Learning Multiple Feature Representations from Natural Image Sequences

Wolfgang Einhäuser, Christoph Kayser, Konrad P. Körding, and Peter König

Institute of Neuroinformatics, University / ETH Zürich
Winterthurerstr. 190, 8057 Zürich, Switzerland
{weinhaeu,kayser,koerding,peterk}@ini.phys.ethz.ch

Abstract. Hierarchical neural networks require the parallel extraction of multiple features. This raises the question how a subpopulation of cells can become specific to one feature and invariant to another, while a different subpopulation becomes invariant to the first but specific to the second feature. Using a colour image sequence recorded by a camera mounted to a cat's head, we train a population of neurons to achieve optimally stable responses. We find that colour sensitive cells emerge. Adding the additional objective of decorrelating the neurons' outputs leads a subpopulation to develop achromatic receptive fields. The colour sensitive cells tend to be non-oriented, while the achromatic cells are orientation-tuned, in accordance with physiological findings. The proposed objective thus successfully separates cells which are specific for orientation and invariant to colour from orientation invariant colour cells.

Keywords. Learning, visual cortex, natural stimuli, temporal coherence, colour UTN: I0109

1 Introduction

In recent years there has been increasing interest in the question on how the properties of the early visual system are linked to the statistics of its natural input. Regarding primary visual cortex (V1), the spatial properties of simple [[1]] as well as complex [[2]] cells could be explained on the basis of sparse coding. A different coding principle based on the trace rule originally proposed by Földiak [[3]], namely temporal coherence or stability, has also been shown to lead to the emergence of simple [[4]] and complex [[5]] type receptive fields when applied to natural image sequences. These studies use greyscale images and thus address spatial properties only. However, a considerable fraction of primate V1 neurons is sensitive to colour and their spatial properties deviate from those of achromatic cells [[6,7]]. We here address whether temporal coherence leads to the emergence of colour-sensitive cells, how their spatial receptive fields are related to their chromatic properties, and how their fraction relative to the whole population is determined.

J.R. Dorronsoro (Ed.): ICANN 2002, LNCS 2415, pp. 21–26, 2002.
© Springer-Verlag Berlin Heidelberg 2002

2 Methods

Natural Stimuli. Sequences of natural stimuli are recorded using a removable lightweight CCD camera (Conrad electronics, Hirschau, Germany) mounted to the head of a freely behaving cat, while the animal is taken for walks in various local environments. The output of the camera is recorded via a cable attached to the leash onto a VCR carried by the experimenter and digitized offline at 25 Hz and 320x240 pixels. A colour-depth of 24-bit is used and encoded in standard RGB format. In each of the 4900 image frames thus obtained, at 20 randomly chosen locations a 30x30 pixel wide patch is extracted. This patch together with the patch from the same location of the following frame constitutes a stimulus pair. Each colour channel of each patch is smoothed with a Gaussian kernel of width 12 pixel to reduce boundary effects. For computational efficiency a principal component analysis (PCA) is performed on the stimuli to reduce the input dimensionality of 2700 (30x30 pixels times 3 colours). Unless otherwise stated, the first 200 principal components are used for further analysis, which explain 97 % of the variance. In order to process the stimuli independently of the global illumination level, the mean intensity is discarded by excluding the first principal component.

Objective function. We analyse a population of N=200 neurons. The activity of each neuron is computed as

$$A_i(t) = \left| \sum_j W_{ij} * I_j(t) \right| \tag{1}$$

where I is the input vector and W the weight matrix. We define the temporal coherence of each neuron as

$$\psi_i^{stable} := -\frac{\left\langle \left(\frac{\mathrm{d}}{\mathrm{dt}} A_i(t)\right)^2 \right\rangle_t}{var_t\left(A_i\left(t\right)\right)} \tag{2}$$

where $\langle\rangle$ denotes the mean and var_t the variance over time. The temporal derivative is calculated as the difference in activity for a pair of stimuli from consecutive frames. We refer to this objective as 'stability', since it favours neurons, whose responses vary little over time. The total stability is defined as sum over the individual stabilities:

$$\Psi^{stable} := \sum_i \psi_i^{stable} \tag{3}$$

Furthermore, we define the decorrelation objective as

$$\Psi^{decorr} = -\frac{1}{(N-1)^2} \left\langle \sum_i \sum_{j \neq i} \left(\sigma_{ij}^2(t)\right) \right\rangle_t \tag{4}$$

where $\sigma_{ij}(t)$ denotes the correlation coefficient between $A_i(t)$ and $A_j(t)$. Combining these objectives we define

$$\Psi^{total} = \Psi^{stable} + \beta \Psi^{decorr} \tag{5}$$

where β is a constant.

The network is trained from random initial conditions by gradient ascent: For each iteration Ψ^{total} is computed over the complete natural stimulus set and the weightmatrix W is changed in the direction of the analytically given gradient, $\frac{d\Psi^{total}}{dW}$ to maximise Ψ^{total}. All presented results are taken after 60 iterations, when the network has converged under all analysed conditions. Simulations are performed using MATLAB (Mathworks, Natick, MA).

Analysis of neuron properties. By inverting the PCA the receptive field representation in input space is obtained from the rows of W. For further analysis we convert this colour channel (RGB) representation into a representation separating the hue, the saturation and the value (brightness) of each pixel (HSV representation) by the standard MATLAB function rgb2hsv using default mapping. To illustrate the colour properties of a receptive field abstracted from its topography, we plot the projection of each of its pixels onto the isoluminant plane. We quantify the colour content of a receptive field by the mean saturation over all pixels and define a cell to be colour sensitive if its mean saturation falls above 0.2. The isotropy of a receptive field is assessed using a standard method [[8]]: The tensor of inertia is computed on the pixel values and anisotropy is defined as the ratio of the difference between the tensor's long and short principal axis divided by the sum of these axes. This measure is 0 for an isotropic structure and approaches 1 for a perfectly oriented structure.

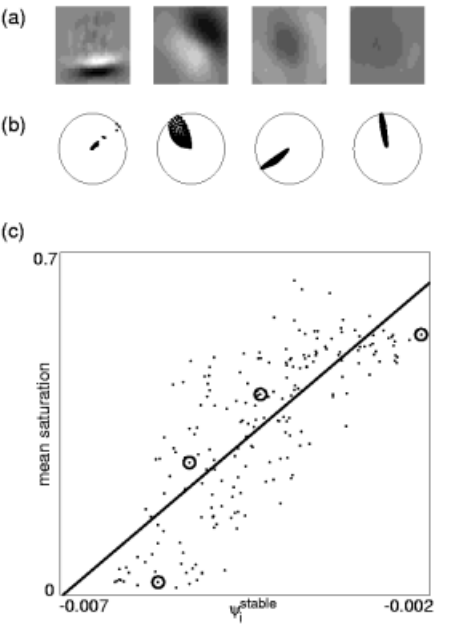

Fig. 1. (a) Four examples of receptive fields of a simulation using $\beta = 5$ after convergence (60 iterations) sorted (from left to right) by increasing ψ_i^{stable}.
(b) All pixels of each receptive field from (a) projected onto the isoluminant plane. Points close to the center indicate low saturation, i.e. little colour content, points to the periphery indicate saturated (coloured) pixels. The scaling is identical for all plots.
(c) Dependence of mean saturation on ψ_i^{stable}. Encircled points represent examples from (a) and (b).

3 Results

We train the network to optimize Ψ^{total} on natural stimuli. After convergence one observes an about equal fraction of chromatic and achromatic cells. The achromatic cells have lower indi-

vidual values of ψ_i^{stable} than the colour sensitive cells, indicating that the stability objective favours chromatic cells (Figure 1).

The relative contribution of the decorrelation and the stability objective can be regulated by changing the parameter β. In order to investigate the influence of the stability objective alone, we set $\beta = 0$. In this case nearly all cells become colour-sensitive (Figure 2 left). This shows that colour-sensitive cells have more stable responses to natural stimuli than achromatic cells.

Strengthening of the decorrelation objective – by increasing β – on the other hand forces the cells to acquire more dissimilar receptive fields. In this case some cells have to acquire receptive fields which are suboptimal with respect to the stability objective, yielding an increasing fraction of achromatic cells (Figure 2). Thereby the parameter β controls the relative fraction of chromatic versus achromatic cells.

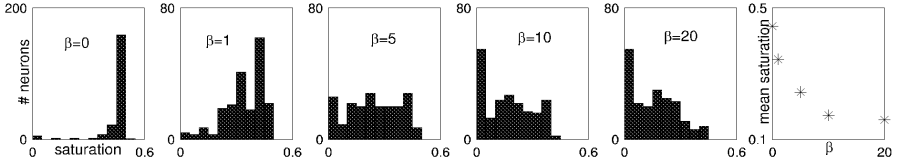

Fig. 2. Histograms of mean saturation for different values of β. Rightmost panel: mean of each histogram vs. β

Robustness with respect to the input dimensionality is a desirable property, especially when embedding the proposed system into a hierarchical model. We thus determine the fraction of colour selective cells in dependence on the PCA dimension used. We find the results to be over a wide range independent of the input dimension (Figure 3). This is in contrast to studies using independent component analysis (ICA), where the fraction of colour sensitive cells strongly depends on the PCA dimension [[9]].

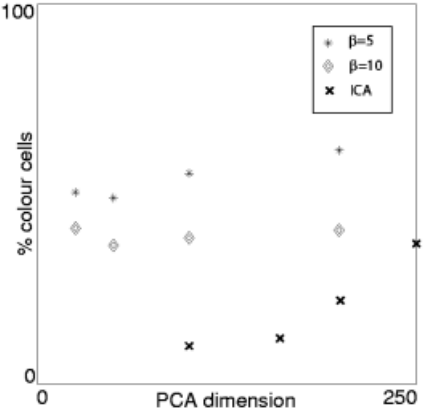

Fig. 3. Percentage of colour cells (mean saturation > 0.2) vs. PCA dimension for two values of β compared to ICA results of reference [[9]].

Unlike achromatic cells, most colour sensitive neurons in primate V1 are non-oriented, i.e. they do not exhibit a pronounced orientation preference [[6, 7]]. We thus analyse the dependence of the model cells' anisotropy (see methods) to their mean saturation. We indeed find a strong tendency for chromatic cells to be non-oriented. On the other

hand most achromatic cells show the orientation tuning typical of V1 simple cells (Figure 4), in compliance with the application of the stability objective on greyscale images.

Concluding, unoriented chromatic cells exhibit optimally stable responses on natural stimuli. When cells are forced by decorrelating to achieve suboptimal stability, a second subpopulation emerges, oriented achromatic cells. This implies that colour is the most stable feature in natural scenes, but that on the other hand achromatic edges are more stable than chromatic edges.

4 Discussion

A common problem in hierarchical networks is the separation of different variables. Here we obtain two distinct populations, one specific for colour and invariant to orientation and the other vice versa. The individual stability ψ_i^{stable} of each cell provides a system inherent measure to distinguish between both populations. Furthermore, the relative size of both populations is regulated by a single parameter. These properties make the proposed stability objective promising for the use at different stages of hierarchical systems.

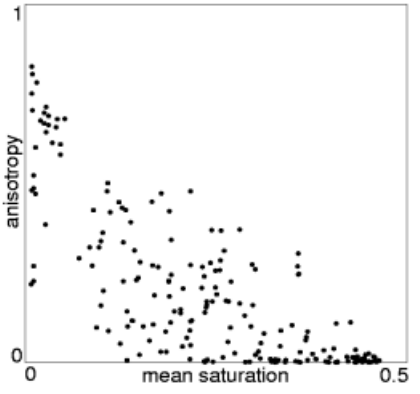

Fig. 4. Value anisotropy vs. mean saturation for a simulation of $\beta = 5$.

We showed in a previous study how the stability objective can be implemented in a physiologically realistic framework [[10]]. A possible implementation of the decorrelation objective in neural circuits is mediated by strong lateral inhibition. Due to response latencies this mechanism would have larger impact with increasing distance from the input layer. The fact that stronger decorrelation leads to less chromatic cells thus is in compliance with the physiological finding that in primate V1 most chromatic cells are found in layer 4.

A number of studies recently addressed the application of ICA on standard colour natural stimuli [[9]] and on hyperspectral images [[11,12]]. These studies find colour-sensitive cells, similar to the ones described here. However, none of the studies quantifies the relation of the cells' spatial to their chromatic properties, leaving an important issue for the comparison to physiology unaddressed. Here we find a strong correlation between spatial and chromatic properties comparable to physiological findings. Another remarkable difference between stability and ICA is the dependence of the latter on the input dimension. This implies that the relative size of the emerging chromatic and achromatic subpopulations might

be influenced by changes in feedforward connectivity. In the case of stability on the other hand, the subpopulations' size is regulated by the relative strength of decorrelation, which might be mediated by lateral connections.

Acknowledgments. This work was financially supported by Honda R&D Europe (Germany), the Center for Neuroscience Zurich (ZNZ), the Collegium Helveticum and the Swiss national fund (SNF – grantNo: 31-65415.01). All procedures are in accordance with national and institutional guidelines for animal care.

References

1. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (1996) 607–609
2. Hyvärinen, A., Hoyer, P.O.: Emergence of phase and shift invaraint features by decomposition of natural images into independent feature subspaces. Neural Comput. 12 (2000) 1705–1720
3. Földiak, P.: Learning Invariance from Transformation Sequences. Neural Computation 3 (1991) 194–200
4. Hurri, J., Hyvärinen A. Simple-Cell-Like Receptive Fields Maximize Temporal Coherence in Natural Video. submitted (2002)
5. Kayser, C., Einhäuser, W., Dümmer O., König P., Körding K.P.: Extracting slow subspaces from natural videos leads to complex cells. In G. Dorffner, H. Bischoff, K. Hornik (eds.) Artificial Neural Networks – (ICANN) LNCS 2130, Springer-Verlag, Berlin Heidelberg New York (2001) 1075–1080
6. Gouras, P.: Opponent-colour cells in different layers of foveal striate cortex. J. Physiol 199 (1974) 533–547
7. Lennie, P., Krauskopf, J., Sclar, G.: Chromatic Mechanisms in Striate Cortex of Macaque. J. Neurosci. 10 (1990) 649–669
8. Jähne, B.: Digital Image Processing - Concepts, Algortihms and Scientific Applications, 4th compl. rev. edn. Springer-Verlag, Berlin Heidelberg New York (1997)
9. Hoyer, P.O., Hyvärinen, A.: Independent Component Analysis Applied to Feature Extraction From Colour and Stereo Images. Network 11 (2000) 191–210
10. Einhäuser, W., Kayser, C., König, P., Körding, K.P.: Learning of complex cells properties from their responses to natural stimuli. Eur. J. Neurosci. 15 (2002) in press.
11. Lee, T.W., Wachtler, T., Sejnowski, T.: Color Opponency Constitutes A Sparse Representation For the Chromatic Structure of Natural Scenes. NIPS. 13 (2001) 866–872.
12. Wachtler, T., Lee, T.W., Sejnowski, T.: Chromatic Structure of natural scenes. J. Opt. Sco. Am. A 18 (2001) 65–77