High Speed Grasping Using Visual and Force Feedback

Akio Namiki, Yoshihiro Nakabo, Idaku Ishii, and Masatoshi Ishikawa

Department of Mathematical Engineering and Information Physics
University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

E-mail : namik@k2.t.u-tokyo.ac.jp
URL : http://www.k2.t.u-tokyo.ac.jp/~sfoc/

Abstract

In most conventional manipulation systems, changes in the environment cannot be observed in real time because the vision sensor is too slow. As a result the system is powerless under dynamic changes or sudden accidents. To solve this problem we have developed a grasping system using high-speed visual and force feedback. This is a multi-fingered hand-arm with a hierarchical parallel processing system and a high-speed vision system called SPE-256. The most important feature of the system is the ability to process sensory feedback at high speed, that is, in about 1ms. By using an algorithm with parallel sensory feedback in this system, grasping with high responsiveness and adaptiveness to dynamic changes in the environment is realized.

Key Words: *grasping, high-speed visual and force feedback, sensor fusion, hierarchical parallel architecture, multi-fingered hand-arm, active vision*

1 Introduction

A grasping motion is one of the most important processes for control of multi-fingered hands. To complete the grasping process multiple types of sensor information are needed such as visual information and force/haptic information, and grasping motion should be controlled by fusing this sensory information. Then the sensory information should be acquired at a rate as the cycle time of grasping control so that the grasping motion has responsiveness to dynamic changes in the environment.

In conventional research, a delay of visual feedback is a serious problem for realizing the above requirements. In most cases a CCD camera is used to acquire visual information, which requires over 33ms to acquire an image. As a result the acquired image is always behind the real world action and prediction is necessary to compensate for the delay. As research based on such an approach, Allen et al. realized grasping of a moving model train on stereo triangulation of optic-flow field [1], and Hong et al. realized grasping of a flying object using stereo vision [2]. In these researches it was assumed that motion of an object was known beforehand. For this reason there was a problem in that the system could not complete a task in an unknown environment.

On the other hand, vision chip systems have recently been developed for high-speed visual processing [3, 4]. Because they have a parallel processing architecture in which each photodetector is directly connected to a corresponding processing element, visual processing is realized at a high rate [5, 6]. The use of such high-speed vision in manipulation can solve the problem of the delay of visual feedback and realize high responsiveness and flexibility in an unknown environment.

Against this background our objective is to develop a grasping system in which high-speed visual feedback by a vision chip system is fused with force/haptic feedback and to realize responsive and flexible grasping in the real world environment.

2 Architecture for High Speed Grasping

In this section a system architecture suitable to fuse sensor information is discussed by analyzing grasping process in the real world. There are two important features to realize grasping in the real world, as follows:

(A) Flexibility under multiple conditions

A grasping system should have flexibility to complete various tasks under various conditions. The grasping process should be suitably changed according to the grasping condition, for example an object's position, an object's shape and an object's motion.

To implement this, a hierarchical parallel processing architecture with several types of sensor is valid. Multiple types of sensory-motor fusion processing coexist in one system based on it. As a result, flexibility in multiple grasping environments is realized.

(B) Responsiveness to dynamic changes

In the real world, the grasping environment changes dynamically and is possible that the object moves at high speed and sudden accidents happen during grasping.

To overcome this, grasping control based on high-speed sensory feedback is effective. High-speed sensory feedback means to return feedback of external sensor information at a rate higher than the rate of control. Because the system can recognize an external environment in real time, responsiveness to dynamic changes in the grasping environment is realized.

We adopt an architecture in which both flexibility and responsiveness are realized. This is a hierarchical parallel architecture in which each element consists of high-speed sensory feedback within 1ms as shown in Figure 1. Because each feedback process is completed within 1ms, adjustment to various conditions is realized at high speed.

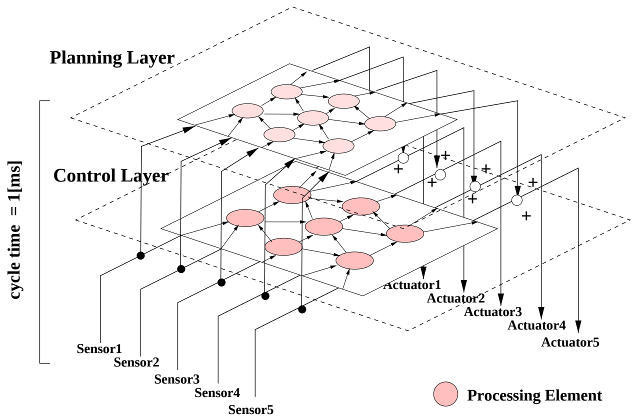

Figure 1. Hierarchical Parallel Architecture Based on High-speed Sensory Feedback

In general the cycle time of 1ms is necessary to prevent mechanical resonance in robotic control. In our architecture we decided that the cycle time of each sensory feedback should be 1ms to ensure stable grasping control.

As a related research Albus proposed a hierarchical parallel architecture based on the model of humans [7], and Brooks proposed a behavior-based hierarchical architecture consisting of layered sensory feedback modules [8]. We adopt a similar hierarchical parallel architecture, but responsiveness based on high-speed sensory feedback is not considered in these architectures.

3 1ms Sensory-Motor Fusion System

Using the idea of a hierarchical parallel architecture based on high-speed sensory feedback described in Section 2, we have developed a system called the "1ms Sensory-Motor Fusion System" to realize high-speed sensory feedback and fusion of sensory information. This system exhibits high performance processing of all sensory feedback, including visual feedback, with a cycle time of 1ms. Because the processing result is directly used to control the manipulator, each task is realized with high responsiveness.

Figure 2 shows the system components and Figure 3 shows a photograph of the system. This system consists of three parts: (1) a DSP subsystem for high-speed sensor fusion processing; (2) an active vision subsystem for high-speed visual processing; and (3) a multi-fingered hand-arm subsystem for high-speed manipulation.

3.1 DSP Subsystem

The DSP subsystem is the main part for fusion processing of sensory feedback within 1ms. It has a hierarchical parallel architecture consisting of 7 DSPs connected to each other, and many I/O ports are installed for inputing various types of information in parallel.

In this system we use a floating-point DSP TMS320C40 which has high performance (275 MOPS) and 6 I/O ports (20 Mbytes/sec). By connecting several C40 processors, a low bottle-neck hierarchical parallel architecture is realized. In our system the following I/O ports are prepared; ADC (12 bit, 64 CH), DAC (12 bit, 24 CH), and Digital I/O (8 bit, 8 ports). These I/O ports are distributed on several DSPs to minimize the I/O bottleneck so that sensor signals are input in parallel.

A parallel programming development environment has been prepared in which multi-process and multi-thread programming is easily realized. This function is useful to program parallel sensory feedback.

3.2 Active Vision with SPE

The active vision subsystem consists of a vision chip system called SPE-256 and a 2-axis actuator moved by DC servo motors. SPE-256 consists of a 16×16 array of processing elements (PE) and PIN photo-diodes (PD). The output of each PD is connected with a corresponding PE. Each PE is a 4-neighbor connected SIMD based processor which has a 24 bit register and a bit-serial arithmetic logic unit capable of AND, OR, and XOR operations etc. Because the visual processing is perfectly executed in parallel, high-speed visual feedback is realized within 1ms [5].

The SPE-256 is a scale-up model of an integrated vision chip and the next generation is currently being developed in which all elements of the vision chip architecture are integrated in one chip [4].

The actuator part of the active vision subsystem has two degrees of freedom; pan and tilt. This is used to move the sensor platform and this is controlled by a DSP assigned for active vision control.

3.3 7-axis Manipulator with Dextrous Multi-fingered Hand

The hand-arm subsystem is a 7-axis manipulator with a dextrous multi-fingered hand.

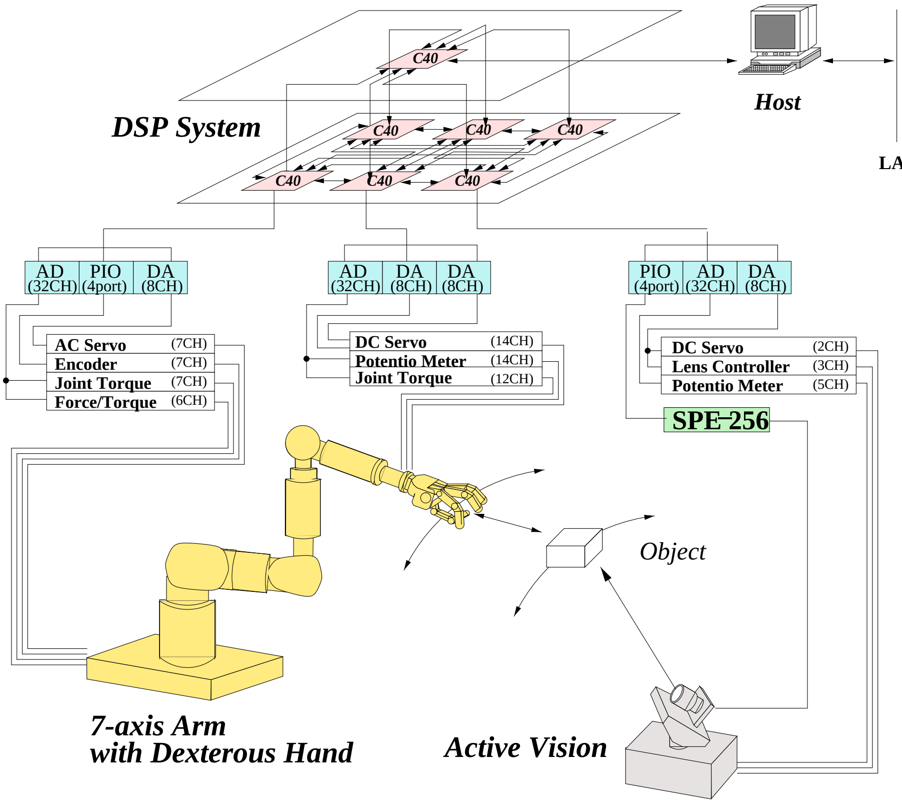

Figure 2. Architecture of 1ms Sensory-Motor Fusion System

The multi-fingered hand has 4 fingers and 14 joints. Its structure is similar to a human hand, in which a thumb finger is installed opposite to the other three fingers. Each joint is controlled by DC servo motors in a remote place using a control cable consisting of an outer casing and an inner wire. Each joint of the hand has a potentiometer for position control and a strain gage for force control.

The arm has 7 joints controlled by AC servo motors. An encoder is installed in each joint and a 6-axis force/torque sensor is installed at the wrist.

Two DSPs are assigned to control the hand and the arm.

4 Algorithm of High Speed Grasping

Using the idea of a hierarchical parallel architecture based on high-speed sensory feedback described in Section 2, we have developed an algorithm to realize responsiveness to dynamic changes of a grasping object.

Figure 4 shows a block diagram of the algorithm. In this algorithm 4 sensory feedbacks and one sensory information processing are executed in parallel; tracking control of the active vision, tracking control of the arm, reaching control of the arm, grasping control of the hand, and visual object recognition. These processes are distributed to 3 DSPs which are respectively assigned to the active vision, the arm, and the hand.

Figure 3. 1ms Sensory-Motor Fusion System

Though no prediction methods are used in this algorithm, our system has a high performance and sufficient responsiveness based on high-speed sensory feedback whose cycle time is less than 1.5ms.

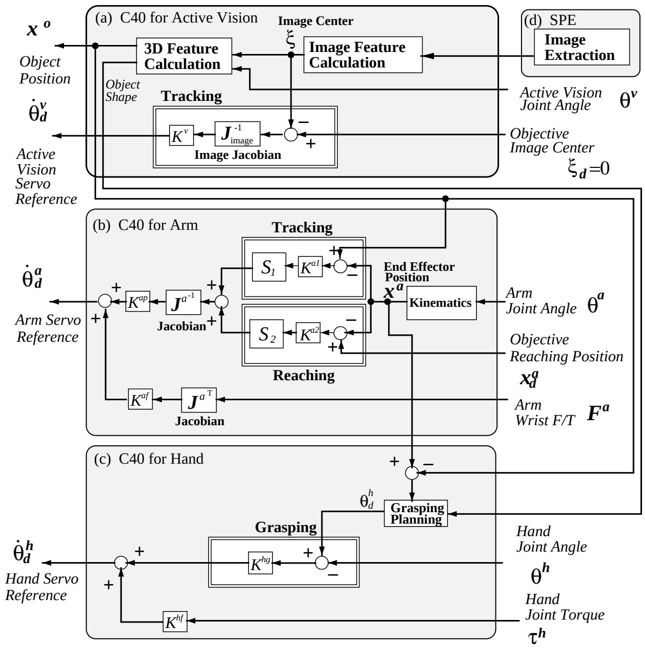

Figure 4. Algorithm of High Speed Grasping

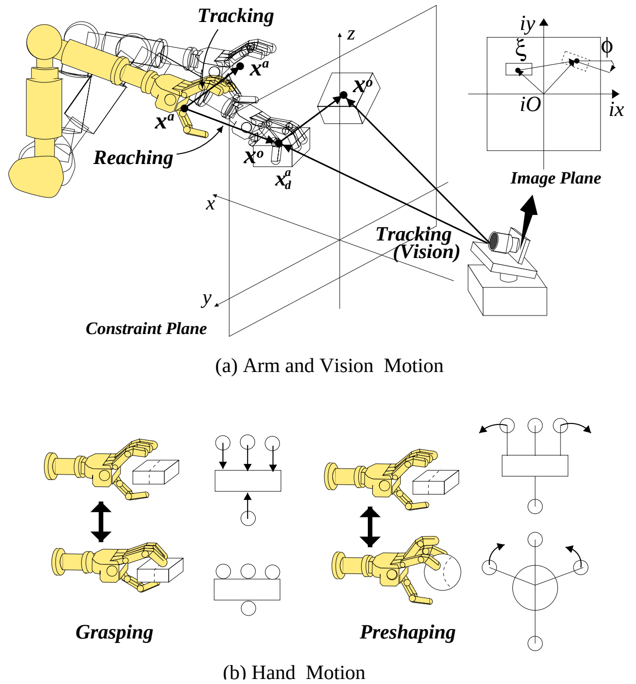

Figure 5. Motion of High Speed Grasping

4.1 Processing for Active Vision

In the DSP for active vision, two processes are executed in parallel: (1) object recognition by visual information, and (2) target tracking control.

4.1.1 Object Recognition by Visual Information

First, in SPE-256 a manipulated object is extracted in the image plane using an algorithm called Self Windowing which is realized by utilizing features of high-speed vision [5]. Next some image features are calculated from the extracted image in the DSP: the center of the image $\boldsymbol{\xi} \in \boldsymbol{R}^2$, and the angle of rotation of the image $\phi \in \boldsymbol{R}$.

Lastly, some 3-D geometrical parameters are calculated in the DSP: the 3-D position and the orientation (role, pitch, and yow angles) of the object: $\boldsymbol{x}^o \in \boldsymbol{R}^6$, and the object size. Then the object shape is detected based on these parameters. In the present configuration we assume that the motion of an object is limited to a constraint plane given beforehand. Using parameters of both the plane and image features, 3-D parameters are calculated.

4.1.2 Tracking Control of Active Vision

In this system the objective of active vision control is to acquire reliable object information by keeping an object in sight. To realize this we adopt feedback control using the image feature $\boldsymbol{\xi}$ written as,

$$\dot{\boldsymbol{\theta}}_d^v = K^v J_{\text{image}}^{-1}(\boldsymbol{\xi}_d - \boldsymbol{\xi}) - K^{vv}\dot{\boldsymbol{\theta}}^v, \qquad (1)$$

where $\dot{\boldsymbol{\theta}}_d^v \in \boldsymbol{R}^2$ is the control input to the active vision servo, $\boldsymbol{\xi}_d \in \boldsymbol{R}^2$ is the objective position on the image plane, $\boldsymbol{\theta}^v \in \boldsymbol{R}^2$ is the joint angle vector of the active vision, and $K^v \in \boldsymbol{R}^{2\times 2}$ and $K^{vv} \in \boldsymbol{R}^{2\times 2}$ are diagonal gain matrices. The matrix $J_{\text{image}} \equiv \dfrac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\theta}^v} \in \boldsymbol{R}^{2\times 2}$ is called the "image Jacobian".

4.2 Processing for Arm

On the DSP for arm control, two sensory feedback controls are executed: (1) tracking control to object motion, and (2) reaching control to the grasping position.

The objective of tracking control is cancelation of the object motion by maintaining the relative position between the hand and the object. In the present configuration it is realized as a servo control so that the relative position error is kept to zero on the constraint plane, as shown in Figure 5(a).

The objective of reaching control is to reach the hand to the object by control of the relative position. In the present configuration this is controlled in the direction orthogonal to the constraint plane, as shown in Figure 5(a). In this control the objective trajectory of the hand is given beforehand.

By integrating tracking motion and reaching motion, the arm control scheme can be written as,

$$\begin{aligned}\dot{\boldsymbol{\theta}}_d^a &= K^{ap} J^{a-1} S K^{a1}(\boldsymbol{x}^o - \boldsymbol{x}^a) \\ &\quad + K^{ap} J^{a-1}(I - S) K^{a2}(\boldsymbol{x}_d^a - \boldsymbol{x}^a) \\ &\quad - K^{av}\dot{\boldsymbol{\theta}}^a + K^{af} J^{aT} \boldsymbol{F}^a \end{aligned} \qquad (2)$$

where $\dot{\boldsymbol{\theta}}_d^a \in \boldsymbol{R}^6$ is the control input to the arm servo, $\boldsymbol{\theta}^a \in \boldsymbol{R}^6$ is the joint angle vector of the arm, and $J^a \in \boldsymbol{R}^{6\times6}$ is the jacobian matrix of the arm. The vector $\boldsymbol{x}^o \in \boldsymbol{R}^6$ is the position and the orientation of the object observed by vision, $\boldsymbol{x}^a \in \boldsymbol{R}^6$ is the position and the orientation of the hand obtained by haptic sensors, and $\boldsymbol{x}_d^a \in \boldsymbol{R}^6$ is the objective trajectory for reaching. The matrix K^{ap}, K^{a1}, K^{a2}, K^{av}, and $K^{af} \in \boldsymbol{R}^{6\times6}$ are diagonal gain matrices. The matrix $I \in \boldsymbol{R}^{6\times6}$ is the unit matrix and $S \equiv \mathrm{diag}\{s_i\}(i = x, y, z, \mathrm{role}, \mathrm{pitch}, \mathrm{yow})$ is a task partition matrix defined as,

$$s_i \equiv \begin{cases} 1 & \text{if } i = y, z, \mathrm{role} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In Eqn.(2) tracking motion and reaching motion respectively correspond to the first term and the second term. Because reaching motion is orthogonal to the tracking motion, there is no interaction. Then the fourth term is force feedback of the wrist force/torque sensor for compliance control.

4.3 Processing for Hand

In this system the objective of grasping control is to fix the object with the hand for manipulation. Using the compliance control method the hand is controlled as follows:

$$\dot{\boldsymbol{\theta}}_d^h = K^{hg}(\boldsymbol{\theta}_d^h - \boldsymbol{\theta}^h) - K^{hv}\dot{\boldsymbol{\theta}}^h + K^{hf}\boldsymbol{F}^h \tag{4}$$

where $\dot{\boldsymbol{\theta}}_d^h \in \boldsymbol{R}^{14}$ is the control input to the hand servo, and $\boldsymbol{\theta}^h \in \boldsymbol{R}^{14}$ is the joint angle vector of the hand. Matrices K^{hg}, K^{hv}, and $K^{hf} \in \boldsymbol{R}^{14\times14}$ are diagonal gain matrices, and $\boldsymbol{F}^h \in \boldsymbol{R}^{14}$ is the joint torque vector. The vector $\boldsymbol{\theta}_d^h \in \boldsymbol{R}^{14}$ is the objective trajectory for grasping and is planned according to reaching motion $\boldsymbol{x}_d^a$.

Furthermore, according to the object shape, preshaping motion is executed to set the appropriate hand shape for grasping. In the present configuration the grasping shape is changed by distinguishing a circle and a rectangle in the 2D image-plane, as shown in Figure 5(b).

5 Experimental Results

We have performed experiments of grasping an object on the 1ms Sensory-Motor Fusion System.

The experimental result is shown in Figure 6 as a continuous sequence of pictures. All sensory feedback is executed in parallel according to the object motion at high speed: tracking motion of the active vision, tracking and reaching motion of the arm, and grasping motion of the hand. In Figure 7 a close-up view of the same motion is shown. In this figure tracking is executed from 0.0ms to 0.5ms and both reaching and grasping motion start at 0.5ms and all motion is completed at 0.8ms. Then in Figure 8 a close-up view of the grasping motion of a spherical object is shown. It is shown that the shape of the hand is changed to a suitable shape for grasping of a sphere.

In Figure 9 the trajectory of the hand is shown when grasping and releasing are alternately executed. In this figure, the Y axis position of the hand and the object show the tracking motion, and the X axis position of the hand and objective trajectory for reaching motion show the reaching motion. This figure shows that both responsive tracking by visual feedback during the releasing phase and stable grasping by visual and force feedback during the grasping phase are realized.

Figure 6. Experimental Result: Grasping of a Hexahedron

In these experiments, because the object is moved by a human hand, its trajectory is irregular and difficult to predict. Using the speed of the sensory feedback this problem is solved.

6 Conclusion

In this paper we describe a grasping system using high-speed sensory feedback with visual and force feedback. The system consists of two parts.

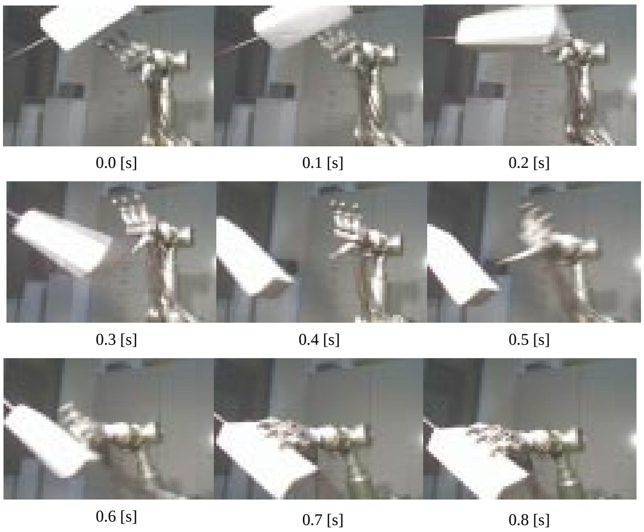

Figure 7. Experimental Result: Grasping of a Hexahedron

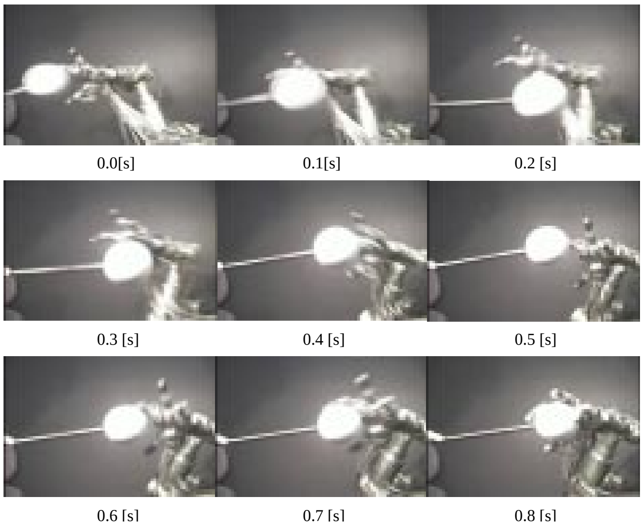

Figure 8. Experimental Result: Grasping of a Sphere

1. 1ms Sensory-Motor Fusion System has been developed to process and fuse sensory information at high speed. This consists of a hierarchical parallel processing subsystem with DSPs, a high-speed active vision subsystem and a manipulator with a dextrous multi-fingered hand. As a result all sensory feedback can be realized in about 1ms.
2. An algorithm for high-speed grasping is proposed. In this algorithm a grasping task is decomposed into some subtasks and each subtask is executed by high-speed sensory feedback in parallel.

As a result, grasping responsive to dynamic changes of object motion is realized.

Now we are developing various types of application on our system to realize responsive and flexible manipulation in the real world environment.

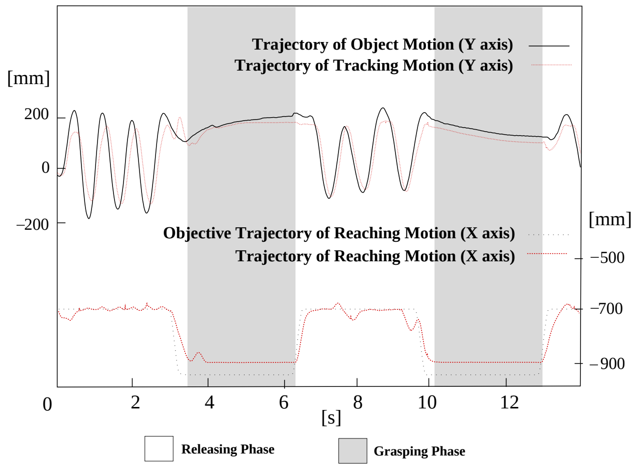

Figure 9. Feedback Response

References

[1] P.K. Allen, B. Yoshimi, and A. Timucenko. Real-time visual servoing. *Proc. IEEE Int. Conf. Robotics and Automation*, pages 2376–2384, 1991.

[2] W. Hong and J.E. Slotine. Experiments in hand-eye coordination using active vision. *Proc. 4th Int. Symp. on Experimental Robotics (ISER'95)*, 1995.

[3] M. Ishikawa, A. Morita, and N. Takayanagi. High speed vision system using massively parallel processing. *Proc. IEEE Int. Conf. on Intelligent Robotics and Systems*, pages 373–377, 1992.

[4] T. Komuro, I. Ishii, and M. Ishikawa. Vision chip architecture using general-purpose processing elements for 1ms vision system. *Proc. 4th IEEE Int. Workshop on Computer Architecture for Machine Perception (CAMP'97)*, pages 276–279, 1997.

[5] I. Ishii, Y. Nakabo, and M. Ishikawa. Target tracking algorithm for 1ms visual feedback system using massively parallel processing vision. *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 2309–2314, 1996.

[6] Y. Nakabo and M. Ishikawa. Visual impedance using 1ms visual feedback system. *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 2333–2338, 1998.

[7] J.S. Albus. Outline for a theory of intelligence. *IEEE Trans. on Systems, Man, and Cybernetics*, 21(3):473–509, 1991.

[8] R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2(1):14–23, 1986.