Visual Servoing Frequency Response of Eye-vergence System in Lateral Motion with Evolutionary Pose Tracking of 3D-Object

Hongzhi Tian¹, Yu Cui¹, Mamoru Minami¹, and Akira Yanou¹

¹Okayama University, Japan
(Tel: 81-86-251-8233, Fax: 81-86-251-8233)
¹psnc8ytd@okayama-u.ac.jp

Abstract: Visual servoing towards moving target with hand-eye cameras fixed at the hand is inevitably affected by hand dynamical oscillations, therefore it’s difficult to make target position always at the center of camera’s view, because nonlinear dynamical effects of whole manipulator stand against tracking ability. To overcome the defects of the fixed-hand-eye system, in hand-eye-vergence system left and right cameras’ directions could be rotated to observe and keep the target object to be seen at the center of camera images, reducing the influences of aberration of cameras lens. The dynamical superiorities of eye-vergence system are verified through frequency response experiments. In theory compared with fixed-hand-eye system the proposed eye-vergence system can track objects more promptly and stably. However there is a lack of understanding of the practical application status of the proposed system. Previous research has been stuck in research and improving about position tracking performance of the system. This paper, for the first time, analyzes the performance of 3D-object orientation tracking. The utilized orientation recognition method using quaternion is introduced and the orientation tracking results are shown for a more comprehensive analysis of system performance.

Keywords: Visual servoing, Eye-vergence, Model-based matching, Quaternion

1 INTRODUCTION

Nowadays, in a field of robot vision, a control method called a visual servoing attracts attentions [1]-[4], which can be classified into three major groups: position-based[5], image-based [6], [7] and hybrid visual servoing [8], [9].

The visual servoing, a method for controlling a robot using visual information in the feedback loop, is expected to be able to allow the robot to adapt to changing or unknown environments. Some methods have already been proposed to improve observation abilities, by using stereo cameras [10], multiple cameras [11], and two cameras; with one fixed on the end-effector, and the other one fixed in the workspace [12]. These methods obtain different views to observe an object by increasing the number of cameras.

Recent researches on visual servoing are limited generally in a swath of tracking an object while keeping a certain constant distance [10], [13], [14]. But the final objective of visual servoing seems to lie in approaching the end-effector to a target object and then work on it, like grasping. In this case, the desired relation between cameras and the object is time varying, so such eye-vergence camera system is indispensable to keep suitable viewpoint all the time during the approaching visual servoing, utilizing the changeable cameras’ eye direction so as to look at the target in the center of camera images. This advantage of eye-vergence can be called “kinematical merit”.

It is easy to catch up to the object in case of the object moving slowly, but when the object become moving faster and faster, human’s face can hardly keep position squarely to the object, while human’s eye can still keep staring at the object because of its small mass and inertial moment. This another advantage of eye vergence can be called “dynamical merit”.

These merits of eye-vergence concerning kinematical and dynamical effects is deemed to be important and useful to keep control stability of closed loop using visual feedback. Therefore in this report the merits of eye-vergence visual servoing for tracking have been confirmed experimentally by using eye-vergence function that enables the target to be seen at the center of images, avoiding aberration of lens.

In the previous study [15], the system performance of position tracking have been analyzed detailedly. However for three-dimensional objects, in addition to the position tracking, it is also needed for manipulator taking the orientation tracking. This time we get the orientation tracking data of the system through the experiment in lateral direction. From the results of experiment, it is clear that the proposed eye-vergence system is faster and more stable than fixed-camera system during orientation and position tracking a target with 3D-pose.

2 3D POSE TRACKING METHOD

In this paper, a 3D-ball-object as shown in Fig.1 is taken whose size and color are known as an example of the target object. However, other different kinds of shape targets can also be measured by model-based matching strategy if their
2.1 Model-based matching

In [17] this approach have been described in detail, h we simply introduce. The 3D solid model named $S_0$ rectangular block is shown in Fig.2 (on the top). The of coordinates inside of the dotted line block model is picted as $S_{in}$, which is composed of each surfaces $S_{in,k}(k = 1, 2, \cdots, n)$, the outside strip space enveloping $S_{in}$ is noted as $S_{out}$. Projecting $S_{in}$ and $S_{out}$ onto the 2D coordinates of left camera $\Sigma_{L}$ and right camera $\Sigma_{R}$. The left and right 2-D searching models, named $S_L$ and $S_R$, are shown in Fig.2(on the bottom). Color information is used to search for the target object in the images. Supposing there are distributed solid models in the searching space in $\Sigma_W$, each has its own pose. To determine which solid model is most close to the real target, a correlation function used fitness function in GA is defined for evaluation. Everyone of $S_{in}$ have three small circles. And everyone of $S_{out}$ have three 3 rings. The relative positions of circles and rings are unchanged. Each pair of circle and ring corresponds with a color, and three pairs of circles and rings are corresponding to red, blue and green. The higher coincidence degree between a circle and corresponding color ball is, the higher fitness is. Conversely, the higher coincidence degree between a ring and the corresponding color ball is, lower fitness will be. When the search-

Fig. 5. Block diagram of the hand visual servoing system

Fig. 6. Cameras’ gazing point

the higher coincidence degree between a ring and the corresponding color ball is, lower fitness will be. When the searching model fits to the target object being imaged in the right and left images, then the fitness function gives maximum value. This optimization problem is solved by GA. And the genes of GA representing possible pose solution is defined as,

$$
\begin{align*}
10 \cdots 10 & \quad 11 \cdots 01 \quad 01 \cdots 10 \\
10 \cdots 10 \quad 10 \cdots 10 \quad \cdots \quad 10 \cdots 10
\end{align*}
$$

2.2 On-line Pose Tracking “1-Step GA”

For real-time visual control purposes, GA have been employed in a way denoted as “1-Step GA” evolution [18].

2.3 Orientation recognition method using quaternion

For representing the orientation of 3D object, widely used methods include Euler angles, axis-angle representation and rotation quaternions. The first two methods are easy to
understand. However, because the orientation singularities exist in which we choose the quaternions finally [19]. But the previous research is on the basic of fixed-hand-eye system. So this time during the experiment it is needed to observe whether the orientation recognition method is also applicable to the new proposed system. The definition of unit quaternion is shown in Eq.7. On the basis of axis-angle representation, a unit vector \( \hat{\epsilon} \) indicating the direction, and an angle \( \theta \) describing the magnitude of the rotation about the axis.

\[
q = \{ \eta, \epsilon \} \tag{1}
\]

\[
\epsilon = \sin \frac{\theta}{2} \hat{k}
\]

here,

\[
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\epsilon_3 \\
\end{bmatrix} = \sin \frac{\theta}{2}
\begin{bmatrix}
k_x \\
k_y \\
k_z \\
\end{bmatrix}
\]

\( \eta \) is the scalar part of the quaternion, and \( \epsilon \) is the vector part of the quaternion. They satisfy the following relationship:

\[
\eta^2 + \epsilon^T \epsilon = 1 \tag{4}
\]

![Fig. 7. Definition of quaternion in the proposed system](image)

3 HAND & EYE VISUAL SERVOING CONTROLLER

3.1 Hand Visual Servoing Controller

The block diagram of our proposed hand & eye-vergence visual servoing controller is shown in Fig.5. The hand-visual servoing is the outer loop.

Based on the above analysis of the desired-trajectory generation, the desired hand velocity \( W\hat{r}_d \) is calculated as,

\[
W\hat{r}_d = K_{P_h} W\hat{r}_{E,Ed} + K_{V_h} W\hat{r}_{E,Ed}, \tag{5}
\]

where \( W\hat{r}_{E,Ed} \) and \( E\hat{T}_{Ed} \) can be calculated from \( E\hat{T}_{Ed} \) and \( E\hat{r}_{Ed} \). \( K_{P_h} \) and \( K_{V_h} \) are positive definite matrix to determine PD gain.

The desired hand angular velocity \( W\hat{\omega}_d \) is calculated as,

\[
W\hat{\omega}_d = K_{P_o} W\hat{R}_E D + K_{V_o} W\hat{\omega}_{E,Ed}, \tag{6}
\]

where \( E\Delta\epsilon \) is a quaternion error [20] calculated from the pose tracking result, and \( W\hat{\omega}_{E,Ed} \) can be computed by transforming the base coordinates of \( E\hat{T}_{Ed} \) and \( E\hat{r}_{Ed} \) to \( \Sigma_W \). Also, \( K_{P_r} \) and \( K_{V_r} \) are suitable feedback matrix gains. The desired hand pose is defined as \( W\phi_d = [W\hat{r}_d, W\hat{\omega}_d]^T \).

The desired joint variable \( q_{Ed} = [q_1, \ldots, q_7]^T \) and \( \hat{q}_{Ed} \) is obtained by

\[
q_{Ed} = f^{-1}(W\phi_d) \tag{7}
\]

\[
\hat{q}_{Ed} = J^+(q) \begin{bmatrix} W\hat{r}_d \\ W\hat{\omega}_d \end{bmatrix} \tag{8}
\]

where \( f^{-1}(W\phi_d) \) is the inverse kinematic function and \( J^+(q) \) is the pseudo-inverse matrix of \( J_E(q) \), and \( J_E(q) = \frac{1}{2}(J_E J_E)^{-1} \). The manipulator is 7 links, and the end-effector has 6-DoF, so \( q_1 \) is made 0 to solve the redundancy problem. Using the inverse kinematics it can make the joint of angles approximately as the desired joint angles. The formula of the desired joint angles was defined in the new controller as

\[
q_{Ed} = k_p(q_{Ed} - q_E) + J^+_E(q) \begin{bmatrix} W\hat{r}_d \\ W\hat{\omega}_d \end{bmatrix} \tag{9}
\]

where \( k_p \) is P positive gain.

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

\[
\tau = K_{SP}(q_S - q) + K_{SD}(q_S - \hat{q}) \tag{10}
\]

where \( K_{SP} \) and \( K_{SD} \) are symmetric positive definite matrices to determine PD gain.

3.2 Eye-vergence Visual Servoing Controller

The eye-vergence visual servoing is the inner loop of the visual servoing system shown in Fig.5. 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, \( q_S \) 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.3 (a) and (b), \( E\hat{r}_{M}, E\hat{r}_{M}, E\hat{z}_M \) express position of the detected object in the end-effector coordinate. The desired angle of the camera joints are calculated by:

\[
q_{sd} = atan2(E\hat{y}_M, E\hat{z}_M) \tag{11}
\]

\[
q_{ad} = atan2(\hat{l}_R, E\hat{r}_{M}, E\hat{z}_M) \tag{12}
\]

\[
q_{ld} = atan2(\hat{l}_L, E\hat{r}_{M}, E\hat{z}_M) \tag{13}
\]

where \( \hat{l}_L = 120 [\text{mm}] \) is the camera location.
For every group, shown in Fig. 4 (a) and (b). And the 3D marker is shown CT-3001, receiving the image from the CCD camera is const stereo cameras is set as 30fps. The image processing board, manufactured by Sony Industries. The frame frequency of manufactured by Mitsubishi Heavy Industries. Two rotat visual servoing system through real robot PA-10 robot arm-4

EXPERIMENT OF HAND & EYE-VERGENCE VISUAL SERVOING

4.1 Experimental system

To verify the effectiveness of the hand & eye visual servoing system through real robot PA-10 robot arm manufactured by Mitsubishi Heavy Industries. Two rotatable cameras mounted on the end-effector are FCB-1X11A manufactured by Sony Industries. The frame frequency of stereo cameras is set as 30fps. The image processing board, CT-3001, receiving the image from the CCD camera is connected to the host computer (CPU: Intel Core i7-3770, 3.40 GHz).

The structure of the manipulator and the cameras are shown in Fig. 4 (a) and (b). And the 3D marker is shown in Fig. 1. The coordinate of the target object and the manipulator in experiment are shown in Fig. 8.

Firstly, an experiment was made in which true object’s, \( x, \ y, \ z, \ \varepsilon_1, \ \varepsilon_2 \) and \( \varepsilon_3 \), are assumed to be given to servoing controller. Then it was carried out that 3 groups of experiments of frequency response. In these experiments, It was made that \( x \)-position, 3-Dof position, and 6-Dof position/orientation are recognized by the cameras respectively. For every group, \( \omega=0.314, \ \omega=0.628, \) and \( \omega=1.256 \) is set separately, which are angular velocities of the object.

4.2 Experiment condition

\( E_0, \ M_0 \) and \( E_5 \) represent the initial hand pose, the initial object pose and the midpoint of round-trip tracking move ments of hand respectively. Therefore their coordinate systems are defined as \( \Sigma_{E_0}, \ \Sigma_{E_5}, \) and \( \Sigma_{M_0} \) separately. The homogeneous transformation matrix from \( \Sigma_W \) to \( \Sigma_{E_0} \) and \( \Sigma_{M_0} \) are:

\[
\begin{align*}
W_{TE_0} &= \begin{bmatrix}
0 & 0 & -1 & -690[mm] \\
1 & 0 & 0 & 0[mm] \\
0 & -1 & 0 & 485[mm] \\
0 & 0 & 0 & 1
\end{bmatrix} \\
W_{TM_0} &= \begin{bmatrix}
0 & 0 & -1 & -1235[mm] \\
1 & 0 & 0 & -150[mm] \\
0 & -1 & 0 & 585[mm] \\
0 & 0 & 0 & 1
\end{bmatrix}
\end{align*}
\]

The target object move according to the following time function

\[
M_{z_M}(t) = 150 - 150 \cos(\omega t)[mm]
\]

Target position and orientation relationship between the object and the end-effector is set as:

\[
E_d \psi_M = [0, -100[mm], 545[mm], 0, 0, 0]
\]

The object is subjected to reciprocating motion of the sine wave in orbit. Pose relationship of the coordinate system of the object and the visual servoing system is shown in Fig. 8.

4.3 Symbol meaning

\( M \) represents the object and \( \hat{M} \) represents the estimated object. Then \( \hat{\Sigma}_M \) denotes the coordinate system that moves along with the object. The relationship between coordinate systems such as the actual pose of the hand \( \hat{\Sigma}_E \) or the recognized pose of the object \( \hat{\Sigma}_M \) which is viewed from the \( x \)-\( z \) plane of the center coordinate system \( \Sigma_{E_5} \) is shown in Fig. 9. \( \hat{\Sigma} \) represents that a coordinate system is running when it is viewed from the world coordinate system \( \Sigma_W \). The coordinate system represented by \( \hat{\Sigma} \) keeps fixed in the world coordinate system \( \Sigma_W \). In other words \( \hat{\Sigma}_E, \ \hat{\Sigma}_{E_5}, \ \hat{\Sigma}_M \) and \( \hat{\Sigma}_{M_0} \) are all moving in the world coordinate system \( \Sigma_W \). On the other hand \( \Sigma_{E_0}, \ \Sigma_{E_5}, \) and \( \Sigma_{M_0} \) keeps fixed in the world coordinate system \( \Sigma_W \). The motion of object \( M \), hand \( \hat{E} \) and gazing point \( \hat{M} \) in the \( x \)-direction of \( \Sigma_{E_5} \) are represented by \( E_{x} x_{M}, \ E_{\hat{x}} x_{E}, \) and \( E_{\hat{x}} x_{M} \). And as shown in Fig. 11.

\[
\Delta s_{E_5} = E_{\hat{x}} s_{E_5} - E_{x} s_{E_5}, (s = x, y, z)
\]
\[
\Delta x_{M} = E_{\hat{x}} x_{M} - E_{x} x_{M}
\]

represents a follow-up error of the hand and the gaze point respectively. And \( E_{\hat{y}} y_{E_5} \) is the target position of the hand in the \( y \)-axis direction of the center coordinate system \( \Sigma_{E_5} \). Furthermore

\[
\Delta u_{E_5} = E_{\hat{y}} u_{E_5} - E_{y} u_{E_5}, (u = y, z)
\]

represents the tracking error between the actual position and the target position of the hand in \( y \)-axis direction or \( z \)-axis.
4.4.2 Position tracking result and analysis of the tracking experiment

Object is reciprocating on the trajectory in lateral direction. Object and the system are shown from the x-z plane of \( \Sigma_E \) of hand. Initial position of the object \( \Sigma_M \), actual object \( \hat{\Sigma}_M \), detected object \( \hat{\Sigma}_M \), initial position of the hand \( \Sigma_E \), actual end effector \( \hat{\Sigma}_E \) and theoretical end effector \( \hat{\Sigma}_E \) viewed from \( \Sigma_E \). As it is shown in Fig.6, the intersection of both cameras’ gazing directions is defined as the gazing point of cameras to examine trackability of the eye-vergence system. Because the gazing point has been calculated on the basis of the recognition result of the object by the 1-step GA, recognition error is included in the Gazing point.

4.4 Experiment results

4.4.1 Relation between position diagram and real machine

Fig.10 shows the positional relationship between the hand and the object in the condition that the tracking all the six position and orientation variables are recognized. And the motion period of the object is \( T = 20 [s] \). Movement trajectory of the object \( M \), hand \( E \) and gazing point \( \hat{M} \) are represented by dashed line, dotted line and solid line respectively. At the time of (a) and (c), since the moving velocity of the object is fast, hand is not able to track the object. Since the tracking state of hand is same as fixed camera system, the dotted line indicated by \( E_c \) in the figure also represents the movement of fixed camera system. At this time, it is clear that the distance between the hand \( E_c \) and the object \( E_c \) on the x-axis direction is farther than that between the gazing point \( E_M \) and target object \( E_c \) of the camera. From the error between the gazing point \( E_M \) and the object \( E_c \) it can be seen that it is easier for eye-vergence system to track the object than the fixed camera system.

4.4.2 Position tracking result and analysis of the tracking experiment

Because the object is reciprocating in the x direction, this time only the result of tracking at the x-axis is given and analyzed as shown in Fig.11. And at this time the movement cycle is 10 seconds (\( \omega = 0.628 \)). As shown in Fig.11 when the cycle is 10 seconds it is clear that the motion of hand

4.4.3 Orientation tracking result and analysis of the tracking experiment

Orientation tracking result of the detected object and hand are shown in Fig.12. Same as the tracking status of x, y and z the quaternion variation of detected object is more frequent than that of hand. Since the camera mass is smaller than manipulator, so the moment of inertia is also smaller than that of manipulator. Therefore it can perform faster adjustment than manipulator. The phases of three variation curves of quaternion of the detected object are all earlier than that of the hand. That result is consistent with the control procedure. Compared with the fixed-camera system, giving the camera freedom can make the camera more quickly track the object during its transform of the orientation. And tracking error of each step in 1-Step GA will not so large that manipulator can track the object stably.

5 CONCLUSION

In this paper the eye-vergence visual servoing controller of eye-vergence system have been described in detail. And for the first time by tracking experiment to the 3D marker with 6 degree of freedoms it has been analyzed that the recognition and control results of both of the orientation and position. Finally it is confirmed that not only position but also orientation trackability of the eye-vergence system have superior performance is and are better than that of the fixed-camera system.

REFERENCES


Fig. 11. Movements of actual object $\mathbf{M}_x$, detected object $\hat{\mathbf{M}}_x$ and end effector $\hat{\mathbf{E}}$ on the x, y and z directions in the center coordinate system of hand $\Sigma_{E}$. The object’s pose $\mathbf{x}, \mathbf{y}, \mathbf{z}, \varepsilon_1, \varepsilon_2, \varepsilon_3$ are recognized by camera.


