Internal Representation of Slip for a Soft Finger with Vision and Tactile Sensors

Koh Hosoda, Yasunori Tada, and Minoru Asada

Department of Adaptive Machine Systems, Osaka University
hosoda@ams.eng.osaka-u.ac.jp, tada@er.ams.eng.osaka-u.ac.jp, asada@ams.eng.osaka-u.ac.ajp

Abstract

To build an adaptive autonomous robot, it must have a certain number of external sensors to observe the environment. One physical phenomenon is observed as sensor signal flows through these sensors. In this paper, we focus on a "slip" phenomenon and try to build up a network representation of slip of an anthropomorphic robot hand. A robot hand with distributed tactile sensors and a vision sensor is built to demonstrate how it acquires the representation of "slip". At the beginning of leaning, only the vision sensor can sense the slip as the movement of target, but after a while the tactile sensors can sense the slip even it is so small that the vision sensor cannot sense.

1 Introduction

To build an autonomous robot that can act in an unknown dynamic environment, it is extremely necessary for the robot to have a certain number of external sensors to observe the environment. For example, a robot hand must have force, tactile, and vision sensors to achieve robust manipulation of the object.

Existing control schemes for robot hands usually need calibration, because it is common for a robot to accomplish a task defined in a Cartesian coordinate frame [1, 2, 3]. The frame is usually defined by a human designer, which is different from that of robot's own sensors. Therefore, the sensors must be carefully calibrated with respect to the Cartesian frame so that the robot can observe its own performance in it. Because of the procedure, task performance is prone to be affected by calibration errors and disturbances.

Also, the finger must be solid in these work otherwise it is extremely severe to calibrate a finger tip with respect to the Cartesian coordinate frame. Soft finger tip, however, plays a great role to achieve stable grasping. Since these facts, Cartesian way of calibration is not effective to achieve adaptive and stable grasping.

Let us consider human grasping and manipulation. A human has several sensor modalities such as vision, force, and tactile. Although he/she may not recognize his/her Cartesian coordinate frame, he/she can realize dexterous manipulation. The reason why he/she can still do it may be that grasping and manipulation are defined not in the Cartesian coordinate frame, but in his/her own sensor spaces [4, 5].

From the view of the robot designer, he/she can understand one physical phenomenon and can extract it to each sensor modality. As a result, every sensor must be calibrated with respect to the robot designer's view, that is, the Cartesian coordinate frame. However, an autonomous agent can only *understand* the phenomenon as sensor signals through several modalities. To make a representation of a physical phenomenon inside the agent, it must correct certain amount of sensor signals and find out the consistency between the signals. The consistency will be a *representation* of the phenomenon.

Although several studies have been devoted to make the tactile sensors, they try to make well-calibrated sensors [6, 7, 8]. The only study on uncalibrated tactile sensors is [9] to the best of our knowledge. However, in the paper, they discussed only on tactile sensors, and not on the other sensor modality. As a result, they did not intend to make internal representation of a physical phenomenon combining several sensor modalities.

In this paper, we build a robot hand with anthropomorphic finger tips, and try to make a representation of "slip" inside the agent. The finger tip is made of silicon gum and has several strain gauges inside. Since the vision and tactile sensors are totally uncalibrated, the robot have to correct data and find out the relation between them. We proposed a network to represent the relation, and this will be the most primitive representation of "slip" inside the agent. At the beginning of leaning, vision sensor can only

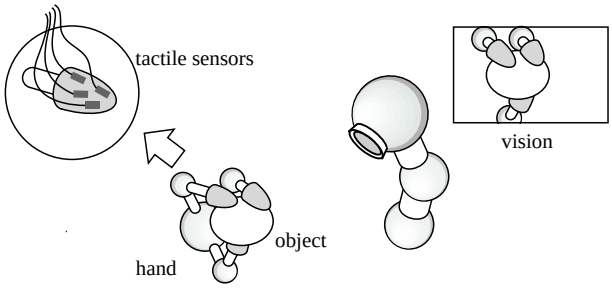

Figure 1: *A robot hand with vision and tactile sensors*

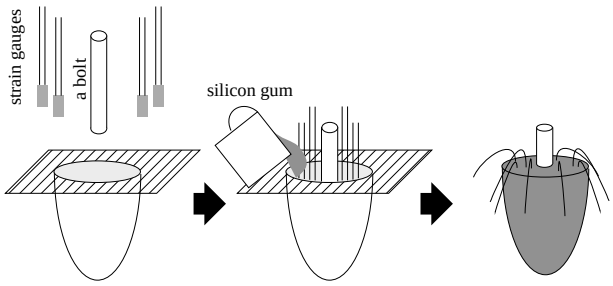

Figure 2: *The finger tip is made of silicon gum and has a bolt at the center, and several strain gauges inside.*

sense the slip as the movement of target, but after a while the tactile sensors can sense the slip even it is so small that the vision sensor cannot sense.

2 Robot hand with vision and tactile sensors

A human has several sensor modalities such as vision, force, and tactile. Although these sensors are not calibrated precisely, he/she can sense the *slip* from different modality, and can realize dexterous manipulation. How the representation looks like? We try to understand the mechanism by building a finger system with uncalibrated tactile and vision sensors (figure 1).

First, we focus on the finger tip. Softness of the finger plays a great role for stable grasping, therefore the finger is made of silicon gum. Several strain gauges are put inside. The figure 2 shows the procedure how the finger tip is made. The reason why we call it an anthropomorphic finger is that it is soft and the sensors are distributed randomly like that of human. The real anthropomorphic finger tip is shown in figure 3.

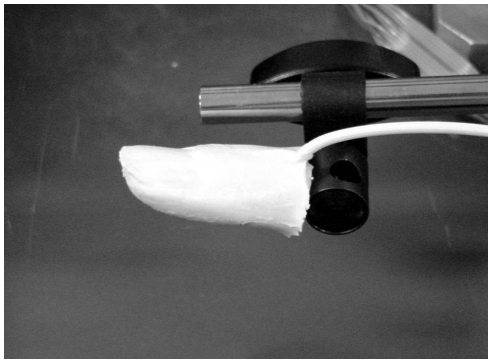

Figure 3: *An anthropomorphic finger is made of silicon gum and has several strain gauges inside.*

3 Representation of slip in sensor spaces

From the view of the robot designer, he/she can understand one physical phenomenon and can extract it to each sensor modality. In this case, however, every sensor must be calibrated with respect to the robot designer's view, that is, the Cartesian coordinate frame. To calibrate all the sensors is tedious and the robot will not be robust against modeling error, noise, and disturbance. Since the anthropomorphic finger is soft and several strain gauges are placed randomly, it is extremely hard to calibrate them with respect to the Cartesian frame. One of the main reasons is that model of the physical phenomenon is the one of the designer, not the one of the robot itself.

Since an autonomous agent cannot *understand* the physical phenomenon, all it can get are only sensor signal flows from different sensors. At the beginning of learning, it can only catch the flows and cannot see any constancy between them. However, after a while, it corrects a certain amounts of sensor signals, and it may be able to find a certain constancy between them that may be caused from one physical phenomenon. The consistency will be a representation of the phenomenon.

In this paper, we focus on "slip." The hand system has a vision sensor and the several tactile sensors. All sensors are not calibrated with a certain coordinate frame. During the slip phenomenon, the system corrects sensory data and finds the consistency between them.

4 Network to acquire the representation of "slip"

We propose to use a simple Hebbian network to get the consistency (figure 4). There are two layers: a tactile sensor layer and a vision sensor layer. Signals

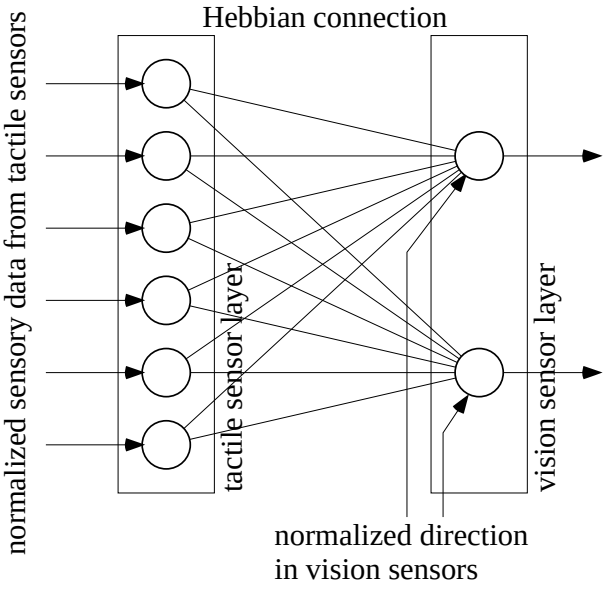

Figure 4: *A Hebbian network to find out the consistency between tactile and vision*

of each tactile sensor are normalized by the maximum value over time and are given to each node as activation. Two neurons in the vision layer are activated by displacement of the image target in the image plane along x and y–axes, respectively.

Since the vision sensor and the tactile sensors are not calibrated, there is no *a priori* knowledge on the relation between them. Therefore, the weights between the nodes are 0 initially. Imagine that the finger is contacting with an object. The vision sensor can observe the object and the finger tip, therefore, when the finger slips, it is observed by the sensor as the difference of displacements of the object and the finger tip in the image plane. Simultaneously, certain amount of strain information can be obtained by the tactile sensors. If the direction of one tactile sensor (in a 3D coordinate frame, say a finger tip frame) happens to be along the slip direction in the image plane, the connection between the nodes is strengthened according to the Hebbian rule. Over time, the connection between the vision node and the tactile node that has the corresponding directions has certain amount according to their cosine.

The consistency existing in the vision and tactile sensors is, therefore, represented as a weight set of the network, not in the symbolic way. Although the representation is difficult for the robot designer to understand, it may be a natural expression and easy to access for the agent.

At the beginning of learning, as discussed above, slip is mainly observed by the vision sensor. After some

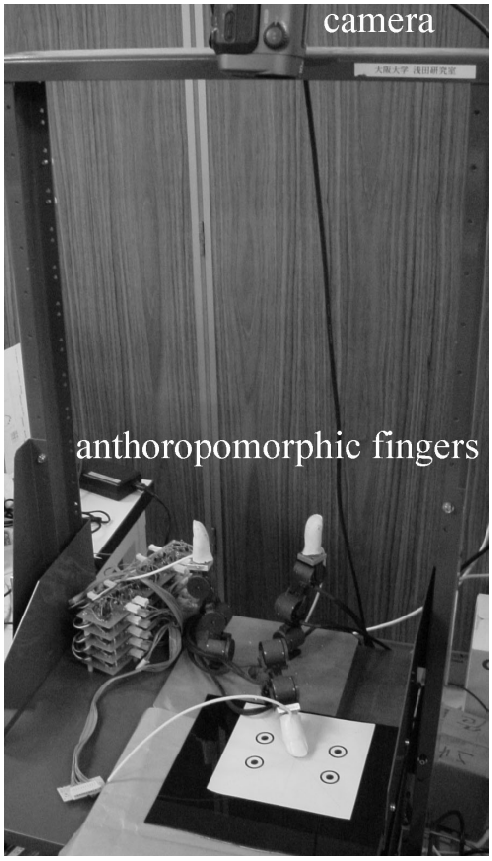

Figure 5: *A robot hand system with a camera: Every finger is equipped with an anthropomorphic finger-tip made of silicon gum. A camera is placed above the hand.*

trials, the direction of slip can be also sensed by the tactile sensors. This provides the system redundancy [10]. That is, even if the vision sensor cannot catch the slip information because of some reasons, for example, because of occlusion, the network still can detect the slip direction.

Even more interesting thing concerns the sensor resolution. Normally, as the device to observe displacement, a tactile sensor is more sensitive than a vision sensor. As a result, the tactile sensor is expected to observe the slip earlier than the vision sensor.

5 Experiment

5.1 Experimental equipment

To show how the network learns through experiences, we build a experimental hand system with anthropomorphic finger tips and a vision camera.

In the experiment, we put six strain gauges in each finger tip. The strain is measured as resistor and amplified by the sensor amplifier (you can see the

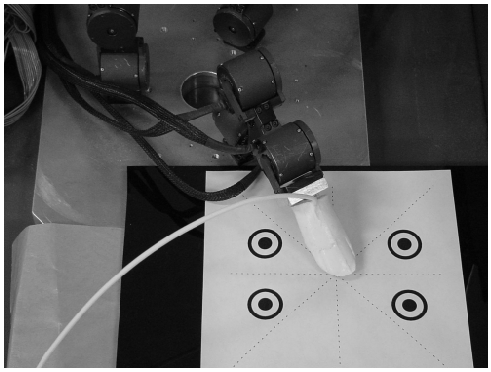

Figure 6: *A finger exerts force against a board. The board is moved by the human operator and the finger senses by vision and tactile sensors.*

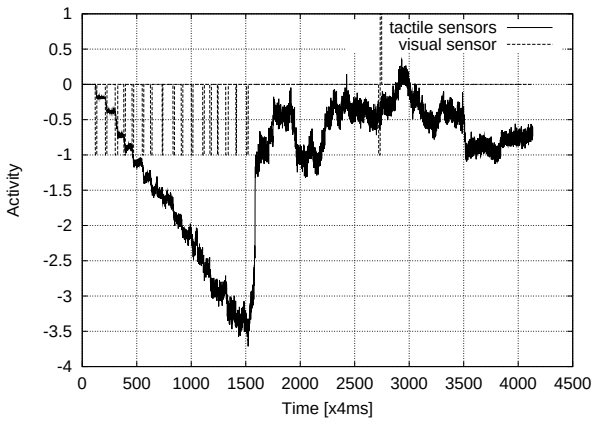

Figure 7: *Experimental result 1 : the input of the vision node 1 corresponding to the x-direction in the image plane from tactile sensors, and the activation by the vision sensor, before learning*

amplifier in the figure next to the hand) and fed to the host computer via an A/D board.

Video signals from a CCD camera are sent via a tracking module equipped with a high-speed correlation processor by Fujitsu (image size : 512[pixel] $\times$ 512[pixel]). We specify a certain region in the image (called a template) to be tracked before starting an experiment. During the experiments the module feeds coordinates, where the correlation measure (it uses a SAD measure, Sum of Absolute Difference) is the smallest with respect to the template, to the host computer.

We put several target marks on a board and it is moved by a human operator. The displacements of these targets are observed by the camera. The finger exerts a certain force on the board, and the information from strain gauges enlarge the corresponding weight of the Hebbian network.

5.2 Experimental results

In figure 7, before learning, the input of the vision node 1 corresponding to the x-direction in the image plane, and the activation by the vision sensor are shown. Since the network is not trained, there is no consistency between these activations.

After 260 learning trials, what the network obtained is shown in figure 8. We can see consistency between the activation by the tactile sensors with that of the vision sensor. From this results, we can say that now the tactile and vision sensors are redundant to sense the "slip." We can also see that at first the activation by the tactile sensors gets larger, and then the vision sensor is activated. The reason may be that the tactile sensors are more sensitive than the vision sensors.

6 Discussion and future work

In this paper, we have built a robot hand with anthropomorphic finger tips, and try to make a representation of "slip" inside the agent. Since the vision and tactile sensors are totally uncalibrated, the robot have to correct data and find out the consistency between them. We have proposed a network to represent the consistency, and suggested that this will be the most primitive representation of "slip" inside the agent.

An intelligent human can understand physical mechanism of slip, and may be able to extract it to each sensor modality. However, this representation is not of the robot itself but of the human, that is, slip is not grounded on robot's own sensor/motor apparatus. Therefore, this way of extracting needs calibration of all sensors with respect to a Cartesian coordinate frame, and as a result, the system loses adaptability. On the other hand, the physical phenomenon is observed by several sensor modalities, therefore, the consistency between sensors can be regarded as the most primitive representation of the phenomenon. After the robot obtain such consistency existing in sensor data, it may be able to "understand" the dynamics underlying the physical phenomenon. In this sense, the obtained weights can be the most primitive representation of "slip" with vision and tactile sensors.

From the sum of activations of the tactile nodes, we cannot distinguish slip from just pressured. This is because the "slip" is only defined the displacement in the image plane in this paper. Basically, our understanding of slip is more complicated. To acquire such complicated concept, the robot must have more

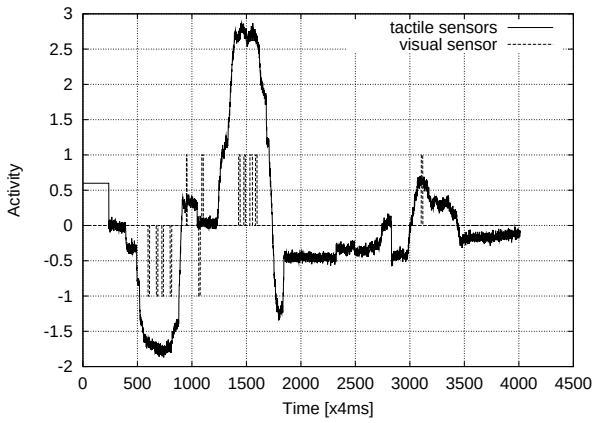

Figure 8: *Experimental result 2 : the input of the vision node 1 corresponding to the x-direction in the image plane from tactile sensors, and the activation by the vision sensor, after 260 learning*

sensors, and have to find consistency between them.

The "slip" representation in this paper is discussed only from the viewpoint of sensing. Actually, the task context is also very important to obtain the representation. In this sense, the representation must be acquired in the context of task, that is, under a certain sensory-motor coordination.

References

[1] P. K. Allen, A. T. Miller, P. Y. Oh, and B. S. Leibowitz. Using tactile and visual sensing with a robotic hand. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, pages 676–681, 1997.

[2] T. Matsuoka, T. Hasegawa, and K. Honda. A dexterous manipulation system with error detection and recovery by a multi-fingered robotic hand. In *Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 1, pages 418–423, 1999.

[3] Y. Yokokohji, M. Sakamoto, and T. Yoshikawa. Vision-aided object manipulation by a multifingered hand with soft fingertips. In *Proceedings of the 1999 IEEE International Conference on Robotics and Automation*, pages 3201–3208, 1999.

[4] K. Hosoda, K. Igarashi, and M. Asada. Adaptive hybrid control for visual servoing and force servoing in an unknown environment. *IEEE Robotics and Automation Magazine*, 5(4):39–43, 1998.

[5] Koh Hosoda, Takuya Hisano, and Minoru Asada. Sensor dependent task definition: Object manipulation by fingers with uncalibrated vision. In *Proceedings of Intelligent Autonomus Systems 6(IAS-6)*, pages 843–850, 2000.

[6] M. Shimojo, M. Ishikawa, and K. Kanayama. A flexible high resolution tactile imager with video signal output. In *Proceedings fo 1991 IEEE International Conference on Robotics and Automation*, pages 384–391, 1991.

[7] Ryosuke Kageyama, Satoshi Kagami, Msayuki Inaba, and Hirochika Inoue. Development of soft and distributed tactile sensors and the application to a humanoid robot. In *Proceedings of the 1999 IEEE International Conference on Systems, Man, and Cybernetics*, pages 981–986, 1999.

[8] Daisuke Yamada, Takashi Maeno, and Yoji Yamada. Artificial finger skin having ridges and distributed tactile sensors used for grasp force control. In *Proc. of International Conference on Intelligent Robots and Systems (IROS2001)*, pages 686–691, 2001.

[9] Mitsuhiro Hakozaki, Katsuhiko Nakamura, and Hiroyuki Shinoda. Telemetric artificial skin for soft robot. In *Proceedings of TRANSDUCERS '99*, pages 844–847, 1999.

[10] Rolf Pfeifer and Christian Scheier. *Understanding Intelligence*. The MIT Press, 1999.