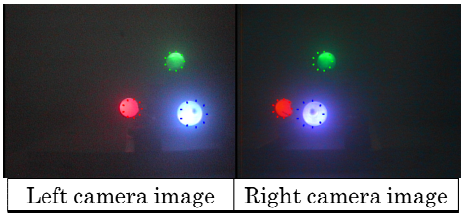
Fig. 4. Comparison of appearances of a lighting marker and a non-lighting one under turbid environment.

### 3.2 Union evaluation function using HSV color information

The previous evaluation method used in [6], [7] was based only on the Hue value of the object. However, using this method, its relative pose cannot be recognized well while using the lighting 3D marker proposed this time in a dark and turbid environment (Fig. 5(A)). The reason is that as the robot approaches the marker, the amount of light reaching the cameras increases, and on the camera image, the strongly emitting part appears white. When using only Hue information for evaluation, there is a disadvantage that an achromatic color such as white or black cannot be evaluated. Therefore, when the amount of light reaching the camera is large and light is diffused due to turbidity, as shown in Fig. 5(A), the outside of the light emitting part is recognized, and as a result deviation from the true value occurs. Here, this problem is solved by adjusting the light quantity of the light emitting part according to the turbidity and the distance between the robot and the marker. However, when constructing such a system, it is inevitable that it becomes a very complicated system. Therefore, we need to think about a simpler method. Another method newly proposed in the paper is a composite type object evaluation method using three color elements of HSV. This is in consideration of the evaluation of "lighting" in addition to the conventional evaluation using "color". As shown in Fig. 6, score is evaluated for each point (pair of points in the case of lighting evaluation) by using correlation function ( Eq. (1), (2) and (3)). Then, by synthesizing the two fitness values according to Eq. (4), "union fitness value ($F_{union}$)" which we newly defined is obtained ultimately. Eq. (4) plays the role of so-called "OR" in which the union fitness value becomes higher if either of the two fitness values (fitness values of color evaluation ($F_{hsv}$) and light emission evaluation ($F_{v_d}$)) is higher. Therefore, the model of the object has redundancy. Here, it is assumed that $N_{pairs}$ has the same value as $N_{out}$ in Eq. (4).




(*A*) Only Hue evaluation in 8.0 FTU




(*B*) Color and light emission evaluation by HSV in 8.0 FTU

Fig. 5. Difference in accuracy between two methods in recognizing the marker.

$$p_{v_d}({}^I\boldsymbol{r}_{in}^{(m)}, {}^I\boldsymbol{r}_{out}^{(m)}) = \begin{cases} 1 & (|V({}^I\boldsymbol{r}_{in}^{(m)}) - V({}^I\boldsymbol{r}_{out}^{(m)})| \geq 10) \\ -1 & (otherwise) \end{cases} \quad (1)$$

$$p_{hsv}({}^I\boldsymbol{r}_{in}^{(n)}) = \begin{cases} 1 & \{H_u({}^I\boldsymbol{r}^{(n)}) \in [h_u - 40, h_u + 40](u = R,G,B), \\ & S({}^I\boldsymbol{r}^{(n)}) \geq 0.10, V({}^I\boldsymbol{r}^{(n)}) \geq 30\} \\ -1 & (otherwise) \end{cases} \quad (2)$$

$$p_{hsv}({}^I\boldsymbol{r}_{out}^{(n)}) = \begin{cases} -1 & \{H_u({}^I\boldsymbol{r}^{(n)}) \in [h_u - 40, h_u + 40](u = R,G,B), \\ & S({}^I\boldsymbol{r}^{(n)}) \geq 0.10, V({}^I\boldsymbol{r}^{(n)}) \geq 30\} \\ 1 & (otherwise) \end{cases} \quad (3)$$

$$\begin{aligned} F_{union} &= F_{v_d} + F_{hsv} - F_{v_d} \cdot F_{hsv} \\ &= \frac{\sum_m^{N_{pairs}} p_{v_d}({}^I\boldsymbol{r}^{(m)})}{N_{pairs}} + \frac{\sum_n^{N_{in}+N_{out}} p_{hsv}({}^I\boldsymbol{r}^{(n)})}{N_{in}+N_{out}} \\ &\quad - \frac{\sum_m^{N_{pairs}} p_{v_d}({}^I\boldsymbol{r}^{(m)})}{N_{pairs}} \cdot \frac{\sum_n^{N_{in}+N_{out}} p_{hsv}({}^I\boldsymbol{r}^{(n)})}{N_{in}+N_{out}} \end{aligned} \quad (4)$$

The fitness function was designed as an evaluation parameter in the pose estimation process. It is a defined correlation between a projected model and a real target in the image. In Fig. 7, the three solid circles and the three circles outlined with dotted lines represent the spheres on the real target and those on the j-th model obtained from 3D-to-2D projection, respectively. The pose $\phi_M^j$ of the 3D model is an unknown variable composed of six parameters $(x, y, z, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ and is determined in the pose estimation process. The 2D projection of each sphere in the model is divided into two regions, as shown by the dashed circles in Fig. 6. Instead of evaluating the positions of all of the points in the model, only select points are considered, as shown in Fig. 6. When the j-th model is projected onto the 2D images of the left and right cameras, the fitness value for that model is calculated. Portions of the target object that lie inside the inner and outer regions of each corresponding sphere of the projected model proportionally increase and decrease the fitness value, respectively. Therefore, the fitness function is maximized when the pose of the model fits that of the target object depicted in the images of the left and right cameras. The evaluation parameters of the fitness function were designed to reduce the effect of noise, which is considered here as peaks in the fitness function that represent incorrect poses of the target. A detailed definition of the fitness function can be found in [7], [8] and [9]. The concept of the fitness function in this study can be said to be extension of the work in [9], in which different models, including a model with rectangular surface strips, were evaluated using images from a single camera.

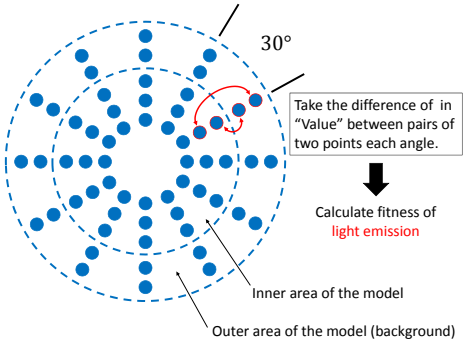



Fig. 6. Projection of the blue sphere of a model with selected sample points. There are a total of 60 points (36(=$N_{in}/3$) and 24(=$N_{out}/3$) points in the inner and outer regions, respectively) in the projection, and the diameter of the inner region is same as that of the actual sphere.

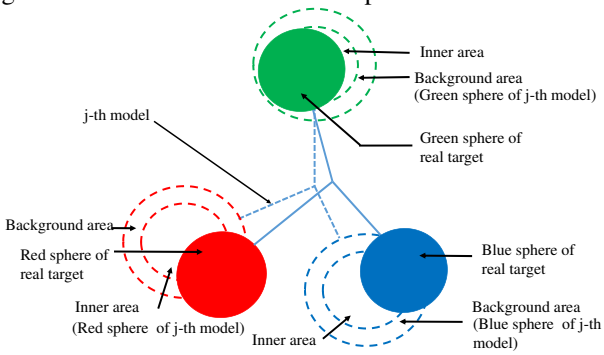



Fig. 7. Real target (solid circles) and projected 3D model (circles with dashed outlines) in a 2D image obtained by the right camera.

## 4 DOCKING EXPERIMENT IN A POOL

### 4.1 A Strategy of Docking Operation

The sea docking strategy is same as the pool test except for the approaching step and stay step. To perform the continuously repeated docking experiments, the proposed docking system includes five steps: (1) Approaching step (approached the docking station by manually until the 3D marker is in the field of view of the cameras), (2) visual servoing step (initial state of the docking), (3) docking step (fitting the docking pole into the docking hole), (4) stay step (staying in front of the docking station for data storing without performing visual servoing), and (5) launching step (go back to the desired position). These steps is explained in [7] in detail.

### 4.2 Experimental Results

Fig. 8 shows the docking experimental results conducted in the pool. (A) is a fitness obtained by synthesizing two values of color fitness and light emission fitness. Based on the synthesized fitness, it is judged whether the robot operates or not. Here, it can be seen that the synthesized fitness oscillates approximately periodically from about 20 seconds onward. To clarify the cause, a comparison with the recog-

nition value in the x direction is shown in (B). Here, as the recognition value in the x direction decreases, the fitness increases and as the recognition value increases, the fitness degree decreases. As can be seen from (G), the light emission fitness, which is one of the fitness which became the origin of (A), is influenced by the distance of x direction between the robot and the 3D marker. Generally, the longer the distance between the light source and the observer becomes, the greater an influence on light affected by turbidity. Therefore, under the influence, the boundary between the light-emitting part of the marker and the background becomes blurred, and the difference in brightness is considered to become small. (C) shows the recognition value and the desired value in the x direction of the robot at the same time. In this time three consecutive dockings were conducted. If the docking conditions shown in (D), (E) and (J) are satisfied, the desired value in the x direction is decreased. According to (C), it can be said that the recognition value follows the desired value. (D) and (E) are recognition values in the y direction and the z direction, respectively. The desired value for these is always constant and is zero. It is understood that the recognition value after entering the docking operation almost always satisfies the docking condition and it is possible to stably recognize even under a turbid environment. (F) and (G) are the color fitness and the light emission fitness. Data is saved in Stay Process, and at that time control is stopped. Here, to speed up data storage, data storage of (F) and (G) is stopped. (F) is around 0.3 to 0.4, which is lower than (G), but shows a stable value without being influenced by turbidity depending on the distance. Regarding (G), as described above, it has a dependency on distance due to turbidity. (H) and (J) are the recognition values of the rotation angle around the x axis and y axis, respectively. Quaternion is used for angle representation, so it has no unit. Robot do not control these because it can maintain a stable posture even if they are not driven. It shows that after the docking operation, it shows stable recognition value as compared with before. Finally, (J) is the recognition value of the rotation angle around the z axis. Although this angle is controlled, it can be said that it shows a stable recognition value after the docking operation like (H) and (I). Besides, regarding the docking condition, all the conditions are almost always satisfied, thereby realizing a stable docking operation.

## 5 CONCLUSION

In this time, we proposed new methods on the measurement of the position and orientation of the object under the dark and turbid environment, and conducted a docking experiment at an indoor pool to demonstrate its utility. It was able to realize recognition and docking operation in turbid water (8.0 FTU) without influence of water flow such as waves. From now on, we will conduct docking experiments in real sea areas with turbid environments using these methods.

## 6 ACKNOWLEDGMENT

The authors would like to thank Monbukagakusho; Mitsui Engineering and Shipbuilding Co., Ltd.; and Kowa Corporation for their collaboration and support for this study.

## REFERENCES

[1] Song W., Minami M. and Aoyagi S., ”Feedforward Online Pose Evolutionary Recognition Based on Quaternion”, *Journal of the Robot Society of Japan*, Vol.28, No.1, pp.55-64 (in Japanese), 2010.

[2] Song W. and Minami M., ”3-D Visual Servoing Using Feedforward Evolutionary Recognition”, *Journal of the Robot Society of Japan* , Vol.28, No.5, pp.591-598 (in Japanese), 2010.

[3] Wei. Song, M. Minami, Fujia Yu, Yanan Zhang and Akira Yanou, ”3-D Hand and Eye-Vergence Approaching Visual Servoing with Lyapunouv-Stable Pose Tracking”, *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 5210-5217, 2011.

[4] Yu F., Minami M., Song W., Zhu J. and Yanou A., ”On-line Head Pose Estimation with Binocular Hand-eye Robot based on Evolutionary Model-based Matching”, *Journal of Computer and Information Technology*, Vol.2, No.1, pp.43-54, 2012.

[5] Suzuki H. and Minami M., ”Visual Servoing to catch fish Using Global/local GA Search”, *IEEE/ASME Transactions on Mechatronics*, V0l.10, Issue 3, pp.352-357, 2005.

[6] Myint, M., Yonemori, K., Yanou, A., Minami, M., Ishiyama, S., Visual-servo-based autonomous docking system for underwater vehicle using dual-eyes camera 3D-pose tracking, In 2015 IEEE/SICE International Symposium on System Integration(SII), (2015), pp. 989-994.

[7] Myo Myint, Kenta Yonemori, Khin Nwe Lwin, Akira Yanou, Mamoru Minami. J Intell Robot Syst (2017). DOI 10.1007/s10846-017-0703-6.

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

[9] Song W, Minami M, Aoyagi S, On-line stable evolutionary recognition based on unit quaternion representation by motion-feedforward compensation, International Journal of Intelligent Computing in Medical Sciences & Image Processing, 2(2), pp.127-139, 2008.

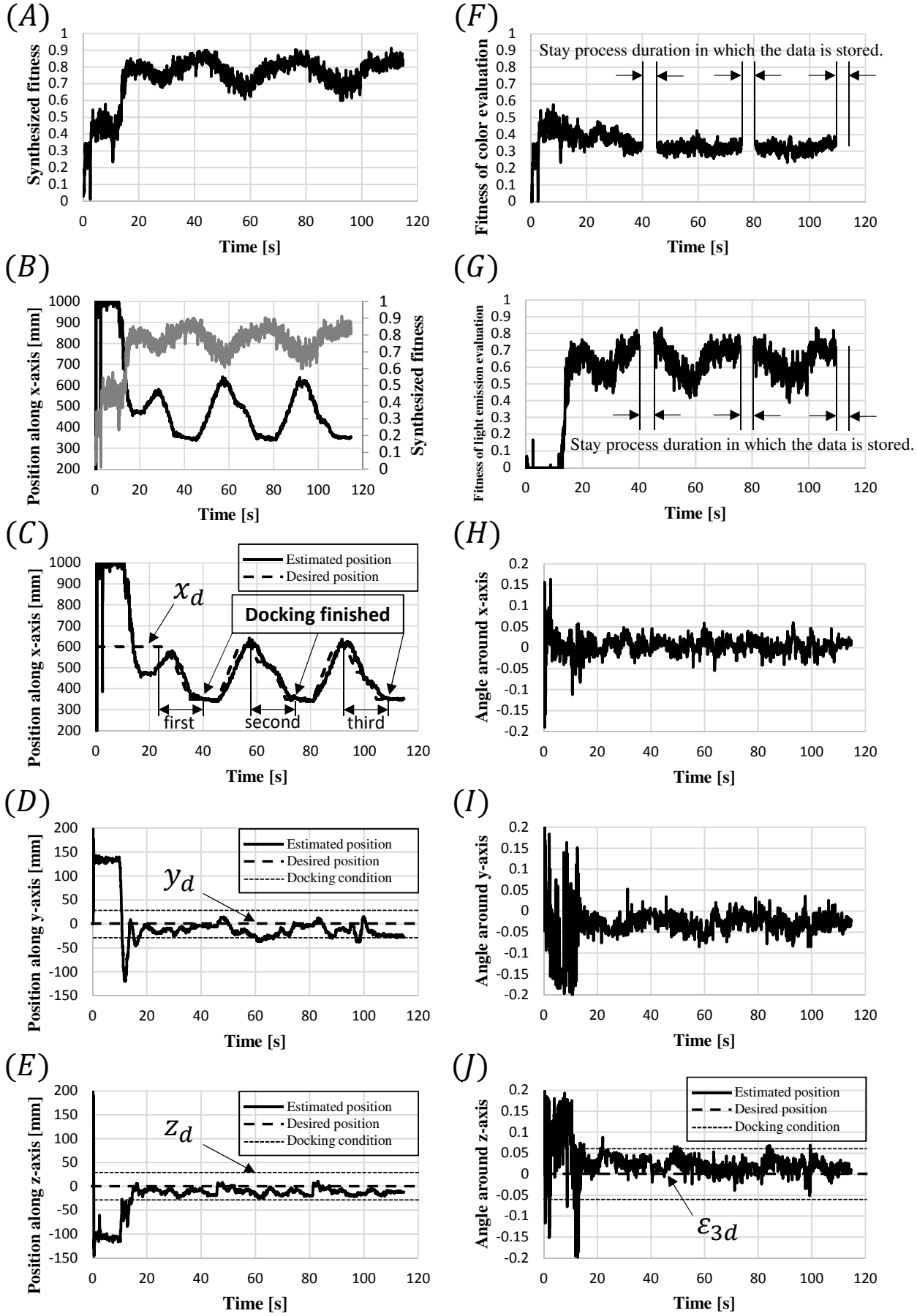



Fig. 8. Experimental result in a pool; (A)Fitness value synthesized of color and light emission fitness, (B)Comparison of (A) and position along x axis, (C)Position along x axis, (D)Position along y axis, (E)Position along z axis, (F)Fitness value of color evaluation, (G)Fitness value of light emission evaluation, (H)Angle around x-axis, (I)Angle around y-axis,and (J)Angle around z-axis.