Attentive Robot Vision

Amirhossein Jamalian, Frederik Beuth and Fred H. Hamker

Technische Universitat Chemnitz, Artificial Intelligence
amirhossein.jamalian@informatik.tu-chemnitz.de
frederik.beuth@informatik.tu-chemnitz.de
fred.hamker@ informatik.tu-chemnitz.de

Visual attention is a smart mechanism to omit the processing of unnecessary data coming into primate's visual system. Based on this mechanism, primates can rapidly find and focus on the searched object in their visual field in presence of several dictractors and in front of a cluttered background. These find and focus processes are performed in the primates' brain through *ventral* and *dorsal streams* depicted in Figure 1 a. The former is involved in the object recognition task whereas the latter is the responsible of saccade generation and saliency map formation (Beuth & Hamker, 2015a).

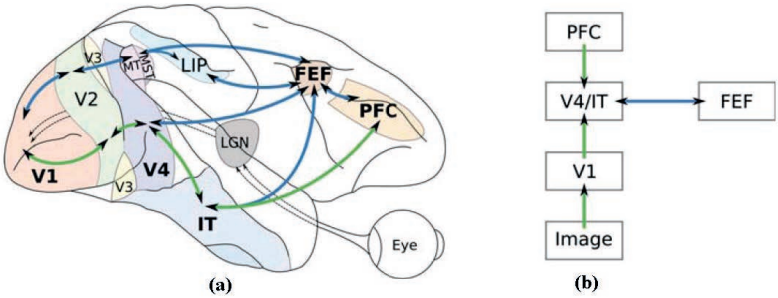

Figure 1. a) Primates' Visual System with ventral and dorsal streams, green and blue arrows respectively. b) Areas and connections which are simulated in the computational model. V1, V4, IT, PFC and FEF are primary visual cortex, fourth visual cortex, inferior temporal cortex, prefrontal cortex and frontal eye field respectively.

Heretofore, the biologically plausible computational model of visual attention, which is an implementation of the ventral stream, has been developed based on primates' ventral stream (Beuth & Hamker, 2015b). Inspired by cortical architecture, this model uses a structure with two layers to implement amplification, normalization and suppression as well as spatial pooling. The diagram of its layers and corresponding connections is illustrated in Figure 2. The input image immediately is preprocessed (as the effect of the Lateral Geniculate Nucleus (LGN)) and stays in the neurons of V1 layer. Then, it is given to the cycle of HVA-FEF to recognize and localize the searched object concurrently (HVA is abbreviation of Higher Visual Area whereas FEF is abbreviation

of Frontal Eye Field). Therefore, it could be categorized in the class of iterative models mentioned in (Jamalian & Hamker, 2016). HVA is the combination of V2, V3, V4 and IT regions in the brain. The performance of current model in Object Recognition of COIL-100 dataset (Nene, Nayar, & Murase, 1996) is 92%, 72% and 42% in black, white-noise affected and real-world backgrounds respectively (Beuth & Hamker, 2015a).

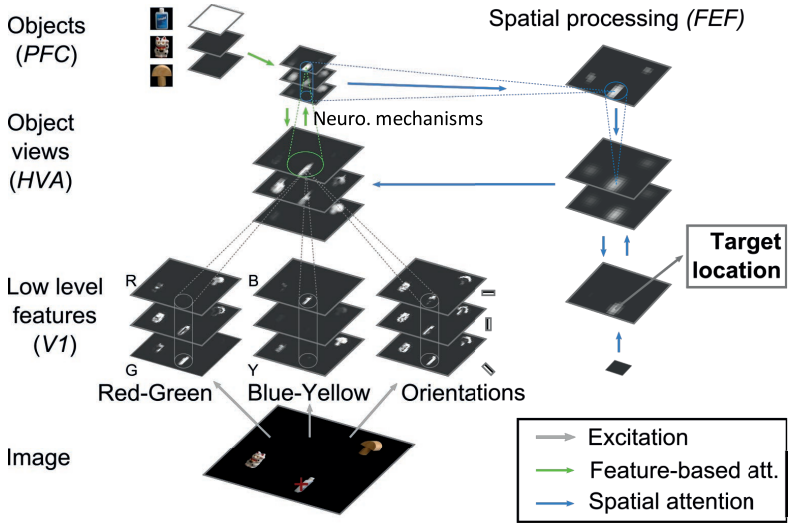

Figure 2. The computational model of visual attention in search of a bottle in COIL-100 dataset (Nene, Nayar, & Murase, 1996). The picture is adapted from (Beuth & Hamker, 2015a).

Recently, this model, individually as well as integrated with the model of dorsal stream, has been tested using a neuro-cognitive agent (Hamker, 2015) in virtual reality (VR) within the European Union project so-called *Spatial Cognition*. This is the first time that the robustness of such models against large scaling for the visual search task (Wolfe, 1994) in semi real-world scenarios is demonstrated. Few models have already been demonstrated with (semi) real-world objects (Chikkerur, Serre, Tan, & Poggio, 2010) & (Walther & Koch, 2007). Yet, they have mostly used static input material.

The VR is based on the VR engine Unity[1], while the neuro-cognitive agent is implemented using ANNarchy which is a new neural network simulator reported in (Vitay, Dinkelbach, & Hamker, 2015). We designed a set of scenarios to test the capability of the agent in object recognition and localization. In each single scenario, the agent has to execute a classical visual search task (Wolfe, 1994), i.e. search a target object among distractors. For instance, Figure 3 a. demonstrates the scene in which the agent should search for a yellow crane truck (target) while all other objects serve as

[1] https://unity3d.com/

distractors. The set of scenarios contains 9000 different tasks; each is a visual field with at least 3 objects (one target and a couple of distractors). The performance of the model at this test set is depicted in Figure 3 b as a confusion matrix. As it can be seen in this figure, in 85% of the cases the model can localize the target correctly (the saccade landed within the object borders or not more than 50 pixels away).

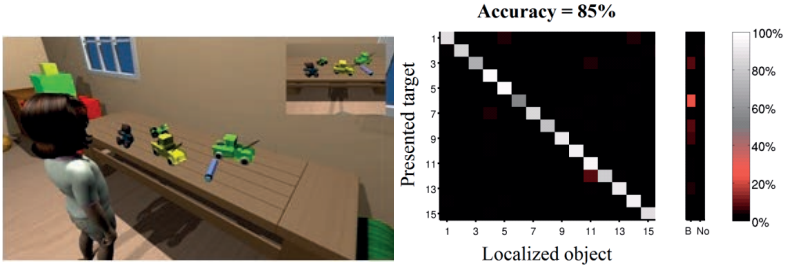

Figure 3. a) Neuro-cognitive agent in VR performing the visual search task. b) The performance of the object recognition / localization model based on 9000 scenarios.

In near future, first, we will implement the model on the iCub robot (a humanoid robot with the capability of saccade) to evaluate its performance in real-world scenarios. Then, we will improve the model based on the following ideas:

- Propagating the feature-based attention back to the V1 layer which is absent in current model. The idea is that, this attention signal could make the model more robust against noises and effect of distractors.
- Considering the robustness of the model against external and internal noises. It has been shown that the proper amount of noise can have positive effect in human visual attention through the mechanism called Stochastic Resonance (Kitajo, Yamanaka, Ward, & Yamamoto, 2006).
- Implementing the dynamic receptive field inside the model.

Acknowledgement

This work has been supported by the European Union's 7th Framework Program (FET, Neuro-Bio-Inspired Systems: Spatial Cognition) under grant agreement *n*° 600785.

References

Beuth, F., & Hamker, F. H. (2015b). A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vision Research* , 241-257.

Beuth, F., & Hamker, F. H. (2015a). Attention as cognitive, holistic control of the visual system. *Workshop of New Challenges in Neural Computation (NCNC 2015)*, (pp. 133-140). Aachen.

Chikkerur, S., Serre, T., Tan, C., & Poggio, T. (2010). What and where: a Bayesian inference theory of attention. *Vision Research , 50* (22), 2233-2247.

Hamker, F. H. (2015). Spatial Cognition of humans and brain-like artificial agents. *Künstliche Intelligenz* , 83-88.

Jamalian, A., & Hamker, F. H. (2016). Biologically-Inspired Models for Attentive Robot Vision: A Review. In R. Pal, *Innovative Research in Attention Modeling and Computer Vision Applications* (pp. 69-98). Hershey, PA: Information Science Reference.

Kitajo, K., Yamanaka, K., Ward, L. M., & Yamamoto, Y. (2006). Stochastic resonance in attention control. *Europhysics Letters* , 1029-1035.

Nene, S. A., Nayar, S. K., & Murase, H. (1996). *Columbia Object Image Library (COIL-100)*. Technical Report: CUCS-006-96.

Vitay, J., Dinkelbach, H. Ü., & Hamker, F. H. (2015). ANNarchy: a code generation approach to neural simulations on parallel hardware. *Frontiers in Neuroinformatics* , 1-20.

Walther, D. B., & Koch, C. (2007). Attention in hierarchical models of object recognition. *Progress in Brain Research , 165*, 57-78.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review* , 202-238.