Visuo-Tactile Face Representation through Self-induced Motor Activity

Sawa Fuke, Masaki Ogino, Minoru Asada
JST ERATO, Osaka university
2-1, Yamadaoka, Suita, Osaka 565-0871, Japan
sawa.fuke@ams.eng.osaka-u.ac.jp, [ogino, asada]@jeap.org

Abstract

This paper presents a learning model that enables a robot to acquire body representation for its face which is invisible to itself while the robot touches it. Two processes are simultaneously carried out: self-organization of tactile sensor unit configuration though the visuo-tactile sensations by self-induced motor behaviors, and facial part detection based on the discontinuity of sensor values.

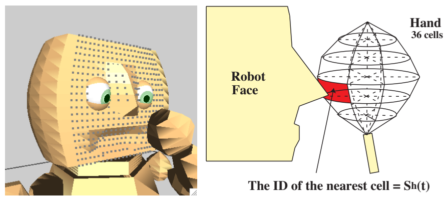
1. Introduction

Robots are expected to learn the flexible body representation in order to adapt themselves to changes in environment and tasks such as tool use and imitation of others' behaviors. In cognitive developmental robotics, the integration of multi modal (visual, tactile and somatic) sensations through double touch experience have been investigated (Nabeshima et al., 2006), (Yoshikawa et al., 2002). However, the existing work lacks two important aspects for face representation. First, they cannot cope with invisible part representation. To enable it, a robot should have an ability to apply the visual experiences in the visible area to the invisible ones. Second, they have not categorized facial parts since they supposed smooth flat body surface.

In this paper, we propose a model that enables a robot to acquire a body representation for its face invisible to itself while touching it. Two processes are simultaneously carried out: self-organization of tactile sensor unit configuration though the visuo-tactile sensations by self-induced motor behaviors, and facial part detection based on the discontinuity in sensations.

2. The proposed model

(Yoshikawa et al., 2002) proposed a method in which a robot can acquire two mappings between the proprioceptive space (joint angles) and vision space, and between the tactile space and vision space by touching its own body parts with its hand and associating the coordinates of the touched part in the camera image with the activated tactile sensor units and the joint angles. However, this method cannot be applied to touched parts that are not visible such as the face and back. In such a case, it is necessary to construct the integrated spatial perception before the association so as to estimate the invisible hand positions. The spatial perception based on motor information is inevitable to construct the body representation. In fact, it is important to associate the invisible hand positions with the visible ones for the robot to explore its invisible parts with its hand.



(a) the front view (b) the robot hand

Figure 1: A new model with the convexties on the surface of the face and a hand of which surface is divided into 36 cells. The ID of the cell that is the nearest to the surface of the face at time can be detected.

In this method, the robot model that is shown in Figure 1 is used. First, Jacobian transformation f from the displacement of joint angles to that of the coordinates is learned by the back-propagation algorithm during the hand motion in the visual field. Then, the robot touches its own face (invisible parts) and estimates the position of the hand $\hat{\boldsymbol{r}}$ by using learned function f and the displacement of joint angles at that time.

As shown in Figure 2, the robot has vision, tactile, and proprioceptive spaces in advance. Each cell of the tactile space is the actual tactile sensor unit on the surface of the face. In the next phase, the mapping is modelled as the learning of the reference vector of the unit in the tactile space,

$$\boldsymbol{r}_i^T = (x_i^T, y_i^T). \qquad (1)$$

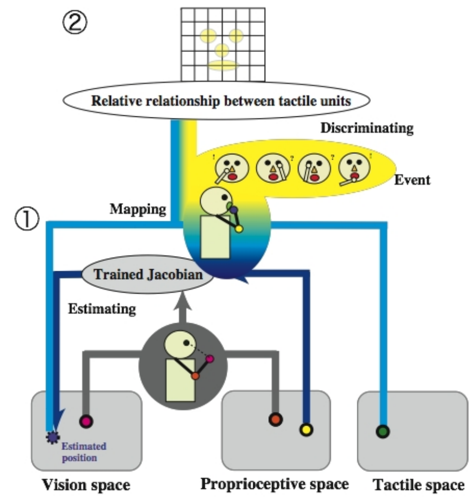


Figure 2: The idea of the proposed model

This reference vector $\mathbf{r}_i^T$ of the i-th unit is updated in the same way as the self organizing map algorithm (Kohonen, 1995). When the current estimated hand position is $\hat{\mathbf{r}}$, the mapping for the i-th unit on the tactile space is updated depending on the distance from the c-th unit on the tactile space,

$$\Delta \boldsymbol{r}_i^T = \alpha_T(t) \exp(-\|\boldsymbol{T}_i - \boldsymbol{T}_c\|/\gamma)(\hat{\boldsymbol{r}} - \boldsymbol{r}_i^T)), \quad (2)$$

where $\alpha_T(t)$ is a learning rate that decays based on the exponential function, $\exp(-(count)/p)$. Here, "*count*" means the learning step and p is a constant. In addition, $\boldsymbol{T}_i$ and $\boldsymbol{T}_c$ are the coordinates of the i-th and the c-th units in the tactile space, respectively. (① in Figure 2).

Furthermore, in the mapping phase, the robot can discriminate the characteristic tactile sensor units in parallel. We assume that the robot can discriminate some tactile units from others by detecting the discontinuity of the sensory information; the detection of the change of the contact cell of the hand by hitting the convexity, the sudden obstruction by hand in right or left eye, and the feeling of the mouth movement.

When certain discontinuity is detected at the i-th unit in the type of the sensor k, the discontinuity level u is updated by the following equation,

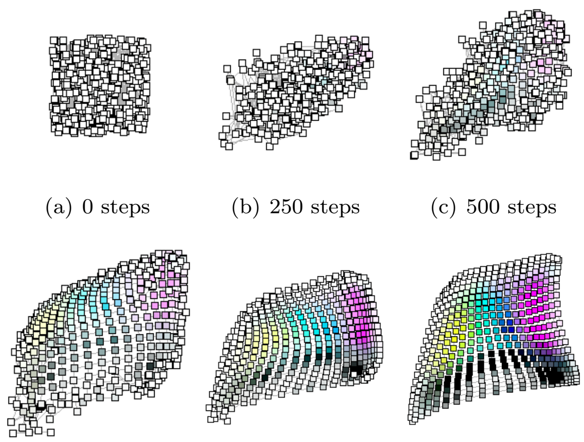
$$\Delta u_i^k(t) = \alpha_p \exp(-\|\boldsymbol{T}_i - \boldsymbol{T}_{c_p}\|/\gamma_p), \quad (3)$$

where c_p is the sensor unit whose mapped position is the nearest to the estimated hand position $\hat{\boldsymbol{r}}$. α_p is a constant in this experiment. The discriminated units can be categorized into the clusters based on the type of the sensory data. (② in Figure 2).

3. Experimental results

The Jacobian function, f, that associates the displacement of the joint angles and that of the hand position in the camera coordinate system is learned by a neural network while the robot draws a circle with the hand (the end effector of the arm) in both clockwise and anticlockwise directions during the training phase. The learning rate is 0.2 and the number of learning steps is 200.

Then, during the mapping phase, the robot touches a random positions on its face with its hand, the vision and the tactile spaces are associated and the discontinuity level are updated at the same time. The learning result of the spatial configuration of the tactile sensors in the visual space is shown in Figure 3. As stated above, there is no difference among all sensor units at the beginning. In addition, the initial mapping positions of sensor units are random and the total number of the learning steps is 5000. In Figure 3, the discontinuity level of the i-th unit is expressed by a color of each unit. The result shows that activation of tactile units on facial parts turns up relative to the other units as learning proceeds by focusing on the change of multi-modal sensor values. Moreover, the hand movement of hitting the convexities seemed to have little effect to acquire the relative relationship among the units while touching its face.



(a) 0 steps (b) 250 steps (c) 500 steps

(d) 1000 steps (e) 2000 steps (f) 5000 steps

Figure 3: The acquired face representation

References

Kohonen, T. (1995). Self-organizing maps. *Springer-Verlag Verlin Heidelverg.*

Nabeshima, C., Kuniyoshi, Y., and Lungarella, M. (2006). Body schema adaptation for robotic tool use. *Advanced Robotics*, 20,10:1105–1126.

Yoshikawa, Y., Kawanishi, H., and Asada, M. (2002). Body scheme acquisition by cross modal map learning among tactile, visual and proprioceptive spaces. *Proc. of the 2nd International Workshop on Epigenetic Robotics 2002*, pages 181–184.