2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Hand & Eye-Vergence Dual Visual Servoing to Enhance Observability and Stability

Wei Song and Mamoru Minami

Abstract— **In this paper, we propose a new two-way visual servoing method, named as hand & eye-vergence visual servoing. This idea stems from animal's evolution history, predator have evolved their eye positions to be at the front face and their eyes turn to gaze at the target prey to be suited to triangulation, enhancing the ability to measure precisely the distance to the prey for catching it. This animal's visual tracking includes motion control by visual servoing and triangular eye vergence. Our proposed method includes two loops: an outer loop for conventional visual servoing that direct a manipulator toward a target object and an inner loop for active motion of binocular camera for accurate and broad observation of the target object. The effectiveness of the hand & eye-vergence visual servoing is evaluated through simulations incorporated with actual dynamics of 7-DoF robot on the view points of how the new idea improved the stability in visual servoing dynamics and the accuracy of hand pose.**

I. Introduction

Tasks in which visual information are used to direct a manipulator toward a target object are referred to visual servoing in [1], [2]. This is the general definition of visual servoing and it can be classified into two major groups: position-based and image-based visual servoing. Position-based visual servoing is to determine the object pose in Cartesian coordinate frame and lead to Cartesian robot motion planning, [3], [4]. On the other hand, in an image-based visual servoing, image features are measured in the 2-D image space, and the robot is controlled to move the image features to a set of desired locations, [5], [6]. The advantages and drawbacks of each visual servoing method have been discussed by a significant amount of researches, listed in Table I. Compare with image-based visual servoing, position-based visual servoing is more understandable, the visual servoing task is described in Cartesian space and it is more like human-being's space perception.

However, most visual servoing researches focus attention on robot control problem, and simplify or omit the object measurement problem, which is thought to be dealt with in another research field: robot vision field. Visual servoing system is generally an eye-in-hand configuration, in which the camera is fixed on the end-effector. The point is that we have to enhance both the camera observability and the robot stability simultaneously, because they affect each other in visual servoing. Keeping suitable viewpoint is important for object observation. It can provide more information of the object for fast and correct recognition. Unsuitable viewpoint may possibly cause a part of the target object or some feature points get out of the image, which will cause the robot unstable. Some methods are proposed to improve observability of the object, like using stereo camera [7], multiple cameras [8], and two cameras: one is fixed on the end-effector, the other is fixed in the workspace, [9]. However, these methods only increase the number of cameras to give different views to observe the object, the cameras in these systems lack the adaptability to a changing environment, that is, the ability to change the viewpoint along with the moving object.

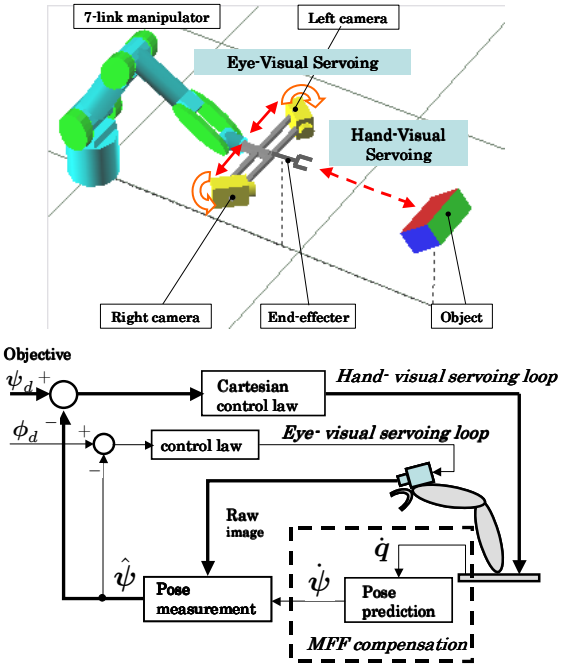


Fig. 1. Hand & Eye Visual servo system

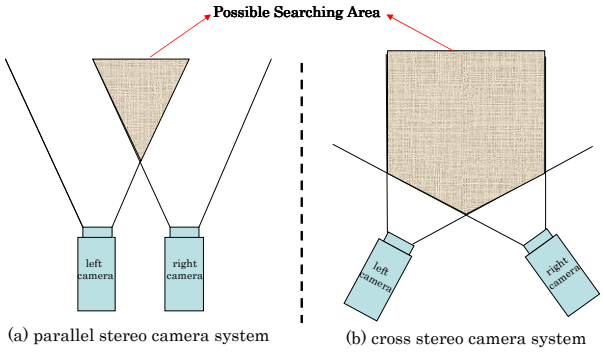
The final objective of visual servoing is generally for the end-effector to approach an target object and then work on it, like grasping. But recent researches on visual servoing are generally tracking an object while keeping a certain distance between the camera and the object, for the reason that the possible searching area becomes small when the target is nearing the camera, as shown in Fig. 2(a). Thus it is necessary to change the pose of the camera to enlarge the possible searching area in the case that the target is close to the camera, as shown in Fig. 2(b). Since the cameras and the end-effector do different tasks - camera is used for observe the target object, and the end-effector is to move toward the target object - it is reasonable to separate the motion of the

Wei Song and Mamoru Minami are with Department of Human and Artificial Intelligence Systems, Graduate School of Engineering, University of Fukui, Fukui, 910–8507, Japan { songwei, minami } @rc.his.fukui-u.ac.jp

TABLE I

POSITION-BASED AND IMAGE-BASED VISUAL SERVOING

	Advantage	Drawback
Position-based	Trajectory planning is done in an intuitive Cartesian coordinate space	Require a model of the target object
	There is a clear separation of the measurement problem and the control problem.	Require camera and robot calibration
	Familiar robot control design is used	
Image-based	do not require a model of the target object	difficult to do trajectory planning in the non-intuitive image plane,
	Best suited to planar motion where the plane is parallel to the image plane	difficult to non-planar motion where the plane is not parallel to the image plane
	Robust to camera and robot calibration errors	



(a) parallel stereo camera system (b) cross stereo camera system

Fig. 2. Possible area of parallel stereo camera system and cross stereo camera system in (b)

camera and the motion of the end-effector. In this paper, we present a hand & eye-vergence dual visual servoing system as shown in Fig. 1, in which the hand-visual servoing loop includes the active motion of binocular camera to maximize accurate and broad observation of the target object.

With an eye-in-hand configuration, a problem exists, that is the motion of the end-effector cause a fictional motion of the object. Here we are interested in how to compensate such a fictional motion of the target object. Consider about the human's action, we can perceive the target pose in the world coordinate by subtracting the fictional motion caused by the motion of ourselves, thus the influence on recognition from the ego motion can be compensated. To realize this compensation of ego motion, it is better to use the kinematics relation of robot to calculate the fictional motion, thus we chose position-based method instead of image-based method. To apply such a intelligence into recognition system, we propose an robust recognition method, called motion-feedforward (MFF) compensation method. The MFF method gives a relation that connects the rate of change of the target pose in end-effector frame to the rate of change of the joint angles. We use the relation to compensate the fictional motion of the target based on the joint velocity, and extract the real motion of the target in the camera images, which can improve the performance of the image recognition unit, making the recognition system dynamically stable. This stable sensing brought additional merit to stabilize the visual servoing motion.

We use model-based method to recognize 3-D target's pose in real-time. The matching degree of the model to the target can be estimated by a fitness function, whose maximum value represents the best matching and can be solved on-line by"1-Step GA" [10]. Unit quaternion is used to represent the orientation of the target object, which has an advantage that can represent the orientation of a rigid body without singularities.

The effectiveness of the hand & eye-vergence dual visual servoing is evaluated through simulations incorporated with actual dynamics of 7-DoF robot on the view points of how the new idea improved the stability in visual servoing dynamics and the accuracy of hand pose.

II. SIMULATOR AND ROBOT DYNAMICS

The Mitsubishi PA-10 robot arm is a 7 DOF robot arm manufactured by Mitsubishi Heavy Industries. Our simulator is model PA-10 (see Fig. 1), by using the actual physical parameters of the PA-10 that are shown in Table II.

The general equation of motion of manipulator is

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{h}(\boldsymbol{q}, \dot{\boldsymbol{q}}) + \boldsymbol{g}(\boldsymbol{q}) + \boldsymbol{d}(\dot{\boldsymbol{q}}) = \boldsymbol{\tau}, \tag{1}$$

where, $\boldsymbol{q}$: the joint displacement and $\boldsymbol{q} = [q_1, q_2, \cdots, q_7]^T$, $\boldsymbol{\tau}$: the joint driving force and $\boldsymbol{\tau} = [\tau_1, \tau_2, \cdots, \tau_7]^T$, $\boldsymbol{M}(\boldsymbol{q})$: the inertia matrix, $\boldsymbol{h}(\boldsymbol{q}, \dot{\boldsymbol{q}})$: the vector representing the centrifugal and coriolis forces, $\boldsymbol{g}(\boldsymbol{q})$: the vector representing the gravity load, $\boldsymbol{d}(\dot{\boldsymbol{q}})$: the vector representing the frictional force. Here, we assumed $\boldsymbol{d}(\dot{\boldsymbol{q}}) = \boldsymbol{0}$.

Two cameras are mounted on the end-effector, modeling CCD-TRV86 manufactured by Sony Industries. The frame frequency of stereo cameras is set as 33fps.

III. ON-LINE EVOLUTIONARY RECOGNITION

First, 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.

A. 3-D Measurement Method

We use a model-based matching method to recognize a target object in a 3-D searching area. A solid models is located in Σ_E, its position and orientation are determined by six parameters, ${}^E\boldsymbol{\psi} = [{}^E\boldsymbol{r}^T, {}^E\boldsymbol{\epsilon}^T]^T$, where ${}^E\boldsymbol{r} = [{}^Ex, {}^Ey, {}^Ez]^T$, ${}^E\boldsymbol{\epsilon} = [{}^E\epsilon_1, {}^E\epsilon_2, {}^E\epsilon_3]^T$. Here, the target's orientation is represented by unit quaternion [13], which has a advantage that can represent the orientation of a rigid body

TABLE II
PHYSICAL PARAMETERS OF THE PA-10

Joint	*Base*	*Link1*	*Link2*	*Link3*	*Link4*	*Link5*	*Link6*	*Link7*
Length(m)	0.200	0.115	0.307	0.143	0.225	0.245	0.080	0.020
Center of mass (m)	0.0750	-0.0518	0.0633	0.0536	0.0461	0.0803	-0.0186	0.0040
mass (Kg)	3.04	9.78	8.41	3.51	4.31	3.45	1.70	0.36
Inertia momentI_{xx} (Kgm^2)	N/A	1.23×10^{-2}	6.86×10^{-2}	3.70×10^{-2}	2.79×10^{-2}	4.07×10^{-2}	1.09×10^{-2}	2.50×10^{-3}
Inertia momentI_{yy} (Kgm^2)	N/A	6.36×10^{-2}	6.86×10^{-2}	2.62×10^{-2}	2.79×10^{-2}	5.83×10^{-3}	1.09×10^{-2}	2.50×10^{-3}
Inertia momentI_{zz} (Kgm^2)	N/A	1.23×10^{-1}	1.19×10^{-2}	3.70×10^{-2}	6.48×10^{-3}	4.07×10^{-2}	6.97×10^{-4}	1.74×10^{-4}

without singularities. The left and right input images from the stereo cameras are directly matched by the left and right searching models, which are projected from 3-D model onto 2-D image plane. The matching degree of the model to the target can be estimated by a fitness function $F({}^E\boldsymbol{\psi})$ by using the color information of the target. Please refer to [12] for a detailed definition of $F({}^E\boldsymbol{\psi})$. When the searching models fit to the target objects being imaged in the right and left images, $F({}^E\boldsymbol{\psi})$ gives the maximum value. Therefore the 3-D object's position/orientation measurement problem can be converted to a searching problem of ${}^E\boldsymbol{\psi}$ that maximizes $F({}^E\boldsymbol{\psi})$. We solve this optimization problem by 1-step GA method that will be explained in the next section.

In the following of this paper, we will omit the left superscipt "E" of the definition of ${}^E\boldsymbol{\psi}$ for abbreviation.

B. 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

$$\boldsymbol{\psi}^{max}(t) = \{\boldsymbol{\psi}(t) \mid \max_{\boldsymbol{\psi}\in\boldsymbol{L}} F(\boldsymbol{\psi}(t))\}, \tag{2}$$

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,

$$\boldsymbol{\psi}_{ga}^{max}(t) = \{\boldsymbol{\psi}_{i,j}(t) \mid \max_{\boldsymbol{\psi}_{i,j}\in\boldsymbol{L}} F(\boldsymbol{\psi}_{i,j}(t))\}. \tag{3}$$

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

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

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

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

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.

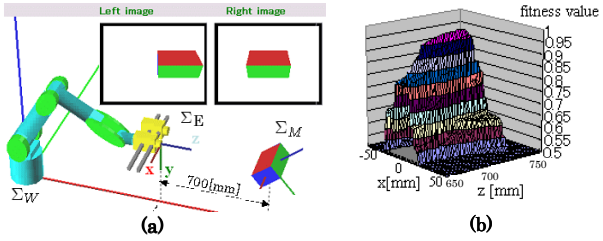


Fig. 3. (a) parallel stereo camera system (b) configuration of $F(\phi)$ in position x and z

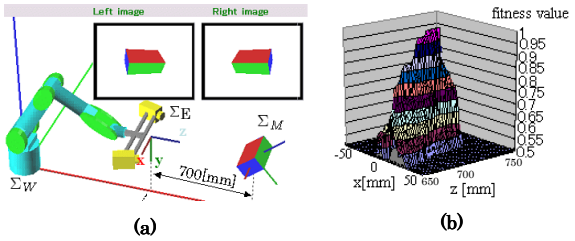


Fig. 4. (a) cross stereo camera system (b) configuration of $F(\phi)$ in position x and z

From [Assumption 2], we have

$$\Delta\dot{F}(\delta\boldsymbol{\psi}(t)) = \dot{F}(\boldsymbol{\psi}^{max}(t)) - \dot{F}(\boldsymbol{\psi}_{ga}^{max}(t)) < 0. \tag{6}$$

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, (5) and (6) 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

$$\boldsymbol{\psi}_{ga}^{max}(t) \rightarrow \boldsymbol{\psi}^{max}(t), (t \rightarrow \infty) \tag{7}$$

Let t_ϵ denotes a convergence time, then

$$|\delta\boldsymbol{\psi}(t)| = |\boldsymbol{\psi}^{max}(t) - \boldsymbol{\psi}_{ga}^{max}(t)| \leq \epsilon, \quad (\epsilon > 0, t > t_\epsilon) \tag{8}$$

In (8), 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

$$\boldsymbol{\psi}_{i,j}(t) \stackrel{\text{evolve}}{\longrightarrow} \boldsymbol{\psi}_{i,j}(t+\Delta t). \tag{9}$$

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

$$|\delta\boldsymbol{\psi}(t)| \leq \epsilon', \quad (\epsilon' > \epsilon > 0), \tag{10}$$

then, it can be considered Δt can manage real-time optimal solution. In (10), 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 [10]. $\boldsymbol{\psi}_{ga}^{max}(t)$ will be output as the measurement result in each generation to control the robot manipulator. We define

$$\hat{\boldsymbol{\psi}}(t) = \boldsymbol{\psi}_{ga}^{max}(t), \hat{\boldsymbol{\psi}} = [\hat{x}, \hat{y}, \hat{z}, \hat{\epsilon_1}, \hat{\epsilon_2}, \hat{\epsilon_3}]^T. \tag{11}$$

C. Maximum Observability

Here, we show an investigation into the influence of viewpoint of observing on 3-D position/orientation measurement. The distance between the origin of one camera to the origin of Σ_E is defined as L_l /L_r, and the angle rotating around y axis of Σ_{CL} /Σ_{CR} is defined as θ_l /θ_r.

We compare two stereo camera systems: parallel stereo camera system (Fig.3(a)), where $L_l = L_r = 75[mm]$, and cross stereo camera system (Fig.4(a)), where $L_l = L_r = 250[mm]$, $\theta_l = \theta_r = 20[deg]$. A fitness value distribution of $F(\boldsymbol{\psi})$ by scanning the x and z position of of Σ_E using the moving model with fixed true values of $(y, \epsilon_1, \epsilon_2, \epsilon_3)$, is shown in Fig.3(b) and Fig.4(b) corresponding to each stereo camera system, Fig.3(b) and Fig.4(b). It can be seen that when the position of the model near to the true position $(x, z) = (0, 700)[mm]$, the fitness function has maximum values in both Fig.3(b) and Fig.4(b). However, it shows that the peak in Fig.4(b) is sharper than the peak in Fig.3(b). As we know, GA find the maximum value fast in sharp mountain. Thus, the real-time recognition is easier to be performed in the case of cross stereo camera system than that of parallel stereo camera system.

We have also done some investigations by comparing cross stereo camera systems with different θ_l /θ_r with fixed L_l and L_r, from which we found that observing the object through both centers of left and right cameras gave the sharpest mountain. Based on the this, we consider that the stereo cameras had better to keep changing their placement toward a moving target object in order to prepare a good viewpoint for a better measurement of a target object, which is the concept of eye-visual servoing.

Since the maximum observability can be achieved in eye-visual servoing, the object recognition can be performed more precisely and the stability of the general visual servoing to an object can be improved.

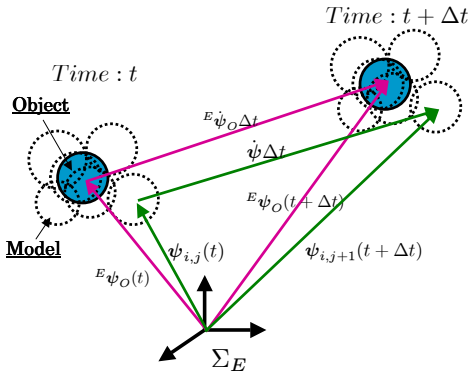


Fig. 5. MFF Compensation. Notice Σ_E and Σ_O are relative coordinates, here we suppose the end-effector is moving and the target is static.

IV. Motion-Feedforward (MFF) Compensation

A. Analysis of target's motion in Σ_E

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.

$$\begin{aligned}
{}^E\dot{\boldsymbol{\psi}}_M &= \begin{bmatrix} {}^E\dot{\boldsymbol{r}}_M \\ {}^E\dot{\boldsymbol{\epsilon}}_M \end{bmatrix} \\
&= \begin{bmatrix} -{}^E\boldsymbol{R}_W(\boldsymbol{q})\boldsymbol{J}_p(\boldsymbol{q}) + {}^E\boldsymbol{R}_W(\boldsymbol{q}) \\ \boldsymbol{S}({}^W\boldsymbol{R}_E(\boldsymbol{q}){}^E\boldsymbol{r}_M)\boldsymbol{J}_o(\boldsymbol{q}) \\ -\frac{1}{2}({}^E\eta_M\boldsymbol{I} - \boldsymbol{S}({}^E\boldsymbol{\epsilon}_M)){}^E\boldsymbol{R}_W(\boldsymbol{q})\boldsymbol{J}_o(\boldsymbol{q}) \end{bmatrix} \dot{\boldsymbol{q}} \\
&\quad + \begin{bmatrix} {}^E\boldsymbol{R}_W(\boldsymbol{q}) & 0 \\ 0 & {}^E\boldsymbol{R}_W(\boldsymbol{q}) \end{bmatrix} \begin{bmatrix} {}^W\dot{\boldsymbol{r}}_M \\ {}^W\dot{\boldsymbol{\epsilon}}_M \end{bmatrix} \\
&= \boldsymbol{J}_M(\boldsymbol{q}, {}^E\boldsymbol{\psi}_M)\dot{\boldsymbol{q}} + \boldsymbol{J}_N(\boldsymbol{q}){}^W\dot{\boldsymbol{\psi}}_M.
\end{aligned} \tag{12}$$

Please refer to [7] for a detailed deduction procedural of (12). The matrix $\boldsymbol{J}_M$ in (12) describes how target pose change in Σ_E with respect to changing manipulator pose in Σ_E. The matrix $\boldsymbol{J}_N$ in (12) describes how target pose change in Σ_E with respect to the pose changing of itself in real word.

In this paper, we do not deal with the prediction of the target's motion in the real world, we take account of the prediction of the target velocity in Σ_E based on the joint velocity $\dot{\boldsymbol{q}}$ of the manipulator, so we can rewrite (12) as

$${}^E\dot{\boldsymbol{\psi}}_M = \boldsymbol{J}_M(\boldsymbol{q}, {}^E\boldsymbol{\psi}_M)\dot{\boldsymbol{q}}. \tag{13}$$

Then the 3-D pose of the target at time $t + \Delta t$ can be predicted based on the motion of the end-effector motion at time t, presented by

$${}^E\boldsymbol{\psi}_M(t+\Delta t) = {}^E\boldsymbol{\psi}_M(t) + {}^E\dot{\boldsymbol{\psi}}_M\Delta t. \tag{14}$$

B. MFF Compensation

In the same way as the above equation (13), (14), the pose of the individuals $\psi_{i,j}$ in the next $j+1$ generation can be predicted based on the current pose,

$$\dot{\psi} = \boldsymbol{J}_M(\boldsymbol{q}, \hat{\psi}(t))\dot{\boldsymbol{q}}. \quad (15)$$

$$\boldsymbol{\psi}_{i,j+1}(t+\Delta t) = \boldsymbol{\psi}_{i,j}(t) + \dot{\boldsymbol{\psi}}\Delta t, \quad (16)$$

where Δt is the time cost in one generation. By using (16), GA group will move together with the motion of the target in Σ_E, never loose it even under a high-speed moving of robot manipulator, as shown in Fig. 5. 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.

V. Hand & Eye Visual Servoing

A. Desired-trajectory generation

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 actual camera pose Σ_E is denoted as ${}^{E}\boldsymbol{T}_{Ed}$. ${}^{E}\boldsymbol{T}_{Ed}$ can be described by

$${}^{E}\boldsymbol{T}_{Ed} = {}^{E}\hat{\boldsymbol{T}}_M(\hat{\psi}(t))\ {}^{M}\boldsymbol{T}_{Ed}(t), \quad (17)$$

Notice that Eq. (17) 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 Eq. (17) with respect to time yields

$${}^{E}\dot{\boldsymbol{T}}_{Ed}(t) = {}^{E}\dot{\hat{T}}_M(t){}^{M}\boldsymbol{T}_{Ed}(t) + {}^{E}\hat{\boldsymbol{T}}_M(t){}^{M}\dot{\boldsymbol{T}}_{Ed}(t), \quad (18)$$

Differentiating Eq. (18) with respect to time again

$${}^{E}\ddot{\boldsymbol{T}}_{Ed}(t) = {}^{E}\ddot{\hat{T}}_M(t){}^{M}\boldsymbol{T}_{Ed}(t) + 2{}^{E}\dot{\hat{T}}_M(t){}^{M}\dot{\boldsymbol{T}}_{Ed}(t) + {}^{E}\hat{\boldsymbol{T}}_M(t){}^{M}\ddot{\boldsymbol{T}}_{Ed}(t), \quad (19)$$

Where, ${}^{M}\boldsymbol{T}_{Ed}$, ${}^{M}\dot{\boldsymbol{T}}_{Ed}$, ${}^{M}\ddot{\boldsymbol{T}}_{Ed}$ are given as the desired visual servoing objective. ${}^{E}\hat{\boldsymbol{T}}_M$, ${}^{E}\dot{\hat{T}}_M$, ${}^{E}\ddot{\hat{T}}_M$ can be observed by cameras using the on-line evolutionary recognition method explained in Section III.

B. Hand-Visual Servoing Controller

The block diagram of our proposed hand & eye visual servoing controller is shown in Fig. 1. The hand-visual servoing is the outer loop. The controller used for hand-visual servoing is proposed by B.Siciliano [13]. Here, we just show main equations of the controller to calculate $\boldsymbol{\tau}$, which is output to control the robot manipulator.

$$\boldsymbol{a}_p = {}^{W}\ddot{\boldsymbol{r}}_{Ed} + \boldsymbol{K}_{D_p}{}^{W}\dot{\boldsymbol{r}}_{E,Ed} + \boldsymbol{K}_{P_r}{}^{W}\boldsymbol{r}_{E,Ed}, \quad (20)$$

$$\boldsymbol{a}_o = {}^{W}\dot{\boldsymbol{\omega}}_{Ed} + \boldsymbol{K}_{D_o}{}^{W}\boldsymbol{\omega}_{E,Ed} + \boldsymbol{K}_{P_o}{}^{W}\boldsymbol{R}_E{}^{E}\Delta\boldsymbol{\epsilon}, \quad (21)$$

$$\ddot{\boldsymbol{q}}_d = \boldsymbol{J}^{+}(\boldsymbol{q})\Big(\begin{bmatrix}\boldsymbol{a}_p\\ \boldsymbol{a}_o\end{bmatrix} - \dot{\boldsymbol{J}}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}\Big) + \big(\boldsymbol{I} - \boldsymbol{J}^{+}(\boldsymbol{q})\boldsymbol{J}(\boldsymbol{q})\big)\big(\boldsymbol{E}_p(\boldsymbol{q}_0 - \boldsymbol{q}) + \boldsymbol{E}_d(\boldsymbol{0} - \dot{\boldsymbol{q}})\big), \quad (22)$$

$$\boldsymbol{\tau} = \boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}}_d + \boldsymbol{h}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}). \quad (23)$$

Here, the error variables in (20), (21) are described in Σ_W , which can be obtained from the vectors in Σ_E in (17), (18), (19) using the rotational matrix ${}^{W}\boldsymbol{R}_E(\boldsymbol{q})$ through coordinate transformation.

And $\boldsymbol{J}^{+}(\boldsymbol{q})$ in (22) is the preudo-inverse of $\boldsymbol{J}(\boldsymbol{q})$ given by $\boldsymbol{J}^{+}(\boldsymbol{q}) = \boldsymbol{J}^{T}(\boldsymbol{J}\boldsymbol{J}^{T})^{-1}$.

It has been proved in [13] that the system is exponentially stable for any choice of positive definite $\boldsymbol{K}_{D_p}, \boldsymbol{K}_{D_o}$ and $\boldsymbol{K}_{P_p}, \boldsymbol{K}_{P_o}$, thus

$$\lim_{t\to\infty}{}^{W}\boldsymbol{r}_{E,Ed} = \boldsymbol{0} \qquad \lim_{t\to\infty}{}^{W}\dot{\boldsymbol{r}}_{E,Ed} = \boldsymbol{0} \quad (24)$$

$$\lim_{t\to\infty}{}^{E}\Delta\boldsymbol{\epsilon} = \boldsymbol{0} \qquad \lim_{t\to\infty}{}^{W}\boldsymbol{\omega}_{CR,CRd} = \boldsymbol{0}. \quad (25)$$

Then we have

$$\lim_{t\to\infty}{}^{E}\boldsymbol{T}_{Ed} = \boldsymbol{I} \qquad \lim_{t\to\infty}{}^{E}\dot{\boldsymbol{T}}_{Ed} = \boldsymbol{0} \quad (26)$$

Substituting Eq. (26) to Eq. (17), we have

$$\lim_{t\to\infty}{}^{E}\boldsymbol{T}_M = \lim_{t\to\infty}{}^{Ed}\boldsymbol{T}_M \quad (27)$$

Eq. (27) proves stable convergence of visual servoing.

C. Eye-Vergence Visual Servoing Controller

The eye-vergence visual servoing is the inner loop of the visual servoing system shown in Fig. 1. In this paper, we use two pan-tilt cameras for eye-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 comman for both cameras. Another DoF for the pose of the cameras is the angle rotating around the z axis of Σ_E, which is the gazing derection of the camera. Our eye-vergence visual servoing system does not include this DoF because the gazing derection of the camera is not changing in this case, which is meaningless for improving target observability.

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

$${}^{CL}\hat{\boldsymbol{T}}_M({}^{CL}\hat{\boldsymbol{\psi}}(t)) = {}^{CL}\boldsymbol{T}_E(\boldsymbol{\phi}_L){}^{E}\hat{\boldsymbol{T}}_M(\hat{\boldsymbol{\psi}}(t)), \quad (28)$$

$${}^{CR}\hat{\boldsymbol{T}}_M({}^{CR}\hat{\boldsymbol{\psi}}(t)) = {}^{CR}\boldsymbol{T}_E(\boldsymbol{\phi}_R){}^{E}\hat{\boldsymbol{T}}_M(\hat{\boldsymbol{\psi}}(t)). \quad (29)$$

The investigations in III-C shows observing the object through both centers of left and right cameras gave the sharpest mountain, which indicates the maximum observability. Based on this, the objective of the eye-visual servoing is given by

$${}^{CL}\boldsymbol{u}_d = [{}^{CL}x_d, {}^{CL}y_d]^T = \boldsymbol{0}, \ {}^{CR}\boldsymbol{u}_d = [{}^{CR}x_d, {}^{CR}y_d]^T = \boldsymbol{0}. \quad (30)$$

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

$$\dot{\boldsymbol{\phi}}_L = \boldsymbol{K}_{P_L}({}^{CL}\boldsymbol{u}_d - {}^{CL}\hat{\boldsymbol{u}}) + \boldsymbol{K}_{D_L}({}^{CL}\dot{\boldsymbol{u}}_d - {}^{CL}\dot{\hat{\boldsymbol{u}}}), \quad (31)$$

$$\dot{\boldsymbol{\phi}}_R = \boldsymbol{K}_{P_R}({}^{CR}\boldsymbol{u}_d - {}^{CR}\hat{\boldsymbol{u}}) + \boldsymbol{K}_{D_R}({}^{CR}\dot{\boldsymbol{u}}_d - {}^{CR}\dot{\hat{\boldsymbol{u}}}), \quad (32)$$

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

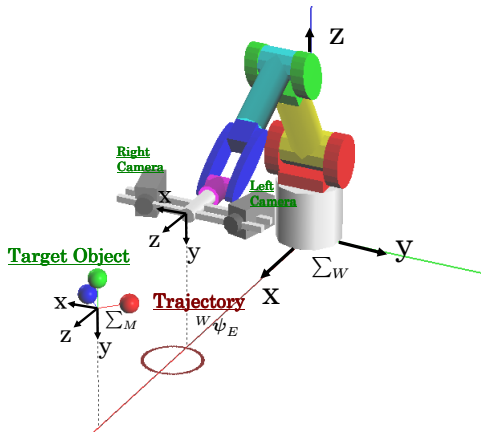


Fig. 6. Simulation of path control in x-z plane (the system is created by OpenGL)

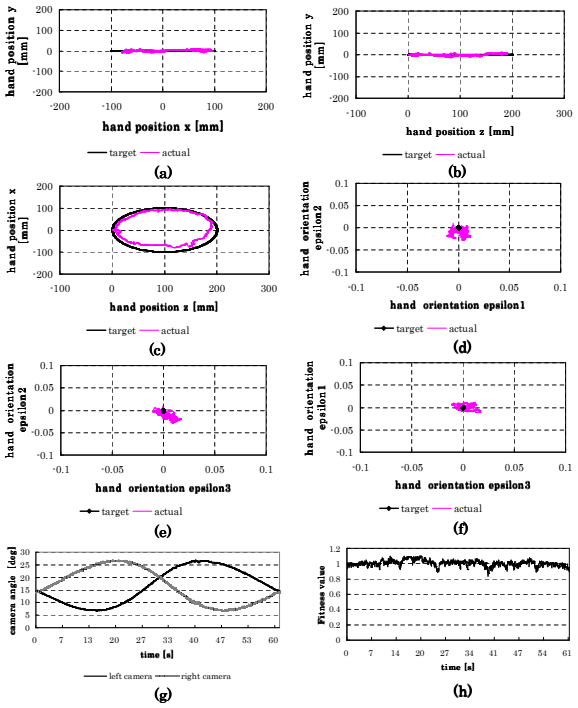


Fig. 7. Results of hand & eye visual servoing in x-z plane by using MFF method in Σ_{E_0}.

VI. Simulation of Hand & Eye Visual Servoing

To verify the effectiveness of the proposed hand & eye visual servoing system, we conduct the simulation of visual servoing to a 3D marker that 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. simulation condition

Visual servoing is usually performed to keep a fixed relation with respect to a static or moving object. The visual servoing described in this paper is that the object remains stationary and the robot is commanded to move through a reference path with respect to it. Such a visual servoing has been performed by William J. Wilson etc. in [4], and they named it as relative path control visual servoing.

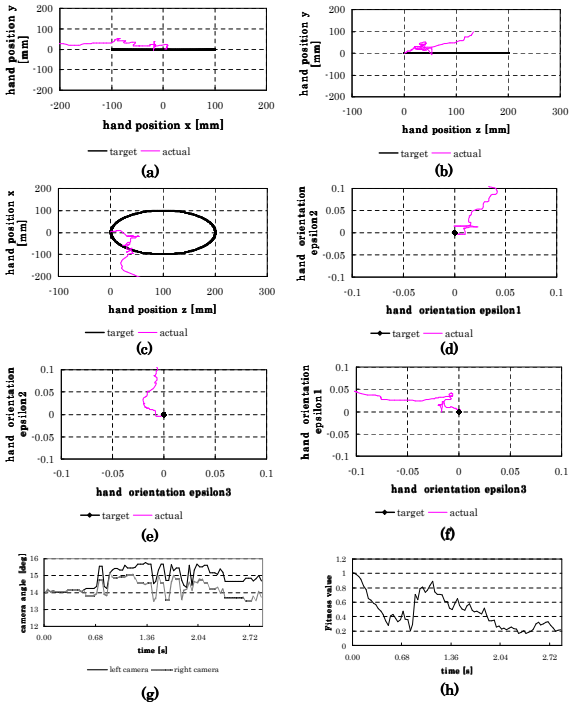


Fig. 8. Results of hand & eye visual servoing in x-z plane without using MFF method in Σ_{E_0}.

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

$$^W\boldsymbol{T}_{E_0} = \begin{bmatrix} 0 & 0 & 1 & 918[mm] \\ -1 & 0 & 0 & 0[mm] \\ 0 & -1 & 0 & 455[mm] \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (33)$$

B. simulation of path control in x-z plane

In this simulation, the desired hand time-varying trajectory expressed in Σ_{E_0} is

$$\begin{cases} {}^{E_0}x_{Ed}(t) = r_0 \ sin\frac{2\pi}{T}t \\ {}^{E_0}y_{Ed}(t) = 0 \\ {}^{E_0}z_{Ed}(t) = r_0 \ cos\frac{2\pi}{T}t \\ {}^{E_0}\epsilon_{1Ed}(t) = 0 \\ {}^{E_0}\epsilon_{2Ed}(t) = 0 \\ {}^{E_0}\epsilon_{3Ed}(t) = 0 \end{cases} \quad (34)$$

where the radius $r_0 = 100[mm]$, period $T = 60[s]$, as shown in Fig. 6.

Firstly, We compare the hand & eye visual servoing in x-z plane with the proposed MFF method and without MFF method separately.

Fig. 7 shows the results of hand & eye visual servoing by using MFF method, all these results are represented in Σ_{E_0}. Fig. 7(a) is the actual end-effector in x and y position compared with the desired x and y. Fig. 7(b) is the end-effector's motion in y and z plane of Σ_{E_0}. Fig. 7(c) is the end-effector's motion in x and z plane of Σ_{E_0}. Fig. 7(d) is the end-effector's motion in the orientation ϵ_1 and ϵ_2. Fig. 7(e) is the end-effector's motion in the orientation ϵ_2 and ϵ_3. Fig. 7(f) is the end-effector's motion in the orientation

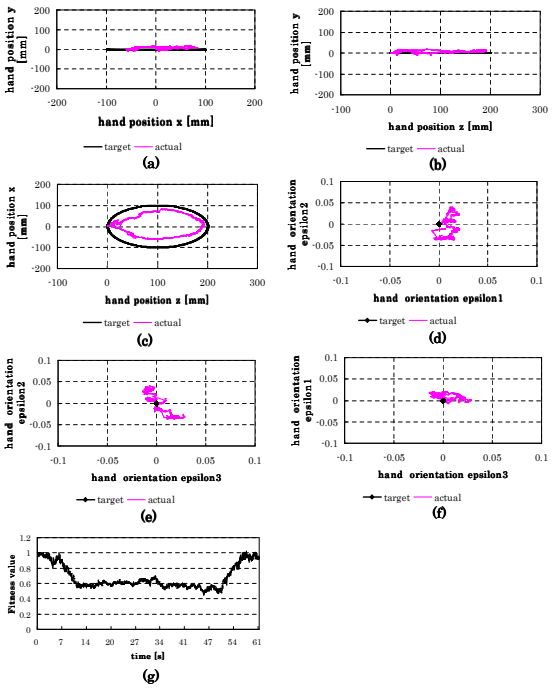


Fig. 9. Results of only hand visual servoing in x-z plane by using MFF method in Σ_{E_0}, using parallel stereo cameras.

ϵ_1 and ϵ_3. Fig. 7(g) shows changing of the pan angles of the left camera and the right camera. Fig. 7(h) shows the changing of fitness value of the target's recognition during visual servoing.

Fig. 7(a) to (f) show us the stable control of the robot manipulator. The errors between the actual and desired position in Fig. 7(a) and (b) is small, less than 10[mm]. As Fig. 7(c) shown, the desired motion of end-effetor in x and z plane is a circle, and the actual positon is close to it. 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 0.02 (about 3[deg]). Fig. 7(g) shows the placement of the left and right cameras keep changing to recognize the object easily. It confirmed the adaptability of the mobile stereo cameras. Fig. 7(h) shows the fitness value keep high, which means precise recognition of the object during visual servoing. It verified that once "1-step GA" finds the closeness model of the target object, the model will keep overlapping the target object, never lose it, because of the MFF method that can compensate the target's fictional motion coming from the robot itself.

On the other hand, as shown in Fig. 8, the visual servoing could not be performed (even in the first 3 seconds) in the case of without using MFF method. When the robot starts to moving, the pose of target object in Σ_E is changed due to the dynamics of the robot manipulator. Without MFF method's compensation, the "1-Step GA" can not recognize precisely, wrong recognition result will lead to wrong control of the robot, which makes the recognition more difficult. As shown in Fig. 8(h), the fitness value of the recognition is decreasing to about 0.2, in such case, the target object is considered as to be lost, and the robot can not be normally controlled. The

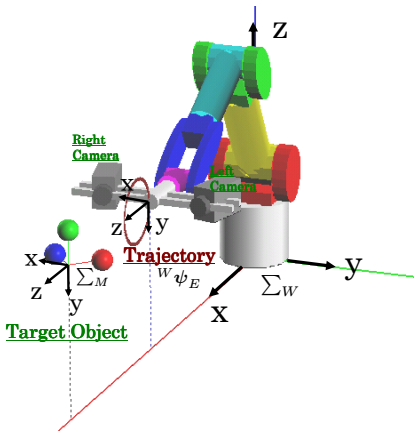


Fig. 10. Simulation of path control in y-z plane (the system is created by OpenGL)

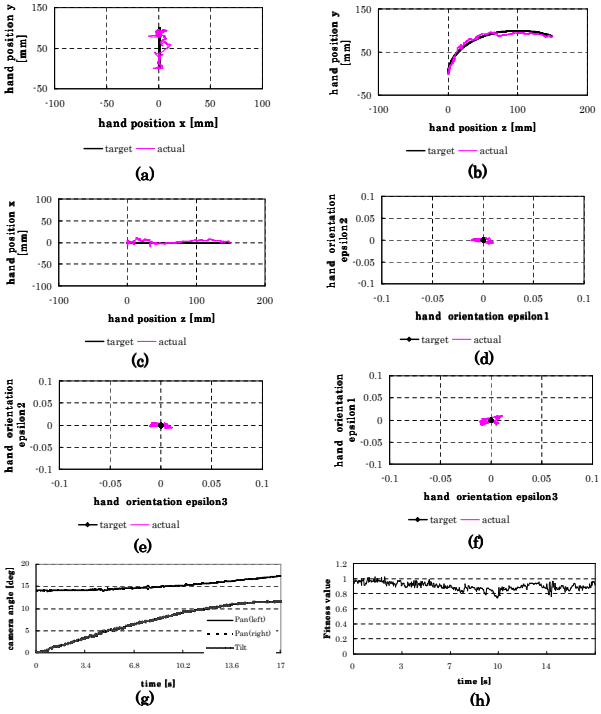


Fig. 11. Results of hand & eye visual servoing in y-z plane by using MFF method in Σ_{E_0}.

reason of target object lost is considered as GA's convergence speed was not faster than the target speed relative to the camera.

Secondly, we compare the visual servoing by using the hand & eye visual servoing system and fixed parallel stereo cameras separately. Fig. 9 shows the results of path control visual servoing with using MFF method, in the case of using parallel stereo cameras. Compare with Fig. 7, we can find that the errors between the actual and desired postion are bigger, especially in Fig. 9(c), the radius of the circle of the actual position in x and z plane is about 20[mm] smaller than the desired one. The distributions of the end-effector's actual orientation ϵ_1,ϵ_2,ϵ_3 in Fig. 7(d) to (f) are around the desired value 0, errora are bigger than using mobile stereo cameras,

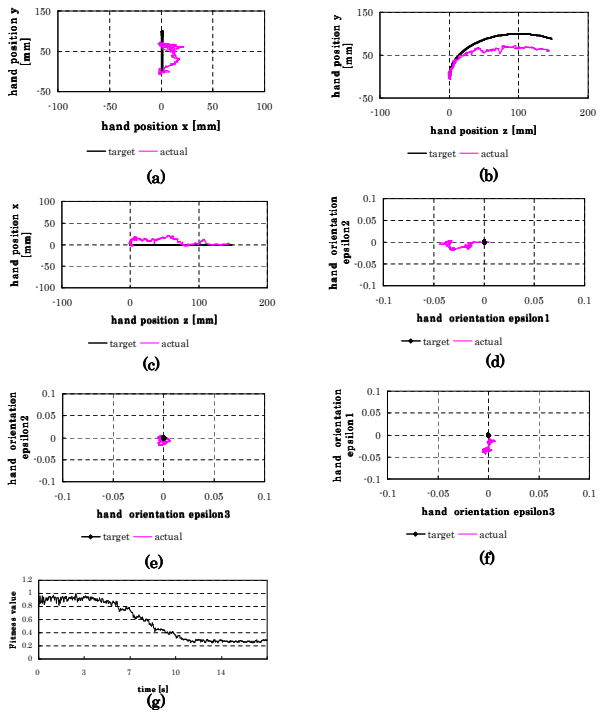


Fig. 12. Results of only hand visual servoing in y-z plane by using MFF method in Σ_{E_0}, using parallel stereo cameras.

about 0.04 (6[deg]). Fitness value of the target's recognition during visual servoing is shown in Fig. 9(g), from 10[s] to 50[s] the fitness value is decreasing to 0.6, where a part of the object got out of the camera view because the stereo cameras are fixed to be parallel, they do not have the adaptability for recognition.

C. simulation of path control in y-z plane

In this simulation, the desired hand path control trajectory expressed in Σ_{E_0} is

$$\begin{cases} {}^{E_0}x_{Ed}(t) = 0 \\ {}^{E_0}y_{Ed}(t) = r_0\ sin\frac{2\pi}{T}t \\ {}^{E_0}z_{Ed}(t) = r_0\ cos\frac{2\pi}{T}t \\ {}^{E_0}\epsilon_{1Ed}(t) = 0 \\ {}^{E_0}\epsilon_{2Ed}(t) = 0 \\ {}^{E_0}\epsilon_{3Ed}(t) = 0 \end{cases} \tag{35}$$

where the radius $r_0 = 100[mm]$, period $T = 60[s]$. In this case, in the hand & eye-vergence system, the tilt angle of the camera that rotating around x axis of Σ_{E_0} should also be controlled to keep the target always in the center of the camea's gazing derection. Fig. 11(g) shows changing of the pan angles of the left camera and the right camera, and the tilt angle of both cameras.

The data shown in Fig. 11 and 12 is about $1/3$ period. Fig. 12 shows the results of path control visual servoing with using MFF method, in the case of using parallel stereo cameras. Compare with Fig. 11, we can find that the errors between the actual and desired postion are bigger, especially in Fig. 12(b), the radius of the circle of the actual position in y and z plane is much smaller than the desired one. The distributions of the end-effector's actual orientation ϵ_1,ϵ_2,ϵ_3 in Fig. 11(d) to (f) are around the desired value 0, error are bigger than using mobile stereo cameras, about 0.04 (6[deg]). Fitness value of the target's recognition during visual servoing is shown in Fig. 12(g), the fitness value is decreasing to 0.3 from 5[s] , where a part of the object got out of the camera view because the stereo cameras are fixed to be parallel, they do not have the adaptability for recognition.

VII. Conclusion

In this paper, we proposed a new two-way visual servoing method, named as hand & eye-vergence dual visual servoing, which includes two loops: an outer loop for conventional visual servoing that direct a manipulator toward a target object and an inner loop for active motion of binocular camera for accurate and broad observation of the target object.

Moreover, we propose a MFF method to compensate the fictional motion of the target based on the joint velocity of manipulator, and extract the real motion of the target for the robot to recognize during visual servoing. Thus, visual recognition preciseness is improved, and the visual servoing become more stable. The effectiveness of the hand & eye-vergence visual servoing has been evaluated through simulations.

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] Wolfgang Sepp, Stefan Fuchs and Gerd Hirzinger, "Hierarchical Featureless Tracking for Position-Based 6-DoF Visual Servoing", IROS2006.

[4] William J. Wilson, Carol C. Williams Hulls and Graham S. Bell, "Relative End-effector Control Using CartesianPosition Based Visual Servoing", IEEE Trans. on Robotics and Automation, vol. 12, No. 5, 1996, pp. 684-696.

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

[6] Tarek Hamel and Robert Mahony, "Visual Servoing of an Under-Actuated Dynamic Rigid-Body System: An Image-Based Approach", IEEE Trans. on Robotics and Automation, VOL. 18, NO. 2, APRIL 2002.

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

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

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

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

[11] D. E.Goldberg, *Genetic algorithm in Search, Optimization and Machine Learning. Reading*, Addison-Wealey, 1989.

[12] 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, Page 127-139 (2007).

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