

3-D Hand & Eye-Vergence Approaching Visual Servoing with Lyapunov-Stable Pose Tracking

Wei Song, Mamoru Minami, Fujia Yu, Yanan Zhang and Akira Yanou

Abstract—In this paper, we focus on how to control the robot end-effector to track an object, meanwhile, to approach it with a suitable posture for grasping. We named it “Approaching Visual Servoing”. A proposed hand & eye-vergence dual control system is used to perform Approaching Visual Servoing, aiming at quick eye-tracking and stable hand servoing and approaching. This idea stems from hammerhead shark whose 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. Moreover, a 3-D pose tracking method that combines “1-Step GA (genetic algorithm)” and hand-motion feedforward compensation is proposed. Our approach differs from known tracking methods using optimization based on Taylor expansion, for it allows the proposed method not to be annoyed by how to sneak out of local minima. A convergence in time domain – whether the 3-D pose tracking error decrease to zero in a successively input images by video rate – , is discussed and verified through Lyapunov method. Both Lyapunov-stable pose tracking and Approaching Visual Servoing are confirmed by experiments using a 7-link manipulator installed with two mobile cameras.

I. INTRODUCTION

Grasping is one of the most common tasks in robot applications. To correctly grasp an object, vision-based techniques are considered to be used to continuously recover the articular pose of the hand. Visual tracking and servoing are such kind of researches. In early stage, a number of researches deal with visual tracking of a moving object [1]-[3], which enabled the robot manipulator to real-time visual tracking of arbitrary 3-D object travelling at unknown velocities in a 2-D space (depth is given as known). Visual tracking satisfies the tracking velocity by decreasing the tracking degree of freedom (DoF). But for grasping, robot needs to approach a moving object with a suitable posture, 2-D visual tracking is not enough to do this. Therefore, recently much research attention turn to focus on full DoF pose regulator of the robot end-effector toward a target object, that is, visual servoing [4]-[7]. Visual servoing is difficult for the full DoF pose measurement and complicated manipulator’s motion control. But most authors only address these two problems as isolated issues. Different kinds of visual servoing, position-based, image-based or $2\frac{1}{2}$ -D visual servoing is usually discussed based on the assumption that pose measurement is known or could be easily received.

Wei Song and Yanan Zhang are with School of Mechatronics Engineering and Automation, Shanghai University, 149 Yanchang Road, Shanghai, 200072, China, e-mail: songwei5726@hotmail.com. Mamoru Minami, Fujia Yu and Akira Yanou are with Graduate School of Natural Science and Technology, Okayama University, Okayama-shi, 700-8530, Japan, e-mail: minami@qumi.sys.okayama-u.ac.jp.

On the other hand, in computer vision field, feature-based, appearance-based and model-based methods are proposed for pose measurement and tracking [8], [9]. But they are seldom discussed together with robot motion or hand-eye motion. In our previous paper [10], we considered the interaction between the sensing unit and the robot motion in the vision feedback control system, and proposed a MFF (hand-motion feedforward method) to compensate the target’s fictional motion seeing in the camera frame that comes from the hand-eye motion.

In visual servoing, when the target moves so fast that the manipulator can not catch up the speed of the target because of the big mass of whole manipulator itself, the object may disappear in the sight of the cameras, resulting in that the visual feedback of the system is cut. Loosing feedback information, the robot may fall in some unexpected motion, being dangerous. So it is important to keep the camera tracking the target. To realize this stable tracking ability against quick and unknown motion of the target, we propose to control the cameras and the manipulator separately. Because of the small mass and inertia moment of the cameras, it can track the target better, like animals track target with eye motion before rotate their heads to the target to improve dynamical tracking ability. Hand & eye-vergence dual visual servoing system was firstly proposed in [11], aiming at enhancing object observability and system stability. It 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 quick tracking and accurate recognition.

This paper is a continuous work of [10] and [11]. The new contents and contributions are:

1) A convergence in time domain – whether the 3-D pose tracking error decrease to zero in successively input images by video rate – , is verified through Lyapunov method.

The tracking problem is treated as an optimization problem of time-varying non-linear function that express the matching degree between the model and the object. We use “1-step GA” recognition method combined with MFF compensation to solve this optimization problem. Our approach differs from known tracking methods using optimization based on Taylor expansion that is annoyed by how to sneak out of local minima [12] [13]. Lyapunov analysis is shown to guarantee that a tracking error of a moving target in task coordinates can be reduced to small value and the assumptions used for the convergence of errors are realistic in practical 3D tracking situation. In addition, experimental verification of the Lyapunov-stable pose tracking is also presented, in which

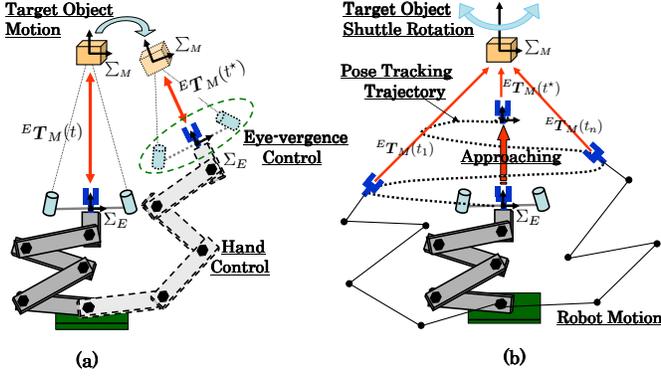


Fig. 1. Approaching Visual Servoing

pose tracking is received under influences of dynamical oscillations of hand-eye cameras.

2) Define “Approaching Visual Servoing” and realize it by hand & eye-vergence dual control system.

Tasks in which visual information is used to direct a manipulator approach a target based on a time-varying relation ${}^E T_M(t)$ (end-effector frame is Σ_E , target frame is Σ_M) are referred to “Approaching Visual Servoing”. Take Fig.1 as an example, in Fig.1(a), the end-effector keep approaching a moved target object and finally stop near to it at time t^* ; and in Fig.1(b), the motion of the target object is shuttle rotation, the end-effector keep approaching the object through a curved pose tracking trajectory given by ${}^E T_M(t_1), \dots, {}^E T_M(t_n), \dots, {}^E T_M(t^*)$, finally it gets near to the object, and then grasping could be performed. Such kind of tasks in Fig.1 will be experimentally realized by a 7-link manipulator installed with a stereo mobile camera system, using the hand & eye-vergence dual control system.

II. LYAPUNOV-STABLE POSE TRACKING

A. “1-step GA + MFF” Pose Tracking Method

The GA-based scene recognition method described here can be designated as an “evolutionary recognition method”, since for every step of the GA’s evolution, it struggles to perform the recognition of a target in the input raw-image to the recognition system. To recognize a target input by CCD camera in real-time, and to avoid time lag waiting for the convergence to a target, we used GA in such manner that only one generation is processed to newly input image, which we called “1-Step GA”. In this way, the GA searching process and the convergence to the target does not consist in one image but the convergence is achieved in the sequence of the input image to recognize it in the continuously input images. While using the elitist model of the GA, the dynamic optimization of the fitness function in every new image, which corresponds to the recognition result of a target can be detected by the top genes in the GA in real-time.

MFF method has been proposed in [10] to compensate the target’s fictional motion seeing in the camera frame that comes from the hand-eye motion. Let matrix J_M describe how target pose change in Σ_E with respect to changing of the

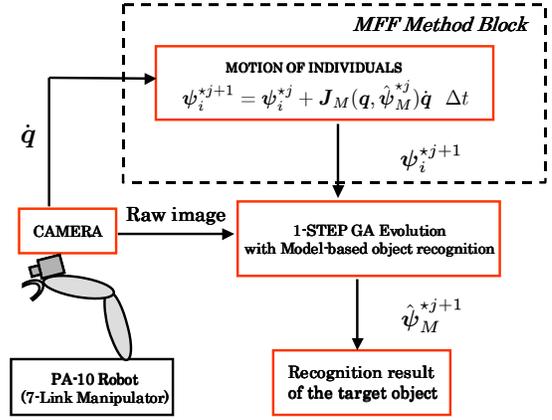


Fig. 2. “1-step GA + MFF” Pose Tracking

manipulator’s joint angles. Then, we can predict the target velocity in Σ_E ($\dot{\psi}$) based on the joint velocity \dot{q} of the manipulator, as

$$\dot{\psi} = J_M(q, \hat{\psi}(t))\dot{q}. \quad (1)$$

Then we use $\dot{\psi}$ to reset the GA individuals once they evolved by one generation, described in Eq.2. An individual of GA is defined as $\psi_i^j(t)$, which means the i -th gene ($i = 1, 2, \dots, p$) in the j -th generation. The mark “*” denotes the MFF reset individual. Δt is the time cost in one generation.

$$\psi_i^{*j+1}(t + \Delta t) = \psi_i^{*j}(t) + \dot{\psi}\Delta t. \quad (2)$$

By using (2), GA group can move together with the motion of the target in Σ_E , never loose it even under a high-speed moving of robot manipulator. Since the effect on the recognition from the dynamics of manipulator can be compensated, recognition by hand-eye cameras is independent of the dynamics of the manipulator, robust recognition can be obtained just like using fixed cameras. This is so called “1-step GA + MFF” pose tracking method, its flow chart is shown in Fig.2. In the next section, we will give a convergence proof through Lyapunov method.

B. Lyapunov-Stable Verify

Theoretically optimal pose $\psi_{max}(t)$ that gives the highest peak of $F(\psi(t))$ is defined as

$$\psi_{max}(t) = \{\psi(t) \mid \max_{\psi \in \mathbf{L}} F(\psi(t))\}, \quad (3)$$

where \mathbf{L} represents 6-DoF searching space.

The genes of GA individual represented by $\psi_i^j(t)$ is defined by binary strings, which are generated randomly in the initial population, with a given individual number p . Denote $\psi_{max}^{GA}(t)$ to be the maximum among the p genes of $\psi_i^j(t)$ in GA process,

$$\psi_{max}^{GA}(t) = \{\psi_i^j(t) \mid \max_{\psi_i^j \in \mathbf{L}} F(\psi_i^j(t))\}. \quad (4)$$

In fact we cannot always guarantee the best individual of GA $\psi_{max}^{GA}(t)$ should coincide with the theoretically optimal

pose $\psi_{max}(t)$, because the number of GA's individuals is not infinite. The difference between $\psi_{max}(t)$ and $\psi_{max}^{GA}(t)$ is denoted as

$$\delta\psi(t) = \psi_{max}(t) - \psi_{max}^{GA}(t). \quad (5)$$

And the difference between $F(\psi_{max}(t))$ and $F(\psi_{max}^{GA}(t))$ is denoted as

$$\Delta F(\delta\psi(t)) = F(\psi_{max}(t)) - F(\psi_{max}^{GA}(t)), \quad (6)$$

Since $F(\psi_{max}(t)) \geq F(\psi_{max}^{GA}(t))$, we have

$$\Delta F(\delta\psi(t)) \geq 0. \quad (7)$$

Based on the definition of $\Delta F(\delta\psi(t))$ in (6), in this research, we let GA's work in the following way:

- (a) GA evolves to minimize $\Delta F(\delta\psi(t))$.
- (b) The elitist individual of GA is preserved at every generation (elitist gene preservation strategy).
- (c) $\psi_{max}^{GA}(t)$ does keep the same value in the evolving when the evolved new gene with different value gives the same value of ΔF .

Here, we present two assumptions.

[Assumption 1] $\Delta F(\delta\psi(t))$ is positive definite.

This means the distribution of $F(\psi(t))$ satisfies $\Delta F(\delta\psi(t)) = 0$ if and only if $\delta\psi(t) = 0$, which indicates there is a single minimum in the searching space \mathbf{L} . ΔF is multipeak distribution having peaks and bottoms with limited number.

[Assumption 2] $\dot{F}(\psi_{max}^{GA}(t)) \geq 0$.

Differentiating (6) by time t , we have

$$\Delta \dot{F}(\delta\psi(t)) = \dot{F}(\psi_{max}(t)) - \dot{F}(\psi_{max}^{GA}(t)). \quad (8)$$

We defined $F(\psi_{max}(t)) = 1$ representing that the true pose of the target object gives the highest peak. Therefore, the time differentiation of $F(\psi_{max}(t))$ will be $\dot{F}(\psi_{max}(t)) = 0$. Thus, from (8) and [Assumption 2], we have

$$\Delta \dot{F}(\delta\psi(t)) = -\dot{F}(\psi_{max}^{GA}(t)) \leq 0. \quad (9)$$

$\psi_{max}^{GA}(t)$ represents current best GA solution. [Assumption 2] means GA can change its best gene $\psi_{max}^{GA}(t)$ to always reduce the value of ΔF regardless of dynamic image or static one, which indicates that the convergence speed to the target in the dynamically continuous images should be faster than the moving speed of the target object.

We cannot guarantee that the above two assumptions always hold, since they depend on some factors such as object's shape, object's speed, definition of $F(\psi(t))$, parameters of GA and viewpoint for observing, lightening environment, et al.. However, we can make efforts to improve the environment and correlation function and so on. Especially, the proposed "1-step GA + MFF" method effectively increases the convergence speed since it removes the GA group to compensate a part of the target motions coming from the hand-eye motion.

Providing above two assumptions be satisfied, (7) and (9) hold, then $\Delta F(\delta\psi(t))$ is so-called Lyapunov function. The

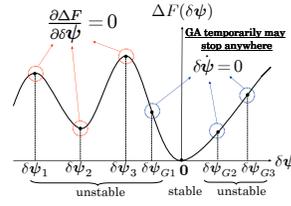


Fig. 3. The invariant set of the solutions of $\Delta \dot{F}(\delta\psi(t)) = 0$.

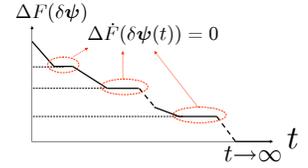


Fig. 4. The changing of $\Delta F(\delta\psi(t))$ with respect to time t in the whole GA's evolution.

objective here is to verify that $\delta\psi(t)$ asymptotically stable, resulting in it converges to $\mathbf{0}$ by using the Lyapunov function of $\Delta F(\delta\psi(t))$, meaning $\psi_{max}^{GA}(t) \rightarrow \psi_{max}(t)$, ($t \rightarrow \infty$), and the following shows how to verify it.

Since $\Delta \dot{F}(\delta\psi(t))$ is only negative semi-definite, in the view of LaSalle theorem, $\delta\psi(t)$ asymptotically converges to the invariant set of the solutions $\delta\psi$ satisfying $\Delta \dot{F}(\delta\psi(t)) = 0$. Considering the following expression,

$$\Delta \dot{F}(\delta\psi(t)) = \frac{\partial \Delta F}{\partial \delta\psi} \cdot \delta\dot{\psi} = 0, \quad (10)$$

the first part $\partial \Delta F / \partial \delta\psi$ describes partial differentiation of ΔF with respect to $\delta\psi$, implying steepest descending direction of ΔF in the space of $\delta\psi$; the second part $\delta\dot{\psi}$ describes the difference between the moving speed of the target object and the evolution speed of the best gene of GA, by the definition in (5).

Equation (10) shows the invariant set of the solutions of $\Delta \dot{F}(\delta\psi(t)) = 0$ includes (1): P_1 , the solution set of $\partial \Delta F / \partial \delta\psi = 0$; (2): P_2 , the solution set of $\delta\dot{\psi} = 0$; and (3): P_3 , the solution set satisfying $\partial \Delta F / \partial \delta\psi \neq \mathbf{0}$, $\delta\dot{\psi} \neq \mathbf{0}$, but their inner product is 0.

As shown in Fig. 3, P_1 includes the points of $\delta\psi$ that give the local maximum or minimum values of the function ΔF including $\mathbf{0}$. The number of these points is finite by [Assumption 1] denoted by p , that is

$$P_1 = \{\mathbf{0}, \delta\psi_1, \delta\psi_2, \dots, \delta\psi_{p-1}\}. \quad (11)$$

The evolving process of GA may stay temporarily at the same ΔF value. If the target object is static, it means the best gene of GA stop at some moments for the reason that the limited individuals of GA could not improve a current solution that gives a smaller fitness function value ΔF during some generations. And when the target object is moving, $\delta\dot{\psi} = 0$ means at these moments that the evolution speed of the best gene of GA is equal to the moving speed of the target object, by (5). The number of these points is assumed to be possibly finite, denoted by q . Thus, we describe the set of P_2 as

$$P_2 = \{\mathbf{0}, \delta\psi_{G1}, \delta\psi_{G2}, \dots, \delta\psi_{G(q-1)}\}. \quad (12)$$

Notice that there is another solution set of $\delta\psi$: P_3 . In this case, the vector of $\partial \Delta F / \partial \delta\psi$ is vertical to the vector of $\delta\dot{\psi}$ since the calculation $(\partial \Delta F / \partial \delta\psi) \cdot \delta\dot{\psi}$ in (10) means inner cross product, which means GA evolves in the direction that keeps a same fitness function value ΔF . This GA's evolution

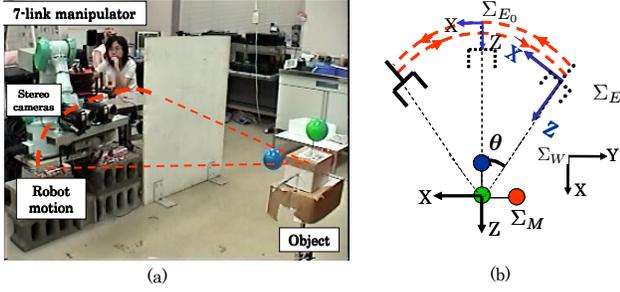


Fig. 5. (a)Experiment system (b) Coordinates system (overlook)

way is forbidden in this research for the GA's work rule (c) that we have stated above. Then, P_3 is null. So the invariant set that $\delta\psi(t)$ asymptotically converges to is

$$P = P_1 \cup P_2. \quad (13)$$

Here, $\delta\psi_1, \delta\psi_2, \dots, \delta\psi_{p-1}$ in P_1 are all unstable, because we define " $F(\psi(t))$ is positive definite and satisfies $\Delta F(\delta\psi(t)) = 0$ if and only if $\delta\psi(t) = 0$ " in [Assumption 1], which means $\Delta F(\delta\psi_i) > 0$ ($i = 1, 2, \dots, p-1$), and only $\delta\psi = 0$ gives minimum value of $\Delta F(\delta\psi(t))$, so only $\delta\psi = 0$ is stable. And in P_2 , all the points are unstable except the point 0 , for the reason that GA always has possibility to get out of these points by its evolving nature, which has been denoted in the GA's work way (a) that GA evolves to minimize $\Delta F(\delta\psi(t))$.

Therefore, 0 is the only stable point in the invariant set of P , that is, $\delta\psi(t)$ will finally converges to 0 . The image of the changing of $\Delta F(\delta\psi(t))$ with respect to time t in the whole GA's evolution is shown in Fig.4.

The above verification shows $\delta\psi(t) \rightarrow 0$, which means

$$\psi_{max}^{GA}(t) \rightarrow \psi_{max}(t), \quad (t \rightarrow \infty) \quad (14)$$

Let t_ϵ denotes a convergence time, then

$$|\delta\psi(t)| = |\psi_{max}(t) - \psi_{max}^{GA}(t)| \leq \epsilon, \quad (\epsilon > 0, t \geq t_\epsilon) \quad (15)$$

In (15), ϵ is tolerable extent that can be considered as an observing error. Thus, it is possible to realize real-time optimization, because $\psi_{max}^{GA}(t)$ can be assumed to be in the vicinity of the theoretically optimal $\psi_{max}(t)$ after t_ϵ .

C. Pose Tracking Experiment

Experimental verification of the proposed "1-step GA + MFF" method is shown here. The target object is a 3D marker that is composed of a red ball, a green ball and a blue ball to identify the 3-DoF orientation. The ball radius is 30[mm]. We compare the tracking result of using just "1-step GA" with that using "1-step GA + MFF" under a given trajectory of the end-effector with dynamical oscillation. To see clearly the effectiveness of the proposed method, here, we keep the target object static, so the target motion in the camera view is purely generated by the motion of the camera.

1) *Experimental Condition:* The robot used in this experimental system is a 7-Link manipulator, Mitsubishi Heavy Industries PA-10 robot. Two 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.

A trajectory of end-effector is given as a circle with a fixed distance to the target and keeping the eye-line (z axis of Σ_E) passes the center of the target, as shown in Fig. 5(a)(b). The initial hand pose is defined as Σ_{E_0} (static frame). The desired hand trajectory expressed in Σ_{E_0} is

$${}^{E_0}\psi_{Ed} = \begin{cases} {}^{E_0}x_{Ed}(t) = d * \sin(\theta_d(t)) \\ {}^{E_0}y_{Ed}(t) = 0 \\ {}^{E_0}z_{Ed}(t) = d - d * \cos(\theta_d(t)) \\ {}^{E_0}\epsilon_{1Ed}(t) = 0 \\ {}^{E_0}\epsilon_{2Ed}(t) = \sin\frac{\theta_d(t)}{2} \\ {}^{E_0}\epsilon_{3Ed}(t) = 0 \end{cases} \quad (16)$$

where $d = 950[mm]$, $\theta_d(t) = 15\sin(\omega t)[deg]$, ω represents the frequency of end-effector's motion. In this paper, we use unit quaternion to represent orientation because of the advantage of no singularities. Unit quaternion is composed of four parameters, since a parameter η can be determined by ϵ , we use only three parameters of ϵ to express the orientation, detailed refer to [14].

Due to transient response made by the dynamics of the manipulator, oscillation happens in the actual hand motion. This makes the tracking the object more difficult. The effectiveness of MFF method to solve this kind of problem will be evaluated in this experiment, where the target object is static in Σ_W , set as ${}^{E_0}\psi_M = [0, 60[mm], 950[mm], 0, 0, 0]^T$.

2) *Experimental Result:* We compare the methods of "1-step GA" and "1-step GA + MFF" under the above given trajectory of the end-effector, the moving speed of the end-effector is set as $\omega = 0.628[rad/s]$, corresponding to the motion period $T = 10[s]$. Time for the experimental process is $t = 0 \sim 40[s]$.

Fig. 6 shows the estimation results by using just "1-step GA" method, compared with the desired target pose in camera frame. The true values of the 3-D pose of the target object in Σ_E are ψ_M . The tracking results using only "1-step GA" without MFF method are represented by $\hat{\psi}_M$. Fig. 7 shows the estimation results by using "1-step GA + MFF" method, compared with the desired target pose in camera frame. The tracking results using "1-step GA + MFF" are denoted by $\hat{\psi}_M^*$. Due to the dynamics of the manipulator, the target object in the images includes the motion caused by hand oscillation. The cosine curves in Fig. 6 and 7(b), (c) are the obvious transient oscillation, and that also exist in (d) and (f) as the orientation oscillation. Such oscillations surely bring difficulty to object recognition.

Comparing the tracking results ψ_M using "1-step GA" and $\hat{\psi}_M^*$ using "1-step GA + MFF", we find that "1-Step GA" method cannot recognize precisely, especially during the oscillation period, which can be seen from the result that the dotted lines representing $\hat{\psi}_M(t)$ without MFF is not

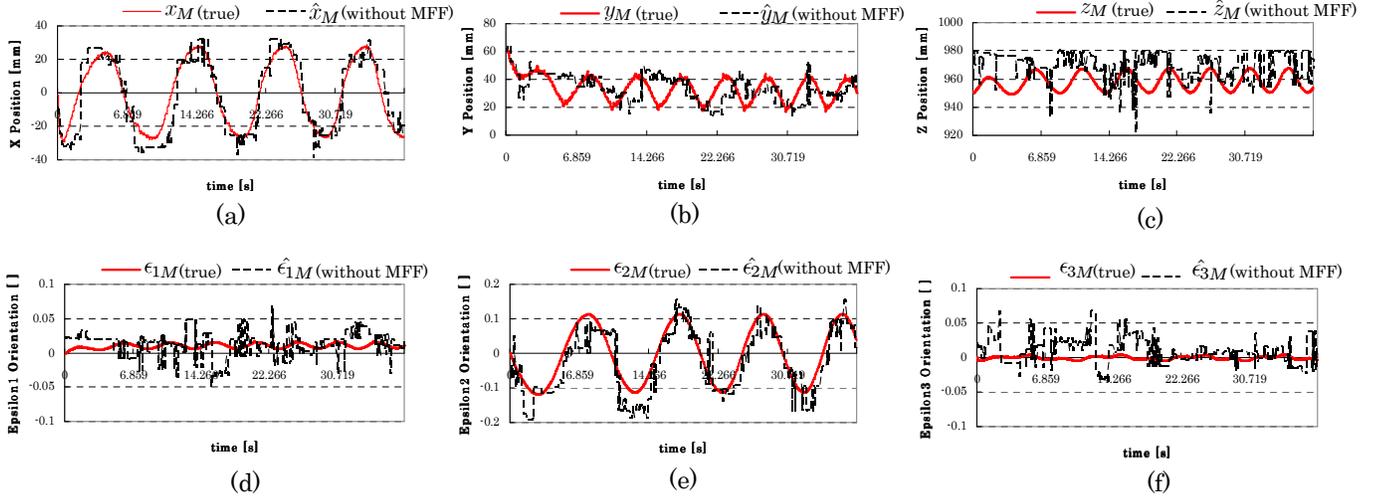


Fig. 6. Desired pose of the target ψ_M and the estimation results $\hat{\psi}_M$ by “1-step GA ” under hand-eye motion $\omega = 0.628[\text{rad/s}]$.

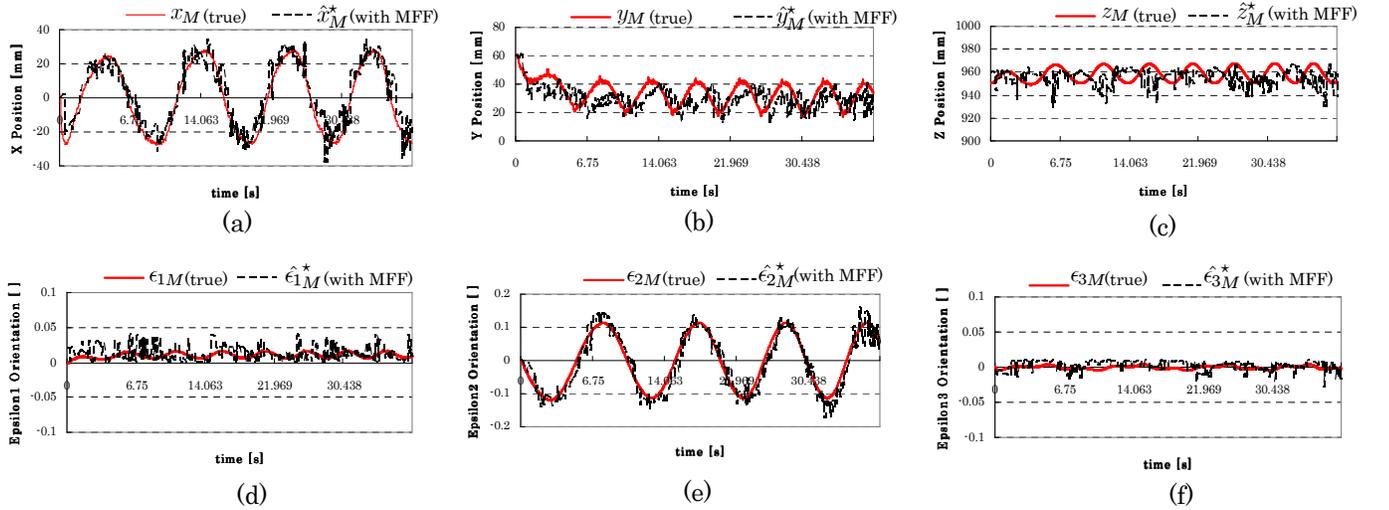


Fig. 7. Desired pose of the target ψ_M and the estimation results by $\hat{\psi}_M^*$ “1-step GA+MFF ” under hand-eye motion $\omega = 0.628[\text{rad/s}]$.

overlapping the true values of $\psi_M(t)$, as shown in Fig. 6. On the other hand, “1-step GA + MFF” method gives more correct result since the dotted lines representing $\hat{\psi}_M^*(t)$ using MFF almost overlaps the true values $\psi_M(t)$ with smaller pose error, as shown in Fig. 7. This experimental result confirmed the recognition accuracy and stability by using “1-step GA + MFF” method.

III. APPROACHING VISUAL SERVOING

A. Hand & Eye-vergence Controller

The block diagram of our proposed hand & eye-vergence dual control system is shown in Fig. 8, 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.

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

$${}^W \dot{r}_d = K_{P_p} {}^W r_{E,Ed} + K_{V_p} {}^W \dot{r}_{E,Ed}; \quad (17)$$

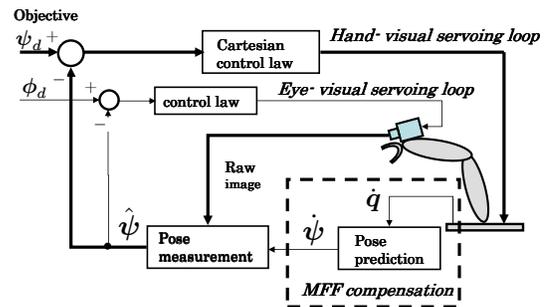


Fig. 8. Block diagram of the hand & eye-vergence visual servoing system

the desired hand angular velocity ${}^W \omega_d$ is calculated as

$${}^W \omega_d = K_{P_o} {}^W R_E {}^E \Delta\epsilon + K_{V_o} {}^W \omega_{E,Ed}, \quad (18)$$

where hand error ${}^W r_{E,Ed}$, error velocity ${}^W \dot{r}_{E,Ed}$, quaternion error ${}^E \Delta\epsilon$ and angular velocity ${}^W \omega_{E,Ed}$ are calculated by using the given servoing objective and the visual pose

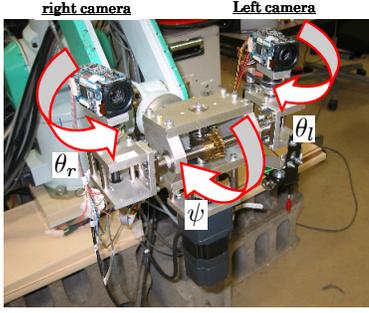


Fig. 9. Eye-vergence stereo camera system

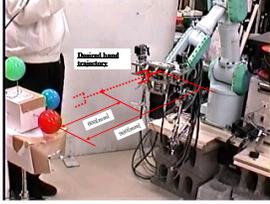


Fig. 10. A photograph of approaching visual servo system.

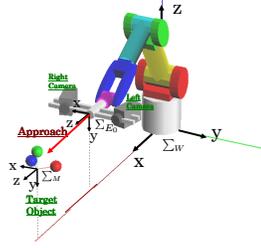


Fig. 11. Coordinate system of approaching visual servo system.

measurement, refer to [11].

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

$$\dot{q}_d = \mathbf{J}^+(\mathbf{q}) \begin{bmatrix} W \dot{r}_d \\ W \omega_d \end{bmatrix}. \quad (19)$$

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

$$\boldsymbol{\tau} = \mathbf{K}_{SP}(\dot{q}_d - \dot{\mathbf{q}}) + \mathbf{K}_{SI} \int_0^t (\dot{q}_d - \dot{\mathbf{q}}) dt \quad (20)$$

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

The eye-vergence visual servoing is the inner loop of the visual servoing system shown in Fig. 8. In this paper, we use stereo mobile camera for eye-vergence visual servoing. This mobile camera system has three DoF (Fig. 9), the left and right camera's poses are defined by $\phi_L = [\theta_l, \psi]^T$, $\phi_R = [\theta_r, \psi]^T$, where θ_l and θ_r are pan angles, and ψ is title angle that is common for both cameras. Here, the positions of cameras are supposed to be fixed.

The objective of the eye-vergence control is given by

$${}^{CL}\mathbf{u}_d = [{}^{CL}x_d, {}^{CL}y_d]^T = \mathbf{0}, \quad {}^{CR}\mathbf{u}_d = [{}^{CR}x_d, {}^{CR}y_d]^T = \mathbf{0}, \quad (21)$$

which means the objects is desired to be in the center of the camera frame. We define ${}^{CL}\hat{\mathbf{u}}$ is the x and y direction of ${}^{CL}\hat{\boldsymbol{\psi}}$ (object pose in left camera), and ${}^{CR}\hat{\mathbf{u}}$ is the x and y direction of ${}^{CR}\hat{\boldsymbol{\psi}}$ (object pose in right camera), then the controller of eye-visual servoing is given by

$$\dot{\phi}_L = \mathbf{K}_{PL}({}^{CL}\mathbf{u}_d - {}^{CL}\hat{\mathbf{u}}) + \mathbf{K}_{DL}({}^{CL}\dot{\mathbf{u}}_d - {}^{CL}\dot{\hat{\mathbf{u}}}), \quad (22)$$

$$\dot{\phi}_R = \mathbf{K}_{PR}({}^{CR}\mathbf{u}_d - {}^{CR}\hat{\mathbf{u}}) + \mathbf{K}_{DR}({}^{CR}\dot{\mathbf{u}}_d - {}^{CR}\dot{\hat{\mathbf{u}}}), \quad (23)$$

where \mathbf{K}_{PL} , \mathbf{K}_{DL} , \mathbf{K}_{PR} , \mathbf{K}_{DR} are positive control gain.

B. Experiment of Approaching Visual Servoing

Here, we conduct the experiments of Approaching Visual Servoing to a 3D marker by using the hand & eye-vergence controller.

1) *Experimental Condition*: A photograph of our experimental system is shown in Fig.10. 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.11 shows the coordinate system corresponding to Fig.10. The initial pose of the end-effector is defined as Σ_{E_0} , and given by

$${}^W\mathbf{T}_{E_0} = \begin{bmatrix} 0 & 0 & 1 & -918 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 455 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (24)$$

position unit: [mm].

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

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

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

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

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.10.

Figs.12(a) to (f) show the actual motion of the end-effector with respect to the fixed frame of Σ_{E_0} , defined as ${}^{E_0}\boldsymbol{\psi}_E$, compared with the desired hand pose ${}^{E_0}\boldsymbol{\psi}_{Ed}$. As shown in Fig.12(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 ${}^{E_0}z_{Ed} = 300[mm]$ after 40[s]. The actual motion of the end-effector shown in Fig.12(c) confirmed that this approaching motion was achieved. The errors between the desired hand pose ${}^{E_0}\boldsymbol{\psi}_{Ed}$ and the actual hand pose ${}^{E_0}\boldsymbol{\psi}_E$ are limited in a small range. 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,

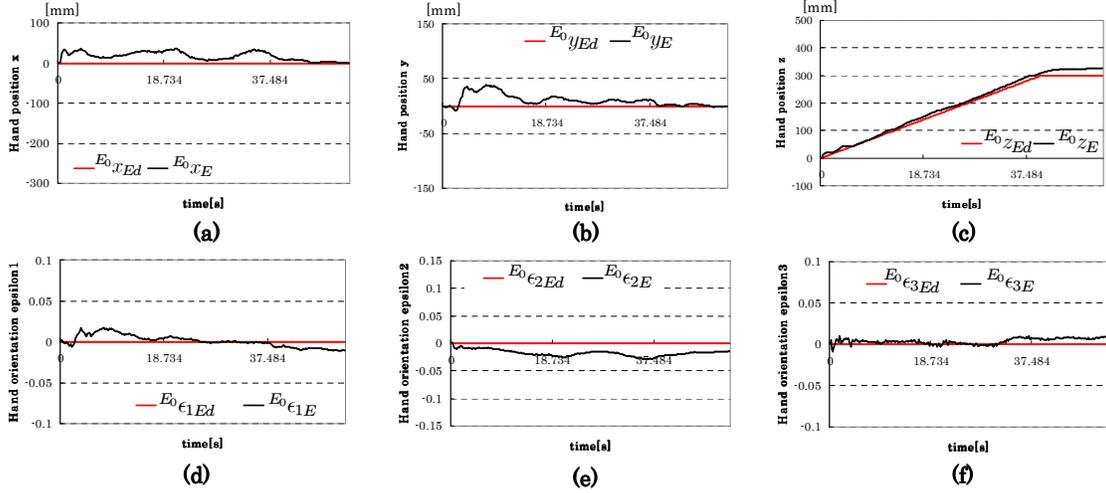


Fig. 12. Hand pose of approaching visual servoing in z-axis, using hand & eye-vergence dual control system. We use millimeter to measure position. When using quaternion to express the orientation of an object, no unit, just values. The object rotation of 1[deg] around x axis corresponding to quaternion representation as $\epsilon_1 = 0.008$, $\epsilon_2 = 0$, $\epsilon_3 = 0$.

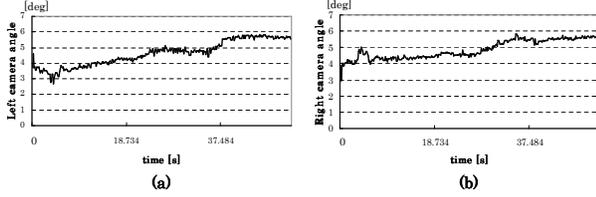


Fig. 13. Camera pose of approaching visual servoing in z-axis.

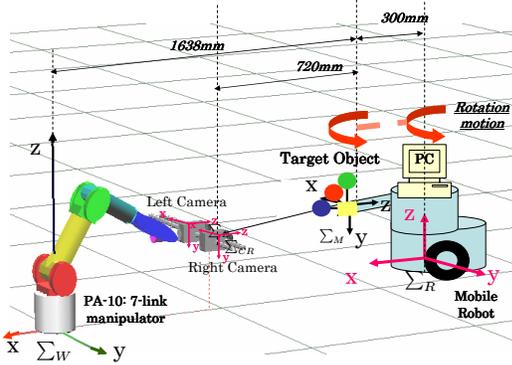


Fig. 14. Approaching Visual Servoing to a moving object, which is fixed on a mobile robot.

which has been confirmed by Figs.13(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], 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.

3) *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. 14. 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

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

where we set $a = 8[\text{deg}]$, $T = 40[\text{s}]$. 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.

So, here Approaching Visual Servoing to a moving target object is performed. The objective of visual servoing is the same as the first experiment, given in (25), (26), but here we set $z_{max} = 900[\text{mm}]$, $z_{min} = 550[\text{mm}]$, $T = 60[\text{s}]$.

Figs. 15(a) to (f) are the experimental results, which show the actual motion of the end-effector defined as ${}^{E0}\psi_E(t)$, compared with the desired hand pose ${}^{E0}\psi_{Ed}(t)$. 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 ${}^{E0}\psi_{Ed}(t)$ 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. 15(a),(e) began to turn to sin/cos curved line.

As shown in Fig.15(c), the end-effector is desired to move 350[mm] in z -axis of Σ_{E0} in the first 60[s]; and keep 350[mm] distance to the target object, no more approach, that is ${}^{E0}z_{Ed} = 350[\text{mm}]$ after 60[s]. The motion image is shown in Fig.1(b). The actual motion of the end-effector shown in Fig.15(c) confirmed that this approaching motion was achieved. And the actual motion of the end-effector shown in Fig.15(a),(e) confirmed that the tracking of the

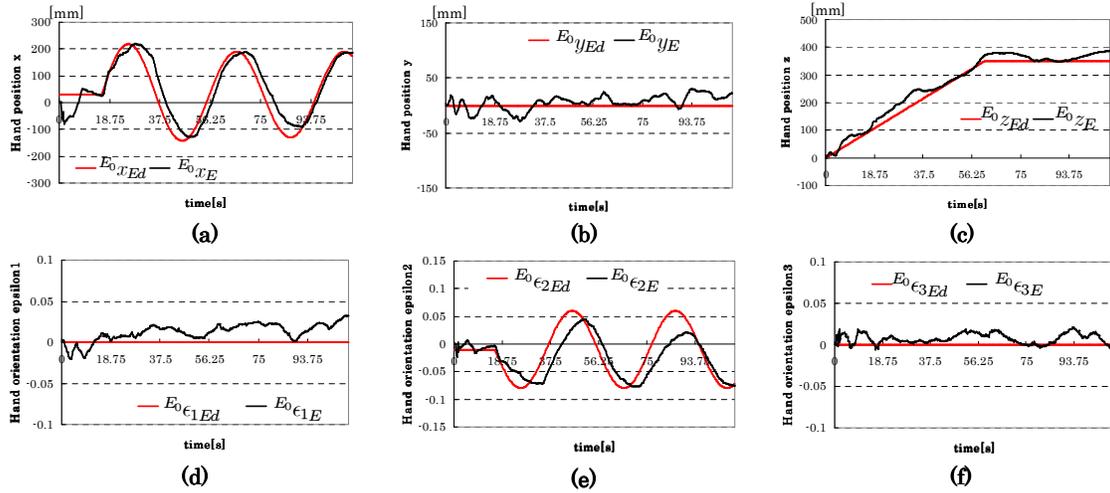


Fig. 15. Hand pose of Approaching Visual Servoing to a moving object, using hand & eye-vergence dual control system.

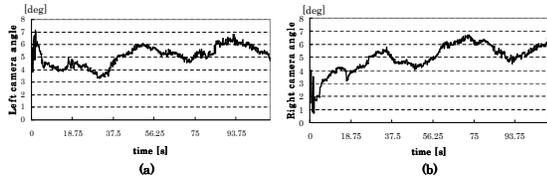


Fig. 16. Camera pose of Approaching Visual Servoing a moving object.

rotating object was achieved, with about 5[s] time delay. The errors between the desired hand pose $E^0\psi_{Ed}$ and the actual hand pose $E^0\psi_E$ are limited in an acceptable range.

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.16(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 that tracking and approaching a moving object can be done by our proposed hand & eye-vergence dual control system, which is a meaningful preparation work for grasping.

IV. CONCLUSION

Intelligent robots are expected to do more service for people, the basic work is grasping. This paper represented two contributions for grasping.

1) A convergence of the 3-D pose tracking error in successively input images is discussed, and we clearly stated what kind of condition leads to Lyapunov-stable pose tracking. Our proposed “1-step GA + MFF” method is verified by both Lyapunov analysis and experiments that shows stable tracking can be realized even under the interference of quick hand-eye motion.

2) “Approaching Visual Servoing” is defined and experimentally realized by using hand & eye-vergence dual control system. This enabled robot end-effector to track an object, meanwhile, to approach it with a suitable posture, which is a part of work for grasping.

In the near future, we have to deal with the problem that how to catch up with a moving object. Object velocity

prediction should be performed to make sure the hand moves faster than the object, else it could not catch it. Other problems, like obstacle avoidance during grasping process, grasping force control are also the subjects we want to research on in future.

REFERENCES

- [1] Nikolaos P. Papanikolopoulos, Pradeep K. Khosla, and Takeo Kanade, *Visual Tracking of a Moving Target by a Camera Mounted on a Robot: A Combination of Control and Vision*, IEEE Trans. on Robotics and Automation, VOL. 9, NO. 1, FEBRUARY 1993.
- [2] R. C. Luo, R. E. Mullen Jr., and D. E. Wessel, *An adaptive robotic tracking system using optical flow*, Proc. IEEE Int. Conf. Robotics Automat., 1988, pp. 568-573.
- [3] J. T. Feddema and C.S.G. Lee, *Adaptive image feature prediction and control for visual tracking with a hand-eye coordinated camera*, IEEE Trans. Syst. Man Cybern., vol. 20, no. 5, pp. 1172-1183, 1990.
- [4] 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.
- [5] E.Malis, F.Chaumette and S.Boudet, *2-1/2-D Visual Servoing*, IEEE Trans. on Robotics and Automation, vol. 15, no. 2, pp. 238-250, 1999.
- [6] 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.
- [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] R.Yang and Z.Zhang, *Model-based head pose tracking with stereovision*, Proc.5th IEEE Int.Conf.on Automatic Face and Gesture Recognition, 2002.
- [9] S.Yamane, M.Izumi, K.Fukunaga, *A Method of Model-Based Pose Estimation*, IEICE, Vol.J79-D-2, No.2, pp.165-173, Feb, 1996.
- [10] Wei Song and Mamoru Minami, *Stability / Precision Improvement of 6-DoF Visual Servoing by Motion Feedforward Compensation and Experimental Evaluation*, IEEE Int. Conf. on Robotics and Automation (ICRA), pp.722-729, 2009.
- [11] Wei Song and Mamoru Minami, *Hand & Eye-Vergence Dual Visual Servoing to Enhance Observability and Stability*, IEEE Int. Conf. on Robotics and Automation (ICRA), pp.714-721, 2009.
- [12] E. Mails, *Improving vision-based control using efficient second-order minimization techniques* in Proceedings of IEEE International Conference on Robotics and Automation, (New Orleans, USA), pp. 1843-1848, Apr. 6-May 1, 2004.
- [13] S. Benhimane and E. Mails, *Real-time image-based tracking of planes using efficient second-order minimization* in Proceedings of IEEE/RSJ Conference on Intelligent Robot and Systems, (Sendai, Japan), pp. 1090-1097, Sep 28-Oct 2, 2004.
- [14] B.Siciliano and L.Villani: *Robot Force Control*, ISBN 0-7923-7733-8.