The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Own Body Perception based on Visuomotor Correlation

Ryo Saegusa, *Member, IEEE*, Giorgio Metta, *Member, IEEE*, Giulio Sandini, *Member, IEEE*

Abstract— **This work proposes a plausible approach for a humanoid robot to define its own body parts based on the correlation of two different sensory signals: vision and proprioception. The high correlation between the motions in vision and proprioception informs the robot that the visually attractive object is related to the motor function of its own body. When the robot finds the highly motor-correlated object during head-arm movements, the visuomotor cues such as the body posture and visual features are stored in a visuomotor memory. Then, developmentally, the robot defines the motor-correlated objects as the own-body parts without prior knowledge on the body appearances and kinematics. It is also adaptable to extended body parts such as a grasped tool. The body movements are generated by stochastic motor babbling. The visuomotor memory biases the babbling to keep the own-body parts in sight. This memory-based bias towards the own-body parts helps the robot explore the large head-arm joint space. The acquired visuomotor memory is also used to anticipate the own-body image from the motor commands in advance of the body movement. The proposed approach was evaluated with two humanoid platforms; iCub and James.**

I. INTRODUCTION

How can a robot know its own body? This is a fundamental question for embodied intelligence and also the early life of primates. We are able to recognize our body under various conditions; for instance, we naturally perceive our own hands with gloves on. In this sense, it would be reasonable to assume that some parts of our body perception are acquired developmentally through the sensorimotor experiences. Our main interest in this work is to realize a primate-like cognitive system to perceive the own body developmentally. The function of the own-body perception is considered essential for robots to identify the self when interacting with people and objects. Also, it is possible to perceive an extended body when using a tool.

An overview of the proposed approach is depicted in Fig.1. The principal idea is to simply move the body and monitor the correlation of the visual and proprioceptive feedback. Then, the robot defines the motor correlated objects as the own body. When the correlation is high, image cues of a visually attractive region are stored in a visuomotor memory with the proprioceptive information. Since the visual movement and the physical movement of the body parts can be assumed dependent, the level of correlation helps to distinguish the own-body from other objects. This correlation is also useful to anticipate the appearance and the location of the own body in sight.

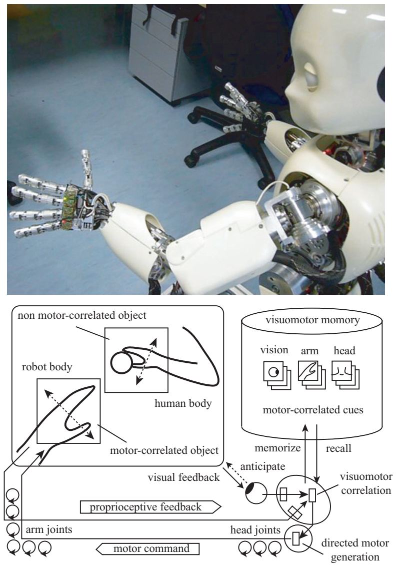

Fig. 1. Visuomotor correlation based own-body definition system. A robot generates arm movements, and senses the visual and proprioceptive feedback. When the feedback is correlated, the robot defines the watching object as its own arm, and memorizes the visuomotor information. After the short-term exploration, the robot anticipates the appearance and the location of the arm by the acquired visuomotor information.

This paper is organized as follows. Section II describes the related works on the body perception in neuroscience and robotics. Section III describes the proposed framework and details of component processes. Section IV describes the experimental results with two humanoid robots; iCub and James. Section V gives conclusions and outlines future tasks.

II. RELATED WORKS

Iriki et al. found in the monkey intraparietal cortex the bimodal (somatosensory and visual) neurons, which seemed to represent the image of the hand into which the tool was incorporated as its extension [1] (Fig.2(a)). This group of the neurons responds the both stimuli from the visual receptive field and the somatosensory receptive field. After the tool use the visual receptive field of these neurons is extended perceptually, as the hand is extended by the tool physically. More recently in [2], they trained the monkey to recognize the image of their hand in a video monitor (Fig.2(b)), and demonstrated that the visual receptive field of these bimodal neurons was projected onto the video screen so as to code the image of the hand as an extension of the self. According to the experimental results, the coincidence of the movement between the real hand and the video-image of the hand

This work was partially supported by European FP7 ICT project CHRIS.

R. Saegusa, G. Metta and G. Sandini are with the Department of Robotics, Brain and Cognitive Sciences, Italian Institute of Technology, Genoa, 16162 Italy, e-mail: ryos@ieee.org, ryo.saegusa@iit.it, giorgio.metta@iit.it, giulio.sandini@iit.it.

978-1-4244-6676-4/10/$25.00 ©2010 IEEE

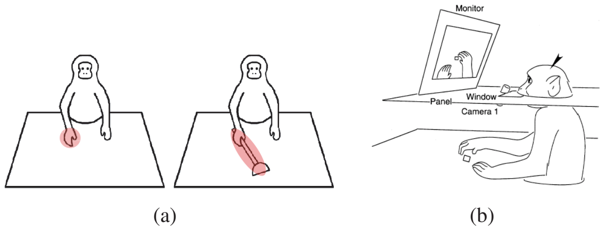

Fig. 2. Body perception of monkeys. (a) Visual receptive field of the bimodal neurons (left: before tool use, right: after tool use) [1][3]. The monkey perceives a tool as an extended body part. (b) The experimental setup of the video-guided manipulation training for a monkey. [2]. The monkeys recognize the image of their hands in a video monitor as an extension of the self.

seemed to be essential for the monkey to use the video-image to guide their hand movements.

In robotics, the sensorimotor coordination is well studied involving neuroscientific aspects and developmental psychology; sensorimotor prediction (Wolpert et al. [4], Kawato et al. [5]), mirror system (Metta et al. [6]), action-perception link (Fitzpatrick et al. [7]), and imitation learning (Schaal et al. [8], Calinon et al. [9]). However, the body detection was often hand coded with predefined rules on body appearances or body kinematics such as visual markers or the joint-link structure. The kind of prior knowledge gives robustness for the body perception, but imposes certain limits too. For instance, a visual model of the hand and the fingers enables the robot to manipulate an object precisely, but once the robot starts to use the object as a tool for another manipulation, it would be difficult for the robot to adapt the body perception to the physically extended hand as the monkeys do it dexterously (Fig.2).

Recently, Stoytchev [10] proposed an approach of video-guided reaching to demonstrate similar tasks to what Iriki et al. examined in [2]. The robot, which was supposed to work on a vertical plane, coordinated its reaching action with a self image projected on a video monitor under the visual transformation. In the experimental setting, the robot was able to identify an object with some different colour markers. This simplification makes us neglect problems in identification of the objects: different appearances of the same object, and similar appearances of different objects. The coincidence of visuomotor information is evaluated by the temporal contingency between the motor command and visual movement; however the time delay in these movements (efferent-afferent delay) must be calibrated in advance, which means that at least the experiment operator must define what the robot hand is.

Hikita et al. [11] proposed a bimodal (visual and somatosensory) representation of the end effector based on Hebbian learning, which simulated the experiments with monkeys in [1]. The visuo-proprioceptive coincidence was evaluated by the contingency between the visual location and hand posture. The approach can be placed as a space-based method compared to the time-based one by Stoytchev. The visual saliency system based on [12] allowed general object detection, although the experiment was not interfered by neither visual disturbance nor sensorimotor noise, since the approach was validated with a robot simulator.

Kemp et al. [13] approached the robot hand discovery utilizing the mutual information between the arm location in joint space and the visual location of the attractive object on the image plane. Here, the mutual information measures statistical dependency between them. The visually detected objects are separated by colour-based off-line image clustering, then the image cluster with high dependency is assigned as the self-image cluster. The proposed approach was validated with a humanoid robot; however the head movements were not considered in the approach. Generally, the head movement affects the motion based object perception, and takes the arm out of sight.

There are other several methods which focus on temporal dependency rather than spatial dependency [14] [15]. The approach by Natale et al. [15] are based on the image differentiation by a periodic (sinusoidal) hand movement, the frequency of which is a robust cue to match the movement of the hand and visually detected object.

Compared to the previous approaches described above, the proposed approach can be characterized as a temporal-coincidence based approach and more condition-tolerant approach to the dynamic change of the camera configuration. The proposed approach does not use prior knowledge for the body definition such as body appearances, kinematics, dynamics, or motor patterns. Instead of these assumptions, the proposed approach only requires general mobility of the body and cross-modal sensing of vision and proprioception to find correlations. Moreover, we will show an enhancement of body search and a benefit for body anticipation. The enhancement in a search problem becomes a popular topic in robotics [16] [17]. The anticipation for motor control is also attracted in the context of internal models [4] [5] [18].

The acquired body images are fundamental memory to identify the end-effector of the robot especially in the learning of visually-guided manipulation [19]. Precise manipulation in a static work cell requires visual markers for reliable hand-eye calibration [20], while we focus rather on the adaptability for the manipulation in a dynamic situation, where the body is supposed to be modified by a tool such the case of the monkeys experiments. Modern techniques of human motion analysis in [21] can be incorporated into the cognitive mechanism of the own-body definition.

III. METHOD

The proposed own-body definition system is outlined in Fig.3. The notation of the most important variables are listed in Table I.

A. Vision

The visual processing is modularized as a set of cascaded image filters, which function in parallel to allow real-time processing. All modules are dually structured for two image streams from the left and right eye cameras, but for the body definition the system simply uses monocular images from

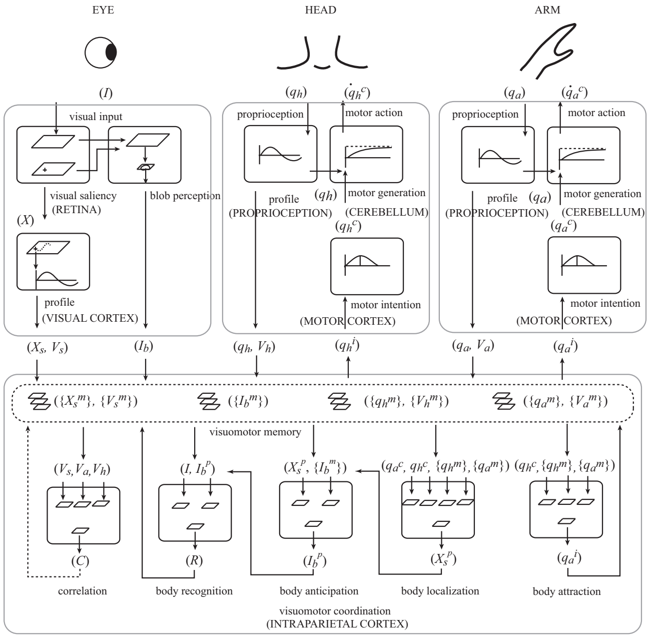

Fig. 3. The own body definition system. The system is composed of the modules; vision, proprioception, motor generation, visuomotor coordination, and visuomotor memory. Each module functions in parallel.

TABLE I

NOTATION OF THE VALIABLES.

t	time of the frame
X	location in sight
X_s	motor-salient point in sight
X_s^p	predicated body location in sight
V_s	speed of the salient location
I_b	motor-correlated image
I_b^p	predicted body image
q_h	head joint angles
q_h^c	motor command of head joint angles
q_h^i	motor intention of head joint angles
V_h	speed of head joint angles
q_a	arm joint angles
q_a^c	motor command of arm joint angles
q_a^i	motor intention of arm joint angles
V_a	speed of arm joint angles

the left-eye camera. Fig.4 illustrates the visual processing procedure. First, the motor-salient point is detected, then a moving blob is extracted. The motor-salient point is traced as a short-term sequence, then the profiles of the position and the speed are given.

The motor-saliency module produces a gray scale image $I_g(X,t)$ at time t from the input colour image by averaging the RGB components at each pixel of location X, then frame subtraction is applied between $I_g(X,t)$ and the previous frame $I_g(X,t-\Delta t)$. The operation is defined as follows:

$$I_f(X,t) = |I_g(X,t) - I_g(X,t-\Delta t)|, \tag{1}$$

where $I_f(X,t)$ denotes the intensity of the subtraction frame at X and t. In the proposed system, the motor-salient point is defined simply as the center of mass such that

$$X_s(t) = \sum_X (X\ I_f(X,t)) / \sum_X I_f(X,t), \tag{2}$$

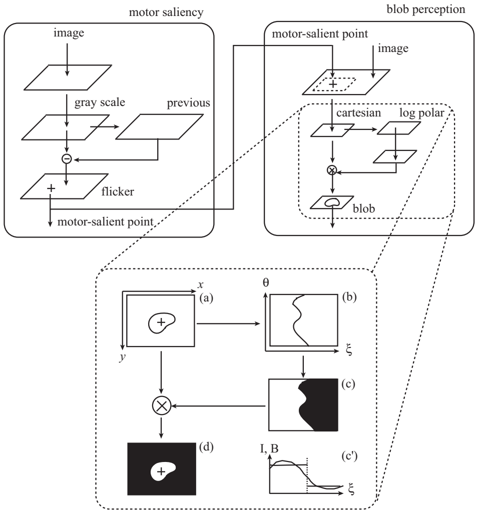

Fig. 4. Visual processing. The motor-saliency module extracts the center of a moving region. The blob perception module grabs a blob located at the motor-salient point by the log-polar transformation.

when the mass $\sum_X I_f(X,t)$ is larger than a threshold. Otherwise, the previous location $X_s(t-\Delta t)$ is given as $X_s(t)$ in order to keep the same location against impulse-like noise. The moving speed V_s is the norm of the velocity defined as:

$$V_s(t) = |\dot{X}_s(t)|, \tag{3}$$

where the upper dot denotes the temporal differential of the variable. In the following formulation, we distinguish the terms of velocity and speed. The speed indicates the norm of the velocity.

A moving blob is extracted based on the motor-salient point. The local region of the input image at the motor-salient point is extracted, then a visual blob is segmented from the region. A colour image $I(x,y)$ presented in a Cartesian coordinate system (x,y) is transformed into the log-polar coordinate system (ξ,θ) as:

$$\xi = \ln\sqrt{x^2+y^2}, \tag{4}$$

$$\theta = \arctan(y/x), \tag{5}$$

where the origin of the Cartesian coordinate system (x,y) is in the center of the image, where a blob is assumed to be located. The log-polar transformation allows us to segment a blob with a curve as illustrated in (b) in Fig.4.

B. Proprioception

In order to realize a visuomotor intelligence, we need a robot platform which has an arm and a head with an eye at least. Fig.5 depicts the child-type humanoid robot used in the following experiments [22][23]. The figure draws the partial joint configuration of the body which we mainly used. In the proposed approach we do not suppose prior knowledge of the body structure such as kinematics and dynamics. Moreover, the system is not aware neither the number of the joints nor body appearances. We assume only that the system is able to distinguish the joint group of the head and the arm.

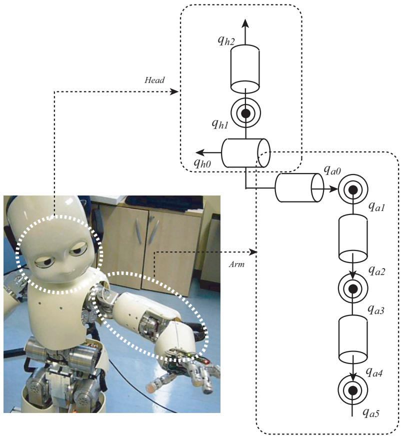

Fig. 5. The body structure of the robot platform iCub [22][23]. The eye, head and left arm of the robot were used in the experiments.

Proprioceptive sensing in a biological system includes many sensory modalities such as tactile, heat, pain, and force sensing. Here, we only use the joint angle given from the joint encoders. The groups of the head joint angles and the arm joint angles are denoted as q_h and q_a, respectively. In the following description, we also use q_p indicating the either joint group ($p = a$ or h). The moving speed of the group of joints are denoted as:

$$V_p(t) = |\dot{q}_p(t)|. \tag{6}$$

C. Motor generation

The motor behaviour of the robot is produced by a biased motor babbling [24]. The motor babbling, which gives random movements of joints, is useful for the robot to explore the learning domain without a structured motor control. The learning domain, however, is huge especially for high degree-of-freedom (DOF) systems such as full-body humanoid robots. For instance, the arm is often located out of sight, when the robot is randomly moving both the head and arm.

The head posture was stationary in the related works as discussed above, while we challenge the own-body definition under natural conditions including head movements without any visual markers. The basic idea to enhance the body search is to bias the randomness of the motor babbling to reduce the search domain.

The motor intention module randomly generates motor commands from the normal distribution, the density function of which is defined as:

$$\mathrm{Prob}(q_p^c) = N(q_p^i, \sigma_p^i), \tag{7}$$

where the mean q_p^i and the deviation σ_k^i are given by the body attraction module in the visuomotor coordination. The arm attraction module gives an arm motor intention coordinated with a head posture, then the lower motor module produces a motor command, which functions to move its arm in sight.

D. Visuomotor coordination

The all sensing data from the eye, head and arm are coordinated in the visuomotor modules. At every moment, the correlation between the speed of a moving blob and the arm proprioception is monitored. The visuomotor correlation is defined as:

$$C(t) = \frac{\sum_{\tau=t-T_c}^{t} V_s'(\tau) V_a'(\tau)}{\sqrt{\sum_{\tau=t-T_c}^{t} V_s'(\tau)^2}\sqrt{\sum_{\tau=t-T_c}^{t} V_a'(\tau)^2}}, \tag{8}$$

$$V_s'(\tau) = V_s(\tau) - \frac{1}{T_c}\sum_{\tau=t-T_c}^{t} V_s(\tau), \tag{9}$$

$$V_a'(\tau) = V_a(\tau) - \frac{1}{T_c}\sum_{\tau=t-T_c}^{t} V_a(\tau), \tag{10}$$

where T_c denotes the size of the sequence. V_s' and V_a' denote the biased values of V_s and V_a by subtracting the average value of each sequence, respectively. $C(t)$ satisfies the formula on the upper and lower boundary;

$$-1 \leq C(t) \leq 1. \tag{11}$$

The visuomotor information described in Fig.3 are memorized, when the visuomotor correlation exceeds a certain threshold. Here, let us define the visuomotor memory as the set of variables; $\{X_s^{m_j}, I_b^{m_j}, q_h^{m_j}, V_h^{m_j}, q_a^{m_j}, V_a^{m_j}\}_{j=1,\cdots,N_l}$. When the capacity of the visuomotor memory reaches the limit N_l, the system forgets the oldest memory and memorizes the new one. Exceptionally, when the head is moving, the visuomotor information is neglected, since the visual motion is always relative to the camera configuration.

E. Visuomotor memory

The visuomotor memory is useful to enhance the body search by directing the motor exploration. Here, we introduce a concept, denoted body attraction. The body attraction is simply realized by recalling an arm position from the acquired visuomotor memory. The robot refers the visuomotor memory and finds the closest head position to the current motor command of the head position q_h^c. Then, the robot recalls the arm position coupled with this closest head position in the memory, and moves the arm towards this position. Since the visuomotor information were memorized when the robot found the motor-correlated object (in most of cases it is the own arm), this association leads the arm into the view field. Note that this motor intention is originated only from the results of sensorimotor exploration.

The motor intention of the arm position is formulated as follows,

$$q_a^i = q_a^{m_k}, \tag{12}$$

$$k = \arg\min_j d_{hj}, \tag{13}$$

$$d_{hj} = |q_h^c - q_h^{m_j}|, \tag{14}$$

where q_a^i denotes the motor intention of the arm, d_{hj} denotes the distance between the motor command of the head position and the jth head position in the visuomotor memory.

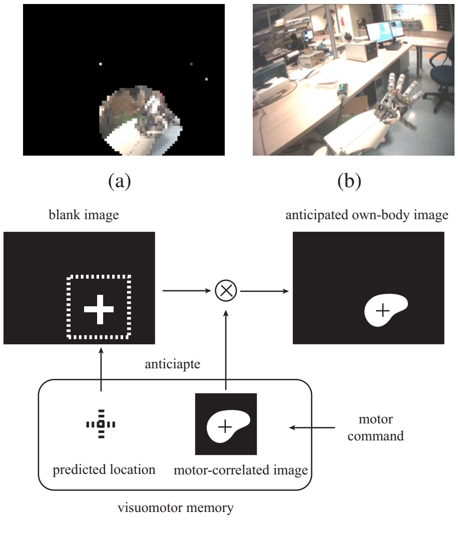

Fig. 6. The snapshot of the body anticipation. (a) the anticipated own-body image I_b^p before a body movement. (b) the input image I after the movement. (c) the anticipation procedure of an own-body image.

This motor intention for the arm is used to generate a motor command as formulated in (7).

The visuomotor memory provides a cue to predict the appearance and location of the body parts in sight. The body posture (the couple of the head and arm posture in this context) geometrically determines the arm location in sight of the robot. However, the arm does not always appear in sight, since the view angle is limited. Also, the arm appearance changes depending on the posture. The visuomotor memory is useful to reconstruct the appearance and location of the own body as a frame of the view, denoted as the own-body image. Moreover, by referring the motor commands, which is the current goal of the body configuration, the visuomotor memory enables the robot to anticipate the own-body image in advance of the movement.

The predicted location of the motor-correlated object in sight is given from the motor commands as follows:

$$X_s^p = X_s^{m_k}, \tag{15}$$

$$k = \arg\min_j d_j, \tag{16}$$

$$d_j = ||[q_h^c, q_a^c] - [q_h^{m_j}, q_a^{m_j}]||, \tag{17}$$

where X_s^p denotes the predicted location of the motor-correlated object (which is the own body if the body definition is successful) in sight. The notation of $[a,b]$ represents the concatenated vector of vector a and b. The anticipated own-body image, denoted as I_b^p, is generated by projecting the motor-correlated image $I_b^{m_k}$ (the body part image in the memory) on the blank frame at the predicted location X_a^p. Fig.6 shows an example of the body anticipation.

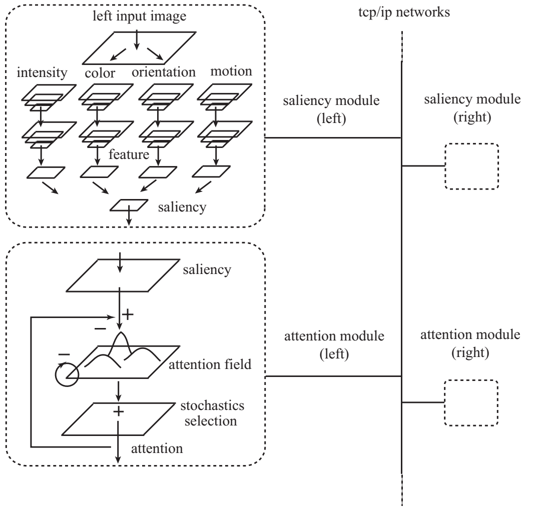

Fig. 7. Visual attention system. The saliency module decomposes an image into basic features as intensity, color, orientation, and motion. The attention module selects an attracted location stochastically depending on the interest of the high level module.

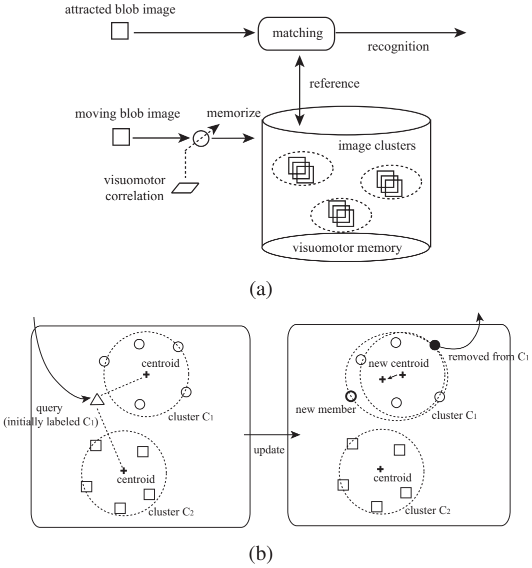

Fig. 8. Visual recognition. (a) The visually attracted region is compared with representatives of motor correlated images, then the own-body is visually recognized. (b) On-line image clustering. The clustering is performed by a modified k-means [26] to allow on-line updating.

The visuomotor memory is also used for visual recognition of own-body parts. Fig.7 illustrates a visually saliency and attention system. The visual saliency in a retina can be modeled as integrated optical difference detectors. Here, we introduce an extended saliency model of [12], which newly includes a motion channel [25]. A visual blob at the attention point is compared with representative images of the motor-correlated images in the visuomotor memory (Fig.8(a)). The representative images are the nearest image of the centroid given by a on-line image clustering. (Fig.8(b)). In the current system, the flexibility of the image scale and rotation is not considered in the recognition, but it can be solved in the pattern matching or the image clustering.

IV. EXPERIMENT

We performed some experiments of the own-body definition with two humanoid robot platforms; iCub and James.

TABLE II

EXPERIMENTAL CONDITIONS.

Identity	Platform	Major condition	Minor conditions
Exp.1(a)	iCub	Body definition	without head, with head
Exp.1(b)	iCub	Body definition	interference
Exp.1(c)	iCub	Body modification	hand wrapping
Exp.1(d)	iCub	Body modification	grasping an object
Exp.2(a)	iCub	Body attraction	{0,25,50,75,100}%
Exp.2(b)	iCub	Body anticipation	-
Exp.3	James	Body recognition	-

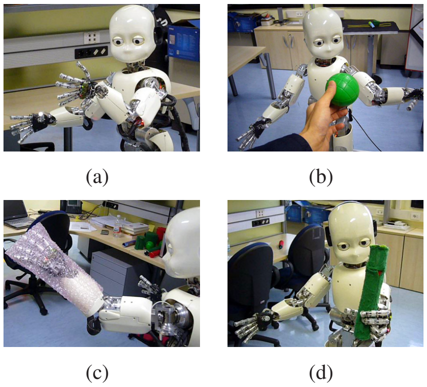

Fig. 9. Experimental scenes in Exp.1. The images of (a)-(d) are the scene of the full-joint movement, human interference, hand wrapping, and stick grasping, respectively.

The contents are listed in Table II.

A. Basic condition

The purpose of the experiments in the basic condition is to validate robustness of the own-body definition against head movements, human interference, and body modification. In the experiment denoted as Exp.1, the robot moved the left arm and/or the head in the random manner by using full joints (a). Then, an experimenter interfered the exploration by presenting a moving object manually (b). Again, the robot moved both the head and the arm as well, while the arm was physically modified by a plastic glove (c) or a grasped object (d). Fig.9 shows the experimental scenes.

Fig.10 shows the snapshots of the profiles during the head and arm movements in Exp.1. The DOF of the head and arm joints for the movements were set as three and six, respectively. The numbers of DOF were the highest DOF configuration. The desired head and arm position were randomly given from the normal distribution of (7) with the constant deviation $\sigma_h^i = \sigma_a^i = 0.3$. The frequency of the head movement (20s) was five-times less than that of the arm movement (4s). The visuomotor coordination module neglected the motion saliency when the head was turning, since the movement of a visually observed object was always relative to the camera movement.

Table III summarizes the time to get the visuomotor memory up to its capacity. The head movements let the body definition slow. Actually, the movements let the arm out of sight, frequently. Experimenter's interference did not affect the time, but reduced the rate of own-body images in the momory. The own-body rate was dependent on the experimenter's moving manner in this case. Basically, the experimenter was requested to move the object randomly in the experiment. On the contrary, when the experimenter mirrored the robot arm movements, the visuomotor correlation was more influenced. That confusion is, however, the reasonable, since the robot defines motor-correlated objects as own-body parts.

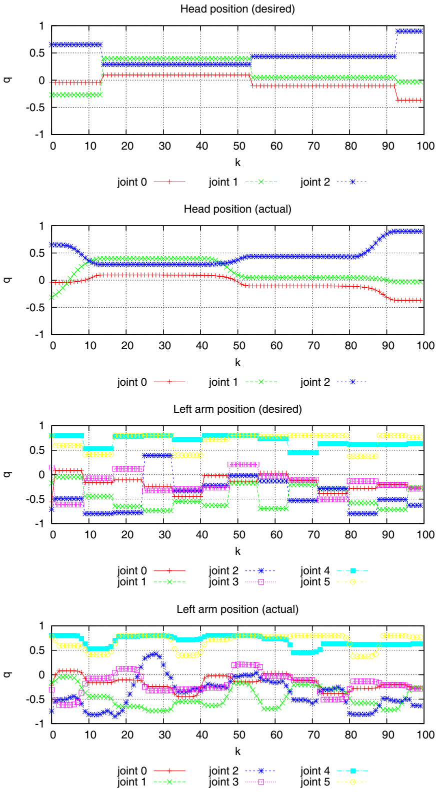

Fig. 10. Snapshots of the profiles during the head and arm movements in Exp.1. All of four profiles plot the values at the same time frames.

TABLE III

OWN-BODY DEFINITION IN THE BASIC CONDITION.

Item	Trial	Capacity	Average	Deviation	Own-body rate
Without head	5	100	351.1	138.1	98.2%
With head	5	100	635.2	155.1	96.2%
Interference	5	100	607.8	79.4	79.8%

Fig.11 shows a set of motor-correlated images acquired in the exploration. The modified body part (wrapped hand) and the extended body part (grasped stick) are successfully defined as own-body parts. These results suggest that the system has potential to the developmental perception of the extended body in the similar manner of the primates [1][2].

B. Advanced condition

The purpose of the experiments in the advanced condition is to exploit the acquired visuomotor memory for body attraction and body anticipation.

In the experiment denoted as Exp.2(a), the body attraction module was activated to attract the robot to move the arm

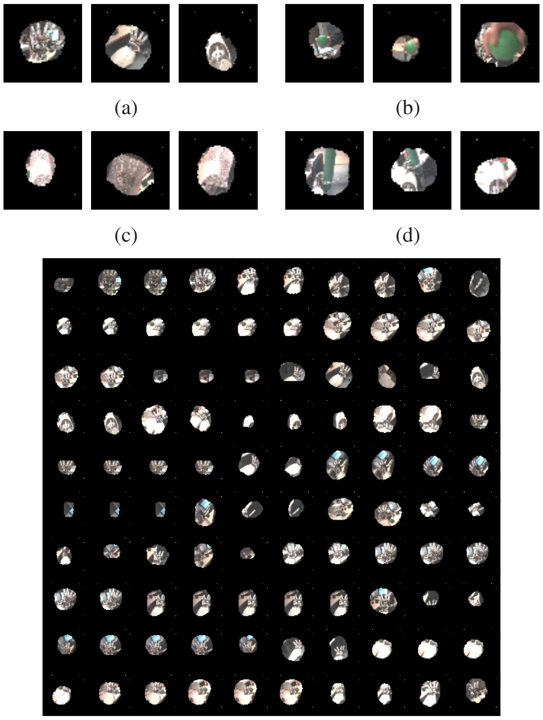

Fig. 11. The motor-correlated images obtained during full-joints exploration in Exp.1. The sets of three images of (a)-(d) are typically sampled in the condition of Fig.9(a)-(d), respectively. Only in (b), the fault samples are presented (objects presented by an experimenter). The bottom 10x10 images are all images obtained during one full-joint exploration in the condition of Fig.9(a).

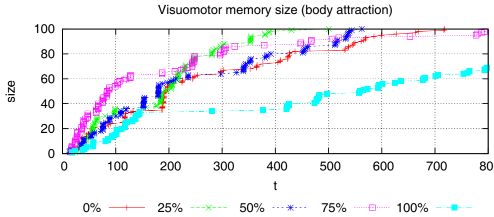

Fig. 12. The evolution of the visuomotor memory in Exp.2(a). The plots correspond to each probability of body attraction (0%: no body attraction, 100%: complete body attraction).

into sight during visuomotor exploration. The body attraction is statistically controlled by the rate to apply the narrow (local) or wide (global) normal distribution to generate the motor intention. Fig.12 plots the evolution of the visuomotor memory against the body attraction rate. According to the result, the middle levels of the probability (25% and 50%) showed a tendency to enhance the body definition more. The condition is considered to balance the global and local search. In the future, sensorimotor exploration of the robot should be designed not only for spacial exploration but also pattern exploration of movements, as the infants develop complex movements through an action and its outcome [27].

In the Exp.2(b), the robot anticipated the own-body image from the visuomotor memory and motor commands. In advance of the anticipation, the robot explored the joint space in the same condition of Exp.1(a) with full-joint head-arm movements, until it acquired 100 recodes of the visuomotor memory. Fig.13 shows snapshots of the body anticipation. The anticipation module gave approximate appearances and locations of the own forearm, successfully. In the experiment, we also got some failure anticipations. The anticipation quality can be improved by a voluntary reconfirmation of the obtained visuomotor memory. The robot can simply reproduce the same configuration of the visuomotor memory and check the results. The non-repeatable memory should be filtered, then.

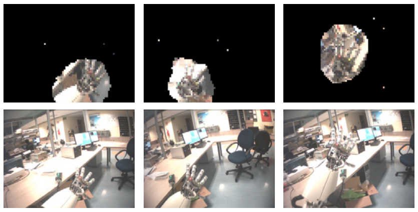

Fig. 13. The snapshot of body anticipation in Exp.2(b). The top and bottom images are the anticipated own-body image before the body movement, and the observed image after the body movement (4.0s later), respectively.

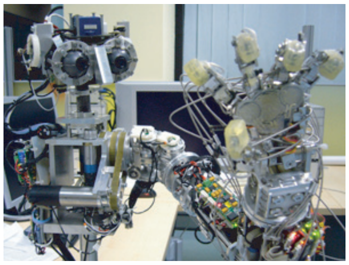

Fig. 14. The humanoid robot James [28] was used for the own-body recognition in Exp.3

C. Another platform

The proposed body definition system is independent from the body structure of a robot. The purpose of the Exp.3 is to experimentally prove it by using the other humanoid robot, James [28]. James is a upper-body robot equipped with binocular vision and other rich of sensors.

In the Exp.3, we performed the body recognition based on the visual saliency and on-line image clustering. Each channel of visual saliency are shown in Fig.15(a). Snapshots of the body recognition are presented in Fig.15(b). As detailed in [25], the motor-correlated objects (own-body parts) and the non motor-correlated objects (other objects) were successfully recognized. The recognition rate was 0.89 for the body parts, and 0.93 for the other objects as shown in Table .IV. The number of clusters was experimentally given, but the optimum number can be determined by some information criterion such as Bayesian Information Criterion (BIC) [29].

V. CONCLUSION

This paper proposed a developmental approach of own-body definition without specific prior knowledge on kine-

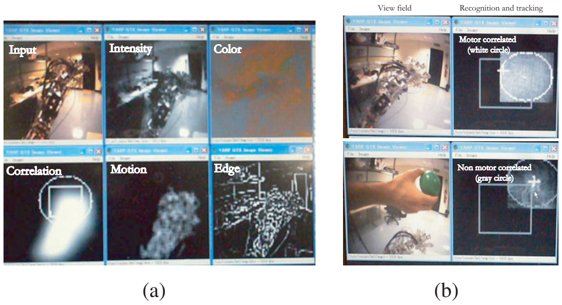

Fig. 15. The snapshots of experimental scenes with James. (a) each chanel of cisual saliency. (b) Visual recognition of the own body (marked by a white circle) and the other object (marked by a gray circle).

TABLE IV
BODY RECOGNITION RATE

Item	Average	Deviation
Own-body parts	0.89	0.19
Other objects	0.93	0.27

matics, dynamics and body appearances. The visuomotor correlation allows the robot to define its own body through the sensorimotor exploration. The robustness of the body definition against the body modification and human interference was experimentally proved. Also the applications of the body definition for the body attraction, body anticipation and body recognition were discussed.

The current body definition system has potential to the binocular perception, but it is not yet examined experimentally. The depth sensing should be included to acquire the three dimensional model of the own body, which is essential for calibrating the environment by the own body. We are also motivated to connect the proposed own-body definition to the learning-based reaching [24] without visual markers.

Another aspect which we should encompass is the haptic information such as tactile and force/torque sensing. In order to distinguish extended body parts from inherent body parts, the haptic information plays an important role. It enables the robot to perceive the extension level of the body structure, and make use of the extended body in a cognitive manner.

References

[1] A. Iriki, M. Tanaka, and Y. Iwamura, "Coding of modified body schema during tool use by macaque postcentral neurones," *Neuroreport*, vol. 7(14), pp. 2325–30., 1996.

[2] A. Iriki, M. Tanaka, S. Obayashi, and Y. Iwamura, "Self-images in the video monitor coded by monkey intraparietal neurons," *Neuroscience Research*, vol. 40, pp. 163–173, 2001.

[3] A. Maravita and A. Iriki, "Tools for the body (schema)," *Trends in Cognitive Sciences*, vol. 8(2), pp. 79–96, 2004.

[4] D. Wolpert, Z. Ghahramani, and M. Jordan, "An internal model for sensorimotor integration," *Science*, vol. 269, no. 5232, pp. 1880–1882, 1995.

[5] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion in Neurobiology*, no. 9, pp. 718–727, 1999.

[6] G. Metta, G. Sandini, L. Natale, L. Craighero, and L. Fadiga, "Understanding mirror neurons: a bio-robotic approach," *Interaction Studies*, vol. 7, no. 2, pp. 197–232, 2006.

[7] P. Fitzpatrick, A. Needham, L. Natale, and G. Metta, "Shared challenges in object perception for robots and infants," *Infant and Child Development*, vol. 17, no. 1, pp. 7 – 24, 2008.

[8] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends in Cognitive Sciences*, vol. 3, pp. 233–242, 1999.

[9] S. Calinon, F. Guenter, and B. Aude, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Transactions on system, man, and cybernetics, Part B*, vol. 37, no. 2, pp. 286–298, 2007.

[10] A. Stoytchev, "Toward video-guided robot behaviors," in *Proceedings of the Seventh International Conference on Epigenetic Robotics (EpiRob)*, L. Berthouze, C. G. Prince, M. Littman, H. Kozima, , and C. Balkenius, Eds., vol. Modeling 135, 2007, pp. 165–172.

[11] M. Hikita, S. Fuke, M. Ogino, and M. Asada, "Cross-modal body representation based on visual attention by saliency," in *IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS)*, 2008.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliencybased visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[13] C. C. Kemp and E. Aaron, "What can i control?: The development of visual categories for a robot's body and the world that it influences," in *Proceedings of the Fifth International Conference on Development and Learning, Special Session on Autonomous Mental Development*, 2006.

[14] P. Fitzpatrick and G. Metta, "Grounding vision through experimental manipulation," *Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences*, vol. 361, no. 1811, pp. 2165–2185, 2003.

[15] L. Natale, "Linking action to perception in a humanoid robot: A developmental approach to grasping." Ph.D. dissertation, LIRA-Lab, DIST, University of Genoa, 2004.

[16] P. Robbel, "Active learning in motor control," Ph.D. dissertation, School of Informatics, University of Edinburgh, 2005.

[17] S. Vijayakumar, A. D'Souza, and S. Schaal, "Incremental online learning in high dimensions," *Neural Computation*, vol. 17, no. 12, pp. 2602–2634, June 2005.

[18] S. Nishide, T. Ogata, R. Yokoya, J. Tani, K. Komatani, and H. G. Okuno, "Object dynamics prediction and motion generation based on reliable predictability," in *Proceedings of IEEE-RAS International Conference on Robots and Automation (ICRA2008)*, May 2008, pp. 1608–1614.

[19] H. Hashimoto, T. Kubota, M. Sato, and F. Harashima, "Visual control of robotic manipulator based on neural networks," *IEEE Transaction on Industrial Electronics*, vol. 39, no. 6, pp. 490–496, 1992.

[20] Y. Motai and A. Kosaka, "Hand-eye calibration applied to viewpoint selection for robotic vision," *IEEE Transaction on Industrial Electronics*, vol. 55, no. 10, pp. 3731–3741, 2008.

[21] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: A review," *IEEE Transaction on Systems, Man and Cybernetics, Part C*, vol. 40, no. 1, pp. 13–24, 2010.

[22] P. Fitzpatrick, G. Metta, and L. Natale, "Towards long-lived robot genes," *Robotics and Autonomous Systems*, vol. 56, no. 1, pp. 29–45, 2008.

[23] G. Metta, G. Sandini, D. Vernon, L. Natale, and N. F., "The icub humanoid robot: an open platform for research in embodied cognition," in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, Washington DC, USA, 2008, pp. 50–56.

[24] R. Saegusa, G. Metta, and G. Sandini, "Active learning for multiple sensorimotor coordinations based on state confidence," in *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2009)*, October 11-15 2009, pp. 2598–2603.

[25] ——, "Self-body discovery based on visuomotor coherence," in *Proc. of 3rd International Conference on Human System Interaction (HSI10)*, May 3-8 2010.

[26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley and Sons, New York, 2001.

[27] H. Watanabea and G. Taga, "General to specific development of movement patterns and memory for contingency between actions and events in young infants," *Infant Behavior and Development*, vol. 29, pp. 402–422, 2006.

[28] L. Jamone, G. Metta, F. Nori, and G. Sandini, "James: A humanoid robot acting over an unstructured world," in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 4-6 Dec. 2006, pp. 143–150.

[29] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.