Body Scheme Acquisition by Cross Modal Map Learning among Tactile, Visual, and Proprioceptive Spaces

Yuichiro Yoshikawa **Hiroyoshi Kawanishi** **Minoru Asada** **Koh Hosoda**
Dept. of Adaptive Machine Systems,
Graduate School of Engineering,
Osaka University
2-1 Yamada-Oka, Suita, Osaka 565-0871, Japan
{yoshikawa, kawa, asada, hosoda}@er.ams.eng.osaka-u.ac.jp

Abstract

How to represent own body is one of the most interesting issues in cognitive developmental robotics which aims to understand the cognitive developmental processes that an intelligent robot would require and how to realize them in a physical entity. This paper presents a cognitive model how the robot acquires its own body representation, that is *body scheme* for the body surface. The *internal observer* assumption makes it difficult for a robot to associate the sensory information from different modalities because of the lacking of references between them that are usually given by the designer in the prenatal stage of the robot. Our model is based on cross-modal map learning among joint, vision, and tactile sensor spaces by associating different pairs of sensor values when they are activated simultaneously. We show a preliminary experiment, and then discuss how our model can explain the reported phenomenon on *body scheme* and future issues.

1. Introduction

We, human beings, have a capability to perform various kinds of complicated tasks by our hands, sometimes with tools. In order to acquire such a capability, we should have known the relationship between our body parts and the external space. Representation of body is called *body scheme* or *body image* (Ramachandran and Blakeslee, 1998), which is supposed to be described in the egocentric reference frame. Although recent studies have revealed that body scheme is not simply a representation of joint angles, but a complex integration of vision, proprioception, touch, and motor feedback (Iriki et al., 1996, Ishibashi et al., 2000, Graziano et al., 2000), little is known about how different sensory modalities are associated in order to construct the body scheme.

On the other hand, it seems a promising way of modeling such cognitive process by building a robot which can acquire a body scheme (Asada et al., 2001), not simply because it is expected to reveal a new way of understanding own body scheme representation but also because the concept of body scheme is also important in robotics. For example, in manipulating objects or avoiding obstacles, it needs to know the relationship between its body and objects in the environment. Although the designer can specify some kinds of body scheme by associating the sensor values with its reference (usually called calibration process), the robot should have a capability to acquire its body scheme by itself in order to adapt itself to accidental changes in its body and/or in its environment. If it is an *internal observer* who can use only its resultant perception of its actuation, it is an interesting but formidable issue to acquire body scheme because it needs to find the relationship between the sensory values and its references. This is a *reference* problem or so-called *internal observer problem.* As the first step to attack this problem, we model a mechanism by which a robot can find its body surface representations. Hereafter, the robot is a learner to obtain its body scheme.

In this paper, we propose a cognitive developmental model for a robot to acquire its body scheme for its body surface. The body scheme consists of a cross modal map among tactile, vision, and proprioceptive sensor spaces and is acquired by learning their association from its experiences of self-touching. Then, we show a preliminary experiment to implement a simplified model. Finally, we discuss how the proposed model can explain the behavioral phenomena related to the body scheme reported in psychophysiology.

2. Cross Modal Map

Suppose that the learner has a capability of actuation and the following sensor modalities (see Fig. 1), such as (a) tactile sensors $T_i (i = 1, \cdots, n_t)$ which are distributed on the learner's body surface, and each of which outputs t_i as ON when it senses pressure, or OFF else, (b) visual sensors $X_i (i = 1, \cdots, n_x)$ which are assigned to visual patterns I_i in the stereo views and output the image coordinates $\boldsymbol{x}_i$ of them (ex. the end-effector, the elbow, chest, and so on) when they are observable, and (c) proprioceptive sensor Θ which describes the posture with its joint angles.

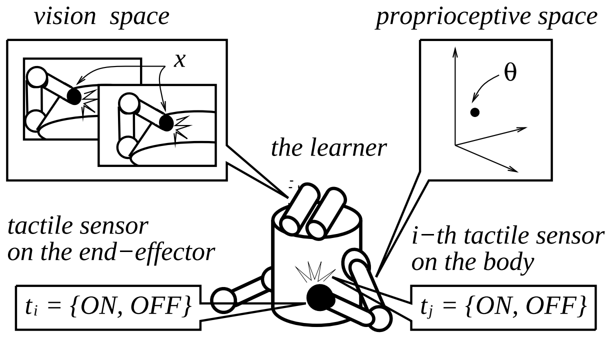


Figure 1: Sensing modalities.

The learner cannot find representations to what extent its body occupies from observing only single modality, even if it is tactile sensors which are distributed on its body surface, because of a reference problem. Then, we consider that it is acquired by finding the receptive field in the different sensor modality corresponding to the tactile one. That is, it can be acquired by a cross modal mapping among different modalities. In principle, however, it is impossible to straightforwardly associate the spatial region of one tactile sensor on the body surface with its view because of no knowledge of the spatial distribution of tactile sensors, kinematic parameters, or camera ones (uncalibrated or uninterpreted).

The first moment we can perceive our own body is when we touch our body by ourselves. At that moment, a pair of tactile sensors are activated (touching and touched), and at the same time the corresponding observable body parts coincide with each other in the visual space.

Therefore, when the robot touches a part on its body surface (T_j) by its another part (T_i), following equations hold,

$$t_i = t_j = 1, \tag{1}$$

$$\boldsymbol{x}_k = \boldsymbol{x}_l, \tag{2}$$

where I_k and I_l are the visual patterns which coincide with each other in the visual space. According to these equations, the learner can associate a pair of the visual sensors (X_k and X_l) with a pair of the tactile sensors (T_i and T_j). However, the robot cannot determine cross modal pairs: ((X_k,T_i) and (X_l,T_j)) or ((X_k,T_j) and (X_l,T_i)). Then, we propose a cognitive model to construct a cross modal map without distinguishing them, in which it associates two pairs ((X_k and X_l) and (T_i and T_j)) and Θ based on the correlation shown in eqs.(1) and (2) when self-touching. After learning the cross modal map, the robot can estimate the receptive areas corresponding to a pair of the tactile sensors in the visual and somatosensory modalities when the two tactile sensors coincide with each other.

From this idea, we provide a cognitive model with *similarity units*($S_{t_{ij}}$, $S_{x_{kl}}$, and S_{θ_m}) which detect simultaneousness of sensors in the same modality (see Fig. 2). Here,

1. unit $S_{t_{ij}}$ judges whether both two tactile sensors (T_i and T_j) outputs ONs,

2. unit $S_{x_{kl}}$ judges the closeness of the image coordinates which are the outputs of two visual sensors (X_k and X_l), and

3. the last one S_{θ_m} is activated when its somatosensory sensor value corresponds to the m-th quantized vector $\boldsymbol{\theta}_m$ that is one of the n_s quantized segments in the somatosensory space.

By associating similarity units which are simultaneously activated based on Hebbian rules, it can find receptive fields of tactile sensors when they are touched each other.

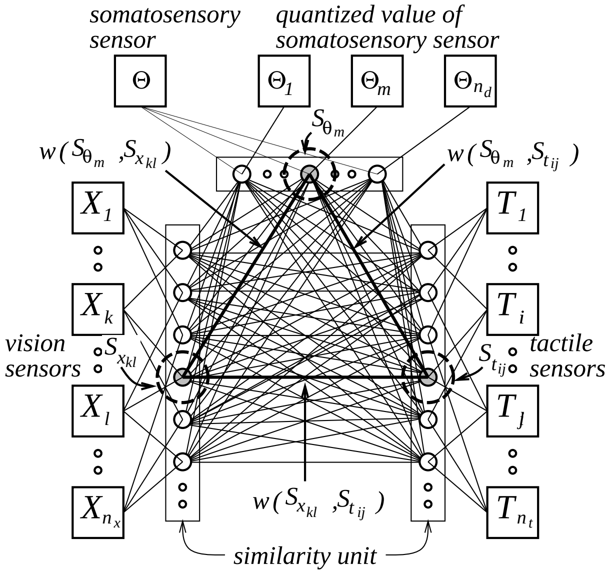


Figure 2: Architecture for cross modal map learning.

3. Preliminary Experiment

We are implementing the process of cross modal mapping with a robot which has a pair of stereo cameras, touch sensors on its body surfaces including the end-effectors of its arms (Fig. 3). There is a obstacle (white box) besides it. At first, the learner has the tactile sensors on its end-effector, forearm, and the upper arm, and has the visual recognition modules which detect the image coordinates of the visual patterns, namely the end-effector, the forearm, the upper arm, and the obstacle. However, it does not know which patterns correspond to its body parts.

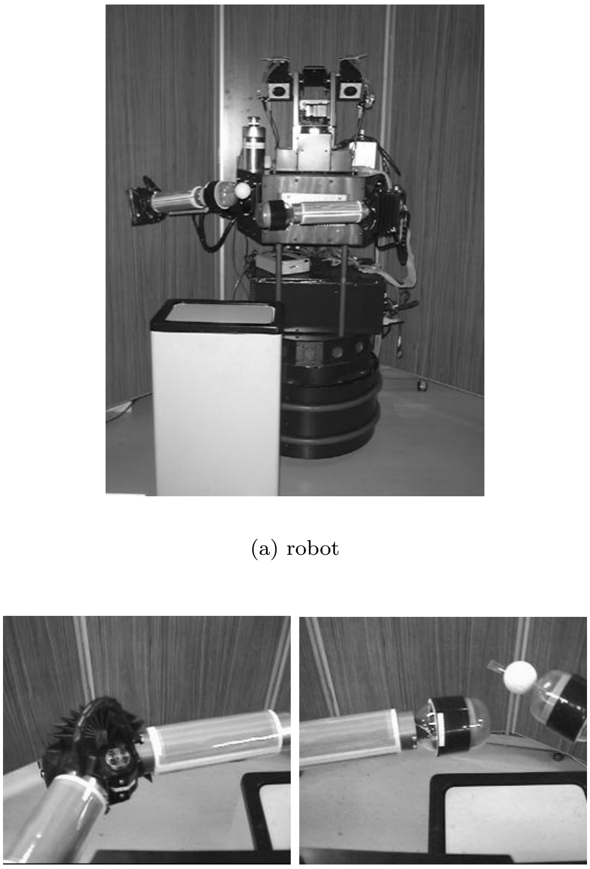
In oredr to learn a cross modal map of the body, we let the learner touch its forearm, the upper arm, and the obstacle with its end-effector. Based on the activation of the similarity units in the experiences, the cross modal map is learned by Hebbian rule. Fig. 4 shows the transitions of the synaptic weights between similarity units. We can see those of incorrelative units are shrinked while correlative ones remain positive. Threrefore, the learner can find receptive fields of the tactile sensors in the visual space by the proposed coginitive model.

4. Discussion

In this section, we discuss how the proposed model can explain the behavioral phenomena related to body scheme reported in psychophysiology. It is said that when one (person A) touches other person's nose (person B) with A's finger, without seeing it, concurrently B touches A's nose in the same way, both A and B have introspection as if their noses extended to opponent real nose position where both feel tactile sense (Ramachandran and Blakeslee, 1998).

Our model can explain this phenomenon if we consider the activation of the similarity units of a cross modal map as the element of introspection. If the touching by A and B are synchronized, the similarity unit between tactile sensors on their noses and fingers are activated, as well as one in the somatosensory modality. But, no units in the visual modality is activated by real sensors since they do not open their eyes. However, the similarity unit in the visual modality can be activated by the propagated activations through the connections of the cross modal map. As a result, both have such introspection because these activation of the similarity units are consistent with each other from the viewpoint of the cross modal map behavior. That is, such introspection seems to be caused by the fact that the similarity unit of nose in the tactile moodily is referred by one of current posture in the somatosensory modality unless one in the visual modality is inhibited by opening its eyes.

The proposed model should be modified by considering following issues.



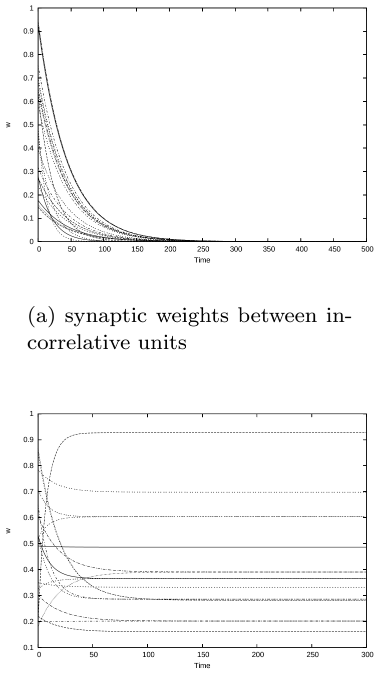
(a) robot

(b) left view

(c) right view

Figure 3: Test-bed robot and its views .

- How the visual recognition modules come from? Althogh we assume the learner already has had them, they should be acquired from his/her experiences.
- How asynchronized activations of the different modalities can be associated? Although we assume the centralizated clock, different modalities have different clocks and different delays to transmit from sensors to information processing parts.
- Since it is the static representation of its posture, it needs another mechanism when it behaves using the cross modal map. It seems caused by the fact that the cross modal map discards the information of the touching motion. Adding units of actuation and making the cross modal map dynamic are needed.
- How spatial representation of the body comes from? Although our model discards the spatial information in the retina, it can be associated with the motion of the eyes or the neck. As well as, it does not utilize the information in the displacement of the tactile sensors yet.



(a) synaptic weights between incorrelative units

(b) synaptic weights between correlative units

Figure 4: A transition of synaptic weight.

- Our final goal is to make a robot which is an internal observer enable to imitate. It may be realized by finding similarity between the observed information and the acquired body scheme. Although the mechanism to compare them which depend on the viewpoint has not been revealed yet, the cross modal map seems to need a function of associative memorization in order to map the observed motion to self one.

References

Asada, M., MacDorman, K. F., Ishiguro, H., and Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, 37:185–193.

Graziano, M. S. A., Cooke, D. F., and Taylor, C. S. R. (2000). Coding the location of the arm by signt. *Science*, 290(5498):1782–1786.

Iriki, A., Tanaka, M., and Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurons. *Neuroreport*, 7:2325–2330.

Ishibashi, H., Hirata, S., and Iriki, A. (2000). Acquisition and development of monkey tool-use: behavioral and kinematic analyses. *Can. J. Physiol. Pharmacol*, 78:958–966.

Ramachandran, V. S. and Blakeslee, S. (1998). *Phantoms in the Brain: Probing the Mysteries of the Human mind.* William Mollow.