Learning of Invariant Representations of Visual Objects in a Model of the Visual System

Reinhard Eckhorn, Basim Al-Shaikhli, Timm Zwickel, Frank Michler, Thomas Wachtler
NeuroPhysics Group, Department of Physics, Philipps-University, Marburg, Germany
reinhard.eckhorn@physik.uni-marburg.de

Introduction and Goal. We recognize known objects independent of their projection onto the retinal surface, and we perceive locations in visual space as stable even though their retinal projections change while we move. Thus, our visual system generates invariant representations of visual objects and their locations from variant retinal projections. How are invariant representations achieved? Different views of the same object are more likely to occur in temporal neighborhood than in any random order. Földiák [1] was the first who used this fact for learning of transformation invariances in a neural network. Wiskott and Sejnowski [2] pointed out that invariant features of visual objects change more slowly than other features. We extended these ideas by starting to develop a biologically plausible model of the visual system in which, during motion of an observer, the *slowly changing features* of visual objects (ventral visual pathway) and their *locations* (dorsal pathway) are extracted by the spatio-temporal properties of the neural network so that they can selectively be used for learning invariant aspects of visual objects. Here, we focus on three major aspects of this approach.

1. ***Learning the retino-cortical representation of visual space*** from dynamic visual inputs due to self-motion. In a first step we used the known retino-cortical mapping and the simplifying assumption that during self-motion object representations cross the same cortical distance within a fixed time interval, independent of retinal eccentricity. This enabled us to calculate the required locations of stimuli in visual space that would activate the cortical neurons optimally. These locations lie close to an arc of an ellipse in front of the observer's view. In a second step we developed a two-level neural network of spiking neurons [3,4] consisting of a retinal input and a cortical projection layer. When the flow field induced by objects at these optimal locations was used as visual input, learning of these sequences with a Hebbian learning rule changed the initially fully forward-connected two-layer model to an inhomogeneous retino-cortical representation in accordance with the real retino-cortical mapping. A high density of small cortical receptive fields (RF) emerged in the central and fewer large RFs in the peripheral representation of the visual field (for details see Abstract *Al-Shaikhli et al.*).

2. ***Coding the presence of visual objects independent of their identity*** by a multilayer cortical network with fast feed-forward, backward and lateral connections. In the lowest layer local contour detectors (like simple cells in V1) with the same preferred contour orientation and location, but encoding opposing object-directions, mutually inhibit each other. Neurons at different positions of the spatial representation are laterally connected by modulatory and inhibitory connections, supporting sensitivity of the representation to Gestalt properties (including *similarity, good continuation, and convexity*). Inter-layer feed-forward connections converge to the highest area in which the location of a visual object is represented largely independent of its form. Top-down feedback modulates the activity of neurons in the lowest layer when they code part of the contour of a visual object, leading to response specificities similar to those of border-ownership (BO) neurons in visual cortex, as described by Zhou and coworkers [5]. The top-down support of object contours improves figure-ground segregation of visual objects markedly. This capability of good figure-ground-segregation is an important condition in our project for learning the invariant properties of visual objects in realisitic scenes (for details see Abstract *Zwickel et al.*).

3. ***Learning of view-independent object representations*** from views during self-motion and ocular fixation of objects (*closed* perception-action *loop*). In a 3-layer model of pulse-coding neurons the *sensory layer* (**S**) receives sequences of visual inputs of fixated objects. The *direction layer* (**D**) codes

for the direction of such sequences, generated during *right*- or *left*-ward passing the fixated object. The *conjunction layer* (**C**) combines the information from both S and D. C is a hetero-associative memory learning the views and their successive occurrence (sequence) such that the views of an object are systematically connected to their neighbor views, forming a continuously connected representation of a visual object. Through learning, each view is most strongly connected to the views that follow when passing the object in right or left directions during fixation. The activation of the directional input D can therefore activate the next (most probable) view or initiate *mental rotation* of the object (for details see Abstract *Michler et al.*).

Discussion and Conclusions. We approach the task of learning invariant object representations from three different directions. All have in common basic principles necessary for the robust generation of invariances.

First, we have shown that the inhomogeneous retino-cortical mapping can be explained by assuming that the cortical representations of objects in visual space travel the same distance, independent of eccentricity, under self-motion of an observer (*open action-perception loop*). This would allow the visual cortex to use the same connectivity and dynamic neural properties, invariant with eccentricity. Interestingly, the inhomogeneous mapping can be achieved qualitatively in a minimal biologically realistic network model through competitive Hebbian learning. In a next step we will test the hypothesis that the inhomogeneous mapping plays an important role in the formation of invariant object representations during active exploration of the environment (closed action-perception loop).

Second, we developed a model in which visual objects are detected invariant of their form (identity), corresponding to the dorsal visual cortical pathway. In this part of our modelling network the representation of a visual object feeds back to the lowest layer, in which neurons represent an object at its specific contour (variant object representation), enhancing those contour neurons that belong to the object. As this feedback in the model appears at short latencies (corresponding to the magnocellular retino-cortical afferents and the dorsal cortical visual pathway) it can label an object's contour early enough to enhance the slower parvocellular input evoked by the same object. Such an interaction of dorsal and ventral visual pathways improve the difficult and important task of figure-ground-segregation which is essential for learning the invariances of visual objects.

Third, we showed that learning recurrent connections between those representations of object views that appear often close in time enables view-independent object recognition. In contrast to the pure feed forward generation of view invariances in other models [1,6,7], our model can also account for the prediction of view changes and mental rotation.

References

[1] Földiák P (1991) Learning invariance from transformation sequences. Neural Computation 3: 194-200

[2] Wiskott L, Sejnowski T (2002) Slow feature analysis: Unsuprervised learning of invariances. Neural Computation 14:715-770

[3] Eckhorn R, Reitboeck HJ, Arndt M, Dicke P (1990) Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. Neural Computation 2:293-307

[4] Saam M, Eckhorn R (2000) Lateral spike conduction velocity in the visual cortex affects spatial experience: a learning model with spiking neurons. Biol Cybernetics 83:L1-L9.

[5] Zhou H, Friedman HS, Heydt Rv (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20:6594-6611

[6] Wallis and Rolls (1996) A model of invariant object recognition in the visual system. *Prog Neurobiol.*, 51:167-194

[7] W. Einhäuser, C. Kayser, P. König, and K. P. Körding. (2002) Learning the invariance properties of complex cells from their responses to natural stimuli. *European Journal of Neuroscience*, 15:475-486

Supported by Deutsche Forschungsgemeinschaft DFG (FOR 254) to R.E and T.W.