Sequence Learning as Basis for Invariant Recognition of Object Views

Frank Michler, Reinhard Eckhorn, Thomas Wachtler

Neuro Physics Group, Department of Physics, University Marburg, Germany
Frank.Michler@physik.uni-marburg.de

Introduction

Our visual system is able to recognize familiar objects from different viewing angles, even though retinal projections of the same object can vary drastically. Conversely similar retinal projections can be caused by different objects. How does the visual system assign the different object views to the correct objects? Powerful cues to solve this problem are temporal and spatial contiguity of an object's retinal projection: when we fixate an object during self motion, the identity of the object remains the same while its retinal projection changes continuously. One approach to take advantage of these systematically changing spatio-temporal correlations is to learn them in a convergent feed-forward network, where after the learning process the representations of different views of one object project to the same output neuron (*cardinal cell representation* [1,2,3]). But in such an approach the spatial and temporal relations between object views, as they occur during self-motion and fixation of the object, cannot be represented.

In order to capture the spatio-temporal order of object views we propose a learning mechanism which connects representations of views by modulating their coupling strengths according to the spatial and temporal distances of the local object features as they occur during self-motion and fixation. These distances are related to the neighborhood in the aspect graph of an object [4]. Our model also provides a basis for invariant recognition because the representation of a single view can reactivate the representations of the other views of the same object. Furthermore it can account for predictive activation of representations of probable future stimuli and mental rotation of object views.

Methods

We demonstrate the proposed mechanism in a network model with pulse coding neurons [5,6]. Synaptic weights are modified with a STDP-like learning rule [7]. We simulate a training phase and a recall phase. As input for the training phase we use pattern sequences which correspond to successive views of visual objects during self-motion in passing a fixated object at its left or right side. This is simulated using a Markov model with two internal state variables, one representing the current object view and the other the movement direction. After presentation of one view a transition occurs to another view: With high probability the self-motion direction is maintained and the neighboring view of this object is presented. With low probability the self-motion direction is changed or the input switches to a view of another object.

The network has two input layers. The *sensory layer* (***S***) is fed with the *visual* patterns; the *direction layer* (***D***) receives the information about the current moving direction (passing the object at left or right side). ***S*** has fast auto-associative connections which enable pattern completion. ***S*** and ***D*** have fast reciprocal connections to a *conjunction layer* (***C***) which combines the sensory and directional information. ***C*** has delayed hetero-associative connections to learn the temporal transitions between the input patterns (object views).

Results

After training the neurons in layer ***C*** have learned to respond selectively to activity patterns in ***S*** and ***D,*** which corresponds to the information about object views and movement direction. The learned feedback connections from ***C*** to ***S*** are mostly reciprocal to the feed-forward connections. When the network is stimulated briefly with a single object view the internal dynamics can recall the other views that belong to the same object. This recall continues until another stimulus is presented. In the presence of a static visual input the sensory input dominates the activity of the network and the internal dynamics only generate the actual object representation and predicts by a preactivation the most probable next view, while in the absence of strong visual input the internal dynamics can recall the stored spatio-temporal sequence as it has been learned.

Discussion

Although correlated sequences of object views have already been used for learning invariant object representations [1,2], in these studies the learning resulted in cardinal neurons which respond to all views of an object while ignoring the spatio-temporal relations between the views. We demonstrate learning of distributed object representations in which consecutively appearing views are systematically mapped so that the representation of one view can activate the neighboring ones, depending on the learned sequence. During ongoing visual stimulation this predictive information can improve signal to noise ratio for the perception of the next stimulus. In the absence of visual input the internal dynamics can support mental rotation, which is not possible with cardinal cell representations.

The proposed mechanism is not a contradiction to the cardinal neuron approach but a complementation. We hypothesize that for the learning of view invariance the formation of associations between object views by means of sequence learning is an early step and provides a basis for the later formation of cardinal cell representations.

References

[1] P Földiak (1991). Learning invariance from transformation sequences. Neural Comput 3: 194–200
[2] G Wallis, E T Rolls (1996). A model of invariant object recognition in the visual system. Progr Neurobiol 51:167–194
[3] W Einhäuser, C Kayser, P König, K. P. Körding (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. Europ J Neurosci 15:475–486
[4] JJ Koenderink, AJ van Doorn (1979). The Internal Representation of Solid Shape with Respect to Vision, Biol Cybernetics 32: 211-216
[5] R Eckhorn, H J Reitboeck, M Arndt, P Dicke(1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. Neural Comput 2:293–307
[6] M Saam, R Eckhorn (2000). Lateral spike conduction velocity in the visual cortex affects spatial range of synchronization and receptive field size without visual experience: A learning model with spiking neurons. Biol Cybernetics 83:L1–L9
[7] G Bi and M Poo (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472

Acknowledgments.
Supported by Deutsche Forschungsgemeinschaft (Ec 53/11-1) to R.E. and T.W.