Artif Life Robotics (2018) 23:131–139
https://doi.org/10.1007/s10015-017-0408-8

ORIGINAL ARTICLE

Improved eye-vergence visual servoing system in longitudinal direction with RM-GA

Yejun Kou[1] · Hongzhi Tian[1] · Mamoru Minami[1] · Takayuki Matsuno[1]

Received: 7 June 2017 / Accepted: 30 October 2017 / Published online: 11 November 2017

Abstract Visual servoing is a control method to manipulate the motion of the robot using visual information, which aims to realize "working while watching." However, the visual servoing towards moving target with hand–eye cameras fixed at hand is inevitably affected by hand dynamical oscillation. To overcome this defect of the hand–eye fixed camera system, an eye-vergence system has been put forward, where the pose of the cameras could be rotated to observe the target object. The visual servoing controllers of hand and eye-vergence are installed independently, so that it can observe the target object at the center of camera images through eye-vergence function. In this research, genetic algorithm (GA) is used as a pose tracking method, which is called "Real-Time Multi-step GA(RM-GA)," solves online optimization problems for 3D visual servoing. The performances of real-time object tracking using eye-vergence system and "RM-GA" method have been examined, and also the pose tracking accuracy has been verified.

Keywords Visual servoing · Eye-vergence · Real-time Multi-step GA · Longitudinal · Optimization

This work was presented in part at the 22ndInternational Symposium on Artificial Life and Robotics, Beppu, Oita, January 19–21, 2017.

✉ Yejun Kou
ptlg9dvi@s.okayama-u.ac.jp

1 Graduate School of Natural Science and Technology, Okayama University, 1-1 Tsushima-naka, 3-Chome Kita-ku, Okayama 700-8530, Japan

1 Introduction

Visual servoing is a control method to control the motion of the robot. By incorporating visual information obtained from visual sensor [1–4] with the feedback loop, visual servoing is expected to enable the robot to adapt the changing environment or unknown environment. Some methods have been proposed already to improve the observation abilities of the robot, for instance using stereo cameras [5], multiple cameras [6], and two cameras: with one fixed on the end-effector, and the other one fixed in the workspace [7]. These methods obtained different views to observe an object by simply increasing the number of cameras, leaving the system less adaptive for changing environment.

A fixed-hand–eye system has some disadvantages, making the observing ability deteriorated because of the relative geometry of the camera and the target. The robot cannot observe the object well in such occasions: (1) when it is near the cameras, (2) with small intersection of the possible sight space of the two cameras, and (3) 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, by lens aberration, reducing the pose measurement accuracy.

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. To the problem to find position/orientation, i.e., pose of an object relatively based on hand–eye camera frame can be transposed to the optimization problem of correlation function. In this research, we use Genetic Algorithm (GA) to solve the optimization problem to find the maximum correlation through dynamic images periodically input with video rate, 33 [ms], called "Real-time Multi-step GA" (RM-GA) algorithm that is an on-line estimation method [8].

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

In the previous works, it has been shown that the lateral direction tracking performance could be improved using eye-vergence system [12]. However, the argument has not been based on how the performance of RM-GA affects on-line optimization process during dynamic recognition. In this research, visual servoing experiment was performed to evaluate it from the view point of frequency response and fitness function distribution that is optimized by RM-GA, whose experimental apparatus is shown in Fig. 1, where target object represented by Σ_M was oscillated by sinusoidal function along the cameras' eyesight direction, and end-effector and eye-vergence system are both controlled to keep desired constant pose relations between the hand pose and the one of target 3D marker. By comparing the results, we have clarified how the dynamical performance can be improved by RM-GA during tracking the moving target object.

2 Pose tracking method

Visual servoing system used in this research needs real-time pose estimation. Dynamic images given by video camera are made of a row of static pictures input successively. Tracking an object in video is realized by estimating the pose of the object through static pictures sequentially input with a time interval of a video rate (33 [ms]). This section explains the outline of the on-line estimation method.

In this section, a model-based matching method was presented. A 3D-ball-object is used as 3D target object, whose color and size information were known, we named it as 3D marker. The 3D marker and models are being shown

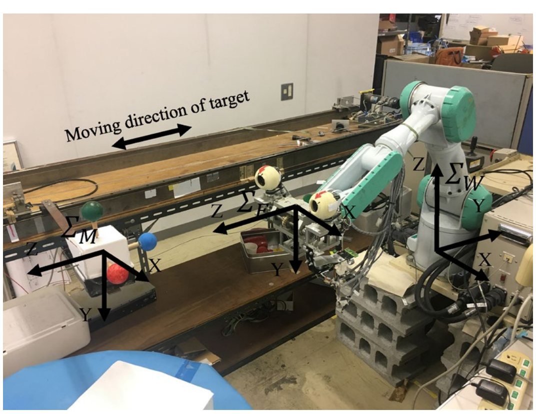


Fig. 1 Object and the visual servoing system for longitudinal tracking experiment, whose the 3D marker's motion are shown by arrow of "⟷"

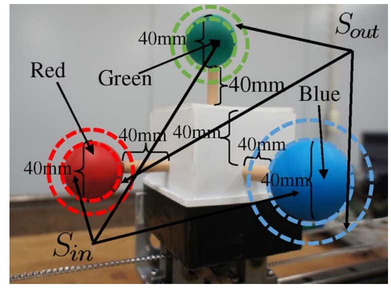


Fig. 2 3D marker and model

in Fig. 2. The pose of 3D model $\phi = (x, y, z, \epsilon_1, \epsilon_2, \epsilon_3)$ (ϵ is orientation variable of quaternion) is estimated by the gene of RM-GA in real time. In this research, the target object should was aimed to be recognized in 3D space; therefore, to describe a object in 3D space, we need at least 6 variables; in this research, the position variables of the target have been presented as x, y, z, and the orientation variables was defined as $\epsilon_1, \epsilon_2, \epsilon_3$.

Figure 3a shows searching models Σ_n without sampling points. The model to detect 3D-ball-object has the same 3D structure. The part of inner circle is named as S_{in}, and part between S_{in} and outer circle is named S_{out}. After a projection of the model to left and right images, it will be shown as (b). Then, we take the sampling points on the images like (c) and calculate the fitness $F(\phi)$.

The dotted line block named $\mathbb{R}$ in Fig. 3a means a searching space.

Through the projection transformation, S_{in} and S_{out} are projected onto the 2D coordinates of left image Σ_{IL} and right image Σ_{IR} named S_{L} and S_{R}, as shown in Fig. 3b.

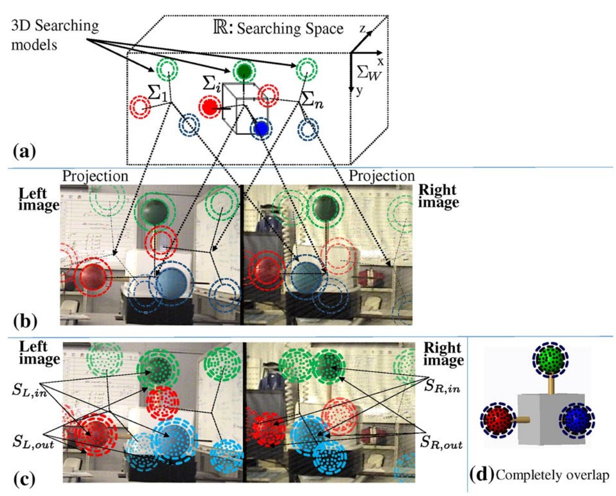


Fig. 3 Definition of a solid model without sampling points (**a**), left/right searching models after projection transformation (**b**), and taking sampling points in two images (**c**). When a model completely overlaps the object (**d**), its fitness function gets the maximum

As shown in Fig. 3d, inner portions of a model corresponding to three balls are $S_{\mathrm{in,R}}$, $S_{\mathrm{in,G}}$, and $S_{\mathrm{in,B}}$. Similarly the three outer portions are $S_{\mathrm{out,R}}$, $S_{\mathrm{out,G}}$, and $S_{\mathrm{out,B}}$. Each pair of circle and ring corresponds with a color, and three pairs of circles and rings are corresponding to red, blue, and green. Each S_{in} is composed by three concentric circles with 36 sampling points. Each S_{out} is composed by two concentric circles with 24 sampling points. On each circle, 12 sampling points are taken at an equal interval. The total number of sampling points is as follows:

$$\Sigma_{\mathrm{s}} = \Sigma_{\mathrm{sin}} + \Sigma_{\mathrm{sout}} = 3 \times (12 \times 3) + 3 \times (12 \times 2) = 180. \quad (1)$$

Hue information of HSV is used to search for the target object in the images. 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 $p({}^{IL}\boldsymbol{r}_i)$. The higher coincidence degree between a circle (inner portion) and corresponding color ball is, the higher fitness is. Conversely, the higher coincidence degree between a ring (outer portion) and the corresponding color ball is, lower fitness will be. When the searching model $\Sigma_{\hat{M}}$ completely overlaps to the target object like (d), then the fitness function gives maximum value. Eq. (2) shows the fitness function that calculates the correlation function between the search model and left and right images:

$$\begin{aligned} F(\phi) = & \left\{ \left(\left(\sum_{\substack{IR\boldsymbol{r}_i \in \\ S_{R,in}({}^{CR}\boldsymbol{\psi}_{\hat{M}})}} p({}^{IR}\boldsymbol{r}_i) - \sum_{\substack{IR\boldsymbol{r}_i \in \\ S_{R,out}({}^{CR}\boldsymbol{\psi}_{\hat{M}})}} p({}^{IR}\boldsymbol{r}_i) \right) \Big/ n_{R,in} \right. \right. \\ & \left. + \left(\sum_{\substack{IL\boldsymbol{r}_i \in \\ S_{L,in}({}^{CL}\boldsymbol{\psi}_{\hat{M}})}} p({}^{IL}\boldsymbol{r}_i) - \sum_{\substack{{}^{IL}\boldsymbol{r}_i \in \\ S_{L,out}({}^{CL}\boldsymbol{\psi}_{\hat{M}})}} p({}^{IL}\boldsymbol{r}_i) \right) \Big/ n_{L,in} \right\} \Big/ 2N \\ = & \left\{ F({}^{CR}\boldsymbol{\psi}_{\hat{M}}) + F({}^{CL}\boldsymbol{\psi}_{\hat{M}}) \right\} \Big/ 2N. \end{aligned} \quad (2)$$

Here, the N means the total number of sampling points.

2.1 Optimal solution searching method using GA

Using a fitness function, the problem that searching for the pose $\boldsymbol{\phi}$ of an object can be transposed to the problem that searches for the maximum of a fitness function $F(\boldsymbol{\phi})$. In this research, we use GA to get the maximum fitness value in the consecutively input dynamic image sequences by video rate. The gene information showing the position/orientation on the individual in this research is shown below.

The position/orientation of the individual gene shows the pose of the solid model in the model-based matching method. Top 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. Less bit number assigned for position and orientation of genes requires less evolving time of RM-GA, enabling the repeating time in one video input period, 33 [ms] increase. However, the rough in bit member assigned of pose induces incorrect estimation. Therefore, the bit assigned length and the performance of RM-GA conflicts each other. The length of 12 bits has been determined empirically.

Next, each individual gene gets a fitness value from the fitness function $F(\boldsymbol{\phi})$ using its assumed pose information $\boldsymbol{\phi}$. Evolution processing is performed based on the superiority or inferiority of this value, and a set of possible solutions of pose $\boldsymbol{\phi}$ for the next generation is modified through GA's process. At this time, the pose whose fitness value was high in former generation, that is, it approaches toward the maximum neighborhood of the fitness function that represents the target object. By repeating this process (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 judged to be completed that means matching with the solid model into the target has been done, the GA is thought that it has found the best result to present the pose. Since usually a time has passed before the GAs convergence, there is a possibility that the surrounding situation may be changed a lot, that means target object may turned into a very different pose. Therefore, we use RM-GA (Fig. 4). RM-GA is on-line

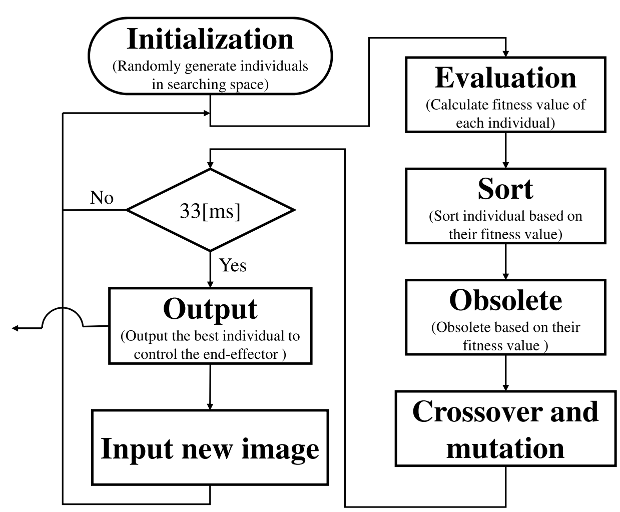


Fig. 4 Real-time multi-step GA

estimation method [8]. Its evolving speed to optimize the fitness function should be faster than the target object's moving speed, and then, we can obtain the best gene in each time point of video rate. Figure 4 shows the flowchart of RM-GA; first, the individuals of the first generation of GA genes are generated in the 3D searching space randomly. Second, each individual's fitness is calculated (Evaluation). Then, based on the calculated results, select the genes with high fitness value by sorting the order of genes from high to low (Sort), and obsolete the weak genes (Obsolete). Genes in next generation are regenerated from the selected genes through two-point crossover and mutation like changing of generation of living beings (Crossover and mutation). If the process written above was completed in 33[ms], then repeat the work again until input a new image. Otherwise, output the best individual to control the end-effector of manipulator (Fig. 5).

Utilizing RM-GA with reasonable performance in one loop and increasing accuracy with repeatable ability within real-time video rate are our approach strategy comparing to others that may provide powerful accuracy but also with computational burden and time-consuming [10].

3 Eye-vergence visual servoing system

Hand–eye composition has a shortcoming that servoing operation may become unstable, triggered by hand vibration or time-delay in pose tracking detected by dynamic video and simultaneous analyses by computer. The merits of hand–eye system is that the viewpoint can be arbitrarily chosen to find a suitable viewpoint. In this paper, the performance of the visual servoing of hand–eye composition with two cameras is experimentally evaluated. With assumption that object 3D shape is known, it is possible to measure six variables of a pose (position/orientation) also by a single eye. However, it is well known that the measurement of the distance between camera and objects tends to be incorrect. However, dual hand–eye cameras composition can overcome this problem well, which is derived from basic feature of dual-eye system, i.e., parallactic nature. Furthermore, eye-vergence mechanism is effective, since it is valid to enhance pose detection accuracy by avoiding influence of lens aberration. The eye-vergence helps that the target object could be projected at the center of images, where the lens aberration hardly exists.

$$\underbrace{\overbrace{10\cdots10}^{t_x}}_{12bit}\underbrace{\overbrace{11\cdots01}^{t_y}}_{12bit}\underbrace{\overbrace{01\cdots10}^{t_z}}_{12bit}\underbrace{\overbrace{11\cdots10}^{\varepsilon_1}}_{12bit}\underbrace{\overbrace{10\cdots10}^{\varepsilon_2}}_{12bit}\underbrace{\overbrace{10\cdots01}^{\varepsilon_3}}_{12bit}$$

Fig. 5 Information of gene

3.1 Hand-visual servoing controll

The block diagram of our proposed hand and eye-vergence visual servoing controller is shown in Fig. 6. The hand-visual servoing is the outer loop.

For the outer loop, the desired hand velocity ${}^W\dot{\boldsymbol{r}}_d$ is calculated as follows:

$$^W\dot{\boldsymbol{r}}_d = \boldsymbol{K}_{P_p}\,{}^W\boldsymbol{r}_{E,Ed} + \boldsymbol{K}_{V_p}\,{}^W\dot{\boldsymbol{r}}_{E,Ed}, \tag{3}$$

where hand error ${}^W\boldsymbol{r}_{E,Ed}$ and error velocity ${}^W\dot{\boldsymbol{r}}_{E,Ed}$ can be calculated from ${}^E\boldsymbol{T}_{Ed}$ and ${}^E\dot{\boldsymbol{T}}_{Ed}$. $\boldsymbol{K}_{P_p}$ and $\boldsymbol{K}_{V_p}$ are positive definite matrix to determine PD gain.

In addition, the desired hand angular velocity ${}^W\boldsymbol{\omega}_d$ is calculated as follows:

$$^W\boldsymbol{\omega}_d = \boldsymbol{K}_{P_o}\,{}^W\boldsymbol{R}_E\,{}^E\Delta\boldsymbol{\epsilon} + \boldsymbol{K}_{V_o}\,{}^W\boldsymbol{\omega}_{E,Ed}, \tag{4}$$

where quaternion error ${}^E\Delta\boldsymbol{\epsilon}$ and angular velocity ${}^W\boldsymbol{\omega}_{E,Ed}$ are calculated by transforming the base coordinates of ${}^E\boldsymbol{T}_{Ed}$ and ${}^E\dot{\boldsymbol{T}}_{Ed}$ from Σ_E to Σ_W. In addition, $\boldsymbol{K}_{PO}$ and $\boldsymbol{K}_{VO}$ are suitable feedback matrix gains, refer to [13].

The desired joint variable $\dot{\boldsymbol{q}}_d$ is obtained by the following:

$$\dot{\boldsymbol{q}}_d = \boldsymbol{J}^+(\boldsymbol{q})\begin{bmatrix} {}^W\dot{\boldsymbol{r}}_d \\ {}^W\boldsymbol{\omega}_d \end{bmatrix}, \tag{5}$$

where $\boldsymbol{J}^+(\boldsymbol{q})$ represents pseudo-inverse matrix of Jacobian matrix $\boldsymbol{J}(\boldsymbol{q})$.

The hardware control system of the velocity-based servo system of PA10 is expressed as follows:

$$\boldsymbol{\tau} = \boldsymbol{K}_{SP}(\dot{\boldsymbol{q}}_d - \dot{\boldsymbol{q}}) + \boldsymbol{K}_{SI}\int_0^t (\dot{\boldsymbol{q}}_d - \dot{\boldsymbol{q}})\mathrm{d}t, \tag{6}$$

where $\boldsymbol{K}_{\mathrm{SP}}$ and $\boldsymbol{K}_{\mathrm{SI}}$ are symmetric positive definite matrix to determine PD gain.

3.2 Eye-vergence visual servoing controller

The eye-vergence visual servoing is conducted by the inner loop of the visual servoing system, as shown in Fig. 6. In this paper, two pan-tilt cameras are used for eye-vergence visual servoing. Here, the positions of cameras are supposed to be fixed on the end effector.

For camera system, q_8 is tilt angle, q_9 and q_{10} are pan angles, and q_8 is common for both cameras.

As it is shown in Fig. 7a, b, ${}^Ex_{\hat{M}}$, ${}^Ey_{\hat{M}}$, ${}^Ez_{\hat{M}}$ express the position of the detected object in the end-effector coordinate. The desired angle of camera joints are calculated by the following:

$$q_{8d} = a\tan 2\left({}^Ey_{\hat{M}}, {}^Ez_{\hat{M}}\right) \tag{7}$$

$$q_{9d} = a\tan 2\left(-l_{8R} + {}^Ex_{\hat{M}}, {}^Ez_{\hat{M}}\right) \tag{8}$$

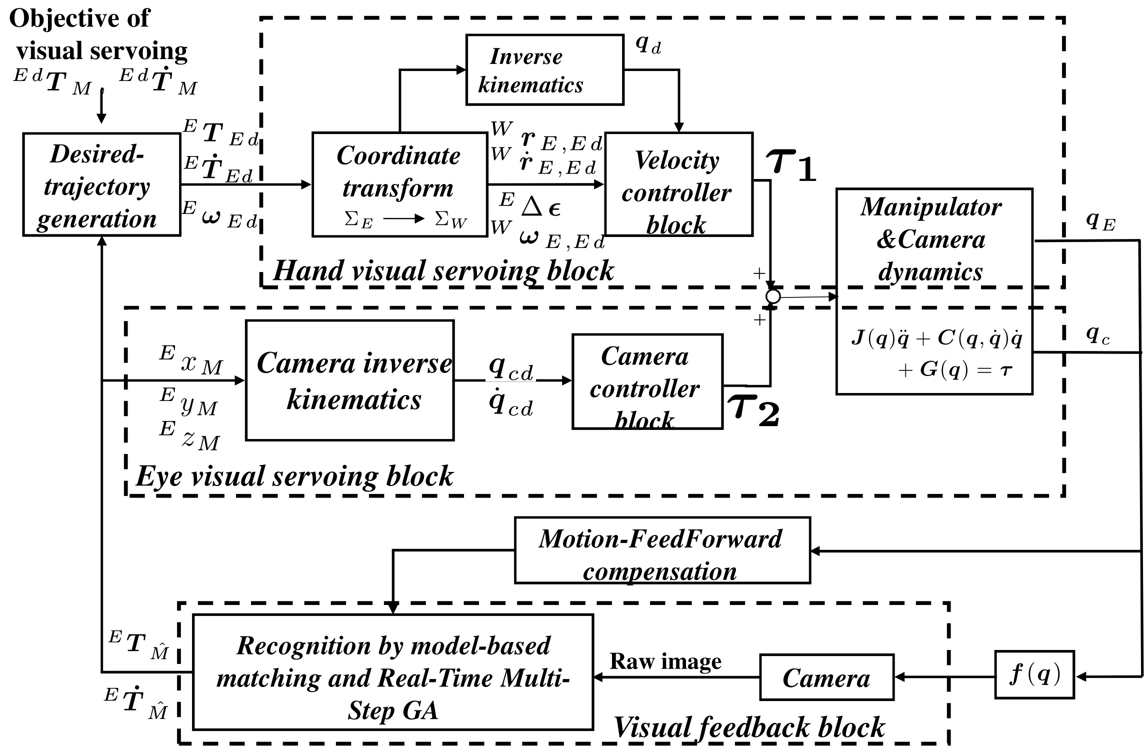
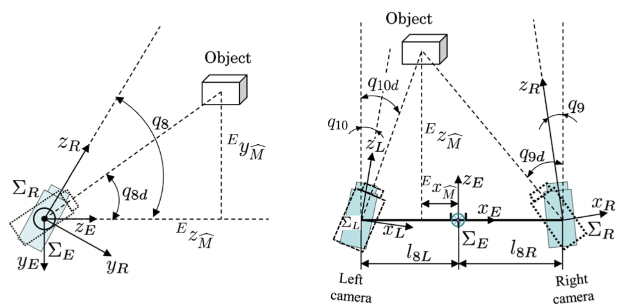


Fig. 6 Hand/eye-vergence visual servo system



(a) Tilt angle **(b)** Pan angle

Fig. 7 Calculation of tilt and pan angles

$$q_{10d} = a\tan 2\left(l_{8L} + {}^{E}x_{\hat{M}}, {}^{E}z_{\hat{M}}\right), \tag{9}$$

where $l_{8L} = l_{8R} = 120$[mm] is the camera location. The controller of eye-visual servoing is given by the following:

$$\dot{q}_{8Cd} = K_P(q_{8d} - q_8) + K_D(\dot{q}_{8d} - \dot{q}_8) \tag{10}$$

$$\dot{q}_{9Cd} = K_P(q_{9d} - q_9) + K_D(\dot{q}_{9d} - \dot{q}_9) \tag{11}$$

$$\dot{q}_{10Cd} = K_P(q_{10d} - q_{10}) + K_D(\dot{q}_{10d} - \dot{q}_{10}), \tag{12}$$

where K_P and K_D are positive control gain.

Because the motion of camera motor is an open loop, it is controlled to rotate a certain degree without getting the actual angle during the rotation, which makes the accurate camera angle cannot be got. Therefore, the desired camera angles are input in every 33 ms limited to a certain value.

4 Experiment

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 of Σ_{E_0} and Σ_{M_0} based on Σ_W are as follows:

$${}^{W}\boldsymbol{T}_{M_0} = \begin{bmatrix} 0 & 0 & -1 & -1435[\text{mm}] \\ 1 & 0 & 0 & 0[\text{mm}] \\ 0 & -1 & 0 & 450[\text{mm}] \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

$${}^{W}\boldsymbol{T}_{E_0} = \begin{bmatrix} 0 & 0 & -1 & -890[\text{mm}] \\ 1 & 0 & 0 & 0[\text{mm}] \\ 0 & -1 & 0 & 450[\text{mm}] \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

The above relations between desired hand pose Σ_{E_d}, actual hand pose Σ_E, estimated target pose $\Sigma_{\hat{M}}$, and actual target pose Σ_M are depicted in Fig. 8. From Fig. 8, we have

$$\Delta^{E}z = {}^{W}x_E - {}^{W}x_{Ed}, \tag{15}$$

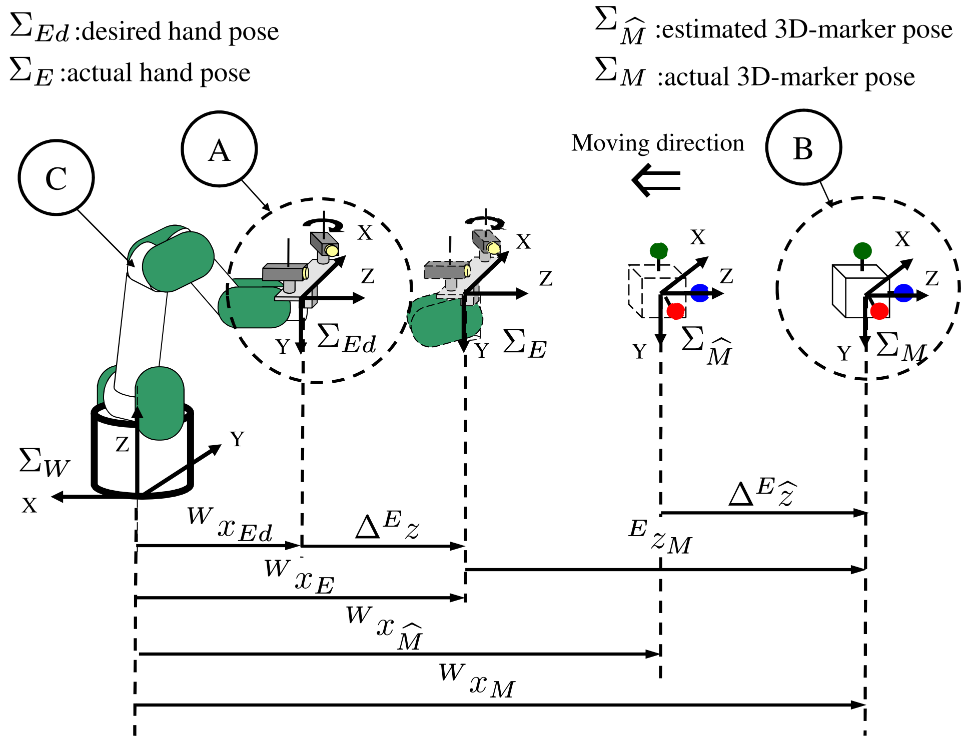


Fig. 8 Schematic coordinates relation concerning analyzed longitudinal visual servoing experiments with eye-vergence pose tracking

$$\Delta^E\widehat{z} = {}^E z_M + {}^E\boldsymbol{R}_W({}^W x_{\widehat{M}} - {}^W x_E), \tag{16}$$

where ${}^W x_{\mathrm{Ed}}$ desired hand x position; ${}^W x_E$ hand x position measured by robot; ${}^W x_M$ target x position measured by actuator of marker's motion; ${}^E z_M$ the tracking error of hand; $\Delta^E\widehat{z}$:the pose estimation error of RM-GA.

The target object moves according to the following time function as follows:

$${}^{M_0} z_M(t) = -150 + 150\cos(\omega t)[\mathrm{mm}]. \tag{17}$$

The relation between the object and the desired end-effector is set to be constant as follows:

$${}^{Ed}\boldsymbol{\psi}_M = [0, -100[\mathrm{mm}], 545[\mathrm{mm}], 0, 0, 0]. \tag{18}$$

Since it is reasonable that when fitness value was low, the estimated object pose would not be reliable. We have set a minimum critical threshold value of fitness as 0.1 and the visual servoing feedback would be cut when fitness values become less than 0.1.

In this paper, to examine reliability of RM-GA, the position of 3D marker (x, z coordinates in Σ_W) was set to be unknown, while the orientation ($\varepsilon_1, \varepsilon_2, \varepsilon_3$) and the y-direction coordinate were set as constant, since those variables do not change in the longitudinal visual servoing experiments. In addition, different from the previous works [12], the distribution of $F(\boldsymbol{\phi})$ was calculated to examine the true pose (the pose that gives the maximum peak of $F(\boldsymbol{\phi})$) and the detected pose by RM-GA whose position is pointed in $F(\boldsymbol{\phi})$ distribution, as shown in Fig. 9. In that analysis, the angle of the two cameras (θ_1, θ_2) has been taken into account to increase the accuracy of calculation of $F(\boldsymbol{\phi})$.

5 Longitudinal visual servoing results

Two kinds of frequency response experiments have been conducted, one with angular velocity ω=0.618[rad/s], shown in Table 1 and the other one with ω=0.309[rad/s], shown in Table 2. During the experiments, left and right camera images have been stored at the time shown by $(A) \sim (F)$ in the leftiest column in Tables 1 and 2, and the fitness value calculated by RM-GA and maximum value are listed next. The positions to give the fitness value listed in column (1) are shown in column (3), and the positions to give the highest peak are also listed in column (4). Figure 9 shows the $F(\boldsymbol{\phi})$ distribution calculated from left and right images taken at 14.078[s] as shown in row with (B) by changing ${}^E x$ and ${}^E z$, with other pose parameters and θ_1, θ_2 fixed at actual values. The white point in the Fig. 9 means the target object's position detected by RM-GA, and the peak shows the target true position. In the figure, the distance between the point designated by white dot that is detected by RM-GA and the full-search peak means the error of desired hand position and actually detected position.

Table 2 Parameters in experiment of $\omega = 0.309$rad/s

Time points[s]	(1) Fitness value of RM-GA	(2) Maximum fitness value	(3) Position of found by GA(x, y, z)[mm]	(4) Position with maximum fitness value (x, y, z)[mm]
(A)6.188	0.8056	0.926	15.332, −112.529, 453.984	17, −112.529, 482
(B)12.672	0.861	0.9259	−5.566, −108.916, 564.922	−9, −108.916, 596
(C)26.766	0.824	0.8519	12.207, −116.24, 450.566	15, −116.24, 468
(D)34.766	0.778	0.8889	−5.371, −117.51, 612.969	0 , −117.51, 652
(E)43.313	0.824	0.8333	14.453, −110.674, 469.219	19, −110.674, 482
(F)51.891	0.926	0.9444	11.719, −121.514, 555.742	12, −121.514, 586

Table 1 Parameters in experiment of $\omega = 0.618$rad/s

Time points[s]	(1) Fitness value of RM-GA	(2) Maximum fitness value	(3) Position of found by GA(x, y, z)[mm]	(4) Position with maximum fitness value (x, y, z)[mm]
(A)6.594	0.907	0.9772	8.887, −108.184, 580.156	7, −108.184, 616
(B)14.078	0.944	0.9907	20.41, −114.238, 418.633	20, −114.238, 448
(C)22.609	0.852	0.9259	8.496, −121.025, 437.285	10, −121.025, 448
(D)32.156	0.972	0.983	9.277, −129.473, 454.961	12, −129.473, 476
(E)42.984	0.833	0.9167	23.535, −125.322, 409.941	29, −125.322, 426
(F)55.047	0.981	0.9907	15.137, −118.779, 484.453	14, −118.779, 518

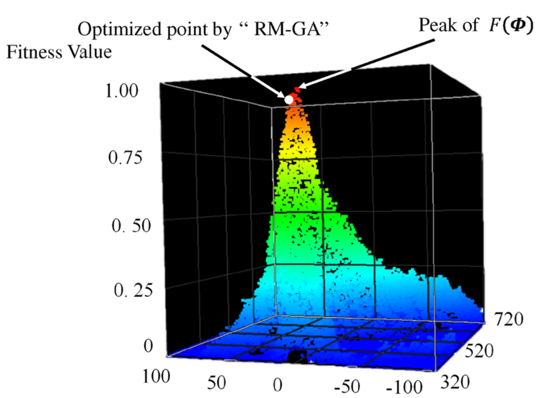


Fig. 9 Fitness distributions, $F(\boldsymbol{\phi})$, given by Eq. (1).

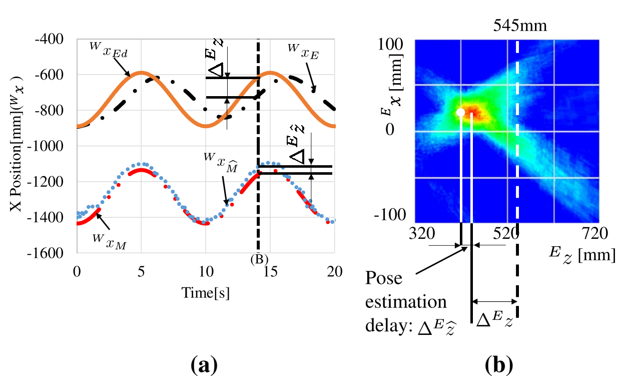


Fig. 10 Relationship between the fitness distributions and frequency respond results

Figure 10b displays Fig. 9 in flat, and Fig. 10a shows the time-profile of ^Wx_d, the desired hand x position represented by the motion of 3D marker shown in Fig. 8. In that figure, the arrow ("⇐") means the moving direction of the 3D marker at the time t=14.078[s] that is listed in Table 1 as (B). ^Wx_E means hand x position in Σ_W, and ^Wx_M and $^Wx_{\hat{M}}$ represent x position of 3D marker and detected x position by RM-GA. The distance $\Delta^E\hat{z}$ means tracking error of the RM-GA that continues to detect the 3D marker's pose, which corresponds to the distance $\Delta^E\hat{z}$ in Fig. 10b. The hand tracking error, $\Delta^E z$ in (a), also corresponds to the distance, $\Delta^E z$, depicted in (b).

According to Fig. 10, we can see that there was delay in the tracking performance ($\Delta^E z$=100 [mm]), while the detected error that was the result of real-time pose estimation by RM-GA and eye-vergence mechanism is $\Delta^E\hat{z}$=30 [mm], less than the hand tracking delay, 100 [mm]. This means the eye-vergence system can track and focus on the target object correctly than the performance of hand tracking, and this feature enhance the system would not lose the image of 3D marker from the camera-seeable area, meaning that the visual servoing control can prevent the feedback would not be lost.

Figure 9 represents frequency response with $\omega = 0.618$ [rad/s] whose numerical data are in Table 1, and Fig. 12 with $\omega = 0.314$[rad/s] whose numerical data are in Table 2. In Figs.11 and 12, all variables are presented in Σ_W, other than the fitness distributions from (A) to (F) that were presented by Σ_E. The fitness distributions with (A) to (F) are calculated using two camera images taken at the time designated by (A) to (F) in the time-profile

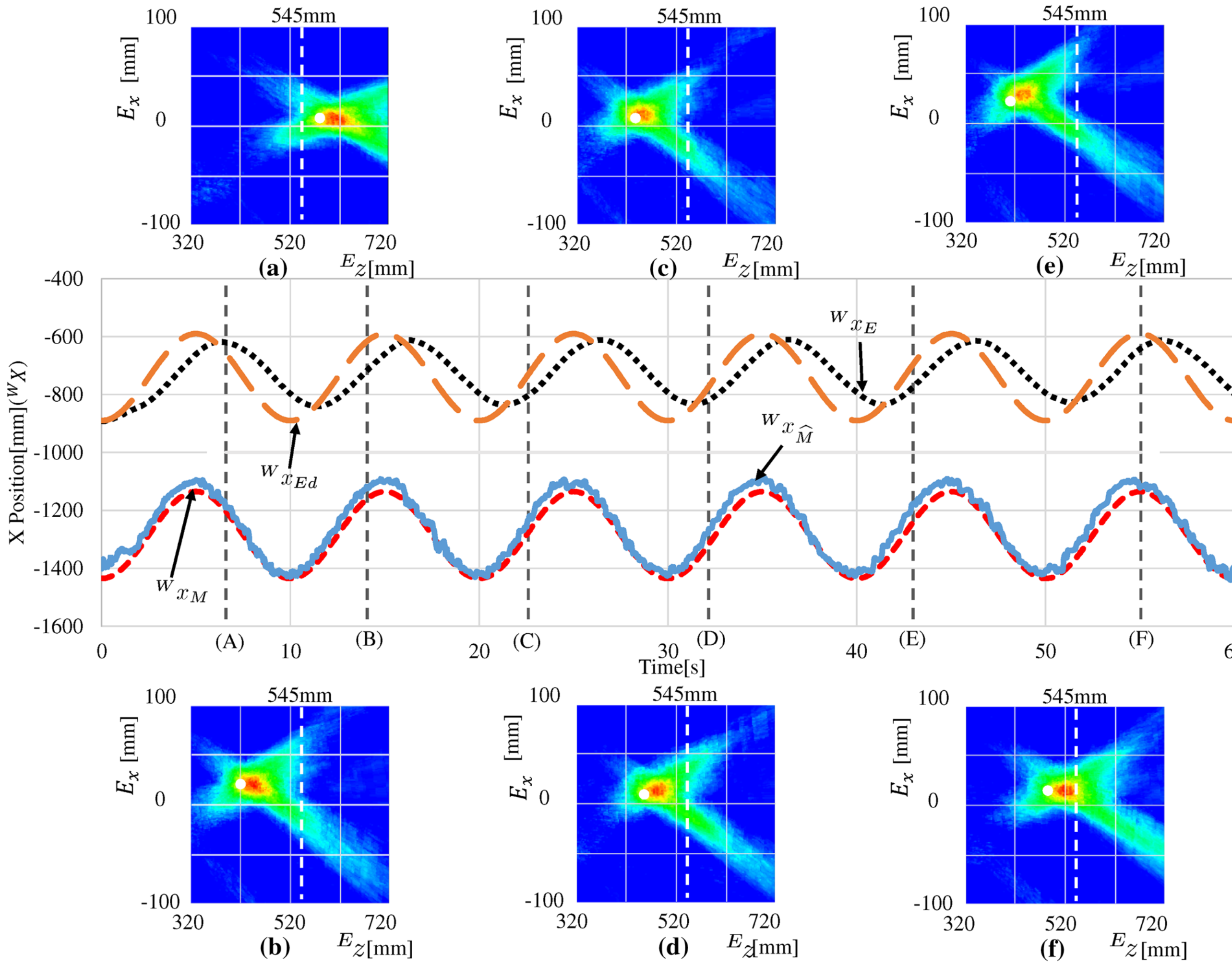


Fig. 11 Relation between fitness value, position of end effector, actual position of target object, and position estimated in z–x plane by GA in $\omega = 0.618$ rad/s, T=10[s]

graphs at the center position. The time-profile graphs are ^Wx_E, $^Wx_{Ed}$, ^Wx_M, and $^Wx_{\widehat{M}}$, which are defined in Fig. 8. The fitness distribution situations were shown in Fig. 11(A)~(F) and Fig. 12 (A)~(F). Similar to the explanation of Fig. 10, the black points in each graph show the marker's x position estimated by RM-GA, and the peak of $F(\boldsymbol{\phi})$ means the true x position of the marker. The distance between the black point and peak means the estimation error. In addition, according to Eq. (11), the end-effector and the marker were set to maintain a distance of 545[mm] between them, so the distance between the white dashed line in (A) to (F) and the peak shows the tracking error of hand.

From the results, we can see that when period of the target's motion is $T = 20[s]$ (Fig. 12), it can be confirmed that the end-effector tracks the target with fewer phase delay, and the RM-GA can recognize the highest peak correctly even when the target is changing its position. Even though the situation $T = 10[s]$ in Fig. 11, to make the correct tracking of end-effector is difficult, but from Tables 1 and 2, the position error of estimated results between full search and "RM-GA" both can be constricted in a certain range (10 ~ 40[mm]), which ensure the cameras keep staring at the target.

6 Conclusion

In this paper, the visual servoing was evaluated in the experiments of frequency response, and the performance of RM-GA in a visual servoing was shown. The real-time estimation tracking error has been grasped by clarifying the results of GA, in detail. It has been confirmed that the eye-vergence system helped the visual servoing system

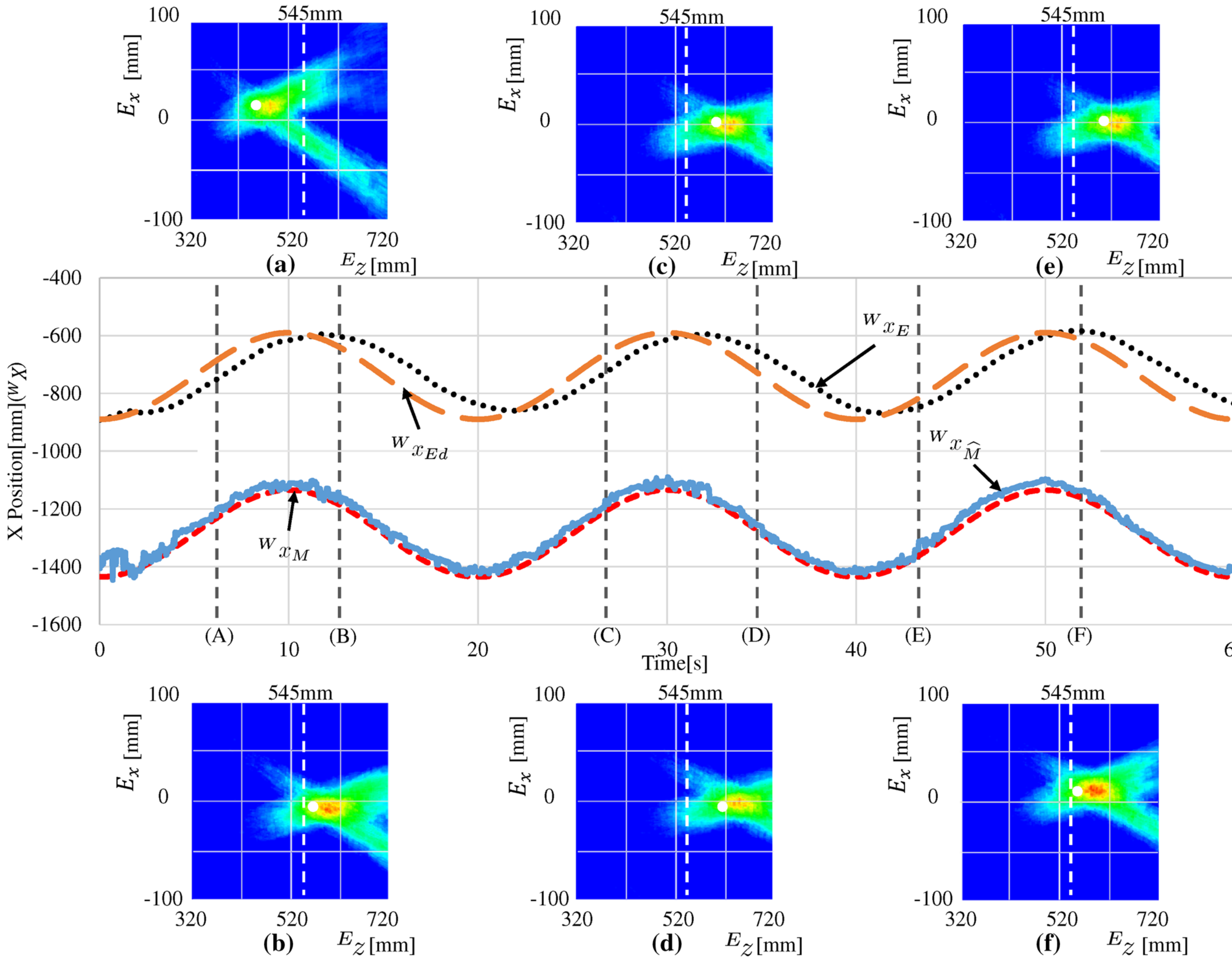


Fig. 12 Relation between fitness value, position of end effector, actual position of target object, and position estimated in z–x plane by GA in $\omega = 0.314$ rad/s, T=20[s]

track the target object. Comparing with the full-search method, the superiority of "RM-GA" can be confirmed.

References

1. Hutchinson S, Hager GD, Corke PI (1996) A tutorial on visual servo control. Robot Autom IEEE Trans 12(5):651–670
2. Oh PY, Allen PK (2001) Visual servoing by partitioning degrees of freedom. Robot Autom IEEE Trans 17(1):1–17
3. Malis E, Chaumentte F, Boudet S (1999) 2–1/2-D visual servoing. Robot Autom IEEE Trans 15(2):238–250
4. Allen PK et al (1993) Automated tracking and grasping of a moving target with a robotic hand-eye system. Autom Robot Autom IEEE Trans 9(2):152–165
5. Song W, Minmami M, Mae Y, Aoyagi S. (2007) On-line evolutionary head pose measurement by feedforward stereo model matching. In: Robotics and automation. ICRA'07. 2007 IEEE international conference on IEEE, pp 4394–4400
6. Stavnitzky J, Capson D (2000) Multiple camera model-based 3-D visual servoing. Robot Autom IEEE Trans 16(6):732–739
7. Dune C, Marchand E, Leroux C (2007) One click focus with eye-in-hand/Eye-to-hand Cooperationh. In: Robotics and automation. 2007 IEEE international conference on IEEE, pp 2471–2476
8. Suzuki H, Minami M (2005) Visual servoing to catch fish using global/local GA Search. Mechatron IEEE/ASME Trans 10(3):352–357
9. Minami Song W (2009) Hand-eye-motion invariant pose estimation with On-line 1-step GA-3D pose tracking accuracy evaluation in dynamic hand-eye oscillation. J Robot Mechatron 21(6):709–719
10. Chesi G, Hung Y (2007) Global path-planning for constrained and optimal visual servoing. IEEE Trans Robot 23(5):1050–1060
11. Song W, Minami M, Yu F, Zhang Y, Yanou A (2011) 3-D hand & eye-vergence approaching visual servoing with lyapunouv-stable pose tracking. In Robotics and Automation,2011,ICRA'11. 2011 IEEE International Conference on IEEE, pp 5210–5217
12. Tian H, Cui Y, Minami M, Yanou A (2016) Frequency response experiments of eye-vergence visual servoing in lateral motion with 3D evolutionary pose tracking. Artif Life Robot 1(8):36–43
13. Song W, Minami M (2009) Hand & eye-vergence dual visual servoing to enhance observability and stability. In: Robotics and Automation, ICRA'09. IEEE International Conference on IEEE, pp 714–721