High Tracking Experimental Performances by Approaching Hand/Eye-Vergence Visual Servoing

Wei Song, Mamoru Minami, Fujia Yu, Akira Yanou and Mingcong Deng
Wei Song is with University of Shanghai, Yanchang Road 149 Shanghai China, songwei5726@hotmail.com.
Mamoru Minami, Fujia Yu, Akira Yanou and Mingcong Deng are with Graduate School of Natural Science and Technology Okayama University, Tsushimanaka3-1-1 Okayama JAPAN, { minami, yufujia, yanou, deng } @suri.sys.okayama-u.ac.jp
.

***Abstract*—In this paper, we focus on how to control the robot's end-effector to track an object, meanwhile, to approach it with a desired posture for grasping, which we named as "Approaching Visual Servoing (AVS)." AVS with binocular cameras needs inheritably eye-vergence motion to keep cameras' sight against the target since approaching motion makes the cameras sight narrow or losing the object. We propose a hand & eye-vergence dual control system to perform AVS. The experiment of full 6-DoF Approaching Visual Servoing experiments to a moving object by a 7-link manipulator installed with a binocular camera system, confirmed the ability of Hand & Eye-vergence control system.**

I. Introduction

Tasks in which visual information are used to direct a manipulator toward a target object are referred to visual servoing [1]. Visual servoing is usually performed to keep a fixed relation with respect to a static or moving object [2], [3]. This kind of visual servoing deals with 3-D pose tracking problem, that is, the robot follows the motion of the target object to keep a fixed given relative pose between the end-effector and the target object ${}^{E}\boldsymbol{T}_{M}$, as shown in Fig.1, and it can be said as "Pose-Regulator Visual Servoing."

We think it is difficult to find papers that confirmed experimentally the full 6-pose servoing ability of position and orientation to moving target object. Most researchers concentrate only on the control of the position of the end effector, while ignoring the orientation control [4]. When trying full pose control, the performances were evaluated by simulations without real experiment [5]. Also in most researches, the target object is static [6]. So we think there are few researches about experimental evaluation of full pose visual servoing.

Besides tracking an object, robots are expected to do more intelligent tasks, a possible task is to grasp a moving object. To complete these kinds of tasks, we have to do two works: one is to estimate the 6-D information (both position and orientation) of the object, the other is to approach the object with a desired grasping posture, while approaching it. We think this kind of visual servoing is necessary in actual application as catching the moving target. We think it has not yet been well discussed. Fig.2 shows an image of how the robot track the object and meanwhile approach it, which we named as "Approaching Visual Servoing (AVS)." In the AVS, the relative relation between the end-effector and the target object is time varying,

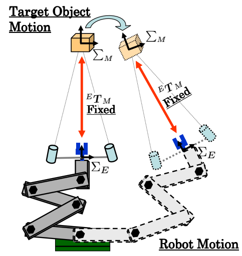

Fig. 1. Pose Regulator Visual Servoing

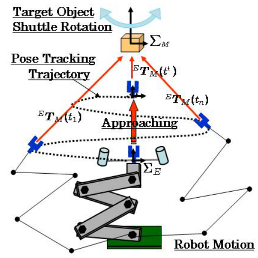

Fig. 2. Approaching Visual Servoing

defined as ${}^{E}\boldsymbol{T}_{M}(t)$. In Fig.2, the motion of the target object is shuttle rotation, the end-effector keep approaching the object through a curved pose tracking trajectory given by ${}^{E}\boldsymbol{T}_{M}(t_1)$, $\cdots$, ${}^{E}\boldsymbol{T}_{M}(t_n)$, $\cdots$, ${}^{E}\boldsymbol{T}_{M}(t^{\star})$, finally it gets near to the object, and then it is possible to grasp the object.

To grasp an object an eye-to-hand system is used in [7]. Since the camera is fixed in work space, the hand motion does not affect the pose estimation, which decouples the stability of visual servoing motion and pose estimation. On the other hand the observing function may also be hazarded, for example, the sight of the camera may be obstructed by the hand when grasping a target. So we think hand-eye visual servoing system is more useful than the system with cameras being static to the floor, for the reason that the hand-eye configuration extends the robot's observable space, which will provide much more visual information to enhance the dexterity of the robot's operation. Moreover, it will improve the adaptive ability, "searching and looking" of the robot. So we use a hand-eye configuration, having two cameras mounted on the robot's end-effector.

In the case of using fixed-hand-eye configuration in AVS, a problem of sight restriction should be discussed first, that is, the possible searching space becomes narrow when the camera approaches to the object (Fig.3(a)), what is worse, a

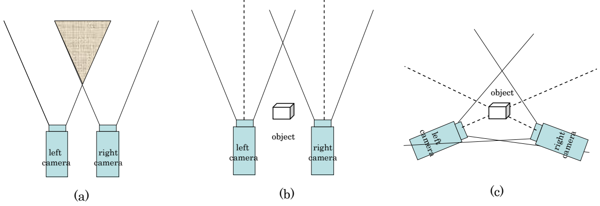

Fig. 3. Sight of fixed camera system and eye-vergence system

part of the object or some feature points possibly get out of the image (Fig.3(b)). Thus it is necessary to change the pose of the camera to enlarge the possible searching space in the case that the camera is close to the target, as Fig.3(c). With a hand-eye configuration, another problem exists, that is the dynamics of the manipulator will deteriorate the on-line hand-eye pose estimation, since the camera's oscillation produces a false motion of the target object in the camera image even though the target is stopping in the task space. We call the false motion as "fictional motion." Here we had been interested in how to compensate such a fictional motion of the target object. Then we had a proposed robust recognition method, called motion-feedforward (MFF) compensation method [**?**], which can provide a pose estimation accuracy with a same precision level of fixed-camera-configuration since it can compensate the influence to pose estimation of dynamical hand-eye motion. We utilized this MFF technique for AVS in this report as a stable pose estimation. A mobile hand-eye stereo camera of our experimental system that can change the pose of the camera to focus on a target object is shown in Fig. 4. This camera system has 3-DoF, one is for pan rotation of the left camera, one is for pan rotation of the right camera, and the other is the tilt rotation of both cameras. Installing this mobile stereo camera system in the end-effector of a 7-link manipulator, we built up a hand & eye-vergence dual visual servoing system. In this presentation, we focus on an experimental evaluation of full 6-DoF approaching visual servoing to moving target without a priori knowledge of the target's motion, with the target shape as a known information to the servoing system. We think it is worth to examine the possibility whether visual servoing technologies can grasp a moving target object in a space by a robot with eye-in-hand configuration. In this presentation we will show the data of our experiment, confirming that, the position error is less than 20[mm], orientation error is less than 3[deg], while approaching to a moving target, so we can get the conclusion that: proposed system can approach with the eye keeping a good observation and with the hand errors maintaining within possible-catching extent, representing that proposed AVS method may possibly be able to catch a target moving in space with full 3-D unknown swinging motion.

II. On-line Evolutionary Recognition

We use a model-based matching method to recognize an object. Different kinds of targets can be measured by this strategy if their shape is given.

Firstly, we give the definitions of coordinate systems used in this paper. World coordinate frame is defined as Σ_W, the end-effector's frame is Σ_E, left/right camera coordinate systems is Σ_{CL} / Σ_{CR} and target coordinate frame is Σ_M.

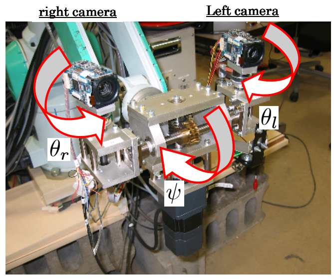

Fig. 4. Eye-vergence stereo camera system

A. GA-based On-line Recognition

The theoretically optimal pose $\boldsymbol{\psi}^{max}(t)$ that gives the highest peak of $F(\boldsymbol{\psi}(t))$ is defined as

$$\boldsymbol{\psi}^{max}(t) = \left\{\boldsymbol{\psi}(t) \;\middle|\; \max_{\boldsymbol{\psi}\in\boldsymbol{L}} F(\boldsymbol{\psi}(t))\right\}, \tag{1}$$

where $\boldsymbol{L}$ represents 6-DoF searching area of $x, y, z, \epsilon_1, \epsilon_2, \epsilon_3$.

Here we use GA to search $\boldsymbol{\psi}^{max}(t)$. The individual of GA is defined as $\boldsymbol{\psi}_{i,j}(t)$, which means the i-th gene ($i = 1, 2, \cdots, p$) in the j-th generation. Denote $\boldsymbol{\psi}_{ga}^{max}(t)$ as the highest peak in GA process,

$$\boldsymbol{\psi}_{ga}^{max}(t) = \left\{\boldsymbol{\psi}_{i,j}(t) \;\middle|\; \max_{\boldsymbol{\psi}_{i,j}\in\boldsymbol{L}} F(\boldsymbol{\psi}_{i,j}(t))\right\}. \tag{2}$$

In fact we cannot always guarantee the best individual of GA $\boldsymbol{\psi}_{ga}^{max}(t)$ correspond to the theoretically optimal pose $\boldsymbol{\psi}^{max}(t)$, because the number of GA's individuals is limited. The difference of $\boldsymbol{\psi}^{max}(t)$ and $\boldsymbol{\psi}_{ga}^{max}(t)$ is denoted as

$$\delta\boldsymbol{\psi}(t) = \boldsymbol{\psi}^{max}(t) - \boldsymbol{\psi}_{ga}^{max}(t). \tag{3}$$

And the difference of $F(\boldsymbol{\psi}^{max}(t))$ and $F(\boldsymbol{\psi}_{ga}^{max}(t))$ is denoted as

$$\Delta F(\delta\boldsymbol{\psi}(t)) = F(\boldsymbol{\psi}^{max}(t)) - F(\boldsymbol{\psi}_{ga}^{max}(t)) \geq 0. \tag{4}$$

Here, we present two assumptions.

[Assumption 1]: Assuming that $F(\boldsymbol{\psi}(t))$ distribution satisfies $\Delta F(\delta\boldsymbol{\psi}(t)) = 0$ if and only if $\delta\boldsymbol{\psi}(t) = 0$.

[Assumption 2]: Assuming that $\dot{F}(\boldsymbol{\psi}_{ga}^{max}(t)) > \dot{F}(\boldsymbol{\psi}^{max}(t))$, which indicates that the convergence speed to the target in the dynamic images should be faster than the changing speed of the dynamic $F(\boldsymbol{\psi}(t))$ distribution as time t varying.

From [Assumption 2], we have

$$\Delta\dot{F}(\delta\boldsymbol{\psi}(t)) = \dot{F}(\boldsymbol{\psi}^{max}(t)) - \dot{F}(\boldsymbol{\psi}_{ga}^{max}(t)) < 0. \tag{5}$$

In the case the object is static, we have $\dot{F}(\boldsymbol{\psi}^{max}(t)) = 0$. Then [Assumption 2] becomes $\dot{F}(\boldsymbol{\psi}_{ga}^{max}(t)) > 0$, which indicates GA is assumed to be able to converge to a optimum value.

These two assumptions depend on some factors such as object's shape, object's speed, definition of $F(\boldsymbol{\psi}(t))$, parameters of GA and viewpoint for observing. We could set such an environment to satisfy or close to the above two assumptions. When above two assumptions are satisfied, (4) and (5) will be satisfied, then $\Delta F(\delta\boldsymbol{\psi}(t))$ is so-called Lyapunov function. That means $\Delta F(\delta\boldsymbol{\psi}(t))$ will be gradually decreased to 0. Thus, from the above definitions, we have $\delta\boldsymbol{\psi}(t)\rightarrow 0$, which means gradual stability in searching space $\boldsymbol{L}$, that is

$$\boldsymbol{\psi}_{ga}^{max}(t)\rightarrow\boldsymbol{\psi}^{max}(t), (t\rightarrow\infty) \tag{6}$$

Let t_ϵ denotes a convergence time, then

$$|\delta\boldsymbol{\psi}(t)| = |\boldsymbol{\psi}^{max}(t) - \boldsymbol{\psi}_{ga}^{max}(t)|\leq\epsilon, \quad (\epsilon > 0, t > t_\epsilon) \tag{7}$$

In (7), ϵ is tolerable extent that can be considered as a observing error. Thus, it is possible to realize real-time optimization, because $\boldsymbol{\psi}_{ga}^{max}(t)$ is or near to the theoretically optimal $\boldsymbol{\psi}^{max}(t)$ after t_ϵ. Notice that the detected pose of the object, $\boldsymbol{\psi}_{ga}^{max}$, is the abbreviation of ${}^{E}\boldsymbol{\psi}_{ga}^{max}$, which is based on the hand coordinate Σ_E.

Above discussion is under the condition of varying time. Here, when we consider evolution time of each generation of GA denoted by Δt. The GA's evolving process is described as

$$\boldsymbol{\psi}_{i,j}(t) \stackrel{\text{evolve}}{\longrightarrow} \boldsymbol{\psi}_{i,j}(t+\Delta t). \tag{8}$$

Obviously, this evolution time Δt will be possible to generate somewhat bad influence. If we assume that this bad influence on $\delta\boldsymbol{\psi}(t)$ can be described as

$$|\delta\boldsymbol{\psi}(t)|\leq\epsilon', \quad (\epsilon' > \epsilon > 0), \tag{9}$$

then, it can be considered Δt can manage real-time optimal solution. In (9), ϵ' is also tolerable extent as a observing error and it is somewhat larger than ϵ. Since the GA process is executed only one time to output the semi-optimal $\boldsymbol{\psi}_{ga}^{max}(t)$, we named this on-line recognition method as "1-step GA".

We have confirmed that the above time-variant optimization problem could be solved by 1-step GA through several experiments [**?**], [**?**]. $\boldsymbol{\psi}_{ga}^{max}(t)$ will be output as the measurement result in each generation to control the robot manipulator. We define

$$\hat{\boldsymbol{\psi}}(t) = \boldsymbol{\psi}_{ga}^{max}(t), \hat{\boldsymbol{\psi}} = [\hat{x}, \hat{y}, \hat{z}, \hat{\epsilon_1}, \hat{\epsilon_2}, \hat{\epsilon_3}]^T. \tag{10}$$

III. Motion-Feedforward (MFF) Compensation

The target coordinate system is represented as Σ_M. Since solid models used to search for the target object are located in the end-effector's coordinate Σ_E, here we discuss the changing of ${}^{E}\boldsymbol{\psi}_M$ based on the changing of ${}^{W}\boldsymbol{\psi}_M$ and the configuration of the robot determined by $\boldsymbol{q}$. Such a relation will be described by the following mathematical function, which can distinguish these two affected motions clearly.

$$\begin{aligned} {}^{E}\dot{\boldsymbol{\psi}}_M &= \begin{bmatrix} {}^{E}\dot{\boldsymbol{r}}_M \\ {}^{E}\dot{\boldsymbol{\epsilon}}_M \end{bmatrix} \\ &= \boldsymbol{J}_M(\boldsymbol{q}, {}^{E}\boldsymbol{\psi}_M)\dot{\boldsymbol{q}} + \boldsymbol{J}_N(\boldsymbol{q}){}^{W}\dot{\boldsymbol{\psi}}_M. \end{aligned} \tag{11}$$

Here $\dot{\boldsymbol{q}}$ is the angular velocity of the joints of the manipulator. The matrix $\boldsymbol{J}_M$ in (11) which describes how target pose change in Σ_E with respect to changing of the manipulator's joint angles can be calculated from the relative position and orientation between the Σ_M, Σ_E and Σ_W. The matrix $\boldsymbol{J}_N$ in (11) which describes how target pose change in Σ_E with respect to the pose changing of itself in real world can be calculated by the rotation matrix from Σ_E to Σ_W. Please refer to [**?**] for a detailed introduction of MFF. Since the effect on the recognition from the dynamics of manipulator can be compensated, recognition by hand-eye cameras will be independent of the dynamics of the manipulator, robust recognition can be obtained just like using fixed cameras.

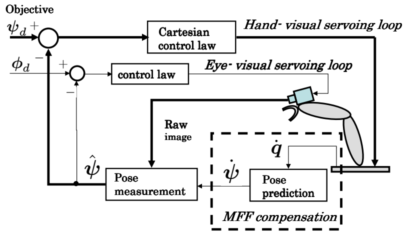

Fig. 5. Block diagram of the hand & eye-vergence visual servoing system

IV. Hand & Eye-vergence Visual Servoing

The block diagram of our proposed hand & eye-vergence dual control system is shown in Fig. 5, which includes two loops. An outer loop for conventional visual servoing that direct a manipulator toward a target object , named as hand visual servoing control, and an inner loop for active motion of binocular camera for accurate and broad observation of the target object, named as eye-vergence visual servoing control.

A. Hand Visual Servoing Controller

Firstly, we explain how to generate the desired hand trajectory. The desired relative relationship of Σ_M and Σ_E is given by Homogeneous Transformation as ${}^{Ed}\boldsymbol{T}_M(t)$, the difference of the desired camera pose Σ_{Ed} and the current camera pose Σ_E is denoted as ${}^{E}\boldsymbol{T}_{Ed}$. ${}^{E}\boldsymbol{T}_{Ed}$ can be described by

$${}^{E}\hat{\boldsymbol{T}}_{Ed}(t) = {}^{E}\hat{\boldsymbol{T}}_M(t) \ {}^{Ed}\boldsymbol{T}_M^{-1}(t), \tag{12}$$

Notice that (12) is a general deduction that satisfies arbitrary object motion ${}^{W}\boldsymbol{T}_M(t)$ and arbitrary objective of visual servoing ${}^{Ed}\boldsymbol{T}_M(t)$.

Differentiating (12) with respect to time yields

$${}^{E}\dot{\boldsymbol{T}}_{Ed}(t) = {}^{E}\dot{\hat{\boldsymbol{T}}}_M(t){}^{Ed}\boldsymbol{T}_M^{-1}(t) + {}^{E}\hat{\boldsymbol{T}}_M(t){}^{Ed}\dot{\boldsymbol{T}}_M^{-1}(t). \tag{13}$$

Here, ${}^{Ed}\boldsymbol{T}_M(t)$, ${}^{Ed}\dot{\boldsymbol{T}}_M(t)$ are given as the desired visual servoing objective. ${}^{E}\hat{\boldsymbol{T}}_M(t)$ is measured by cameras using the on-line recognition method proposed in Section II and MFF compensation in Section III. ${}^{E}\dot{\hat{\boldsymbol{T}}}_M(t)$ is calculated by

$${}^{E}\dot{\hat{\boldsymbol{T}}}_M(t) = ({}^{E}\hat{\boldsymbol{T}}_M(t) - {}^{E}\hat{\boldsymbol{T}}_M(t-\Delta t))/\Delta t, \tag{14}$$

which is output periodically with a time of Δt regardless the object is moving or not. Notice that (14) can not be used to calculate ${}^{E}\dot{\hat{\boldsymbol{T}}}_M(t)$ because it assumes the object is stationary.

Based on the above analysis of the desired-trajectory generation, the desired hand velocity ${}^{W}\dot{\boldsymbol{r}}_d$ is calculated as,

$${}^{W}\dot{\boldsymbol{r}}_d = \boldsymbol{K}_{P_p}{}^{W}\boldsymbol{r}_{E,Ed} + \boldsymbol{K}_{V_p}{}^{W}\dot{\boldsymbol{r}}_{E,Ed}, \tag{15}$$

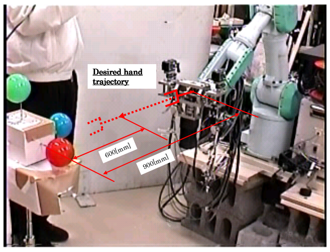

Fig. 6. A photograph of approaching visual servo system.

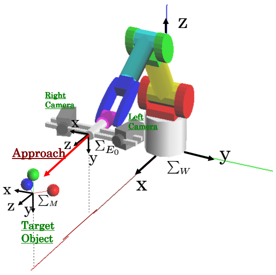

Fig. 7. Coordinate system of approaching visual servo system.

where ${}^{W}\boldsymbol{r}_{E,Ed}$, ${}^{W}\dot{\boldsymbol{r}}_{E,Ed}$ are given by transforming ${}^{E}\boldsymbol{T}_{Ed}$ and ${}^{E}\dot{\boldsymbol{T}}_{Ed}$ from Σ_E to Σ_W. $\boldsymbol{K}_{P_p}$ and $\boldsymbol{K}_{V_p}$ are positive definite matrix to determine PD gain.

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

$$ {}^{W}\boldsymbol{\omega}_d = \boldsymbol{K}_{P_o}{}^{W}\boldsymbol{R}_E{}^{E}\Delta\boldsymbol{\epsilon} + \boldsymbol{K}_{V_o}{}^{W}\boldsymbol{\omega}_{E,Ed}, \tag{16} $$

where ${}^{E}\Delta\boldsymbol{\epsilon}$ is the quaternion error that from the recognition result directly, and ${}^{W}\boldsymbol{\omega}_{E,Ed}$ can be calculated by transforming ${}^{E}\boldsymbol{T}_{Ed}$ 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.

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

$$ \dot{\boldsymbol{q}}_d = \boldsymbol{J}^{+}(\boldsymbol{q}) \begin{bmatrix} {}^{W}\dot{\boldsymbol{r}}_d \\ {}^{W}\boldsymbol{\omega}_d \end{bmatrix}. \tag{17} $$

where $\boldsymbol{J}^{+}(\boldsymbol{q})$ is the pseudoinverse matrix of $\boldsymbol{J}(\boldsymbol{q})$, and $\boldsymbol{J}^{+}(\boldsymbol{q}) = \boldsymbol{J}^{T}(\boldsymbol{J}\boldsymbol{J}^{T})^{-1}$. The hardware control system of the velocity-based servo system of PA10 is expressed as

$$ \boldsymbol{\tau} = \boldsymbol{K}_{SP}(\dot{\boldsymbol{q}}_d - \dot{\boldsymbol{q}}) + \boldsymbol{K}_{SI}\int_0^t (\dot{\boldsymbol{q}}_d - \dot{\boldsymbol{q}})dt \tag{18} $$

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

B. 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, we use pan-tilt stereo camera for eye-vergence visual servoing. Here, the positions of cameras are supposed to be fixed. The left and right camera's poses are defined by $\boldsymbol{\phi}_L = [\theta_l, \psi]^T$, $\boldsymbol{\phi}_R = [\theta_r, \psi]^T$, where θ_l and θ_r are pan angles, and ψ is title angle that is common for both cameras.

Since the object's measurement result $\hat{\boldsymbol{\psi}}$ is described in Σ_E, it can be transformed to Σ_{CL} and Σ_{CR} by Homogeneous

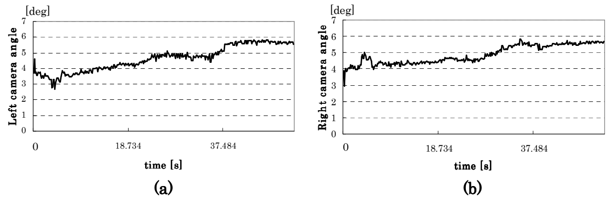

Fig. 9. Camera pose of approaching visual servoing in z-axis.

Transformations as,

$$ {}^{CL}\hat{\boldsymbol{T}}_M({}^{CL}\hat{\boldsymbol{\psi}}) = {}^{CL}\boldsymbol{T}_E(\boldsymbol{\phi}_L){}^{E}\hat{\boldsymbol{T}}_M(\hat{\boldsymbol{\psi}}(t)), \tag{19} $$

$$ {}^{CR}\hat{\boldsymbol{T}}_M({}^{CR}\hat{\boldsymbol{\psi}}) = {}^{CR}\boldsymbol{T}_E(\boldsymbol{\phi}_R){}^{E}\hat{\boldsymbol{T}}_M(\hat{\boldsymbol{\psi}}(t)). \tag{20} $$

In previous research, we investigated that observing the object through both centers of left and right cameras gave the maximum observability. Based on this, the objective of the eye-visual servoing is given by

$$ {}^{CL}\boldsymbol{u}_d = [{}^{CL}x_d, {}^{CL}y_d]^T = \boldsymbol{0}, \quad {}^{CR}\boldsymbol{u}_d = [{}^{CR}x_d, {}^{CR}y_d]^T = \boldsymbol{0}. \tag{21} $$

We define ${}^{CL}\hat{\boldsymbol{u}}$ is the x and y direction of ${}^{CL}\hat{\boldsymbol{\psi}}$, and ${}^{CR}\hat{\boldsymbol{u}}$ is the x and y direction of ${}^{CR}\hat{\boldsymbol{\psi}}$, then the controller of eye-visual servoing is given by

$$ \dot{\boldsymbol{\phi}}_L = \boldsymbol{K}_{P_L}({}^{CL}\boldsymbol{u}_d - {}^{CL}\hat{\boldsymbol{u}}) + \boldsymbol{K}_{D_L}({}^{CL}\dot{\boldsymbol{u}}_d - {}^{CL}\dot{\hat{\boldsymbol{u}}}), \tag{22} $$

$$ \dot{\boldsymbol{\phi}}_R = \boldsymbol{K}_{P_R}({}^{CR}\boldsymbol{u}_d - {}^{CR}\hat{\boldsymbol{u}}) + \boldsymbol{K}_{D_R}({}^{CR}\dot{\boldsymbol{u}}_d - {}^{CR}\dot{\hat{\boldsymbol{u}}}), \tag{23} $$

where $\boldsymbol{K}_{P_L}$, $\boldsymbol{K}_{D_L}$, $\boldsymbol{K}_{P_R}$, $\boldsymbol{K}_{D_R}$ are positive control gain.

V. Experiment of Approaching Visual Servoing by Hand & Eye-Vergence Dual Control System

The visual servoing described in this paper is that the end-effector of the robot is commanded to approach the object, while keeping a given relative pose with respect to the target object. We conduct the experiments of Approaching Visual Servoing to a 3D marker to verify the effectiveness of the proposed hand & eye-vergence dual control system. The 3D marker is composed of a red ball, a green ball and a blue ball. The radiuses of these three balls are set as 30[mm].

A. Experimental Condition

A photograph of our experimental system is shown in Fig.6. The robot used in this experimental system is a 7-Link manipulator, Mitsubishi Heavy Industries PA-10 robot. Two mobile cameras are mounted on the robot manipulator's end-effector. The image processing board, CT-3001, receiving the image from the CCD camera is connected to the DELL Optiplex GX1 (CPU: Pentium2, 400 MHz) host computer.

Fig.7 shows the coordinate system corresponding to Fig.6. The initial pose of the end-effector is defined as Σ_{E_0}, and given by

$$ {}^{W}\boldsymbol{T}_{E_0} = \begin{bmatrix} 0 & 0 & 1 & -918 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 455 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{24} $$

position unit: [mm].

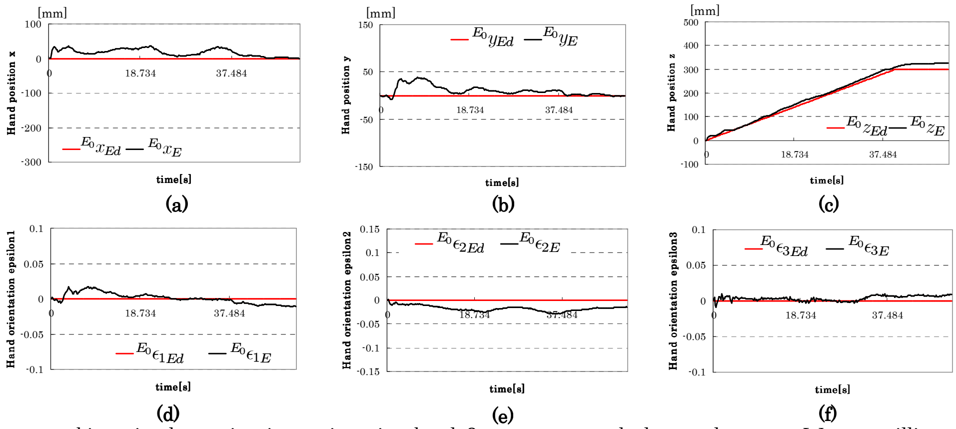

Fig. 8. Hand pose of approaching visual servoing in z-axis, using hand & eye-vergence dual control system. We use millimeter to measure position. When using quaternion to express the orientation of an object, no unit, just values. The object rotation of 1[deg] around x axis corresponding to quaternion representation as $\epsilon_1 = 0.008, \epsilon_2 = 0, \epsilon_3 = 0$.

B. Experiments

1) Approaching Visual Servoing in z-axis: Here, a static object is set as ${}^{E_0}\psi_M = [0[mm], 50[mm], 900[mm], 0, 0, 0]^T$. The objective of the Approaching Visual Servoing is given by

$$\begin{cases} {}^{Ed}z_M(t) = z_{max} - (z_{max} - z_{min})t/T \quad if(t \le T) \\ {}^{Ed}z_M(t) = z_{min} \quad if(t > T) \end{cases} \tag{25}$$

where we set $z_{max} = 900[mm]$, $z_{min} = 600[mm]$, $T = 40[s]$. The other objective parameters are given the same as beginning (${}^{E_0}\psi_M$), that is,

$$\begin{cases} {}^{Ed}x_M(t) = 0 \\ {}^{Ed}y_M(t) = 50[mm] \\ {}^{Ed}\epsilon_{1M}(t) = 0 \\ {}^{Ed}\epsilon_{2M}(t) = 0 \\ {}^{Ed}\epsilon_{3M}(t) = 0 \end{cases} \tag{26}$$

The above objective of the Approaching Visual Servoing given in (25), (26) means observing the target object from a 900[mm] faraway place to a 600[mm] distance, as shown in Fig.6.

Figs.8(a) to (f) show the actual motion of the end-effector with respect to the fixed frame of Σ_{E_0}, defined as ${}^{E0}\psi_E$, compared with the desired hand pose ${}^{E0}\psi_{Ed}$. As shown in Fig.8(c), the end-effector is desired to move 300[mm] in z-axis of Σ_{E_0} in the first 40[s]; and keep 600[mm] distance to the target object, no more approach, that is ${}^{E0}z_{Ed} = 300[mm]$ after 40[s]. The actual motion of the end-effector shown in Fig.8(c) confirmed that this approaching motion was achieved. The errors between the desired hand pose ${}^{E0}\psi_{Ed}$ and the actual hand pose ${}^{E0}\psi_E$ are limited in a small range. Position error is about 30[mm], orientation error is about 0.02 (3[deg]). When the end-effector became nearer to the target object, the hand motion errors became smaller, since the target object is bigger in the camera images, which is easier for recognition.

Meanwhile, as the end-effector approach the target object, the cameras change their pan angles to focus on the object, which has been confirmed by Figs.9(a) and (b). From 0[s] to 40[s], the angles of both left and right cameras are changed from 4[deg] to 5.7[deg]. The pose changing of the cameras look very small, less than 2[deg] totally, however, consider the short distance from the cameras to the target object, which is only 600[mm] in the last, even small rotation of the cameras is enough to make sure the object is observable. After 40[s], both the hand motion and cameras' motion are converged, which also confirmed the stability of our hand & eye dual control system.

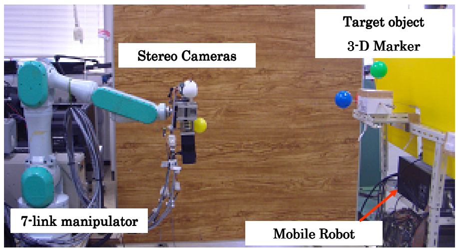

Fig. 10. A photograph of approaching visual servo to a moving object, which is fixed on a mobile robot.

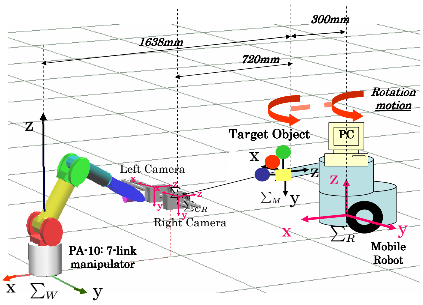

Fig. 11. Coordinate system of approaching visual servo system in Fig.10.

2) Approaching Visual Servoing to A Moving Object: In this experiment, the target object is fixed on a mobile robot, and moves together with the mobile robot, as shown in Fig. 10. The coordinate system corresponding to Fig. 10 is shown in Fig. 11. The coordinate system of the mobile robot is represented as Σ_R. Here, the motion of the mobile robot is a shuttle rotation around the z axis of Σ_R given by

$$\theta_d[deg] = a\ sin(\frac{2\pi}{T})t, \tag{27}$$

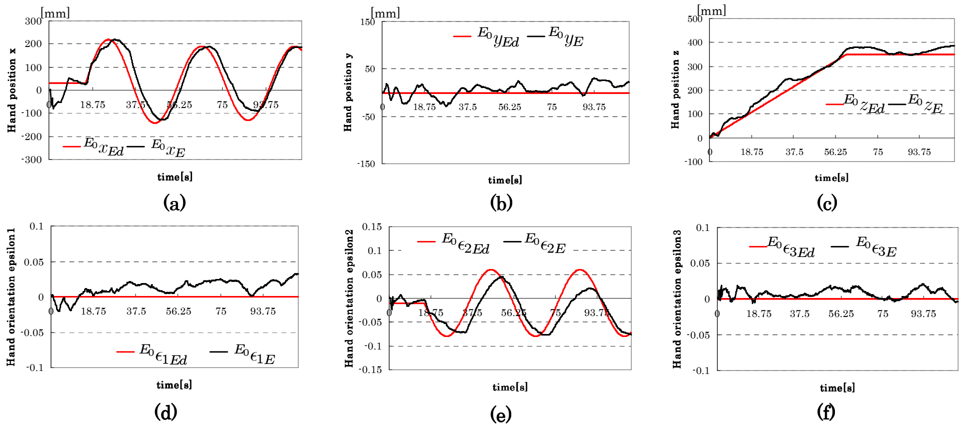

Fig. 12. Hand pose of Approaching Visual Servoing to a moving object, using hand & eye-vergence dual control system.

where we set $a = 8[deg], T = 40[s]$. The voltage of the right and left wheel is given by

$$V_R = kp(\theta_d - \theta) + kv(\dot{\theta}_d - \dot{\theta}), \quad (28)$$

$$V_L = -V_R, \quad (29)$$

where kp and kv are suitable feedback PD control gains.

Here, the effectiveness of the proposed hand & eye-vergence dual control system is evaluated through Approaching Visual Servoing to the moving target object.. Here, the objective of visual servoing is same as the first experiment, given in (25), (26), but here we set $z_{max} = 900[mm]$, $z_{min} = 550[mm]$, $T = 60[s]$.

Figs. 12(a) to (f) are the experimental results, which show the actual motion of the end-effector with respect to the fixed frame of Σ_{E_0}, defined as ${}^{E0}\psi_E$, compared with the desired hand pose ${}^{E0}\psi_{Ed}$. In the first 15[s], the mobile robot did not move, Approaching Visual Servoing to a static object (the same with the first experiment) was performed, so the trajectory of ${}^{E0}\psi_{Ed}$ is a straight line from 0[s] to 15[s]. Then in the moment the mobile robot started to move, the desired trajectory in Fig. 12(a),(e) began to turn to curved line of sin/cos function.

As shown in Fig.12(c), the end-effector is desired to move 350[mm] in z-axis of Σ_{E_0} in the first 60[s]; and keep 550[mm] distance to the target object, no more approach, that is ${}^{E0}z_{Ed} = 350[mm]$ after 60[s]. The motion image is shown in Fig.2(b). The actual motion of the end-effector shown in Fig.12(c) confirmed that this approaching motion was achieved. And the actual motion of the end-effector shown in Fig.12(a),(e) confirmed that the tracking of the rotating object was achieved, with about 5[s] time delay. The errors between the desired hand pose ${}^{E0}\psi_{Ed}$ and the actual hand pose ${}^{E0}\psi_E$ are limited in a small range. Position error is about 20[mm], orientation error is about 0.02 (3[deg]).

As the end-effector approach the target object, the cameras change their pan angles to focus on the object, which has been shown in Figs.13(a) and (b). The pan angle of the left camera is changed from 4[deg] to 6[deg], and the right one is changed from 2[deg] to 6[deg].

This experiment has shown the effectiveness of our proposed hand & eye-vergence dual control system, to keep the end-effetor tracking a moving target object while approaching it.

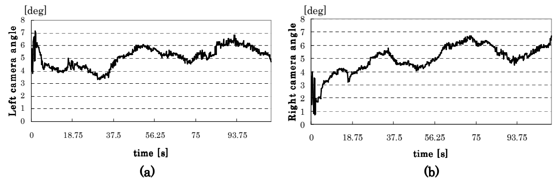

Fig. 13. Camera pose of Approaching Visual Servoing a moving object.

3) Approaching Visual Servoing in x-z plane: In the first experiment, the end-effector is controlled to approach the object in just z-axis of Σ_{E_0}. Here, we set the approaching trajectory as a circle in x-z plane of Σ_{E_0}. And the 3D marker is static. The objective of this visual servoing is given by

$$\begin{cases} {}^{Ed}x_M(t) = r_0 \ sin\frac{2\pi}{T}t \\ {}^{Ed}z_M(t) = d \ + \ r_0 \ cos\frac{2\pi}{T}t \end{cases} \quad (30)$$

where we set $d = 700[mm]$, $r_0 = 100[mm]$, $T = 60[s]$. The other objective parameters are given the same as beginning (${}^{E_0}\psi_M$), that is,

$$\begin{cases} {}^{Ed}y_M(t) = 50[mm] \\ {}^{Ed}\epsilon_{1M}(t) = 0 \\ {}^{Ed}\epsilon_{2M}(t) = 0 \\ {}^{Ed}\epsilon_{3M}(t) = 0 \end{cases} \quad (31)$$

We compare the visual servoing by using the hand & eye-vergence visual servoing system and fixed parallel stereo camera system separately. In both cases, the distance between the left and right cameras is set as 250[mm].

Fig.14 shows the experiment results of Approaching Visual Servoing in x-z plane, using parallel stereo cameras, all these results are represented in Σ_{E_0}. Fig.14(a) is the actual end-effector in x and y position compared with the desired x and y. Fig.14(b) is the end-effector's motion in y and z plane of Σ_{E_0}. Fig.14(c) is the end-effector's motion in x and z plane of Σ_{E_0}. Fig.14(d) is the end-effector's motion in the orientation ϵ_1 and ϵ_2. Fig.14(e) is the end-effector's motion in the orientation ϵ_2 and ϵ_3. Fig.14(f) is the end-effector's motion in the orientation ϵ_1 and ϵ_3.

We can find that in Fig.14, the errors between the actual and desired position are big to about 80[mm]. Especially in

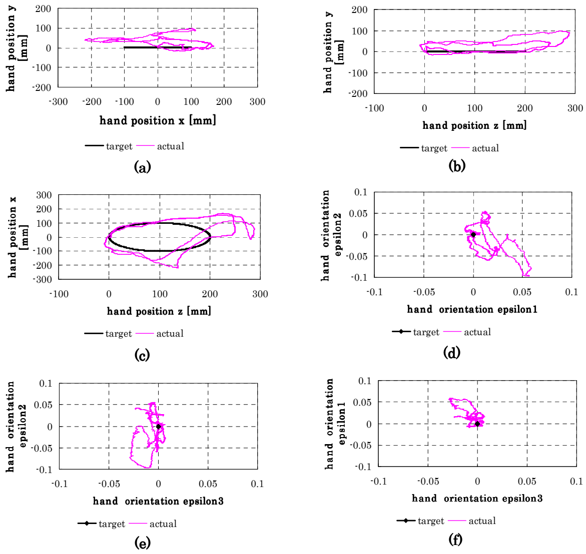

Fig. 14. Hand pose of Approaching Visual Servoing in x-z plane, using parallel stereo cameras.

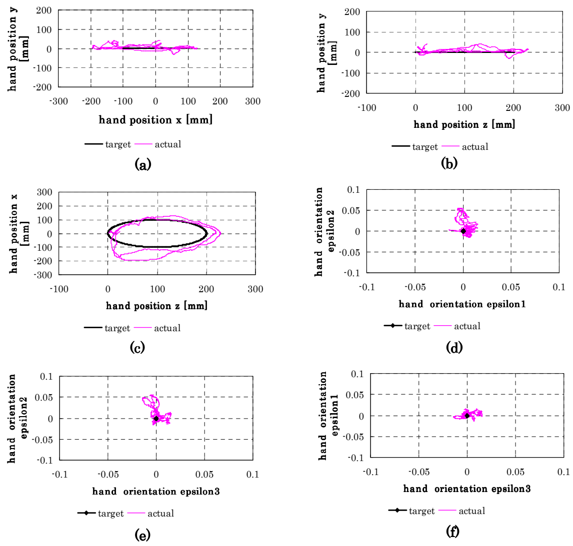

Fig. 15. Hand pose of Approaching Visual Servoing in x-z plane, using hand & eye-vergence dual control system.

Fig. 14(c), it shows a part of the actual trajectory goes far away from the desired trajectory. The distributions of the end-effector's actual orientation ϵ_1,ϵ_2,ϵ_3 in Fig.14(d) to (f) are around the desired value 0. However, error of ϵ_2 is big to about 0.1 (14[deg]), and errors of ϵ_1 and ϵ_3 are about 0.05 (7[deg]). It was found that since the stereo cameras were fixed to be parallel, a part of the object got out of the camera view when the cameras approached the target object. Therefore, the recognition in such cased could not be correct, so the errors of the hand trajectory become bigger.

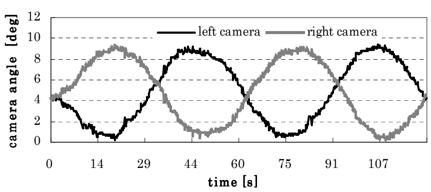

Fig. 16. Camera pose of Approaching Visual Servoing in x-z plane.

Fig.15 shows the Approaching Visual Servoing in x-z plane in Σ_{E_0}, using hand & eye-vergence dual control system. Fig.15(a) to (f) is the same meaning as Fig.14(a) to (f). The changing of the pan angles of the left camera and the right camera is shown in Fig.16. Comparing with Fig.14, the errors between the actual and desired position in Fig.15(a) and (b) is small, less than 20[mm], and the actual position of end-effetor is close to the desired circle trajectory in Fig.15(c). The distribution of the end-effector's actual orientation ϵ_1,ϵ_2,ϵ_3 is converged to the desired value 0, error is less than 0.05 (about 7[deg]). It was found that the target object could be observed by both cameras all the time because the cameras are changing their poses to keep gazing on the target object. Therefore, the errors of the hand pose became smaller.

VI. Conclusion

Besides tracking an object, intelligent robots are expected to do more service for people, the basic work is to grasp an object and carry it to a given place. In this paper, we focus on how to control the robot's end-effector to track an object, meanwhile, to approach it with a suitable posture for grasping, which we named as "Approaching Visual Servoing". We proposed a hand & eye-vergence dual control system to perform Approaching Visual Servoing, which includes two loops: an outer loop that direct a manipulator toward a target object and an inner loop that direct active motion of binocular camera for accurate and broad observation of the target object. The experiment of full 6-DoF Approaching Visual Servoing has confirmed the effectiveness of our proposed the hand & eye-vergence dual control system. As future research, we will try to perform the grasping operation by installing a multifingered mechanical hand on the end-effector. This work was supported by Grant-in-Aid for Scientific Research (C) 19560254.

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] 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.

[3] Omar Tahri and François Chaumette, "Point-Based and Region-Based Image Moments for Visual Servoing of Planar Objects", IEEE Tran. on Robotics, vol. 21, no. 6, Dec 2005.

[4] Lippiello, V, Siciliano, B, Villani, L, "Position-Based Visual Servoing in Industrial Multirobot Cells Using a Hybrid Camera Configuration", IEEE Transactions on Robotics, Vol. 23, NO.1, pp. 73-86, FEB,2007

[5] Tarek Hamel, Robert Mahony "Visual Servoing of and Under-Actuated Dynamic Rigid-Body System: An Image-Based Approach" IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, VOL. 18, NO.2, APRIL 2002

[6] 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

[7] 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.

[8] B.Siciliano and L.Villani: *Robot Force Control*, ISBN 0-7923-7733-8.