Learning invariant representations in the visual system from natural stimuli

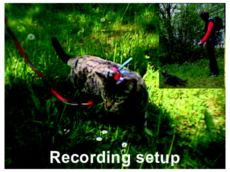
Wolfgang Einhäuser, Christoph Kayser, Konrad P. Körding & Peter König
Institute of Neuroinformatics - UNI/ETH Zürich - Switzerland (weinhaeu@ini.phys.ethz.ch)

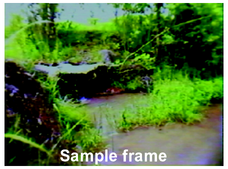
1 Introduction

Sensory systems extract multiple features in parallel. This implies that different groups of neurons exhibit specificity to different features, while being invariant to others. How can this invariance be explained by neurons being optimized for the representation of the stimuli they naturally encode? And how can an initially homogenous neuronal population split in groups displaying such distinct selectivities?

We show here that optimizing the principle of stability on natural visual input, yields the invariance properties of complex cells. Furthermore, the same principle can explain the desired functional segregation of neurons into one group specific to one feature and invariant to another, and a second group with the complementary response properties.

2 Natural Stimuli





All of the presented models are trained with stimuli obtained by a camera mounted to the head of a freely behaving cat. These 'CatCam' stimuli preserve the spatial and temporal structure of real world input to an animals visual system.

3 Objective functions

The stability objective function is maximized by neurons, whose responses vary slowly over time. For each neuron it is formulated as squared temporal derivative, which is divided by the temporal variance to avoid trivial solutions.

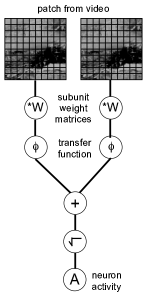
$$\Psi_{stab} = -\sum_{\substack{i \text{ over all} \\ \text{neurons}}} \left\langle (A_i(t)-A_i(t-\tau))^2 \right\rangle_t / \mathrm{var}_t(A_i);$$ with A_i activity of i-th neuron

As stability is separated for each neuron, lateral coupling is introduced by adding a de-correlation term, which favors dissimilar receptive fields:

$$\Psi_{decorr} = -\sum_{\substack{i \text{ over all} \\ \text{neurons}}} \sum_{i<j} CC^2_{ij};$$ where CC denotes the correlation-coefficient

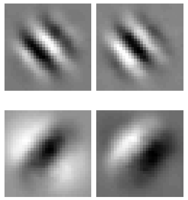
The sum of both functions is maximized by adaptive gradient ascent.

4 Cell models



Each neuron consists of n subunits. The activity of a subunit at a specific timepoint is given by the inner product of its weightmatrix W (receptive field) with the input passed through a non-linear transfer-function φ. The activities of the subunit are then added to yield the activity of the neuron A . The data shown here are obtained with n=2, $\phi(x) = x^2$ (two subunit energy detector, panels 5&6) and n=1, $\phi(x) = x^2$ (panel 8).

5 Simple & Complex cells

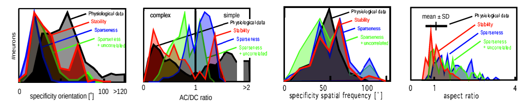


subunits of two example complex cells, that emerge in the simulation after convergence

After training the model neurons with gray-scale CatCam stimuli, the subunits acquire Gabor like receptive fields, as typical for simple cells in V1.

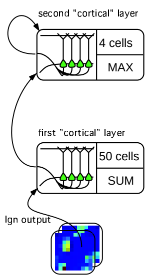
For most of the neurons their subunits obtain a relative phase-difference of 90 degrees. Thus the complete neuron's activity is insensitive (invariant) to phase and polarity of a stimulus, a characterizing property of cortical complex cells.

6 Comparison to physiology and other objectives



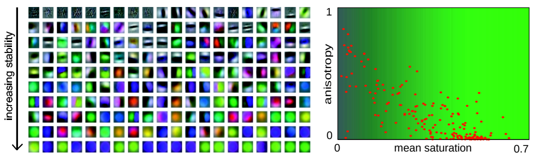
The proposed stability objective matches most properties typically measured in electrophysiolgical experiments better than other commonly employed objective functions Especially, only stability shows a clear bimodality in the response-modulation (AC/ DC) ratio, that is usually used to classify simple from complex cells.

7 Physiological Implementation



The stability objective can be implemented by physiological mechanisms. Our recently proposed model (Eur. J. Neurosci.15 (3) 475-486) employs physiological results on neurons with two integration sites, separating the learning signal from a signal gating learning to implement timing dependent plasticity. This model is shown to lead to the emergence of simple and complex cells when applied to the natural videos.

8 Functional Segregation of Visual Pathways



Using a single subunit cell model and colored stimuli, chromatic as well as achromatic receptive fields emerge. The chromatic cells tend to be non-oriented, while the achromatic show a pronounced orientation-tuning. The individual contribution of a cell to the stability objective serves as inherent criterion to separate those two groups from each other.

9 Conclusions

The stability objective successfully explains how properties of the early visual system are optimized for processing natural input:

- stability leads to the emergence of simple as well as complex cell receptive fields
- the distribution of those better matches the physiological data than other recently proposed goal-functions
- stability can segregate different stimulus-dimensions, e.g. orientation from color
- stability can be implemented by physiological mechanisms

These properties of the stability make it a promising objective, not only for the explanation of early visual system properties, but also to yield predictions on higher areas.

This work was supported by Honda R&D Europe (Germany), the Center of Neuroscience Zurich (ZNZ), the Collegium Helveticum, the Swiss National Fund (SNF, grant-no: 31-65415.01) and the FENS/EJN travel grant.

References: Einhäuser et al. (2002) Eur. J. Neurosci. 15(3):475-486;
Einhäuser et al. (2002) Network, submitted manuscript