Neural Dynamic Architecture for Action Parsing using SCAMP.

Julien Martel[1], David Lobato[2], and Yulia Sandamirskaya*[1]

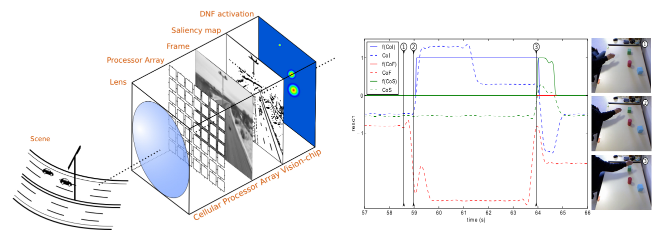
Neural networks have received a lot of attention recently for their success in pattern classification, leading to systems capable to recognise objects in images and video streams, or to understand speech. However, application of neural networks to represent and eventually learn actions and action plans, required in robotics, is not straight-forward, since common structure among actions is rather abstract (e.g., the roles of subject and object in an action), while on the level of appearance, even actions from the same class might differ considerably. Consequently, more structure needs to be imposed on a neuronal network in order to realise a neuronal action representing system.

In our previous work, we have explored neural-dynamic framework in the domain of representing and learning actions and action sequences. In this paper, we combine this neural-dynamic framework with efficient on-sensor processing to achieve real-time performance, which will help to scale-up the framework to real-world scenarios.

The neural-dynamic framework uses dynamic neural fields (DNFs) [9] to represent each action as a tuple of intention and condition of satisfaction (CoS) neural fields. Both intention and CoS neural fields are coupled to sensors and motors of the robotic agent: activation of the intention of an elementary behavior (EB) leads to initiation of an action by shaping an attractor for the motor dynamics, whereas CoS is activated by sensory input that signals the successful accomplishment of the action.

In our previous work, we have demonstrated how such EB-modules can be used to generate flexible – i.e. adaptive to changing environmental situations – sequences of object-directed actions [8]. We have demonstrated how such sequences can have hierarchical structure [3], how planning [1], reinforcement learning [4], and action parsing [5] can be realised in this architecture. Moreover, the same framework has been used in a system, capable to generate spatial language descriptions of a scene [7].

All these examples used either very simplified robotic settings or simulations to demonstrate functionality of the system. Here, we present our first steps towards scaling-up this method to real-time processing by using a focal-plane processor of a smart camera SCAMP-5 [2]. On SCAMP, each pixel is able to perform simple "arithmetic" operations on a local analog processor. All processors operate simultaneously on the data captured at each pixel. Along with a resistive network for direct communication with neighbouring pixels, SCAMP fits nicely as substrate for an ultra-fast implementation of neural-dynamics, along with image processing, achieving several hundreds iterations at 25 FPS [6].



(a) (b)

Fig. 1. **Left:** DNF implemented on SCAMP [6]. **Right:** Neural-dynamic architecture for action parsing, recognising an action "reach for green object". CoF: condition of failure, CoI: condition of initiation, CoS: condition of success; lines show activation of the respective neuronal nodes [5].

Here, we demonstrate how SCAMP device can be used as a sensory front-end for the neuro-dynamic action parsing architecture [5], providing fast detection of salient regions and movement patterns, required to detect the onset and the end of object-directed actions in a table-top scenario (Fig. 1).

We believe that this work is a first step towards real-time implementation of the neuro-dynamic architecture for action understanding [5], which can not run efficiently on conventional computing platforms. Implementation of the neural-dynamic architecture on a massively parallel focal-plane processor will facilitate its deployment in real world applications[1].

References

[1] E. Billing, R. Lowe, and Y Sandamirskaya. Simultaneous planning and action: Neural-dynamic sequencing of elementary behaviours in robot navigation. *Adaptive Behvaior*, pages 1–22, 2015.

[2] S. J. Carey, D. R. W. Barr, A. Lopich, and P. Dudek. A 100'000 fps vision sensor with embedded 535 GOPS/W 256×256 SIMD processor array. In *Proc. of the VLSI Circuits Symp.'13*, pages C182–C183, 2013.

[3] B Duran and Y Sandamirskaya. Neural dynamics of hierarchically organized sequences: a robotic implementation. In *Humanoids*, 2012.

[4] S. Kazerounian, M. Luciw, M. Richter, and Y. Sandamirskaya. Autonomous reinforcement of behavioral sequences in neural dynamics. In *IJCNN*, 2013.

[5] D Lobato, Y Sandamirskaya, M Richter, and G Schöner. Parsing of action sequences: A neural dynamics approach. *Paladyn*, 6, 2015.

[6] J N P Martel and Y Sandamirskaya. A neuromorphic approach for tracking using dynamic neural fields on a programmable vision-chip. In *Proceedings of the 10th International Conference on Distributed Smart Camera (ICDSC), ACM*, 2016.

[7] M Richter, J Lins, S Schneegans, Y Sandamirskaya, and G Schöner. Autonomous neural dynamics to test hypotheses in a model of spatial language. In *The Annual Meeting of the Cognitive Science Society, CogSci*, 2014.

[8] M. Richter, Y. Sandamirskaya, and G Schöner. A robotic architecture for action selection and behavioral organization inspired by human cognition. In *IROS*, 2012.

[9] G. Schöner and J. P. Spencer, editors. *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press, 2015.

[1]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland, [2]VISLAB, University of Algarve, Faro, Portugal *Contact: ysandamirskaya@ini.uzh.ch

[1]SNF grant PZOOP2_168183_1 "Ambizione" is gratefully aknowledged