Body Definition Based on Visuomotor Correlation

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

***Abstract*—This work proposes a plausible approach for a humanoid robot to define its own body based on visuomotor correlation. The high correlation of motion between vision and proprioception informs the robot that a visually moving object is related to the motor function of its own body. When the robot finds a motor-correlated object during motor exploration, visuomotor cues such as body posture and the visual features of the object are stored in visuomotor memory. Then, the robot developmentally defines its own body without prior knowledge on body appearances and kinematics. Body definition is also adaptable for an extended body such as a tool that the robot is grasping. The body movements are generated in the manner of stochastic motor babbling, whereas visuomotor memory biases the babbling to keep the body parts in sight. This ego-attracted bias helps the robot explore the joint space more efficiently. After motor exploration, visuomotor memory allows the robot to anticipate a visual image of its own body from a motor command. The proposed approach was experimentally evaluated with humanoid robot iCub.**

***Index Terms*—Body perception, visuomotor coordination.**

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 sensorimotor experiences. Our main interest in this work is to realize a primatelike cognitive system for perceiving own body developmentally. The function of body perception is considered essential for robots to identify their selves when interacting with people and objects. In addition, it allows perceiving 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 visual and proprioceptive feedback. Then, the robot defines motor-correlated objects as its own body. When the correlation is high, visual cues of the attractive region are stored in visuomotor memory with the proprioceptive information. Since the visual movement and the physical movement of the body parts are assumed dependent, the level of correlation helps the robot distinguish its own body from other objects.

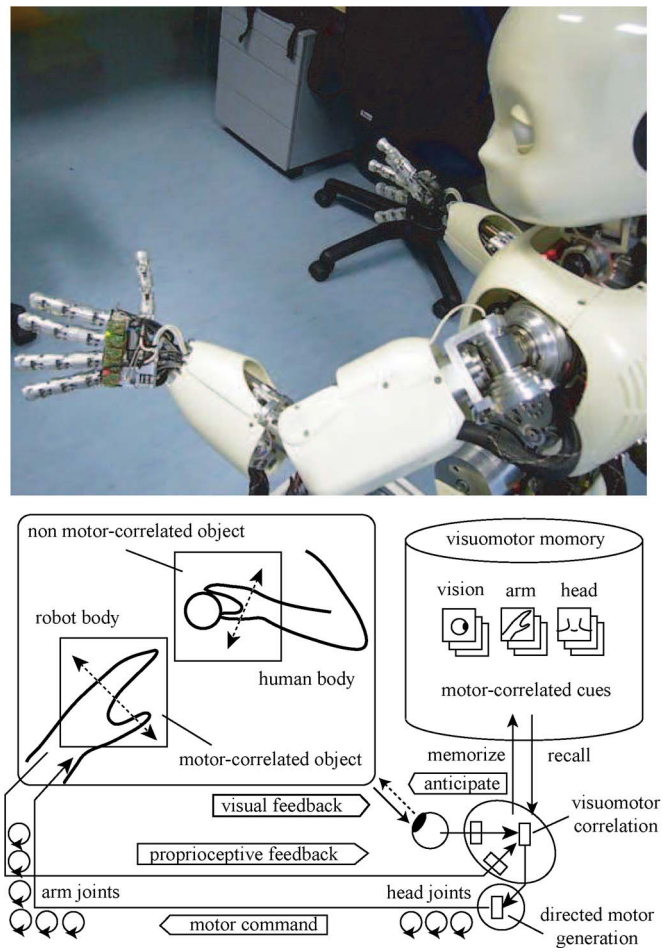


Fig. 1. Body definition system. A robot generates arm movements and senses visual and proprioceptive feedback. When the feedback is correlated, the robot defines the moving object as its own body part and memorizes the related visuomotor information. Through motor exploration, the robot obtains the ability to anticipate visual images of its own body.

This correlation is also useful in anticipating the appearance and location of the body in sight.

This paper is organized as follows: Section II describes related works on body perception in neuroscience and robotics. Section III describes the proposed framework and details of the component processes. Section IV describes experimental results with the humanoid robot iCub. Section V discusses a search problem in body definition. Section VI concludes the work and outlines some future tasks.

II. Related Work

Iriki *et al.* found bimodal neurons (somatosensory and visual neurons) in the intraparietal cortex of monkeys, which incorporated a tool into a mental image of the hand [1]. This

Manuscript received June 13, 2010; revised March 15, 2011; accepted April 20, 2011. Date of publication May 19, 2011; date of current version March 30, 2012. This work was supported in part by the European Union FP7 Project (Cooperative Human Robot Interaction Systems (CHRIS) FP7 215805).

The authors are with the Department of Robotics, Brain and Cognitive Sciences, Italian Institute of Technology, 16162 Genoa, Italy (e-mail: ryos@ieee.org; ryo.saegusa@iit.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2011.2157280

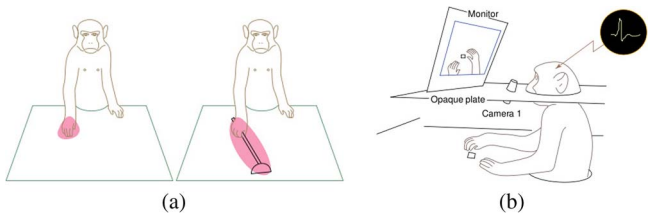


Fig. 2. (a) Visual receptive field of the bimodal neurons (left: before tool use, right: after tool use). The monkey perceives a tool as an extended body part [1]. (b) Video-guided manipulation. After training, the monkey correctly recognizes the hands projected on the monitor as its own hands [2]. The figures were reproduced from [4] under permission.

group of the neurons responds to stimuli from both the visual receptive field and the somatosensory receptive field. After use of the tool, the visual receptive field of these neurons is extended to include the tool [Fig. 2(a)]. More recently, in [2], they trained a monkey to recognize the image of its hands in a video monitor and demonstrated that the visual receptive field of these bimodal neurons was projected onto the video screen [Fig. 2(b)]. Experimental results suggested that the coincidence of the movement between the real hand and the video-image of the hand seemed essential for the monkey to use the video-image for guiding its hands. It is known that this type of manipulative actions is mirrored with the observed similar actions in the premotor cortex of monkeys [3].

In robotics, sensorimotor coordination is well studied, involving neuroscientific aspects and developmental psychology such as sensorimotor prediction [5], [6], mirror systems [7], action–perception links [8], and imitation learning [9], [10]. However, body perception was hand coded with predefined rules on body appearances or body kinematics such as visual markers and the joint-link structure. This kind of prior knowledge gives robustness for body perception but imposes certain limits as well. For instance, it would be difficult for the robot to adapt body perception to the physically extended hand.

Recently, Stoytchev proposed video-guided robot reaching [11], which successfully simulated similar tasks for monkeys in [2]. However, some practical limits still remain, e.g., movements of the robot were constrained on a plane and object identification was neglected by using colored markers. The time delay of temporal contingency between the motor command and the visual movement must be calibrated in advance, which means that the system or the experimenter needs to know what the robot hand is.

Hikita *et al.* proposed a bimodal (visual and somatosensory) representation of an end effector based on Hebbian learning [12], which simulated the experiments with monkeys in [1]. The visual saliency system based on [13] allowed general object detection. However, the approach was evaluated only with a simulator. Then, it is not clear how much visual disturbance and sensorimotor noise interfere in body perception.

Kemp *et al.* approached robot hand discovery by utilizing mutual information between the arm location in joint space and the visual location of the attracted object in sight [14]. The proposed method functioned successfully, even though the robot was interacting with a person; however, it did not consider a head movement, which could interfere in motion-based object perception and push the arm out of sight.

Some other methods focus on temporal dependency rather than spatial dependency [15], [16]. The work of Natale *et al.* [16] was based on image differentiation using periodic (e.g., sinusoidal) hand movements, the frequency of which is a robust cue to match the movement of the hand and the visually detected object.

Compared to previous studies, our method can be a tolerant approach toward dynamic change in camera configuration. It does not require prior knowledge such as body appearances, kinematics, dynamics, or motor patterns. It rather requires mobility and cross-sensory modality. Moreover, we introduce ego-attracted body exploration and body anticipation. Efficient motor exploration and sensory anticipation obtain more important topics in motor learning [17]–[19].

Body information is fundamental to identifying the end effector of robots, particularly when working in visually-guided manipulation [20] with learning mechanisms [21], [22]. For manipulation tasks in a fixed location, the use of visual markers gives an advantage for reliable hand–eye calibration [23], whereas we focus, rather, on adaptability in a dynamic situation, where the body is supposed to be modified by grasping a tool. Modern techniques of visual motion analysis [24], [25] can be incorporated to improve our body definition paradigm.

The ability for body perception is still important for the safety of robots, even if the effector position is given by kinematics. The body perception system allows adaptability for body modification such as coating and exchange of the end effector. It can contribute to tele-operated manipulation [26], [27] and adaptive manipulation [28], [29] in terms of verification of the operator's control and self-diagnosis of motor functions.

III. Method

The body definition system is outlined in Fig. 3. The system is composed of vision, proprioception, motor generation, visuomotor coordination, and visuomotor memory. Each function is detailed in the succeeding sections. The proposed system is an extension of the system in [30].

A. Vision

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

The procedure of image processing is illustrated in Fig. 4. First, a visually salient point is detected; then, a visual blob is extracted from the input image. The salient point is traced as a short-term sequence; then, the profiles of the position and the velocity are given.

In the previous study, visual saliency was modeled as an optical difference between the center and surrounding regions in terms of shape and color cues [13]. In the proposed system,

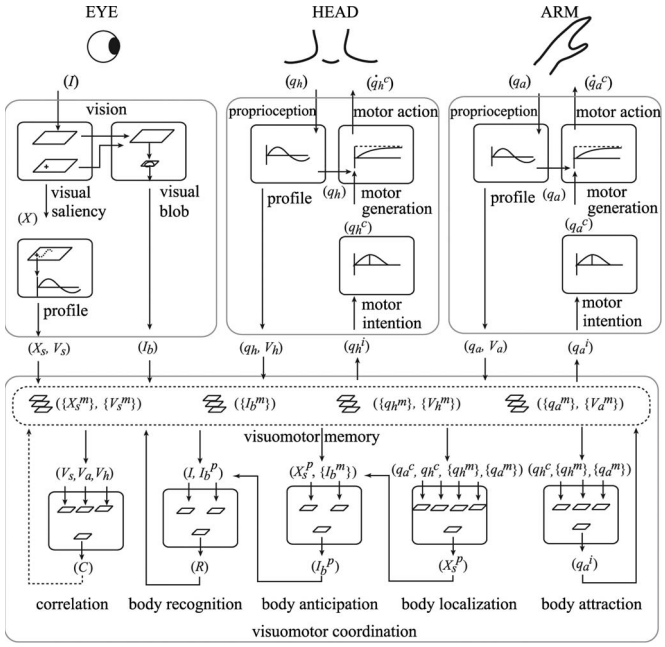


Fig. 3. Cognitive architecture. The system is composed of vision, proprioception, motor generation, visuomotor coordination, and visuomotor memory.

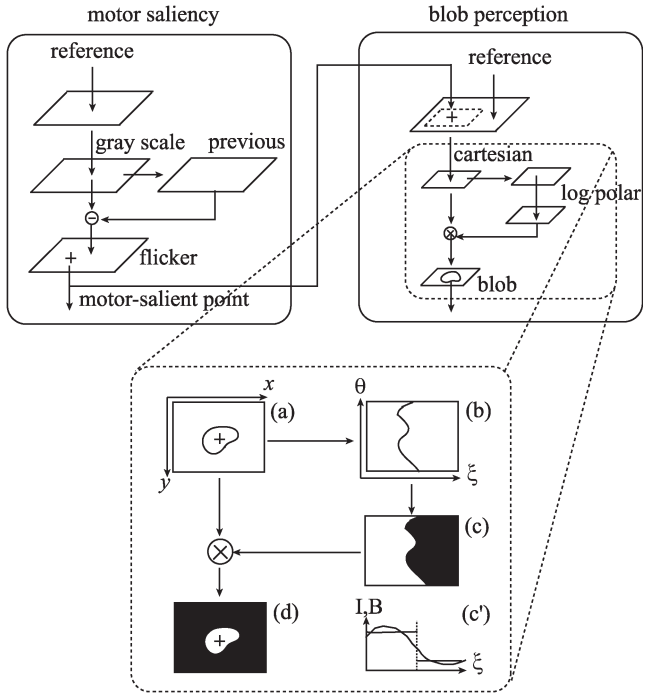


Fig. 4. Visual processing. The saliency module detects a salient point based on a motion cue; then, the blob perception module grabs a visual blob by the log-polar transformation.

we focus on the motion cue. The saliency module produces a gray scale image $I_g(X,t)$ from the input color image by averaging the RGB components. $X=(x,y)$ and t denote the image coordinates and frame time, respectively. Frame subtraction is applied between $I_g(X,t)$ and the previous frame $I_g(X,t-\Delta t)$ 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. When the optical mass $\sum_X I_f(X,t)$ is enough, the center of mass is regarded as the salient point. The coordinates are given by

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

Otherwise, the previous point $X_s(t-\Delta t)$ is given as $X_s(t)$. Moving velocity V_s is the norm of the velocity vector defined as

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

where the upper dot denotes the temporal differential of the variable. In the following formulation, we use the term *velocity* to indicate the norm of the velocity vector.

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

The log-polar transformation allows the system to segment a blob, as illustrated in Fig. 4(b). On the log-polar image, each horizontal line at θ is segmented into two domains, as shown in Fig. 4(c′). We defined the segment border as the curve to minimize the segmentation error on the log-polar image. On each horizontal line at θ, the position of the segmentation point $\xi_s(\theta)$ is given by the following equations:

$$\xi_s(\theta) = \arg\min_{\xi'} E(\xi',\theta) \tag{6}$$

$$E(\xi',\theta) = \sum_{\xi} \{I(\xi,\theta) - B(\xi,\theta;\xi')\}^2 \tag{7}$$

where $E(\xi',\theta)$ denotes the segmentation error when segmented by the point ξ' on the horizontal line at θ. $B(\xi,\theta;\xi')$ denotes the binary function on the line, which has two values corresponding to the left $(\xi \le \xi')$ and right domains $(\xi' < \xi)$, respectively. The value of each domain is the average level in the domain in Fig. 4(c′). The curve drawn by $\xi_s(\theta)$ are smoothed by local averaging and then exploited to segment the blob. The extracted blob image is denoted as the motor-correlated image I_b in Fig. 4(d).

Some examples of the blob perception are shown in Fig. 5. The segmentation is applied to each color channel of R, G, B, and Y and also to intensity channel I. The value of each channel is given by $R=r-(g+b)/2$, $G=g-(b+r)/2$, $B=b-(r+g)/2$, $Y=(r+g)/2-|r-g|/2-b$, and $I=(r+g+b)/3$, where r, g, and b denote the color components

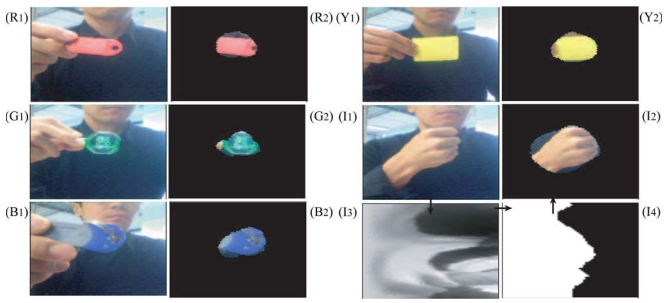


Fig. 5. Blob perception. R_1, G_1, B_1, and Y_1 are the input images, including an object with the particular color of red, blue, green, and yellow, respectively. R_2, G_2, B_2, and Y_2 are segmented images based on the corresponding color channel. I_1 is also an input image but without the particular color. I_2 is the segmented image based on the intensity. I_3 and I_4 are the log-polar images of I_1 and its segmentation, respectively.

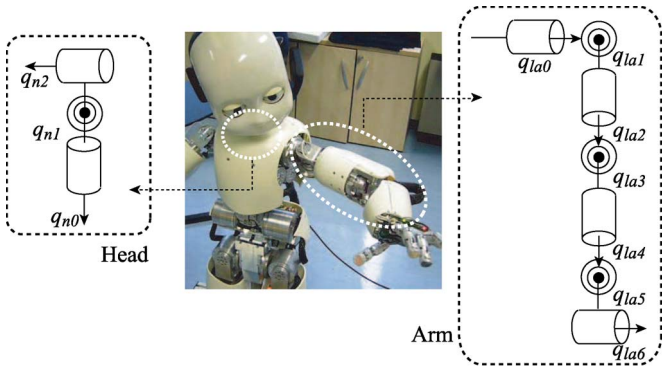


Fig. 6. Body structure of the robot platform iCub [31]. The eye, head, and left arm of the robot were used in the experiments.

of the standard RGB image format. The channel with minimum segmentation error is applied for blob segmentation. Therefore, if the object has a particular color in R, G, B, or Y, it is segmented by this color channel. Otherwise, the intensity channel is applied.

B. Proprioception

The proprioceptive function is based on the humanoid robot platform depicted in Fig. 6 [31]. The figure draws the partial joint configuration of the body that we mainly used. In the proposed approach, we do not suppose prior knowledge of the body structure such as kinematics and dynamics. The system is aware of neither the number of the joints nor the body appearances. We assume only the identity of joint groups to distinguish the head joints from the arm joints.

In biological systems, proprioceptive sensing includes many sensory modalities such as tactile, heat, pain, and force sensing. Here, we only employ joint angles and velocities given by the joint encoders. The groups of the head and arm joint angles are denoted as q_h and q_a, respectively. In the following description, we also use q_p, which indicates either joint group ($p = a$ or h). The velocity of the joints is denoted as

$$V_p(t) = \left|\dot{q}_p(t)\right|. \tag{8}$$

C. Motor Generation

The motor behavior of the robot is produced by biased motor babbling [32]. Motor babbling, which gives random movements of joints, is useful for the robot to explore the learning domain without a structured motor control.

The head posture was stationary in the previous studies, whereas we challenge body definition under natural conditions including head movements. The key idea in enhancing efficiency in body search is to bias the randomness of motor babbling in the search domain. The low-level module, which is denoted as the motor generation module, produces a velocity motor command from the position motor commands. We adopted a simple proportional–integral differential control as formulated in the following:

$$\begin{aligned}\dot{q_p}^c(t) = {} & K_p e\left(q_p^c, q_p, t\right) \\ & + K_i \sum_{\tau=t-T_e}^{t} e\left(q_p^c, q_p, \tau\right) \\ & + K_d\left\{e\left(q_p^c, q_p, t\right) - e\left(q_p^c, q_p, t-\Delta t\right)\right\} \end{aligned} \tag{9}$$

$$e\left(q_p^c, q_p, t\right) = q_p^c(t) - q_p(t) \tag{10}$$

where $\dot{q_p}^c$ denotes the velocity motor command for the robot. $e(q_p^c, q_p, t)$ denotes the position error of the joint angle at time t. K_p, K_i, and K_d denote the constant weight for the proportional, integral, and differential factors. Note that q_p is given by the joint encoder, whereas the position motor command q_p^c is given by the middle-level module denoted as the motor intention module.

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

$$\text{Prob}\left(q_p^c\right) = N\left(q_p^i, \sigma_p^i\right) \tag{11}$$

where the mean q_p^i and the deviation σ_p^i are given by the body attraction module. 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. The details of the body attraction are described in Section III-E.

D. Visuomotor Coordination

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

$$C(t) = \frac{\sum_{\tau=t-T_c}^{t} V_s'(\tau) V_a'(\tau)}{\sqrt{\sum_{\tau=t-T_c}^{t} V_s'(\tau)^2}\sqrt{\sum_{\tau=t-T_c}^{t} V_a'(\tau)^2}} \tag{12}$$

$$V_s'(t) = V_s(t) - \frac{1}{T_c}\sum_{\tau=t-T_c}^{t} V_s(\tau) \tag{13}$$

$$V_a'(t) = V_a(t) - \frac{1}{T_c}\sum_{\tau=t-T_c}^{t} V_a(\tau) \tag{14}$$

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 lower and upper boundaries such that $-1 \leq C(t) \leq 1$.

We define the visuomotor memory as the set of variables $\{X_s^{m_j}, I_b^{m_j}, q_h^{m_j}, V_h^{m_j}, q_a^{m_j}, V_a^{m_j}\}_{j=1,\cdots,N_l}$. The visuomotor information is memorized when the visuomotor correlation exceeds a certain threshold. 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 not reliable.

E. Visuomotor Memory

Visuomotor memory is useful in enhancing body search by directing motor exploration. Here, we introduce the concept of body attraction. Body attraction is simply realized by recalling an arm position from the acquired visuomotor memory. The robot refers to 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 the closest head position in the memory and moves the arm toward this position. Since the visuomotor information was 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 originated only from visuomotor memory, which is the result of self-generated motor exploration.

The arm attraction is formulated as follows:

$$q_a^i = q_a^{m_k} \tag{15}$$

$$k = \arg\min_j d_{hj} \tag{16}$$

$$d_{hj} = \left| q_h^c - q_h^{m_j} \right| \tag{17}$$

where q_a^i denotes the motor intention of the arm, and d_{hj} denotes the distance between the motor command of the head position and the jth head position in the visuomotor memory. This motor intention is given to (11) in order to generate a motor command of the arm.

The visuomotor memory provides a prediction of the appearance and location of the body. By referring the motor commands, which is the current goal of the body configuration, a frame of a body image can be anticipated in advance of the action. The procedure of the anticipation is illustrated in Fig. 7. The predicted location of the motor-correlated object in sight is given by the motor commands, i.e.,

$$X_s^p = X_s^{m_k} \tag{18}$$

$$k = \arg\min_j d_j \tag{19}$$

$$d_j = \left| [q_h^c, q_a^c] - \left[q_h^{m_j}, q_a^{m_j} \right] \right| \tag{20}$$

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

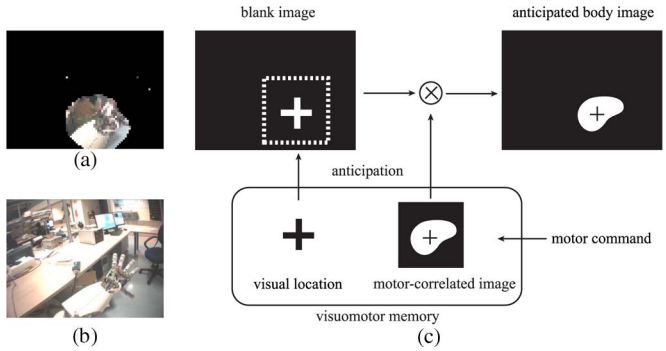


Fig. 7. Body anticipation. (a) Anticipated body image I_b^p before a movement. (b) Actual image I after the movement. (c) Procedure of body anticipation.

TABLE I
EXPERIMENTAL CONDITIONS

Notation	Type	Major condition	Minor conditions
Exp.1(a)	Limited	Motor patterns	Rect., Sin., Random
Exp.1(b)	Limited	Joint activation	Shoulder, Elbow, List
Exp.2(a)	Standard	Natural movement	No head, Head, Interfered
Exp.2(b)	Standard	Body modification	Wrapping, Grasping
Exp.3(a)	Advanced	Body attraction	{0,25,50,75,100}%
Exp.3(b)	Advanced	Body anticipation	Random

IV. EXPERIMENT

We performed experiments on the body definition with the humanoid robot platform. The conditions of the experiments are listed in Table I. In the limited condition, the head posture was fixed, and the arm was moved in low degree of freedom (DOF). Under the standard condition, the robot moved both the head and the arm under visual disturbance by arm modification and human interference. Under the advanced condition, the obtained visuomotor memory was applied to body attraction and body anticipation.

A. Limited Condition

The purpose of the experiments in the limited condition is to validate the robustness of the body definition against different motor patterns and joint activation. In order to focus on these basic characteristics, the head posture was fixed in the experiments. In experiment 1(a), the robot moved a single shoulder joint q_{a0} of the left arm in rectangular, sinusoidal, and random manners. In experiment 1(b), the robot moved a single joint of shoulder q_{a0}, elbow q_{a3}, and list q_{a5} in rectangular manner. The joint structure is illustrated in Fig. 6. Joint position q_{ai} is the ith arm joint angle normalized in $[-1, 1]$, where the upper and lower boundaries correspond to the mechanical lower and upper limits of the joint angle.

The parameters used in the experiment are listed in Table II. T_c, Δt_c, and N_c denote the reference interval, sampling interval, and the number of intervals for calculation of visuomotor correlation. The visuomotor correlation is calculated at each sampling with a sliding reference window of length $N_c \Delta t_c$.

TABLE II
EXPERIMENTAL PARAMETERS

Parameter	Notation	Value
Reference interval	T_c	1.00 s
Sampling interval	Δt_c	100 ms
Number of intervals	N_c	10
Motor pattern period (head)	T_h	20.0 s
Motor pattern period (arm)	T_a	4.00 s
Command interval	Δt_{mc}	100 ms
Execution interval	Δt_{mx}	$\sim$20 ms

These parameters are constrained as $T_c = N_c \Delta t_c$. τ in (12) is interpreted as $\tau = t - k\Delta t_c$, $k = \{0, 1, \cdots, N_c - 1\}$.

The parameters of the motor patterns are listed in Table II. Here, the motor pattern means a segment of sequential motor commands q_p^c. T_h, T_a, Δt_{mc}, and Δt_{mx} denote the period of the motor pattern for the head, the period for the arm, the command interval of the system, and the execution interval of the motor. T_h was not used in Exp.1 but used in Exp.2 (see Section IV-B). The execution interval is the theoretical value determined by the control board of joint motors.

The profiles used when performing continuous rectangle movements are shown in Fig. 8. According to the desired and actual position profile, the robot performed rectangle movements in acceptable delay. Although the velocity of the arm and the moving blob was roughly estimated, it was enough to detect a high correlation of more than threshold 0.8 between them. The visuomotor information was memorized only when the correlation was over the threshold. Note that the appearance of the moving blob and background in sight were not modeled. Then, the salient point was influenced by instantaneous changes of the blob appearances (shape, illumination, and color) and lighting condition in the scene. However, this influence was not serious for body detection in the experiment. The visual modules can be improved by installing more sophisticated visual processing algorithms [24]. Velocity estimation can also be improved by involving an interpolation method such as B-spline approximation [33].

We performed the sinusoidal and random movements in the same manner as the rectangular movements. The profiles during the sinusoidal and random movements are shown in Fig. 9. The sinusoidal movement has a similar profile as the rectangular movement, but the waveform is different. The random movement gives a rectangular-shaped profile as well, but at each interval, the desired position was randomly selected from the normal distribution. The mean q_a^i and deviation σ_a^i of the distribution were set as the joint home position and a constant 0.3, respectively.

The evolution of the visuomotor memory in each condition of Exp. 1(a) and 1(b) is shown in Figs. 10 and 11. Table III summarizes the results of the body definition with respect to variations in the motor pattern and actuated joint. Each trial of the rectangular, sinusoidal, and random movements was performed until the robot acquired 25 recodes of the visuomotor memory. In the table, the columns Trial, Average, and Deviation denote the number of trials and the average and deviation of the time to finish a trial. The column Body Rate is the number of the body-part images divided by the number of all the motor-correlated images (25 images). We defined the body-part image as the image that included a part of the robot arm. The records marked with an asterisk in the table originated from identical experimental data (rectangular movements of the shoulder joint).

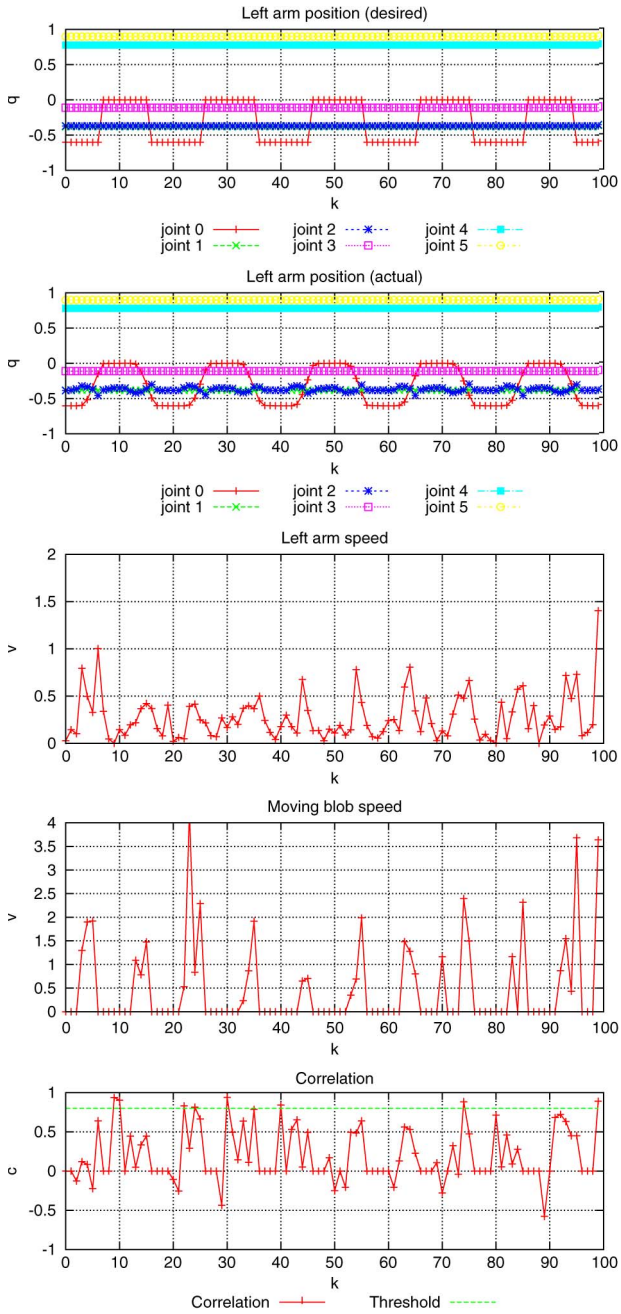


Fig. 8. Profiles of the rectangular movement of the shoulder joint q_{a0} in Exp.1(a). The plots from the top to the bottom correspond to the desired joint position, actual joint position, estimated joint velocity, estimated velocity of a moving blob, and visuomotor correlation, respectively.

According to Fig. 10 and Table III, experimental results positively supported robustness of the body definition against motor pattern variations. In case of the random motor pattern, the deviation of the time to finish a trial was larger than the others. As shown in Fig. 9, the joint positions in the random movement were mostly varied, and the arm trajectory was not uniform. Because of this randomness, the evolution patterns of the visuomotor memory were diverged (Fig. 10 random),

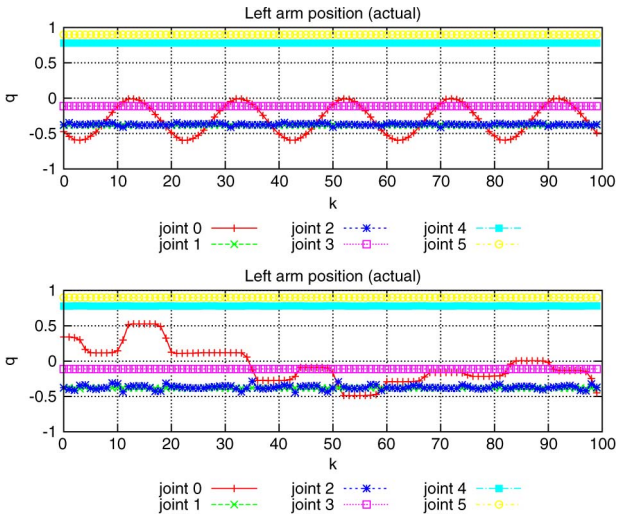


Fig. 9. Profiles of the (top) sinusoidal movement and (bottom) random movement of shoulder joint q_{a0} in Exp.1(a).

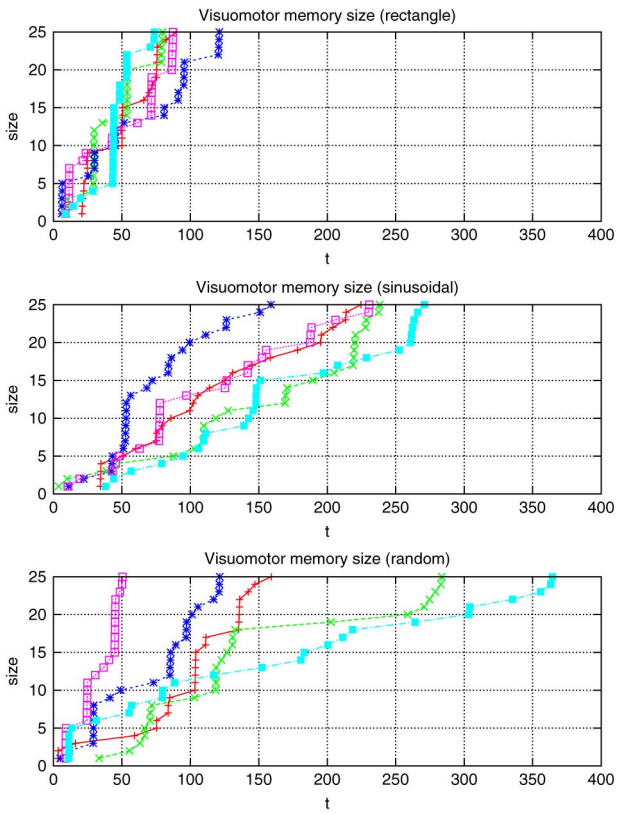


Fig. 10. Evolution of the visuomotor memory against motor pattern variation in Exp.1(a). The plots from the top to bottom correspond to the rectangular, sinusoidal, and random motor patterns. The unit of axis t is second.

and the deviation became larger than the deterministic cases (Table III random).

Moreover, we performed the body definition by using different joints of the arm. The basic configuration was the same as the configuration of the rectangular movement, but the elbow and the list joint, instead of the shoulder, were activated. As shown in Fig. 11 and Table III, experimental results positively supported robustness of the body definition with respect to

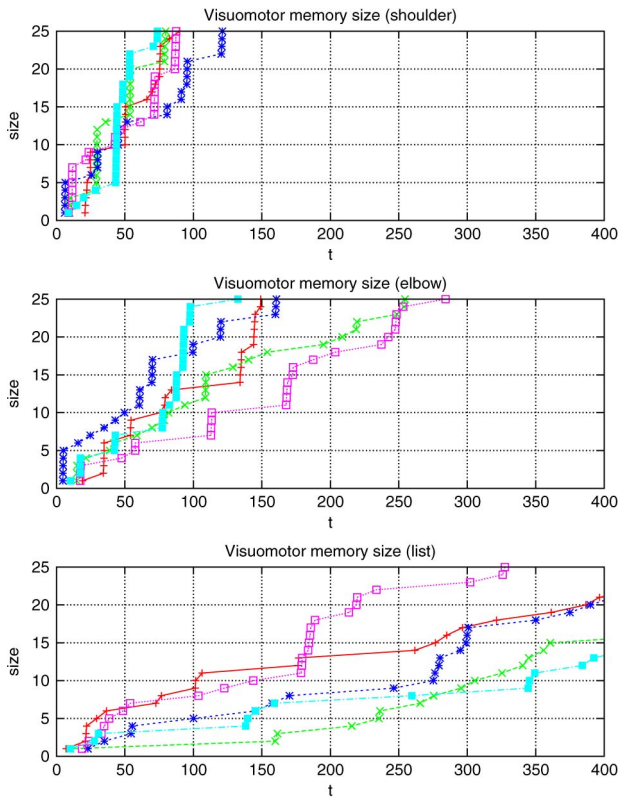


Fig. 11. Evolution of the visuomotor memory against the active joint variation in Exp.1(b). The plots from the top to bottom correspond to the shoulder, elbow and list joint activation. Note that the shoulder plot is same as the plot of the rectangular motor pattern in Fig. 10.

TABLE III
BODY DEFINITION AGAINST MOTOR PATTERNS AND ACTIVE JOINTS

Item	Trial	Capacity	Average	Deviation	Body rate
Rectangular *	5	25	90.4	18.4	100%
Sinusoidal	5	25	224.6	40.9	100%
Random	5	25	195.8	126.6	100%
Shoulder *	5	25	90.4	18.4	100%
Elbow	5	25	196.2	68.3	100%
List	5	25	542.1	188.7	100%

variations of the activated joint. When the robot moved the list joint, more hand images were collected than arm images.

B. Standard Condition

The purpose of the experiments under the standard condition is to validate robustness of the body definition against human interference and body modification. In Exp.2(a), the robot moved both the arm and the head in the random manner by using full joints. First, body definition was performed without human interference. Then, an experimenter interfered with the exploration by presenting a moving object manually. In Exp.2(b), the robot moved both the head and the arm as well, whereas the arm was physically modified by a plastic glove or a grasped object. The experimental scenes of the human interference and the body modification are shown in Fig. 12.

The profiles of the random head and arm movement in Exp.2(a) are presented in Fig. 13. The DOF of the head and arm joints for the movements were set as 3 and 6, which

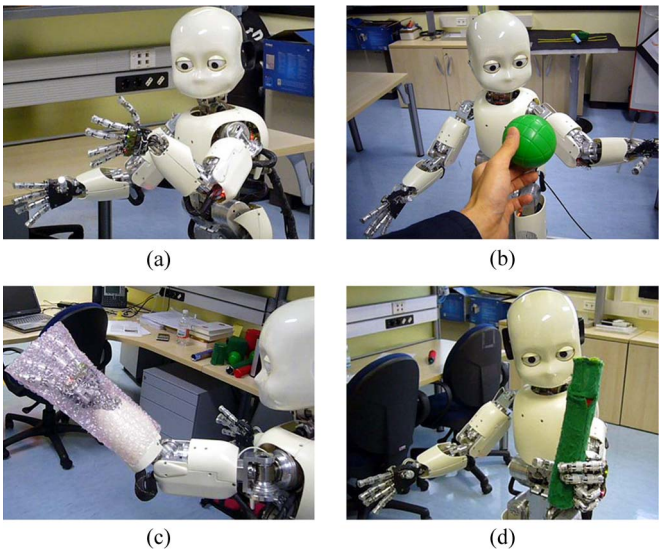


Fig. 12. Experimental scenes in Exp.2. (a)–(d) Scenes of the full-joint movement, human interference, arm wrapping, and stick grasping, respectively.

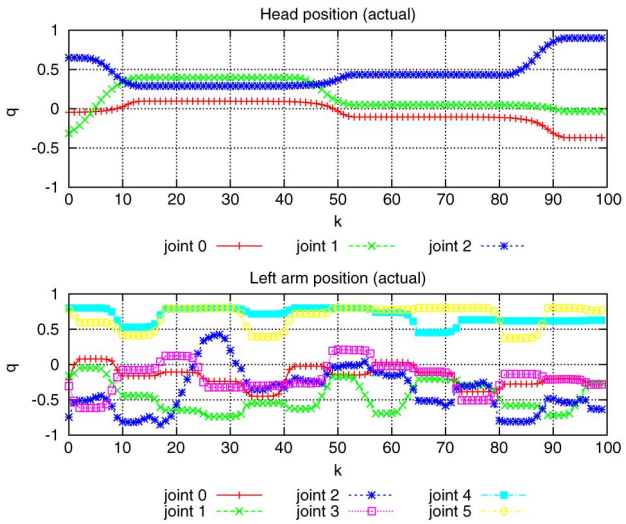


Fig. 13. Profiles of the head–arm random movement by using full joints in Exp.2(a). The plots of the top and the bottom correspond to the head and arm joint positions, respectively.

were the highest DOF configuration. The desired head and arm position were randomly given by the normal distribution in (11) with the constant deviation $\sigma_h^i = \sigma_a^i = 0.3$. The period of the motor patterns T_h and T_a listed in Table II was used for motor generation. In this case, the period of the head movement was five times longer than that of the arm movement. The visuomotor coordination module neglected the motion saliency when the head was turning since the visual movement was not reliable during the change in camera configuration.

A set of motor-correlated images acquired when moving both the head and arm is presented in Fig. 14. The body rate of the images is summarized in Table IV. According to Fig. 14(a) and Table IV, the head movement slightly affected body definition. The head movements allowed the arm to be out of sight frequently. When the arm was moving near the outer frame of the view, the blob perception module failed to capture the arm. The influence of human interference is shown in Fig. 14(b)

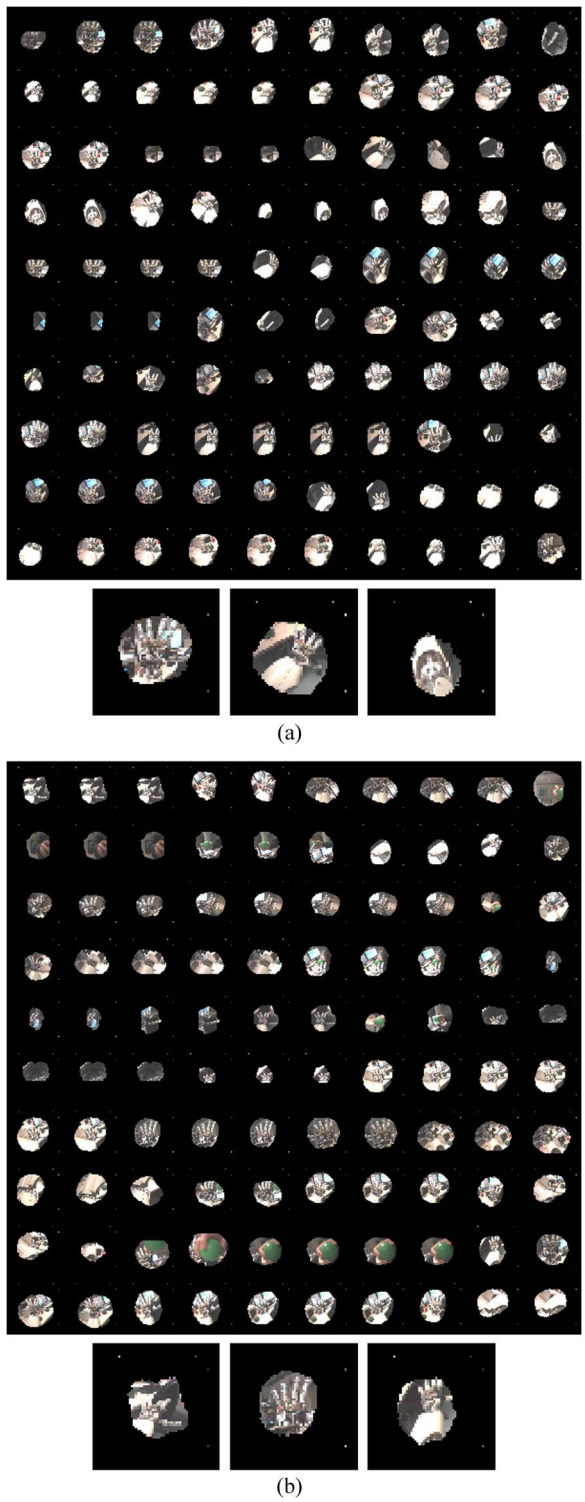


Fig. 14. Motor-correlated images obtained during full-joint movements of the head and the arm in Exp.2(a). The image set of (a) and (b) corresponds to the condition without and with human interference, respectively.

TABLE IV
BODY DEFINITION UNDER STANDARD CONDITIONS

Item	Trial	Capacity	Average	Deviation	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%

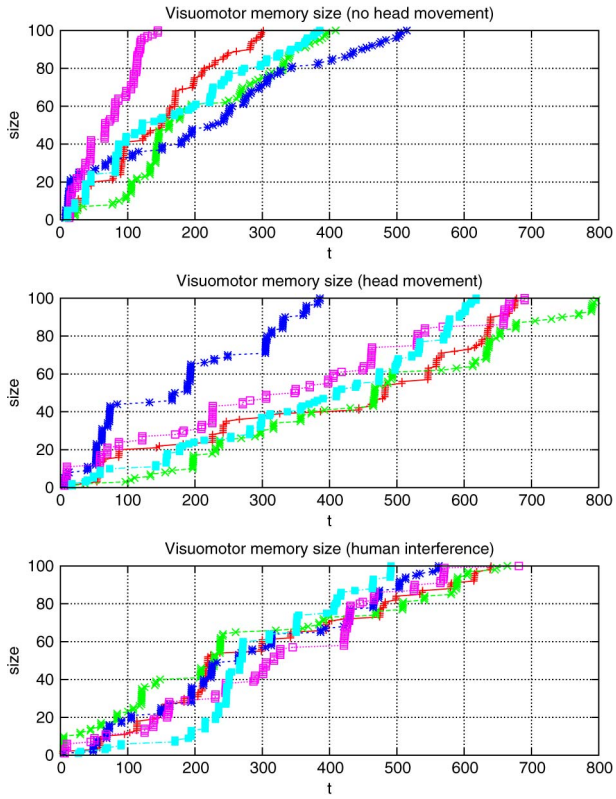


Fig. 15. Evolution of the visuomotor memory without head movements, with head movements, and with head movements under human interference in Exp.2(a).

and Table IV. Under this condition, the experimenter frequently moved a green ball in front of the robot. The set of motor-correlated images includes the experimenter's hand and the presented ball. As compared in Fig. 15, the visuomotor memory with head movements evolved more slowly than the case of no head movements; however, the visuomotor exploration was still successful. The results of the evolution are summarized in Table IV.

The body rate under the human interference condition was dependent on the experimenter's moving manner. Basically, the experimenter moved the object randomly in the experiment. When the experimenter mirrored the robot arm movement, the visuomotor correlation was more influenced. When the experimenter moved the ball close to the eyes of the robot, the influence was stronger as well. The interference also influenced the visually salient location since the current version of the visual saliency module did not take care of the multiplicity of the moving regions. This noise can be reduced by applying a conventional clustering approach such as k-means clustering [33].

The motor-correlated images were obtained, even when the arm was physically modified by wrapping with a plastic glove and grasping a stick. The images are presented in Fig. 16. The head and arm movements were generated in the same manner as in the previous case. As shown in the figure, the own body was defined successfully, even if the arm was modified. Moreover, the wrapped hand and the grasped stick were defined as the body parts and the inherent body parts. These results suggest

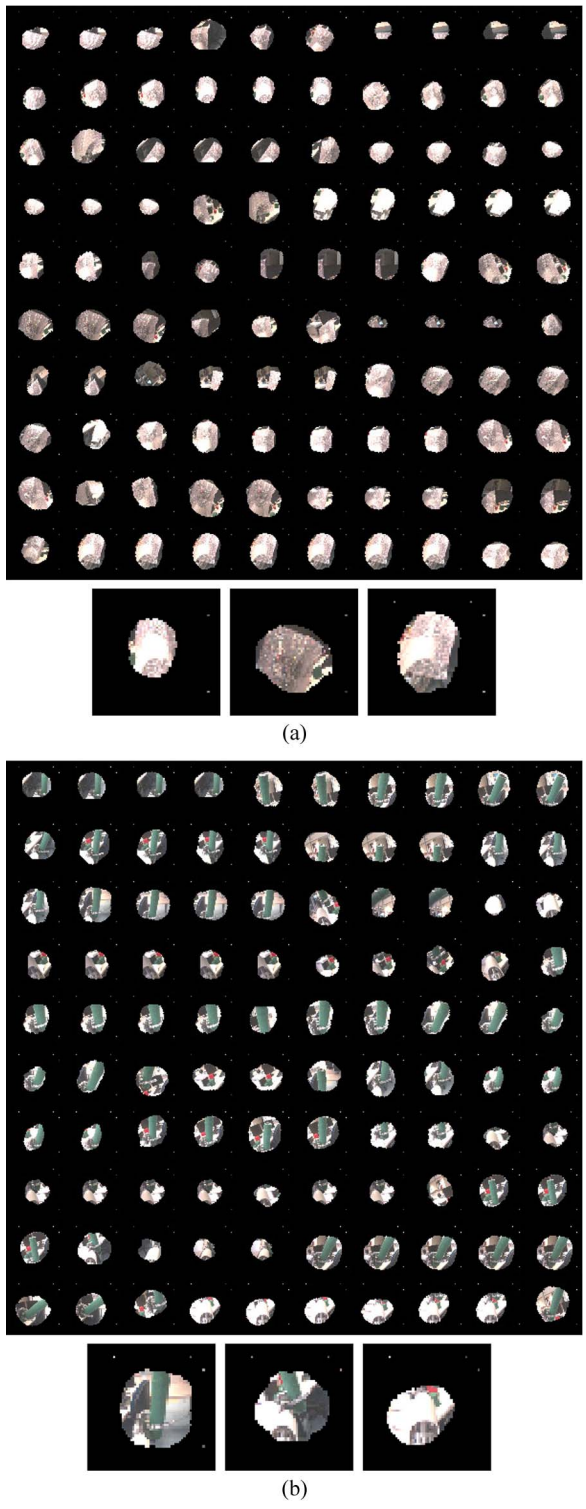


Fig. 16. Motor-correlated images obtained during the full-joint movements of the head and modified arm in Exp.2(b). The image set of (a) and (b) corresponds to the arm wrapping condition and tool grasping condition, respectively.

that the system has the potential for developmental perception of the extended body as monkeys do [1], [2].

C. Advanced Condition

The purpose of the experiments under advanced condition is to exploit the acquired visuomotor memory for body attraction

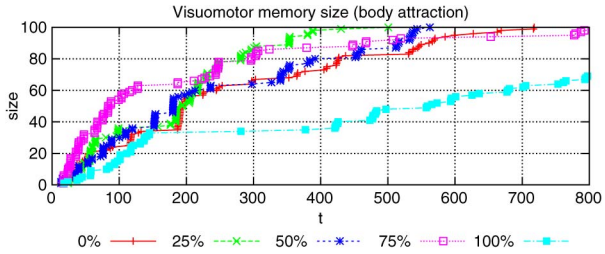


Fig. 17. Evolution of the visuomotor memory in Exp.3(a). The plots correspond to each probability of body attraction (0%: no body attraction, 100%: complete body attraction).

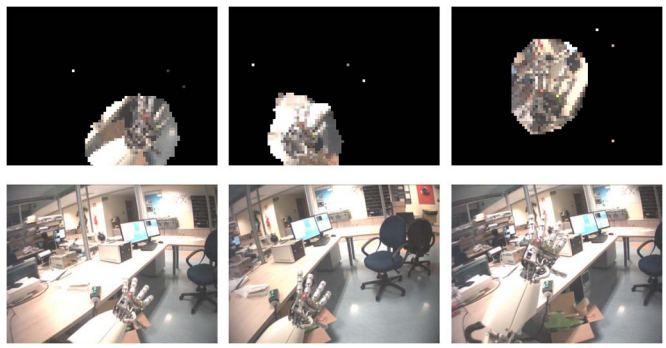


Fig. 18. Body anticipation in Exp.3(b). The top and bottom images in the same column present the anticipated body image before the body movement and the observed image after the body movement (5.0 s later), respectively.

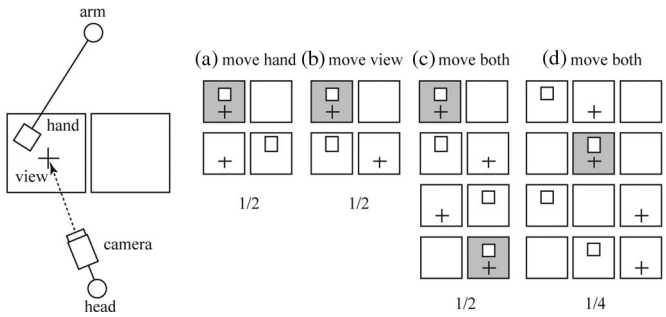


Fig. 19. Search problem to get the hand in sight. (a) The hand is moved, and the view is fixed. (b) The hand is fixed, and the view is moved. (c) Both the hand and the view are moved. (d) Both the hand and the view are moved, but half of the location is not completely overlapped. In the illustrations, the gray box means the case in which the hand appears in sight.

and body anticipation. In Exp.3(a), body attraction was applied to visuomotor exploration. In (11), we used two sets of parameters, i.e., S_g ($\{(q_h^i, \sigma_h^i) = (q_h^h, 0.3),\ (q_a^i, \sigma_a^i) = (q_a^h, 0.3)\}$) and S_l ($\{(q_h^i, \sigma_h^i) = (q_h^h, 0.3), (q_a^i, \sigma_a^i) = (q_a^{m_k}, 0.1)\}$). q_h^h and q_a^h denote the constant vector of the home position of the head and the arm joints, respectively. $q_a^{m_k}$ denotes the arm position associated from the visuomotor memory. S_g gives a global random search of the robot's own body, whereas S_l gives experience-based local random search aiming at the arm to move in sight. One of the modes was selected with constant probability (body attraction rate) in each interval of motor generation.

The evolution of the visuomotor memory with respect to the body attraction rate is plotted in Fig. 17. According to the result, the middle levels of body attraction (25% and 50%) enhanced body definition more. As the results suggest, experience-based action bias accelerates sensorimotor exploration. The paradigm of learning enhancement by learning results may bring a reasonable speed of motor skill development as infants demonstrate [34].

In Exp.3(b), we demonstrated body anticipation. In the advance condition of the experiment, the robot performed body definition under the condition of Exp.2(a) until it acquired 100 recodes of the visuomotor memory. After this learning, the robot generated a head and arm motor command and anticipated the corresponding body image. After the anticipation, the robot executed this motor command and obtained an actual image.

Examples of the body anticipation are presented in Fig. 18. The approximate appearance and location of the robot's own arm were successfully anticipated. In the experiment, we obtained a few failure anticipations. In order to improve the reliability, we are designing a voluntary verification system of the visuomotor memory. The robot can simply reproduce the configuration of a visuomotor memory and filter memory noise in the stochastic manner.

V. Discussion

Body definition with head movements is not a trivial problem, but it was not well considered in the previous studies. Let us assume a simplified situation of body search, as illustrated in Fig. 19. In the figure, the head and arm movement are modeled to direct the view and the hand in two or three discrete locations. Now, we shall neglect the time to move the body to a certain location. When the robot moves only the arm or the head, the probability of the case in which the hand appears in sight after the movement is 1/2. When the robot moves both the arm and head, the probability is still 1/2, but the variation of the patterns becomes twice (from two to four patterns). If the possible locations of the hand and view are not completely overlapped, the probability lowers to 1/4, and the number of patterns remains four.

This simulation suggests that movements of both the head and the arm in the same domain do not decrease the possibility of finding the hand in sight but increase the number of patterns to search. Moreover, when all of the head and arm locations are not overlapped spatially, the possibility of finding the hand in sight decreases. For this reason, body attraction (the motivation of action toward the motor correlations), which we demonstrated in the experiment, is considered as an effective nature to be embedded into the robot.

Another important issue to be discussed is cross-modality in body perception. Since an appearance of a body part depends on the body posture and the view point, it is natural to integrate proprioceptive sensing with visual identification of the robot's own body. In addition, modality integration with tactile and force/torque sensing has the potential to improve visually dominant body definition.

VI. Conclusion

We have proposed a developmental approach of body definition without prior knowledge on kinematics, dynamics, and body appearances. The visuomotor correlation allows the robot

to define its own body through sensorimotor exploration. The robustness of body definition with respect to variation in motor patterns and actuated joints, body modification, and human interference has been experimentally proven. Moreover, body attraction and body anticipation have been demonstrated.

The current body definition system has the potential for binocular perception, but it is not yet examined experimentally. Depth sensing should be included to acquire a 3-D model of the robot's own body. We have also been motivated to embed the proposed body definition into learning-based reaching [32].

Another aspect that we should encompass is haptic information such as tactile and force/torque sensing. In distinguishing an extended body from the inherent body, haptic information plays an important role.

REFERENCES

[1] A. Iriki, M. Tanaka, and Y. Iwamura, "Coding of modified body schema during tool use by macaque postcentral neurones," *Neuroreport*, vol. 7, no. 14, pp. 2325–2330, Oct. 1996.

[2] A. Iriki, M. Tanaka, S. Obayashi, and Y. Iwamura, "Self-images in the video monitor coded by monkey intraparietal neurons," *Neurosci. Res.*, vol. 40, no. 2, pp. 163–173, Jun. 2001.

[3] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cogn. Brain Res.*, vol. 3, no. 2, pp. 131–141, Mar. 1996.

[4] A. Maravita and A. Iriki, "Tools for the body (schema)," *Trends Cogn. Sci.*, vol. 8, no. 2, pp. 79–86, Feb. 2004.

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

[6] M. Kawato, "Internal models for motor control and trajectory planning," *Curr. Opin. Neurobiol.*, vol. 9, no. 6, pp. 718–727, Dec. 1999.

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

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

[9] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends Cogn. Sci.*, vol. 3, no. 6, pp. 233–242, Jun. 1999.

[10] S. Calinon, F. Guenter, and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 286–298, Apr. 2007.

[11] A. Stoytchev, "Toward video-guided robot behaviors," in *Proc. 7th Int. Conf. EpiRob*, L. Berthouze, C. G. Prince, M. Littman, H. Kozima, and C. Balkenius, Eds., vol. Modeling 135, 2007, pp. 165–172.

[12] M. Hikita, S. Fuke, M. Ogino, and M. Asada, "Cross-modal body representation based on visual attention by saliency," in *Proc. IEEE/RSJ Int. Conf. IROS*, 2008, pp. 2041–2046.

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

[14] 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 *Proc. 5th Int. Conf. Develop. Learn.—Special Session on Autonomous Mental Development*, 2006.

[15] P. Fitzpatrick and G. Metta, "Grounding vision through experimental manipulation," *Philos. Trans. Roy. Soc.: Math., Phys., Eng. Sci.*, vol. 361, no. 1811, pp. 2165–2185, Oct. 2003.

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

[17] P. Robbel, "Active learning in motor control," Ph.D. dissertation, School Informat., Univ. Edinburgh, Edinburgh, U.K., 2005.

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

[19] 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 *Proc. IEEE-RAS ICRA*, May 2008, pp. 1608–1614.

[20] C. Tsai, H. Huang, and S. Lin, "FPGA-based parallel DNA algorithm for optimal configurations of an omnidirectional mobile service robot performing fire extinguishment," *IEEE Trans. Ind. Electron.*, vol. 58, no. 3, pp. 1016–1026, Mar. 2011.

[21] B. Wilamowski, "Neural network architectures and learning algorithms," *IEEE Ind. Electron. Mag.*, vol. 3, no. 4, pp. 56–63, Dec. 2009.

[22] B. M. Wilamowski, N. J. Cotton, O. Kaynak, and G. Dundar, "Computing gradient vector and Jacobian matrix in arbitrarily connected neural networks," *IEEE Trans. Ind. Electron.*, vol. 55, no. 10, pp. 3784–3790, Oct. 2008.

[23] Y. Motai and A. Kosaka, "Hand-eye calibration applied to viewpoint selection for robotic vision," *IEEE Trans. Ind. Electron.*, vol. 55, no. 10, pp. 3731–3741, Oct. 2008.

[24] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: A review," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 13–24, Jan. 2010.

[25] Y. Chen, B. Wu, H. Huang, and C. Fan, "A real-time vision system for nighttime vehicle detection and traffic surveillance," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2030–2044, May 2011.

[26] M. Al-Mouhamed, O. Toker, and A. Al-Harthy, "A 3-D vision-based man-machine interface for hand-controlled telerobot," *IEEE Trans. Ind. Electron.*, vol. 52, no. 1, pp. 306–319, Feb. 2005.

[27] J. Kofman, X. Wu, T. Luu, and S. Verma, "Teleoperation of a robot manipulator using a vision-based human-robot interface," *IEEE Trans. Ind. Electron.*, vol. 52, no. 5, pp. 1206–1219, Oct. 2005.

[28] L. Ren, L. Wang, J. Mills, and D. Sun, "Vision-based 2-D automatic micrograsping using coarse-to-fine grasping strategy," *IEEE Trans. Ind. Electron.*, vol. 55, no. 9, pp. 3324–3331, Sep. 2008.

[29] C. Lee and J. Lee, "Multiple neuro-adaptive control of robot manipulators using visual cues," *IEEE Trans. Ind. Electron.*, vol. 52, no. 1, pp. 320–326, Feb. 2005.

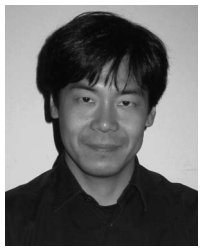
[30] R. Saegusa, G. Metta, and G. Sandini, "Own body perception based on visuomotor correlation," in *Proc. IEEE/RSJ Int. Conf. IROS*, Taipei, Taiwan, Oct. 18–22, 2010, pp. 1044–1051.

[31] G. Metta, P. Fitzpatrick, and L. Natale, "Yarp: Yet another robot platform," *Int. J. Adv. Robot. Syst.*, vol. 3, no. 1, pp. 43–48, 2006.

[32] R. Saegusa, G. Metta, and G. Sandini, "Active learning for multiple sensorimotor coordinations based on state confidence," in *Proc. IEEE/RSJ Int. Conf. IROS*, St. Louis, MO, Oct. 11–15, 2009, pp. 2598–2603.

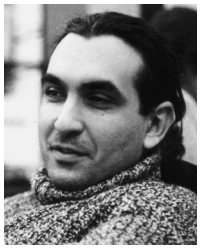
[33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.

[34] H. Watanabe and G. Taga, "General to specific development of movement patterns and memory for contingency between actions and events in young infants," *Infant Behav. Develop.*, vol. 29, no. 3, pp. 402–422, Jul. 2006.



Ryo Saegusa (S'04–M'05) received the B.Eng., M.Eng., and D.Eng degrees in applied physics from Waseda University, Tokyo, Japan, in 1999, 2001, and 2005, respectively.

He is currently a Senior Postdoctral Researcher with the Department of Robotics, Brain, and Cognitive Sciences, Italian Institute of Technology, Genoa, Italy. From 2004 to 2007, he was a Research Associate with the Department of Applied Physics, Waseda University. His research interests include machine learning, computer vision, signal processing, and cognitive robotics.



Giorgio Metta received the M.S. degree (with honors) and the Ph.D. degree in electronic engineering from the University of Genoa, Genoa, Italy, in 1994 and 2000, respectively.

He is currently a Senior Scientist with the Department of Robotics, Brain and Cognitive Sciences, Italian Institute of Technology, Italian Institute of Technology (IIT), Genoa, and an Assistant Professor with the University of Genoa, where he teaches courses on anthropomorphic robotics and intelligent systems for the bioengineering curricula. He has been an Assistant Professor with the University of Genoa since 2005 and with IIT since 2006. From 2001 to 2002, he was a Postdoctoral Associate with the Massachusetts Institute of Technology Artificial Intelligence Laboratory, where he worked on various humanoid robotic platforms. His research interests include biologically motivated and humanoid robotics, particularly developing life-long developing artificial systems that show some of the abilities of natural systems. His research develops in collaboration with leading European and international scientists from different disciplines such as neuroscience, psychology, and robotics.

Giulio Sandini received the B.S. degree in electronic engineering (bioengineering) from the University of Genova, Genoa, Italy.

He is currently a Director of Research with the Italian Institute of Technology, Genoa, and Full Professor of bioengineering with the University of Genoa. His research interests are computational and cognitive neuroscience and robotics with the objective of understanding the neural mechanisms of human sensorimotor coordination and cognitive development from a biological and an artificial perspective. He has been Assistant Professor with the Scuola Normale Superiore, Pisa, Italy, and a Visiting Scientist with the Department of Neurology, Harvard Medical School, Boston, and the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge. Since 2006, he has been a Director of Research with the Italian Institute of Technology, where he leads the Department of Robotics, Brain and Cognitive Science.