Eye-Vergence Visual Servoing on Moving Target

○ Pengyu Liu Mamoru Minami Sen Hou Koichi Maeda Akira Yanou
Okayama University

Abstract Visual servoing system is developed to control robot with visual information. It can be used to make robot adaptive to changing and unknown environment. A method of Eye-vergence visual servoing (EVS in short) is proposed to track moving target in this paper. The EVS system is superior to fixed-hand-eye system in a view point that eye-vergence system enables the hand-eye camera to look the target object always at the center of camera lens. The EVS system can be used to make the camera rotate to a proper position, and get clear images of the target object. In this paper, target object recognition method based on 1-step Genetic algorithm is presented, position-based controller is designed, and the trackability of the EVS system is tested with the frequency response experiments in depth directional visual servoing. The experiments show that the EVS system is stable and effective, and the moving target can be well tracked.

1 Introduction

In recent years, visual servoing attracts our interest in the field of robot vision [1]. Visual servoing can be classified into three major groups: position-based [2], image-based [3] and hybrid visual servoing [4].

Visual servoing, a method to control a robot with visual information in the feedback loop, is expected to be able to allow the robot adapt to the changing and unknown environment. Some methods [5] have already been proposed to improve observation abilities by multiple cameras, stereo cameras, etc.

Visual servoing system is generally used as a hand-eye configuration, in which the camera is fixed on the end-effector. However, the end-effector is not flexible enough for rapid tracking in this case, stemming from dynamical effects against hand's motions. Therefore, A method of Eye-vergence visual servoing was proposed to track rapidly moving target in this paper. The Eye-vergence system is superior to fixed-hand-eye system in a view point that eye-vergence system enables the hand-eye camera rotate flexibly so as to find the target object always at the center of camera lens.

2 Object Recognition Method

2.1 Model-based Matching Method

In this part, a model-based matching method was presented. The searching model is shown as Fig.1. In the figure, $S_{R,in}$, and $S_{L,in}$ are the inside spaces of coordinate on the surface of the block model. Accordingly, $S_{R,out}$, and $S_{L,out}$ are the outside spaces. The H value of right image at the position ${}^{IR}\boldsymbol{r}_i$ is expressed as $p({}^{IR}\boldsymbol{r}_i)$, and the H value of left image at the position ${}^{IL}\boldsymbol{r}_i$ is expressed as $p({}^{IL}\boldsymbol{r}_i)$. ${}^{IR}\boldsymbol{r}_i$ and ${}^{IL}\boldsymbol{r}_i$ are the positions of the pixels.

If it was defined that $\boldsymbol{m}(\boldsymbol{r}) = 1(\boldsymbol{r} \in \boldsymbol{S}_{R,in}$, or $\boldsymbol{r} \in \boldsymbol{S}_{L,in})$, $\boldsymbol{m}(\boldsymbol{r}) = -1(\boldsymbol{r} \in \boldsymbol{S}_{R,out}$, or $\boldsymbol{r} \in \boldsymbol{S}_{L,out})$, and $\boldsymbol{\phi}$ was the position/orientation of the model, then the fitting evaluation function was

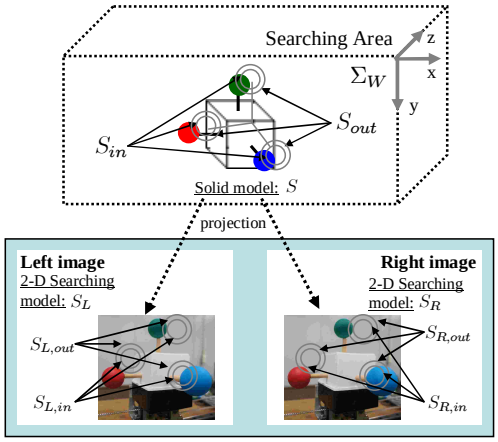


Fig. 1: Searching model

$$
\begin{aligned}
&F_{site}(\boldsymbol{\phi}) \\
&= \left\{ \left(\sum_{{}^{IR}\boldsymbol{r}_i \in S_{R,in}(\boldsymbol{\phi})} m * p({}^{IR}\boldsymbol{r}_i) - \sum_{{}^{IR}\boldsymbol{r}_i \in S_{R,out}(\boldsymbol{\phi})} m * p({}^{IR}\boldsymbol{r}_i) \right) \right. \\
&\left. + \left(\sum_{{}^{IL}\boldsymbol{r}_i \in S_{L,in}(\boldsymbol{\phi})} m * p({}^{IL}\boldsymbol{r}_i) - \sum_{{}^{IL}\boldsymbol{r}_i \in S_{L,out}(\boldsymbol{\phi})} m * p({}^{IL}\boldsymbol{r}_i) \right) \right\} /2 \\
&= \{F_{R,site}(\boldsymbol{\phi}) + F_{L,site}(\boldsymbol{\phi})\} /2 \qquad (1)
\end{aligned}
$$

Equation (1) was used as a fitness function in GA process. When the moving searching model fitted to the target object being imaged in the right and left images, the fitness function $\boldsymbol{F}_{site}(\boldsymbol{\phi})$ got maximum value.

Therefore, the problem of finding a target object and detecting its pose can be converted to searching $\boldsymbol{\phi}$ that maximize $\boldsymbol{F}_{site}(\boldsymbol{\phi})$. This problem will be solved with 1-step GA algorithm.

2.2 "1-step GA" Algorithm for Evolutionary Recognition

"1-step GA" algorithm was used to find the target object in this paper. The GA evaluational iteration was applied one time to every newly input image.

Firstly, several model individuals, as the gene of GA, were generated at random in the searching area, and each of them had its own position/orientation information. The fitness function was taken to calculate the fitness of every individual. And then, the individuals were sorted according to their fitness values, and the best individual which had the maximum fitness value was gotten as the recognition result, and outputted as the real-time position/orientation information of the target object. Finally, some top ranking individuals were selected to create new genes of GA by the way of crossover and mutation.

3 Eye-Vergence Visual servoing Controller

3.1 Desired-trajectory generation

In Fig.2, the world coordinate frame is denoted by Σ_W, the target coordinate frame is denoted by Σ_M, and the desired and actual end-effector coordinate frame is denoted by Σ_{Ed}, Σ_E respectively. The desired relation between the target and the end-effector is given by Homogeneous Transformation as ${}^{Ed}\boldsymbol{T}_M$, the relation between the target and the actual end-effector is given by ${}^{E}\boldsymbol{T}_M$, then the difference between the desired end-effector pose Σ_{Ed} and the actual end-effector pose Σ_E is denoted as ${}^{E}\boldsymbol{T}_{Ed}$, which can be described by:

$$ {}^{E}\boldsymbol{T}_{Ed}(t) = {}^{E}\boldsymbol{T}_M(t)\,{}^{Ed}\boldsymbol{T}_M^{-1}(t) \tag{2} $$

(2) is a general representation of hand pose tracking error that satisfies arbitrary object motion ${}^{W}\boldsymbol{T}_M(t)$ and arbitrary visual servoing objective ${}^{Ed}\boldsymbol{T}_M(t)$. The relation ${}^{E}\boldsymbol{T}_M(t)$ can be estimated by 1-step GA [6], having been presented as an on-line model-based pose estimation method. Let $\Sigma_{\hat{M}}$ denote the detected object, It is natural there should always exist an error between the actual object Σ_M and the detected one $\Sigma_{\hat{M}}$. So in visual servoing, (2) will be rewritten based on $\Sigma_{\hat{M}}$ that includes the error ${}^{M}\boldsymbol{T}_{\hat{M}}$, as

$$ {}^{E}\boldsymbol{T}_{Ed}(t) = {}^{E}\boldsymbol{T}_{\hat{M}}(t)\,{}^{Ed}\boldsymbol{T}_{\hat{M}}^{-1}(t). \tag{3} $$

Differentiating (3) with respect to time yields

$$ {}^{E}\dot{\boldsymbol{T}}_{Ed}(t) = {}^{E}\dot{\boldsymbol{T}}_{\hat{M}}(t)\,{}^{\hat{M}}\boldsymbol{T}_{Ed}(t) + {}^{E}\boldsymbol{T}_{\hat{M}}(t)\,{}^{\hat{M}}\dot{\boldsymbol{T}}_{Ed}(t). \tag{4} $$

Differentiating (4) with respect to time again

$$ {}^{E}\ddot{\boldsymbol{T}}_{Ed}(t) = {}^{E}\ddot{\boldsymbol{T}}_{\hat{M}}(t)\,{}^{\hat{M}}\boldsymbol{T}_{Ed}(t) + 2\,{}^{E}\dot{\boldsymbol{T}}_{\hat{M}}(t)\,{}^{\hat{M}}\dot{\boldsymbol{T}}_{Ed}(t) + {}^{E}\boldsymbol{T}_{\hat{M}}(t)\,{}^{\hat{M}}\ddot{\boldsymbol{T}}_{Ed}(t), \tag{5} $$

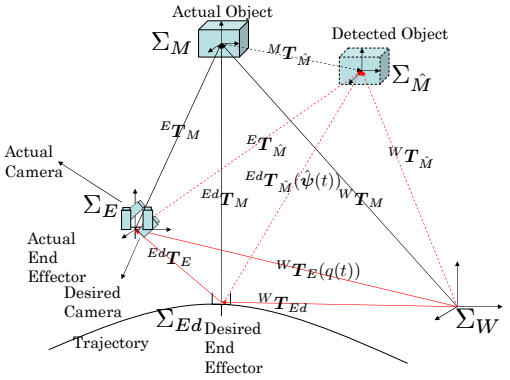


Fig. 2: Motion of the end-effector and object

where ${}^{\hat{M}}\boldsymbol{T}_{Ed}$, ${}^{\hat{M}}\dot{\boldsymbol{T}}_{Ed}$, ${}^{\hat{M}}\ddot{\boldsymbol{T}}_{Ed}$ are given as the desired visual servoing objective. ${}^{E}\boldsymbol{T}_{\hat{M}}$, ${}^{E}\dot{\boldsymbol{T}}_{\hat{M}}$, ${}^{E}\ddot{\boldsymbol{T}}_{\hat{M}}$ can be observed by cameras. As shown in Fig.2, there are two errors left to be decreased in the visual system. One is the error between the actual object and the detected one, ${}^{M}\boldsymbol{T}_{\hat{M}}$, and the other is the error between the desired end-effector and the actual one, ${}^{E}\boldsymbol{T}_{Ed}$. The error of ${}^{M}\boldsymbol{T}_{\hat{M}}$ is decreased by pose tracking method of the "1-step GA" [6], and the eye-vergence camera system, and the error of ${}^{E}\boldsymbol{T}_{Ed}$ depends on the performances of the hand visual servoing controller.

3.2 Hand Visual Servoing Controller

The block diagram of the eye-vergence visual servoing controller is shown as Fig.3. Based on the above analysis of the desired-trajectory generation, the desired hand velocity ${}^{W}\dot{\boldsymbol{r}}_d$ is calculated as,

$$ {}^{W}\dot{\boldsymbol{r}}_d = \boldsymbol{K}_{P_p}\,{}^{W}\boldsymbol{r}_{E,Ed} + \boldsymbol{K}_{V_p}\,{}^{W}\dot{\boldsymbol{r}}_{E,Ed}, \tag{6} $$

where ${}^{W}\boldsymbol{r}_{E,Ed}$, ${}^{W}\dot{\boldsymbol{r}}_{E,Ed}$ 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.

The desired hand angular velocity ${}^{W}\boldsymbol{\omega}_d$ is calculated

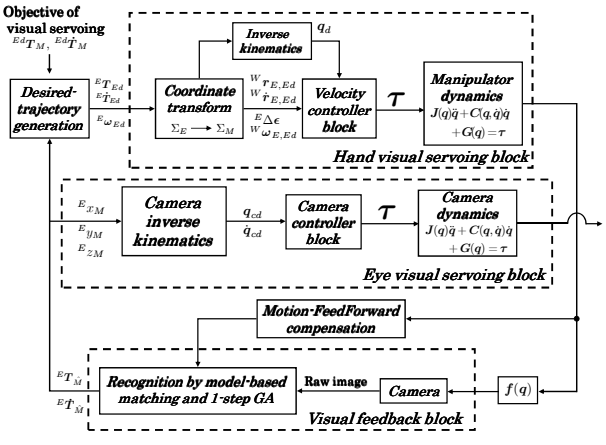


Fig. 3: Block diagram of Eye-Vergence visual servoing controller

as,

$$^{W}\boldsymbol{\omega}_{d} = \boldsymbol{K}_{P_o}{}^{W}\boldsymbol{R}_{E}{}^{E}\Delta\boldsymbol{\epsilon} + \boldsymbol{K}_{V_o}{}^{W}\boldsymbol{\omega}_{E,Ed}, \quad (7)$$

where $^{E}\Delta\boldsymbol{\epsilon}$ is a quaternion error [6] calculated from the pose tracking result, and $^{W}\boldsymbol{\omega}_{E,Ed}$ can be computed by transforming the base coordinates of $^{E}\boldsymbol{T}_{Ed}$ and $^{E}\dot{\boldsymbol{T}}_{Ed}$ from Σ_E to Σ_W. Also, $\boldsymbol{K}_{P_o}$ and $\boldsymbol{K}_{V_o}$ are suitable feedback matrix gains. We define the desired hand pose as $^{W}\boldsymbol{\psi}_{d}^{T} = [^{W}\boldsymbol{r}_{d}^{T}, ^{W}\boldsymbol{\epsilon}_{d}^{T}]^{T}$

The desired joint variable $\boldsymbol{q}_{Ed} = [q_{1d}, ..., q_{7d}]^{T}$ and $\dot{\boldsymbol{q}}_{Ed}$ is obtained by

$$\boldsymbol{q}_{Ed} = \boldsymbol{f}^{-1}(^{W}\boldsymbol{\psi}_{d}^{T}) \quad (8)$$

$$\dot{\boldsymbol{q}}_{Ed} = \boldsymbol{J}_{E}^{+}(\boldsymbol{q}) \begin{bmatrix} ^{W}\dot{\boldsymbol{r}}_{d} \\ ^{W}\boldsymbol{\omega}_{d} \end{bmatrix} \quad (9)$$

where $\boldsymbol{f}^{-1}(^{W}\boldsymbol{\psi}_{d}^{T})$ is the inverse kinematic function and $\boldsymbol{J}_{E}^{+}(\boldsymbol{q})$ is the pseudo-inverse matrix of $\boldsymbol{J}_{E}(\boldsymbol{q})$, and $\boldsymbol{J}_{E}^{+}(\boldsymbol{q}) = \boldsymbol{J}_{E}^{T}(\boldsymbol{J}_{E}\boldsymbol{J}_{E}^{T})^{-1}$.

The robot arm is a 7 links manipulator, and the end-effector has 6-DoF, so it has a redundance. In the research before, we only calculated the position of the manipulator's end-effector, but not considering the joint angles through the position of the manipulator's end-effector. For one end-effector pose, there may exist infinite kinds of shapes, which will make the system dangerous. In this report, we made q_1 is 0, and used the inverse kinematics to calculate all joint angles. It can solve the redundancy problem. Meanwhile we took a controller to make the joint of angles approximately as the desired joint angles. So we defined the formula of the desired joint angles in the new controller as

$$\dot{\boldsymbol{q}}_{Ed} = \boldsymbol{k}_{p}(\boldsymbol{q}_{Ed} - \boldsymbol{q}_{E}) + \boldsymbol{J}_{E}^{+}(\boldsymbol{q}) \begin{bmatrix} ^{W}\dot{\boldsymbol{r}}_{d} \\ ^{W}\boldsymbol{\omega}_{d} \end{bmatrix} \quad (10)$$

where $\boldsymbol{k}_{p}$ is P positive gain.

3.3 Eye-vergence Visual Servoing Controller

Two pan-tilt cameras were used for eye-vergence visual servoing in this paper. The positions of cameras were supposed to be fixed on the end-effector. For camera system, q_8 is tilt angle, q_9 and q_{10} are pan angles,

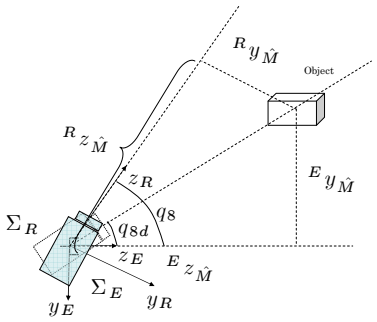


Fig. 4: Calculation of tilt angle

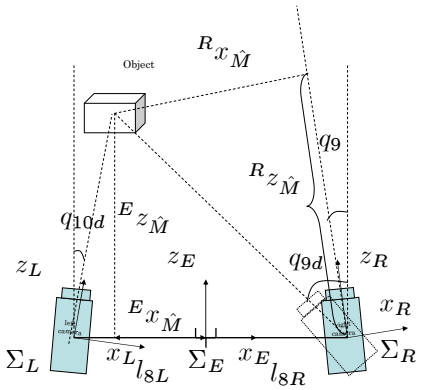


Fig. 5: Calculation of pan angles

and q_8 is common for both cameras. As it is shown in Fig.4 and Fig.5, $^{E}x_{\hat{M}}$, $^{E}y_{\hat{M}}$, $^{E}z_{\hat{M}}$ express position of the detected object in the end-effector coordinate. The desired angle of the camera joints are calculated by:

$$q_{8d} = atan2(^{E}y_{\hat{M}}, ^{E}z_{\hat{M}}) \quad (11)$$

$$q_{9d} = atan2(-l_{8R} + ^{E}x_{\hat{M}}, ^{E}z_{\hat{M}}) \quad (12)$$

$$q_{10d} = atan2(l_{8L} + ^{E}x_{\hat{M}}, ^{E}z_{\hat{M}}) \quad (13)$$

where $l_{8L} = l_{8R} = 120[mm]$ that is the camera location. We set the center line of the camera as the z axis of each camera coordinate, so the object will be in the center of the sight of the right camera when $^{R}x_{\hat{M}} = 0$ and $^{R}y_{\hat{M}} = 0$, $^{R}x_{\hat{M}}$, $^{R}y_{\hat{M}}$, $^{R}z_{\hat{M}}$ express the position of the detected object in the right camera coordinate. The controller of eye-visual servoing is given by

$$\dot{q}_{8Cd} = K_P(q_{8d} - q_8) + K_D(\dot{q}_{8d} - \dot{q}_8) \quad (14)$$

$$\dot{q}_{9Cd} = K_P(q_{9d} - q_9) + K_D(\dot{q}_{9d} - \dot{q}_9) \quad (15)$$

$$\dot{q}_{10Cd} = K_P(q_{10d} - q_{10}) + K_D(\dot{q}_{10d} - \dot{q}_{10}) \quad (16)$$

where K_P, K_D are positive control gain. Where $\dot{q}_{Cd} = [\dot{q}_{8Cd}, \dot{q}_{9Cd}, \dot{q}_{10Cd}]^{T}$

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

$$\boldsymbol{\tau} = \boldsymbol{K}_{SP}(\boldsymbol{q}_{d} - \boldsymbol{q}) + \boldsymbol{K}_{SD}(\dot{\boldsymbol{q}}_{d} - \dot{\boldsymbol{q}}) \quad (17)$$

where $\boldsymbol{K}_{SP}$ and $\boldsymbol{K}_{SD}$ are symmetric positive definite matrices to determine PD gain, and $\boldsymbol{q}_{d} = [q_{Ed}^{T}, q_{Cd}^{T}]^{T}$.

Because the motion of camera motor is an open loop, we can only make it rotate a certain degree without getting the actual angle during the rotation, which make us cannot get the accurate camera angle. So the desired camera angles are input in every 33ms, and the input is limited to a certain value.

4 Experiment of hand & eye-vergence visual servoing

4.1 Experimental system

To verify the effectiveness of the eye-vergence visual servoing system through real robot, a robot with

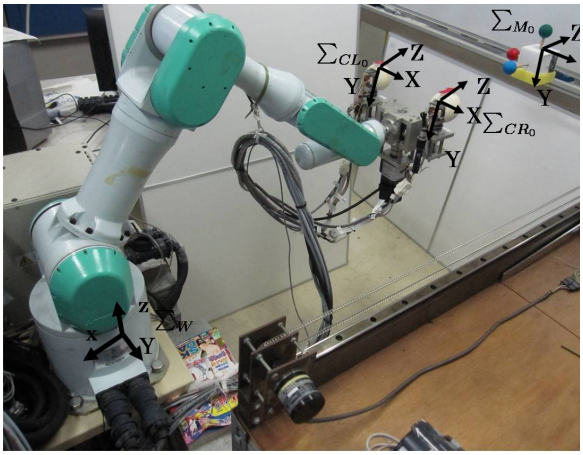


Fig. 6: Experiment environment

two cameras on the end-effector was used, as shown in Fig.6. The 3D marker used for the target object was composed of a red ball, a green ball and a blue ball. The coordinate of the target object and the manipulator in experiment were shown in Fig.6.

4.2 Experiment condition

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

$$ {}^{W}\boldsymbol{T}_{M_0} = \begin{bmatrix} 0 & 0 & -1 & -1435[mm] \\ 1 & 0 & 0 & 0[mm] \\ 0 & -1 & 0 & 499[mm] \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{18} $$

$$ {}^{W}\boldsymbol{T}_{E_0} = \begin{bmatrix} 0 & 0 & -1 & -890[mm] \\ 1 & 0 & 0 & 0[mm] \\ 0 & -1 & 0 & 499[mm] \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19} $$

The target object Σ_{M_0} was driven to moves along z axis, and the time function was defined as below.

$$ {}^{M_0}z_M(t) = -150 + 150\cos(\omega t)[mm] \tag{20} $$

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

$$ {}^{Ed}\boldsymbol{\psi}_M = [0, -90[mm], 545[mm], 0, 0, 0] \tag{21} $$

In order to test the trackability of visual servoing system, the 1-Dof position, 3-Dof position, 6-Dof position/orientation recognition experiments were done with the angular velocities of ω=0.209, ω=0.314, ω=0.628, and ω=1.256 separately.

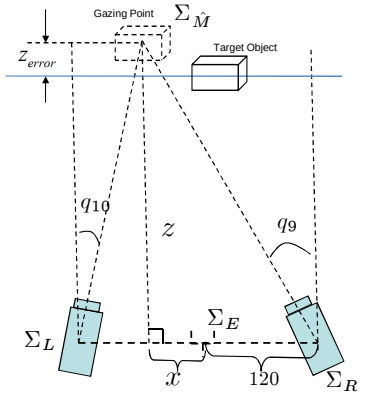


Fig. 7: Calculation of Gazing point

4.3 Definition of trackability

In order to evaluate the trackability of the cameras, a gazing point was defined as the intersection of the gazing line of both cameras. The gazing point was calculated according to Fig.7. q_9 and q_{10} were the rotation angles of both cameras. The distance between two cameras was 240 mm. From Fig.7, the following equations can be gotten.

$$ \frac{z}{120 - x} = tan(\frac{\pi}{2} - q_{10}) \tag{22} $$

$$ \frac{z}{120 + x} = tan(\frac{\pi}{2} - q_9) \tag{23} $$

Therefore, x and z can be gotten as:

$$ x = \frac{-120tan(\frac{\pi}{2} - q_9) + 120tan(\frac{\pi}{2} - q_{10})}{tan(\frac{\pi}{2} - q_9) + tan(\frac{\pi}{2} - q_{10})} \tag{24} $$

$$ z = \frac{240tan(\frac{\pi}{2} - q_9)tan(\frac{\pi}{2} - q_{10})}{tan(\frac{\pi}{2} - q_9) + tan(\frac{\pi}{2} - q_{10})} \tag{25} $$

The distances between hand and camera were 50 mm along z axis, and 95 mm along y axis. The coordinate in world coordination can be calculated as:

$$ {}^{E}\boldsymbol{P} = \begin{bmatrix} \frac{120tan(\frac{\pi}{2} - q_9) - 120tan(\frac{\pi}{2} - q_{10})}{tan(\frac{\pi}{2} - q_9) + tan(\frac{\pi}{2} - q_{10})} \\ 95 \\ \frac{240tan(\frac{\pi}{2} - q_9)tan(\frac{\pi}{2} - q_{10})}{tan(\frac{\pi}{2} - q_9) + tan(\frac{\pi}{2} - q_{10})} - 50 \end{bmatrix} \tag{26} $$

4.4 Experiment Results

Every experiment was done in 40 seconds. According to the experiments, 6 Dof position/orientation can be recognized with the eye-vergence visual servoing system when the motion period was 5 seconds. The experiment results are shown in Fig.8-Fig.10. According the results, it is obviously shown that the error of gazing point is smaller than that of end effector. The trackability of eye-vergence visual system is superior to that of fixed-hand-eye visual system. Whtat's more, the errors of gazing point and end effector are enlarged with

the increase of variable number. And the target object can be well tracked with the eye-vergence system when ω is below 1.256rad/s.

5 Conclusion

In this paper, eye-vergence visual servoing controller was developed to track longitudinally moving target object. The rapidly moving object can be kept in sight with eye-vergence system in case that hand-eye system is not flexible enough to track it.

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] Wolfgang Sepp, Stefan Fuchs and Gerd Hirzinger, "Hierarchical Featureless Tracking for Position-Based 6-DoF Visual Servoing", Proceedings of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robotics and Systems (IROS), pp.4310-4315, 2006.

[3] Toshifumi Hiramatsu, Takanori Fukao, Keita Kurashiki, Koichi Osuka "Image-based Path Following Control of Mobile Robots with Central Catadioptric Cameras" IEEE International Conference on Robotics and Automation Kobe, Japan, May 12-17, 2009

[4] Dae-Jin Kim, Ryan Lovelett, and Aman Behal "Eye-in-Hand Stereo Visual Servoing of an Assistive Robot Arm in Unstructured Environments" IEEE International Conference on Robotics and Automation Kobe, Japan, May 12-17, 2009

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

[6] W. Song, M. Minami, S. Aoyagi, "On-line Stable Evolutionary Recognition Based on Unit Quaternion Representation by Motion-Feedforward Compensation", International Journal of Intelligent Computing in Medical Sciences and Image Processing (IC-MED) Vol. 2, No. 2, pp. 127-139 (2007).

[7] 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)

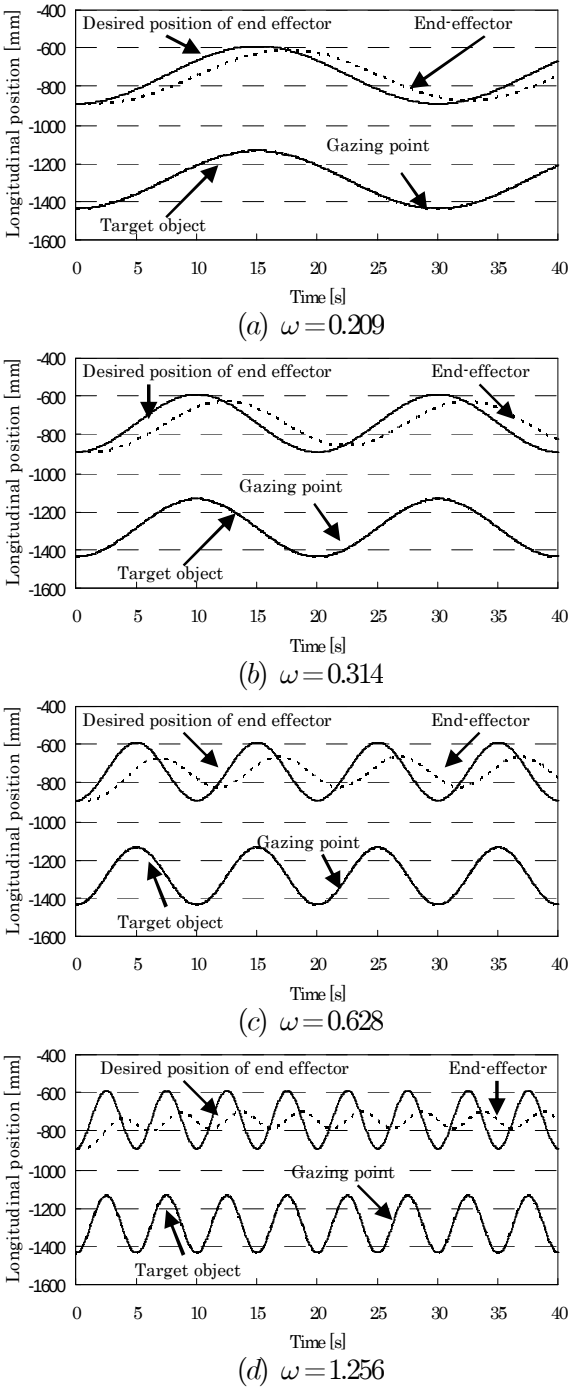
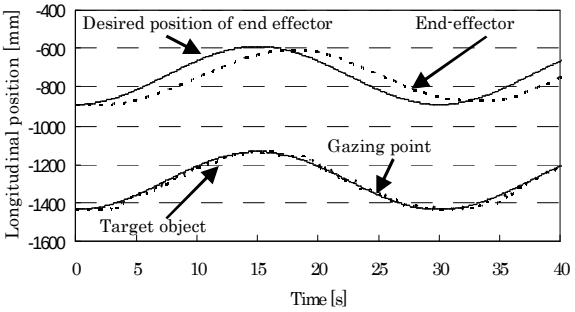
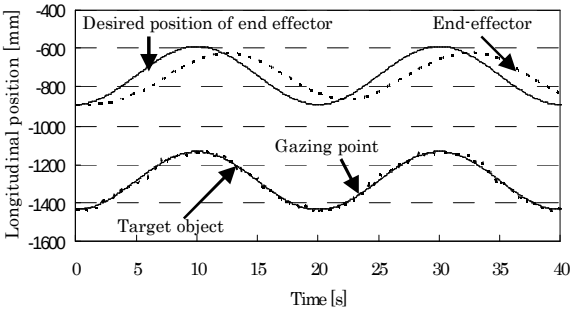


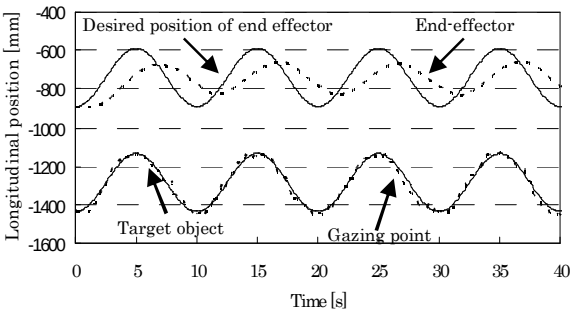
Fig. 8: The position of Gazing point and end-effector when the pose parameters of x, y, z, ε_1, ε_2, ε_3, are set on servoing controller



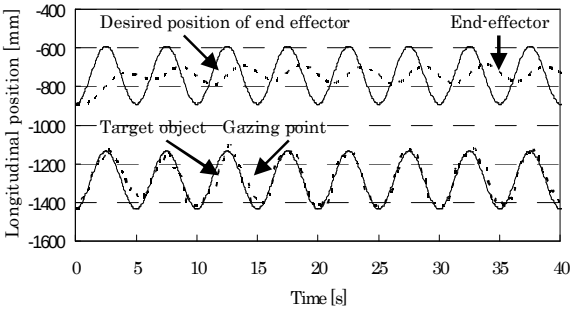
(a) $\omega = 0.209$



(b) $\omega = 0.314$

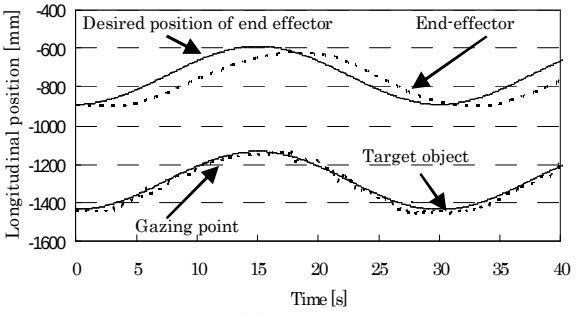


(c) $\omega = 0.628$

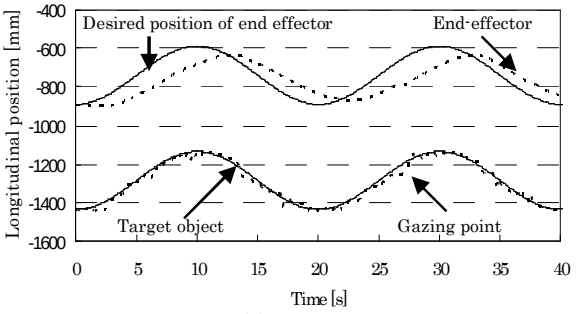


(d) $\omega = 1.256$

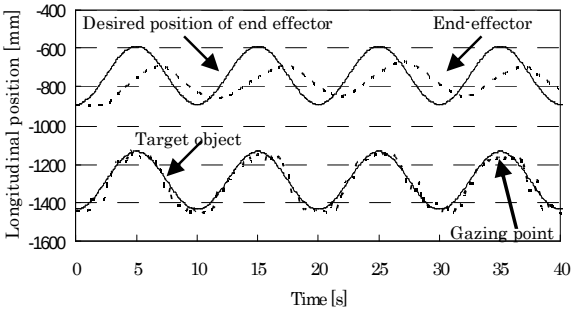
Fig. 9: The position of Gazing point and end-effector when the pose parameters of x, y, z are recognized, and ε_1, ε_2, ε_3, are set on servoing controller



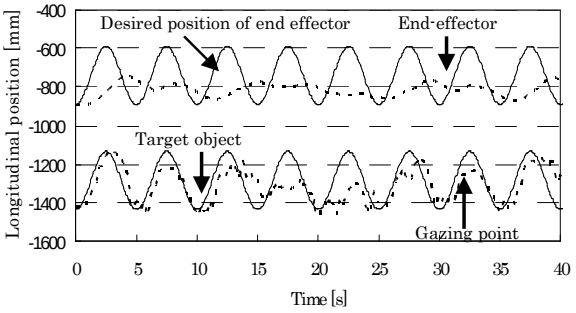
(a) $\omega = 0.209$



(b) $\omega = 0.314$



(c) $\omega = 0.628$



(d) $\omega = 1.256$

Fig. 10: The position of Gazing point and end-effector when all the pose parameters of x, y, z, ε_1, ε_2, ε_3 are recognized