Proceedings of 2013 IEEE
International Conference on Mechatronics and Automation
August 4 - 7, Takamatsu, Japan

Analyses on On-line Evolutionary Optimization Performance for Pose Tracking while Eye-vergence Visual Servoing

Kenta Nishimura, Sen Hou, Koichi Maeda, Mamoru Minami and Akira Yanou
Graduate School of Natural Science and Technology
Okayama University
Okayama, Japan 700-8530
{*nishimurak, hou, maeda, minami, yanou*}@*suri.sys.okayama-u.ac.jp*

***Abstract*— In this research, Genetic algorithm (GA) is used as pose-tracking method "1-step GA," to solve on-line optimization problem for 3-D visual servoing. A correlation function between the target object projected in camera flame and model defined in the control computer is used for a fitness function to be optimized by the 1-step GA. The optimization process for real-time object tracking has been examined on a view point of realizing real-time pose estimation utilizing eye-vergence function. We have confirmed that the 1-step GA optimization method together with eye-vergence correlation fitness function worked cooperatively and how the eye-vergence helped real-time optimization processes in time-domain during visual servoing.**

***Index Terms*— Visual servoing, Eye-vergence, GA.**

I. Introduction

Visual servoing is a control method of robot's motion through visual information in the feedback loop, which is obtained from visual cameras [1]-[4]. Some methods have already been proposed to improve observation abilities, by using stereo cameras [5], multiple cameras [6], and two cameras; with one fixed on the end-effector, and the other done in the workspace [7]. These methods obtain different views to observe the object by increasing the number of cameras, leaving the system less adaptive for changing environment.

The problem to find position/orientation , i. e, poseof an object relatively based on hand-eye camera flame can be transposed to the optimization problem of correlation function. In this research, we use Genetic Algorithm(GA) to get the maximum correlation value within less than video rate, "1-step GA" algorithm that is on-line estimation method [8].

The pose of an actual object are expressed with the pose of the coordinate system Σ_M being fixed to the object, and the object's pose estimated by 1-step GA is expressed as $\Sigma_{\hat{M}}$. It is natural there should always exist an error, then ${}^{M}\boldsymbol{T}_{\hat{M}}$ 4 $\times$ 4 homogeneous Matrix representing the pose relation between the actual object Σ_M and the detected one $\Sigma_{\hat{M}}$ is usually not identity matrix. It is necessary to decrease this estimation error ${}^{M}\boldsymbol{T}_{\hat{M}}$. Then, we compared the pose of the object calculated by full search of fitness value with the pose estimated by 1-step GA. The fitness function means correlation between projected target object to left and right camera image and 3-D target model predefined in the computer.

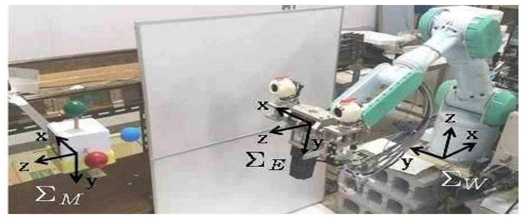

Fig. 1. Object and the visual-servoing system

By a previous work, it has been shown that the tracking performance could be improved by using eye-vergence system. However, the argument is not based on what kind of preferable influence the eye-vergence system can affect on-line optimization process of 1-step GA. In this research, visual servoing experiment was performed on the point of frequency response as shown in Fig. 1, where target object represented by Σ_M was oscillated by sinusoidal function and end-effector and eye-vergence systems are both controlled to keep desired pose relations that it constant desired pose relations. We clarify how the eye-vergence system can improve the dynamical performance of 1-step GA to tracking the moving target object, through comparing time-varying distribution of correlation function and the pose found by 1-step GA in the time-varying correlation distribution. GA and a object through the optimizing action of GA's gene during visual servoing with eye-vergence.

II. Object Estimation Method

Visual servoing system used in this research needs real-time estimation. Dynamic image given by video camera constitutes of still pictures input successively in order by a time series. Tracking an object in video is realized by estimating the object in a still picture continuously within a video rate (33 [ms]). Therefore, this section explains the outline of the estimation technique for the still picture of one sheet.

The pose of 3-D model $\boldsymbol{\phi} = (x, y, z, \epsilon_1, \epsilon_2, \epsilon_3)$ (ϵ is orientation variable of quaternion) is determined by the gene of GA. The 2-D models are obtained by projecting the 3-D model onto either side of left and right camera flame. This 2-D model is evaluated by calculating a fitness function defined by correlation between the 2-D model and

978-1-4673-5560-5/13/$31.00 ©2013 IEEE

the input image. And when the pose ϕ of the solid model coincides with the pose of the object, the correlation function has been designed to have maximum value. Therefore, the pose estimation problem of the object is convertible into optimization problem to find ϕ maximizing the correlation function, i.e., fitness function [8]. There are various methods to solve optimization problem that the maximum of a given distribution is searched for and discovered. Simplest and easy method is full search method. Although full search method can discover the maximum by calculating all the values of the function and can certainly discover the maximum value, it is inefficient. That is, it has the fault of spending much computation time. It is important for real-time pose tracking that needs calculation time to be very short–within video rate, 33[ms]–in this research, GA has been applied for the pose tracking in a form of "1-step GA" [8]. If GA's gene should give high correlation function value,F it could be assumed that the gene's pose almost coincides with the actual real target object's pose. Thereby, the pose ϕ of the object are measurable in real-time.

A. Model-based Matching Method

In this section, a model-based matching method was presented. The images input from a right-and-left video cameras are composed by hue value ranging from 0 to 255. The searching model is shown as Fig. 2. The model constituted inside spaces $S_{R,in}(\phi)$ and $S_{L,in}(\phi)$, and outsize spaces $S_{R,out}(\phi)$ and $S_{L,out}(\phi)$, in order to evaluate difference of hue value between the object and the circumference. The hue value of right image at the position ${}^{IR}\boldsymbol{r}_i$ is expressed as $p({}^{IR}\boldsymbol{r}_i)$, and the hue value of left image at the position ${}^{IL}\boldsymbol{r}_i$ is expressed as $p({}^{IL}\boldsymbol{r}_i)$. Equation (1) shows the fitness function that calculate the correlation function between the search model and image.

$$
\begin{aligned}
F(\phi) = \Bigg\{ & \Bigg(\sum_{{}^{IR}\boldsymbol{r}_i \in S_{R,in}(\phi)} p({}^{IR}\boldsymbol{r}_i) - \sum_{{}^{IR}\boldsymbol{r}_i \in S_{R,out}(\phi)} p({}^{IR}\boldsymbol{r}_i) \Bigg) \\
& + \Bigg(\sum_{{}^{IL}\boldsymbol{r}_i \in S_{L,in}(\phi)} p({}^{IL}\boldsymbol{r}_i) - \sum_{{}^{IL}\boldsymbol{r}_i \in S_{L,out}(\phi)} p({}^{IL}\boldsymbol{r}_i) \Bigg) \Bigg\} /2 \\
= & \ \{F_R(\phi) + F_L(\phi)\}/2 \qquad (1)
\end{aligned}
$$

The projected model area, $S_{p,q}(p = L, R; q = in, out)$ are all depending on the model's assumed pose that in designated by gene of GA's evolutionary processes. In the right imaging range, Eq. (1) deducts the total value in $S_{R,out}(\phi)$, from $S_{R,in}(\phi)$ and obtains the fitness value of the right image from the total value of the hue value $p({}^{IR}\boldsymbol{r}_i)$ of an input image. The left imaging range is also the same. The fitness value of the left and the right is added and an average is taken. The image of the left and the right is simultaneously evaluated using this fitness function. This fitness function composed of "in" and "out" area aimed at to determine correlation value more emphasized then calculating correlation merely by solid model. When a solid model $S_{R,in}(\phi)$ and $S_{L,in}(\phi)$

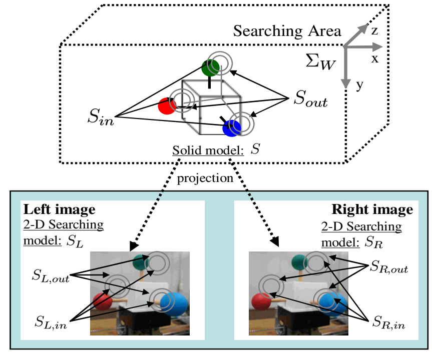

Fig. 2. Searching model

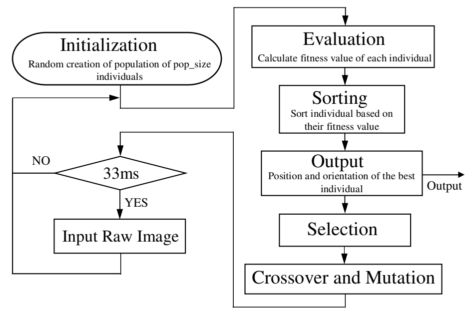

Fig. 3. 1-step GA

are cprrectly in agreement with the object, both in the right-and-left image, the object and the search model must be in agreement in 3-D pose. Though there is no gurantee that the variables to give the highest peek of the correlation function, i .e. , fitness function ?coincider? with the true pose of object, but we can make effects to realize such conditions; righting condition, shape and ?? of the target, simple backdrop, and so on. Then we assume in this paper that highest peak of the fitness function indicates that the variables to give the peak represents true pose of the target object. It is defined as $F_R(\phi) = 0$ if $F_R(\phi) \leq 0$ and $F_L(\phi) = 0$ if $F_L(\phi) \leq 0$.

B. The Optimal Solution Searching Method using GA

By using a fitness function, the problem searches for the pose of an object can be transposed to the problem which searches for the maximum of a fitness function $F(\phi)$. In this research, we use GA to get the maximum fitness value within less that video rate. Moreover, the gene information showing the position/orientation on the individual in this research is shown below.

$$
\underbrace{\overbrace{011000100111}^{t_x}}_{12bit}\underbrace{\overbrace{000011000111}^{t_y}}_{12bit}\underbrace{\overbrace{001100111101}^{t_z}}_{12bit}
$$

$$
\underbrace{\overbrace{110101001001}^{\epsilon_1}}_{12bit}\underbrace{\overbrace{000101111001}^{\epsilon_2}}_{12bit}\underbrace{\overbrace{001101111001}^{\epsilon_3}}_{12bit}
$$

The position/orientation of the individual shows the pose of the solid model in the Model-based Matching method. Top

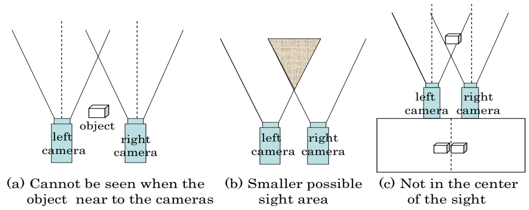

Fig. 4. Disadvantage of fixed camera system

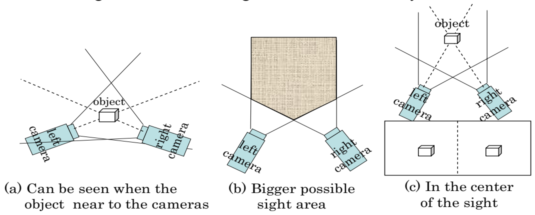

Fig. 5. Advantage of Eye-vergence system

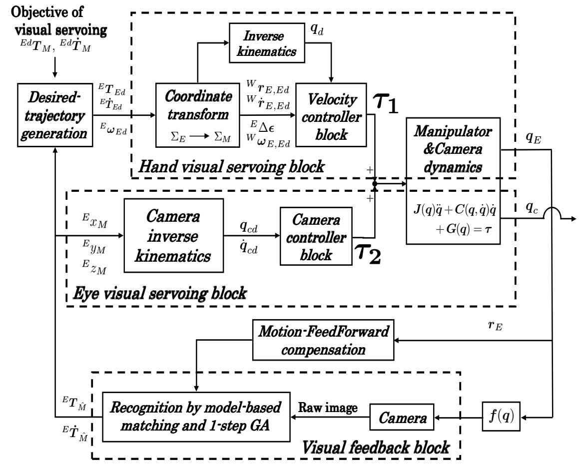

Fig. 6. Hand & Eye-Vergence Visual Servo System

36 bits with every 12 bits of this gene express the position coordinate of a solid model, and remainder 36 bits with every 12 bits expresses the orientation of the solid model, where the orientation is defined by quaternion. Bit used at this time may be reduced for searching time shortening.

Next, each individual gene get fitness value from the fitness function $F(\boldsymbol{\phi})$ using its pose information. Evolution processing is performed based on the superiority or inferiority of this value, and a set of the next generation is generated through GA's process. At this time, the pose in which fitness value was high in former generation, that is, it approaches toward the maximum neighborhood of the fitness function showing object. By repeating this processing (change of generation), GA discovers the maximum value showing the true pose of the target object.

However, normal GA needs to wait for convergence for a definite period of time. When a fitness function shows a value high enough and estimation of object is completed often more time has passed by, then, there is a possibility that the surrounding situation is changing a lot,that means target object exists at very different pose. Therefore we use 1-step GA (Fig. 3). 1-step GA is on-line estimation method [8]. While using the elitist model of the GA, the position/orientation of a target can be detect in every new image by that of the searching model given by the best individual in the population. Of the 1-step GA's evolving speed to optimize the fitness function should be faster than the target object's moving speed, the we can pose that best gene can track the moving optimal point of the fitness function where variables represent the target pose [9]. This feature happens to be favorable for real-time visual estimation.

III. Servoing System

A. Eye-vergence System

Although hand-eye composition has a deficit that servo operation may become unstable easily by vibration of the hand and time-delay of visual pose detection compared with floor-fixed camera composition, there is good point in which a viewpoint can be chosen accommodatively. In this paper, the visual servoing of hand-eye composition with two cameras is considered. If it assumes that object form is known, it is possible to measure six variables of a position/orientation also by a simple eye. However, it is known well that there is a problem in the precision for measurement of the distance between camera and objects, and compound eye composition is used for it here.

A fixed-hand-eye system has some disadvantages, making the observing ability deteriorated depending on the relative geometry of the camera and the target. Such as: the robot cannot observe the object well when it is near the cameras (Fig. 4 (a)), small intersection of the possible sight space of the two cameras (Fig. 4 (b)), and the image of the object cannot appear in the center of both cameras, so we could not get clear image information of target and its periphery, reducing the pose measurement accuracy (Fig. 4 (c)). To solve the problems above, in this research, we have chosen Eye-Vergence system that gives the cameras an ability to rotate themselves to project a target at center of the images.

Thus it is possible to change the pose of the cameras in order to observe the object better, as it is shown in Fig. 5, enhancing the measurement accuracy in trigonometric calculation and avoiding peripheral distortion of camera lens by observing target at the center of lens.

B. Hand & Eye Visual Servoing Controller

The block diagram of our proposed hand & eye-vergence visual servoing controller is shown in Fig. 6. Each joint angle of a manipulator is set to $\boldsymbol{q}_E = [q_1, \cdots, q_7]$, pan tilt angle of a camera is set to $\boldsymbol{q}_c = [q_8, q_9, q_{10}]$, the desired angle of each link is set to $\boldsymbol{q_d}$. The hardware control system of the velocity-based servo system of Mitsubishi Heavy Industries, Ltd PA10 is expressed as

$$\boldsymbol{\tau} = \boldsymbol{K}_{SP}(\boldsymbol{q}_d - \boldsymbol{q}) + \boldsymbol{K}_{SD}(\dot{\boldsymbol{q}}_d - \dot{\boldsymbol{q}}) \tag{2}$$

where $\boldsymbol{K}_{SP}$ and $\boldsymbol{K}_{SD}$ are symmetric positive definite matrices to determine PD gain. Moreover, the target angle of a

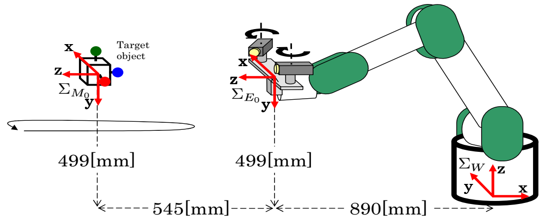

Fig. 7. Coordinate system

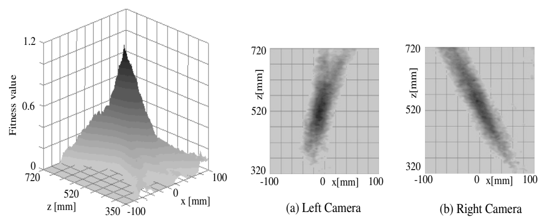

Fig. 8. Fitness value of the left camera and right camera

camera is set with $\boldsymbol{q}_{cd}$ and the error of an angle is defined as

$$\Delta \boldsymbol{q}_c = \boldsymbol{q}_{cd} - \boldsymbol{q}_c. \tag{3}$$

The controller of eye-visual servoing is given by

$$\dot{\boldsymbol{q}}_{cd} = \boldsymbol{K}_{P_c} \Delta \boldsymbol{q}_c \tag{4}$$

where $\boldsymbol{K}_{P_c}$ are positive control gain. The visual servoing is performed using these controllers [10].

IV. Pose Tracking by 1-step GA

In this part, analyses of real-time pose tracking performance of 1-step GA are extended using experiments of frequency responses.

A. Experiment Condition

The initial hand pose is defined as Σ_{E_0}, and the initial object pose is defined as Σ_{M_0}. The homogeneous transformation matrix from Σ_W to Σ_{E_0} and from Σ_W to Σ_{M_0} are:

$$^{W}\boldsymbol{T}_{M_0} = \begin{bmatrix} 0 & 0 & -1 & -1435[mm] \\ 1 & 0 & 0 & 0[mm] \\ 0 & -1 & 0 & 499[mm] \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

$$^{W}\boldsymbol{T}_{E_0} = \begin{bmatrix} 0 & 0 & -1 & -890[mm] \\ 1 & 0 & 0 & 0[mm] \\ 0 & -1 & 0 & 499[mm] \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6}$$

where the above relations between Σ_{M_0}, Σ_{E_0}, Σ_W are depicted in Fig. 7. The target object moves according to the following time function as:

$$^{M_0}z_M(t) = -150 + 150\cos(\omega t)[mm]. \tag{7}$$

The relation between the object and the desired end-effector is set as:

$$^{Ed}\boldsymbol{\psi}_M = [0, -90[mm], 545[mm], 0, 0, 0]. \tag{8}$$

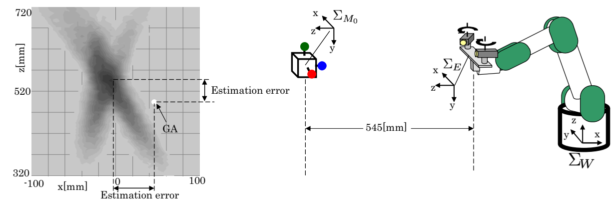

Fig. 9. The relation of the position of a target object and a hand (A target relation)

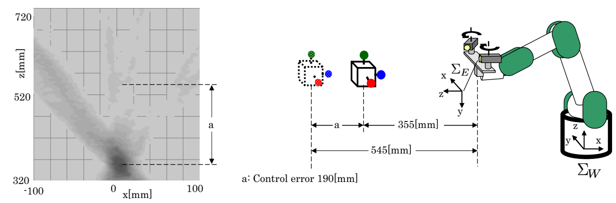

Fig. 10. The relation of the position of a target object and a hand (nearer than target relations)

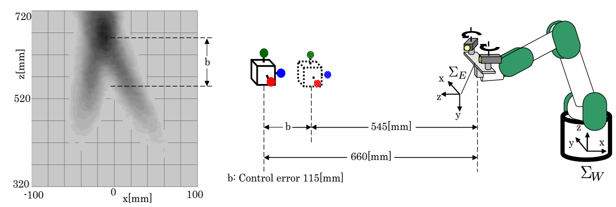

Fig. 11. The relation of the position of a target object and a hand (further than target relations)

Since it is reasonable that the target object is incorrectly estimated when fitness value was low, visual servoing motion will be stopped in consideration of the safety during experiments on condition that fitness value decreased to 0.1 or below. The fitness function is made of correlation function of target object in the camera scene and the model predefined in the control computer, then the correlation function value presents convicting level of the detected peak by 1-step GA expressing the real target moving arround the robot in the environment. Table I shows information set on GA used in this experiment.

B. Experiment Results

In this experiment, position (x, y, z coordinates) was unknown and a true value was given to posture. In this case, the variables in genes of the 1-step-GA pose-tracking-system are only x, y, z, positions and the variables for orientation defined by quaternion are fixed to the right value and not updated by 1-step GA. The desired relation between the target

TABLE I

GA

Number of individual	20
Length of gene	72[bit]
Selection rate	0.40
Mutation	0.10
Search range (Σ_E)	$-200 \leq t_x \leq 200$ $-195 \leq t_y \leq 5$ $350 \leq t_z \leq 750$ $-0.3 \leq \epsilon_1, \epsilon_2, \epsilon_3 \leq 0.3$

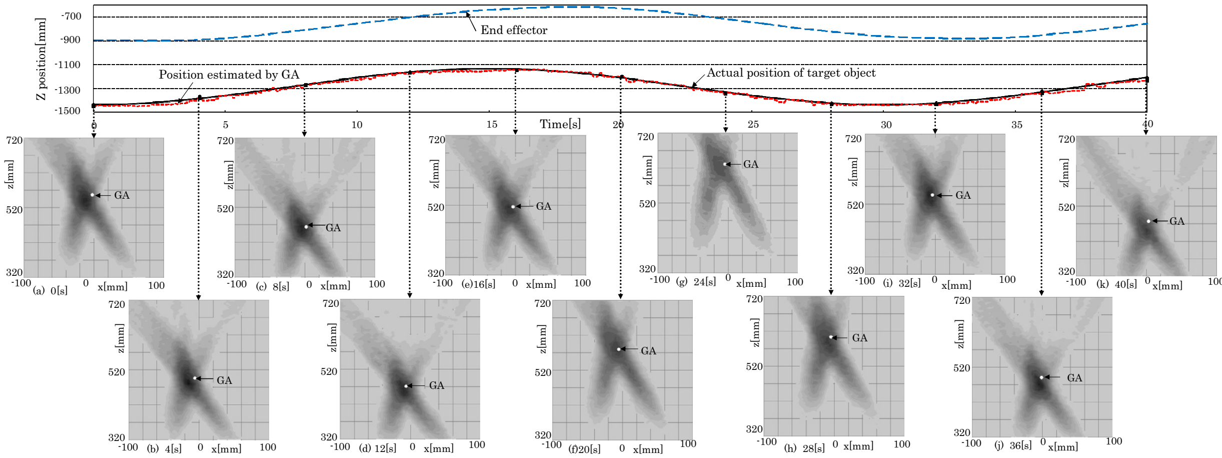

Fig. 12. Relation between z-x fitness value, end effector ,actual position of target object and position estimated by GA in $\omega = 0.209$

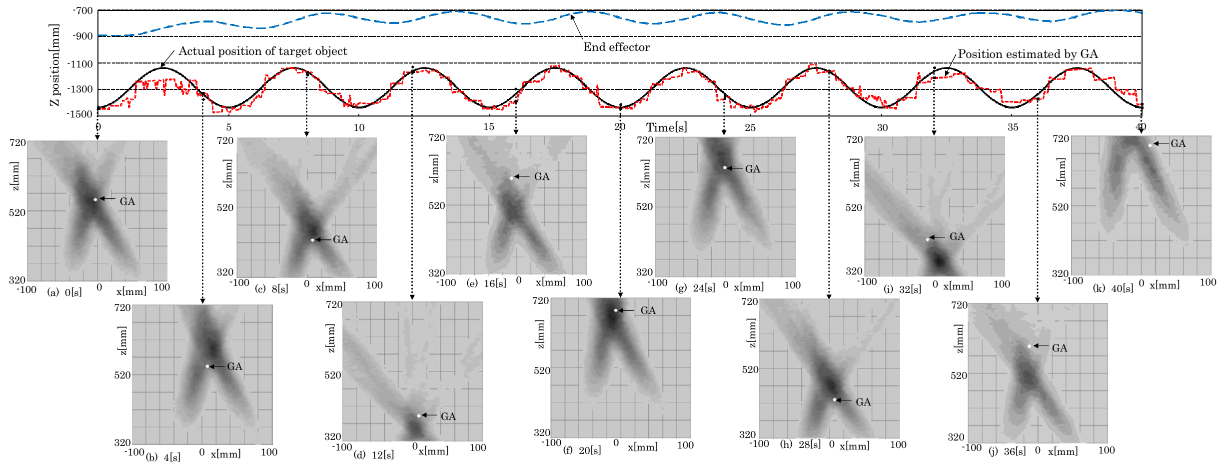

Fig. 13. Relation between z-x fitness value, end effector ,actual position of target object and position estimated by GA in $\omega = 1.256$

object and the desired end-effector is set as Eq.(8). In this paper, with the two kinds of ω- $\omega = 0.209$ (Period: $T = 30[s]$) and $\omega = 1.256$ (Period: $T = 5[s]$)- the visual servoing experiments were performed, and distribution of the fitness value is calculated against all possible positions of x-z plane, where the value of x-z position that gives the highest peak means the target's x-z position. Then the difference of the pose of GA and the variable to give the highest peak shows the detection error of pose by 1-step GA.

Fig. 8 shows the fitness value distribution in x-z plane of a right-and-left camera, where densely dotted area means high level of fitness function and Figs. 9- 11 shows the results of full search by the relation between a hand and an object. In Fig. 9, the time varying fitness function distribution combined by both left and right camera is depicted, and the position of the white circle with a notation "GA" represents that the 1-step GA has found the indicated position represents most possible position.

Therefore the deviation between the peak (target real position) and the GA's position means on-line tracking error of object in 3-D space as shown in Fig. 9.

Fig. 10 depicts the situation that the object place nearer than prescribed desired position relation of the object and the hand, and Fig. 11 shown vice versa. The circular frequency of the target object of Fig. 12 is $\omega = 0.209$, and the one of the target object of Fig. 13 is $\omega = 1.256$. It is the result obtained by full search in the plane of z-x in Σ_W for fitness value once every 4 seconds. GA's gene gotten highest fitness value is displayed on these results by the white circle. The position of the depth direction (z-direction) of the hand in Σ_W, the estimation results and the relation of the position of the depth direction of an actual object are shown in the graph. When period of the target's motion is $T = 30[s]$, it can be seen that the end-effector tracks the target with some place delay. but the 1-step GA tracks correctly the highest peak changing its position in real time. Even though the condiction $T = 5[s]$ in Fig. 13 makes the correct tracking of end-effector difficult, 1-step GA maintains real-time tracking of target through real-time optimization of time-varying fitness distributions. And the relation with full search results is shown. When fitness value is so high, it is deep-colored; inversely, it is light-colored if fitness value is low.

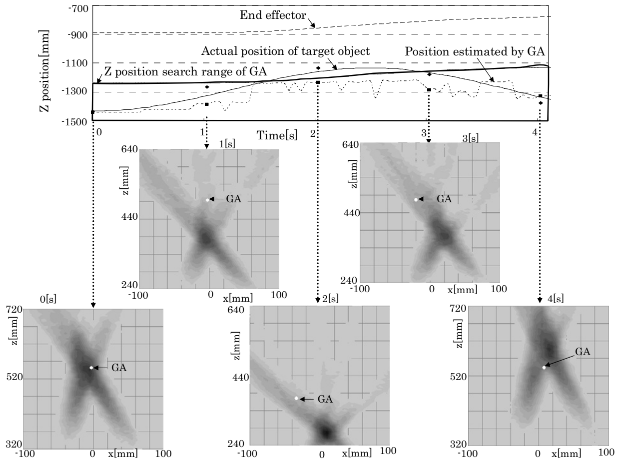

Fig. 14. Relation between z-x fitness value, end effector ,actual position of target object and position estimated by GA in $\omega = 1.256$ (b)

C. Discussion

Full search results of fitness value show that fitness value is distributed in the shape of x centering on a deep-colored point (fitness value is high). This visual servoing is compound eye composition, and it is asking for fitness value, using simultaneously the image obtained from the right-and-left camera (Fig. 8). thereby, GA becomes easy to discover an object.

Fig. 9 - Fig. 11 is a figure showing the relation of the result of full search, and the position of a hand and a target object. In Fig. 9, it is a result of full search when the hand and the target object are maintaining target relations. The error of the highest point of fitness value and GA is defined as estimation error. Fig. 10 shows full search results when the relation between a target object and a hand is nearer than target relations. Fig. 11 shows full search results when the relation between a target object and a hand is further than target relations. In these cases, the control error exists.

Fig. 12 show a result when the speed of a target object is slow. GA is in a deep-colored point (fitness value is high). That is, a target object is discovered, and it follows, without being behind. Fig. 13 show a result when the speed of a target object is quick. GA is in a light-colored point (fitness value is low). That is, it follows later than a target object. Moreover, the position/orientation of the target object in full search results of fitness value, especially, the relation between the position of the direction of z and a hand cannot be maintaining the interval of 545[mm], which is the preset value shown in the Eq.(8). Fig. 13 (d), (i) show the relation in which the distance of the direction of z of a target object and a hand is closer than a preset value (Fig. 10). Fig. 13 (f), (k) show the relation in which the distance of the direction of z of a target object and a hand is further than a preset value (Fig. 11).Because the torque of a hand is large, a phase delay has arisen to the quick motion of a target object.

Like Fig. 10, even when a hand is in a near position to a target object, the maximum of the fitness value obtained by full search shows a high value like Fig. 13 (d), (i). This shows the advantage of the Eye-vergence system in Fig. 5 (a).

In 0 second to 4 seconds in Fig. 13, the error of the estimation result by the position of an actual object and GA is large. Then, full search of fitness value was performed for 0 second to 4 seconds for every second. A result is shown in Fig.14. In Fig.13, when the estimation error of GA and a target object was large, distribution of fitness value was thin on the whole, and tends to find a target object was not able to be said. However, Fig.14 showed that GA may not be able to discover the target object, even if distribution of fitness value was deep and was in the state of being easy to find a target object. It is thought that it has slow flattery of a hand and it starts since the object is located out of the searching area of GA.

V. Conclusion

In this paper, the visual servoing was performed in the experiment of frequency response, and the action of GA gene in a visual servoing was shown. The real-time estimation tracking error has been grasped by clarifying relationship of GA and an object. 1-step GA became easy to calculate the optimal solution with eye-vergence system. Full search results were able to shown one of the characteristics of eye-vergence system. There was a case that the object was out of the search range of GA. As a future subject, it is necessary to consider change of the searching area of GA.

References

[1] S.Hutchinson, G.Hager, and P.Corke, “A Tutorial on Visual Servo Control”, IEEE Trans. on Robotics and Automation, vol. 12, no. 5, pp. 651-670, 1996.

[2] P.Y.Oh, and P.K.Allen, “Visual Servoing by Partitioning Degrees of Freedom”, IEEE Trans. on Robotics and Automation, vol. 17, no. 1, pp. 1-17, 2001.

[3] E.Malis, F.Chaumentte and S.Boudet, “2-1/2-D Visual Servoing”, IEEE Trans. on Robotics and Automation, vol. 15, no. 2, pp. 238-250, 1999.

[4] P.K.Allen, A.Timchenko, B.Yoshimi, and P.Michelman, “Automated Tracking and Grasping of a Moving object with a Robotic Hand-Eye System”, IEEE Trans. on Robotics and Automation, vol. 9, no. 2, pp. 152-165, 1993.

[5] W. Song, M. Minami, Y. Mae and S. Aoyagi, “On-line Evolutionary Head Pose Measurement by Feedforward Stereo Model Matching”, IEEE Int. Conf. on Robotics and Automation (ICRA), pp.4394-4400, 2007.

[6] J. Stavnitzky, D. Capson, ”Mutiple Camera Model-Based 3-D Visual Servoing”, IEEE Trans. on Robotics and Automation, vol. 16, no. 6, December 2000.

[7] C. Dune, E. Marchand, C. leroux, “One Click Focus with Eye-inhand/Eye-to hand Cooperation”, IEEE Int. Conf. on Robotics and Automation (ICRA), pp.2471-2476, 2007.

[8] H. Suzuki, M. Minami, “Visual Servoing to catch fish Using Global/local GA Search”, IEEE/ASME Transactions on Mechatronics, Vol.10, Issue 3, 352-357 (2005.6).

[9] M.Minami, W.Song, “Hand-eye-motion Invariant Pose Estimation with On-line 1-step GA -3D Pose Tracking Accuracy Evaluation in Dynamic Hand-eye Oscillation”, Journal of Robotics and Mechatronics, Vol.21, No.6, pp.709-719 (2009.12)

[10] Wei. Song, M. Minami, Fujia Yu, Yanan Zhang and Akira Yanou “3-D Hand & Eye-Vergence Approaching Visual Servoing with Lyapunouv-Stable Pose Tracking ”, IEEE Int. Conf. on Robotics and Automation (ICRA), pp.11, 2011.