

# Acquisition of Multi-Modal Expression of Slip through Pick-Up Experiences

Yasunori Tada\* and Koh Hosoda\*\*

\* Dept. of Adaptive Machine Systems, Osaka University

\*\* Dept. of Adaptive Machine Systems, HANDAI FRC, Osaka University  
tada@er.ams.eng.osaka-u.ac.jp, hosoda@ams.eng.osaka-u.ac.jp

*Keywords: expression of slip, tactile sensor, multi-modal sensing, adaptive manipulation*

## Abstract

To realize adaptive and robust manipulation, a robot should have several sensing modalities and coordinate their outputs to achieve the given task based on underlying constraint in the real environment. This paper discusses on acquisition of multi-modal expression of slip consisting of vibration, pressure, and vision sensations through pick-up experiences. A sensor network is proposed to acquire the expression, whose learning ability is demonstrated by a real experiment. The applicability of the learned network is also demonstrated by experiments to realize robust and adaptive picking.

## 1 Introduction

We can utilize our fingers to touch, pick up, and manipulate various kinds of objects making use of tactile, force, and vision sensors. Although there have been an enormous number of studies on robot hands trying to reproduce such adaptive and dexterous behaviors [1], so far the performance is not satisfactory. The reason is supposed to be not only lack of sophisticated control strategy, but poor sensing abilities: dynamics existing among the fingers and the object seems to be too complicated to be observed by the existing sensor system.

A slip is one of such dynamic phenomena that often occurs during manipulation, therefore, should be observed by the sensor system. Numerous attempts have been made to produce sensors that can observe slips. Some studies utilized piezoelectric films embedded in soft materials, which could

sense vibration [2] [3] [4] [5] [6]. They detected initial slips by processing the output of the films. Vibration information from piezoelectric receptors only helps to detect micro slips, but not to detect the direction of the slip. Yamada and Cutkosky proposed to use not only piezoelectric receptors but a force sensor to sense the direction of the slip [7]. Several researches utilized strain gauges embedded in soft materials and differentiated the output with respect to space and/or time to detect slips[8][9]. Accelerometers [10] and air pressure sensors [11] were also used to detect slips by making use of the softness of the fingers. Since the initial micro slips are local phenomena, some studies utilized distributed array sensors and detected slips by finding local changes on them [12][13][14][15].

These sensor systems can observe micro slips and can be utilized to avoid them: not to drop the object. However, the designer should analyze micro slip phenomena and make a model to translate the vibration information into slip information by utilizing, for example, a FEM analysis. As a result, positions of the receptors should be controlled precisely when the sensor is produced, and the system is prone to the modeling error. Moreover, once a macro slip occurs, the robot should use a global sensor such as a vision sensor. These macro and micro slips are, actually, not really independent but continuous physical phenomena. Therefore, if the robot learns the correlation between the tactile sensor and the vision sensor, the tactile sensor is expected to observe the slip without the precise manufacturing and the modeling.

In this paper, we propose a sensor network consisting of not only one modality but three modalities, piezoelectric films, strain gauges, and a vision sensor, each of which provides sensation of vibration, pressure, and vision, respectively. The network is trained to acquire multi-modal expression of slips autonomously through pick-up experiences. Before learning, the robot does not know the relation between these sensations and the slip can only be detected by the vision sensor. Through pick-up experiences, it correlates the output of the vision with those of other receptors, and finally can learn to detect slips by vibration and pressure receptors without any physical modeling.

The remainder of this paper is organized as follows. First, we discuss about the multi-modal expression of the slip observed by a few sensations. Then, we propose a sensor network to acquire the relation between these sensations through experiences. The learning ability of the proposed network is demonstrated by a real experiment. Finally, we demonstrate that the learned network can be utilized to realize adaptive grasping by sensing micro slips. In the experiments, each experiment is repeated and shows that the proposed system has robustness for the detecting slip.

## **2 Multi-modal expression of the slip**

### **2.1 Macro and micro slips**

If the finger is rigid, a slip is observed as a relative movement between the finger and the object, and therefore, can be easily observed by sensors such as a vision sensor or strain gauges pasted on a surface of a finger [16]. However, once we introduce softness to the finger to increase robustness of the grasping and manipulation, it contacts with the object in certain area and phenomenon between them becomes complicated: at the beginning of the slip, there are few micro slips between the finger and the object, but there is no relative movement between them in a macro scale. As the exerted force grows, the number of micro slips increases gradually, and then, suddenly the finger begins to move relatively with the object since the number of micro slips catastrophically increases. The micro slips should be observed to predict the macro slip, and the macro slip should also be observed to control amount of the slip, therefore, it is crucial to observe these slips to achieve adaptive manipulation.

Although these slips are continuous phenomena, physical properties of sensors to observe them are different: the micro slips can be observed as vibrations by piezoelectric films or as spatial differentiation of a strain gauge array whereas the macro slips can be observed by a vision sensor. To utilize these receptors for smooth manipulation, therefore, the robot should know the relation between them. In the existing work, they did not deal these slips as a continuous process, and the sensors are calibrated by the robot designer. As a result, the sensor system is prone to the modeling error. If the robot can acquire the relation between them through experiences, it can utilize their continuity and obtain robust sensor system for both macro and micro slips.

### **2.2 Sensations of vibration and pressure**

If the finger has only the sense of vibration, it can detect the occurrence of the slip, but cannot observe its direction. On the other hand, the sense of pressure only gives the direction and strength of applied local force and cannot detect the occurrence of the slip. We could enhance the sensing ability of one of these sensations by making use of an array structure, but it will be more robust to utilize two different modalities together. In our implementation, the piezoelectric films and the strain gauges are used to provide the sense of vibration and pressure, respectively.

By introducing three different modalities, vision, vibration, and pressure, the sensing is expected to observe various contact conditions, but on the other hand, it is difficult to integrate these sensations

for realizing a given task. In the previous work, the relation between expressions in different modalities is ignored or calibrated by a human designer. Therefore, the resultant system becomes brittle against the modeling error. In this paper, we propose a sensor network that can learn the relation between the modalities through experiences. In the early stage of learning, the robot detects the slip as relative motion in the vision sensor, that is, a macro slip. The other modalities, sensations of vibration and pressure, will be trained through experiences. After learning, the robot can sense the micro slip and its direction as well even if the designer does not calibrate the receptors.

### 2.3 A sensor network that can learn multi-modal expression of the slip

In Figure 1, we show a system sketch that consists of a robot hand equipped with tactile receptors and a vision sensor. In Figure 2, we show a sensor network to acquire the multi-modal expression of the slip. The outputs of vibration and pressure receptors are normalized by their maximum values and are given as activations of tactile nodes. The visual information is coded as activations  $v_1$  and  $v_2$  denoting the relative movement between the hand and the object and the movement of the hand, respectively:

$$v_1 = \begin{cases} 1 & \text{(there is no relative motion between the hand and the object in the vision sensor)} \\ 0 & \text{(both the hand and the object do not move)} \\ -1 & \text{(there is relative motion between them)} \end{cases}, \quad (1)$$

$$v_2 = \begin{cases} 1 & \text{(the hand is moving upward in the vision sensor)} \\ 0 & \text{(it does not move)} \\ -1 & \text{(it is moving downward)} \end{cases}. \quad (2)$$

The tactile nodes  $t_j$  are connected to the output nodes  $o_i$  by weights  $w_{ij}$ :

$$o_i = f\left(\sum_j w_{ij} t_j\right), \quad (3)$$

where  $f(x)$  is a saturation function:

$$f(x) = \begin{cases} 1 & (x > 1) \\ x & (|x| < 1) \\ -1 & (x < -1) \end{cases}. \quad (4)$$

The structure of the proposed network which learns the relation between the sensors is suited for Hebbian learning. Hebbian learning is fast learning algorithm and is able to learn the

correlation on line. Therefore, the weights  $w_{ij}$  are updated basically based on the Hebbian learning rule according to the activations of tactile nodes and vision nodes [17], but it is slightly modified:

$$\Delta w_{ij} = \alpha r v_i t_j - \beta w_{ij}, \quad (5)$$

where  $\alpha$  and  $\beta$  are a learning rate and a forgetting rate, respectively.  $r$  is a variable learning rate:

$$r = (|w_{ij}| + \delta) / (\sum_j |w_{ij}| + \delta), \quad (6)$$

that accelerates the learning of a connection that has large weight, and decelerates the learning of other connections. This term helps to eliminate the effect of steady state offsets of receptors.

### 3 Experiments: picking up an object

#### 3.1 A robot system used for experiments

A robot system used for experiments is shown in Figure 3. It has a 7-DOF manipulator, PA-10 (Mitsubishi Heavy Industry) as an arm, two 2-DOF fingers (Yasukawa Electric Corporation) equipped with anthropomorphic fingertips, and a vision sensor. The detailed description of the anthropomorphic fingertip is shown in Figure 4[18]. It is basically imitating the human's finger, which has a metal rod as a bone, inner and outer layers as cutis and epidermis layers. We adopted PVDF (polyvinylidene fluoride) films as vibration receptors and foil strain gauges (Kyowa sensor system solutions) as pressure receptors. The absolute value of a PVDF film is adopted as the output of a vibration receptor since the sign of the film data has no sense about vibration. We embedded 6 films and 6 strain gauges in each layer, that is, one fingertip has totally 24 receptors. The control rate is 1 [kHz]. Data from the tactile receptors and the vision sensor are updated in 1 [kHz] and 30 [Hz], respectively. 1 [pixel] in the vision sensor equals 1.32 [mm] in the world coordinate frame.

#### 3.2 A Learning procedure

If the behavior of the robot is random, it takes so much time to learn. To accelerate learning, we embedded a simple pick-up behavior to the robot system shown in Figure 5: (1) The arm moves the hand upward in its Cartesian frame while distance between fingers is controlled to be smaller

gradually, (2) the fingers slip along the surface of the object while the distance between fingers is not small enough, (3) the hand succeeds to pick up the object, (4) after it succeeds to pick up, the arm moves the hand downward in its Cartesian frame, (5) it continues to move while the fingers slip along the surface of the object downward. Meanwhile, the sensor network learns the relation between the receptors and the vision sensor. In the learning, the object is a cup that weight is 450 [g]. In the learning procedure and other experiments after learning, the contact area on the fingertip, initial position and posture of the hand are the same, and the moving speed of the hand is 2.5 [cm/s].

We recorded the coded output of the vision sensor and output of tactile receptors during the behavior (Figure 6). The numbers on the top of the figures indicate steps of the learning procedure. The robot repeats the behavior 2 times in 20 [s]. Figure 6 (a) and (b) show the coded output of the vision sensor,  $v_1$  and  $v_2$ , when there is relative motion in the vision sensor between the hand and the object and when the hand moves upward/downward in the vision sensor, respectively. In these figures, the output of the vision sensor seems like chattering because of a following reason. The vision sensor is updated in 30 [Hz]. Therefore, if the vision sensor detects the motion of the object, it continues to output -1 or 1 at least 33 [ms]. However, if the motion of the object is slower than the sampling rate of the vision sensor, the vision sensor does not exactly output -1 or 1 in every frame. As a result, the output of the vision sensor seems like chattering. Figure 6(c) and (d) show two typical time courses of the normalized output of strain gauges. Some of the receptors only generate positive values like (d). We can speculate that those receptors which only generate positive values are measuring grasping force. Other receptors like (c) sense tangential force. Figure 6 (e) and (f) show two typical time courses of the unsigned normalized output of PVDF films. Depending on the depth of the receptor, the sensitivity may change. From comparing Figure 6(a) with (e), only when the relative motion is observed by the vision sensor  $v_1$ , a vibration receptor outputs large signal. Therefore, the vibration receptor is expected to become a slip sensor.

### 3.3 Learning expression of the slip through experiences

Before learning, the output of the network  $o_i$  is 0 since we set the initial values of connection weights  $w_{ij} = 0$ . Therefore, the robot can detect occurrence of the slip and its direction by only the vision sensor before learning. During the learning process, the network finds the correlation between output of the vision sensor and the tactile receptors. The learning is iterated until that the output of the network becomes large sufficiently. Figure 7 shows the detected occurrence of slip by the vision sensor and that by the learned network after 7 learning trials. The learned network

can sense the slip earlier (0.76[s]) than just using the vision sensor (0.94[s]). In this experiment system, the resolution of the vision sensor and the moving speed of the hand are 1.32 [mm/pixel] and 2.5 [cm/s], respectively. Thus, the vision sensor needs at least 2 frames (66 [ms]) to observe the macro slip. The time difference of detected slip between the vision sensor and the proposed network is 0.18 [s] which is more than 5 frames. Therefore, we conclude that the network can detect the micro slip before occurrence of the macro slip whereas the vision sensor can detect only the macro slip.

In the top left graph, the network  $o_1$  does not output 1 from 2 to 6 [s] and is discord from the output of the vision sensor  $v_1$ . A reason is that the tactile sensor can observe the vibration only when the slip occurs. Therefore, the output of the network  $o_1$  equals 0 when the slip does not occur. Additionally, a reason that the network  $o_2$  continues to output 1 from 4.5 to 6 [s] whereas the vision sensor  $v_2$  outputs 0 in the top right graph is as follows. The hand is stopped at 4.5 [s] by the designer's directive but continues to grasp the object at this time. Therefore, the network  $o_2$  continues to output 1 whereas the vision sensor  $v_2$  outputs 0 because there is no motion of the object.

We repeat the experiment and verify whether the network can detect the slip in different objects from the learning phase. The objects are the cup (450[g]) in the learning phase, another cup which has the different friction coefficient, square timber of 250, 350, 450, 550, and 650 [g]. The robot repeats picking up the objects 50 times in each object. Figure 8 shows averages and standard deviations of the first occurrence time of the slip. This result shows that the learned network can adapt to different objects and can detect the slip earlier than the vision sensor.

### 3.4 Pick up experiments utilizing the learned network

By utilizing the learned network, the robot can successfully pick up the object without slips. We implemented a simple controller: when the network or the vision sensor detects a slip, the robot increased the grasping force by changing the distance between the fingers. In Figure 9, we show movements of the object in the vision sensor (a) by utilizing the neural network and (b) by utilizing only the vision sensor. We could find that if the network is utilized to detect the slip, the robot can pick up the object 0.3 [s] earlier than just using the vision sensor (1.65[s]).

We repeat the experiment and verify whether the network adapts to another object. The robot repeats the experiment 50 times in the objects which are used in the learning phase and the square timber of 450 [g]. Figure 10 shows averages and standard deviation of time of picking up the

objects. This figure shows that the robot using the network can pick up the object earlier than using the vision sensor and adapt to another object. Moreover, we applied the learning method to other fingertips. Some of them were capable to learn to pickup the object, but some were not. The ability of the proposed method is obviously dependent on distribution of receptors. Since we do not control it, the learning ability will change accordingly.

We conducted another experiment. At the beginning of the experiment, the robot picks up the cup and holds the condition of grasping. While the robot grasps the cup, we pour water into it to increase the weight. The robot will detects a slip and increase the grasping force not to drop it. In **Figure 11**, we compare two cases: with the proposed network and without the network but only utilizing the vision sensor. We can conclude that the hand can grasp the cup by adapting to the slip and does not drop if we utilize the learned network whereas the slippage is larger if we use only the vision information to detect the slip. We also repeat the experiment and measure averages and standard deviations of the slippage pixels. **Figure 12** shows that the proposed network adapts to the slip and the amount of the slip is smaller than when the hand is controlled by the vision sensor.

## **4 Conclusions and Discussion**

In this paper, we have proposed a network that can acquire the multi-modal expression of slips by making use of three different modalities: vibration, pressure, and vision sensations. Through grasping experiences, the network is trained to sense not only macro slips but micro ones. Experimental results have demonstrated that the learned network can be utilized for adaptive grasping.

Since the aim of this paper is to show basic learning ability of the proposed network, the task given for the robot is extremely simple: grasping and lifting up the object. Further goal for developing such a sensor system is to deal with a variety of tasks. Therefore, we should demonstrate further ability of the network by achieving more tasks, and hopefully really dexterous manipulation. In the human development, the human may learn the complex relationship between the vision and the tactile through the experience, and will achieve the dexterous task without the vision. In this sense, we should discuss further what kind of information should be processed from the vision sensor. If the given task is simple like this paper, the robot achieves the task by the simple vision information. However, if the robot achieves the more complex task, the robot will need the complex visual information to learn the neural network. We should also consider the procedure for learning.

In the proposed method, the learning and executing phases are distinguished. We should further



consider the network architecture that can learn while it performs the given task. If the network can learn in the context of sensory-motor coordination, the expression of phenomena in the network should be different since we do not have to reinforce the network by a certain sensor (in this case, a vision sensor) but just utilize the performance of the task.

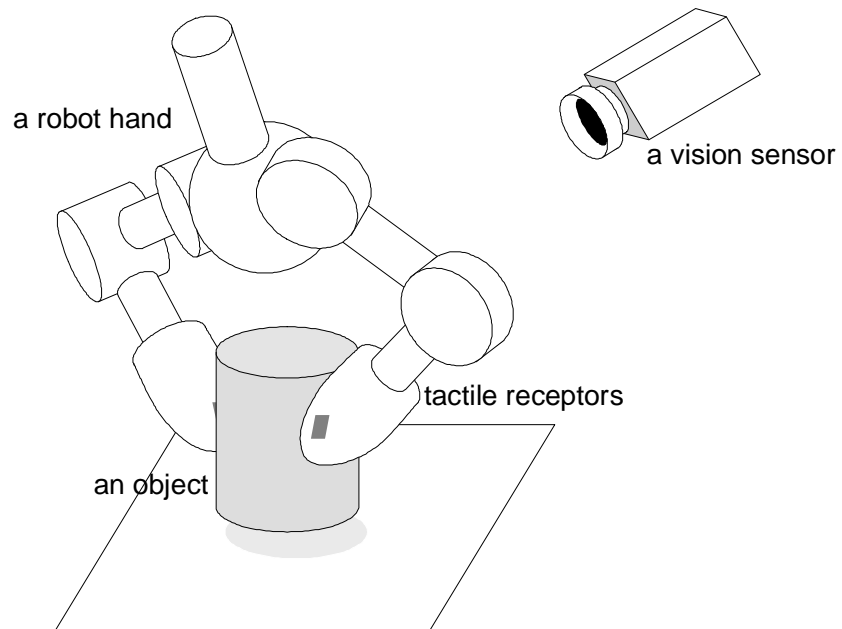
## Acknowledgement

The authors would like to thank their colleagues, Dr. Minoru Asada and Mr. Atsushi Fukuda, for valuable discussions, comments, and helping the experiments. This study was partly supported by the Advanced and Innovational Research Program in Life Science, and partly supported by Grant-in-Aid for Scientific Research (B) #16300056 from the Ministry of Education, Science, Sports, and Culture of the Japanese Government.

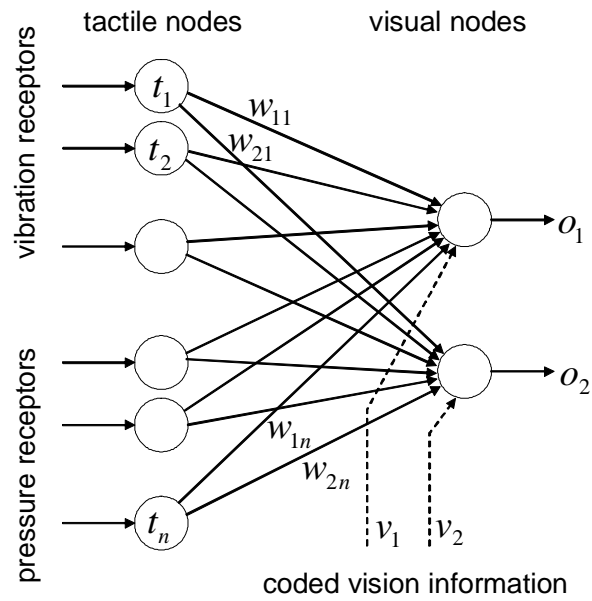
## References

- [1] A. Bicchi and V. Kumar, "Robotic Grasping and Contact: A Review", IEEE Int. Conf. on Robotics and Automation, pp.348-353, 2000.
- [2] J. S. Son, E. A. Monteverde, and R. D. Howe, "A Tactile Sensor for Localizing Transient Events in Manipulation", In Proc. of the 1994 IEEE Int. Conf. on Robotics and Automation, pp.471-476, 1994.
- [3] J. Jockusch, J. Walter, and H. Ritter, "A Tactile Sensor System for a Three-Fingered Robot Manipulator", In Proc. of the 1997 IEEE Int. Conf. on Robotics and Automation, pp.3080-3086, 1997.
- [4] D. J. O'Brien and D. M. Lane, "Force and Slip Sensing for a Dexterous Underwater Gripper", In Proc. of the 1998 IEEE Int. Conf. on Robotics and Automation, pp.1057-1062, 1998.
- [5] Y. Yamada, H. Morita, and Y. Umetani, "Vibrotactile Sensor Generating Impulsive Signals for Distinguishing Only Slipping States", In Proc. of the 1999 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.844-850, 1999.
- [6] I. Fujimoto et al., "Development of Artificial Finger Skin to Detect Incipient Slip for Realization of Static Friction Sensation", In Proc. of IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, pp.15-20, 2003.
- [7] Y. Yamada and M. R. Cutkosky, "Tactile Sensor with 3-Axis Force and Vibration Sensing Functions and Its Application to Detect Rotational Slip", In Proc. of the 1994 IEEE Int. Conf. on Robotics and Automation, pp.3550-3557, 1994.
- [8] T. Maeno, S. Hiromitsu, and T. Kawai, "Control of Grasping Force by Detecting Stick/Slip Distribution at the Curved Surface of an Elastic Finger", In Proc. of the 2000 IEEE Int. Conf. on Robotics and Automation, pp.3896-3901, 2000.

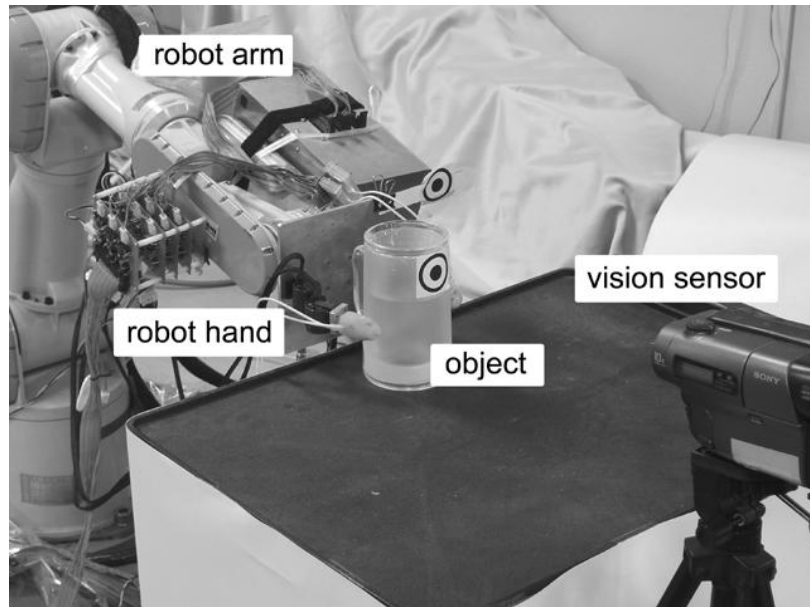
- [9] D. Yamada, T. Maeno and Y. Yamada, "Artificial Finger Skin having Ridges and Distributed Tactile Sensors used for Grasp Force Control", In Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.686-691, 2001.
- [10] M. R. Tremblay and M. R. Cutkosky, "Estimation Friction Using Incipient Slip Sensing During a Manipulation Task", In Proc. of the 1993 IEEE Int. Conf. on Robotics and Automation, pp.429-434, 1993.
- [11] H. Shinoda, M. Uehara, and S. Ando, "A Tactile Sensor using Three-Dimensional Structure", In Proc. of the 1993 IEEE Int. Conf. on Robotics and Automation, pp.435-441, 1993.
- [12] E. G. M. Holweg et al., "Slip Detection by Tactile Sensors: Algorithms and Experimental Results", In Proc. of the 1996 IEEE Int. Conf. on Robotics and Automation, pp.3234-3239, 1996.
- [13] C. Melchiorri, "Slip Detection and Control Using Tactile and Force Sensors", IEEE/ASME Trans. on Mechatronics, Vol. 5, No.3, pp.235-243, 2000.
- [14] A. Sano et al., "Multi-Fingered Hand System for Telepresence Based on Tactile Information", In Proc. of the 2004 IEEE Int. Conf. on Robotics and Automation, pp.1676-1681, 2004.
- [15] N. Tsujiuchi et al., "Slip Detection with Distributed-Type Tactile Sensor", In Proc. of 2004 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.331-336, 2004.
- [16] L. Birglen and C. M. Gosselin, "Fuzzy Enhanced Control of an Underactuated Finger Using Tactile and Position Sensors", In Proc. of the 2005 IEEE Int. Conf. on Robotics and Automation, pp.2331-2336, 2005.
- [17] R. Pfeifer and C. Scheier, "Understanding Intelligence", The MIT Press, 1999.
- [18] K. Hosoda, "Robot Finger Design for Developmental Tactile Interaction - Anthropomorphic Robotic Soft Fingertip with Randomly Distributed Receptors", Embodied Artificial Intelligence, Fumiya Iida et al. Eds., Springer-Verlag, pp.219-230, 2004.



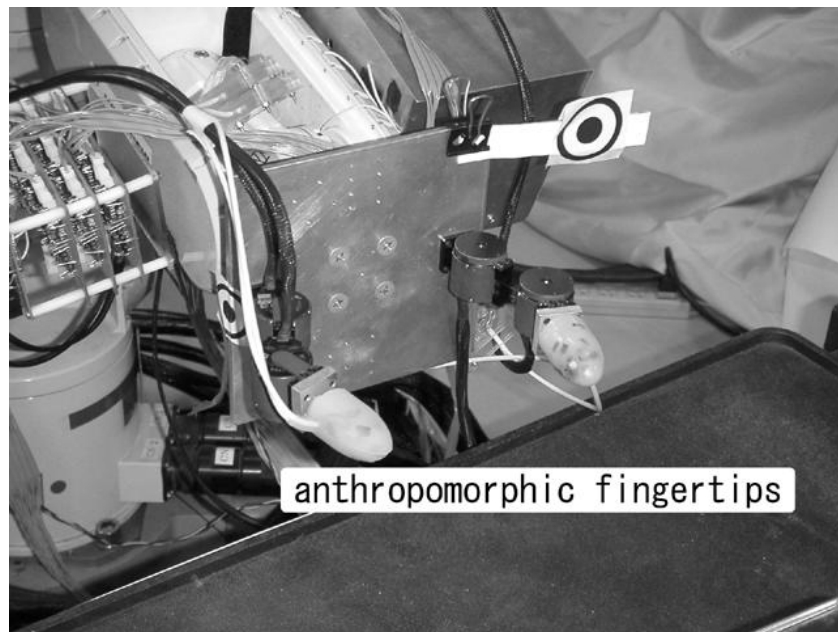
**Figure 1: A robot system consists of fingers equipped with tactile receptors and a vision sensor. The relation between the sensor and receptors is not known beforehand. The task for the robot is to pick up the object.**



**Figure 2:** A sensor network that learns multi-modal expression of the slip. The weights between the tactile nodes and the visual nodes are updated by a Hebbian rule.

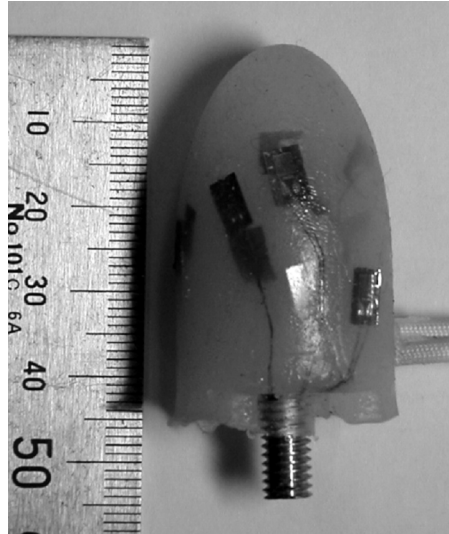


(a) A robot system used for experiments. The robot has an arm, two fingers equipped with anthropomorphic fingertips, and a vision sensor. The task for the robot is to pick up an object.

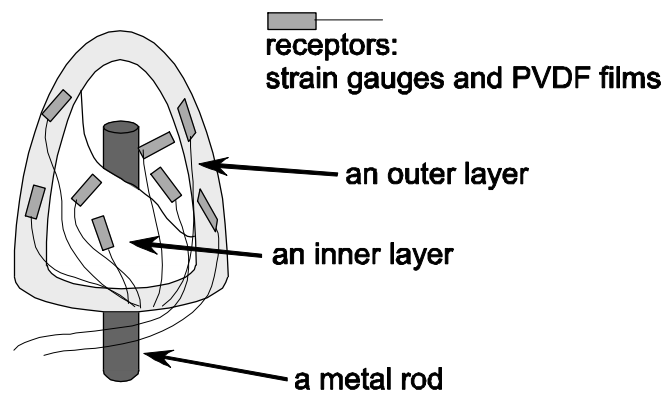


(b) The robot hand with anthropomorphic fingertips. The each finger has two degrees of freedom.

**Figure 3: The real robot system used for experiments.**

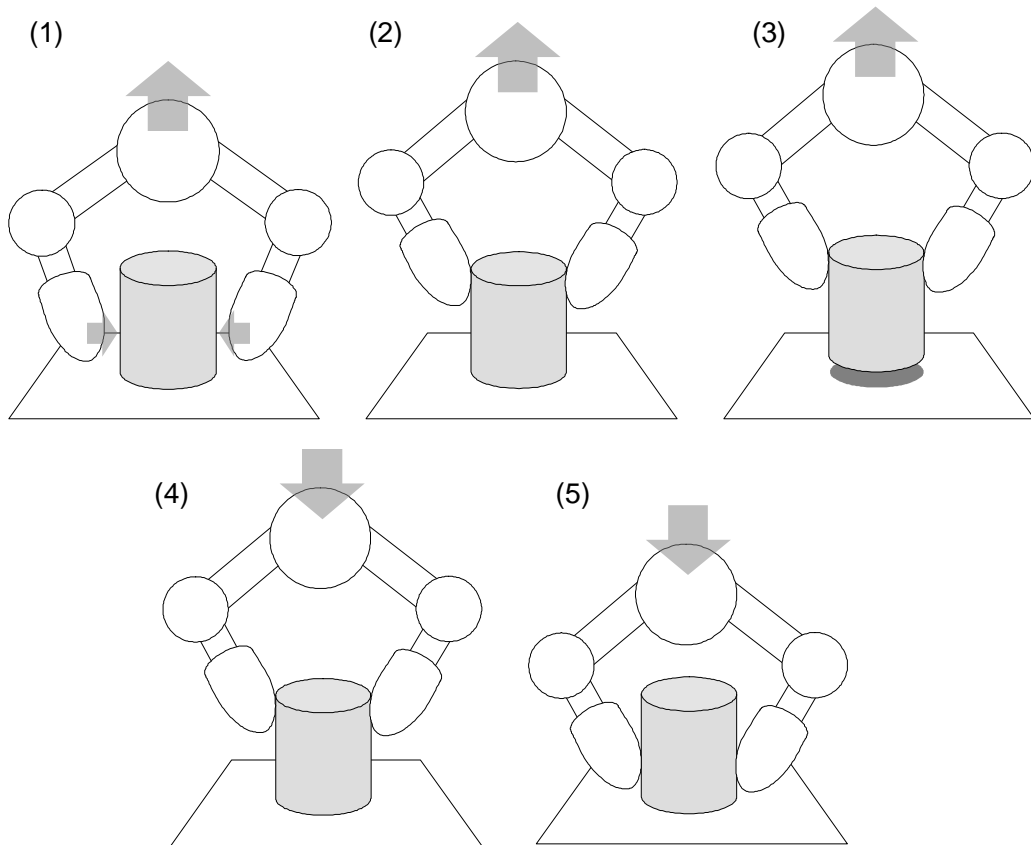


(a) A photo of the anthropomorphic fingertip of the robot system. Its length and diameter are 45mm and 25mm, respectively.

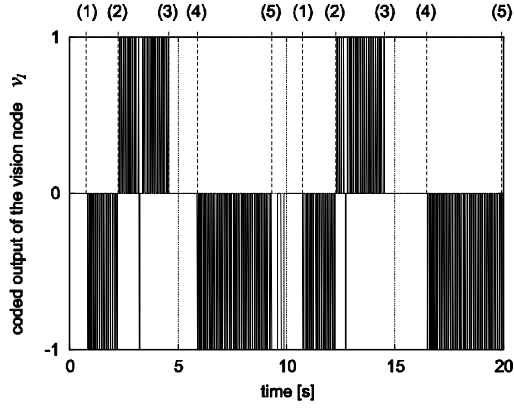


(b) A cross sectional sketch of the fingertip. The fingertip has two layers and a metal rod imitating the structure of the human's.

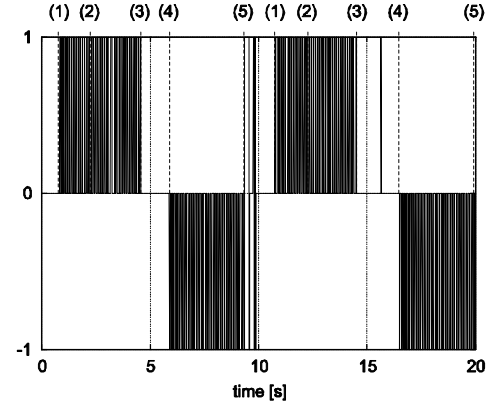
**Figure 4: An anthropomorphic fingertip used for the experiments**



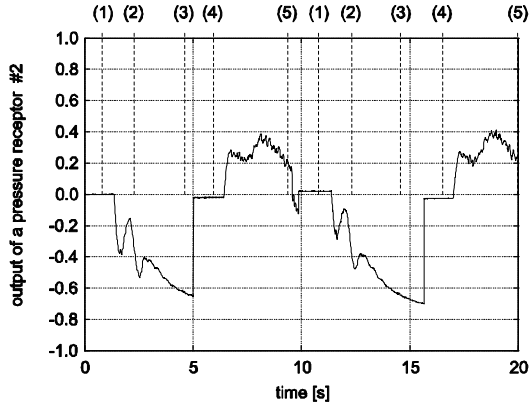
**Figure 5: An embedded behavior for the robot system to learn multi-modal expression of the slip. (1) The arm moves the hand upward in its Cartesian frame while the fingers are position-controlled to close, (2), (3) the hand succeeds to pick up the object. (4) After it succeeded to pick up, the arm moves the hand downward in its Cartesian frame, (5) it continues to move the hand downward while the fingers keep to touch the object.**



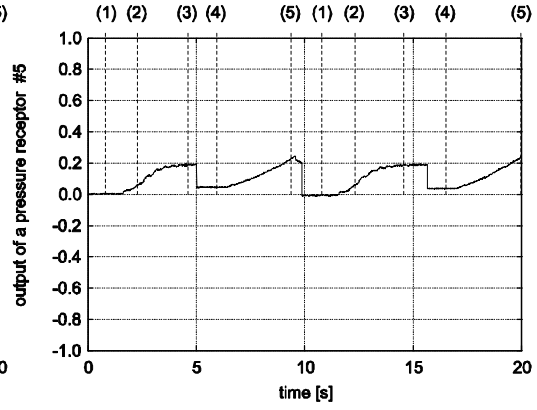
(a) the coded output of the vision sensor  $v_1$  when there is relative motion in vision the sensor between the hand and the object.



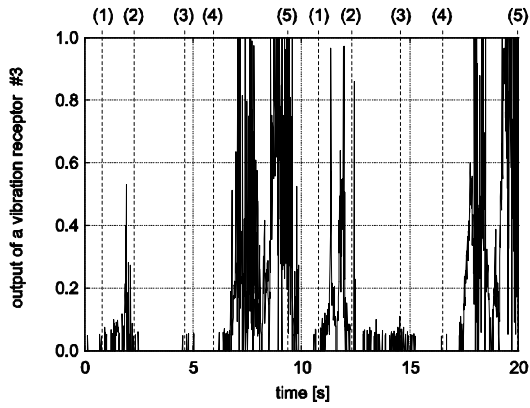
(b) the coded output of the vision sensor  $v_2$  when the hand moves upward/downward in the vision sensor.



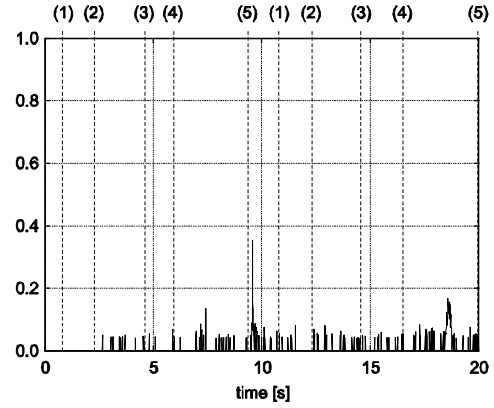
(c) output of a pressure receptor #2



(d) output of a pressure receptor #5



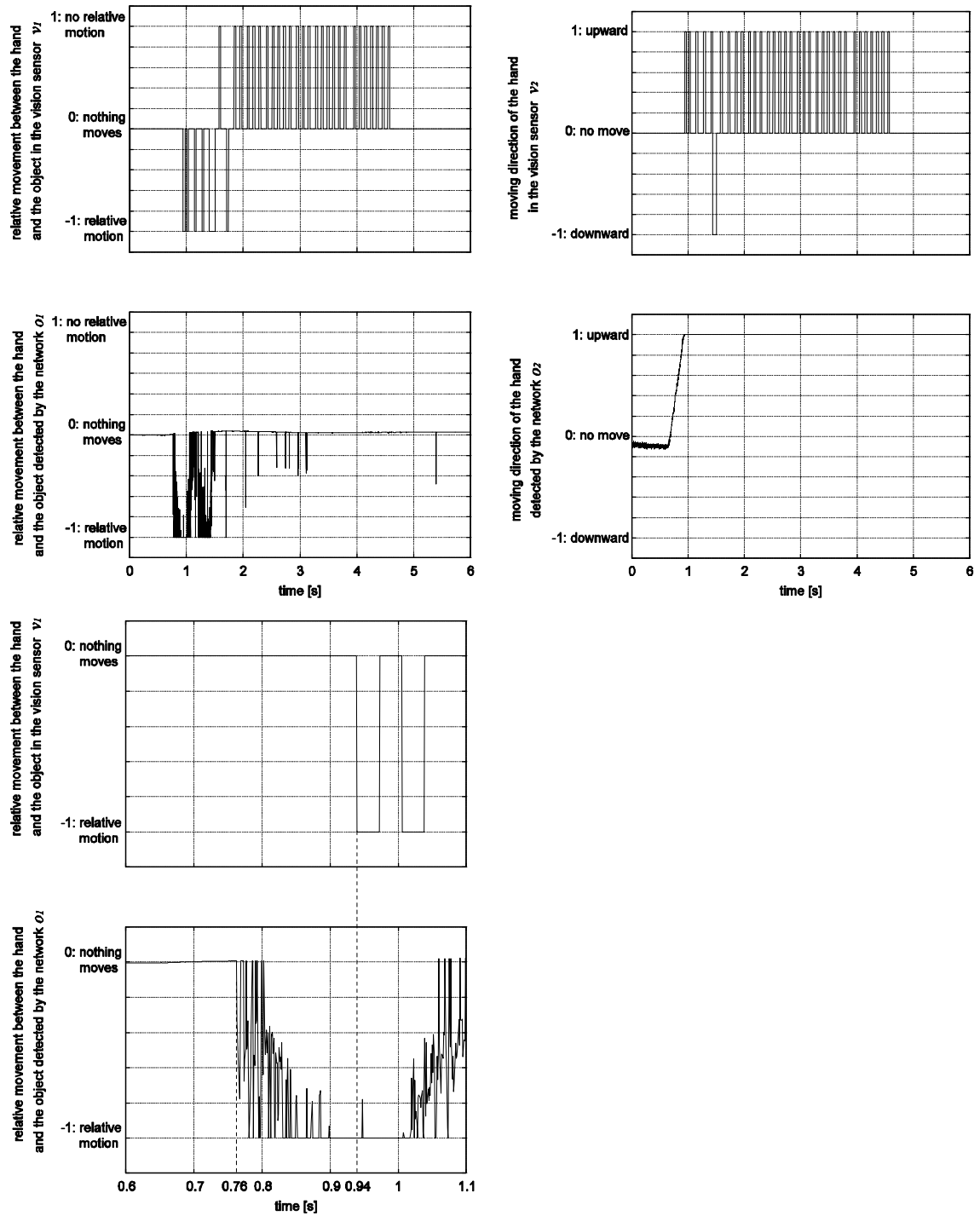
(e) output of a vibration receptor #3



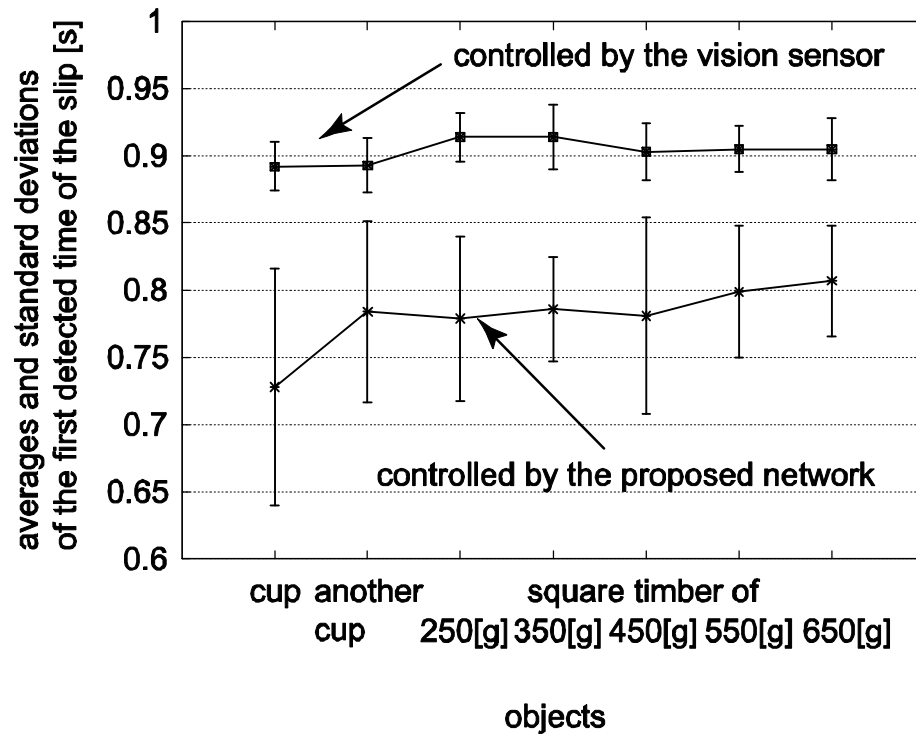
(f) output of a vibration receptor #4

**Figure 6:** the coded output of the vision sensor and output of tactile receptors during repeating the behavior 2 times.

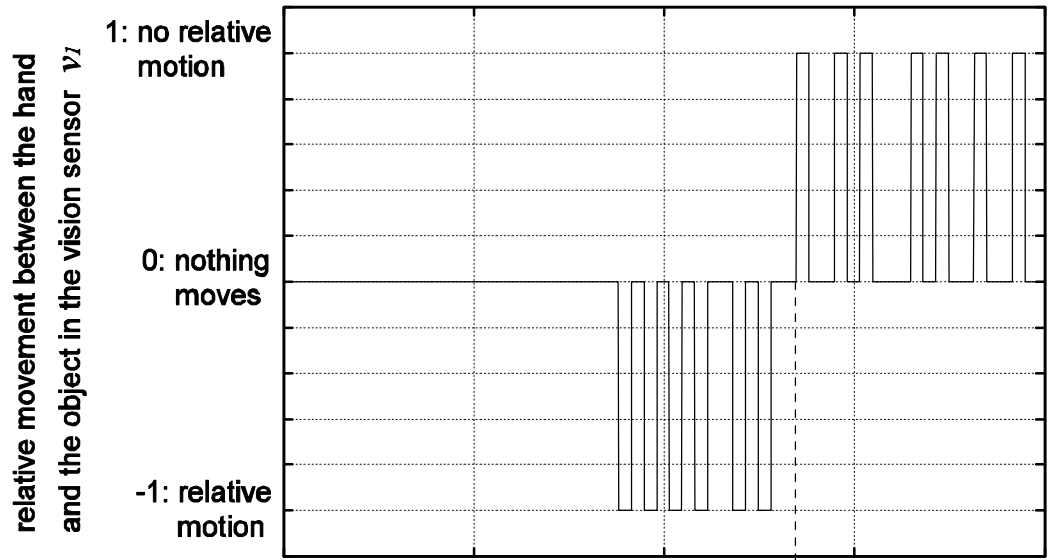




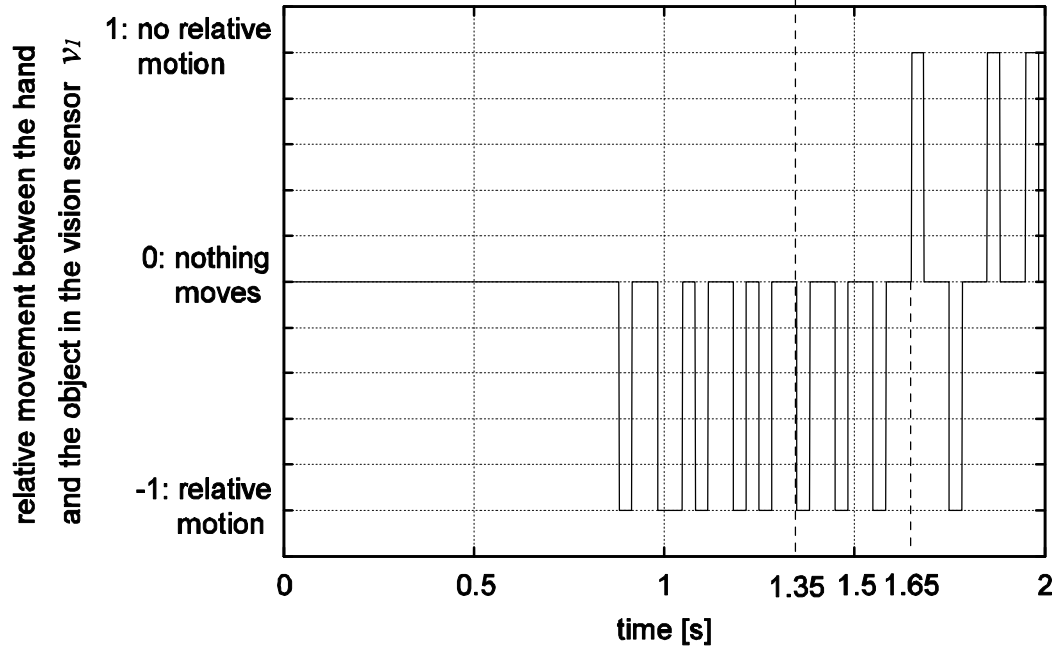
**Figure 7:** Detected occurrence of slip by the vision sensor and that by the learned network after 7 learning trials. Two graphs of top left show the occurrence of slip detected by a vision sensor (top) and by the proposed sensor network (bottom). Since it is difficult to see the detailed difference between them, we magnify these graphs into [0.6, 1.1] in bottom two graphs. Two graphs of top right show the detected direction of the slip by the vision sensor (top) and by the proposed network.



**Figure 8:** Averages and standard deviations of detected slip time show that the proposed network can detect the slip on the different objects from learning phase and can detect the slip earlier than the vision sensor.



(a) controlled by the neural network



(b) controlled by the vision sensor

Figure 9: Observed macro slips in the vision sensor of pick-up experiments, by utilizing the proposed network (top) and by utilizing only vision sensor (bottom). If we use proposed network, the macro slip stops at 1.35[s] whereas it does at 1.65[s] if we use only the vision sensor.

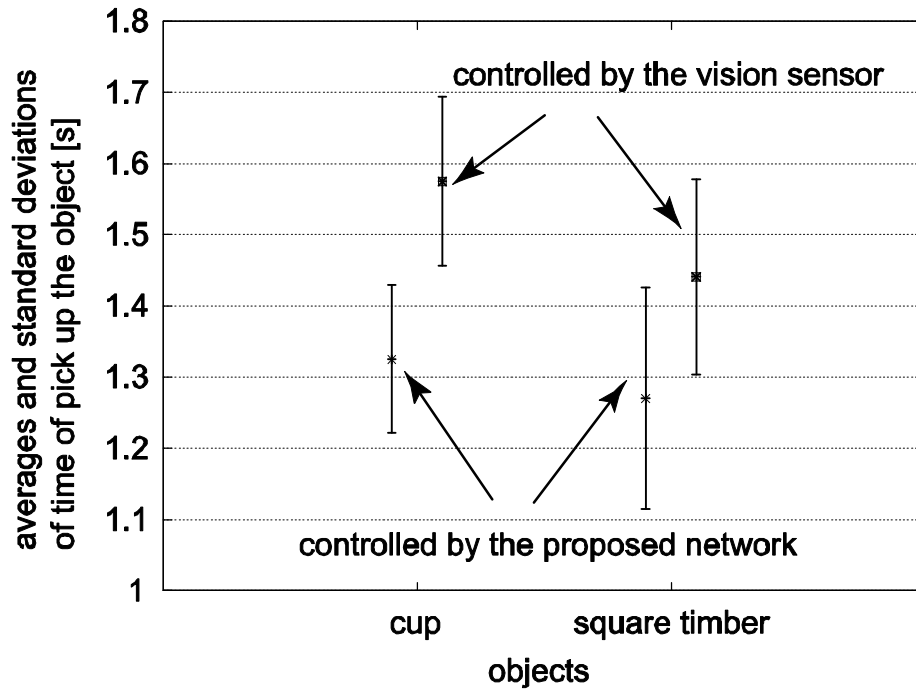


Figure 10: The proposed network is utilized in the controlling of the grasping force and adapts to another object which is used in the learning phase. The network can pick up the objects earlier than the vision sensor because the network can detect the micro slip.

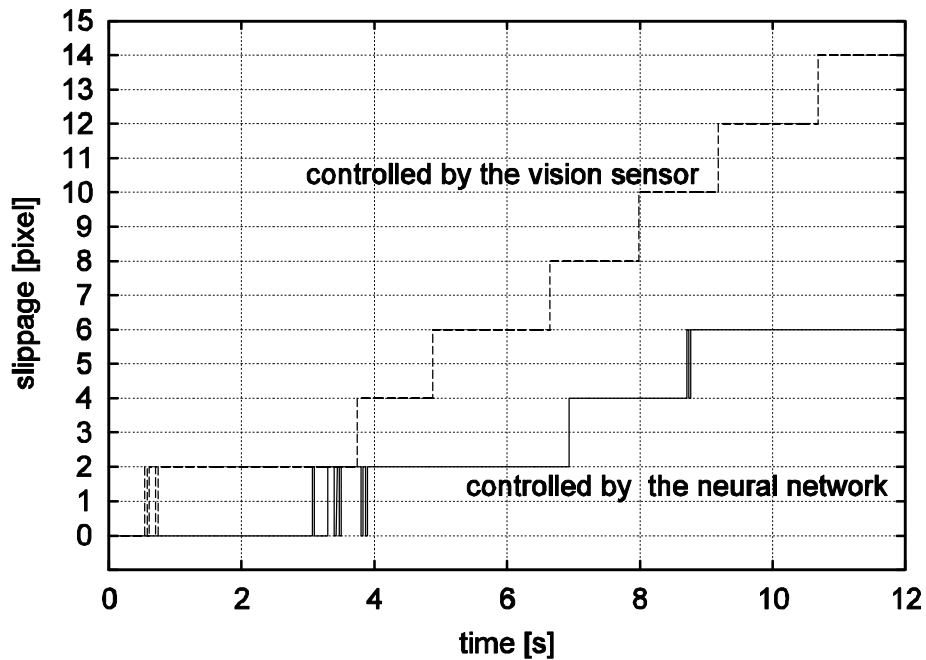
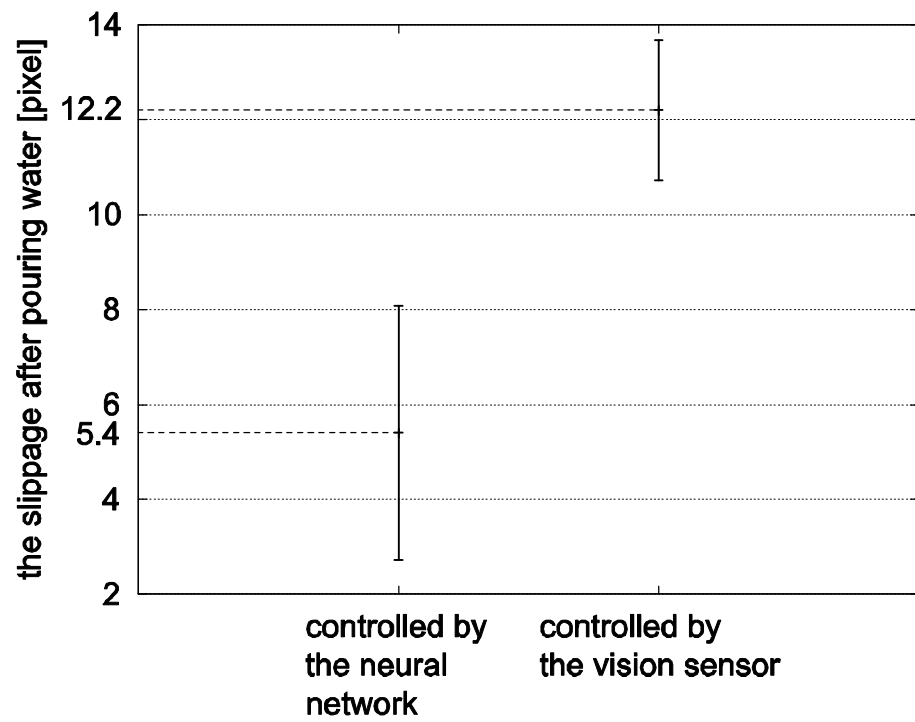


Figure 11: Pick-up experiment 2. The experimenter poured water into the cup that was grasped by the fingers. The amount of slip is smaller when it is controlled by the proposed neural network than when it is controlled by the vision sensor.



**Figure 12:** The proposed network adapts to the slip and the amount of the slip is smaller than when the hand is controlled by the vision sensor.