2012 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 7-12, 2012. Vilamoura, Algarve, Portugal

Robust Visual Servoing for Object Manipulation with Large Time-Delays of Visual Information

Akihiro Kawamura, Kenji Tahara, Ryo Kurazume and Tsutomu Hasegawa

Abstract— **This paper proposes a new visual servoing method for object manipulation robust to considerable time-delays of visual information. There still remain several problems in visual servoing methods although they are quite useful and effective for dexterous object manipulation. For instance, time-delays to obtain necessary information for object manipulation from visual images induce unstable behavior. The time-delays are mainly caused by low sampling rate of visual sensing system, computational cost for image processing, and latency of data transmission from visual sensor to processor. The method makes it possible to avoid such unstable behavior of the systems due to considerable time-delays using virtual object frame defined by only each joint angle. Firstly, a new control scheme for object manipulation using the virtual object frame is designed. Next, numerical simulations are conducted to verify the effectiveness of the control scheme. Finally, experimental results are shown to demonstrate the practical usefulness of proposed method.**

I. Introduction

Visual information is useful for dexterous object manipulation by a sensory feedback controller [1–4]. However in general, time-delays to obtain necessary information from visual images should be considered carefully since they make a visual servoing system unstable. The time-delays are mainly caused by low sampling rate, computational cost, and data transmission latency. It is known that the sampling rate of a standard NTSC camera which is often used in the visual sensing system is fixed at 30 [Hz]. However, it is not enough to use directly in the visual servoing since the servo-loop generally requires a few milliseconds ordered sampling rate. In addition, other inevitable time-delays are caused by a latency of data transmission from a visual sensor to a processor, and a computational cost to extract necessary information from images.

To date, several methods have been proposed to improve the robustness of the visual servoing against these time-delays. They are based on state estimation techniques, such as Kalman filter [5], AR model [6], nonlinear observer [7] and multi-rate control [8], [9]. The state estimation techniques are based on an assumption that the overall dynamics can be approximated as a linear system. However, the overall dynamics of an object manipulation system is too complicated to be approximated as a linear system since many geometrical and physical constraints have to be taken into consideration. Also in these methods, reaching or moving tasks of the end-effector are mainly addressed and the effect of the time-delays for dexterous object manipulation has not been discussed.

Recently, we have proposed a stable object grasping and manipulation method for an arbitrary polyhedral object using a multi-fingered hand-arm system [10], [11]. The advantage of the method is that only proprioceptive sensory information, such as joint angles or angular velocities, is required for stable object grasping and no other information such as the object shape and grasping points is necessary. After that, it has been extended to a robust visual servoing method for object manipulation against temporary loss of visual information when occlusion occurs or the object goes out-of-sight [12]. However, any time-delay of visual information is not considered in this method explicitly.

In this paper, a new visual servoing method for object manipulation by a multi-fingered hand-arm system is proposed. The new method is robust not only to the temporary loss of the visual information, but also the time-delays of the visual information. This method utilizes a virtual object frame that is defined by a position and attitude of each fingertip obtained by proprioceptive sensors similar to the previous method [12]. The difference between the new one and previous one is that information of an actual object frame obtained by a visual sensor is not directly used in a feedback controller, but used in the design of desired virtual object frame in the new one. Namely, the new controller is designed so as to manipulate the position and attitude of the virtual object frame toward its desired value designed by the information of the actual object frame.

In what follows, the overall model composed of a multi-fingered hand-arm system and a grasped object is shown in section II. The virtual object frame is given in section III. The new control scheme for robust object manipulation using the information of both the virtual and actual object frame is designed in section IV. Numerical simulation results are shown to verify the performance in section V, and experimental results to show the practical usefulness are presented in section VI.

II. A Multi-Fingered Hand-Arm System and a Grasped object

In this section, a hand-arm system composed of an arm and a multi-fingered hand is given. In this study, assume that a robotic hand-arm system has enough number of fingers and DOFs to achieve stable object grasping and manipulation

This work was partially supported by Grant-in-Aid for JSPS Fellows.

A. Kawamura is with Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, 819-0395, JAPAN kawamura@irvs.ait.kyushu-u.ac.jp

K. Tahara is with Faculty of Engineering, Kyushu University, Fukuoka, 819-0395, JAPAN tahara@ieee.org

R. Kurazume and T. Hasegawa are with Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, 819-0395, JAPAN {kurazume, hasegawa}@ait.kyushu-u.ac.jp

978-1-4673-1736-8/12/S31.00 ©2012 IEEE

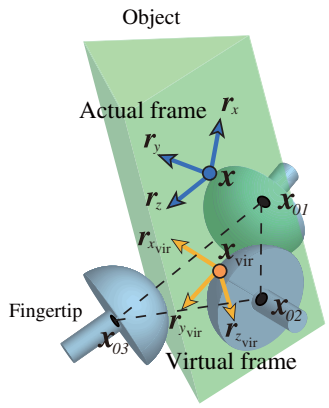

Fig. 1. Actual and virtual object frame

of a grasped object. The symbol N_a and N_i denote the number of DOFs of the arm and the i^{th} finger respectively, and $N_D = N_a + \sum_{i=1}^{N} N_i$ is the total number of DOFs in the system where N is the number of the fingers. The subscript of i in all variables and equations refers to the i^{th} finger hereinafter. Each fingertip is hemispheric and made of some soft material. Then, assume that all fingertips maintain rolling contact with the object surfaces, and do not slip and detach from the surfaces during manipulation. Also assume that fingertips roll within the ranges of its hemispheric surfaces, and they do not deviate from each initial contact surface. Note that the gravity effect of a grasped object is ignored for the sake of modeling easier at this stage.

III. Virtual Object Frame

The virtual object frame has been reported by Wimböck *et al.* [13] and Tahara *et al.* [14], independently. In these methods, the position and attitude of the virtual object frame defined by a position of each fingertip is utilized as controlled variables instead of an actual object frame. We also have proposed another virtual object frame defined not only by a position, but also an attitude of each fingertip in our previous work [12]. In this section, the virtual object frame defined in our previous work [12] is introduced.

Firstly, the position of the virtual frame $\boldsymbol{x}_{\text{vir}}$ is defined as

$$\boldsymbol{x}_{\text{vir}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_{0i} \in \mathbb{R}^3, \tag{1}$$

where $\boldsymbol{x}_{0i} \in \mathbb{R}^3$ denotes the position of the center of each fingertip. Equation (1) means that $\boldsymbol{x}_{\text{vir}}$ is the centroid of a polygon fixed by $\boldsymbol{x}_{0i}$. The attitude of the virtual object frame located at $\boldsymbol{x}_{\text{vir}}$ is expressed as a rotational matrix $\boldsymbol{R}_{\text{vir}} = [\boldsymbol{r}_{x_{\text{vir}}}, \boldsymbol{r}_{y_{\text{vir}}}, \boldsymbol{r}_{z_{\text{vir}}}] \in \mathrm{SO}(3)$ as shown in Fig. 1, and its components are given in the following way:

$$\boldsymbol{r}_{x_{\text{vir}}} = \frac{\tilde{\boldsymbol{r}}_{x_{\text{vir}}}}{\|\tilde{\boldsymbol{r}}_{x_{\text{vir}}}\|} \in \mathbb{R}^3, \quad \left(\tilde{\boldsymbol{r}}_{x_{\text{vir}}} = \textstyle\sum_{i=1}^{N} \boldsymbol{r}_{x_{\text{f}i}}\right) \tag{2}$$

$$\boldsymbol{r}_{y_{\text{vir}}} = \frac{\tilde{\boldsymbol{r}}_{x_{\text{vir}}} \times \tilde{\boldsymbol{r}}_{y_{\text{vir}}}}{\|\tilde{\boldsymbol{r}}_{x_{\text{vir}}} \times \tilde{\boldsymbol{r}}_{y_{\text{vir}}}\|} \in \mathbb{R}^3, \quad \left(\tilde{\boldsymbol{r}}_{y_{\text{vir}}} = \textstyle\sum_{i=1}^{N} \boldsymbol{r}_{y_{\text{f}i}}\right) \tag{3}$$

$$\boldsymbol{r}_{z_{\text{vir}}} = \boldsymbol{r}_{x_{\text{vir}}} \times \boldsymbol{r}_{y_{\text{vir}}} \in \mathbb{R}^3, \tag{4}$$

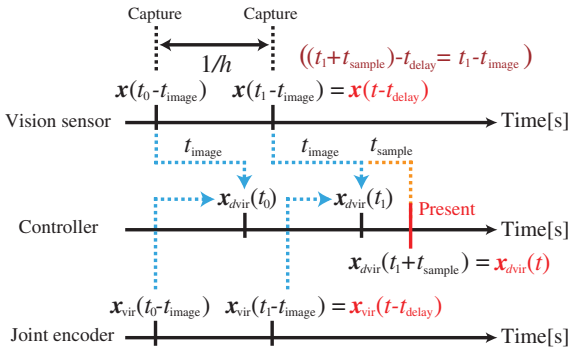

Fig. 2. Time-line chart of the vision sensor, the servo-loop of the proposed controller and the proprioceptive sensor at each joint (rotary encoder)

where $\boldsymbol{R}_{\text{f}i} = [\boldsymbol{r}_{x_{\text{f}i}}, \boldsymbol{r}_{y_{\text{f}i}}, \boldsymbol{r}_{z_{\text{f}i}}] \in \mathrm{SO}(3)$ indicates a local frame located at the position of the center of each fingertip $\boldsymbol{x}_{0i}$, and denotes its attitude in Cartesian coordinates.

IV. Control Law

In this section, a control input to regulate position and attitude of a grasped object by visual servoing with considerable time-delays is designed. The basic idea of the design of the input is based on our previous method that is robust to the temporary loss of the visual information using the virtual object frame [12]. The advantage of the new one compared with the previous one is that it is robust not only to the temporary loss of the visual information, but also to the considerable time-delays. This advantage is brought by the use of the virtual object frame as a controlled variable all of the time during manipulation. Its availability is nothing to do with such temporary loss and time-delay of visual information because it is defined only using the information from proprioceptive sensors. Therefore, the new method makes the overall system more stable than the previous method. In addition, a desired virtual object frame is continuously designed and updated using the information of the actual object frame from the visual sensor. The error between the actual object frame and the virtual object frame is reduced in real-time by the updating manner.

A. Time-Delay

In this paper, assume that the time-delays are caused by low sampling rate, computational cost for image processing, and data transmission latency, and that their time-delay can be measured roughly in real-time. Note that any unstable behavior is not induced even if the measured delay time is not accurate since the delay time is not related to the current value of the virtual object frame. In this method, the time-delay due to the low sampling rate is considered independently of other time-delays induced by the computational cost for image processing, and the data transmission latency. This is because these features and their countermeasures are different. The length of the time-delay due to the low sampling rate is fluctuated by a timing of when an instantaneous visual image is captured. For example, the time-delay is fluctuated from 0 [ms] to 33 [ms] in the case of using 30 [Hz] NTSC camera. The variance of it is now

expressed as $t_{\rm sample}$, and it satisfies the following inequality.

$$0 \leq t_{\rm sample} < \frac{1}{h}, \tag{5}$$

where h denotes the sampling rate.

The other time-delay caused by the computational cost for image processing, and data transmission latency is expressed as $t_{\rm image}$, and the total time-delay is expressed as $t_{\rm delay}$. It is given as

$$t_{\rm delay} = t_{\rm sample} + t_{\rm image}. \tag{6}$$

The sampling rate h and the total time-delay $t_{\rm delay}$ is shown in Fig. 2.

B. Control Input

The proposed control input $\boldsymbol{u}(t) \in \mathbb{R}^{N_D}$ is given by the summation of three control inputs, one is for stable object grasping $\boldsymbol{u}_s(t)$, another is for position control $\boldsymbol{u}_p(t)$, and the other is for attitude control $\boldsymbol{u}_o(t)$. Eventually the total control input $\boldsymbol{u}(t)$ is given as

$$\boldsymbol{u}(t) = \boldsymbol{u}_s(t) + \boldsymbol{u}_p(t) + \boldsymbol{u}_o(t). \tag{7}$$

The control input for stable grasping $\boldsymbol{u}_s(t)$ is designed so that the center of each fingertip approaches each other, and it accomplishes stable grasping of an arbitrary polyhedral object [10]. It is given as follows:

$$\boldsymbol{u}_s(t) = K_s \sum_{i=1}^{N} \boldsymbol{J}_{0i}(t)^{\rm T} \left(\boldsymbol{x}_{\rm vir}(t) - \boldsymbol{x}_{0i}(t)\right) - \boldsymbol{C}\dot{\boldsymbol{q}}(t) + \boldsymbol{g}(t) \tag{8}$$

$$K_s = \frac{f_d}{\sum_{j=1}^{N} r_j}, \tag{9}$$

where $\boldsymbol{J}_{0i}(t) \in \mathbb{R}^{3\times N_D}$ denotes the Jacobian matrix for the velocity of the position of the center of each fingertip $\boldsymbol{x}_{0i}$ with respect to each joint angular velocity $\dot{\boldsymbol{q}}$, $\boldsymbol{C} \in \mathbb{R}^{N_D\times N_D} > 0$ is a positive definite diagonal matrix that plays a role of the damping gain for each joint, $\boldsymbol{g}(t) \in \mathbb{R}^{N_D}$ signifies gravity compensation term for the robot, and f_d indicates a nominal desired grasping force.

The control input to regulate the position of the virtual object frame $\boldsymbol{u}_p(t)$ is designed as

$$\boldsymbol{u}_p(t) = K_p \sum_{i=1}^{N} \boldsymbol{J}_{0i}(t)^{\rm T} \left(\boldsymbol{x}_{d_{\rm vir}}(t) - \boldsymbol{x}_{\rm vir}(t)\right), \tag{10}$$

where K_p is a positive scalar constant, $\boldsymbol{x}_{d_{\rm vir}}(t)$ denotes a desired position of the virtual object frame and it is designed by considering the time-delays in the following way:

$$\boldsymbol{x}_{d_{\rm vir}}(t) = \boldsymbol{x}_{\rm vir}(t - t_{\rm delay}) + \left(\boldsymbol{x}_d - \boldsymbol{x}(t - t_{\rm delay})\right), \tag{11}$$

where $\boldsymbol{x}_d$ denotes a desired position of the actual object frame, $\boldsymbol{x}(t - t_{\rm delay})$ indicates a position of the actual object position which is obtained by a visual sensor at $t_{\rm delay}$ [s] before, and $\boldsymbol{x}_{\rm vir}(t - t_{\rm delay})$ stands for a position of the virtual object frame which is obtained at the same timing of obtaining $\boldsymbol{x}(t - t_{\rm delay})$. The relationship between these values is shown in Fig. 2.

The control input to regulate the attitude of the virtual object frame $\boldsymbol{u}_o(t)$ is designed as

$$\begin{aligned}\boldsymbol{u}_o(t) =& K_o \sum_{i=1}^{N} \boldsymbol{J}_{\Omega i}(t)^{\rm T} \left\{\boldsymbol{r}_{x_{\rm vir}}(t) \times \boldsymbol{r}_{xd_{\rm vir}}(t)\right. \\ &\left. + \boldsymbol{r}_{y_{\rm vir}}(t) \times \boldsymbol{r}_{yd_{\rm vir}}(t) + \boldsymbol{r}_{z_{\rm vir}}(t) \times \boldsymbol{r}_{zd_{\rm vir}}(t)\right\}, \end{aligned} \tag{12}$$

where $K_o > 0$ is a positive scalar constant, $\boldsymbol{J}_{\Omega_i}(t) \in \mathbb{R}^{3\times N_D}$ denotes the Jacobian matrix for the attitude angular velocity of each fingertip with respect to each joint angular velocity $\dot{\boldsymbol{q}}$. The desired attitude of the virtual object frame is expressed as a rotational matrix $\boldsymbol{R}_{d_{\rm vir}}(t) = [\boldsymbol{r}_{xd_{\rm vir}}(t), \boldsymbol{r}_{yd_{\rm vir}}(t), \boldsymbol{r}_{zd_{\rm vir}}(t)] \in {\rm SO(3)}$. The summation of each cross product in (12) implies a desired instantaneous rotational axis of the virtual object. Namely, the attitude error between the present virtual object frame and the desired virtual object frame can be reduced by rotating the virtual object frame around the axis. The desired attitude of the virtual object frame $\boldsymbol{R}_{d_{\rm vir}}(t)$ is designed in the following manner:

$$\boldsymbol{R}_{d_{\rm vir}}(t) = \boldsymbol{R}_d \boldsymbol{R}(t - t_{\rm delay})^{\rm T} \boldsymbol{R}_{\rm vir}(t - t_{\rm delay}), \tag{13}$$

where the desired attitude of the actual object frame is expressed as a rotational matrix $\boldsymbol{R}_d = [\boldsymbol{r}_{xd}, \boldsymbol{r}_{yd}, \boldsymbol{r}_{zd}] \in {\rm SO(3)}$. The attitudes of both the actual and the virtual object frame $t_{\rm delay}$ [s] one time-step before the present time are expressed as $\boldsymbol{R}(t - t_{\rm delay})$ and $\boldsymbol{R}_{\rm vir}(t - t_{\rm delay})$ respectively, and they are obtained at the same timing of obtaining $\boldsymbol{x}(t - t_{\rm delay})$.

C. Conditions for convergence of a desired state

In order to make the position and attitude of the actual object frame converge to each desired value in the final state by the control signal $\boldsymbol{u}(t)$, it is necessary to fulfill the following inequalities. They are given as follows:
For the object position:

$$\begin{aligned}&\left|\left\{\boldsymbol{x}(t_B) - \boldsymbol{x}_{\rm vir}(t_B)\right\} - \left\{\boldsymbol{x}(t_A) - \boldsymbol{x}_{\rm vir}(t_A)\right\}\right| \\ &\qquad \leq \left|\boldsymbol{x}(t_B) - \boldsymbol{x}(t_A)\right|, \end{aligned} \tag{14}$$

For the object attitude:

$$\left|\cos^{-1}\left\{{\rm tr}\left(\frac{\boldsymbol{R}_\alpha - 1}{2}\right)\right\}\right| \leq \left|\cos^{-1}\left\{{\rm tr}\left(\frac{\boldsymbol{R}_\beta - 1}{2}\right)\right\}\right|, \tag{15}$$

where

$$\boldsymbol{R}_\alpha = \left(\boldsymbol{R}(t_B)\boldsymbol{R}_{\rm vir}(t_B)^{\rm T}\right)\left(\boldsymbol{R}(t_A)\boldsymbol{R}_{\rm vir}(t_A)^{\rm T}\right)^{\rm T} \tag{16}$$

$$\boldsymbol{R}_\beta = \boldsymbol{R}(t_B)\boldsymbol{R}(t_A)^{\rm T}, \tag{17}$$

and both t_A and t_B denote an instantaneous arbitrary time between one time-step and the next time-step of an acquisition interval of the visual image respectively, and they satisfy the following inequality.

$$0 \leq t_B - t_A \leq \frac{1}{h}. \tag{18}$$

Equations (14) and (15) mean that the error norm between the difference of the actual object frame and the virtual object frame at t_A and that at t_B is smaller than that between the

TABLE I

PHYSICAL PARAMETERS FOR THE SIMULATION

Triple-fingered hand-arm system

1^{st} link length	l_{a1}	1.300[m]	l_{i1}	0.300[m]
2^{nd} link length	l_{a2}	1.000[m]	l_{i2}	0.200[m]
3^{rd} link length	l_{a3}	0.175[m]	l_{i3}	0.140[m]
1^{st} mass center	l_{ga1}	0.650[m]	l_{gi1}	0.150[m]
2^{nd} mass center	l_{ga2}	0.500[m]	l_{gi2}	0.100[m]
3^{rd} mass center	l_{ga3}	0.0875[m]	l_{gi3}	0.070[m]
1^{st} mass	m_{a1}	1.300[kg]	m_{i1}	0.250[kg]
2^{nd} mass	m_{a2}	1.000[kg]	m_{i2}	0.150[kg]
3^{rd} mass	m_{a3}	0.400[kg]	m_{i3}	0.100[kg]
1^{st} Inertia $\boldsymbol{I}_{a1}$	$\text{diag}(7.453, 7.453, 0.260)\times10^{-1}$[kg·m^2]			
2^{nd} Inertia $\boldsymbol{I}_{a2}$	$\text{diag}(3.397, 3.397, 0.128)\times10^{-1}$[kg·m^2]			
3^{rd} Inertia $\boldsymbol{I}_{a3}$	$\text{diag}(0.291, 0.291, 0.500)\times10^{-1}$[kg·m^2]			
1^{st} Inertia $\boldsymbol{I}_{i1}$	$\text{diag}(7.725, 7.725, 0.450)\times10^{-3}$[kg·m^2]			
2^{nd} Inertia $\boldsymbol{I}_{i2}$	$\text{diag}(2.060, 2.060, 0.120)\times10^{-3}$[kg·m^2]			
3^{rd} Inertia $\boldsymbol{I}_{i3}$	$\text{diag}(0.538, 0.538, 0.031)\times10^{-3}$[kg·m^2]			
Radius of fingertip r_i	0.070[m]			
Stiffness coefficient k_i	1.000×10^5[N/m^2]			
Damping function ξ_i	$1.000\times\left(r_i^2-\Delta r_i^2\right)\pi$[Ns/m^2]			

Object

Mass m	0.037[kg]
Y_1	0.092[m]
Y_2	0.048[m]
Y_3	0.048[m]
θ_{t1}	1.309[rad]
θ_{t2}	1.309[rad]
θ_{t3}	0.524[rad]
Inertia $\boldsymbol{I}$	$\text{diag}\,(1.273, 0.193, 1.148)\times10^{-3}$[kg · m^2]

TABLE II

NOMINAL DESIRED GRASPING FORCE AND GAINS FOR THE SIMULATION

f_d	10.0[N]
K_p	4.762
K_o	0.238
$\boldsymbol{C}_a$	$\text{diag}(1.003, 0.651, 0.735, 0.278, 0.177)\times10^{-1}$[Ns·m/rad]
$\boldsymbol{C}_1$	$\text{diag}(0.606, 0.687, 0.786, 0.642, 0.198)\times10^{-2}$[Ns·m/rad]
$\boldsymbol{C}_2$	$\text{diag}(0.468, 0.780, 0.318, 0.099)\times10^{-2}$[Ns·m/rad]
$\boldsymbol{C}_3$	$\text{diag}(0.648, 0.780, 0.318, 0.099)\times10^{-2}$[Ns·m/rad]
$\boldsymbol{x}_d$	$(0.100, 0.500, 0.700)^{\text{T}}$ [m]
$\boldsymbol{R}_d$	$\begin{bmatrix} 0.88 & -0.32 & -0.34 \\ 0.34 & 0.94 & 0.00 \\ 0.32 & -0.12 & 0.94 \end{bmatrix}$

actual object frame at t_A and at t_B, where the error norm of the object attitude is represented by a rotational angle around the equivalent rotational axis. Namely, a physical meaning of the conditions is that the change of the difference between the actual object frame and the virtual object frame is smaller than the discretization error due to the low sampling rate. The actual object frame converges to the desired state finally if the conditions are satisfied in every servo-loop.

V. NUMERICAL SIMULATION

This section shows numerical simulation results to verify the proposed controller. A triple-fingered hand-arm model is utilized in the simulations (the detail of its dynamic model has been given in [10]). It consists of an arm part which has 5 DOFs and a triple-fingered hand part which has one 5 DOFs finger and two 4 DOFs fingers. A triangular prism is utilized for the grasped object as an example of polyhedral objects. Specific parameters of the hand-arm system and the grasped object are shown in Table I. In this Table, Y_i is a perpendicular distance from the center of the object mass $O_{c.m.}$ to each surface of the object, and θ_{ti} is the internal angle of a cross-sectional triangle of the object. Table II shows the desired nominal grasping force and gains. In order to focus on the visual servoing for object manipulation, assume that the hand-arm system has already grasped the object in the initial condition. The parameters of this system in initial state are shown in Table III.

TABLE III

INITIAL CONDITION OF THE SIMULATION

$\dot{\boldsymbol{q}}$	$\boldsymbol{0}$[rad/s]
$\boldsymbol{q}_a$	$(-0.183, -1.369, 1.898, 1.343, -0.787)^{\text{T}}$ [rad]
$\boldsymbol{q}_{01}$	$(1.007, 0.235, -0.771, 1.338, 0.328)^{\text{T}}$ [rad]
$\boldsymbol{q}_{02}$	$(0.242, -0.733, 1.122, 0.754)^{\text{T}}$ [rad]
$\boldsymbol{q}_{03}$	$(2.019, -0.924, 0.912, 1.088)^{\text{T}}$ [rad]
$\dot{\boldsymbol{x}}$	$\boldsymbol{0}$ [m/s]
$\boldsymbol{x}$	$(0.158, 0.501, 0.681)^{\text{T}}$ [m]
$\boldsymbol{\omega}$	$\boldsymbol{0}$ [rad/s]
$\boldsymbol{R}$	$\begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$

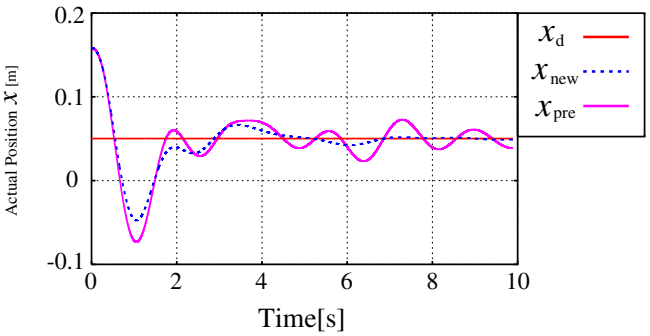

Fig. 3. Transient responses of the x-component of the position of the actual object frame $\boldsymbol{x}$: The sampling rate is $h = 20$ [Hz], the time-delay due to the computational cost for image processing, and data transmission latency is $t_{\text{image}} = 50$ [ms]

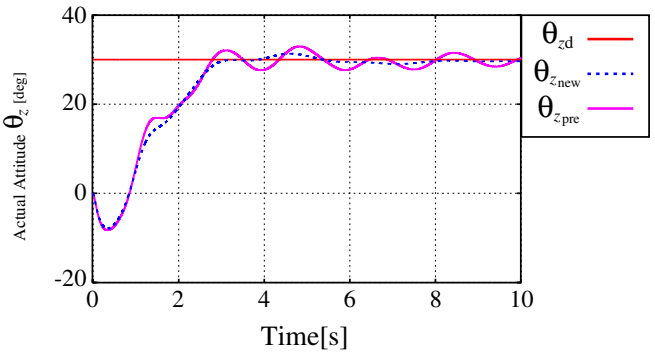

Fig. 4. Transient responses of the θ_z which means the actual rotational angle around the z-axis expressed by XYZ Euler angles: The sampling rate is $h = 20$ [Hz], the time-delay due to the computational cost for image processing, and data transmission latency is $t_{\text{image}} = 50$ [ms]

Two types of the simulations are conducted to show the advantage of the newly proposed method. One is using the newly proposed visual servoing method. The other is using our previous visual servoing method which utilizes the position and attitude of the actual object frame as a controlled variable directly [11]. The previous method uses control inputs for position control $\boldsymbol{u}_{p_{\text{real}}}$ and attitude control $\boldsymbol{u}_{o_{\text{real}}}$ of the actual object frame instead of $\boldsymbol{u}_p$ and $\boldsymbol{u}_o$. They

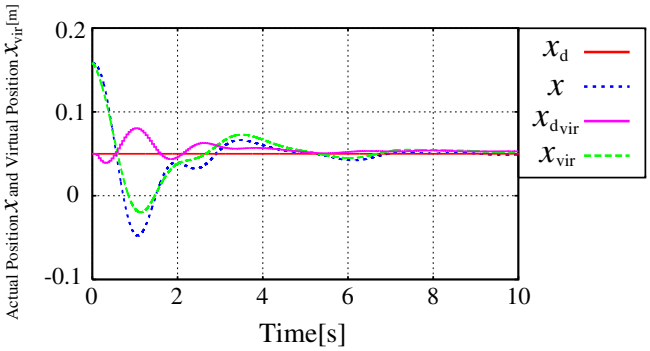

Fig. 5. Transient responses of the x-component of the position of the actual object frame $\boldsymbol{x}$, the position of the virtual object frame $\boldsymbol{x}_{\rm vir}$, and the position of the desired virtual object frame $\boldsymbol{x}_{d_{\rm vir}}$ when the control inputs $\boldsymbol{u}_p$ and $\boldsymbol{u}_o$ are used: The sampling rate is $h = 20$ [Hz], the time-delay due to the computational cost for image processing, and data transmission latency is $t_{\rm image} = 50$ [ms]

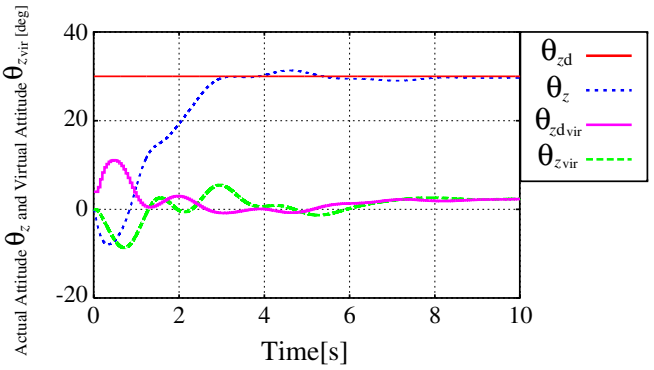

Fig. 6. Transient responses of the θ_z of the attitude of the actual object frame $\boldsymbol{R}$, and $\theta_{z_{\rm vir}}$ of the attitude of the virtual object frame $\boldsymbol{R}_{\rm vir}$ expressed by XYZ Euler angles when the control inputs $\boldsymbol{u}_p$ and $\boldsymbol{u}_o$ are used: The sampling rate is $h = 20$ [Hz], the time-delay due to the computational cost for image processing, and data transmission latency is $t_{\rm image} = 50$ [ms]

are given as follows:

$$\boldsymbol{u}_{p_{\rm real}}(t) = K_p \sum_{i=1}^{N} \boldsymbol{J}_{0i}(t)^{\rm T} \left(\boldsymbol{x}_d - \boldsymbol{x}(t - t_{\rm delay})\right) \tag{19}$$

$$\boldsymbol{u}_{o_{\rm real}}(t) = K_o \sum_{i=1}^{N} \boldsymbol{J}_{\Omega i}(t)^{\rm T} \left\{(\boldsymbol{r}_x(t - t_{\rm delay}) \times \boldsymbol{r}_{xd}) + (\boldsymbol{r}_y(t - t_{\rm delay}) \times \boldsymbol{r}_{yd}) + (\boldsymbol{r}_z(t - t_{\rm delay}) \times \boldsymbol{r}_{zd}) \right\}, \tag{20}$$

where the attitude of the actual object frame is expressed as a rotational matrix $\boldsymbol{R}(t) = [\boldsymbol{r}_x(t), \boldsymbol{r}_y(t), \boldsymbol{r}_z(t)] \in {\rm SO}(3)$. The sampling rate is $h = 20$ [Hz] and the time-delay due to the computational cost for image processing, and data transmission latency is $t_{\rm image} = 50$ [ms], respectively.

Figures 3 and 4 show the transient responses of the x-component of the actual $\boldsymbol{x}$, and θ_z which means the actual rotational angle around the z-axis expressed by XYZ Euler angles respectively. In these figures, $x_{\rm new}$ and $\theta_{z_{\rm new}}$ indicate the transient responses of x and θ_z in the case that the newly proposed control inputs $\boldsymbol{u}_p(t)$ and $\boldsymbol{u}_o(t)$ are utilized, and $x_{\rm pre}$ and $\theta_{z_{\rm pre}}$ indicate the transient responses of x and θ_z in the case that the previously proposed control inputs $\boldsymbol{u}_{p_{\rm real}}(t)$ and $\boldsymbol{u}_{o_{\rm real}}(t)$ are utilized. We see from these figures that in the case of using the newly proposed method, an

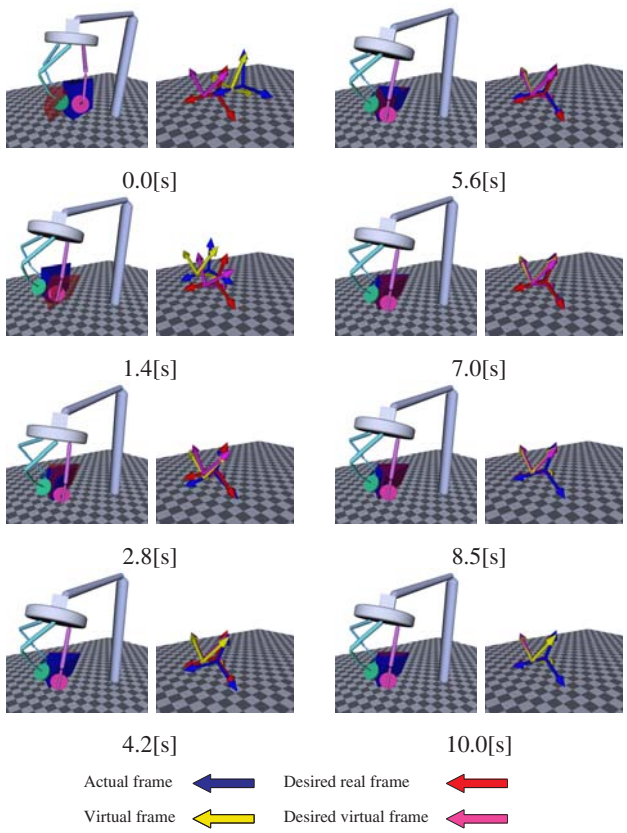

Fig. 7. Snapshots of the simulation in the case that the sampling rate $h = 20$ [Hz] and the time-delay $t_{image} = 50$ [ms]

oscillation is reduced and convergent rate is faster compared with the case of using the previous method, even with the low sampling rate and large time-delays.

Figures 5 and 6 show the behavior of the position and attitude of the actual and virtual object frame when the control inputs $\boldsymbol{u}_p(t)$ and $\boldsymbol{u}_o(t)$ are used in the case that the sampling rate $h = 20$ [Hz] and the time-delay due to the computational cost for image processing, and data transmission latency is $t_{\rm image} = 50$ [ms]. Figure 5 shows the x-component of the positions of the actual object frame $\boldsymbol{x}$, the virtual object frame $\boldsymbol{x}_{\rm vir}$, and the desired virtual object frame $\boldsymbol{x}_{d_{\rm vir}}$. Figure 6 shows the θ_z of the attitudes of the actual object frame $\boldsymbol{R}$, $\theta_{z_{\rm vir}}$ of the attitude of the virtual object frame $\boldsymbol{R}_{\rm vir}$, and $\theta_{zd_{\rm vir}}$ of the attitude of the desired virtual object frame $\boldsymbol{R}_{d_{\rm vir}}$ expressed by XYZ Euler angles. We see from these figures that $\boldsymbol{x}_{d_{\rm vir}}$ and $\boldsymbol{R}_{d_{\rm vir}}$ are actually updated with a step-like pattern in every interval of visual image acquisition. Figure 7 shows the snapshots of the simulation. In this figure, both the actual and virtual frames approach and finally converge to each desired value simultaneously.

Additionally, it is confirmed that the conditions (14) and (15) to guarantee the convergence of the desired state are satisfied in all the simulations.

VI. Experiments

Experiments of object manipulation using a prototype are performed to demonstrate the practical usefulness of the

TABLE IV

PHYSICAL PARAMETERS

Three-fingered robotic hand	
1st link length l_{i1}	0.064[m]
2nd link length l_{i2}	0.064[m]
3rd link length l_{i3}	0.030[m]
1st mass center l_{gi1}	0.023[m]
2nd mass center l_{gi2}	0.035[m]
3rd mass center l_{gi3}	0.010[m]
1st mass m_{i1}	0.038[kg]
2nd mass m_{i2}	0.024[kg]
3rd mass m_{i3}	0.054[kg]
(Fingertip)	
Radius r_i	0.015[m]
Physical properties s_i	2.390×10^6 [N/m^2]

TABLE V

DETAILS OF THE GRASPED OBJECT

Triangular prism	
Mass m	0.0015[kg]
Material	Styrene foam
(Figure)	
Length of side of triangle	0.060[m]
Height	0.039[m]

TABLE VI

NOMINAL DESIRED GRASPING FORCE AND GAINS

f_d	1.1
K_p	70.0
K_o	4.0×10^{-2}
$\boldsymbol{C}_1$	$\mathrm{diag}(0.07, 0.07, 0.05, 0.03)\times 10^{-2}$[Ns·m/rad]
$\boldsymbol{C}_2$	$\mathrm{diag}(0.07, 0.07, 0.05, 0.03)\times 10^{-2}$[Ns·m/rad]
$\boldsymbol{C}_3$	$\mathrm{diag}(0.07, 0.07, 0.05, 0.03)\times 10^{-2}$[Ns·m/rad]

proposed method. Each parameter of the system is shown in Table IV. The grasped object is a triangular prism, and it is made of styrene form. Therefore, its weight is light enough to ignore the effect of gravity. Each parameter of the grasped object is shown in Table V. A stereo optical tracking system (Micron Tracker H3-60, Claron, Inc.) is utilized to measure the position and attitude of the grasped object. The system can measure the position and attitude of the grasped object in every 50 [ms] by tracking a specific marker on the object. The overall system which includes a prototype robotic hand consists of three 4 DOFs fingers and the visual sensing system is shown in Fig. 8. This visual sensing system involves two types of the time-delays, t_{sample} and t_{image} as is mentioned in section IV-A. The sampling rate of the system $h = 20$ [Hz] is not enough to use a real-time feedback manner directly. An average of the time-delay due to image processing and latency of data transmission is $t_{\mathrm{image}} \approx 50$ [ms]. The time-delay due to image processing is observed in real-time by the measurement of the time from capturing an image to the beginning of data transmission. The other time-delay due to data transmission from the vision sensing system to the controller is measured in advance.

Two types of experiments are performed similar to the numerical simulations. In these experiments, the robotic hand has already grasped the object initially. One is using the proposed method and the other is using the previous method. Figure 9 shows the transient responses of the x-component of the position of the actual object frame $\boldsymbol{x}$. In this figure,

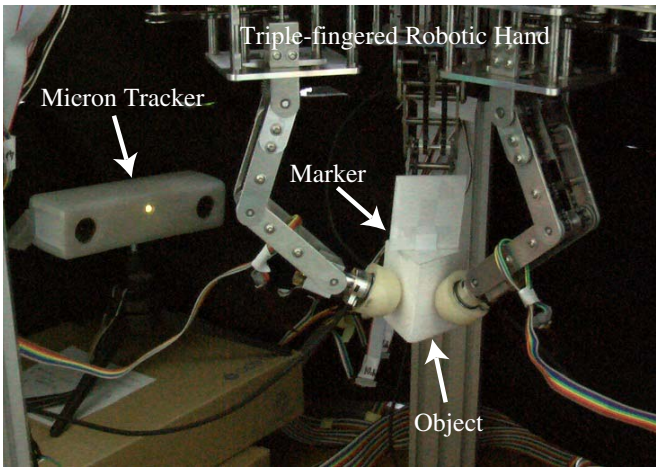

Fig. 8. Object manipulation using the three-fingered robotic hand and the visual sensing system

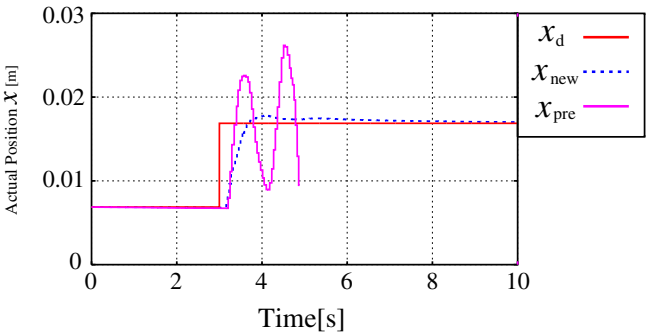

Fig. 9. Transient responses of the x-component of the position of the actual object frame $\boldsymbol{x}$

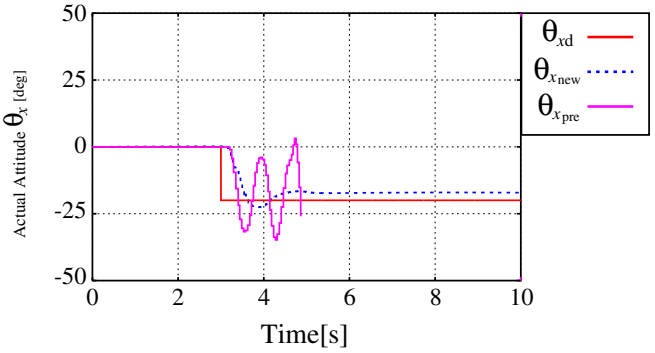

Fig. 10. Transient responses of θ_x of the attitude of the actual object frame $\boldsymbol{R}$ expressed by XYZ Euler angles

x_{new} indicates $x(t - t_{\mathrm{delay}})$ in the case that the new control inputs $\boldsymbol{u}_p(t)$ and $\boldsymbol{u}_o(t)$ are utilized, and x_{pre} indicates the transient responses of $x(t - t_{\mathrm{delay}})$ in the case that the previous control inputs $\boldsymbol{u}_{p_{\mathrm{real}}}(t)$ and $\boldsymbol{u}_{o_{\mathrm{real}}}(t)$ are utilized. Figure 10 shows the transient responses of θ_x which means the rotational angle around the x-axis expressed by XYZ Euler angles. In this figure, $\theta_{x_{\mathrm{new}}}$ indicates $\theta_x(t - t_{\mathrm{delay}})$ in the case that the newly proposed control inputs $\boldsymbol{u}_p(t)$ and $\boldsymbol{u}_o(t)$ are utilized, and $\theta_{x_{\mathrm{pre}}}$ indicates $\theta_x(t - t_{\mathrm{delay}})$ in the case that the previous control inputs $\boldsymbol{u}_{p_{\mathrm{real}}}(t)$ and $\boldsymbol{u}_{o_{\mathrm{real}}}(t)$

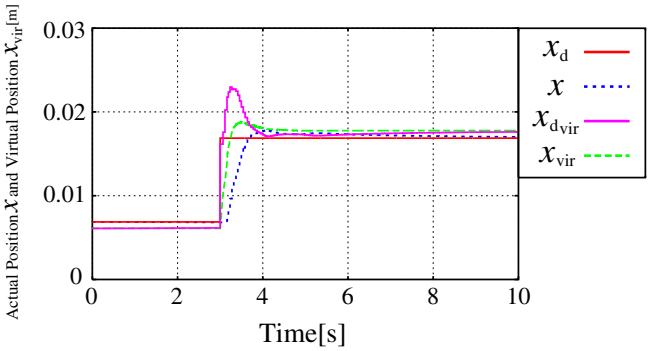

Fig. 11. Transient responses of the x-component of the positions of the actual object frame $\boldsymbol{x}$ and the virtual object frame $\boldsymbol{x}_{\rm vir}$

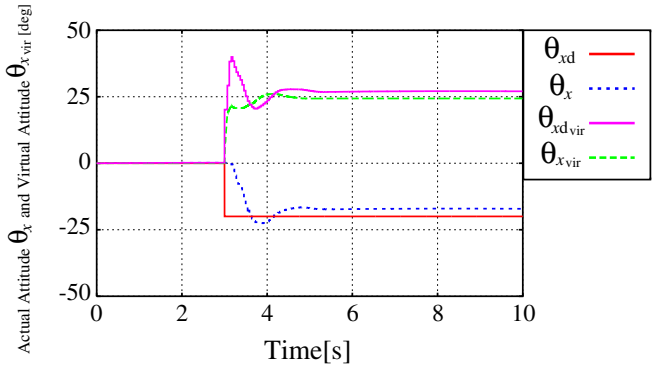

Fig. 12. Transient responses of θ_x of the attitude of the actual object frame $\boldsymbol{R}$ and $\theta_{x_{\rm vir}}$ of the attitude of the virtual object frame $\boldsymbol{R}_{\rm vir}$ expressed by XYZ Euler angles

are utilized. In these figures, the position and the attitude of the actual object frame in the case of using the new control inputs converge to each desired value. However, behavior of the overall system becomes unstable and the object grasping is failed finally in the case of using the previous method. Thus, it is confirmed that the new visual servoing method accomplishes stable object grasping and manipulation even if the sampling rate is low and there exists considerable time-delays, although the previous method cannot realize it.

Figure 11 shows the transient responses of the x-component of the position of the actual object frame, the virtual object frame and the desired virtual object frame when the control inputs $\boldsymbol{u}_p(t)$ and $\boldsymbol{u}_o(t)$ are used. Also Fig. 12 shows the transient responses of θ_x which means the actual rotational angle around the x-axis expressed by XYZ Euler angles. We see from these figures that the desired position and attitude of the virtual object frame are actually updated with a step-like pattern in every interval of visual image acquisition. Finally, both the actual and virtual object frames mostly converge to each desired value respectively with no unstable behavior.

Thus, it is confirmed that the new visual servoing method is practically useful through these experimental results.

VII. Conclusion

This paper presented a novel visual servoing method for object manipulation that is robust not only to temporary loss, but also to considerable time-delays of visual information. Firstly, the new control scheme of visual servoing for object manipulation was proposed to avoid unstable behavior caused by two types of time-delays. In this scheme, the virtual object frame is complementarily utilized as a nominal controlled variable in the interval of visual image acquisition. Next, the effectiveness of the method was verified by the numerical simulations. Finally, its practical usefulness was demonstrated through the experimental results.

In our future work, the stability of the overall system must be analyzed deeply. In addition to the detailed stability analysis, necessary conditions to realize the robustness against time-delays must be revealed in more detail.

Acknowledgment

This work was partially supported by Grant-in-Aid for JSPS Fellows.

References

[1] L. E. Weiss, A. C. Sanderson and C. P. Neuman, "Dynamic sensor-based control of robots with visual feedback," IEEE Trans. on Robot. and Automation, Vol. 3, No. 5, pp. 404–417, 1987.

[2] J. T. Feddema and O. R. Mitchell, "Vision-guided servoing with feature-based trajectory generation," IEEE Trans. on Robot. and Automation, Vol. 5, No. 5, pp. 691–700, 1989.

[3] S. Hutchinson, G. D. Hager and P. I. Corke, "A tutorial on visual servo control," IEEE Trans. on Robot. and Automation, Vol. 12, No. 5, pp. 651–670, 1996.

[4] N. R. Gans and S. A. Hutchinson, "Stable visual servoing through hybrid switched-system control," IEEE Trans. on Robotbotics, Vol. 23, No. 3, pp. 530–540, 2007.

[5] P. I. Corke and M. C. Good, "Dynamics effects in visual closed-loop systems," *IEEE Trans. on Robot. and Automation,* Vol. 12, No. 5, pp. 671–683, 1996.

[6] A. J. Koivo and N. Houshangi, "Real-time vision feedback for servoing robotic manipulator with self-tuning controller," *IEEE Trans. Systems, Man, and Cybernetics,* Vol. 21, No. 1, pp. 134–142, 1991.

[7] K. Hashimoto and H. Kimura, "Visual servoing with nonlinear observer," *Proc. IEEE Int. Conf. Robot. Automat.*, pp. 484–489, Nagoya, Japan, 1995.

[8] H. Fujimoto and Y. Hori, "Visual servoing based on intersample disturbance rejection by multirate sampling control - time delay compensation and experiment verification -," *Proc. IEEE Conf. Decision Contr.*, pp. 334–339, Orlando, Florida, 2001.

[9] K. Sasajima and H. Fujimoto, "6 dof multirate visual servoing for quick moving objects," *Proc. of the 2007 American Contr. Conf.*, pp. 1538–1543, New York, USA, 2007.

[10] A. Kawamura, K. Tahara, R. Kurazume and T. Hasegawa, "Dynamic grasping for an arbitrary polyhedral object by a multi-fingered hand-arm system," *Proc. of the 2009 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2264–2270, St. Louis, Missouri, 2009.

[11] A. Kawamura, K. Tahara, R. Kurazume and T. Hasegawa, "Dynamic Object Manipulation using a Multi-Fingered Hand-Arm System: Enhancement of a Grasping Capability using Relative Attitude Constraints of Fingers," *Proc. Int. Conf. on Advanced Robotics,* pp. 8–14, Estonia, Tallinn, 2011.

[12] A. Kawamura, K. Tahara, R. Kurazume and T. Hasegawa, "Robust manipulation for temporary lack of sensory information by a multi-fingered hand-arm system," *Proc. of the 2011 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. – , San Francisco, CA, 2011 (to appear).

[13] T. Wimböck, C. Ott, and G. Hirzinger, "Passivity-based Object-Level Impedance Control for a Multifingered Hand," *Proc. of the 2006 IEEE/RSJ on Intelligent Robots and Systems*, pp. 4621–4627, Beijing, China, 2006.

[14] K. Tahara, S. Arimoto and M. Yoshida, "Dynamic object manipulation using a virtual frame by a triple soft-fingered robotic hand," *Proc. IEEE Int. Conf. Robot. Automat.*, pp. 4322–4327, Anchorage, AK, 2010.