Cross-anchoring for binding tactile and visual sensations via unique association through self-perception

Yuichiro Yoshikawa, Koh Hosoda*, and Minoru Asada*
* HANDAI Frontier Research Center,
Graduate School of Engineering,
Osaka University
Yamada-Oka 2-1, Suita, Osaka Japan
yoshikawa@er.ams.eng.osaka-u.ac.jp, {hosoda, asada}@ams.eng.osaka-u.ac.jp

Abstract

Binding is one of the most fundamental cognitive functions, how to find the correspondence of sensations between different modalities such as vision and touch. Without a priori knowledge on this correspondence, binding is regarded to be a formidable issue for a robot since it often perceives multiple physical phenomena in its different modal sensors, therefore it should correctly match the foci of attention in different modalities that may have multiple correspondences each other. Learning the multimodal representation of the body is supposed to be the first step toward binding since the morphological constraints in self-body-observation would make the binding problem tractable. The multimodal sensations are expected to be constrained in perceiving own body so as to configurate the unique parts in the multiple correspondence reflecting its morphology. In this paper, we propose a method to match the foci of attention in vision and touch through the unique association by cross-anchoring different modalities. Simple experiments show the validity of the proposed method, and future issues are discussed.

1 Introduction

Binding is one of the most fundamental cognitive functions, how to find the correspondence of sensations between different modalities such as vision and touch, both of which are major sources of perception not only for the external world but also for the agent's body itself. The latter is closely related to the body representation which is often given by the designer and fixed but has much influence on the adaptability to the changes in the environment and the robot body itself. Assuming that the designer does not give any explicit knowledge on the body representation, a robot should construct its body representation only from its uninterpreted multimodal sensory data. In this process, *binding* has a significant role.

Recently, researchers in other fields focus on the binding problem, which concerns the capability to integrate information of different attributes [1]. Although there are already some models of binding, for example based on attention [2], firing in synchrony [3, 4], and so on, it has not been still clear how to bind different sensor modalities since they focused on the binding problem between visual attributes. To propose the model for the cross-modal binding mechanism of humans based on a constructivist approach, we should start with an assumption that the designer does not give any *a priori* knowledge on what the robot's sensors receive, but the robot can discriminate the different sensor modalities such as vision and touch. Generally, receptive fields for touch and vision are simultaneously stimulated, but often respond to different physical phenomena since the foci of attention in these modalities are often different. In other words, the robot does not always watch its touching region. Therefore, to bind different modalities, the robot should correctly match the foci of attention in different modalities that may have multiple correspondences each other.

We suppose that learning the multimodal representation of body should be the first step toward binding since the morphological constraints in self-body-observation would make the binding problem tractable. The multimodal sensations are expected to be constrained in perceiving own body so as to configurate the unique parts in the multiple correspondence reflecting its morphology. Therefore, building a robot that can acquire the representation from multimodal sensory data is an interesting issue from a viewpoint of a constructivist approach towards both establishing the design principle for an intelligent robot and understanding the process how humans acquire their body representation [7]. In this study, as an example of the binding problem, we focus how it can learn to watch its body part when it detects the collision on it. The previous work on the issue of acquiring body representation escaped from this kind of problem by assuming that it can observe only matched sensations in different modalities (ex. [5, 6]).

Yoshikawa et al. have proposed the method to learn the multimodal representation of the body surface through *double-touching*, that is touching its body with its own body part [8]. It is based on the fact that double-touching co-occurs with *self-occlusion*, that is the occlusion caused by covering its body with its own body part in its view. Although they did not take multiple self-occlusions caused by the physical volume of the body into account, which makes the binding problem remain formidable, it seems still reasonable to utilize the fact that double-touching co-occurs with self-occlusions. In this paper, we presents a method to match the foci of attention to its own body in vision and touch by virtue of the morphological constraint in the relationship between double-touching and self-occlusion. In the proposed method, the mismatched responses in these modalities can be discarded through the process of *unique association* where corresponding pairs of subsets in different attributes are exclusively connected with each other by what we call *cross-anchoring*.

In the rest of the paper, first what kind of problem is to be solved for *binding* in different modalities is explained. Then, a possible developmental course towards binding and a basic idea utilizing the morphological constraint of the human-like robot's body to perform binding are argued. After introducing an anchoring Hebbian learning rule to perform unique association, we show preliminary computer simulations of the robot which has 1-DoF or 2-DoFs arm and a camera head to test the proposed learning rule works. Finally, discussion with future issues is given.

2 The binding problem in different modalities

In order to propose the model for the binding mechanism of humans based on a constructivist approach, we should start with an assumption that the designer does not give any *a priori* knowledge on the robot body representation, but a robot can discriminate the different sensor modalities such as vision and touch. Therefore, the problem for the robot is how to associate these different sensations to build its body representation needed to accomplish various tasks such as collision avoidance and object manipulation. In this study, as an example of the binding problem, we focus how it can learn to watch its body part in which it detects the collision. So what is the problem?

There is a series of studies on modeling the brain mechanism of binding different visual attributes that are processed in segregated areas, such as form, shape and color [3, 4]. They built a system capable of binding these attributes through the dynamic process called reentry with topographic connections between the segregated groups of neurons and made several suggestions to understand the mechanism of binding in vision. However, since the topographical relationship between different modal sensors depends how they are embedded in the body, we need to release the assumption of *a priori* topographic connection in order to cope with the binding problem between different modalities.

This kind of binding problem has not been focused so far although it is an important issue for acquiring the multimodal representation of the body. In some previous studies on own body representation with both tactile and visual sensors, the designer provided the agent with the competence to detect the position of the touch sensors in its view [6] or assumed that there is only one object which can collide with its body [5]. In other words, the binding problem is solved by the designer instead of the robot itself by assuming that it can observe only matched sensations in different modalities. However, there are often multiple visual responses to both the body and nonbody that co-occur with the tactile one on the body since the agent watches multiple objects at a moment. Therefore, the robot must determine which visual ones should be bound.

On the issue how to find the matched sensations in these modalities, Yoshikawa et al. have proposed the cross-modal map among vision, touch, and proprioception to learn the representation of the body surface [8]. It is based on the idea that the tactile sensors which collide with each other also coincide with each other in its vision. In other words, touching its body with its own body part (i.e., *double-touching*) always co-occurs the occlusion caused by covering its body with its own body part in its view (i.e., *self-occlusion*). They assumed that there is only one self-occlusion at a moment. However, there can be multiple self-occlusions since the body occupy a certain volume in the physical space. For example, when the agent touches its body trunk with its hand, not only the hand but also its arm cover its body trunk from its sight, therefore, multiple self-occlusions occur. Therefore, there still remains the binding problem where it must determine which self-occlusion should be bound to the double-touching and vice versa.

3 A basic idea

As suggested from the previous work [8], it seems reasonable to utilize the fact that double-touching co-occurs with self-occlusion although they did not take the physical volume of the body into account, which makes the binding problem remain formidable. We will explain a basic idea how the robot can correctly match double-touching and self-occlusion based on this fact. In the following, first we introduce the assumptions what kinds of cognitive competences it should possess and argue a possible developmental course to acquire them. Then, we show a basic idea of cross-anchoring to solve the binding problem by virtue of the morphological constraint. In the following argument, we suppose that it has a human-like configuration in which it has a trunk with a camera and an end-effector connected through serial links, that is, the robot consists of its trunk, a camera head and an arm.

3.1 A possible developmental course of prerequisite competences for binding

We assume that the robot has acquired the competences to detect double-touching and self-occlusion. However, it is worthy to argue how such assumptions are justified from a viewpoint of robotics and analogy to the human development. We propose a possible developmental course of the prerequisite competences, which consists of three stages: 1) learning to detect double-touching, 2) finding the major components of visual changes caused by its own camera head motions, and 3) learning to detect self-occlusion.

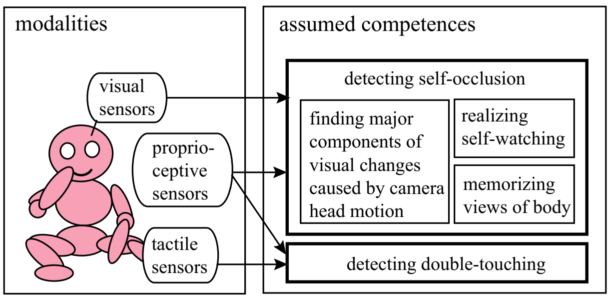
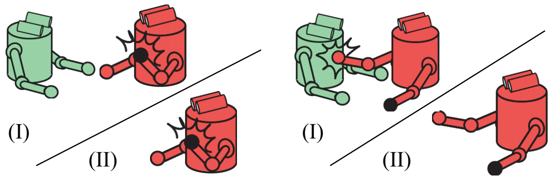


Figure 1. The modalities of the robot and the competences supposed to possess

Learning to detect double-touching At the first stage, a robot learns to detect double-touching through the iterations of double-touching. According to the assumption in the previous work [9] that the sensation of its body is invariant with its posture (see Fig. 2), it can judge whether the tactile sensation is caused by double-touching after learning the tactile-proprioceptive map that represents how invariant with its postures the tactile sensations are. Therefore, it can detect double-touching when it occurs after this stage.

Since it is observed that a human fetus touches its body with its hands in the womb [10] and it is reported that a human neonate can distinguish double-touching from being touched by the other in the study on rooting reflex [11], it seems biological plausible to assume that an infant has already possessed the competence to detect double-touching before the following stages with the visual sensation.



(a) a posture with the invariant touch

(b) a posture with the variant touch

Figure 2. Examples of the invariance (a) and variance (b) of the tactile sensation with certain postures in the different environment (I) and (II)

Finding major components of visual changes caused by camera head motion This is prerequisite for detecting self-occlusion. When the robot moves, an optical flow is induced in its vision. It is considered that the motion of the camera mainly contribute to this optical flow due to the following two facts: (1) the motion of the camera usually induces larger optical flow components in wider region than one of the arm and (2) what a human-like robot observes is not usually its arm but the environment due to its configuration. In other words, the motion of the camera head would often predict the changes in the optic flow of a robot with human-like configuration.

Therefore, the robot can find that the major components of visual changes by finding the principle component of the motion to predict the optic flow as performed in the previous work [12]. Human infants can not hold their heads up in their first several month [13], and therefore, usually lie on the bed. It can be conjectured that this kind of immaturities constraints infant's motion so that the neck motion is easily found since it usually watches external world rather than its arm in such a situation.

Learning to detect self-occlusion According to the same idea used in the case of double-touching, a robot learns to detect the occurrence of self-watching, that is watching its own body, by judging whether the visual sensations is invariant with its posture. If the robot learned the invariance with respect to the posture of the camera head, it is expected that it can detect the occurrence of self-watching of its trunk since the visual sensations of the trunk is invariant with the posture of the camera head. Once it learned to detect the occurrence of self-watching and memorized the invariant visual sensations with each posture, it can detect the occurrence of self-occlusion by comparing the current sensations and the memorized one.

3.2 A learning rule for the binding problem

Since the robot does not have *a priori* knowledge how to bind, we suppose that it keeps changing the posture both of its arm and its camera head at random to explore for binding. It perceives its posture and the view in the center region of its camera. Fig. 3 illustrates an example of self-watching view of a human-like robot (imagine that the robot watches its body in sitting). Fig. 4 illustrates the simplified situations of the robot's exploration for binding in Fig. 3, where DoFs for the arm is simplified to one to slide its end-effector while DoFs for the camera head is simplified to one to displace the camera. Note that the notations in Fig. 4 correspond to those in Fig. 3.

Concerning a point on the trunk, in this case B, there are following five types of experience of the robot (See Fig. 4). In each experience, the robot detects the posture of its arm, namely, Θ_B in Fig. 4(a)-(c), Θ_C in Fig. 4(d), and Θ_A in Fig. 4(e). In the case where the robot double-touches with B and tries to self-watch B (Fig. 4(a)), it detects the occurrence of self-occlusion. In the other cases where the robot double-touches B, it sometimes detects self-occlusion (Fig. 4(b)) but sometimes does not (Fig. 4(c)) depending on the posture of its camera head. In the other cases where it self-watches B, it sometimes detects self-occlusion (Fig. 4(d)) but sometimes does not (Fig. 4(e)) depending on the posture of its arm. The robot can not distinguish the self-occlusion by its end-effector (Fig. 4(a)) from self-occlusions by the link (Fig. 4(b) and 4(d)). Note that these kinds of experiences uniformly occur for each part on the trunk since the robot explores at random.

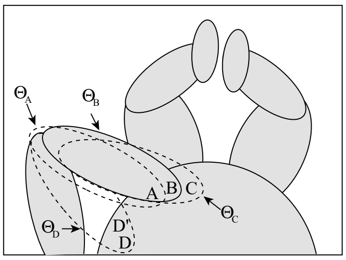
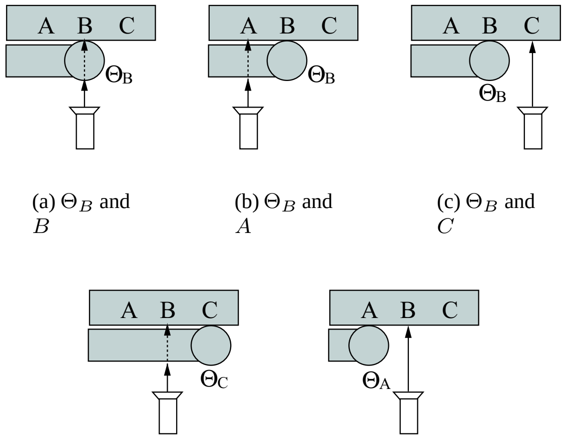


Figure 3. An example of self-watching view of the robot in double-touching: three ellipsoids with solid and broken lines are links in the postures (Θ_A, Θ_B, and Θ_C) by which it touches with the parts labeled A, B, and C, respectively. The rest ellipsoid with the symbols D and D' is one in the posture (Θ_D) by which it touches the part labeled D with occlusion at D and D'.



(a) Θ_B and B

(b) Θ_B and A

(c) Θ_B and C

(d) Θ_C and B

(e) Θ_A and B

Figure 4. Five simplified situations in the robot exploration: A top rectangle indicates the robot's trunk while a rectangle with a circle in the middle indicates the robot's arm that has a sliding DoF in the horizontal axis. The bottom object shows the posture of the robot's camera head while the arrow indicates its focus of attention. Notations correspond to those in Fig. 3.

A problem in the statistical approach As mentioned in section 2, the fact that the body occupies a certain volume in the physical space remains binding problem formidable. For example, a double-touching posture causes self-occlusions at multiple parts (see Figs. 4 (a) and (b)) while a self-occlusion at a part is caused by several double-touching postures (see figs. 4 (a) and (d)).

In the explorations, the robot sometimes experiences the matched responses in the different modalities which are caused by focusing on the same region, in this case detecting the self-occlusion at the double-touching point (see Fig. 4 (a)). However, such experiences of the matched response are not significantly frequent compared to mismatched responses (see Fig. 4 (b) or (d)) since it explores at random instead of utilizing *a priori* knowledge. In other words, the correctly matched responses are not significantly major in the obtained data. Therefore, it is difficult to associate them by considering all obtained data through the exploration. Then, we need a mechanism to narrow down the influence of the mismatched data on learning while augmenting the influence of the matched one.

Cross-anchoring association Due to the morphological constraints on the human-like configuration, we can utilize the following two morphological constraints: 1) how many double-touching postures occlude a certain part on the trunk depends on the location of the part to be occluded, and 2) how many parts the robot occlude by a double-touching posture depends on the location of the contact part.

These facts indicate that there exist *cue* nodes which have fewer candidates for matched response in other modalities to be bound. In this case, the self-occlusion at D can be the cue of double-touching in Θ_D while a double-touching in Θ_A can be the cue of the occlusion at A. Note that the matched response of a cue node is also experienced with mismatched response. For example, the robot sometimes detects a self-occlusion also at D' during the cue double-touching in Θ_D while it sometimes detects a double touching in Θ_B during detecting the cue self-occlusion at A. Since the desired correspondence between touch and vision can be found by unique association in this case, we can utilize such cue nodes as anchors of the unique association. Therefore, we introduce a learning rule with an anchoring mechanism which can adapt the learning rate according how much the responses simultaneously observed are regarded as unique to each other.

4 Cross-anchoring Hebbian learning rule

In this section, we introduce an cross-anchoring Hebbian learning rule as an implementation of the learning rule with the anchoring mechanism. The architecture consists of two layers called the double-touching layer and the self-occlusion one (see Fig. 5). In the double-touching layer, there are N_t nodes each of which is responsible for a set of certain posture of the arm $\Theta_i, (i = 1, \cdots, N_t)$ which is assumed to be quantized in advance. When the posture of the arm is $\boldsymbol{\theta} \in \Re^m$, the activation of the i-th node is calculated by

$$^{t}a_i(\boldsymbol{\theta}) = \begin{cases} 1 & \boldsymbol{\theta} \in \boldsymbol{\Theta}_i \\ 0 & \text{otherwise} \end{cases}. \quad (1)$$

On the other hands, in the self-occlusion layer, there are N_o nodes each of which is responsible for the self-occlusion in a set of certain posture of the camera head $\Phi_j, (j = 1, \cdots, N_o)$ which is assumed to be quantized in advance. When the posture of the camera head is $\boldsymbol{\phi} \in \Re^n$, the activation of the j-th node is calculated by

$$^{o}a_j(\boldsymbol{\phi}) = \begin{cases} 1 & \boldsymbol{\phi} \in \boldsymbol{\Phi}_j, O \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where O is the phenomenon of detecting occlusion.

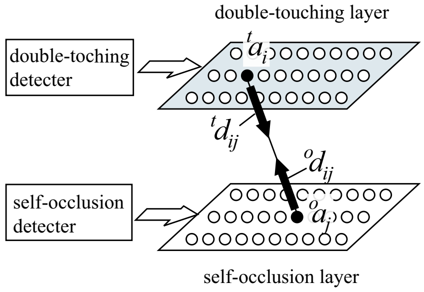


Figure 5. The architecture

Let a connection weight between the i-th neuron in the double-touching layer and the j-th neuron in the self-occlusion layer be w_{ij}. By the cross-anchoring Hebbian learning rule, $w_{i^*j^*}$ is updated as following:

$$\Delta w_{i^*j^*} = \eta({}^{t}d_{i^*j^*}\,{}^{t}a_{i^*} \cdot {}^{o}d_{i^*j^*}\,{}^{o}a_{j^*} - w_{i^*j^*}), \quad (3)$$

where i^* and j^* are the most activated units in the double-touching and the self-occlusion layer, η is a constant learning rate. The dynamic anchoring rates, ${}^{t}d_{ij}$ and ${}^{o}d_{ij}$, determine the degrees of anchoring on the j-th node in the self-occlusion layer from the i-th nodes in the double-touching layer and on the i-th node in the double-touching layer from the j-th nodes in the self-occlusion layer, respectively. They are calculated by

$$\begin{aligned} {}^{t}d_{ij} &= \exp\left(-\frac{\sum_{k,k\neq j} w_{ik}}{{}^{t}\sigma^2}\right), \\ {}^{o}d_{ij} &= \exp\left(-\frac{\sum_{k,k\neq i} w_{kj}}{{}^{o}\sigma^2}\right), \end{aligned} \quad (4)$$

where ${}^{t}\sigma$ and ${}^{o}\sigma$ are parameters that determine the degree of anchoring. Meanwhile, the remaining connection weights are decreased because they lost the competition;

$$\begin{aligned} w_{ij^*}(t+1) &= w_{ij^*}(t) - \eta_t(1 - {}^{t}d_{ij^*})\Delta w_{i^*j^*}, \\ w_{i^*j}(t+1) &= w_{i^*j}(t) - \eta_o(1 - {}^{o}d_{i^*j})\Delta w_{i^*j^*}, \end{aligned} \quad (5)$$

where η_t and η_o are constant coefficients of the competition.

In such an anchoring process, more unique combinations of double-touching and self-occluded are bound earlier. Meanwhile, some of the rest combinations become more unique since the other responses decrease the number of candidates to be bound by losing the responses that are already bound to others. Therefore, the process of binding proceeds step by step. This process is expected to converge since it is considered that there exist anchoring sensations in each modality due to the constraint in its human-like configuration.

5 Simulation results

As preliminary experiments, we tested the cross-anchoring Hebbian learning rule works so that the robot solves the binding problem by using the computer simulation. First, we examined a robot with a single DoF to show how learning proceeds. Then, we examined a robot with more DoFs.

5.1 Simulation with 1-D robot

In the first experiment, we simulated a robot with a 1-DoF sliding arm, a 1-DoF rotating camera head. During the exploration for binding, it moves its arm and camera head at random and detects self-occlusion and double-touching. Fig. 6 shows an example of self-watching view of the simulated robot in double-touching. For the reader's understanding, we quantized the posture space both of the arm and the camera head so that the nodes in both layers were matched with each other in one-to-one manner. The robot was trained for binding in 4,000 double-touching trials with the following network parameters: $N_t = N_o = 10$, $\eta = 0.1, \eta_t = \eta_o = 0.5$, and ${}^t\sigma = {}^o\sigma = 1.0$.

Figs. 7 (a):(I) $\sim$ (V) show the process of learning connection between double-touching layer and self-occlusion one. It can be seen that it starts with multiple connections and finally succeeded in binding since it obtained the correct one-to-one mapping at the 4,000-th step. Furthermore, we can see that the connections grew up both from the right and left ends to the center. It seems to show the process that cross-anchoring between a pair of nodes seems to make neighbor pairs of nodes more unique to each other and therefore guides cross-anchoring between the neighbor pairs. Such propagation of cross-anchoring starts from the pairs of nodes, either of which is a cue node. Consistently with the analysis of the learning procedure, the left end node in the bottom layer and the right end node in the top layer were cue nodes due to the morphological constraints. In this case, since the camera and the end-effector were connected through a serial link, how to double-touch and how to self-occlude were constrained. For example, the double-touching at the left end of the trunk could guide the self-occlusion only at the same part while the self-occlusion at the right end could be caused by the double-touching only at the same part.

Figs. 7 (b):(I) $\sim$ (IV) show the process of the learning connections in the case that the posture spaces of the camera head and the arm were quantized in different resolutions. In this case, the resolution of the double-touching was twice in the case of self-occlusion. The parameters were $N_t = 12$, $N_o = 6$, $\eta = 0.1, \eta_t = 0.5, \eta_o = 0.25$, and ${}^t\sigma = {}^o\sigma = 1.0$. Since it finally obtains the desired one-to-many mapping, we may conclude that it succeeds in binding despite the different resolutions.

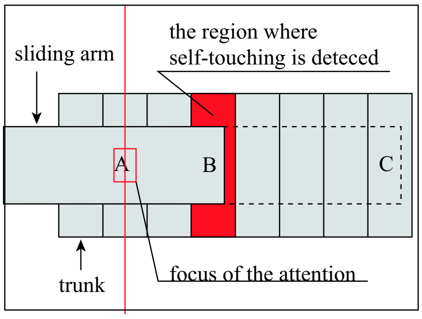


Figure 6. An example of self-watching view of the simulated 1-D robot in double-touching: The biggest gray box is its trunk while the other gray box is its sliding arm. Although each part labeled by a symbol is corresponding to Fig. 3, it is supposed that its trunk is a plane and the DoFs for the motion of its arm is one for the simplicity. The small box on the horizontal line indicates the focus of the attention in vision. The black box indicates the region where double-touching is detected.

After these processes, when the robot double-touches its body trunk, it can use the acquired mapping to know how to shift the focus of attention in the vision to the double-touching part by propagating the activation of the nodes responsible for the double-touching through the learned connection. Shortly, it can watch its touching part on its body.

5.2 Simulation of 2-D robot

In the second experiment, we simulated a more realistic robot with a 2-DoF rotating arm, a 2-DoF rotating camera head. Taking a posture of the camera head was emulated by changing the focus of the attention in vision. Fig. 8 shows an example of self-watching view of the simulated robot in double-touching. For the reader's understanding, we quantized the posture space both of the arm and the camera head so that the nodes in both layers were matched with each other in one-to-one manner. We let the robot learn the connections 100 times in each of 100,000 double-touching trials with the following network parameters: $N_t = 25$, $N_o = 25$, $\eta = 0.1, \eta_t = \eta_o = 2.0$, and ${}^t\sigma = {}^o\sigma = 3.0$.

Fig. 9 shows the process of learning connections between double-touching layer (top) and self-occlusion one (bottom). It can be seen that it started with multiple connections and finally succeeded in binding since it obtained the correct one-to-one mapping at the 100,000-th step. Fig. 10 shows the averaged time course of the matching errors of

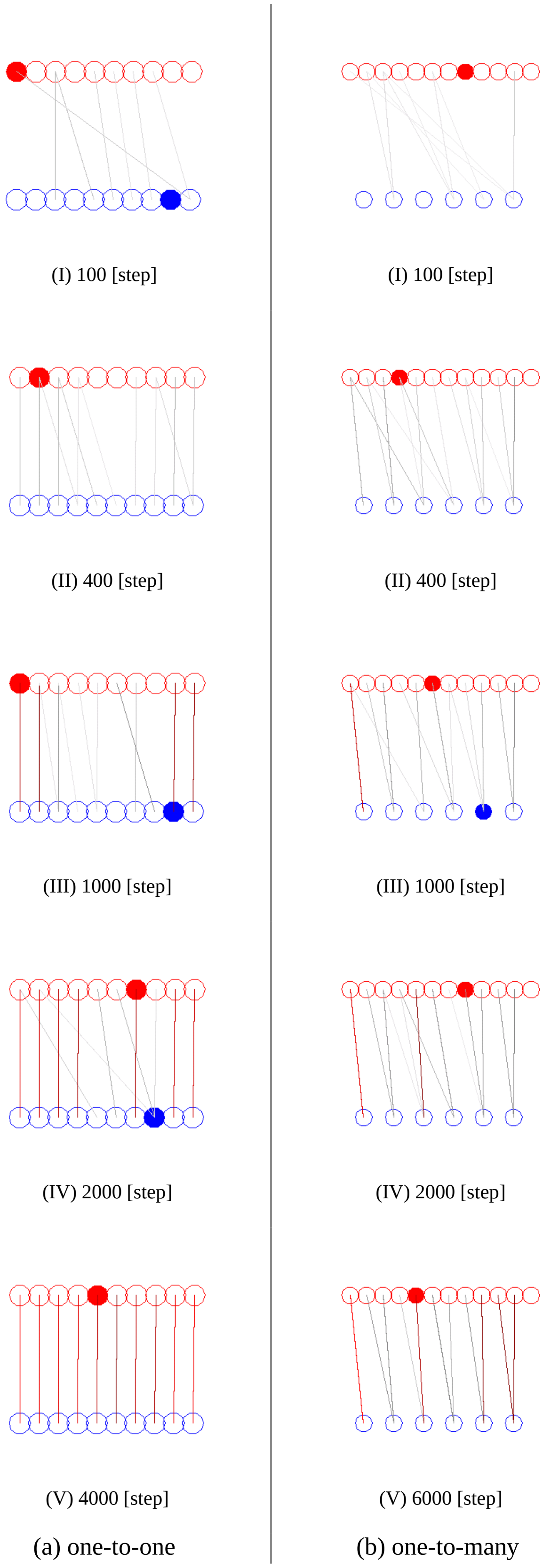


Figure 7. The process of learning connection between the layers ((I) $\sim$ (V)) with the same resolution (a) and with the different one (b).

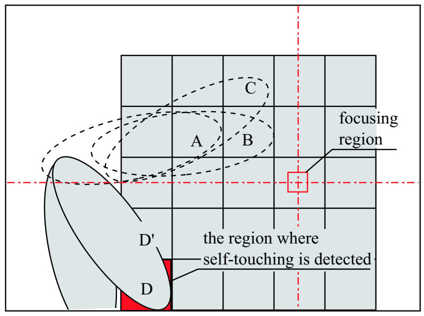


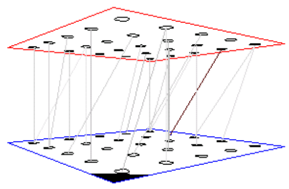
Figure 8. An example of self-watching view of the simulated robot in double-touching: Three ellipsoids with broken curves and one with the symbols D and D' are the postures of the robot arm by which it touches with its trunk. Although each part labeled by a symbol is corresponding to Fig. 3, the trunk is supposed to be a plane for the simplicity. The cross point of the vertical and the horizontal chain lines indicates the focus of the attention in vision. The black box indicates the region where double-touching is detected.

100 times in which the processes of exploration are different. Note that only the experimenter knows the desired connection and can determine the matching error. We can see that learning for binding converged to almost completely correct one-to-one mapping with small variance. Therefore, we may conclude that the robot can robustly find the matched response in different modalities by the anchoring Hebbian learning rule in a more realistic embodiment.

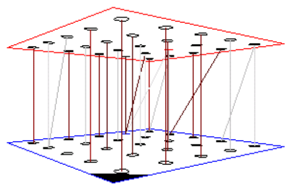
6 Conclusion

In this paper, we address the issue how to solve the binding problem in different modalities for body representation. We propose a method called cross-anchoring Hebbian learning rule to perform binding by virtue of the morphological constraints of its human-like configuration in perceiving the self. In the preliminary computer simulations, we showed that the robot can bind its tactile and visual sensations through the exploration by the iterations of self-watching and double-touching.

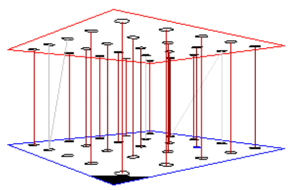
There are parameters in the proposed learning rule that determine how much the degree of anchoring is. Since it should be well selected to obtain the unique association, we should put a mechanism to adapt it when the system fail to bind. Topographical constraint caused by the the receptive fields with continuity that reflects the physical continuity could be a criteria for adaptation. Furthermore, the robot



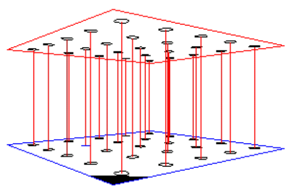
(I) 2500 [step]



(II) 12500 [step]



(III) 25000 [step]



(IV) 100000 [step]

Figure 9. The process of learning connection between the layers with the same resolution in the 2-D robot simulation

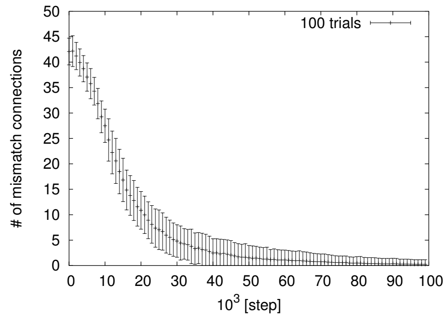


Figure 10. The number of the mismatched connections: A connection is counted as mismatched if it is to a mismatched neuron with more than 0.1 strength and if it is to the matched one with less than 0.8 strength.

needs the competence of binding in the case where it learns multimodal representation of the external objects. Although we concentrated on the binding problem concerning the self body in this paper, extending the proposed method for the binding problem involving tactile sensations of being touched by others is one of our future work.

Acknowledgment The Advanced and Innovational Research program in Life Sciences of the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government and a Research Fellowship for Young Scientists from Japan Society for the Promotion of Science supported this research.

References

[1] A. Treisman. Solutions to the binding problem: Progress through controversy and convergence. *Neuron*, Vol. 24, pp. 105–110, 1999.

[2] J. H. Reynolds and R. Desimone. The role of neural mechanism of attention in solving the binding problem. *Neuron*, Vol. 24, pp. 19–29, 1999.

[3] G. Tononi, O. Sporns, and G.M. Edelman. Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, Vol. 2, pp. 310–335, 1992.

[4] A.K. Seth, J.L. McKinstry, G.M. Edelman, and J.L. Krichmar. Visual binding, reentry, and neuronal synchrony in a physically situated brain-based device. In *Proc. of the 3rd Intl. Workshop on Epigenetic Robotics*, pp. 177–178, 2003.

[5] K.F. MacDorman, K. Tatani, Y. Miyazaki, and M. Koeda. Proto-symbol emergence. In *Proc. of the Intl. Conf. on Intelligent Robot and Systems*, pp. 1619–1625, 2000.

[6] A. Sotytchev. Development and extension of the robot body schema. In *Proc. of the Intl. Workshop on Epigenetic Robotics*, pp. 179–180, 2003.

[7] M. Asada, K.-F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, Vol. 37, pp. 185–193, 2001.

[8] Y. Yoshikawa, H. Kawanishi, M. Asada, and K. Hosoda. Body scheme acquisition by cross modal map learning among tactile, image, and proprioceptive spaces. In *Proc. of Intl. Workshop on Epigenetic Robotics*, pp. 181–184, 2002.

[9] Y. Yoshikawa, K. Hosoda, and M. Asada. Does the invariance in multi-modalities represent the body scheme? - a case study with vision and proprioception -. In *Proc. of Intl. Symp. on Adaptive Motion of Animals and Machines*, 2003.

[10] J. W. Sparling, J. V. Tol, and N. C. Chescheir. Fetal and neonatal hand movement. *Physical Therapy*, Vol. 79, No. 1, pp. 24–39, 1999.

[11] P. Rochat and S.J. Hespos. Differential rooting response by neonates: Evidence for an early sense of self. *Early Development and Pareting*, Vol. 6, pp. 105–112, 1997.

[12] T. Nakamura and M. Asada. Motion sketch: Acquisition of visual motion guided behavior. In *Proc. of Intl. Joint Conf. on Artificial Intelligence*, pp. 126–132, 1995.

[13] G. Bremner. *Infancy*. Blackwell, 1994.