Development of Robot Self-Identification Based on Visuomotor Prediction

Tao Zhou[1], Piotr Dudek[2] and Bertram E. Shi[1]

***Abstract* — We propose a developmental method that enables a robot to identify visual locations associated with its own body from a cluttered visual image based on the concept of visuomotor predictors. A set of statistical predictors are trained by linear regression to predict the visual features at each visual location from proprioceptive input. By measuring each predictor's predictability using the R^2 statistics, the algorithm can determine which visual locations correspond to the robot's body parts. Visual features are extracted using biologically plausible visual motion processing models. We demonstrate that while both orientation selective and motion selective visual features can be used for self-identification, motion selective features are more robust to changes in appearance.**

I. INTRODUCTION

Dexterous visually guided movement is a natural ability that primate animals possess. The key to enabling this ability is the body schema which is an adaptive representation of the body's position in space [1]. It is acquired developmentally. Correlation between different sensory modalities plays a key role in this process as demonstrated by the "rubber hand illusion" experiment [2]. A representation of correlation between the visual consequences of action and proprioception is critical for development of visually guided reaching.

In past approaches to developing body schema in robotics, some researchers used explicit markers to identify the end effector of a robot body to simplify visual processing, and focused on learning the parameters and structure of the robot's kinematic model [3-6]. Some recent works proposed to extract visuomotor correlation by different techniques such as Hebbian-style learning, image blob category classification, Bayesian models and saliency maps [7-12].

Here we propose an approach that enables a robot to develop its body schema based on building predictors of the visual consequences of motor actions. This approach is inspired by Wolpert's prediction based model of human motor control [13]. Rather than linking proprioception or motor actions with image patches or blobs as in [7-10], our representation links proprioception to visual locations.

To extract visual features, we consider two biologically plausible vision processing mechanisms: an area V1 complex cell model [14] and an area Middle Temporal (MT) cell model [15]. While both models can be used to develop robot self-identification if the appearance of the robot is constant over time, the MT model is more robust if the appearance changes.

[1] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, HONG KONG, eebert@ee.ust.hk

[2] School of Electrical and Electronic Engineering, The University of Manchester, Manchester, UK.

II. METHOD

A. Self Identification System Architecture

The proposed robot self identification system consists of two spaces: proprioception and vision. We consider two degrees of freedom in the robot arm (elbow and shoulder joint). The two dimensional proprioception space is uniformly divided into small regions. The vision space consists of visual features extracted at each pixel in the image. We assume that the camera geometry is fixed in body centered coordinates, so that each pixel location corresponds to a gaze direction in body centered coordinates. For each pair of a proprioceptive region and a pixel, we train a visuomotor predictor: a statistical model predicting how visual features extracted from the neighborhood of the pixel change with proprioception. This local approach approximates the nonlinear mapping between the proprioception and vision spaces by a set of learned local linear predictors. If the local visual features vary systematically with the arm position, they are predictable from the local proprioception. Since the robot arm movement determines the variation in the visual features at visual locations related to the robot arm, a measure of predictability should indicate which visual locations are related to the robot arm. Here we adopt the R^2 statistics of the learned predictor as the measure local predictability. The robot's body schema is represented by the set of proprioceptive region/visual location pairs with high predictability.

B. Visual feature extraction

The first model we adopted is an orientation selective V1 complex cell model. Two dimensional Gabor filters are used to model orientation selective simple cell responses [16]. A quadrature pair of simple cell responses are summed and squared to model the responses of orientation selective complex cells [14, 17]. A normalization operation among simple cell responses is also introduced to explain cross orientation inhibition [18].

The second model we adopted is an velocity selective MT cell model [15]. It is used to extract features that are selective to image velocity, but relatively invariant to spatial texture. The basic building blocks of the MT cell model are spatiotemporal orientation and frequency selective simple cells modeled using spatial receptive field profiles which are 2D Gabor filters and temporal receptive field profiles which are sinusoids modulated by Gamma function envelopes. Simple cells are combined into spatiotemporal directional selective complex cells in a similar manner as V1 model. The velocity selectivity of MT cell model is achieved by summing spatiotemporally localized receptive fields of complex cells

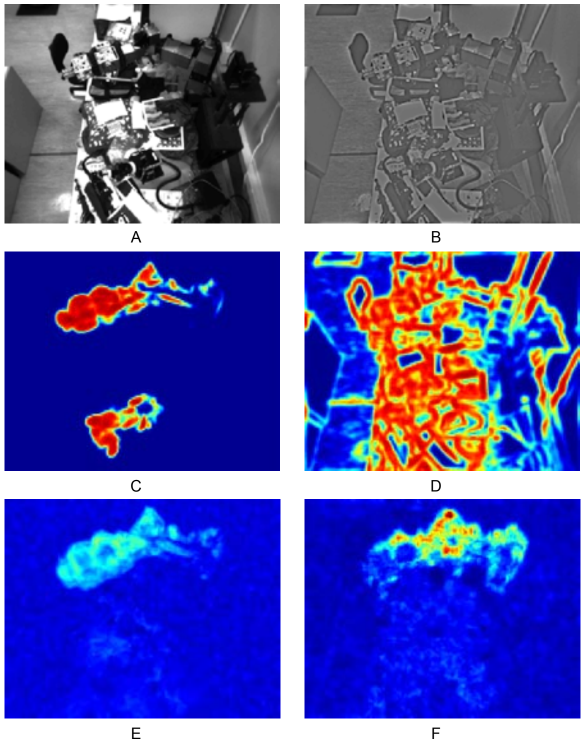

Fig. 1 Experimental results from a real robotic system. (A) An example image obtained by the camera. (B) The whitened image. (C) The L_1 norm of MT motion features at each pixel. (D) The L_1 norm of the orientation feature at each pixel. (E) The R^2 statistic for the MT motion feature; (F) The R^2 statistic of the orientation feature

that lying along a certain velocity plane in the spatiotemporal frequency space.

III. Experimental Results

In the experimental setup, a camera is set at a fixed location. A large robot arm moves its two joints inside the camera view while a small robot arm is moving randomly in the background. As the robot system only receives proprioception feedback from the large robot arm, the large robot arm is treated as “self” and is the target for self identification task. Without proprioception feedback, the small robot provides some visual distractions, which should be ignored as belonging to “others”. The two joints of the large robot arm sweep through their predefined ranges independently. Thus, we obtain a set of uniform distributed samples of joint angles. The proprioception space is divided into small local regions according to the two joint angles.

To compare the properties between the motion and the orientation visual features in real experiment, we attach different textures to the wrist and the tip of the big robot arm during the collection of training data. The collected images are then processed by the vision processing models to obtain orientation selective features and motion velocity selective features.

The result of self identification is shown in Fig. 1. From the orientation feature shown in Fig. 1D, we can see that the environment is indeed cluttered. The motion feature in Fig. 1C responds to both moving robot arms. However, the R^2 statistic of learned predictors as shown in Fig. 1E only responds to the large robot arm, and not the small robot arm. Fig. 1F shows the result of self identification using the R^2 statistic and the orientation feature. The area around the robot arm wrist and tip where the texture is changing cannot be identified, while the area near the elbow joint, which has a more consistent appearance, can be identified.

References

[1] M. Hoffmann, H. Marques, A. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer, "Body Schema in Robotics: A Review," *Autonomous Mental Development, IEEE Transactions on,* vol. 2, pp. 304-324, 2010.

[2] M. Botvinick and J. Cohen, "Rubber hands 'feel' touch that eyes see," *Nature,* vol. 391, p. 756, Feb 19 1998.

[3] M. Hersch, E. Sauser, and A. Billard, "Online Learning of the Body Schema," *International Journal of Humanoid Robotics,* vol. 5, pp. 161-181, 2008.

[4] R. Martinez-Cantin, M. Lopes, and L. Montesano, "Body Schema Acquisition through Active Learning," *IEEE International Conference on Robotics and Automation ICRA,* pp. 1860-1866, 2010.

[5] J. Sturm, C. Plagemann, and W. Burgard, "Body schema learning for robotic manipulators from visual self-perception," *Neurorobotics,* vol. 103, pp. 220-231, 2009.

[6] A. Stoytchev, "Toward Video-Guided Robot Behaviors," in *7th International Conference on Epigenetic Robotics*, 2007, pp. 165-172.

[7] R. Saegusa, G. Metta, and G. Sandini, "Own body perception based on visuomotor correlation," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 1044-1051.

[8] A. Dearden and Y. Demiris, "Learning forward models for robots," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2005, pp. 1440-1445.

[9] M. Hikita, S. Fuke, M. Ogino, T. Minato, and M. Asada, "Visual attention by saliency leads cross-modal body representation," in *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on*, 2008, pp. 157-162.

[10] C. C. Kemp and A. Edsinger, "What can i control?: The development of visual categories for a robot's body and the world that it influences," in *In 5th IEEE International Conference on Development and Learning (ICDL-06), Special Session on Autonomous Mental Development*, 2006.

[11] K. Gold and B. Scassellati, "Using probabilistic reasoning over time to self-recognize," *Robotics and Autonomous Systems,* vol. 57, pp. 384-392, 2009.

[12] Y. Yoshikawa, K. Hosoda, and M. Asada, "Does the invariance in multi-modalities represent the body scheme? - A case study with vision and proprioception," in *2nd International Symposium on Adaptive Motion of Animals and Machines*, 2003.

[13] D. Wolpert, Z. Ghahramani, and M. Jordan, "An internal model for sensorimotor integration," *Science,* vol. 269, pp. 1880-1882, 1995.

[14] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *Journal of the Optical Society of America A,* vol. 2, pp. 284-299, 1985.

[15] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vision research,* vol. 38, pp. 743-761, 1998.

[16] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J Neurophysiol,* vol. 58, pp. 1233-58, Dec 1987.

[17] A. B. Watson and J. A. J. Ahumada, "Model of human visual-motion sensing," *Journal of the Optical Society of America A,* vol. 2, pp. 322-341, 1985.

[18] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience,* vol. 9, pp. 181-197, 1992.