Self-body Discovery Based on Visuomotor Coherence

Ryo Saegusa[†], Giorgio Metta[†‡], Giulio Sandini[†‡]
[†]Robotics, Brain and Cognitive Sciences Dept.,
Italian Institute of Technology, Genoa, Italy
[‡]LIRA-lab, University of Genoa, Genoa, Italy
ryos@ieee.org, ryo.saegusa@iit.it, pasa@liralab.it, giulio.sandini@iit.it

***Abstract*— This paper proposes a plausible approach for a humanoid robot to discover its own body part based on the coherence of two different sensory feedbacks; vision and proprioception. The image cues of a visually salient region are stored in a visuomotor base with the level of visuo proprioceptional coherence. The high coherence between the motions in the vision and proprioception suggests the visually attracted object in the view is correlated to its own motor functions. Then, the robot can define the motor correlated objects in the view as the self-body parts without prior knowledge on the body appearances nor the body kinematics. The acquired visuomotor base is also useful to coordinate the head and arm posture to bring the hand inside the view, and also recognize it visually. The adaptable body part perception paradigm is effective when the body is possibly extended by the tool use. Each visual and proprioceptional processes are distributed in parallel, which allows on-line perception and real-time interaction with people and objects in the environment.**

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 in general sense; 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 self body developmentally. The function of self body perception is considered essential for robots to identify the self when interacting with people and objects. Also, it is potential to perceive the extended body when using a tool.

The overview of our approach is depicted in Fig.1. The principal idea is to simply move the body and monitor the coherence of the visual and proprioceptional feedbacks. Here we assume that a robot moves a body part, then notices the motor correlated objects as the self-body. At every moment image cues of visually attracted region are stored in a visuomotor base with the visual and joint information. The visuomotor base manages the visuomotor correlation on a movement. Since the visual movement and the physical movement of the body part is theoretically dependent, the level of correlation helps to recognize the self-body from an image input. This correlation is also useful to predict the location of the self body in the view, and recognize it visually.

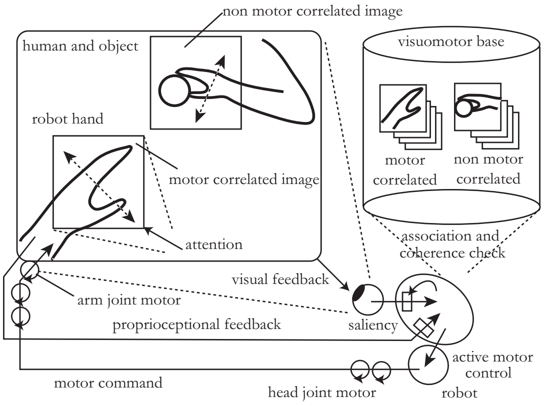

Fig. 1. Visuomotor coherence based self-body discovery system. A robot generates the arm movements, and senses the visual and proprioceptional feedbacks. When these feedbacks are coherent, the image of the attracted region is stored with visuomotor properties. After the short term arm exploration, the robot can visually recognize the body parts. The visuomotor system is allowed to localize the detected body part on-line.

This paper is organized as follows: Section II describes the related works in robotics and neuroscience. Section III describes the proposed framework and details of component processes. Section IV describes the experimental results with the humanoid robot James [1]. Section V gives the conclusion and outlines some 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 [2] (Fig.2). 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 [3], they trained the monkey to recognize the image of their hand in a video monitor (Fig.3), 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 the RobotCub project (IST-2004-004370) funded by the European Commission through the Unit E5 Cognitive Systems.

978-1-4244-7561-2/10/$26.00 ©2010 IEEE

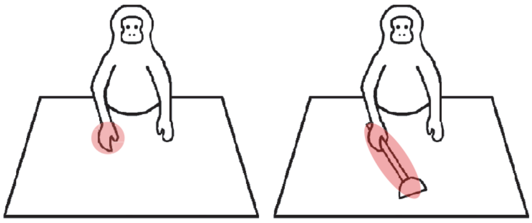

Fig. 2. Visual receptive field of the bimodal neurons (left: before tool use, right: after tool use) [2][4]. The monkey perceives the tool as an extended body part.

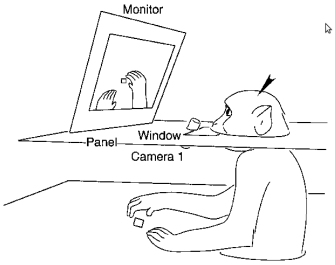

Fig. 3. The experimental setup of the video-guided manipulation training for a monkey demonstrated by Iriki et al. [3]. The monkeys recognized the image of their hands in a video monitor as an extension of the self.

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. [5], Kawato et al. [6]), mirror system (Metta et al. [7]), action-perception link (Fitzpatrick et al. [8]), and imitation learning (Schaal et al. [9], Calinon et al. [10]). However, the body detection was often hand coded with predefined rules on appearances or body kinematics such as visual markers and the joint-link structure. These kinds of prior knowledge give robustness for the body detection as well as certain limits. For instance, a precise manipulation task with robot fingers may require a visual model of the fingers for perception. In other situation; when the robot manipulates something with a tool, it would be difficult to adapt the physically extended hand as the monkeys are dexterously doing it (Fig.2 and Fig.3)

Recently, Stoytchev [11] proposed an approach of video-guided reaching to demonstrate the similar tasks to what Iriki et al. examined in [3]. The robot, which is supposed to work on a plane, coordinates its reaching action with a self-image projected on a video monitor under the visual disturbance. In the experimental setting, the robot identifies an object with prepared different color markers. This simplification is effective to neglect multiple appearances of the same object, and also similar appearances of different objects. The coincidence of visomotor information is evaluated by the temporal contingency between the motor command and visual movement; however the delay in these movements (efferent-afferent delay) must be calibrated in advance, which means that at least the experiment operator must define the robot hand.

Hikita et al. [12] proposed a bimodal (visual and somatosensory) representation of the end effector based on Hebbian learning, which simulated the experiments with monkeys in [2]. The coincidence of the vision and proprioception is evaluated by the contingency between the visual location and hand posture, which can be placed at the more spatial approach compared to the one by Stoytchev, and visual detection based on the saliency [17] seems more general; however, the approach is validated only 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 attracted object on the image plane, which gives a measure of these statistical dependency. The visually detected objects are separated by off-line image clustering, then the image cluster with high dependency is assigned as the self-image cluster. The proposed approach is well validated with a humanoid robot with a fixed head posture. The limitation is that head movements are not considered in the approach. Generally the head movement affects the motion based object perception and broad search space where the arm locates rather out of view.

There are other several methods which focuse on temporal dependency rather than space dependency [14] [15] [16]. The approach by Natale et al. [16] are based on the image differentiation by the periodic (sinosoidal) hand movements, which detects a visually moving object with the similar frequency. In these approaches, a movement pattern functions as the feature to measure the visuomotor coincidence.

Compared to the previous approaches described above, our approach can be characterized as the rather temporal coincidence approach including the action-perception aspects; the efficient exploration of motor correlated objects under the head movements. The developmental visuomotor experience allows to define the self-body, predict the body location, and visual body recognition.

III. METHOD

The self-body discovery system is depicted in Fig.4. The system is composed of some modules categoried in the four types: vision, proprioception, visuomotor coordination, and motor generation. This section describes each function of the modules.

A. Vision processing

The overview of the visual processing is shown in Fig.5. The visual processing is modularized as a set of cascaded image filters: saliency, attention, and tracking, which are distributed in the networks to allow real time processing. All modules are dually structured for two image streams from the left and right eye cameras.

The saliency module is the extended saliency model of Itti et al. [17], newly including the flicker channel (motion channel) which responds a moving object in the view (Fig.5, saliency module). Generally, the principal of the visual saliency is to detect a difference of the intensity from the

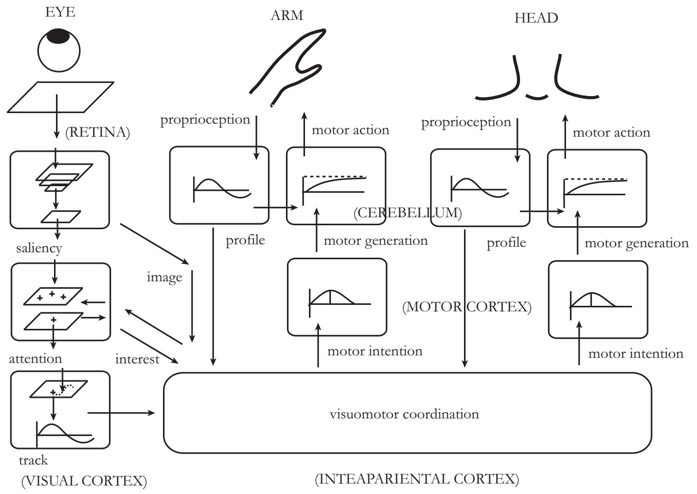

Fig. 4. The self-body discovery system. The system is composed some modules categorized into the four types: the vision, proprioception, visuomotor coordination, motor generation. Each modules functions in parallel.

surround region (center-surround difference), and sum up these differences given by the multi-scaled images.

The input RGB image from the left or right eye camera of the robot is decomposed into the four channels; intensity(I), color (R,G,B,Y), orientation (0,45,90,135 degree), and flicker(F). Here, intensity, color and orientation channels are based on the ones in [17], while the newly included channel of flicker image is a simple temporal difference of the subsequent two frames. The final saliency output S is the combination of the saliency of each visual channels; intensity S_i, color S_o, orientation S_o, and flicker S_f as follows,

$$S = \frac{1}{4}\{S_i(I) + S_c(R,G,B,Y) \\ +S_o(0,45,90,135) + S_f(F)\}. \tag{1}$$

The attention module functions to select the attracted location from the saliency image (Fig.5, attention). The most N_s salient locations p_i are preselected, and these salient levels s_i are weighted by the interest (w_i) of the high level module (visuomotor base). The final attracted location p_a is selected stochastically from the $\{p_i\}_{i=1,\cdot,N_s}$ with the probability $Prob(p_a = p_i)$ as defined,

$$Prob(p_a = p_i) = w_i s_i \ / \ \sum_{j=1}^{N_s} w_j s_j. \tag{2}$$

This attention signal is suppressed when the robot is turning the head. The attention module monitors the head movement by the proprioceptional feedback (although it should be from a gyrosensor biologically), then cancels the attention when the head is turning.

The track module tracks the attracted location on the image denoted as $X(t)$. Here, the attracted location is given from the attention module (Fig.5, track). The tracking is realized by matching of the tracking regions in the subsequent frames on the intensity (I) with a saliency image (S) filtering. The tracked location is updated by the new attracted region in a certain period. The tracked location is temporally profiled with binary reshaping at each cycle of time t and re-sampled with a constant sampling frequency denoted as $\bar{\dot{X}}(t)$.

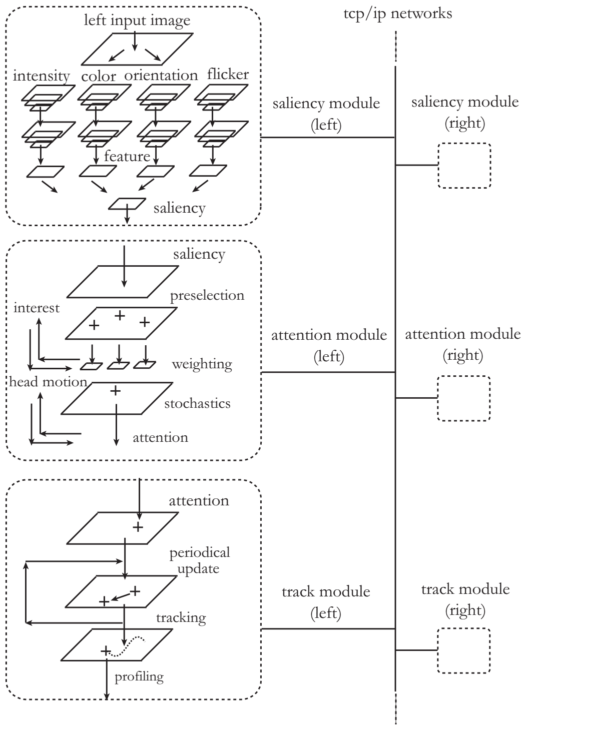

Fig. 5. Visual processing. The visual processes are distributed to allow real-time processing. The saliency module decomposes a left/right input image into the basic features channel as intensity, color, orientation, and motion. The attention module selects an attracted location stochastically depending on the interest of the high level module. The track module tracks the attracted location and gives the motion profile. All modules are dually structured for two video streams from the left and right eye cameras.

B. Proprioceptional processing

The proprioceptional feedback is produced from the velocity signal of the motor encoders. In general the velocity profile is affected by many factors such as the motor torque, trajectory, gravity force, and external force applied by contact objects. In order to define a simple and robust proprioceptional motion feedback, we reshaped the velocity profile like an on-off signal by the equation:

$$\bar{\dot{J}}(t) = 1, \quad \text{if } |\dot{J}(t)| \geq v_0, \tag{3}$$

$$\bar{\dot{J}}(t) = 0, \quad \text{if } |\dot{J}(t)| < v_0, \tag{4}$$

where $\bar{\dot{J}}$, $\dot{J}$, v_0 are the proprioceptional motion feedback, original velocity, and velocity threshold at time t, respectively. The velocity profile and proprioceptional motion feedback are contrasted in Fig 6. The proprioceptional feedbacks of each joint are unified as a matrix with joint rows and time columns.

C. Visuomotor coordination

Visual motion feedback $\bar{\dot{X}}(t)$ from the track module and proprioceptional motion feedback $\bar{\dot{J}}_a(t)$ from the proprioceptional module of the arm are synthesized with the coherence

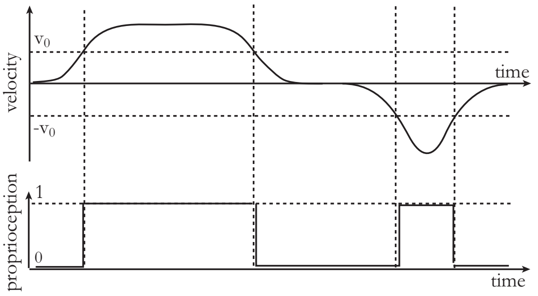

Fig. 6. Velocity profile and proprioceptional motion feedbacks. The velocity is binarized by the threshold to simplify the value like activation and deactivation, which helps to check visuomotor coherence in noise more robustly.

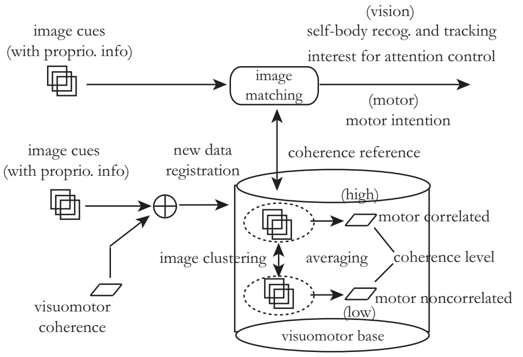

Fig. 7. Visuo proprioceptional coordination. The image of the visually attracted region is registered in the visuomotor base coupled with a coherence level of the proprioceptional motor feedback. The image base updates image clustering on-line, then assigns the rank of the motor correlation by comparing the average of each cluster's coherence level. The highly ranked clusters are regarded as the motor correlated image cluster. Independently, the motor correlation of a query image is recognized by referring the centroid of the image clusters.

level $C(t)$, and stored in an visuomotor base with the image patch of the attracted region $I_a(t)$ and joint position of the arm $J_a(t)$ and head $J_h(t)$. Overview of the visuo proprioceptional coordination is depicted in Fig.7. The image base accepts the on-line image registration and reference independently. The coherence level is the maximum value of the simple temporal correlation as follows,

$$C(t) = \max_{\tau_0 < \tau < \tau_1} \sum_{k=t-T}^{t} \bar{\dot{X}}(k-\tau)\bar{\dot{J}}_a(k) \ / \ |\bar{\dot{X}}(k)||\bar{\dot{J}}_a(k)|. \quad (5)$$

The registration of the image cues are triggered by the input from visual and proprioceptional modules (Fig.4). The visuomotor base updates image clustering on-line, then ranks the motor correlation of each cluster. The rank is defined as the order of the each average value of the coherence level of the cluster members. The highest-ranked cluster is regarded as the most motor correlated cluster.

The online image clustering is illustrated in Fig.8. The clustering is based on the Kmeans method [18], but modified to allow on-line updating. Kmeans is a classical clustering method with a given cluster number. All data are labelled randomly in the beginning. Then, centroids of the clusters are calculated, and reassigned the nearest centroid cluster's label. This update is repeated until any label does not change.

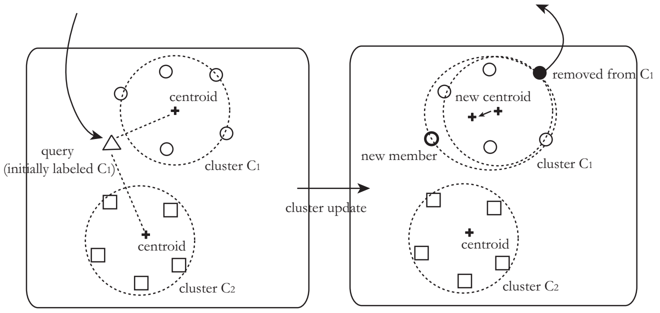

Fig. 8. Online image clustering. The clustering is based on the Kmeans method [18], but modified to allow on-line updating. The query (new member) is initially assigned the label of the nearest centroid's cluster. Then, the clusters are updated in the Kmeans manner. After the update, a member is randomly selected from the cluster of the new member and removed.

In our image clustering, the new member is registered one by one until the total member number reaches the limit number. Let the limit number and cluster number denote N_K and K, respectively. After it reaches the limit, when a new member is registered, a member is removed from the image base. The query (new member) is initially assigned the label of the nearest centroid's cluster. Then, the clusters are updated in the Kmeans manner on-line. After the update, a member is randomly selected from the cluster of the new member and removed. The last operation functions for keeping the total number constant and avoiding reducing a member from the minor cluster. The metric of the images is measured as the multiplication of the canonical correlation value $C(I_a, I_b)$ as defined:

$$C(I_a, I_b) = \frac{I_a(x,y) \cdot I_b(x,y)}{|I_a(x,y)||I_b(x,y)| + \epsilon} \quad (6)$$

where I_a and I_b denote the vectors of images, and ϵ is a small positive constant to avoid zero division.

D. Motor Processing

The motor behavior of the robot is produced by a biased motor babbling. The motor babbling gives random movements of joints, which is useful for the robot to explore the environment without a structured motor control. However, the internal space to explore is large for the robot when including both the arm and head movements.

In the most of related works described above, the head posture is set up stationally, while in this study we are challenging more natural self-body discovery including head movements. The head movement makes two technical problems in our case; (1) the arm posture does not match the visual location of the hand under interference of the head movement, and (2) the hand does not always stays in the view. The first problem is solved by combining the head posture with the arm posture. The second problem is solved by introducing the bias of the motor exploration.

The visuomotor base can suggest the head posture with highest coherence level coupled with the currently similar

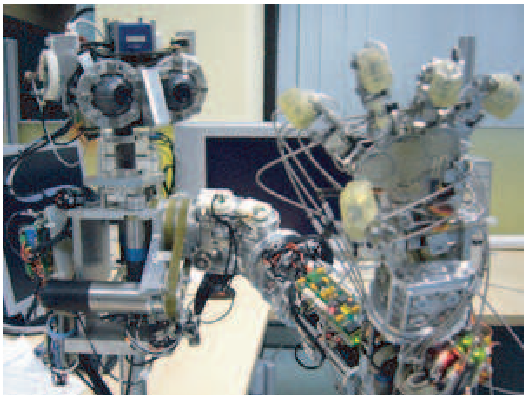

Fig. 9. The humanoid robot James used for experimental validation of the proposed body discovery system. The shoulder motors of the pitch and yaw are actuated to generate hand movements.

arm posture, which functions as a high level motor intention for the motor generator module to direct its exploration to the domain where the attracted object (the self-body) might appear. In the proposed system, the motor generator module utilizes this intention to bias the next desired head posture in the motor babbling if the coherent level is sufficient. The motor generator module stochastically produces the next desired head posture $J_h^{cmd}(t)$ by the gaussian distribution with the center of intended head posture $J_h^{int}(t)$ (Fig.4, motor intention).

IV. EXPERIMENT

We performed the experiment of robot hand discovery using a humanoid robot. In this experiment, we also challenged the hand tracking based on the hand detection. Through the movements, the robot was autonomously getting the appearances of the motor correlated object, which is its own hand, based on the visuo proprioceptional coherence. The robot also collects the appearances of motor noncorrelated objects and people, which are considered useful for general object recognition.

A. Experimental setting

The humanoid robot James [1] is a fixed upper-body robotic platform dedicated to vision-based manipulation studies. It is composed of a 7dof arm with a dexterous 9dof hand and a 7dof head as shown in Fig.9. It is equipped with binocular vision, force/torque sensors, tactile sensors, inertial sensors and motor encoders.

In this experiment, we mainly used the shoulder motors of the pitch and yaw to generate hand movements in the view field, while the other motors such as the wrist and elbow were stationary to make a suitable arm posture. The shoulder encoders were used to sense the proprioceptional feedbacks of the arm movement based on the velocity profile. The visual effects were extracted from the image streams of the left eye camera mounted on the head. The neck also moves in small magnitude and less frequently, but the visual effect of the head turning is suppressed in the attention module.

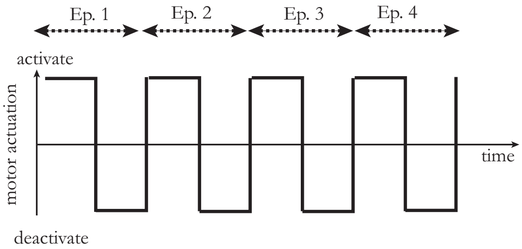

Fig. 10. Time line of the robot arm activation. The schedule is composed of same periods denoted episode T. In the first half of each episode, the motors of the shoulder pitch and yaw were activated to generate random trajectories by its own hand. The shoulder joint motors were deactivated in the next half of the episode. Independent from the robot motor activation, a human partner randomly shows some moving objects in each episode.

TABLE I
EXPERIMENTAL PARAMETERS

N_s	3
N_k	1000
K	2
T	20 sec
v_0	0.05
τ_0	0
τ_1	10
ϵ	0.001

The time line of the robot arm activation is shown in Fig.10. The robot activates and deactivates the shoulder motors alternatively, collecting the appearances of the moving objects and its coherence to the proprioceptional feedbacks. During the activation period, the robot arm movements were generated by random motor babbling of shoulder motors. Let one cycle of activation and deactivation denote episode T. We also activated less frequent head movements.

Independent from the robot arm activation, a human partner presented some moving objects almost randomly. Then the attention module detected not only the hand region but also an object region and the human partner. The binocular object recognition is possible, but we did not perform it in this experiment to focus on a basic condition. The parameters used in the experiment is presented in TABLE I.

B. Results

The outputs from the visual processing modules are shown in Fig.11, while the hand recognition and tracking result are shown in Fig.12, respectively. The readers of this article are recommended to refer the real robot behavior in the attached video clip.

Fig.13 shows some images stored in the visuomotor base at the end of Ep.4. Most of the motor correlated images were the robot hand images, in which we can see the texture of the black-arc-like tendon wires of the arm. On the other hand, the motor noncorrelated images included ball images which the human partner presented in interaction, and also the human's hand. Each cluster includes the images of the same objects but different appearances. The variety of the image clusters allows to recognize an object with different view such as the other side of the hand. This robustness is an

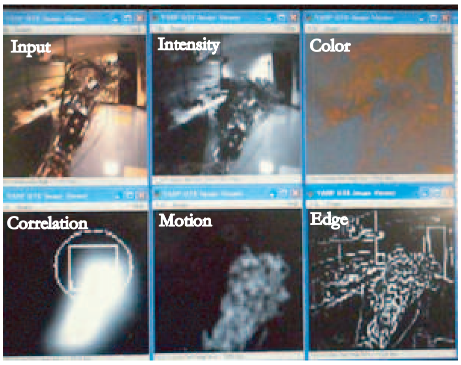

Fig. 11. Samples of filtered images. The white rectangle in the image of correlation indicates that the body discovery system registers the image patches of this region. The white circle indicates that the image is recognized as a motor correlated object.

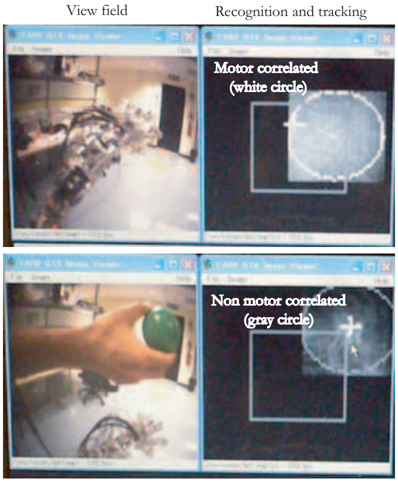

Fig. 12. Image recognition and tracking. The image tracking is an independent process from the robot hand detection process. Moving objects in the view triggers to recognize the motor correlation, and sent to the tracking module as an attracted point with the correlation label. Then the tracking module tracks the new attracted point or keeps tracking the previous point. The white circle shows that the tracked image was recognized as the motor correlated (above figures), while the gray circle indicates that it is recognized as motor noncorrelated (below figures). The blurred white square show the intensity of image correlation in the neighbor, and the white cross indicates the position of the maximum correlation pixel, which is set as the next tracking point.

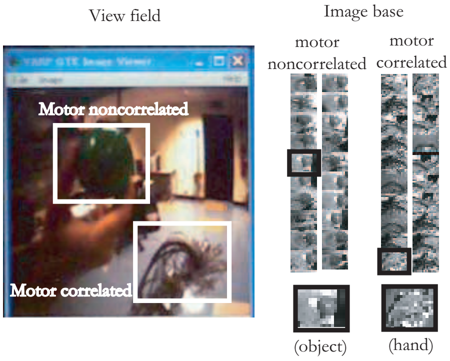

Fig. 13. Registered images in the image base. The right figures show the registered motor correlated and noncorrelated images. In this experiment, the human trainer showed the green ball often, therefore the green ball with the trainer's hand appeared in many motor noncorrelated images Motor correlated images are mainly the robot hand, visually characterized with many cables of tendon which look a black arc.

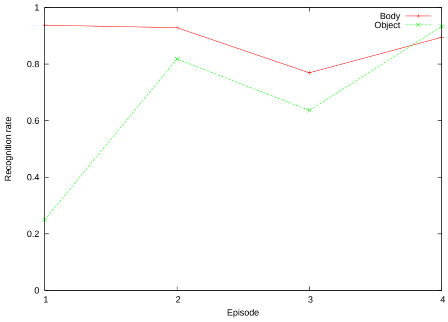

Fig. 14. Recognition rate. Each episode includes the 40 to 50 times of visual recognition. The episode was down sampled into 10 to 12 samples to evaluate the recognition rate. The complexity of motion trajectories and image textures are considered to have influenced the performance of the recognition.

advantage of the proposed approach against the others using predefined visual markers. Those approaches are generally weak for occlusion of markers.

Fig.14 plots the recognition rate of each episode. This recognition was performed only with the visual information. After watching the moving hand and object enough times, the body discovery system gives better hand recognition. The proprioceptional feedback gives a confidential motor sensing of its own body, but it does not give spatial and texture information on the body and its motion. On the other hand, the visual feedback gives object appearances and spatial effects of body movements, but it does not tell the robot whether the moving object is related to its own body. Then, the visuo proprioceptional coherence is an essential bridge to associate the appearance of the moving object to sense of self-generated body movements. This bimodal bridge allows the robot to discover its own body part.

V. Discussion and Conclusions

This paper proposed an approach of an autonomous self-body discovery without specific prior knowledge on the body appearances. The visuomotor coherence informs a robot of the link between observed objects and its own body parts. The robot is also enabled to keep the appearances of the motor correlated objects in memory to recognize them visually as the private movables; the self body.

The current body discovery system allows the binocular object recognition, but not examined experimentally. It should be proved also with the depth sensing, which will be helpful for understanding of spatial body structure and calibration of the environment by the body. Technically, we are also interested in large-scale image clustering and object recognition with the visuomotor base.

The system visually recognizes motor correlated objects including a grasping tool which moves together with the hand. The next work should encompass a tool use application based on the self-body separation. In order to distinguish the extended body part from the inherent body part, tactile information plays an important role. This developmental approach also includes motor learning of the body part after its discovery. The active learning approach which we previously proposed [19] would be helpful for this issue.

As a general conclusion from the series of these studies, the maximization of the cross-modal sensorimotor correlation can be placed as a reasonable inherent desire or motivation for embodied intelligence to generate their motor behaviours in the sense of the self-body perception and control.

References

[1] 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.

[2] 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.

[3] 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.

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

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

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

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[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] A. Arsenio and P. Fitzpatrick, "Exploiting cross-modal rhythm for robot perception of objects." in *In Proceedings of the Second International Conference on Computational Intelligence, Robotics, and Autonomous Systems*, December 2003.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.