Is it my body? - body extraction from uninterpreted sensory data based on the invariance of multiple sensory attributes -

Yuichiro Yoshikawa[1], Yoshiki Tsuji[1], Koh Hosoda[1,2], Minoru Asada[1,2]
[1]Dept. of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, Osaka, Japan
[2]HANDAI Frontier Research Center, Osaka University, Osaka, Japan
Email: {yoshikawa, tsuji}@er.ams.eng.osaka-u.ac.jp, {hosoda, asada}@ams.eng.osaka-u.ac.jp

***Abstract*— Finding the body in uninterpreted sensory data is one of the fundamental competences to construct the body representation that influences on adaptabilities of the robot to the changes in the environment and the robot body. The invariance of sensations in self-observation seems a promising key information to find the body. However, since each sensory attribute can be invariant only in the observation of a part of the body, the robot should complementarily utilize the invariance of the multiple sensory attributes. In this paper, we propose a method of body-nonbody discrimination by complementarily utilizing multiple sensory attributes based on a conjecture about the distribution of the variance of sensations for each observing posture, where it can be approximated by a mixture of two Gaussian distributions which are for observing the body and the nonbody, respectively. By estimating the distribution, the robot can automatically find a discrimination hyperplane to judge whether it observes its body in the current observing posture. Simple experiments show the validity of the proposed method.**

I. INTRODUCTION

Adaptability to the changes in the environment and the robot body fundamentally depends on the robot body representation that is often given by the designer. However, it is difficult for the designer to prepare the universal body representation that can be applied to any kinds of robots since the designing process involves understanding robot's embodiment (i.e., how the robot is embodied) that depends on its body structure, the environment, the sensory attributes, and so on. Therefore, a robot should possess the competence to acquire its body representation from its uninterpreted sensory data by itself. On the other hands, the representation of the body in human beings, that is so-called *body scheme* or *body image*, is one of the hottest topics in the brain science [1], [2] but the acquisition process of it has not been revealed yet. Therefore, building a robot that can acquire its body representation by itself is an interesting issue from the viewpoint of a constructivist approach towards both establishing the design principle for an intelligent robot and understanding human intelligence [3].

In the previous work on acquiring body representation of the robot, the designer usually gives *a priori* knowledge to find a part that can be interpreted as the sensation of the body in the sensory data [4], [5], [6], [7]. However, finding the body in its uninterpreted sensory data seems the first step to acquire body representation since how to find it depends on the robot's embodiment. Although the method based on the correlation of an optic flow with the motion is proposed to find the position of the end-effector [8], it is not clear whether it works in the case where the robot moves both its end-effector and its camera head, in other words, where the robot is not given *a priori* knowledge of the physical meanings of its DoFs. Without such *a priori* knowledge, Yoshikawa et al. have proposed a method to find the body based on the invariance of a sensory attribute with its observing posture [9]. Since they implemented the idea based on Hebbian learning between the nodes for the external sensor and the robot's proprioception, it is limited to utilize the invariance of one single sensory attributes. However, since each attribute can be invariant only in the observation of a part of the body, the robot should complementarily utilize the invariance of the multiple sensory attributes in parts of self-observation.

In this paper, we propose a method of body-nonbody discrimination by complementarily using multiple sensory attributes based on the invariance of sensations in self-observation. Sensations can be caused by observing the body or by observing the external world while the sensations of the body involves less variance than one of the external world. Therefore, it is conjectured that the distribution of the variance of sensations for each observing posture can be approximated by a mixture of two Gaussian distributions which are for observing the body and the nonbody. After estimating the distribution via the EM algorithm [10], the robot can automatically find a discrimination hyperplane to judge whether it observes its body in current observing posture by applying the linear discriminant method [11].

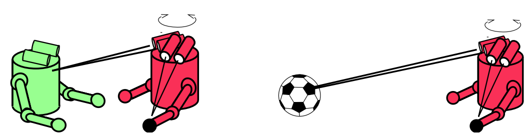
In the rest of this paper, first we revisit the basic idea of the invariance in self-observation that is proposed in the previous work [9] and its limitation. Then, to perform automatic body-nonbody discrimination with the multiple sensory attributes, we introduce the mixture of Gaussian distributions as a model of the invariance in self-observation. By using a test-bed robot with plural types of texture on its body surface, we show the experiments to test the proposed method work.

II. An approach to finding self body based on the invariance in self-observation

On the issue how to find the body of the robot from uninterpreted sensory data, it seems reasonable to follow the idea in the previous work that the sensations of its body are invariant with its observing posture [9] although it is limited to utilize the invariance of only one sensory attribute. In this section, first we describe the assumptions in the problem to find the body from uninterpreted sensory data, and then revisit the previous approach [9] and point out its limitation.

a) The problem and the assumptions: The robot observes its body or the external world with multiple sensory attributes and can shift its focus of attention by changing its posture. Although discrimination between proprioceptive sensors and external ones is one of formidable issues as addressed by Philipona et al. [12], here we assume that the robot can distinguish them as a starting point. It perceives the current posture but does not know the physical meanings of each DoF, namely, it can not distinguish the DoFs for the camera head and ones for the arm. In these assumptions, the problem to be attacked is to learn to judge whether the sensation in the current posture is caused by observing its body or by observing the external world. For the simplicity of the results, we concentrate on the visual attributes at the center region of its camera image in this paper.

b) The basic idea of the previous work and its limitation: Sensations by an external sensor, that is a camera in this paper, are related with the robot's proprioception since the camera is embedded in its body. For example, when it fixates one object in the environment, the view changes depending on the environmental changes (see Fig. 1). However, when it fixates its body, the view is independent of the environment. The basic idea of the previous work to find the body is learning the invariance of the relationship between the sensory data of the external sensor and one of its proprioception. In other words, the robot judges what is always observed as its body.



(a) (b)

Fig. 1. The invariance/variance of the relationship among view and the observing posture in different environments ((a) and (b)).

The learning mechanism was implemented by adopting Hebbian learning between the nodes which are sensitively activated to a specific sensor signal both from the external sensors and the proprioceptive ones. Therefore, the connection weights between the nodes of the proprioception and the external sensor which are activated in self-observation increase while ones not in self-observation decrease. Although the robot can judge whether the fixated object is its body based on the amount of the learned connection weights, the designer must prepare different thresholds for different robot's embodiment.

Furthermore, they do not cope with the case where the robot has multiple sensory attributes. However, each attribute can be invariant only in the observation of a part of the body since the sensing process depends on the relationship between the type of the sensory attributes and the fixated object. For example, the luminance pattern of a body part with a fine texture easily varies with the translation of the fixating point while the disparity of one with a repeated texture easily varies with mismatching between views. Therefore, the robot should use the multiple sensory attributes so that they complementarily work for varieties of textures on the robot body. Furthermore, compared to the case with a single attribute, it is more complicated to determine the threshold to find the body in the case with multiple ones since how to let them complementarily work should be considered.

III. Body-nonbody discrimination with multiple sensory attributes

To discriminate the body and nonbody with multiple sensory attributes based on the invariance in self-observation, we need to introduce a new way to complementarily use multiple attributes complementarily. For this purpose, we conjecture that the distribution of the variance of sensations for each observing posture can be approximated by a mixture of two Gaussian distributions.

A. Mixture of Gaussian distribution model of the observing variance of sensations

Suppose that the robot can observe the fixated object with D types of the sensory attributes such as disparity, luminance patterns, chroma, and so on. It moves the fixating point by changing its posture $\boldsymbol{\theta} \in \Re^N$ which is measured by the encoders. Let the i-th sensory attribute in $\boldsymbol{\theta}$ be $\boldsymbol{x}_i(\boldsymbol{\theta}) \in \Re^{M_i}, (i = 1, \cdots, D)$.

The robot can measure $\sigma_i^2(\boldsymbol{\theta}_j)$ that is the variance of the sensation with i-th attribute for each posture $\boldsymbol{\theta}_j, (j = 1, \cdots, q_\theta)$ where $\boldsymbol{\theta}_j$ is the j-th posture in which the posture is quantized in q_θ postures. We define observing variance vector

$$\boldsymbol{z}(\boldsymbol{\theta}) = [\tilde{\sigma}_1(\boldsymbol{\theta})^2, \cdots, \tilde{\sigma}_D(\boldsymbol{\theta})^2]^T \in \Re^D, \tag{1}$$

where $\tilde{\sigma}_i^2(\boldsymbol{\theta}_j)$ is a normalized variance of the i-th attribute in j-th posture $\sigma_i^2(\boldsymbol{\theta}_j)$.

Since sensation can be caused by observing the body or by observing the external world while the sensations of the body involves less variance than one of the external world according to the previous work [9], it is conjectured that the distribution of observing variance vectors can be regard as a mixture of two Gaussian distributions which are for observing the body and the nonbody, respectively (see Fig. 2). In other words, the distribution of $\boldsymbol{z}$ is given by

$$p(\boldsymbol{z}; \alpha) = w_b \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) + w_e \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_e, \boldsymbol{\Sigma}_e) \tag{2}$$

where $\mathcal{N}(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denote a normalized Gaussian distribution of $\boldsymbol{z}$ with the average $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ and suffices b and e indicate the body and the environment, respectively, and $\alpha = \{w_b, \boldsymbol{\mu}_b, \Sigma_b, w_e, \boldsymbol{\mu}_e, \Sigma_e\}$. The weights w_b and w_e are satisfied

$$\begin{aligned} w_b + w_e = 1, \\ 0 \leq w_b, w_e \leq 1. \end{aligned} \tag{3}$$

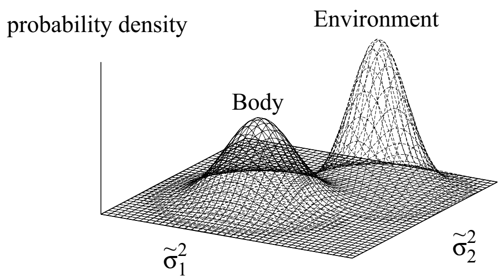


Fig. 2. Mixture of Gaussian distribution model of observing variace vector

B. Estimation of the distribution

Since the robot can measure $\boldsymbol{z}(\boldsymbol{\theta})$ but does not know which of two distributions generates the measured $\boldsymbol{z}(\boldsymbol{\theta})$, it must estimate the distribution (eq. 2) from the incomplete data, that is $Z = \{\boldsymbol{z}(\boldsymbol{\theta}_1), \cdots, \boldsymbol{z}(\boldsymbol{\theta}_{q_\theta})\}$. We can apply the EM algorithm [10] to this kind of problem, which is a theoretical paradigm to estimate the maximum likelihood parameters from an incomplete data.

According to the EM algorithm, to obtain the parameters that maximize the logarithmic likelihood function such as

$$\mathcal{L} = \log p(Z|\alpha), \tag{4}$$

the expectation process and the maximization process are iterately performed until they converge for given initial parameters. In the expectation process, the expectation of the logarithmic likelihood function of the complete data in a given condition of Z and $\alpha^{(t)}$,

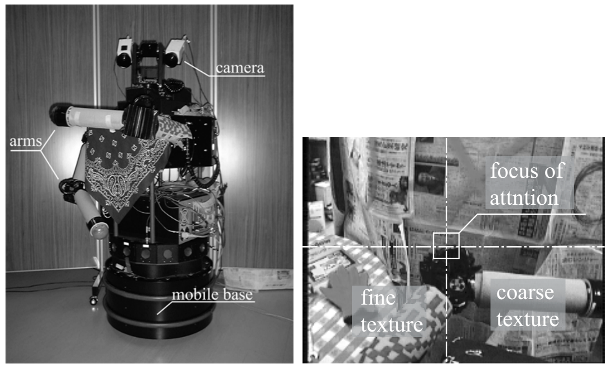
$$Q(\alpha|\alpha^{(t)}) = E_Z\{\log p(Z, H|\alpha)|Z, \alpha^{(t)}\} \tag{5}$$

is calculated where $\alpha^{(t)}$ is the estimated parameter set until the t-step and H is a set of hidden parameters that identify which of two distributions generates $\boldsymbol{z}(\boldsymbol{\theta})$. In the maximization process, α is updated so that the new α maximize $Q(\alpha|\alpha^{(t)})$. It is guaranteed that each iteration of the expectation and maximization process of the EM algorithm increases the logarithm likelihood function [10].

C. Body-nonbody discrimination with the estimated distribution

For the sake of body-nonbody discrimination, that is judging whether the current sensation in $\boldsymbol{\theta}$ is caused by the self-observation, the robot judges which of two distributions more likely generates a given $\boldsymbol{z}(\boldsymbol{\theta})$. We apply the Fisher's linear discriminant method [11] to judge the cause of $\boldsymbol{z}(\boldsymbol{\theta})$. The discrimination function is described as

$$g(\boldsymbol{z}(\boldsymbol{\theta})) = \boldsymbol{\kappa}^T \boldsymbol{z}(\boldsymbol{\theta}) + \kappa_0 \tag{6}$$



(a) The robotic test-bed

(b) An egocentric view of the robot

Fig. 3. The robotic test-bed: (a) the whole body and (b) an egocentric view

where $\boldsymbol{\kappa}$ is the vertical vector of the discriminant hyperplane, $g(\boldsymbol{z}(\boldsymbol{\theta})) = 0$, and κ_0 is the offset of this hyperplane. According to the linear discriminant method, $\boldsymbol{\kappa}$ to discriminate two Gaussian distributions is satisfied with

$$\boldsymbol{\kappa} \propto (\omega_b \boldsymbol{\Sigma}_b + \omega_e \boldsymbol{\Sigma}_e)^{-1}(\boldsymbol{\mu}_b - \boldsymbol{\mu}_e). \tag{7}$$

Selecting κ_0 is equivalent to determine the point where the discriminant hyperplane intersects. Here, we select it so that the discriminant hyperplane intersects on the middle point of two distributions, therefore

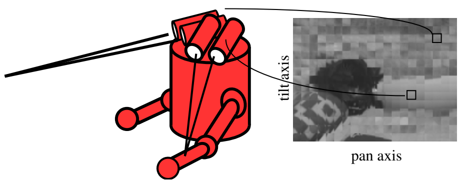
$$g(\boldsymbol{z}) = (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^T(\omega_b\boldsymbol{\Sigma}_b + \omega_e\boldsymbol{\Sigma}_e)^{-T}(\boldsymbol{z} - \frac{\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2}{2}), \tag{8}$$

where $\boldsymbol{A}^{-T}$ indicates the transpose of the inverse matrix $\boldsymbol{A}$ and $\{\hat{w}_b, \hat{\boldsymbol{\mu}}_b, \hat{\boldsymbol{\Sigma}}_b, \hat{w}_e, \hat{\boldsymbol{\mu}}_e, \hat{\boldsymbol{\Sigma}}_e\}$ are the estimated parameters via the EM algorithm. Finally, the robot can regard what it perceives in $\boldsymbol{\theta}$ is its body if $g(\boldsymbol{\theta}) > 0$.

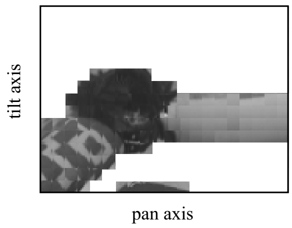
IV. Experiments

In this section, we show a series of experiments using a robotic test-bed (see Fig. 3(a)). To test whether the proposed method works independently of the robot's embodiment, we paste different textures, one is fine and the other is coarse, on the different parts of the arm as shown in an egocentric view of the robot (see Fig. 3(b)). The robot consists of two cameras on the camera head which can rotate in pan and tilt axes, the 4-DoF arms, and the mobile base and detect four visual attributes at the center region of the left camera, namely, disparity, luminance pattern, chroma, and direction of edges. In the following experiments, however, we only show the case where the arm is fixed in a certain posture as shown in Fig. 3(a) for understanding.

It learns the variance of sensations with each attribute for each posture quantum by randomly changing its posture both of its camera head. During the learning process, we let the robot move around to make the external world varies. Fig. 4(a) shows schematic examples of observing posture and the correction of the average luminance patterns at



(a) schematic examples of observing posture and the correction of the average luminance pattern in each posture



(b) the extracted body by human experimenter

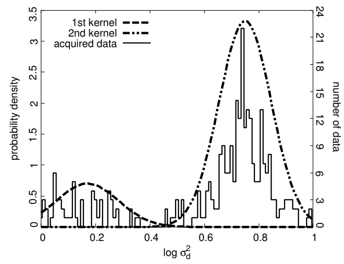
Fig. 4. An schematic explanation of (a)the learning process and (b)the desired extraction of the body

the center region of the left camera for each quantized posture that are arranged in two dimension, pan (the horizontal axis) and tilt (the vertical one). Blank boxes on the correction image are the average luminance patterns in the example of observing posture. Note that the correction image in Fig. 4(a) is slightly different from an egocentric view in Fig. 3(b) since the latter is a entire image of the camera while the former is a correction of a part of the image in various postures of the camera head. The task to be attacked is to extract the posture quanta of camera head in which the robot observes its body. Fig. 4(b) shows the extracted posture quanta with the average luminance pattern by the correct body-nonbody discrimination. Note that although we, human being, can easily distinguish the body and the environment by looking at the correction image in Fig. 4(a) and obtain correct extraction result (Fig. 4(b)), it is formidable for a robot since it is not given any *a priori* knowledge about the difference between its body and the external world.

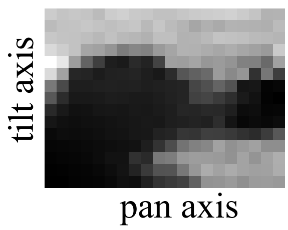
After learning the variance of sensations for each posture, we test how the body-nonbody discrimination by the proposed method extracts the robot's body with a single visual attribute, two of them, and all of them in order. Note that the variances are normalized in the logarithmic scale.

A. Body-nonbody discrimination with single attributes

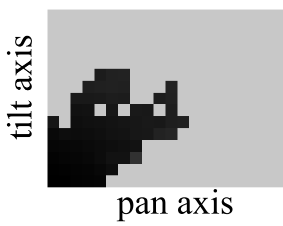
First we test the proposed method with two single attributes, namely, disparity and luminance pattern. Fig. 5 shows the result of the body-nonbody discrimination with disparity. Fig. 5(a) shows the distribution of observing variances of the disparity (solid histogram) and the estimated mixture of Gaussian distributions (broken curve). We can see that the distribution at lower observing variance corresponds to the self-observation while the other corresponds to observing the external world. Fig. 5(b) shows the average disparity in each posture of the camera head, where deeper black indicates larger disparity. Although we can see that the robot detects larger disparity for the posture in which it observes its arm, we cannot preset the threshold of disparity to separate the arm and the external world since it depends on the robot's embodiment. Fig. 5(c) illustrates the average disparity of the extracted body in the posture where the observing variance is regarded to be caused by the distribution of self-observation. By comparing Fig. 4(b) and Fig. 5(c), We can see that the body part with fine texture is correctly extracted while one with coarse texture is not. It seems because the robot tends to fail in stereo matching needed to detect the disparity at the coarse texture.



(a) distribution of observing variance and its estimation



(b) the correction of the average disparity

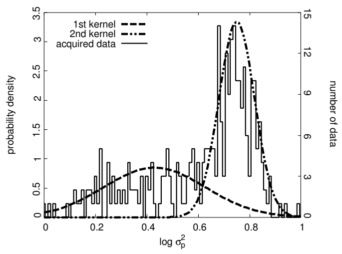


(c) the extracted body

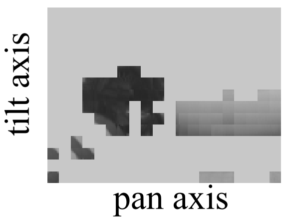
Fig. 5. Body-nonbody discrimination with disparity

Then we test the case using luminance pattern, that is a vector which consists of the luminance of the image elements at the center region (8×8 [pixel]) of the left camera. Fig. 6(a) shows the distribution of variances of the luminance pattern for each posture (solid histogram) and the estimated mixture of Gaussian distributions (broken curve). We can see that the distribution at lower observing variance corresponds to the one for self-observation while the other corresponds to the one for observing the external world. Fig. 6(b) shows the average luminance pattern in each posture of the camera head. Fig. 6(c) illustrates the averaged luminance pattern of the extracted body in the case with the observing variance is regarded to be caused by the distribution of self-observation. By comparing Fig.

4(b) and Fig. 6(c), We can see that the body part with coarse texture is correctly extracted while one with fine texture is not. It seems because the observed luminance pattern of the fine texture sensitively varies with the slight changes of observing posture.



(a) distribution of observing variance and its estimation



(b) the correction of the average disparity

(c) the extracted body

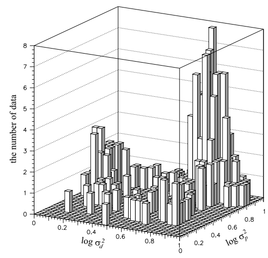
Fig. 6. Body-nonbody discrimination with luminance pattern

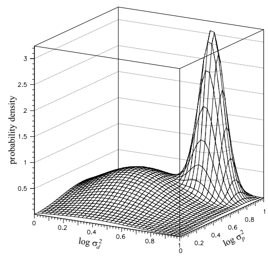
B. Body-nonbody discrimination with two attributes

To show how multiple visual attributes complementarily work to find the body, we experiment with two attributes, namely the disparity and the luminance pattern. Fig. 7(a) shows the distribution of observing variance vectors each of which consists of the variance of disparity and one of luminance pattern for each posture. Fig. 7(b) is the estimated mixture of Gaussian distributions. We can see that the distribution at lower observing variance corresponds to the one for self-observation while the other corresponds to the one for observing the external world. Fig. 7(c) illustrates the average luminance pattern of the extracted body where observing variance is regarded to be caused by the distribution of self-observation. We can see that the body parts both with coarse and fine textures are almost extracted. It is considered that the two attributes make up the loss of the extraction which are the body parts with coarse texture in disparity and one with fine texture in luminance pattern.

C. Body-nonbody discrimination with several attributes

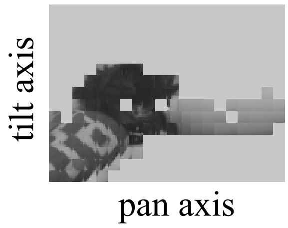
To show whether the proposed method work in the case with several attributes including one by which result in incorrect extraction, we add two more attributes, namely, chroma and direction of edges, each of which is an averaged value at the image element in the center region





(a) distribution of observing variance

(b) the estimated distribution



(c) the correction of the average disparity

Fig. 7. Body-nonbody discrimination with both disparity and luminance patter

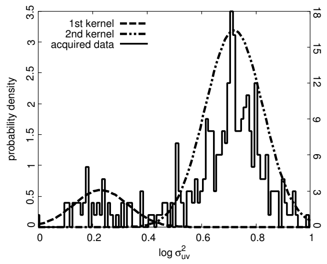
(8×8[pixel]) of the left camera. In the rest of this section, first we show the result of the body-nonbody discrimination with a new single attributes and then one with all attributes.

Fig.s 8 (a) and 9 (a) show the distribution of variances of chroma and direction of edges (solid histogram) and their estimations (broken curve). We can see that the distribution at lower observing variance corresponds to the one for self-observation while the other corresponds to the one for observing the external world as in the previous experiments. Fig. 8(b) and Fig. 8(c) are the average chroma in each posture and the extracted body with the chroma while Fig. 9(b) and Fig. 9(c) show the similar graph of the result with the direction of the edge. From the comparison to Fig. 4(b), we can see that there is some loss of the extraction by chroma at the body part with the fine texture as luminance pattern while there is some incorrect extraction of the external world by the direction of the edge.

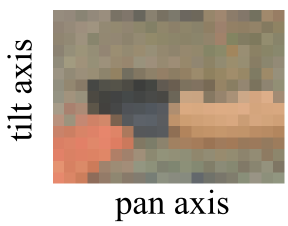
Fig. 10 shows the result of body-nonbody discrimination with all attributes, disparity, luminance pattern, chroma and direction of edges. We can see that the multiple attributes complementarily work to find the body even if there is some incorrect extraction of the external world. Therefore, we can conclude that the proposed method works even if the robot use some not appropriate attributes with which body-nonbody discrimination result in incorrect extraction.

V. Conclusion

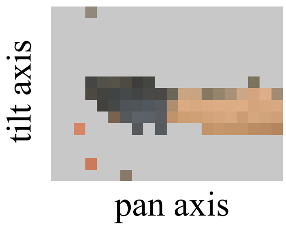
In this paper, we extend the previous study to find the body of the robot from uninterpreted sensory data for



(a) distribution of observing variance and its estimation



(b) the correction of the average disparity



(c) the extracted body

Fig. 8. Body-nonbody discrimination with chroma

the complemental use of multiple attribute based on the conjecture that the distribution of the observing variance can be regarded as a mixture of two Gaussian distributions. By the experiment using a real robot with varieties of textures on the body surface, we confirm that the proposed method works even if the robot use some not appropriate attributes with which body-nonbody discrimination result in incorrect extraction. The competence of body-environment discrimination seems a basis of constructing the representation of the body which 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

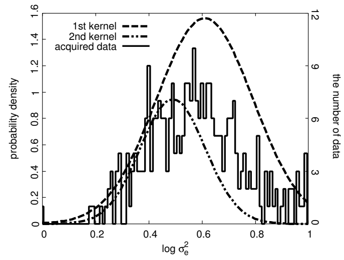
[1] A. Iriki, M. Tanaka, and Y. Iwamura, "Coding of modified body schema during tool use by macaque postcentral neurons," *Neuroreport*, vol. 7, pp. 2325–2330, 1996.

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

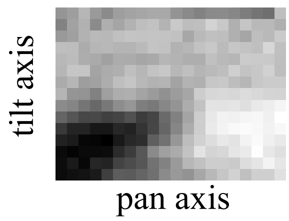
[3] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi, "Cognitive developmental robotics as a new paradigm for the design of humanoid robots," *Robotics and Autonomous System*, vol. 37, pp. 185–193, 2001.

[4] 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 the Second Intl. Workshop on Epigenetic Robotics*, 2002, pp. 181–184.

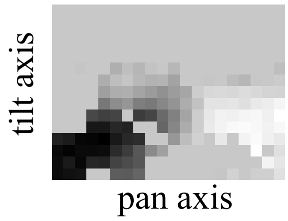
[5] K. F. MacDorman, K. Tatani, Y. Miyazaki, and M. Koeda, "Protosysmbol emergence," in *Proc. of the Intl. Conf. on Intelligent Robot and Systems*, 2000, pp. 1619–1625.



(a) distribution of observing variance and its estimation



(b) the correction of the average disparity



(c) the extracted body

Fig. 9. Body-nonbody discrimination with direction of the edge

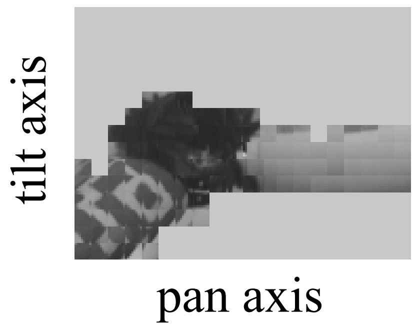


Fig. 10. Body-nonbody discrimination with multiple sensory attributes, disparity, luminance pattern, chroma, and direction of the edge

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

[7] P. Morasso and V. Sanguineti, "Self-organizing body schema for motor planning," *Journal of Motor Behavior*, vol. 27, no. 1, pp. 52–66, 1995.

[8] P. Fitzpatrick and G. Metta, "Toward manipulation-driven vision," in *Proc. of the Intl. Conf. on Intelligent Robot and Systems*, 2002, pp. 43–48.

[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 the 2nd Intl. Symposium on Adaptive Motion of Animals and Machines*, 2003, pp. SaP–II–1.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society*, vol. B-39, pp. 1–38, 1977.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification Second Edition*. John Wiley & Sons, Inc, 2001, ch. Maximum-Likelihood and Bayesian Parameter Estimation.

[12] D. Philipona, J. O'Regan, and J.-P. Nadal, "Is there something out there? inferring space from sensorimotor dependencies." *Neural Computation*, vol. 15, no. 9, pp. 2029–2049, 2003.