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

Yasunori Tada
Department of Adaptive Machine Systems
Osaka University
Osaka, Japan
Email: tada@er.ams.eng.osaka-u.ac.jp

Koh Hosoda
Department of Adaptive Machine Systems
Osaka University
HANDAI FRC
Osaka, Japan
Email: hosoda@ams.eng.osaka-u.ac.jp

***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 adaptive picking.**

I. 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 that should be also calibrated with the tactile receptors to preserve the observation consistency between them.

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.

II. Multi-modal expression of the slip

A. 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

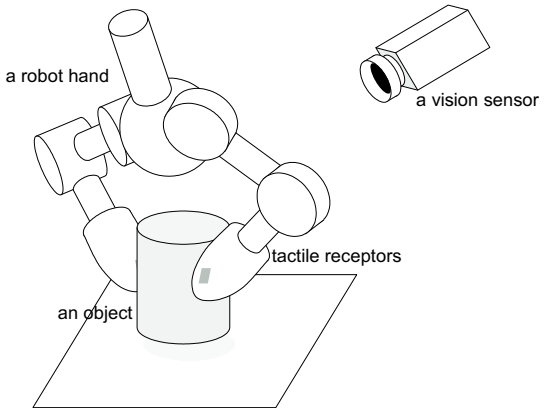


Fig. 1. A robot system consists of fingers equipped with tactile receptors and a vision sensor.

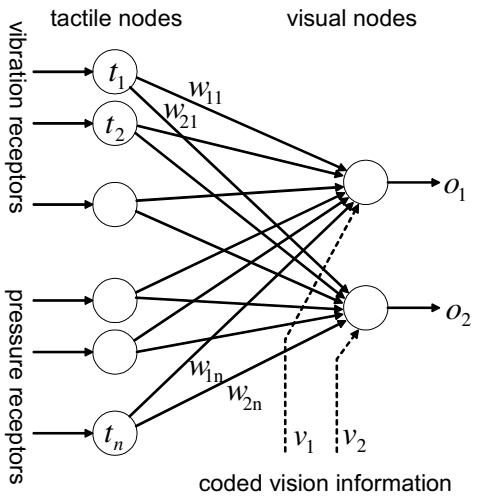


Fig. 2. A sensor network that learns multi-modal expression of the slip. The weights between the tactile nodes and the vision nodes are updated by a Hebbian rule.

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.

B. 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 improve the sensing ability to utilize two 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 modalities, vision, vibration, and pressure, the sensing is expected to improve the sensing ability and to be adaptive, 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.

C. A sensor network that can learn multi-modal expression of the slip

In Fig. 1, we show a system sketch that consists of a robot hand equipped with tactile receptors and a vision sensor. In Fig. 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 and denoting the relative movement between the hand and the object and the movement of the hand, respectively:

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

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

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

$$o_i = f\Big(\sum_j t_j w_{ij} + v_i\Big) \quad (3)$$

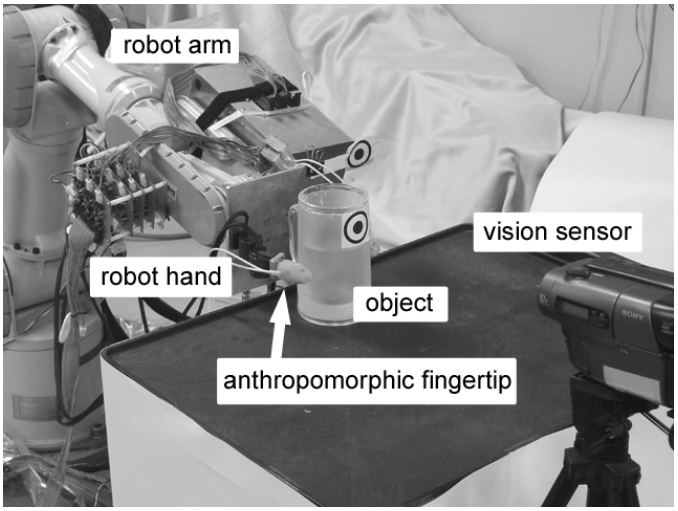


Fig. 3. A robot system used for experiments. The robot has an arm, two fingers equipped with anthropomorphic fingertips, and a vision sensor.

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 weight w_{ij} are updated basically based on the Hebbian learning rule according to the activations of tactile nodes and vision [17]. The Hebbian learning is a fast learning algorithm and is able to learn in online. Additionally, the structure of the network is understood viscerally. Thus, we expect that the network and the Hebbian learning are suitable for acquiring the relation between the sensors. In this paper, however, the algorithm is slightly modified:

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

where α and β are a learning rate and a forgetting rate, respectively. r is a variable learning rate:

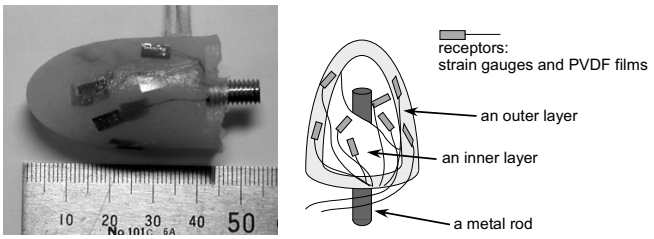
$$r = (|w_{ij}| + \delta)/(\sum_j |w_{ij}| + \delta) \quad (6)$$

where δ is an arbitrary positive small value. r 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 offsets of receptors.

III. Experiments: Picking Up an Object

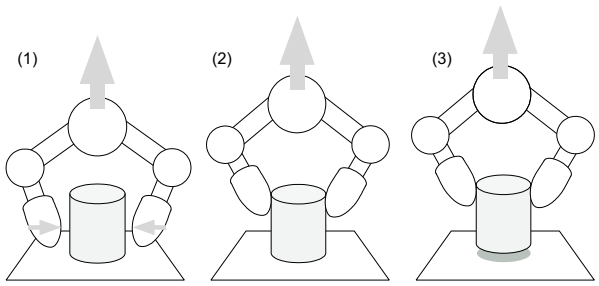
A. A robot system used for experiments

A robot system used for experiments is shown in Fig. 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 Fig. 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 detecting the 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.

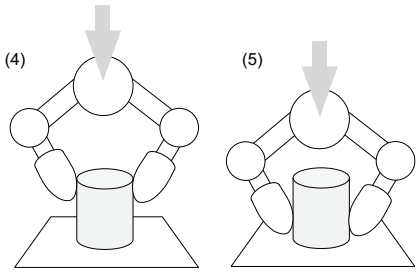


(a) A photo of the anthropomorphic fingertip (b) A cross sectional view of the fingertip

Fig. 4. An anthropomorphic fingertip used for the experiments



(a) The robot hand picks up an object: (1) the arm moves the hand upward while the fingers are position-controlled to close, (2), (3) the hand succeeds to pick up the object.

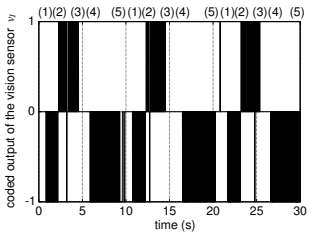


(b) After the picking up the object, (4) the arm moves the hand downward, (5) it continues to move the hand while the fingers keep to touch the object.

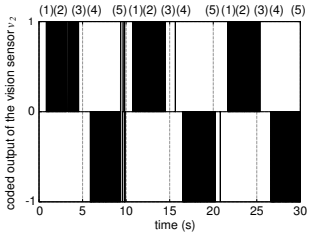
Fig. 5. An embedded behavior for the robot system to learn multi-modal expression of the slip.

B. A learning procedure

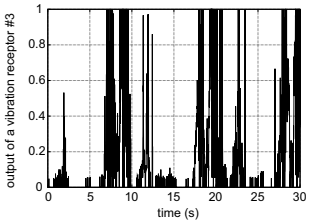
If the behavior of the robot is random, it takes so much time to learn. To accelerate learning, we embedded a simple pick-up behavior to the robot system shown in Fig. 5: (1)



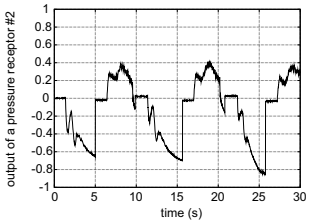
(a) The coded output of the vision sensor v_1 when there is relative motion in vision between the hand and the object.



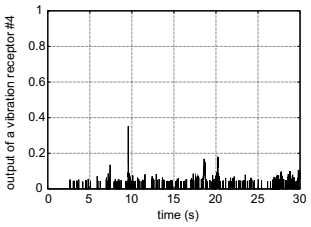
(b) The coded output of the vision sensor v_2 when the hand moves upward/downward in the vision.



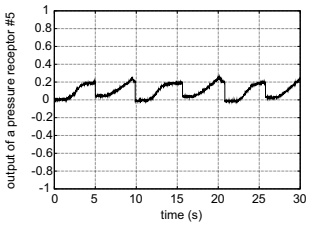
(c) Output of a vibration receptor #3



(d) Output of a pressure receptor #2

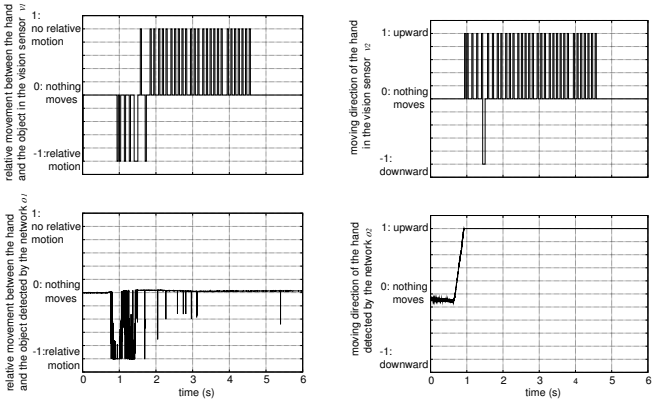


(e) Output of a vibration receptor #4



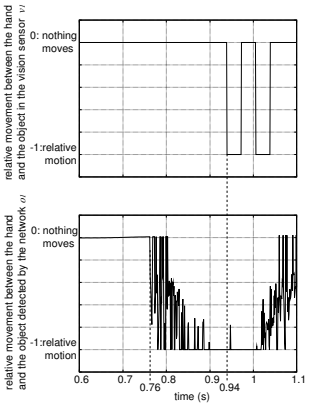
(f) Output of a pressure receptor #5

Fig. 6. The coded output of the vision sensor and output of tactile receptors during repeating the behavior 3 times.

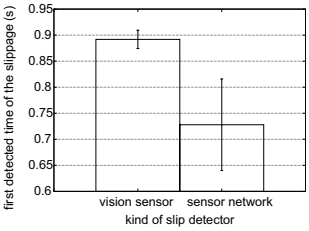


(a) Output of the vision sensor v_1 and the sensor network o_1

(b) Output of the vision sensor v_2 and the sensor network o_2



(c) Magnified graph (a) from 0.6 to 1.1 (s)



(d) Averages and standard deviations of the first detected time of the slippage by 50 trials

Fig. 7. Detected occurrence of slip by the vision sensor and that by the learned network after 7 learning trials.

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 grasping force is not 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. Additionally, contact areas at the fingertips is the same through the experiments in the learning and after the learning. Meanwhile, the sensor network learns the relation between the receptors and the vision sensor.

We recorded the coded output of the vision sensor and output of tactile receptors during the behavior (Fig. 6). The robot repeats the behavior 3 times in 30 s. Fig. 6(a) and (b) show the coded output of the vision sensor when there is relative motion in vision sensor between the hand and the object and when the hand moves upward/downward in the vision sensor, respectively. Numerals at the top of the figures indicate the number of the learning procedure in Fig. 5. A reason that there is chattering in the coded output of the vision sensor is as follows. The output of the vision sensors is updated by each 33 ms. Because of the frame rate, the vision sensor continues to output values -1, 0, or 1 at least 33 ms. If the vision sensor detects the motion of the target mark on the arm and the object, the sensor outputs the values -1 or 1. However, if the motion of the target mark is slow, the vision sensor does not exactly output the values -1 or 1 in every frame because of the quantization error. As a result, the output of the vision sensor seems like chattering.

Fig. 6(c) and (e) show two typical time courses of the unsigned normalized output of PVDF films. Depending on the depth of the receptor, the sensitivity may change. Fig. 6(d) and (f) show two typical time courses of the normalized output of strain gauges. Some of the receptors only generate positive values like (f). We can speculate that those receptors which only generate positive values measure grasping force. Other receptors like (d) are sensing friction force. Additionally, from Fig. 6(a) and (c), the output of the vibration receptor is large when the vision sensor v_1 equals -1: the slippage is observed. In contrast, the output of the vibration receptor is very small when the vision sensor v_1 equals 1: the slippage is not observed. Therefore, we expect that the outputs of the vibration receptors are response to the slip.

C. Leaning expression of the slip through experiences

Before learning, the output of the network is since we set the initial values of connection weights. Therefore, the robot can detect occurrence of the slip and its direction only by the vision sensor before learning. During the learning process, the network finds the correlation between output of the vision sensor and those of tactile receptors. We iterated the learning procedure 7 times. Fig. 7(a) shows the detected occurrence of the slip by the vision sensor and the learned network when the robot picks up the object. In the figure, the network o_1 does not output the "no relative motion" from 2 to 6 s whereas the vision sensor v_1 outputs "no relative motion" at that time. A reason is as follows. In the learning phase, vibration receptors output the large signal only when the relative motion occurs. As a result, connection weights between the vibration receptors and o_1 are reinforced. After the learning, the output of the network o_1 depends on the output of the vibration receptors only. Additionally, the vibration receptor does not output a signal when there is no relative motion because of no vibration. Therefore, the network does not output the "no relative motion" from 2 to 6 s. On the other hand, in Fig. 7(b) which shows the moving direction, the vision sensor v_2 outputs "nothing moves" during 4.5 and 6 s because the robot hand is stopped at 4.5 s. However, the robot hand continues to grasp the object at this time. As a result, the pressure receptors continue to output the signal. Therefore, even if the vision sensor v_2 outputs "nothing moves", the network o_2 outputs "up ward". The vision sensor outputs the signal when the sensor only detects the motion. However, if the robot hand touches the object, the vibration and the pressure receptors output the signal. This is a difference of the characteristic between the vision and tactile sensor.

Fig. 7(c) shows the magnified graph (a) during 0.6 and 1.1 s. 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.

Moreover, we iterated the same experiment 50 times, and measured the first detected time of the slippage. As a result, averages and standard deviations of the first detected time of the slippage are shown in Fig. 7(d). The sensor network detects the slip at 0.73 s whereas the vision sensor detects the slip at 0.89 s.

D. 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

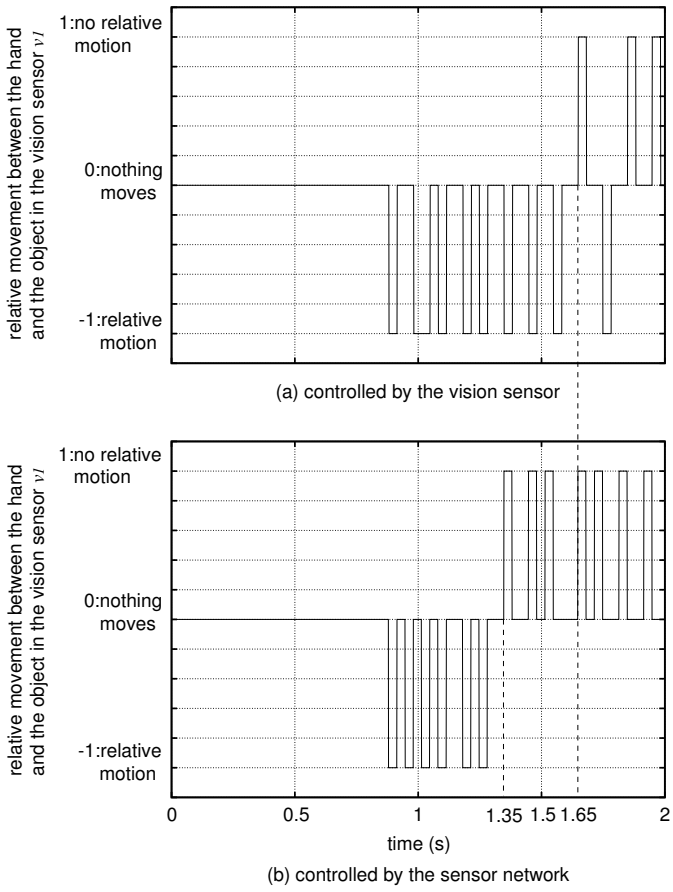


Fig. 8. Observed macro slips in the vision sensor of pick-up experiments, (a) by utilizing only the vision sensor and (b) by utilizing the proposed network

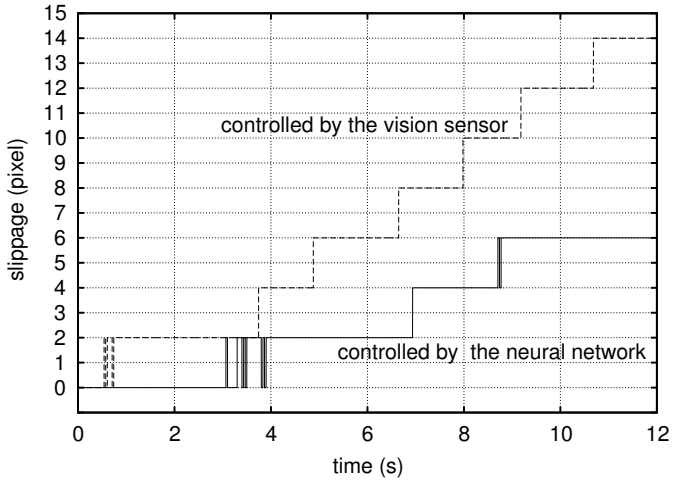


Fig. 9. The experimenter poured water into the cup that grasped by the fingers. The amount of slip is smaller when it is controlled by the proposed network than when it is controlled by the vision sensor.

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 Fig. 8, we show movements of the object in the vision sensor (a) by utilizing only vision sensor, and (b) by utilizing the proposed network. 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 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. Fig. 9 shows the distance of the slip in the vision sensor. In the figure, we compare two cases: with the proposed network and without the network but only utilizing the vision sensor. We can conclude that the slippage is reduced if we utilize the learned network whereas the slippage is larger if we only use the vision information to detect the slip.

IV. 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 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 this sense, the information provided by the vision sensor is also too poor, whether there is relative movement or not and its direction, up or down. To deal with more complicated tasks, we should discuss further about what kind of information should be processed from the vision sensor. If the robot achieves the complex tasks, the robot may need the suitable visual information for learning the neural network. Additionally, 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.

Acknowledgment

The authors would like to thank their colleague, Dr. Minoru Asada, for valuable discussions and comments. 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. Bicci and V. Kumar, "Robotic grasping and contact: A review," In Proceedings of the 2000 IEEE International Conference 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 Proceedings of the 1994 IEEE International Conference 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 Proceedings of the 1997 IEEE International Conference on Robotics and Automation, pp. 3080–3086, 1997.

[4] D. J. O'brien and D. M. Lane, "Force and slip sensing for a dextrous underwater gripper," In Proceedings of the 1998 IEEE International Conference 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 Proceedings of 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 844–850, 1999.

[6] I. Fujimoto, Y. Yamada, T. Maeno, T. Morizono, and Y. Umetani, "Development of artificial finger skin to detect incipient slip for realization of static friction sensation," In Proceedings of the IEEE Conference 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 Proceedings of the 1994 IEEE International Conference 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 Proceedings of the 2000 IEEE International Conference 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," Journal of Robotics and Mechatronics, vol. 14, no. 2, pp. 140–146, 2002.

[10] M. R. Tremblay and M. R. Cutkosky, "Estimating friction using incipient slip sensing during a manipulation task," In Proceedings of the 1993 IEEE International Conference on Robotics and Automation, pp. 429–434, 1993.

[11] H. Shinoda, M. Uehara, and S. Ando, "A tactile sensor using three-dimensional structure," In Proceedings of the 1993 IEEE International Conference on Robotics a nd Automation, pp. 435–441, 1993.

[12] E. G. M. Holweg, H. Hoeve, W. Jongkind, L. Marconi, C. Melchiorri, and C. Bonivento, "Slip detection by tactile sensors: Algorithms and experimental results," In Proceedings of the 1996 IEEE International Conference on Robotics and Automation, pp. 3234–3239, 1996.

[13] C. Melchiorri, "Slip detection and control using tactile and force sensors," IEEE/ASME Transactions on Mechatronics, vol. 5, no. 3, pp. 235–243, 2000.

[14] A. Sano, H. Fujimoto, K. Nishi, and H. Miyanishi, "Multi-fingered hand system for telepresence based on tactile information," In Proceedings of the 2004 IEEE International Conference on Robotics and Automation, pp. 1676–1681, 2004.

[15] N. Tsujiuchi, T. Koizumi, A. Ito, H. Oshima, Y. Nojiri, Y. Tsuchiya, and S. Kurogi, "Slip detection with distributed-type tactile sensor," In Proceedings of 2004 IEEE/RSJ International Conference 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 Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 2331–2336, 2005.

[17] R. Pfeifer and C. Sheier, "Understanding Intelligence," MIT Press, 1999.

[18] K. Hosoda, "Robot Finger Design for Developmental Tactile Interaction - Anthropomorphic Robotic Soft Fingertip with Randomly Distributed Receptors," Embodied Artificial Intelligence, F. Iida et al. Eds., Springer-Verlag, pp. 219–230, 2004.