Constructing bimodal receptive field based on visual attention by saliency

Masaki Ogino[1], Mai Hikita[2], Sawa Fuke[2], Takashi Minato[1] and Minoru Asada[1,2]

[1]ERATO Synergistic Intelligence Project, JST.
[2] Graduate School of Engineering, Osaka University.

Body representation is one of the most fundamental issues for physical agents (humans, primates, and robots) and also robots to perform various kinds of tasks. Especially during tool-use by monkeys, neurophysiological evidence shows that the representation can be dynamically reconstructed by spatio-temporal integration of different sensor modalities so that it can be adaptive to environmental changes [1]. However, to construct such representation, how to associate which information among various sensory data is an issue to be solved.

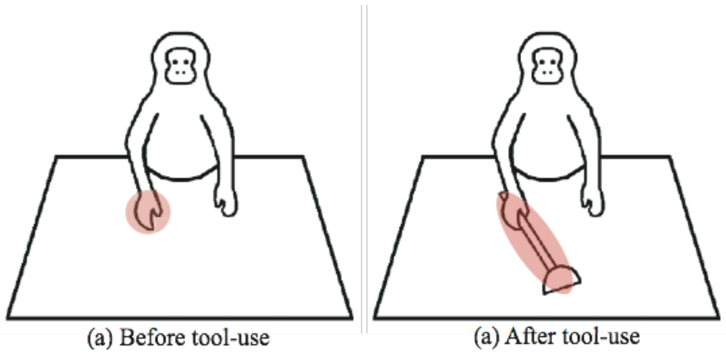


Fig. 1 Changes in bimodal receptive field properties

We suppose that the visual attention is a key issue to realize the visual receptive field not simply because the visual attention mechanism can focus on the salient features (bottom-up flow) but also because such a field can be activated when attention is directed to it in some way (top-down flow) like the activation of the monkey's visual receptive field by the visual pointer. To ease the cross-modal association between vision and proprioception, we utilize the tactile sensation to trigger the process, that is, when it touches something, a robot associates the visual salient features with the proprioceptive data.

An overview of the proposed system is shown in Fig. 2 where three modules are involved. The arm posture module corresponds to the proprioception, representing various kinds of postures in terms of joint angels that are collected and structured as SOM (self organizing map). The attention module detects the salient features in the camera image as the candidates for attention

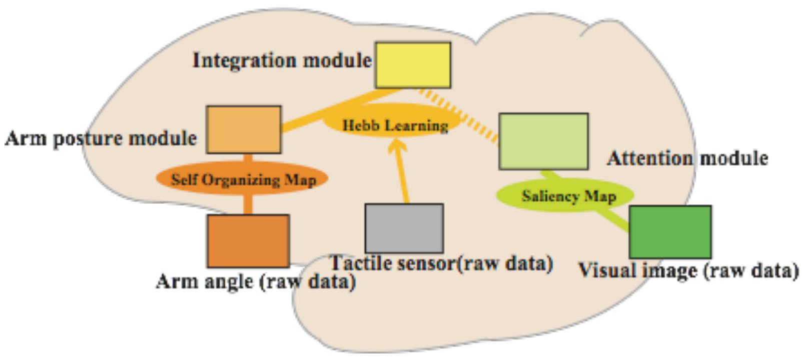
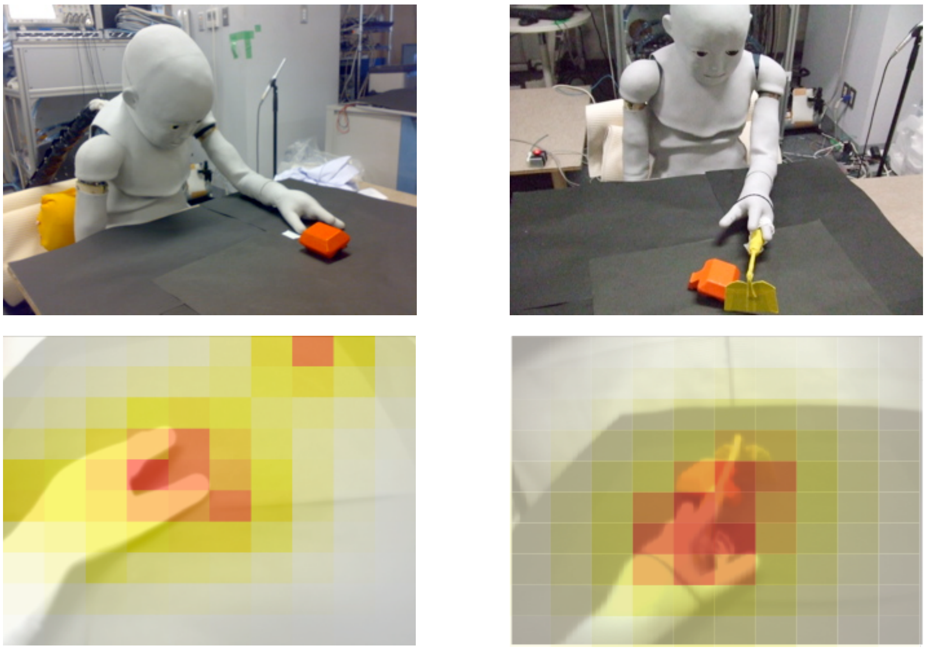


Fig. 2 Overview of the proposed model

point based on saliency map algorithm [2] in every step. The integration module associates the arm posture with visual attention point by Hebbian Learning when the robot detects the tactile sensation by hitting a target with its hand or a tool. This module can be regarded as a model of the neuron in the parietal cortex.

The proposed model is applied to the real robot, CB^2 [3], with the similar experimental environment as that with macaque monkey by Iriki et al. [1]. The target object colored orange is positioned randomly everytime the robot touches the object with his own body (including a tool).



(a) with hand (b) with tool

Fig. 3 Experimental results: learned receptive field of the integration module

Fig. 3 (a) shows that the connection weights converge most strongly to the area around the end effector (the hand of the robot). This implies that the hand area is the most salient for this robot when it touches the object. Fig. 3 (b) shows that the connection weights are extended to the tool area. These results are comparable to those of the experiments with macaque monkeys.

References

[1] A. Iriki, M. Tanaka, and Y. Iwamura, “Coding of modified body schema during tool use by macaque postcentral neurons,” Neuroreport, Vol. 7, pp. 2325-2330, 1996.
[2] L. Itti and F. Pigin, “Realistic avatar eye and head animation using a neurobiologiical model of visual attention,” in Proceedings of SPIE 48th Annual Internationala Symposium on Optical Science and Technology, Vol. 5200, 2003, pp. 64-78.
[3] T. Minato, Y. Yoshikawa, T. Noda, S. Ikemoto and H. Ishiguro, “Cb2: A child robot with biomimetic cbody for cognitive developmental robotics,” in Proceedings of the IEEE-RAS International Conference on Humanoid Robots, 2007. CDROM.