1ms Sensory-Motor Fusion System

Masatoshi Ishikawa, Takashi Komuro, Akio Namiki, and Idaku Ishii

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

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

Abstract

Recently, there has been growing interest in sensory-motor integration for new behavior of intelligent robots. And the key component to this work is sensory information processing technology which is based on recent progress in the integration of electronic circuits, providing increased computing power at low cost. In this talk , a new type of vision chip which has a general purpose parallel processing array with photo detector in a single silicon chip will be discussed. The vision chip achieves 1ms image processing, so mechanical systems can be controlled by using visual information with a 1ms sampling rate. Additionally a 1ms sensory-motor fusion system, a new type of a hierarchical parallel sensory processing system, will be discussed. The system consists of integrated sensor modules and a parallel processing system providing for sensory feedback and novel performance. A demonstration of high speed grasping using visual and force feedback will be described.

1. Introduction

There is a growing interest in sensory-motor integration for achieving new behaviors in intelligent robots. The key to achieving high-level functions is sensory information processing technology. With recent progress in the integration of electronic circuits, great changes will occur in the role and function of the sensor. The most important point to be noted is that with the progress of this integration the communication cost exceeds the computation cost. In other words, the sensor is no longer considered simply as a signal- transformation device transfering a physical value to an electrical value, as in conventional sensors, but rather as an information processing module including processing inherent in the sensor.

In this new design, there must be some processing architectures integrated with the detection function. When viewed as a whole system, the parallel processing architecture is configured so that the processing is distributed among the sensors. As a result of this, a new hierarchical parallel distributed processing must be introduced that corresponds to such a processing architecture. This paper considers primarily the processing architecture for sensory information in robotics from a new perspective integrating massive parallel processing vision, high speed vision, active vision, and sensor fusion. In addition, some applications are presented, and the direction of future sensor technology is discussed.

2. Vision chip

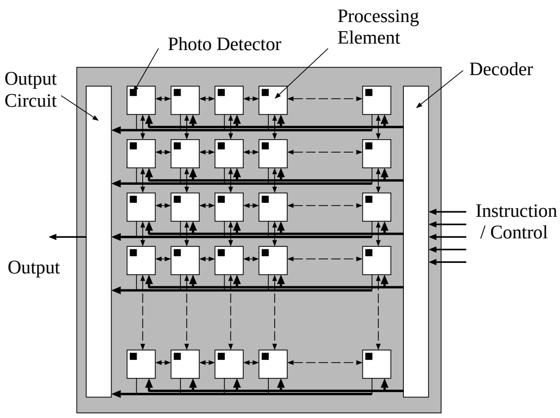
For real-time machine vision such as robot control using high speed visual feedback, traditional vision systems have an I/O bottleneck problem due to scanning and transmitting a large amount of image data. Further, the sampling rate is limited to video rates (NTSC 30 Hz / PAL 25 Hz).

To solve this problem, we have developed a new vision chip architecture called S^3PE (Simple and Smart Sensory Processing Elements) [1]. In the vision chip architecture, photo detectors (PDs) and parallel processing elements (PEs) are integrated in a single chip without the I/O bottleneck, and the parallel PEs have general-purpose processing capability and are controlled by programs using digital circuits for real-time machine vision of robot control.

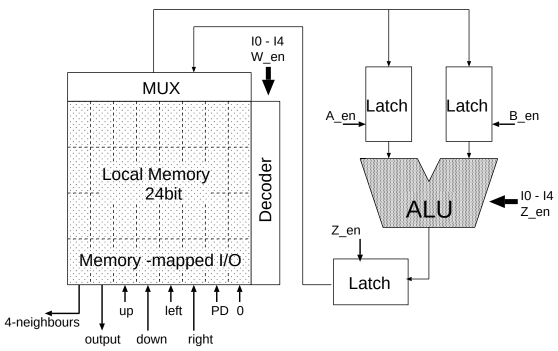
The block diagram of the whole chip is shown in Figure 1(a). General-purpose PEs are arranged in a massive parallel 2D array. Each PE is directly connected to a PD, an output circuit, and its four neighboring PEs. Image signals from the PDs are A/D converted and transmitted in parallel to all the PEs. Instruction codes are decoded, transmitted to all the PEs, and executed simultaneously (SIMD type processing). The calculated result is transmitted to the output circuit and feature values such as moments are extracted and transmitted to an external system.

The block diagram of the PE is shown in Figure 1(b). Each PE has an ALU, a local memory, and three registers. The ALU consists of a full adder, four multiplexers and a D-flipflop for holding a carry bit, and can execute 10 logical and 8 arithmetic binary operations. Multi-bit operations are implemented by repeating single operations serially (bit serial operation). The local memory has a 5-bit address space and consists of a 24-bit RAM and an 8-bit memory-mapped I/O which is connected to a PD, the output circuit, and four-neighboring PEs. Each bit can be randomly accessed.

The instruction set of the S^3PE is very simple. In the S^3PE all operations including I/O and calculations are done by accessing the local memory. Each instruction consists of a 5-bit operation code, two 5-bit read addresses, and a 5-bit write address. Wiring is reduced by sharing the bus and by reducing the area of the circuit. The 20-bit instruction codes are time-multiplexed into four parts and transmitted to the PEs through five hardware lines.



(a) the whole chip



(b) PE

Figure 1: Block diagram of vision chip architecture S^3PE

Table 1: Number of steps and time of sample programs on S^3PE

algorithm	steps(time[1])
4-neighbor edge detection (binary)	11 (0.72 μs)
4-neighbor smoothing (binary)	14 (1.0 μs)
4-neighbor edge detection (8bit)	70 (5.6 μs)
4-neighbor smoothing (8bit)	96 (7.7 μs)
4-neighbor thinning (binary) [2]	23 (1.9 μs)
Convolution (3×3, binary input)	40 (3.2 μs)
Convolution (3×3, 4-bit input)	372 (30 μs)
Poisson equation (4-neighbor, 4-bit) [3]	63 (5.0 μs)

[1] Calculated regarding an instruction cycle of 80 ns
[2] The process is repeated 10 times
[3] The process is repeated 200 times

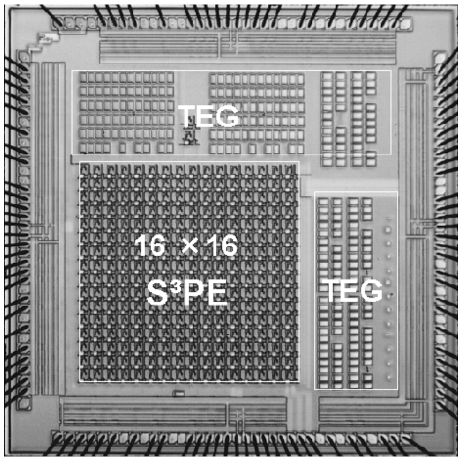


Figure 2: Photograph of the test chip

In the vision chip, the main operation of the PEs is 2D pattern processing. In other words, 2D to 2D pattern transformations are performed in the PEs. Therefore, the total amount of data is still large. If the 2D pattern data were output directly to external pins, we would face the I/O bottleneck problem again. To avoid this problem, we introduced an output circuit which extracts feature values such as moments. To integrate the circuit together with the PEs requires a compact and homogeneous circuit design using digital circuits.

As shown above, the vision chip with the S^3PE architecture has general-purpose processing capabilities and can implement various algorithms. We developed some sample programs for the S^3PE and simulated them using a vision chip simulator we developed. The sample programs and the results of the simulations are shown in Table 1. Assuming an instruction cycle of 80 ns, all of these programs are executed in much less than 1 ms, which is enough for robot control.

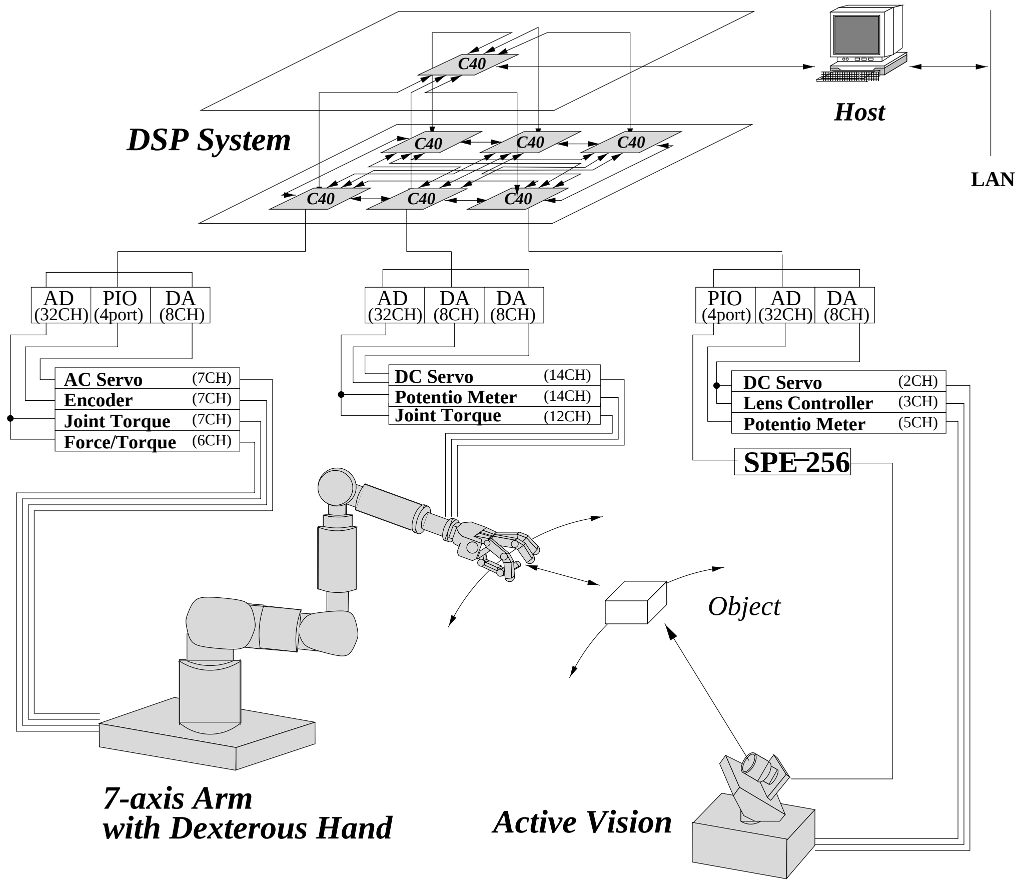


Figure 3: 1ms sensory-motor fusion system

The test chip fabricated using 0.8μm CMOS process has 8$\times$8 pixels and each pixel are is 240$\times$240 μm^2 [2]. We have also developed a test chip using a 0.35μm CMOS process. It has 16$\times$16 pixels and each pixel area is 150$\times$150 μm^2, which tells that 64$\times$64 pixels can be integrated in a single chip. Figure 2 shows a photograph of the chip.

We have already realized many applications such as target tracking, human interface using high speed vision system using vision chip architecture[3,4,5,6].

3. 1ms sensory-motor fusion system

A 1ms sensory-motor fusion system is a robot system for multi-sensor information to obtain real-time robot control in the real world. The system architecture is designed to realize high speed sensory feedback and multi-sensor fusion[7].

High speed sensory feedback with sensing and recognition is necessary for good responses to the real world, assures the stability and safety of a system, and derives the highest responsiveness of a robot mechanism. Therefore the rate of sensory feedback is one of the most important specifications for the system in order to obtain sensor information, process it, and control actuators at high speed.

Multi-sensor fusion is a technique in which information from various kinds of sensors is used to extract meaningful integrated information that cannot be obtained from a single sensor. Multi-sensor fusion technology is effective in obtaining reliable and accurate sensor feedback information for robot control.

Based on these concepts we have developed the 1ms sensory-motor fusion system with a hierarchical parallel processing system for sensory information in which the cycle time of all sensory feedback iterations can be less than 1 ms. This is known as the minimum cycle time to stably control mechanical systems. The 1ms sensory-motor fusion system is shown in Figure 3. The whole system has three subsystems; (1) a hierarchical parallel processing system, (2) a manipulator system with a 7-axis arm and a 4-fingered dexterous hand, and (3) a high speed active vision system.

The parallel processing system consists of 7 DSPs (TMS320C40) which are connected to each other. The DSPs have a processing power of 275 MOPS, communication speed of 120 Mbytes/s, and have a architecture suitable for real-time parallel processing. The system has many I/O ports which are connected using about 80 channels to actuators and sensors via many A/D, D/A and parallel I/O interfaces.

The manipulator system has a 7 axis robot arm with encoders for joint angle. At the end of the manipulator a dexterous 4-fingered hand is fitted with a 6-axis force/torque wrist sensor for compliant motion. The 4-fingered hand has 14 joints in total and a structure similar to a human hand. Each joint of the hand is controlled by a tendon-driven servo-motor, with a potentiometer and three strain gages per finger for grasping with compliance using the force information.

The high speed active vision system consists of a 2-axis actuator and the SPE-256 which can achieve visual processing in less than 1 ms[3].

4. High speed grasping using visual and force feedback

We have choosen to use grasping as an application of the 1ms sensory-motor fusion system[8]. Our objective is to realize high responsiveness to dynamical motion of a manipulated object by high speed visual feedback and force feedback with contact.

Figure 4 shows the block diagram of the grasping algorithm and Figure 5 shows a system configuration for high speed grasping. The manipulator with the dextrous hand and the active vision system are located face-to-face. The manipulated object moves between the manipulator and the active vision system, and the hand catches it by observing its position. Here we use two dimensional image features for the Y-Z plane as visual feedback information.

Four feedback loops are executed in parallel to realize high performance processing in the high speed grasping system.

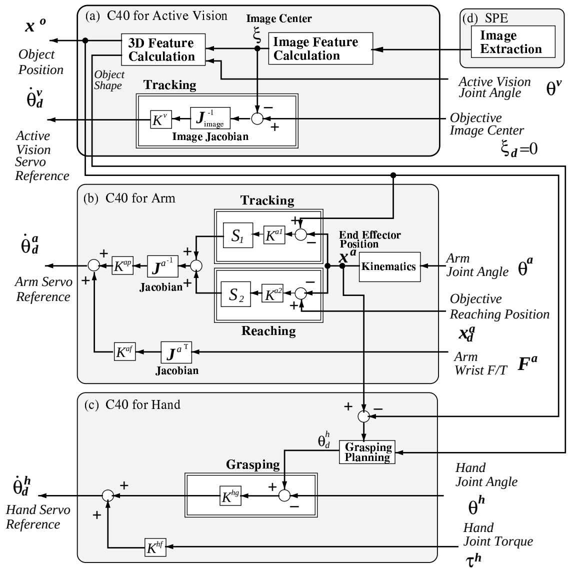


Figure 4: Block diagram of the grasping algorithm

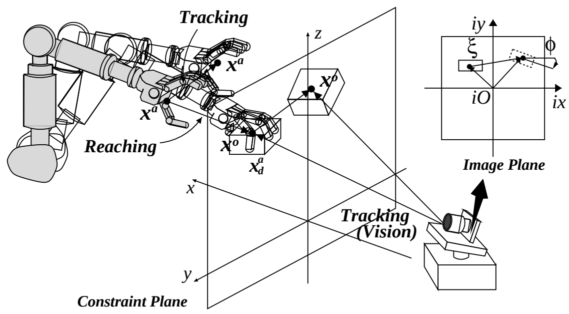
(a) Tracking (Active Vision): Tracking is done to acquire reliable object information. The active vision system is controlled so that the center of the observed object is always kept in the center of the image plane.

(b) Tracking (Arm): By canceling the object motion, tracking of the arm is done to keep the arm in a position suitable for grasping. In the algorithm, the relative position errors and the relative orientation error between the hand and the object, observed by active vision, are maintained at zero on the Y-Z plane.

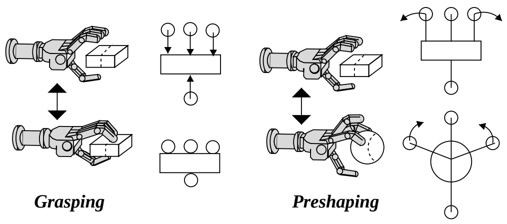
(c) Reaching (Arm): Reaching of the arm is done to control the relative position between the hand and the object. In the algorithm, the arm moves from the initial position to the grasping position along the X axis. The initial position and the trajectory along the X axis can be given beforehand. This motion is orthogonal to the tracking motion of the arm.

(d) Grasping (Hand): Grasping of the hand is performed according to the relative distance between the object and the end-effector. Force sensor compliance control is used to achieve stable grasping at each joint. The hand shape can be suitably adjusted for grasping according to the object shape obtained by visual information.

These four feedback controls are executed in parallel. Each cycle time of the feedback loops is less than 1.5ms, and adequate responsiveness to



(a) Motion of the arm and the active vision



(b) Motion of the hand

Figure 5: System configuration for high speed grasping

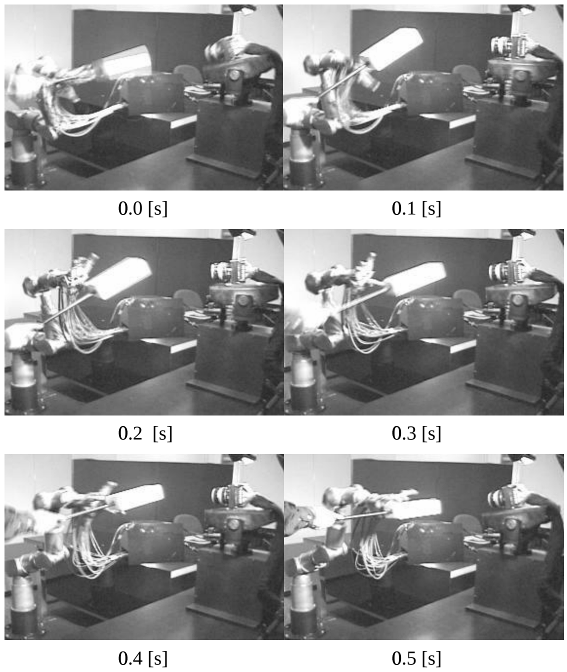


Figure 6: Experimental result: grasping of a hexahedron

the real world is achieved without using prediction.

The experimental result is shown in Figure 6 as a continuous sequence of pictures[8,9,10]. All sensory feedback is executed in parallel at high speed: tracking motion of the active vision, tracking and reaching motion of the arm, and grasping motion of the hand. Figure 7 is a close-up view of the same motion. In this figure tracking is executed from 0.0s to 0.5s and both reaching and grasping motion start at 0.5s and all motion is completed at 0.8s. Figure 8 shows a close-up view of the grasping motion of a spherical object. Notice that the shape of the hand is changed to a suitable shape for grasping 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 X axis objective trajectory 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 achieved.

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

5. Conclusion

This paper is based on the proposition that parallel processing and high speed sensory information processing should be introduced into sensor feedback systems. A special type of architecture is discussed in the context of some applications.

References

[1] Takashi Komuro, Idaku Ishii, and Masatoshi Ishikawa : "Vision Chip Architecture Using General-Purpose Processing Elements for 1ms Vision System," Proc. IEEE Int. Workshop on Computer Architecture for Machine Perception, pp.276-279, 1997.

[2] M. Ishikawa, K. Ogawa, T. Komuro, and I. Ishii: A CMOS Vision Chip with SIMD Processing Element Array for 1ms Image Processing, 1999 Dig. Tech. Papers of 1999 IEEE Int. Solid-State Circuits Conf. (ISSCC'99) /Abst. pp.206–207.

[3] Yoshihiro Nakabo, Idaku Ishii, and Masatoshi Ishikawa : "High Speed Target Tracking Using 1ms visual Feedback System," Abst. Video Proc. IEEE Int. Conf. Robotics and Automation, p.6, 1996.

[4] Idaku Ishii, Yoshihiro Nakabo, and Masatoshi Ishikawa : "Target Tracking Algorithm for 1ms

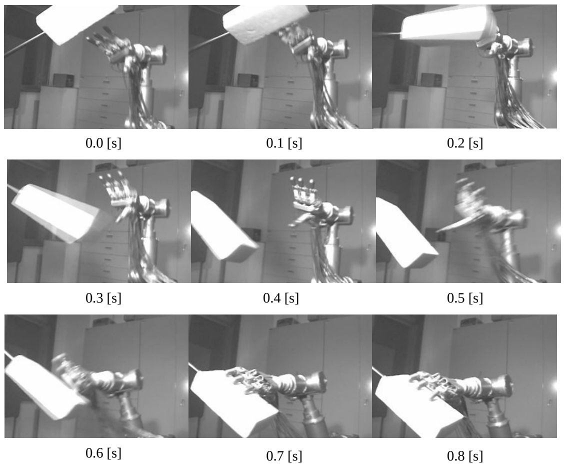


Figure 7: Experimental result: grasping of a hexahedron (close view)

Visual Feedback System Using Massively Parallel Processing," Proc. IEEE Int. Conf. Robotics and Automation , pp.2309-2314, 1996.

[5] Yoshihiro Nakabo and Masatoshi Ishikawa : "Visual Impedance Using 1ms Visual Feedback System," Proc. IEEE Int. Conf. Robotics and Automation pp.2333-2338, 1998.

[6] Takashi Owaki, Yoshihiro Nakabo, Akio Namiki, Idaku Ishii, and Masatoshi Ishikawa : "Real-time System for Virtually Touching Objects in the Real World Using a High Speed Active Vision System," Abst. and Ref. Video Proc. IEEE Int. Conf. Robotics and Automation, p.2, 1998.

[7] Akio Namiki, Yoshihiro Nakabo, Idaku Ishii, and Masatoshi Ishikawa: 1ms Sensory-Motor Fusion System, IEEE/ASME Trans. on Mechatronics, vol.5, no.3, pp.244-252, 2000.

[8] Akio Namiki, Yoshihiro Nakabo, Idaku Ishii, and Masatoshi Ishikawa: High Speed Grasping Using Visual and Force Feedback, Proc. IEEE Int. Conf. Robotics and Automation, pp.3195-3200, 1999.

[9] Akio Namiki, Yoshihiro Nakabo, Idaku Ishii, and Masatoshi Ishikawa: 1ms Grasping System Using Visual and Force Feedback, Abst. and Ref. Video Proc. IEEE Int. Conf. Robotics and Automation, p.12, 1999.

[10] http://www.k2.t.u-tokyo.ac.jp/~sfoc/

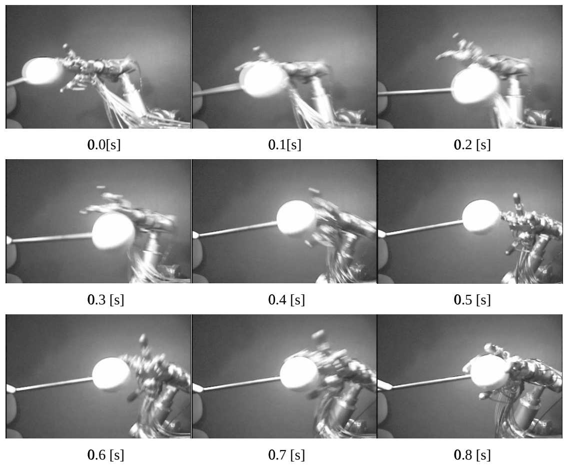


Figure 8: Experimental result: grasping of a sphere (close view)

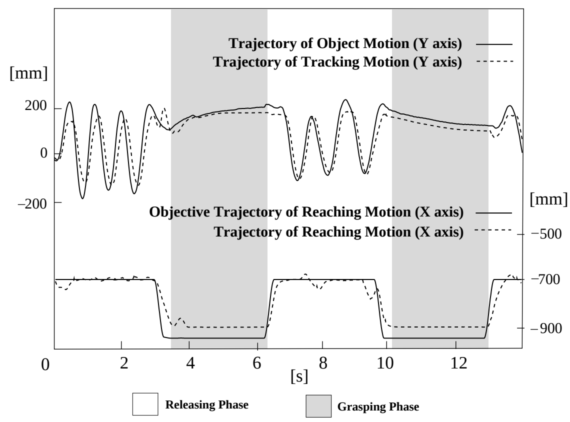


Figure 9: Response of the position