

# Learning from natural videos: From complex cells to object classification.



W. Einhäuser, C. Kayser, M. Franzius, J. Hipp, G.U. Moeller, K.P. Körding & P. König  
 Institute of Neuroinformatics (University of Zurich/ETH Zurich)



## 1 Introduction

When proceeding through the hierarchy of the visual system, neuronal responses tend to get more and more selective to some stimulus features, but at the same time more and more independent from other features. This raises the question, which principles underlie this interplay of invariance and selectivity throughout the cortical hierarchy? Using natural videos recorded by a camera mounted to the head of a freely behaving cat as well as standard image libraries, we train neurons to achieve optimally stable responses. Using this implementation of the 'temporal coherence' principle, we replicate properties of the early visual system and transfer the same principle to invariant object recognition.

## 6 Simple and Complex Cells

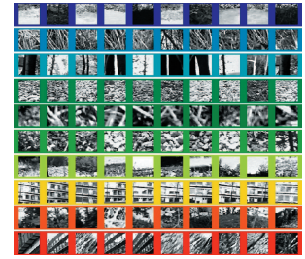
After training the model neurons with gray-scale CatCam stimuli, the subunits acquire Gabor like receptive fields, as typical for simple cells in V1. For most of the neurons their subunits obtain a relative phase-difference of 90 degrees. Thus the complete neuron's activity is insensitive (invariant) to phase and polarity of a stimulus, a characterizing property of cortical complex cells.



Subunits of two example complex cells, that emerge in the simulation. Note the similarity of preferred orientation and spatial frequency, in contrast to the large shift (-90°) in phase between the two subunits of the same neuron.

Ref: Kayser, Einhäuser et al., ICANN 2001;  
 Körding, Kayser, et al., J Neurophysiol. 2004

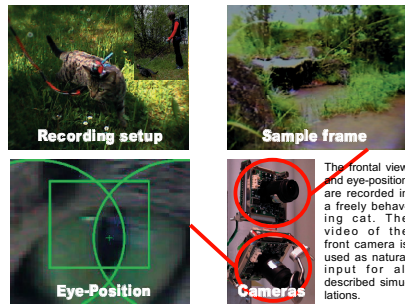
## 11 Beyond V1 - Texture cells



Repeated application of the stability objective in a two-layer hierarchical network leads to the emergence of texture-sensitive cells. Shown on the left are the best stimuli for ten different cells.

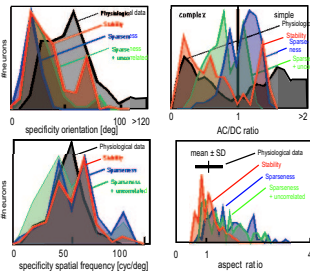
Ref: Franzius, Körding et al., submitted manuscript

## 2 Natural Stimuli



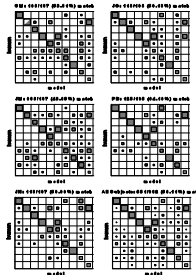
The frontal view and eye-position are recorded in a freely behaving cat. The video of the front camera is used as natural input for all described simulations.

## 7 Comparison to Physiology



The proposed stability objective matches most of the commonly used electrophysiological measures better than other popular objective functions.

## 12 Psychophysics

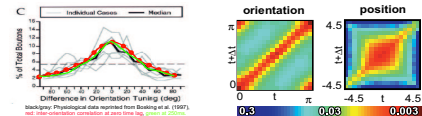


Having shown that properties of our simulated cells match physiological data, we extend the analysis to the system level. Comparing the classification achieved by texture cells to human classification shows a good match of 58% when classifying 200 textures into 10 classes. (Chance level at 10%; classification on pixel-wise-distance: 14%).

## 3 Statistics of Natural Scenes

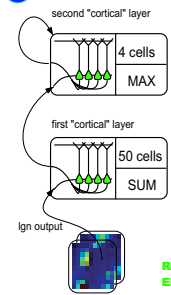
The spatiotemporal statistics of natural scenes matches several physiological properties of the primary visual cortex:

- (1) Cardinal orientations occur more frequent than oblique ones ('oblique effect').
  - (2) Collinear contours prevail over parallel shifted contours.
  - (3) Local position varies faster than local orientation.
- The latter result is the decisive property for the emergence of position-invariant, orientation-selective (i.e. complex) cells.



Ref: Kayser, Einhäuser et al., Neurocomput. 2003;  
 Betsch, Einhäuser et al. Biol. Cybern. 2004

## 8 Physiological implementation

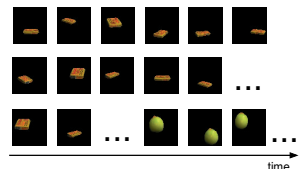


The stability objective can be implemented by physiological mechanisms. Our recently proposed model employs physiological results on neurons with two integration sites, separating the learning signal from a signal gating learning to implement timing dependent plasticity. This model is shown to lead to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

Ref: Körding and König, Neural Comput. 2001  
 Einhäuser, Kayser et al., Eur. J. Neurosci. 2002

## 13 Object Recognition - Stimuli

The experiments described so far have shown that a stability objective can explain physiological and psychophysical results. But can this principle be transferred to invariant object recognition in artificial systems? On top of a complex cell layer, we add another set of neurons trained with the stability objective ('object cells'). To test generalization performance, the network only sees a subset of views of a stimulus (9 in the example) for training, but is tested on all views.



Stimuli for object recognition are drawn from the Columbia object image library (COIL-100), which offers 72 views of 100 different objects.

## 4 Objective functions

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

$$\Psi_{stab} = - \sum_{i \text{ over all neurons}} \frac{\langle (A_i(t) - A_i(t-\tau))^2 \rangle_t}{\text{var}_t(A_i)}$$

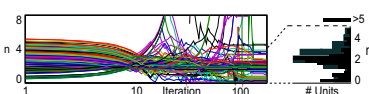
As stability is separated for each neuron, lateral coupling is introduced by adding a de-correlation term, which favors dissimilar receptive fields:

$$\Psi_{decorr} = - \sum_{i \text{ over all neurons}} \sum_{j < i} CC_{ij}^2$$

where CC denotes the correlation-coefficient

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

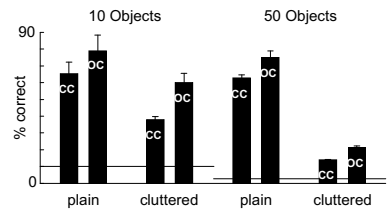
## 9 Learning the non-linearity



While in most simulations the cell model is fixed, this is not a necessary restriction. We show, that the power n of the transfer-function  $\phi(x) = x^n$  can be left free and optimized together with the receptive fields. On optimizing slowness, most cells converge to  $n=2$  as proposed for complex cells (black bars, time course of convergence to the left); but also a large fraction of cells exhibits very high or very low exponents. The distribution of n is similar to recent physiological results (gray bars, reprinted from Lau et al., 2002).

Ref: Kayser et al., Neural Comput. 2003

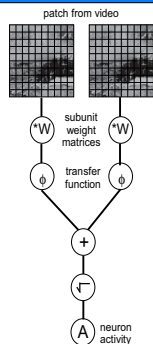
## 14 Object Recognition - Results



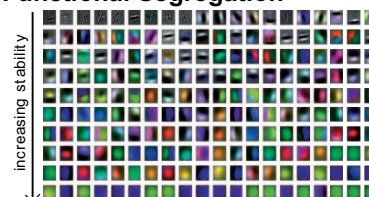
Classification performance, measured with an unsupervised clustering algorithm, is always higher for object cells (OC) trained by the stability objective than for their afferents, complex cells (CC). This shows that the stability objective improves rotation invariant object classification.

## 5 Cell Models

Each neuron consists of n subunits. The activity of a subunit at a specific time-point is given by the inner product of its weightmatrix W (receptive field) with the input passed through a non-linear transfer-function  $\phi$ . The activities of the subunit are then added to yield the activity of the neuron  $A_i$ . The data shown here are obtained with  $n=2$ ,  $\phi(x) = x^2$  (two subunit energy detector, panels 6&7) and  $n=1$ ,  $\phi(x) = x^2$  (panel 10). The power of the transfer-function can also be learnt (panel 9)



## 10 Functional Segregation



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.

Ref: Einhäuser, Kayser et al., Rev. Neurosci. 2003

## 15 Summary

We show here, that our implementation of the temporal coherence principle learns invariant visual representations on various levels:

- (1) Training a single layer network with grayscale natural videos leads to the emergence of complex cells.
- (2) The same architecture trained with colored natural videos learns to segregate cells, which are orientation invariant but color selective, from cells with complementary properties.
- (3) When using flexible non-linear neuron models, the distribution of learnt non-linearities closely matches physiology.
- (4) A two layer-network trained with natural videos achieves texture selectivity, while being invariant to local image features.
- (5) The same two-layer architecture learns to classify objects taken from a standard library, even from previously unseen viewpoints.

weinhau@ini.phys.ethz.ch; http://www.ini.unizh.ch/~weinhau