Visuo Motor Control Learning using Gated Networks

Julien Abrossimoff[1], Alexandre Pitti[1] and Philippe Gaussier[1]

***Abstract*— We present an auto-encoder version of gated networks for modeling inverse-forward transformations in the visuomotor space of a robot arm. Gated networks learn the motor correspondence between related images based on multiplicative interactions. Used as an inverse model, each output neuron encodes separately one motor transformation, a motor primitive, which are combined for visual reaching. Used as a forward model, the gated network serves then for action-based prediction of visual motion for estimation of arm location, aka motor 'imagination'. We apply its features in three experiments including the learning of visual transformations from motor primitives, the construction of the robot arm's peri-personal space and the control of the inverse-forward network for reaching nearby objects. We discuss then the pertinence of our network to related works and its relevance for modeling the so-called gain-field neurons in the parieto-motor cortices for visuomotor transformation, action observation and action understanding of physical effects.**

I. INTRODUCTION

In image processing, multiplicative or sigma-pi networks have been emphasized recently because product images can be viewed as energy functions that can be discriminated better than deep networks on affine transformations for example; see Fig. 1 a). Recently, this technique has been applied extensively by Memisevic to the learning of optical flow, of rotational shifts as well as of spectral filters and spatio-temporal patterns for action recognition [1], [2], [3]. In these researches, the transformation from unseen images is assumed to be hidden to the experimenter and is estimated afterwards [4].

In the case of robotics, however, we know exactly *which* actions have been performed and *which* effects have been caused on sensors. In regard to image processing, the extra information of having motors –, that is, embodiment,– can serve multiplicative networks to estimate better which transformation has been performed between co-variating sensorimotor signals (inverse model) and to predict better how the sensory signals will evolve based on the learned transformations (forward model); resp. the blue and red networks in Fig. 1 c).

On the one hand, these multiplicative networks can be used in robots to learn motor primitives to estimate the minimal motor transformation necessary to control co-variating incoming sensory signals in one desired state (e.g. to move in one visual direction) [5]. On the other hand, they may serve to simulate motor activity in order to infer the most plausible transformations performed from two or more incoming variables; e.g., as during action observation [6].

[1] ETIS, UMR 8051 / ENSEA, University of Cergy-Pontoise, CNRS, F-95000, Cergy-Pontoise, France `surname.name@u-cergy.fr`

To our knowledge, only few teams proposed recently the use of gated networks in robotics [6]; respectively Sigaud and colleagues for categorization and retrieving of motor sequences with the ICub robot [7], [8] and Pitti and colleagues for learning audio-visual integration in a robot head for speech location [9] and visuomotor transformation in a robot arm for body representation [10], [11].

In this paper, we propose to exploit these characteristics of multiplicative networks also known as gain-field or gated networks in robots for action-based prediction of visual transformation, action recognition, spatial representation and visuomotor control; see Fig. 1 c).

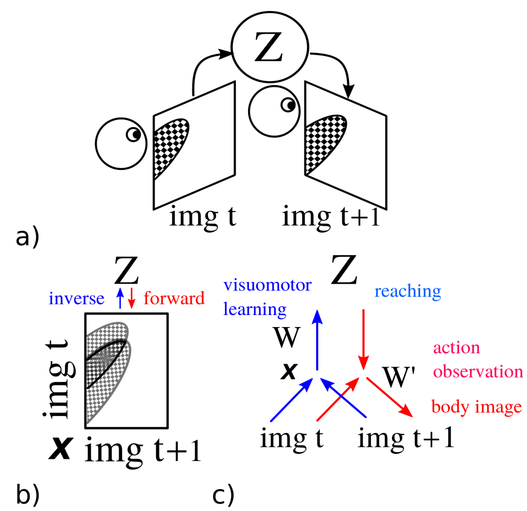

Fig. 1. Description of our architecture and its use for inverse-forward predictions in robotics. (a) The transformation **Z** between two related images at time t and $t+1$ can be learned and estimated by processing the by-product of the two images as an energy landscape, which is seen in (b) as the intersection of the two images. The gated network can be used as an auto-encoder to reconstruct back a predicted image; resp., the black lines and the red lines in (c). In robotics, the estimated **Z** function corresponds to the applied motor activity during hand motion for instance. This information can serve to select the correct primitive **Z** in order to reach one desired target in space (inverse model) and/or to estimate the location of the hand position when a motor primitive **Z** is applied (forward model).

II. CONCLUSION

We have presented auto-encoder gated networks for learning visuo-motor transformations and inverse-forward models. Gated networks are based on multiplicative interaction between related images for infering the corresponding transformation. In robotics, these transformations can be discriminated easily as they correspond to the robot motor activity, which can be learned through supervised learning. Used as an inverse model, the network predicts the most

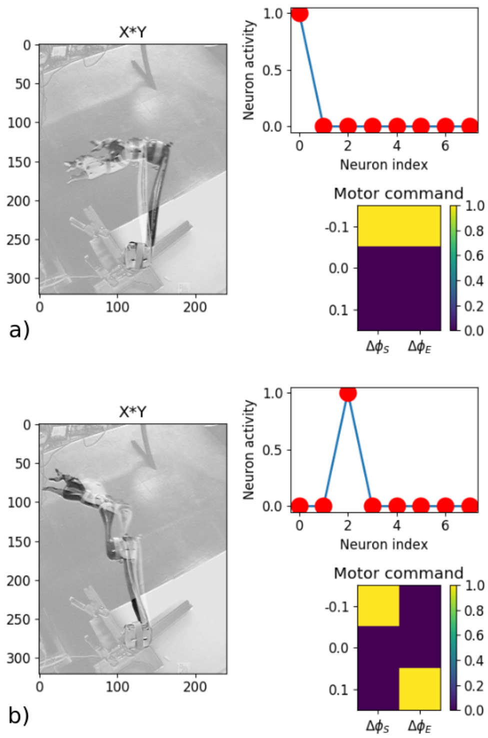

Fig. 2. Action-based prediction of visual motion at two different locations. The inverse gated network is able to discriminate the visual motion between two related images X and Y through multiplicative function (left chart). The prediction of the associated motor unit is presented in the activity level of top chart and the corresponding motor command table is displayed in the bottom chart. In a), motor unit $\#0$ estimation of the visual displacement of the robotic arm on the upright direction. In b), motor unit $\#2$ estimation of the visual displacement of the robotic arm on the upper-left direction. The correspondence between the visual transformation and the associated motor units are nonlinear (see text).

probable motor command within a repertoire of eight motor primitives from its visual motion. The network performs few motor errors and can be employed on reaching tasks. Used as a forward model, the network uses its repertoire of eight motor primitives to perform 'motor simulation' in order to geta density distribution of the robot arm's spatial location; i.e., its peripersonal space. The combination of the inverse and forward networks permits to track objects within reach.

REFERENCES

[1] R. Memisevic, "Learning to represent spatial transformations with factored higher-order boltzmann machines," *Neural Computation*, vol. 22, pp. 1473–1493, 2010.
[2] ——, "Gradient-based learning of higher-order image features," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1591–1598.
[3] ——, "Learning to relate images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1829–1846, 2013.
[4] O. Sigaud, C. Masson, D. Filliat, and F. Stulp, "Gated networks: an inventory," *arXiv:1512.03201v1*, 2016.
[5] P. Baraduc, E. Guigon, and Y. Burnod, "Recording arm position to learn visuomotor transformations," *Cerebral Cortex*, vol. 11, p. 906917, 2001.
[6] D. Bullock, S. Grossberg, and F. Guenther, "A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm," *Journal of Cognitive Neuroscience*, vol. 5, no. 4, pp. 408–435, 1993.

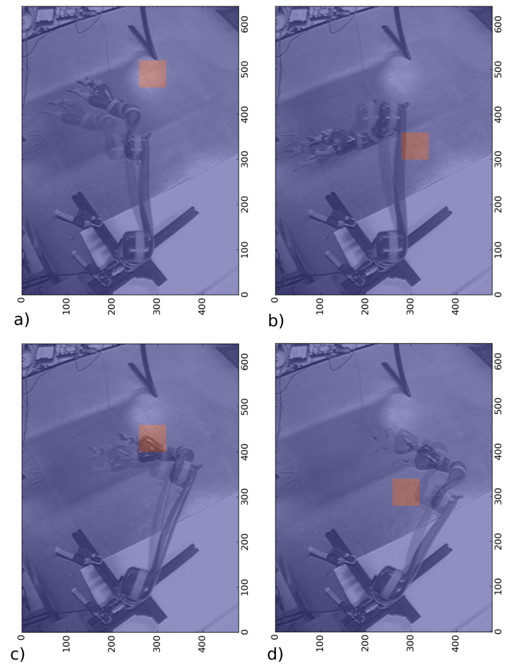

Fig. 3. Body-centered reaching for different locations. The fading representation corresponds to initial posture of the robot arm. The red square correponds to the goal. The final posture of the robot arm is also displayed.

[7] A. Droniou, O. Sigaud, and I. Serena, "A deep unsupervised network for multimodal perception, representation and classification," *Robotics and Autonomous Systems*, vol. 71, p. 8398, 2015.
[8] O. Sigaud and A. Droniou, "Towards Deep Developmental Learning," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 2, pp. 99–114, 2016.
[9] A. Pitti, A. Blanchard, M. Cardinaux, and P. Gaussier, "Gain-field modulation mechanism in multimodal networks for spatial perception," *12th IEEE-RAS International Conference on Humanoid Robots Nov.29-Dec.1, 2012. Business Innovation Center Osaka, Japan*, pp. 297–302, 2012.
[10] S. Mahe, P. Braud, R. Gaussier, M. Quoy, and A. Pitti, "Exploiting the gain-modulation mechanism in parieto-motor neurons application to visuomotor transformations and embodied simulation," *Neural Networks*, vol. 62, pp. 102–111, 2015.
[11] R. Braud, A. Pitti, and P. Gaussier, "A modular dynamic sensorimotor model for affordances learning, sequences planning and tool-use," *IEEE TCDS*, p. to appear, 2017.