

# Towards unsupervised learning of invariance to a large class of transformations

**Author(s):**
Einhäuser, W.

**Publication Date:**
2001

**Permanent Link:**
https://doi.org/10.3929/ethz-a-004266558 →

# Towards unsupervised learning of invariance to a large class of transformations

W. Einhäuser, C. Kayser, K. P. Körding, P. König
Institute of Neuroinformatics (UNI/ETH Zürich)

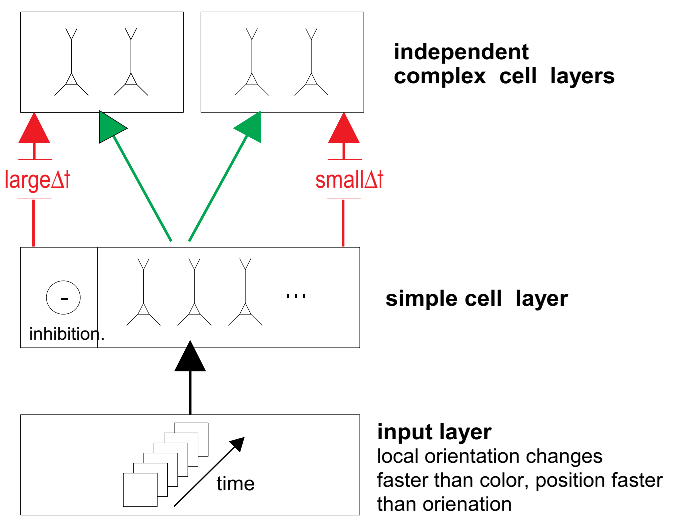
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

**Abstract**

Humans and animals can recognize objects nearly invariantly to various transformations, such as translations, scaling or changes in illumination or viewpoint. However, it is not a priori obvious, which features of an object are relevant to its recognition. This is the rationale to build systems which themselves learn to extract the relevant information from generic properties of their input. Here we present a model which exploits the fact, that different features of natural scenes change on different temporal scales. In a self-supervised scheme cells of the model integrate information over an extended period of time. By only maintaining information which is common to stimuli over this period, fast varying information is discarded, while slowly changing information is preserved, yielding invariance to fast variables. We show, that this model is capable of learning the translation invariance of complex cells from natural image sequences. Furthermore, invariance to other variables such as orientation and color can be obtained by just varying a single parameter, namely by increasing the time constant the system is operating on. Additionally, by decorrelating groups of cells with different time constants, cells can be obtained which are invariant to slower variables, while still specific to faster ones. Extending the proposed model into a hierarchical scheme might thus be a fruitful approach in pushing forward invariant object recognition.

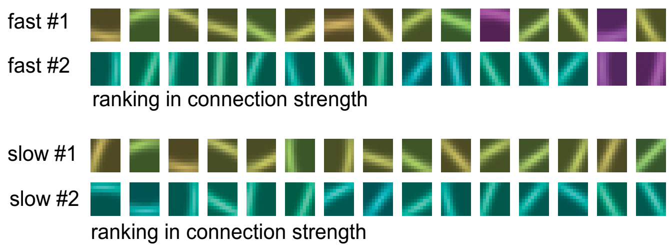
**The model**

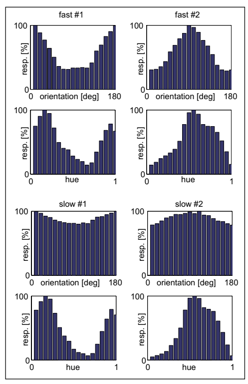
The network model implements three cell layers. The input layer is provided a continuous stream of image data taken either from natural videos (see adjacent poster) or from an artificially generated sequence. The "simple cell" layer is fully connected to the input layer and acquires its properties by competitive Hebbian learning. The "complex cell" layer is divided in sub-layers, each of which is characterized by a specific time constant $\Delta t$. Learning in each complex cell layer is dependent on the present activity in the simple cell and its activity a time $\Delta t$ ago. This enables the network to separate properties of the stimuli which are correlated over $\Delta t$ from those changing faster.

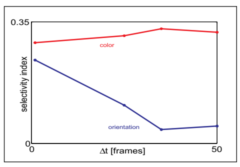



**Results**

For the following analysis colored bars at random positions are used to train the network. Orientation of subsequent stimuli changes with a probability of 10%, color with 1%, mimicking the different time courses of these variables in natural scenes. One observes that the time constant $\Delta t$ of a given "complex cell" population determines the degree of invariance versus specificity a cell acquires. The smaller $\Delta t$, the less invariance a population obtains. $\Delta t=0$ results in simple cells which are specific to all variables (position, orientation and color). "Fast" cells (finite but small $\Delta t$) acquire invariance to position while remaining specific to orientation and color. "Slow" cells (large $\Delta t$) remain only specific to color and obtain invariance to both position and orientation.




The figure above provides a qualitative impression of the results. Receptive fields of simple cells connecting strongest to a given complex cell are shown, sorted by their connection strength. One observes the specificity to color and orientation for the "fast" cells and the orientation invariant color detection of the "slow" cells. This behavior is quantified on the left, showing the orientation and color tuning of the respective cells. A specificity index obtained from these plots reveals the described dependence of invariance versus specificity on the network integration time constant $\Delta t$.

**Conclusion**

The results presented here in combination with the simulations performed on natural stimuli (see adjacent poster) show the development of a system capable of learning invariant object representations by exploiting generic properties of natural scenes. Employed in a hierarchical network structure and combined with top-down interactions this will set the stage for a system capable of learning relevant object representations while moving in a natural environment.

Acknowledgements
This work is financially supported by Honda R&D Europe (WE), the "Neuroscience Center Zurich" (CK), the "Swiss National Fonds" (PK - grant no: 3100-51059.97) and the "Boehringer Ingelheim Fond" (KPK).