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

Pose Estimation by Optimizing Real-time Multi-step GA's Parameters

Khin Nwe Lwin[1], Myo Myint[1], Naoki Mukada[1] Daiki Yamada[1], Takayki 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]pdoj8yez@s.okayama-u.ac.jp

Abstract: 3D pose estimation using dynamic images input by video rate should be conducted in short time when the estimated pose would be used for real-time feedback control. Controlling with 3D pose estimated through single camera images has been studied so far ardently, but it has been confirmed that estimated position accuracy in camera depth direction is not enough. The authors have proposed a new 3D position and orientation (pose) estimation method with dual-eye cameras that exploits the parallactic nature that enables reliable 3D pose estimation in real-time, named as "Real-time Multi-step Genetic Algorithm (RM-GA)." This paper focuses on improving dynamic performance of dual-eye real-time pose tracking by tuning parameters used in RM-GA, having confirmed that the dynamical performance in time domain to estimate target marker's pose in real-time has been optimized.

Keywords: Real-time multi-step GA, Visual-servoing, Pose estimation, Dual-eye cameras, Underwater vehicle

1 INTRODUCTION

The studies on visual-servoing-based underwater vehicle have been conducted all over the world in recent years. Most of them have used a single camera to estimate the pose of the target object [1], [2]. A binocular vision was used in some of these studies in order to detect the target position of the vehicle [3], [4]. Even though two cameras were used in [3], one was facing downward for shooting the sea-floor images and the second camera was pointed forward for the purpose of obstacle avoidance. In another study [4], the vehicle's position was estimated using two cameras and a sonar system. In that approach, the position of the vehicle was calculated by combining data from a Doppler sonar and CCD cameras, where the constructed system does not deal with the orientation for controlling the vehicle. This means that 3D pose estimation using parallactic character of dual-cameras has not been realized.

We also developed a visual-servoing type underwater vehicle using dual-eye cameras and a 3D marker for real-time pose tracking by means of visual servoing as shown in Fig. 1. Visual servoing using stereo vision and parallactic character for the underwater vehicle utilizing 3D model-based recognition and Real-time Multi-step Genetic Algorithm (RM-GA) has been initiated by our research group, and have confirmed the effectiveness [5]-[10].

Researchers have discussed optimization of parameters in GA for specific problems [11]-[13]. Avni Rexhepi [11] applied genetic algorithm for travelling salesman problem (TSP) for Kosovo municipalities, with different settings for the parameters of the genetic algorithm. In that approach, they studied the reasonable amount of time for the TSP by analysing the GA parameters based on number of generation.

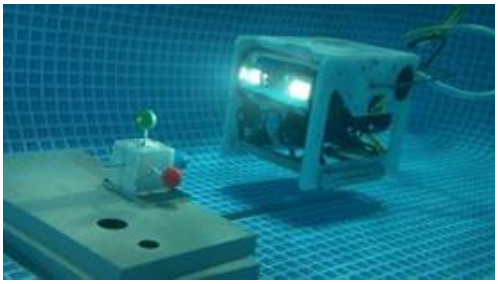

Fig. 1. Underwater vehicle and 3D marker.

Boyabatli, O. [12] analyzes the effect of numerical parameter of GA on its performance and reported that the effect of high mutation rates give better performance. Tabassum, M. [13] utilized the GA for image optimization and the capability of solving the knapsack problem is demonstrated. They studied on how GA parameters affect the reproduction of the original images.

These discussions have been based on iteration number concerning stationary fitness distribution, not on time length used for the optimization. The relative pose estimation in dynamic images of underwater vehicle should be solved optimization problem against time-changing multi-peak fitness distribution as fast as possible for the closed loop stability. Since the quickness of the estimation is related to the time used for the optimization calculation, not the iteration number, the optimization performances should be evaluated on the convergence response measured in time domain. However, most of optimization methodologies have focused on accuracy and iteration number rather than the time used. As discussed above, apart from other approaches that are based on static fitness distribution, pose estimation of moving vehicle should be conducted in time-changing fitness distribution for real time feedback control. To the best of the authors' knowledge, there is no study in the literature analyzing the GA parameters based on real-time performance.

In this context, the contribution of this paper is that the relative pose estimation performance has been confirmed to be improved by optimizing GA's parameters through real time pose estimation experiments. This optimization improves time response performance of the RM-GA to track the moving target 3D marker relative to the vehicle —even though the marker is stationary in space, the pose measured based on the vehicle coordinates dynamically moves—, bringing about the improved stability of the vehicle with the feedback control. This effect of optimization of GA's parameters has significant meaning on stability improvement.

The remainder of the paper is organized as follows: Section 2 describes the method of pose estimation. Experiment results are reported in section 3 with discussion and conclude in section 4.

2 POSE ESTIMATION METHOD

2.1 3D Model-based Matching Method Using Dual-eyes Vision System

The detailed explanation of the 3D pose estimation method using dual-eye cameras and 3D marker are already introduced and explained in previous work [14]. We would like to discuss about 3D pose estimation method briefly for reader convenience in this section. In some approach, epipolar constraints were used to search for corresponding points from a pair of cameras in order to measure the pose of the 3D objects, which uses 2D-to-3D inverse projection. The inverse projection based on 2D-to-3D can cause the wrong mapping point in images. To avoid the effects of incorrect mapping points in images, a three-dimensional model-based pose estimation approach is used in the proposed system with dual-eye cameras based on 3D-to-2D forward projection.

3D model-based pose estimation using dual-eye vision system is shown in Fig. 2. $\sum_{CR}$ and $\sum_{CL}$ are the reference frames of the left and right cameras which are mounted in front of the vehicle respectively. $\sum_{H}$ is the reference frame of the ROV as shown in Fig. 2 and 3. $\sum_{M}$ is the reference frame of the real target object. The search space of the vision system is already defined as shown in Fig. 3. A model of the three-dimensional marker contains three-dimensional shape and color information, which are predefined in a computer system. A 3D marker which is composed of three spheres whose color are red, green and blue is used as a target object. There are many 3D models that have the same 3D information such as shape, color, and size with different poses allocated randomly in the search area. The real target object in space is captured by dual-eye camera and the poses of the dotted-line model, which are given by one of RM-GA

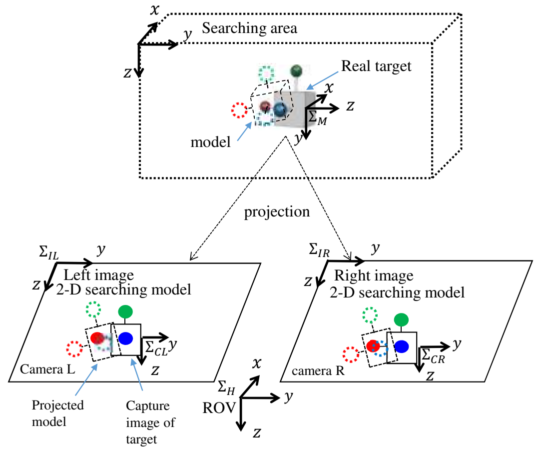

Fig. 2. 3D Model-based matching method using dual-eye vision system.

genes, are projected to 2D image. The matching degree between the projected solid model and the dotted line model is evaluated in 3D space through left and right projected 2D images. The different relative pose is calculated by comparing the projected 2D image and the solid model captured by the dual-eye cameras. Finally, the best model of the target object that represents the true pose can be obtained based on its highest fitness value.

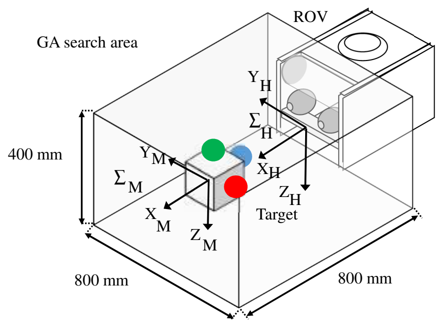

Fig. 3. GA search space.

2.2 Fitness Function

The fitness function has been used to measure how much the degree of matching between the captured image and its projected model with its pose. In other words, the fitness function is the correlation function of 3D pose of individual model with the real 3D target in 3D space. The good fitness functions affect GA to explore the search space and convergence speed more effectively and efficiently. In this system, only hue value is used for recognition of 3D marker because of less sensitive to the environment. Figure. 4 shows the real 3D maker and a model of the target 3D marker. The 3D marker is constructed using red, green, and blue spheres (diameter: 40 mm). Each model consists of three spherical balls (red, green, and blue). The dimensions of the real marker are shown in Fig. 4(a), the model of each sphere is shown in Fig.

4(b) and Fig. 4(c) is the enlarged view of the blue ball model.

Each model consists of two portions, the first portion is the inner area which is the same size as the target and the second portion is the background area. The captured image (pixel) is detected in a 2D image as (green or blue or red) in hue space. If the captured image (on a blue ball of a 3D marker model) situated in the inner portion, the fitness value will increase and the captured image (on a blue ball of a 3D marker model) is situated in the outer portion, the fitness value will decrease. Similarly, the green ball and the red ball are evaluated in the same way. Therefore, the fitness value will be maximum when the model and the real target exactly coincide. The true pose of the real target 3D marker is obtained with maximum fitness value. A concept of the fitness function can be found in our previous study [15].

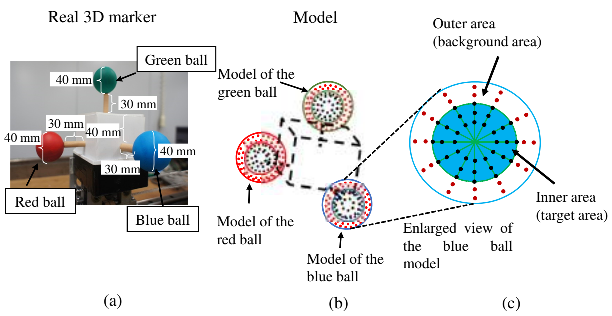

Fig. 4. Real 3D marker and model: (a) real 3D marker, (b) model, (c) enlarged view of the blue ball model, where the inner area is the same size as the real target object (blue ball) and the outer area is the background area. The dots in (c) mean points to calculate the correlation degree on how much the inner area overlaps the blue ball and the outer area does not overlap the blue ball.

2.3 Real-time Multi-step GA Evolution

The reason for choosing RM-GA and how RM-GA works is explained in detail in [14]. Figure 5 represents the flowchart of real-time multi-step GA recognition process and Fig. 3 shows the GA search area. In the flowchart, the recognition process of the target pose is evaluated in 2D and convergence occurs in 3D as shown in Fig. 5(a). The flowchart of the RM-GA is shown in Fig. 5(b). The best pose of the target object is evaluated within 33 ms through the GA process. A population of genes is first generated in random, and a new pair of left and right images is input. The RM-GA procedure is performed within 33 ms. Convergence of the genes to the maximum peak of the fitness distribution, which moves as the input image changes, can be achieved through the RM-GA procedure. The gene information is transferred to the next generation in order to optimize the genes over successively input images. In the evolution, elitism strategy and two-point crossover are used in the genetic algorithm. An

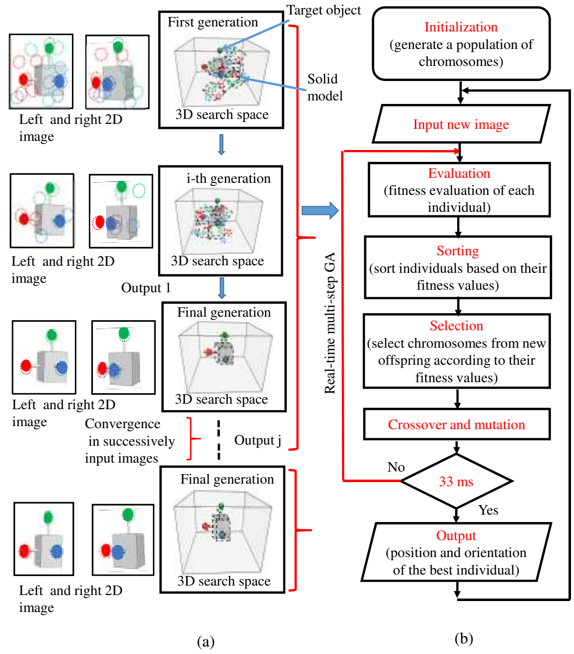

Fig. 5. Flowchart of the real-time multi-step GA: (a) the recognition process of the target pose is evaluated in 2D, convergence occurs in 3D (b) the flowchart of the RM-GA, the best solution is evaluated within 33 ms through the GA process.

elitism preservation strategy of GA, which preserves a small portion of the fittest chromosomes is copied without changing into the next generation. Finally, the best pose of the individual can be made to present the real target's pose.

An individual of GA populations is presented by six parameters (x, y, z, $\epsilon_1, \epsilon_2, \epsilon_3$) as the pose of the target object as shown in Fig. 6. The former 36 bits (12 bits for each x, y, z) represents the position coordinate of the three-dimensional model of the gene. The remaining 36 bits (12 bits for each $\epsilon_1, \epsilon_2, \epsilon_3$) describes the orientation defined by a quaternion. For the next input, a new video image is used. Selection of evolution time, population size, the probability of selection rate and the probability of mutation rate will be discussed in the next section. The effectiveness of the GA was demonstrated in a previous study on visual servoing for catching fish using a GA search [16].

3 RESULTS AND DISCUSSION

3.1 Experiment Environment

The static environment is defined as the environment which doesn't contain any moving objects while the dynamic is the environment which has dynamic moving objects (i.e., moving target object and moving robots). In dynamic environment, square and fair comparisons are very difficult for convergence time because the vehicle motion or marker motion can disturb the performance of GA convergence. There-

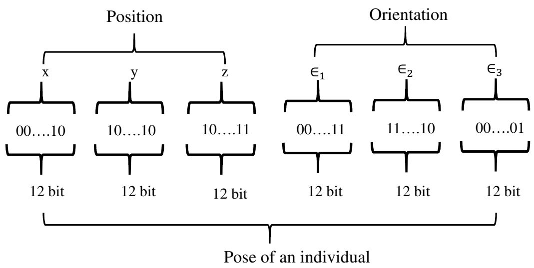

Fig. 6. An individual of GA population: 12 bits for each x, y, z represents the position coordinate of the three dimensional model of the gene and 12 bits for each $\epsilon_1, \epsilon_2, \epsilon_3$ describes the orientation defined by a quaternion.

fore, the performance of GA convergence is analyzed in static environment which is the vehicle and the marker is fixed in position.

The layout of the experiment environment is shown in Fig. 7. The frame is designed to set up firmly the experimental devices. The two cameras (imaging element CCD, 380,000 pixels, signal system NTSC, minimum Illumination 0.8[1X], without zoom) were aligned horizontally and vertically on the frame and these are used to perform a three-dimensional object recognition. A pool (length 2870 mm × width 2010 mm × height 1000 mm) is used as an experimental tank which was filled with clear water. The horizontal distance between two cameras is 178 mm which is the same length of the ROV's camera distance. According to the search area, as shown in Fig. 3, the distance between the 3D marker and the two cameras was prepared for a fixed setting 415 mm. The distance between center of the 3D marker and the bottom of the tank is 120 mm. The distance between from the center of the camera to the bottom of the tank is 130 mm. Power supply and transmission of the control signal from the PC is made through a tether cable. In this experiment, the dynamic image is used in the static environment. The specifications of the PC used for 3D pose estimation are Intel Core(TM) i7-3517UE CPU @ 1.70GHz, RAM 4096 MB, system type 64 bits. Two interfacing boards, PCI 5523, are installed in the PC to receive images from the two cameras.

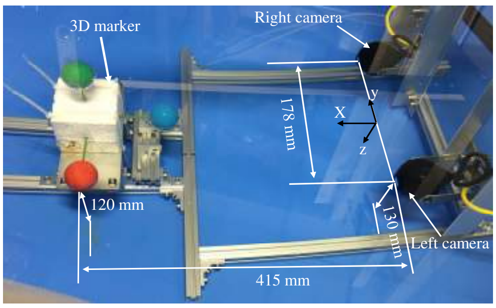

Fig. 7. Layout of the experimental devices using 3D marker and the two cameras.

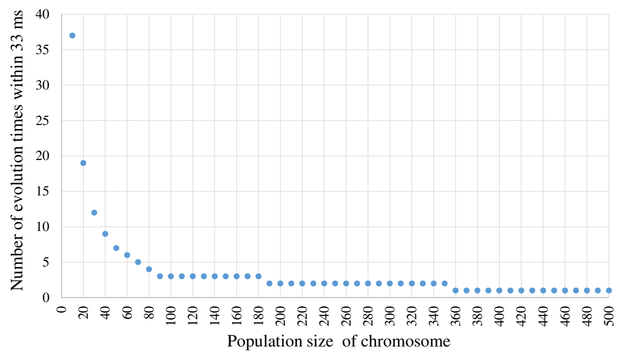

Fig. 8. Number of evolution times of RM-GA within 33 ms based on population size of chromosome.

3.2 Number of Evolution Times based on Population Sizes of Chromosome

Firstly, how many evolution times will be generated within 33 ms based on different population sizes of chromosome were analyzed. Figure 8 shows the number of evolution times based on different population sizes of chromosome from 10 to 500. The horizontal axis is the population sizes of chromosome and the vertical axis is the number of evolution times within 33 ms . The use of correct population size is an important factor for successful GA applications. According to the graph, the more the population sizes , the less the number of evolution times within 33 ms. The maximum number of evolution times are 37 within 33 ms for population size 10 and the minimum evolution time is 1 for population sizes of chromosome from 360 to 500 for real-time performance. In this case, we have to chose the optimum number of population size with a reasonable evolution times within 33 ms for real-time performance. Based on the experimental results of the number of evolution times, the convergence performance of real-time multi-step GA was analyzed by using dynamic image in the next section.

3.3 Convergence Performance of RM-GA Using Dynamic Images

The convergence performance of RM-GA was analyzed with different population sizes of chromosome, the probability of selection and the probability of mutation using the dynamic images. The GA recognition process for the dynamic images is that the GA is applied with respect to the number of evolution times within 33 ms to the new images. Genetic parameters namely as selection, crossover, mutation and population size are the key factors to obtain the optimum accuracy of the system. These parameters are considered as primary parameters. In every generation of GA process, selection, crossover and mutation operators are conducted to evolve the best individual towards the true pose of the target object (x, y, z, $\epsilon_1, \epsilon_2, \epsilon_3$) within 33 ms.

The convergence performance of RM-GA was analyzed based on the time for convergence of the fitness value of 0.6 for the cases where the different population sizes (10, 20, 40, 60, 80) and for each of these populations, the selection rate of (0.2, 0.4, 0.6 and 0.8) and mutation rate of (0.05, 0.1, 0.15). Therefore, totally 60 combinations of GA parameters were conducted. In this experiment, our criteria is to have the fitness value have 0.6 or more for good recognition performance of GA. This threshold fitness value 0.6 was defined experimentally. The different evolution times are used with respect to the population sizes as shown in Fig. 8. We did not described the detailed experiment figures of 60 combinations in this paper. The best convergence time for each population sizes was chosen from 60 combinations of GA parameters as shown in Fig. 9. These time are indicated with "A"-"E" for each population size with different selection rates and mutation rates. The vertical axis is the time domain in seconds and the horizontal axis is the different population sizes. The summary results of the best parameters and convergence time from 60 combinations of GA parameters that indicated with "A"-"E" are shown in Table 1.

There are three different lines in Fig. 9 such as ①, ②,and ③ which have different selection rates and mutation rates. In the first line ①, the population size 60, the selection rates 0.4 and mutation rate 0.1 has the slowest convergence time, it can converge to the real solution above 0.5 s as shown in Fig. 9. Among the different population sizes in line ①, the convergence time "B" for the population sizes 20 with selection rate 0.4 and mutation rate 0.1 can convergence to the solution within 0.156 s. In line ②, the selection rates and the mutation rates of "C", "D" and "E" are same but the population size is different. Among them, the fastest convergence time can be seen in population size 40, selection rate 0.6 and mutation rate 0.1 (see "C"). In a pare of selection rate 0.8 and mutation rate 0.15 ③, the convergence time is nearly same in different population sizes. But, the population size 10 can converge to the solution faster than the other population sizes within 0.187 s.

According to the results, there is a different times for convergence for fitness value of 0.6 in the case of different population sizes 20, 40, 60, and 80, selection rates 0.2,0.4, 0.6, 0.8 and mutation rates 0.05, 0.1, 0.15. Among them, the fastest time for convergence of GA recognition can be seen in population size 40, selection rate 0.6 and mutation rate 0.1, GA can converge to the solution within 0.125 s than the other selection rates and mutation rates as shown in Fig. 9.

Considering the effect of changing the selection and mutation rate for population size 10, 20, 40, 60 and 80, the effect of the mutation is important because it supports to improve new generation in the solution space. The effect of the mutation is noticed in the case of mutation rate 0.05 in different population sizes, there was no enough fitness value

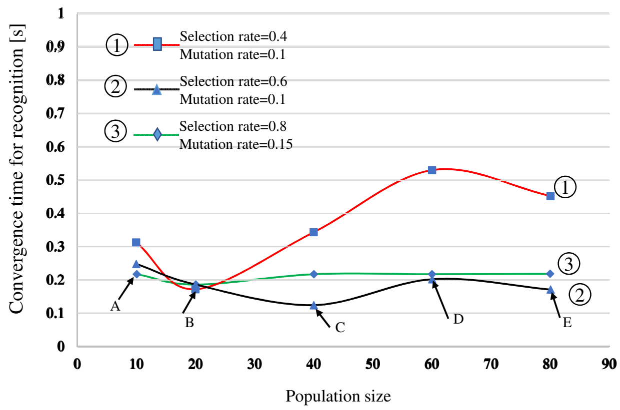

Fig. 9. Comparison of convergence time for the fitness value of 0.6 with parameters variation : population size= 10, 20, 40, 60, 80, selection rate= 0.4, 0.6, 0.8, mutation rate= 0.1,0.15.

to converge to the solution during the short time. It means that there is a slow convergence time for solutions and the small mutation rate does not bring many new and better solutions. It is obvious that the probability of mutation 0.1 is better mutation rate in most of the population size to be close to the optimal solution than the probability of mutation rate 0.15. Even though the fitness value of the different population sizes are maintained above 0.6, the time for convergence of population size 40, selection rate 0.6, mutation rate 0.1 can converge rapidly to the solution within 0.125 s in GA evolution process for real-time performance. The optimum parameters of RM-GA to perform the real-time pose estimation are shown in Table2.

Table 1. Summary results of the best parameters and convergence time in each population from 60 combination of GA parameters.

No.	Population size	Selection rate(%)	Mutation rate(%)	Convergen -ce time [s]
A	10	0.8	0.15	0.187
B	20	0.4	0.1	0.156
C	40	0.6	0.1	0.125
D	60	0.6	0.1	0.203
E	80	0.6	0.1	0.172

4 CONCLUSION

In the present paper, performance analyses and optimization in real-time pose estimation for the underwater vehicle

Table 2. Best parameters for GA.

Number of genes	40
Search area [mm]	$[x, y, z] = \pm 400 \pm 200 \pm 400$
Selection rate [%]	60
crossover probability	Two points
Mutation rate [%]	10
Number of gene evolutions [times/33 ms]	9
Control Period [ms]	33
Evolution strategy	Elitism preservation

by using a 3D marker and dual-eye cameras is presented. The optimum parameters of RM-GA were selected for real-time pose estimation based on the time for convergence of RM-GA using dynamic images with static environment. It can be confirmed that the proposed system can converge to the real solution in RM-GA evolution process using dynamic images. The real battery recharging experiment was conducted in the sea by using the real-time 3D pose tracking system in the future.

REFERENCES

[1] Eustice, R.M., Pizarro, O. and Singh, H.,“ Visually augmented navigation for autonomous underwater vehicles,” IEEE Journal of Oceanic Engineering, 33(2), pp.103-122, 2008.

[2] Ghosh, S., Ray, R., Vadali, SR., Shome, SN., Nandy, S,“ Reliable pose estimation of underwater dock using single camera: a scene invariant approach,” Machine Vision and Applications, 27(2), pp.221-36, 2016.

[3] Kume, A., Maki, T., Sakamaki, T. and Ura, T.,“A method for obtaining high-coverage 3D images of rough seafloor using AUV-real-time quality evaluation and path-planning,” Journal of Robotics and Mechatronics, 25(2), pp.364-374, 2013.

[4] Guo, J.,“Mooring cable tracking using active vision for a biomimetic autonomous underwater vehicle,” Journal of marine science and technology, 13(2), pp.147-15, 2008.

[5] Myint, M., Yonemori, K., Yanou, A., Minami, M. and 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),pp. 989-994, 2015.

[6] Myint, M., Yonemori, K., Yanou, A., Lwin, K.N., Minami, M. and Ishiyama, S.,“Visual-based deep sea docking simulation of underwater vehicle using dual-eyes cameras with lighting adaptation,” Proceedings of OCEAN 2016-Shanghai, pp.1-8, 2016.

[7] Myint, M., Yonemori, K., Yanou, A., Lwin, K.N., Mukada, N. and Minami, M.,“Dual-eyes visual-based sea docking for sea Bottom battery recharging,” OCEAN, 2016.

[8] Lwin, K.N., Yonemori, K., Myint, M., Yanou, A. and Minami, M.,“Autonomous docking experiment in the sea for visual-servo type underwater vehicle using three-dimensional marker and dual-eyes cameras,” In Society of Instrument and Control Engineers of Japan (SICE), 2016 55th Annual Conference ,pp 1359-1365, IEEE ,2016.

[9] Myint, M., Yonemori, K., Yanou, A., Lwin, K.N., Minami, M. and Ishiyama, S.,“ Visual servoing for underwater vehicle using dual-eyes evolutionary real-time pose tracking,” Journal of Robotics and Mechatronics, 28(4), pp.543-558, 2016.

[10] Lwin, K.N., Yonemori, K., Myint, M., Mukada, N., Minami, M., Yanou, A. and Matsuno, T.,“Performance analyses and optimization of real-time multi-step GA for visual-servoing based underwater vehicle,” Techno-Ocean 2016, IEEE 2016.

[11] Rexhepi, A., Maxhuni, A. and Dika, A., “Analysis of the impact of parameters values on the Genetic Algorithm for TSP,” International Journal of Computer Science Issues, 10(1), pp.158-164, 2013.

[12] Boyabatli, O. and Sabuncuoglu, I.,“Parameter selection in genetic algorithms,” Journal of Systemics, Cybernetics and Informatics, 4(2), p.78, 2004.

[13] Tabassum, M. and Mathew, K.,“A genetic algorithm analysis towards optimization solutions,” International Journal of Digital Information and Wireless Communications (IJDIWC), 4(1), pp.124-142, 2014.”

[14] Myint, M., Yonemori, K., Lwin, K N., Yanou, A., Minami, M., J Intell Robot Syst (2017), DOI 10.1007/s10846-017-0703-6.

[15] Minami, M., Agbanhan, J. and Asakura, T., In Soft Computing in Measurement and Information Acquisition, Evolutionary scene recognition and simultaneous position/orientation detection, pp.178-207. Springer Berlin Heidelberg, 2003.

[16] Suzuki, H. and Minami, M., “Visual servoing to catch fish using global/local GA search,” IEEE/ASME Transactions on Mechatronics, 10(3), pp.352-357, 2005.