International Journal of Humanoid Robotics

Body Image Constructed from Motor and Tactile Images with Visual Information

Sawa Fuke

Dept. of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan,
sawa.fuke@ams.eng.osaka-u.ac.jp

Masaki Ogino

Asada Synergistic Intelligence Project, ERATO, JST, FRC1, Graduate School of Engineering, Osaka-Univ. 2-1 Yamadaoka, Suita Osaka 565-0871, Japan
ogino@jeap.org

Minoru Asada

Dept. of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan,
Asada Synergistic Intelligence Project, ERATO, JST, FRC1, Graduate School of Engineering, Osaka-Univ. 2-1 Yamadaoka, Suita Osaka 565-0871, Japan
asada@jeap.org

Received 25 09 2006
Revised 23 12 2006
Accepted 22 03 2007

This paper proposes a learning model that enables a robot to acquire a body image for parts of its body that are invisible to itself. The model associates spatial perception based on motor experience and motor image with perception based on the activations of touch sensors and tactile image, both of which are supported by visual information. The tactile image can be acquired with the help of the motor image, which is thought to be the basis for spatial perception, because all spatial perceptions originate in motor experiences. Based on the proposed model, a robot estimates the invisible hand positions using the Jacobian between the displacement of the joint angles and the optical flow of the hand. When the hand touches one of the invisible tactile sensor units on the face, the robot associates this sensor unit with the estimated hand position. The simulation results show that the spatial arrangement of tactile sensors is successfully acquired by the proposed model.

Keywords: Body Image; Sensor Fusion; Learning and Adaptive System

1. Introduction

It is paramount that robots become capable of body representation if they are to develop further, and there are two standard approaches to represent the relationship between the robot body and the space around it. One is a model-based approach wherein knowledge about the parameters of the links and cameras of a robot is given in advance. The other is more adaptive in that the robot estimates these parameters based on its experience in the environment [1 2 3]. The latter approach is closely related to human body representation; recent brain and medical studies have revealed that biological systems have flexible body representation, so-called body image. Ramachandran showed that patients suffering from phantom limb pain could alleviate their pain by observing the visual feedback of the good limb in a mirror box. He also suggested that the cortical representation of a patient's body might be restructured after the loss of a limb [4]. Iriki et al. showed that the receptive field of the bimodal (somatosensory and visual) neurons in the intraparietal cortex is extended when monkeys use tools to obtain food [5]. Moreover, these body images are thought to represent the relationship between an animal's own body and the external world. This may suggest that body image is the spatio-temporally integrated image of various modalities, such as auditory and visual perceptions and somatic (including tactile) sensations as well.

In developmental cognitive science, it has not yet been revealed how and when humans acquire their body images. Human newborns can imitate gestures such as mouth-opening and tongue-protrusion within a few hours after their birth [6]. This may suggest that a newborn can be aware of the parts of the face of their parents that correspond to its own face. This has lead to discussion on how this is possible at such an early stage of development.

Meltzoff and Moore proposed the active intermodal mapping (AIM) model to explain this form of early imitation [7]. In their model, organ-identification, through which newborns can associate the sensory perception of invisible parts with the features of parts of others in visual information, is a prerequisite. This model suggests that newborns are able to compare gestures and produce facial expressions regardless of differences in modality. However, recent studies reveal the possibility of fetus learning in the womb [8]. Recent sonographic observations have revealed that the fetus' eyes open after about 26 weeks of gestation and that the fetus often touches its face with its hands during embryonic weeks 24 and 27 [16]. Moreover, it is reported that visual stimulation from outside the maternal body can activate the fetal brain [9]. Thus, it does not seem unreasonable to suppose that infants acquire a primitive body image through experiences in womb.

Cognitive developmental robotics has been proposed aiming at discovering a new way of understanding ourselves, especially focusing on the human cognitive developmental process by building a robot that can reproduce the process. In case of body representation, a robot should adaptively acquire the relationship between its own body and the external world. Nabeshima et al. [10] proposed a model that explains the behavior of the neuron observed in the experiment of Iriki et al. [5]. In their model, a robot detects the synchronization of the visuo-tactile sensations based on an associative memory module and acquires a body image. Yoshikawa et al. [11] proposed a model in which a robot develops an association among visual, tactile, and somatic sensations based on Hebbian learning, while touching its own body with its hand. However, in these studies, body parts to be integrated are limited to visible ones, and their methods cannot be applied to the acquisition of body images of invisible parts, such as the face or the back.

In order to represent the invisible body parts, spatial perception that represents the relationship between the body and the external world appears to be important because the locations of invisible parts might be predicted from experiences in the visible area. Studies in brain science suggest that the hippocampus is involved in coding a specific location in space by integrating motion information such as optical flow to localize its own position in space [12]. In this paper, we apply this to the issue of representing the invisible body parts of the robot. We refer to spatial perception based on motor experience as the motor image and propose a learning method to acquire a body image of an invisible face part based on the motor image. An invisible hand position is estimated by integrating the Jacobian of the hand displacement and the resultant optical flow. Thus, a robot can associate the tactile sensor unit with the visual information through touching experiences using its own hand.

The remainder of the present paper is organized as follows. First, an overview of the proposed system is presented, and the details of the system and the learning algorithm are given. Then, the simulation results are presented. Finally, a discussion and conclusions are given.

2. Estimation of the invisible hand position based on motor image

A body image is thought to be an integration of spatial perceptions in terms of different modalities. We define "X image" as the spatial perception based on modal X. We assume that a body image consists of the principal two images: tactile and motor images (Fig. 1). A motor image is the spatial perception based on motor experience. An optical flow of the hand is the result of the motor commands, and therefore the flow with motor command is one example of the motor image. A tactile image is thought to be the sensation of spatial perception when some tactile sensors are touched. Thus, the visual information is utilized to construct motor and tactile images. These images are not acquired at the same time in the developmental process. Rather, the maturation of one image can assist the development of the other. The motor image is thought to be the most important and precedent spatial perception, because it seems that all spatial perceptions originate in motor experiences. Here, we show that the tactile image can be acquired with the help of the motor image (arrow in Fig. 1).

The motor image is more concretely defined as the mapping between the proprioceptive space (joint angles) and vision space, and the tactile image is defined as the mapping between the tactile space and vision space. A robot can acquire such mapping by touching its own body parts with its hand and associating the coordinates of the touched part in the camera image with the identification of activated tactile sensor units and the joint angles. However, this approach cannot be used for touched parts that are not visible, such as the face and back. In these cases, it is

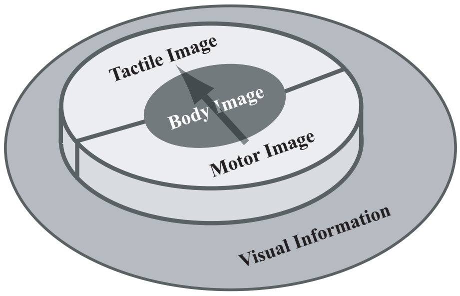


Fig. 1. A body image comprising tactile and motor images supported by visual information

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 (motor image) is inevitable to construct the body image. We suppose that even an infant who has not yet experienced locomotion has achieved primitive spatial perception by associating its hand motion with the resultant information. It has already been shown that the hand can be used as a probe to explore the world [5]. Thus, for the robot to explore its invisible parts with its hand, it is important to associate the invisible hand positions with the visible ones. We propose a learning process: first, the displacement of the hand position related to the motion is learned (spatial learning phase (Fig. 2 (a))), then the invisible tactile sensor units are associated with the spatial perception with the hand probe, based on the learned spatial perception of the hand (mapping phase (Fig. 2 (b))).

In the first phase (Fig. 2 (a)), while a robot moves its hand in front of its face, it learns the Jacobian, f, the relationship between the displacement of the hand position in the camera image, $\Delta\mathbf{r}_m$, and the displacement of the joint angles, $\Delta\theta_m$,

$$\Delta\mathbf{r}_m = f(\Delta\theta_m). \tag{1}$$

In this phase, the body image mapping is not learned, because the tactile information is not available.

In the second phase (Fig. 2 (b)), the robot touches its face with its hand. In this phase, visual information is not available, and the imaginary hand position, $\hat{\mathbf{r}}$, is estimated by the learned Jacobian and its integration,

$$\hat{\mathbf{r}} = \mathbf{r_0} + \int f(\Delta\theta)dt. \tag{2}$$

Although the accurate Jacobian cannot be obtained directly through the experience, we assumed that the learned Jacobian is a reasonable approximation of the Jacobian in invisible space if the joint angles are similar to each other. In the second phase (Fig. 2 (b)), based on the estimated hand position in the visible area, a robot can associate the hand position with the touched sensor units and the joint angles.

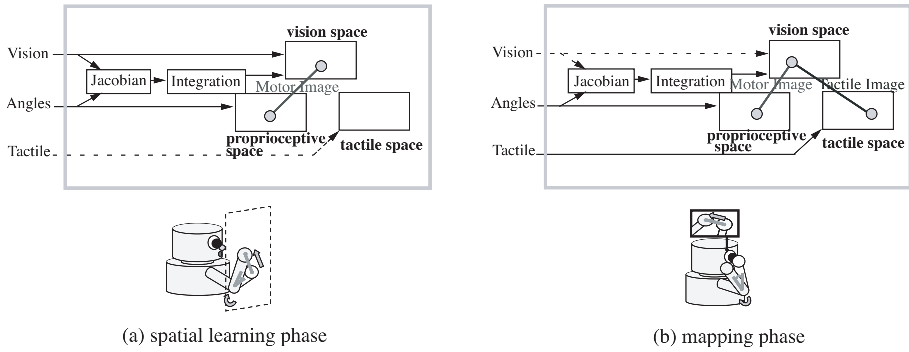


Fig. 2. The proposed model to learn mapping between invisible parts and tactile sensation: (a) In the spatial learning phase, a robot constructs a motor image, which is the association between the vision space and proprioceptive space, through the experience of observing its own hand moving in front of the face. At the same time, the Jacobian, which is the relationship between the displacement of the joint angles of the arm and the resultant optic flow of the hand, is also learned. (b)In the mapping phase, the robot constructs a tactile image, which is the association between the vision space and the tactile space, while touching its face. The invisible position of the face is estimated by integrating the virtual displacement that is calculated by the Jacobian learned in the spatial learning phase.

3. Acquisition of facial body image

The preconditions for a robot to acquire a body image of its face are as follows,

- The robot can detect the hand position in its camera coordinate system and know the posture of the arm by its joint angles.
- The tactile sensors are arranged in a grid pattern, and the robot can detect the activated tactile sensor units.
- The robot does not know in advance the relationships among the vision, proprioceptive, and tactile spaces.

3.1. *vision, tactile, and proprioceptive spaces*

3.1.1. *proprioceptive space*

The joint angles of an arm constitute the proprioceptive space.

$$\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n), \tag{3}$$

where n is the number of joint angles. The data during the face touching are collected and a self organizing map (SOM) is constructed before the mapping phase, as shown in Fig. 3. Thus, a unit of the proprioceptive space is a representative vector of this SOM,

$$\boldsymbol{\Theta}_{\mathbf{i}} = (\theta_1^i, \theta_2^i, ..., \theta_n^i). \tag{4}$$

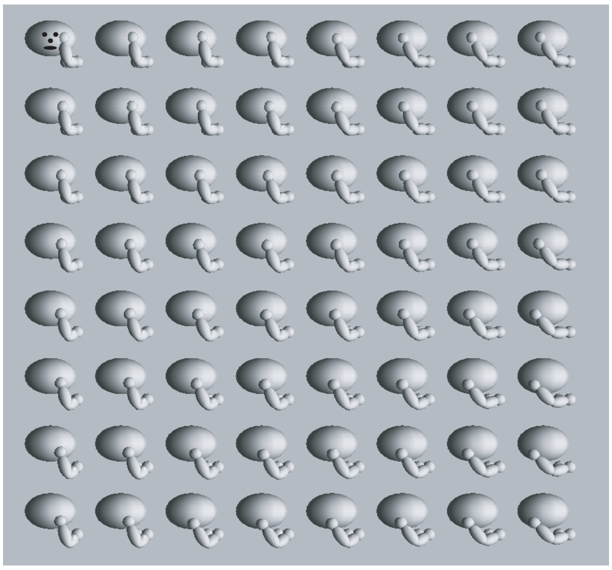


Fig. 3. A self organizing map of the joint angles; 8x8 figures show the representative vectors of the self organizing map as the posture of the arm with the face (other body parts such as the other arm, legs, and the body are not shown). The joint angles are collected as training data for the SOM while the robot touches its face randomly.

3.1.2. *tactile space*

The tactile sensor units are arranged in a grid pattern on the robot face. Thus, the unit of the tactile space is each tactile sensor unit,

$$\mathbf{T_i} = (T_{x_i}, T_{y_i}) \tag{5}$$

where T_{x_i} and T_{y_i} are the coordinates on the face. We arrange the units on the tactile space in the same way as on the face (in a grid pattern).

3.1.3. *vision space*

Unlike the other two modalities, the representative unit is not prepared for vision space. Instead, the continuous value is used to represent the visual information. The visual information is represented as the position of the robot hand in the camera coordinate system,

$$\mathbf{r} = (r_x, r_y). \tag{6}$$

3.2. *Learning Jacobian and Estimation of Invisible Hand Position by Integration*

The Jacobian transformation from the displacement of joint angles to that of the coordinates of the hand, f, is represented by a neural network. The relationship is learned by the back-propagation algorithm during the probe hand motion just in front of the face (Fig. 2 (a)). In the mapping phase, the robot touches its face (invisible parts), and the position of the probe hand is estimated by the following equation. When a robot touches its face during the time period between t_0 and t_1, the estimated hand position, $\hat{\mathbf{r}}$, is calculated as

$$\hat{\mathbf{r}} = F(\theta_{t_0}) + \int_{t_0}^{t_1} f(\Delta\theta)dt, \tag{7}$$

where F is the mapping from the proprioceptive space to the vision space, and $\Delta\theta$ is the displacement of the joint angles ($\Delta\theta = \theta_{t_1} - \theta_{t_0}$). Although the different postures have different Jacobian transformations, in general, here we postulate that the posture difference has little effect on the Jacobian, because the joint angles of the arm during the hand motion in front of the face are close to those during touching the face. The convergence of the learning is evaluated by the total error in teacher data, and the neural network could always approximate the training data well, starting with different random connection weights.

3.3. *Learning the mapping from the tactile space to the vision space*

In the simplest model, the mapping between the tactile space and the vision space can be described as the simple Hebbian learning of the connection weight:

$$\Delta w_{ik} = \alpha A_i^T A_k^r, \tag{8}$$

where α is the learning rate, and A_i^T and A_k^r are the activation levels of the i-th unit of the tactile sensor and the k-th unit of the vision sensor

$$A_i^T = \begin{cases} 1 & \text{the } i\text{-th sensor unit is touched} \\ 0 & \text{else} \end{cases}, \tag{9}$$

$$A_k^r = \begin{cases} 1 & \begin{array}{l}\text{the hand is detected at the position in the camera image} \\ \text{that corresponds to the } k\text{-th unit}\end{array} \\ 0 & \text{else} \end{cases}. \tag{10}$$

Supposing that the units near to the touched sensor unit are also activated, the learning equation becomes,

$$\Delta w_{ik} = \alpha \exp(-\|\mathbf{T}_i - \mathbf{T}_c\|/\gamma) A_k^r, \tag{11}$$

where $\mathbf{T}_c$ is the coordinate of the touched tactile sensor unit.

Since the vision does not have exact unit representation, the mapping from the tactile space to the vision space is modelled as the learning of the reference vector of the unit in tactile space,

$$\mathbf{r}_i^T = (x_i^T, y_i^T). \tag{12}$$

Instead of using the update function of the above mentioned connection weight, this reference vector is updated in the same way as the self organizing map algorithm [13]. 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,

$$\Delta \mathbf{r}_i^T = \alpha_T(t) \exp(-\|\mathbf{T}_i - \mathbf{T}_c\|/\gamma)(\hat{\mathbf{r}} - \mathbf{r}_i^T)), \tag{13}$$

where $\alpha_T(t)$ is a learning rate that decays as the learning goes and $\mathbf{T}_i$ and $\mathbf{T}_c$ are the coordinates of the i-th and c-th units on the tactile space, respectively.

3.4. *Learning the mapping from the proprioceptive space to the vision space*

In the same way as the mapping from the tactile space to the vision space, the mapping from the proprioceptive space to the vision space is defined as,

$$\mathbf{r}_i^\Theta = (x_i^\Theta, y_i^\Theta). \tag{14}$$

The update algorithm is also done in the same manner as in the mapping $\mathbf{r}^T$. When the current estimated hand position is $\hat{\mathbf{r}}$, the mapping for the i-th unit on the proprioceptive space is updated depending on the distance from the c-th unit on the proprioceptive space,

$$\Delta \mathbf{r}_i^\Theta = \alpha_\Theta(t) \exp(-\|\mathbf{u}_i - \mathbf{u}_c\|/\gamma)(\hat{\mathbf{r}} - \mathbf{r}_i^\Theta), \tag{15}$$

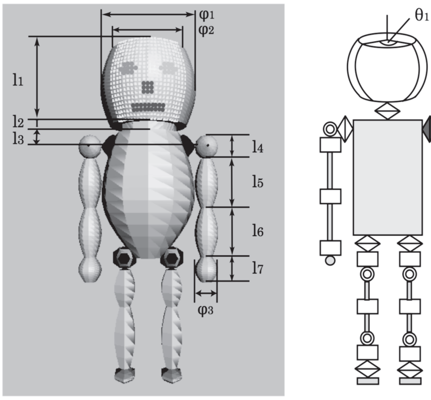
where $\alpha_\Theta(t)$ is a learning rate that decays as the learning goes and $\mathbf{u}_i$ and $\mathbf{u}_c$ are the coordinates of the i-th and c-th units on the proprioceptive space, respectively.

4. Experimental result

To validate the proposed method, computer simulations are conducted in a dynamics simulator. The robot model used in this experiment and its specification are shown in Fig. 4. In this experiment, five joint angles of the left arm, which are colored black in Fig. 4 constitute the proprioceptive space. The robot has 21×21 tactile sensor units arranged in a grid array on its face as shown in Fig. 5. The sensors that belong to the eye, nose, and mouth, can be differentiated by kind of marks for the reader's convenience, but there is no difference among all tactile sensor units. When the face is touched with hand, the nearest sensor unit is most activated. In addition, the neighbor sensor units are activated depending on the distance on the tactile space,

$$I(i) = exp(-(\mathbf{T}_i - \mathbf{T}_c)^2/\gamma)) \tag{16}$$

Here, γ is a scaling factor. We apply a Gaussian function to simulate the face touched by a human hand that is not restricted to one point but a regional area. This property of activated levels of tactile sensors are used in learning of the mapping as mentioned in section 3.



θ1	90°
φ1	0.123[m]
φ2	0.0984[m]
φ3	0.03[m]
l1	0.0891[m]
l2	0.00990[m]
l3	0.015[m]
l4	0.03[m]
l5	0.06[m]
l6	0.06[m]
l7	0.03[m]

Fig. 4. The robot model and its specifications used in the experiments. The robot has five degrees of freedom in each arm and seven degrees of freedom in each leg, and one freedom in the neck. In this experiment, the robot touches its face with the left hand. Five joint angles of the left arm (colored black) constitute the proprioceptive space.

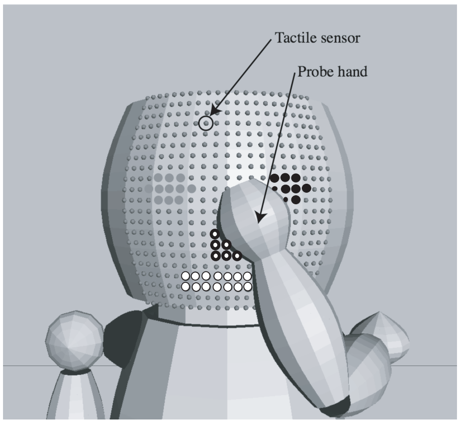


Fig. 5. The close-up figure of the face of the robot. It uses the hand as a probe while touching its own face. There are 21×21 tactile sensor units on the surface of the face.

In this simulation, since we assume a monocular vision system (the right eye), the visual target is projected on the screen just in front of the face. Fig. 6 shows the coordinate system when the screen is $z = h_z$ and the origin of this coordinate system is the center of the right eye. The position of the hand in space is,

$$\mathbf{h}' = (h'_x, h'_y, h'_z). \tag{17}$$

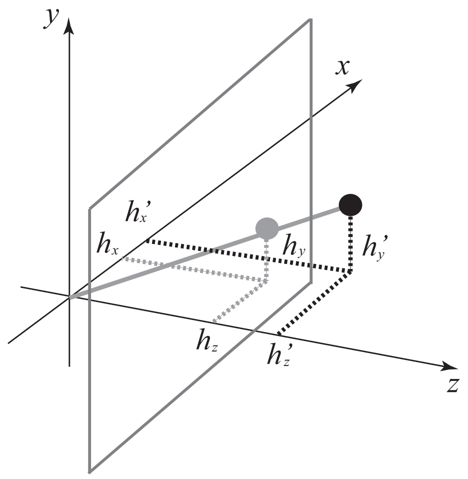


Fig. 6. The coordinate system for the vision space; The visual targets are projected on the virtual screen $z = h_z$. (h'_x, h'_y, h'_z) is the position in the space and (h_x, h_y, h_z) is the one on the screen.

The projected position of the hand is given by;

$$h_y = \frac{h_z \times h'_y}{h'_z}, \tag{18}$$

$$h_x = \frac{h_z \times h'_x}{h'_z}. \tag{19}$$

4.1. *Estimation of the hand position*

In the first experiment, the estimation of the hand position is evaluated. As explained earlier, 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. This neural network is trained by the back-propagation method [14] with the data collected while the robot moves the probe hand in front of its face as shown in Fig. 7. In this case, 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 external force to follow such a desired trajectory is applied to the hand link. The other links move passively. In this simulation, the parameters are set as shown in Table. 1. After the training phase, the velocity estimated by the Jacobian and the actual velocity are compared while the robot moves its hand in the following sequence of positions: mouth, nose, right eye, left eye and nose. Figs. 8 (a) and (b) are the velocities of the hand along with x and y axes in the camera coordinate system shown in Fig. 6, respectively. In addition, in order to find out whether this model can be applied to the real robot, the random torque is added to the hand in each step to simulate the noise using a normal distribution with

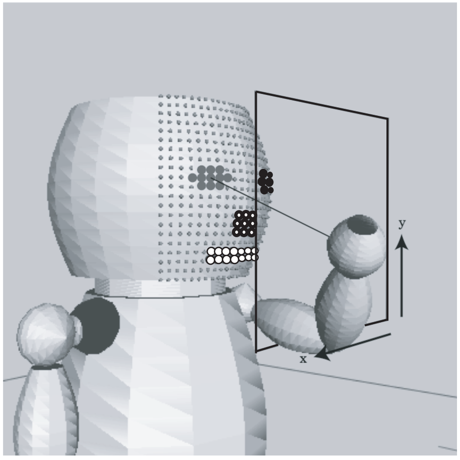


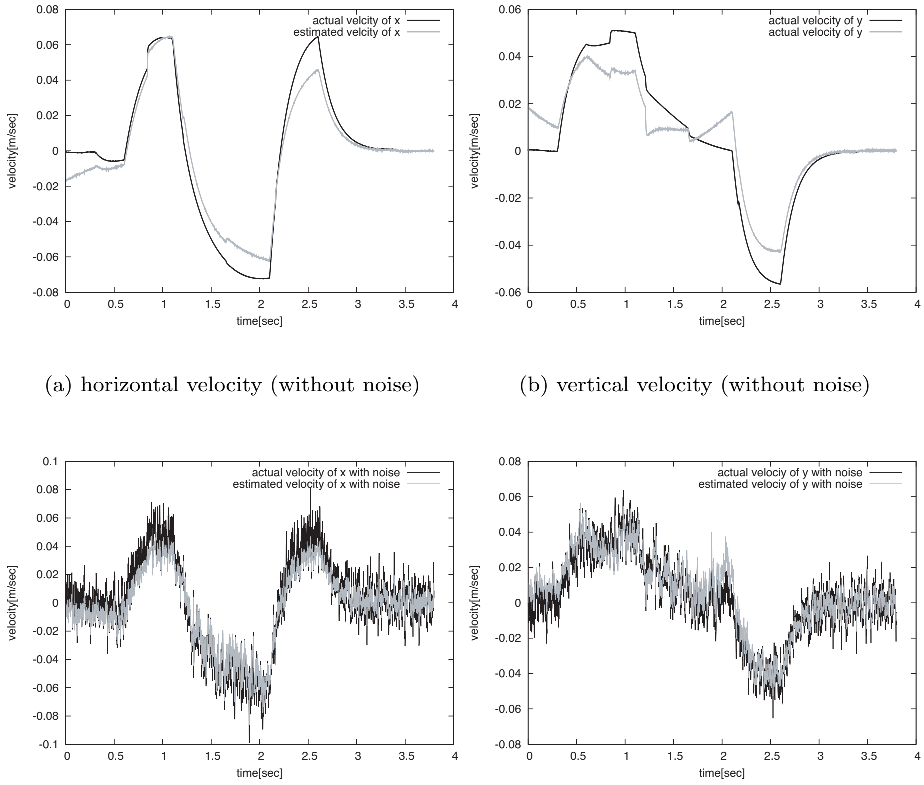
Fig. 7. The robot learns the Jacobian during the hand motion in front of its face; The square in front of its face indicates the virtual screen shown in Fig.6

the mean and variance shown in Table. 1. They are added both in training and estimating phases. Therefore, the learning steps increase from 200 to 500. Figs. 8 (c) and (d) are the the velocities of the hand along x and y axes when the noises of the joint angles are added.

Fig. 9 shows the actual and estimated trajectories of the hand when the hand moves in the same manner as shown in Fig. 8. Fig. 9 (a) shows the result without noises and (b) with noises. For comparison, the initial position for integration of the velocity is set at the position of the mouth. The results show that the Jacobian trained with the visible data estimates the hand position well, because the relative positions of the tactile sensors seem fixed although the long-time integral may accumulate errors. In other words, these results imply the possibility that the robot can use the Jacobian to recognize the topological relationships among the facial organs such as the eyes, the nose and the mouth.

4.2. *Acquiring the facial image of tactile sensors*

Based on the learned Jacobian, the mapping between vision and proprioceptive spaces and the mapping between vision and tactile spaces are associated. While the robot touches a random positions on its face with its hand, the mappings are updated with the algorithm explained in 3.4. The Jacobian trained with the data of the arm with noises is used to estimate the displacement of the hand position from the right eye. Since there is a possibility that the error of the estimated position increases because of the long-time integral, the integral is reset each time the hand stops at the right eye position. The initial estimated position on the camera coordinate system of tactile sensor units are random. In order to simulate the real robot



(a) horizontal velocity (without noise)

(b) vertical velocity (without noise)

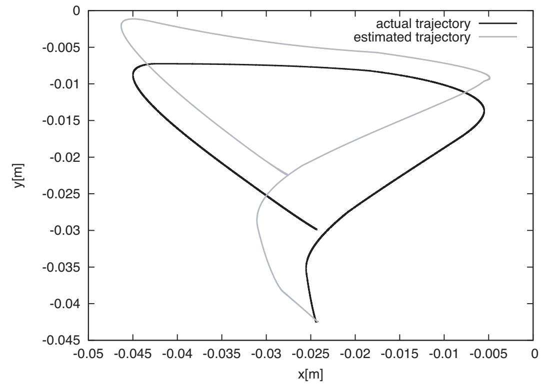
(c) horizontal velocity (with noise)

(d) vertical velocity (with noise)

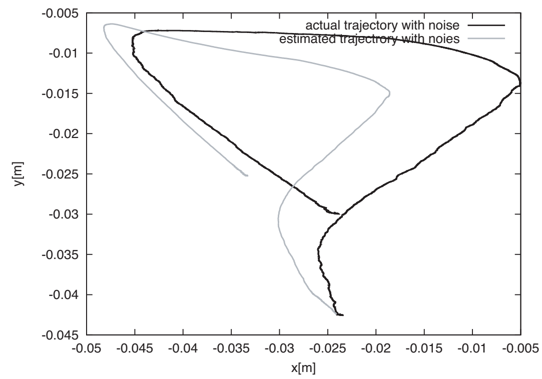
Fig. 8. The actual and estimated velocities of the hand in each direction; The black curve is a actual velocity and grey one is a estimated velocity using the learned Jacobian and joint angles while touching its own face. (a) and (b) are the results without noise and (c) and (d) are the ones with noise, respectively.

experiments where tactile sensors are sometimes inactivated even though the corresponding area are touched, the tactile sensors output no signal with the probability of 20% when the sensor units are actually touched.

The learning time is 800 [sec] in simulation time and the mapping is updated every 0.1 [sec], thus the total number of the learning steps is 8000. Fig. 10 shows the estimated coordinates of the tactile sensors in visual space. The x and y axes in this figure are the same ones in the camera coordinate system shown in Fig. 6. The same units are colored in red, green, yellow, and blue to indicate the positions of sensor units corresponding to the left and right eyes, the mouth, and the nose, respectively, purely for the reader's convenience. As the learning steps proceeds, the relative positions between sensor units gradually become plausible.



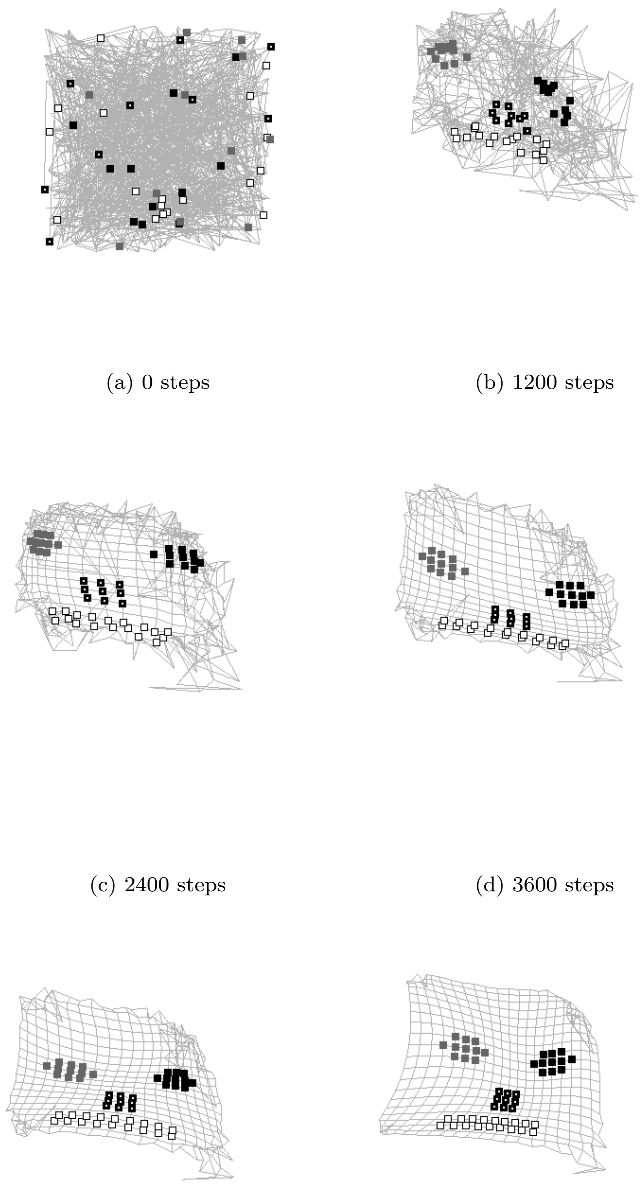
(a) with noise



(b) without noise

Fig. 9. The actual and estimated velocities of the hand; (a) is the result without noise and (b) is the one with noise

To show the validity of the methods, we iterated the experiments 10 times with different initial positions of the tactile sensors and measured the topological error as shown in Fig. 11. Topological error refers to the number of the sensor units whose positions on the camera coordinate system are relatively incorrect to the neighbors ($x_i^T > x_{i+1}^T$ or $y_i^T > y_{i+1}^T$) assuming that the correct positions are aligned with the grid pattern. In Fig. 11, the average value with the standard deviation is shown. The error decrease as the learning steps proceed although the relative position seem distorted partially in Fig. 10. It could be argued that the reason why the error remains is that the tactile sensors that are located on the edges of the face



(a) 0 steps (b) 1200 steps

(c) 2400 steps (d) 3600 steps

(e) 4800 steps (f) 7200 steps

Fig. 10. Tactile sensor units mapped on the imaginary visual space (2D-plots); It is the camera coordinates system shown in Fig. 6. The initial positions of the sensors are random.

scaling factor γ in formula (16)	40
screen position h_z in the camera coordinate system in Fig. 6	$0.04[m]$
learning rate : back-propagation (without noise)	0.2
learning steps : back-propagation (without noise)	200
number of hidden layer : back-propagation (without noise)	10
learning rate : back-propagation (with noise)	0.2
learning steps : back-propagation (with noise)	500
number of hidden layer : back-propagation (with noise)	10
mean value of the normal distribution of noise	0.0
variance of the normal distribution of noise	0.01

Table 1. Parameters used in this experiment

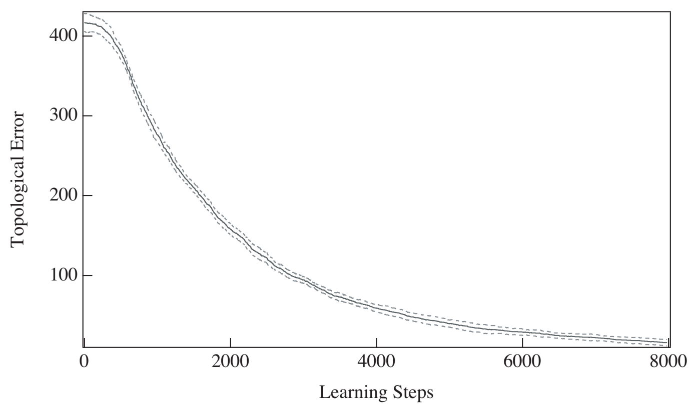


Fig. 11. The topological error during the learning shows that the average with the standard deviation error decreases as the learning steps proceed.

and rarely touched, have not been updated well.

5. Discussion

As well as the acquisition of body image, the proposed model is related to the early imitation of infants. As mentioned in Section 1, the AIM model is thought to be one of the representative models for early facial imitation of infants. However, while organ identification is fundamental to the AIM model, it has not yet been clarified how organ identification fits into in the developmental process, or how and when infants acquire this capability. Using a robot, Breazeal et al. proposed a model of facial imitation based on the AIM model [15]. In this model, in order to acquire the organ identification ability, the robot learns the relationship between the tracking data of the features of the face of another robot and the joints of its own face while imitating another robot. However, it remains unclear how infants understand that their gestures are the same as those of the person being imitated.

Recent sonographic observations have revealed that the fetus often touches its face with its hands during embryonic weeks 24 and 27 [16]. It is thought that the proposed model can allow a robot to acquire the organ identification ability. Assuming that an infant associates its arm movements with its tactile experiences in the womb, it seems reasonable to hypothesize that the infant has developed a topological relationship among his/her own body parts. As such, after birth, the infant might be able to associate the topological relationships of his/her own body parts with those of their parents. It has been reported that one-month-old infants show a preference for viewing the full face of their mother, but no preference for her profile [17]. This fact implies that younger infants are not aware of their mothers from the side view of the faces. This means that they have not associated the full faces with the side faces yet, and in this sense it can be said that their recognition remains a planar one.

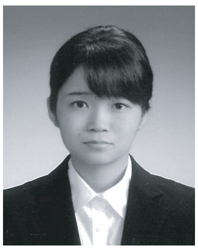
However, there still remain a number of obstacles to realizing organ identification. One is how to find the unit of organ. For visual information, it is known that infants have a preference for patterns like faces [18]. However, how do they feel for tactile sensing? One possibility is that the organs have high sensitivity and thus can be easily separated from other tactile sensors. The other possibility is that the irregularities on the facial surface, such as nose, mouth, and eyes, can be easily sensed with tactile sensation of the hand. We are now constructing a model with more accurate facial structure and tactile sensation for investigating these two possibilities. Another challenge for realizing organ identification is making an appropriate evaluation function to recognize the topographic relationship between organs. The proposed method achieves mapping between the different modalities, enabling a robot to compare different kinds of modal signals. However, this mapping is not so accurate and thus will need geometrical evaluation such as the topological relationship between organs rather than exact pattern matching.

In the present paper, we proposed a learning model to acquire body images for invisible body parts. The invisible hand position is estimated based on the Jacobian between the displacement of the joint angles and the optical flow of the hand. A general idea is to use the Jacobian and its integration for estimating the invisible space. For future work, we are planning to extend the method to acquire the body image of the back.

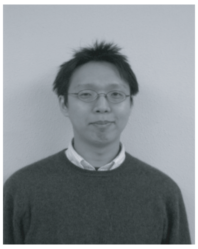
References

1. K. Hosoda and M. Asada, Versatile Visual Servoing without Knowledge of True Jacobian, in *Proc. of IEEE/RSJ/GI Int. Conf. on Intelligent Robots and Systems (IROS)* (Munich, Germany, 1994), pp.186–193.
2. D. Bullock, S. Grossberg and F. H. Guenther, A self-organized neural model of motor equivalent reaching and tool use by a multijoint arm, *Journal of Cognitive Neuroscience* **5**(4) (1993) 408–435.

3. C. G. Sun and B. Scassellati, A Fast and Efficient Model for Learning to Reach, *International Journal of Humanoid Robotics* **2**(4) (2005) 391–414.
4. V. S. Ramachandran and S. Blakeslee, *Phantoms in the Brain: Probing the Mysteries of the Human mind.* (William Mollow, New York, 1998).
5. A. Iriki, M. Tanaka, S. Obayashi and Y. Iwamura, Self-images in the video monitor coded by monkey intraparietal neurons, *Neuroscience Research* **40** (2001) 163–173.
6. A. N. Meltzoff and M. K. Moore, Newborn infants imitate adult facial gestures, *Child Development* **54** (1983) 702–709.
7. A. N. Meltzoff and M. K. Moore, Explaining Facial Imitation: A Theoretical Model. *Early Development and Parenting* **6** (1997) 179–102.
8. J. L. Hopson, Fetal Psychology, *Psychology Today, Sep/Oct, 1998* **31**(5) 44.
9. H. Eswaran, J. Wilson, H. Preissl, S. Robinson, J. Vrba, P. Murphy, D. Rose and C. Lowery, Magnetoencephalographic recordings of visual evoked brain activity in the human fetus, *The Lancet* **360**(9335) (2002) 779–780.
10. C. Nabeshima, M. Lungarella and Y. Kuniyoshi, Timing-Based Model of Body Schema adaptation and its Role in Perception and Tool Use: A Robot Case Study, in *Proceedings of the 4th IEEE Int. Conf. on Development and Learning (ICDL)* (Osaka, Japan, 2005), pp. 7–12.
11. Y. Yoshikawa, H. Kawanishi, M. Asada and K. Hosoda, Body Scheme Acquisition by Cross Modal Map Learning among Tactile, Visual and Proprioceptive Spaces. *Proceedings of the 2nd Int. Workshop on Epigenetic Robotics* (Edinburgh, Scottland, 2002) pp. 181–184.
12. B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser and M. Moser, Path integration and the neural basis of the 'cognitive map', *Nature Reviews neuroscience* **7** (2006) 663–678.
13. T. Kohonen, *Self-Organizing Maps*, (Springer-Verlag Verlin Heidelverg, 1995).
14. D.E.Rumelhart, G.E. Hinton and T.J. Sejnowski, Learning representation by back-propagation errors, *Nature* **323** (1986) 533–536.
15. C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, B. Blumberg, Learning from and about others: towards using imitation to bootstrap the social understanding of others by robots, *Artificial Life* **11** (2005) 1–32.
16. S. Campbell. *Watch Me Grow, A Unique, 3-Dimensional Week-by-Week Look at Your Baby's Behavior and Development in the Womb*, (Carroll & Brown Publishers, London, 2004).
17. F. Sai and I. W. R, Bushnell, ; The Perception of Faces in Different Poses by One-Month-Olds. *British Journal of Developmental Psychology* **6** (1988) 35–41.
18. M. H. Johnson Subcortical face processing, *Nature Review Neuroscience* **6** (2005) 766–774.



Sawa Fuke received the B.E. degree from the University of Electro-Communications, Tokyo, Japan, in 2005. She is currently a Master's course student in the Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University. Her research interests are body-image, spatial perception using visual information, and sensor fusion.



Masaki Ogino received the B.S., M.S. and Ph.D. degrees from Osaka University, Osaka, Japan, in 1996, 1998, and 2005 respectively. From 2002 to 2006, He was a Research Associate in the Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University. He is currently a researcher in 'ASADA Synergistic Intelligence Project" of ERATO (Exploratory Research for Advanced Technology by Japan Science and Technology Agency).

His research interests are humanoid robot control, biped walking, and cognitive issues involving humanoid robots.

Minoru Asada received the B.E., M.E., and Ph.D., degrees in control engineering from Osaka University, Osaka, Japan, in 1977, 1979, and 1982, respectively. From 1982 to 1988, he was a Research Associate of Control Engineering, Osaka University, Toyonaka, Osaka, Japan. In April 1989, he became an Associate Professor of Mechanical Engineering for Computer-Controlled Machinery, Osaka University, Suita, Osaka, Japan. In April 1995, he became a Professor of the same department.

Since April 1997, he has been a Professor of the department of Adaptive Machine Systems at the same university. From August 1986 to October 1987, he was a visiting researcher at the Center for Automation Research, University of Maryland, College Park, MD. Since 2002, he has been the president of the International RoboCup Federation. In 2005, he was elected a Fellow of the IEEE for contributions to robot learning and applications. In the same year, he was also elected the research director of "ASADA Synergistic Intelligence Project" of ERATO (Exploratory Research for Advanced Technology by Japan Science and Technology Agency)