ELSEVIER

Neuroscience Letters 228 (1997) 155–158

NEUROSCIENCE LETTERS

Neural network models and the visual cortex: the missing link between orientation selectivity and the natural environment

Christoph Zetzsche[a,*], Erhardt Barth[a], Gerhard Krieger[a], Bernhard Wegmann[b,1]

[a]*Institut für Medizinische Psychologie, Ludwig-Maximilians-Universität München, Goethestrasse 31, D-80336 München, Germany*
[b]*Lehrstuhl für Nachrichtentechnik, Technische Universität München, Arcisstrasse 21, D-80290 München, Germany*

Received 21 March 1997; revised version received 19 May 1997; accepted 19 May 1997

Abstract

Orientation selectivity is a basic property of neurones in the visual cortex of higher vertebrates. Such neurones can be seen to act as 'feature detectors', which provide an efficient cortical representation of the outside world. More recently, the removal of correlations between the signals of cortical neurones has been suggested as suitable theoretical concept for explaining the development of receptive fields. Corresponding neural network simulations yielded oriented 'receptive field' structures resembling those observed by neurophysiologists. The findings suggest that the 'decorrelation approach' can provide a causal relationship between characteristics of the physical world and brain function. However, we were able to reveal a basic deficit of the decorrelation approach which we illustrate by the construction of two artificial 'worlds', a 'Gaussian' one and an 'orientation-only' one. We show that, according to the decorrelation approach, oriented environmental features would be neither necessary nor sufficient for the development of oriented receptive fields. Thus the link between environmental structure and cortical orientation selectivity still awaits a theoretical explanation. © 1997 Elsevier Science Ireland Ltd.

Keywords: Receptive fields; Orientation selectivity; Natural environment; Neural networks; Image statistics; Visual cortex; Adaptation

The efficient neural representation of the outside world is an important aspect of information processing in any organism. Since the discovery of cells in the mammalian visual cortex being selective to stimulus properties like colour, symmetry, or orientation [9], it has been believed that this selectivity is directly related to environmental properties. In particular, it has been shown that these cells can be described as 'feature detectors' for the oriented local edge and line segments, into which the shapes of objects can be decomposed [2]. The alternative concept, the notion of 'spatial frequency analysis' has been shown to be basically compatible with the feature detector hypothesis and currently, cortical simple cells are seen as linear filter mechanisms with a high degree of selectivity with respect to oriented features (e.g. [13]). Similar orientation selective filters are employed for the compression of digital images [22] and in the closely related 'wavelet' approach, a branch of applied mathematics suitable to such diverse domains as fractals and oceanography [14]. Understanding the relation between the structure of images and orientation selective processing is, therefore, a key feature not only for vision research but also for multi-dimensional signal processing and modern telecommunication [24].

In recent years, a promising source for new concepts regarding cortical orientation selectivity has been the research on neural network models. This led to increasing evidence for a general neural decorrelation strategy that is directed towards the removal of redundant correlations in sensory input data [1–3,10,16,19,20]. And indeed, decorrelation approaches have been successfully applied to the modelling of important aspects of the sensory adaptation of organisms to their environment. In particular, it has been found that oriented receptive field structures can emerge as the result of a neural strategy which aims at the removal of second-order correlations between the input units (e.g. [8,12,15,19,20]).

The link between neural orientation selectivity and environmental structure seems thus to have received its theore-

* Corresponding author. Tel.: +49 89 5996638; fax: +49 89 5996615; e-mail: chris@imp.med.uni-muenchen.de
[1] Now with Siemens AG, München, Germany.

0304-3940/97/$17.00
PII S0304-3940(97)00386-8

tical explanation: the brain has adapted to the environment by creating cortical structures which are able to decorrelate the statistical second-order dependencies as established by the environment. This interpretation, however, rests on the implicit hypothesis that the occurrence of certain parallel relations in empirical observations and in computer simulations is already sufficient for an understanding of the true mechanisms. The critical question to be asked is, therefore, whether a logical connection can be demonstrated between the frequent occurrence of oriented features in natural images and the specific statistical properties exploited in decorrelation models.

We put the decorrelation hypothesis to test by constructing two possible worlds with opposite statistical properties (Fig. 1). One is a perfectly decorrelated 'orientation-only world', i.e. a world, which consists entirely of oriented features but lacks any of the second-order correlations claimed to give rise to oriented receptive field structures. The other

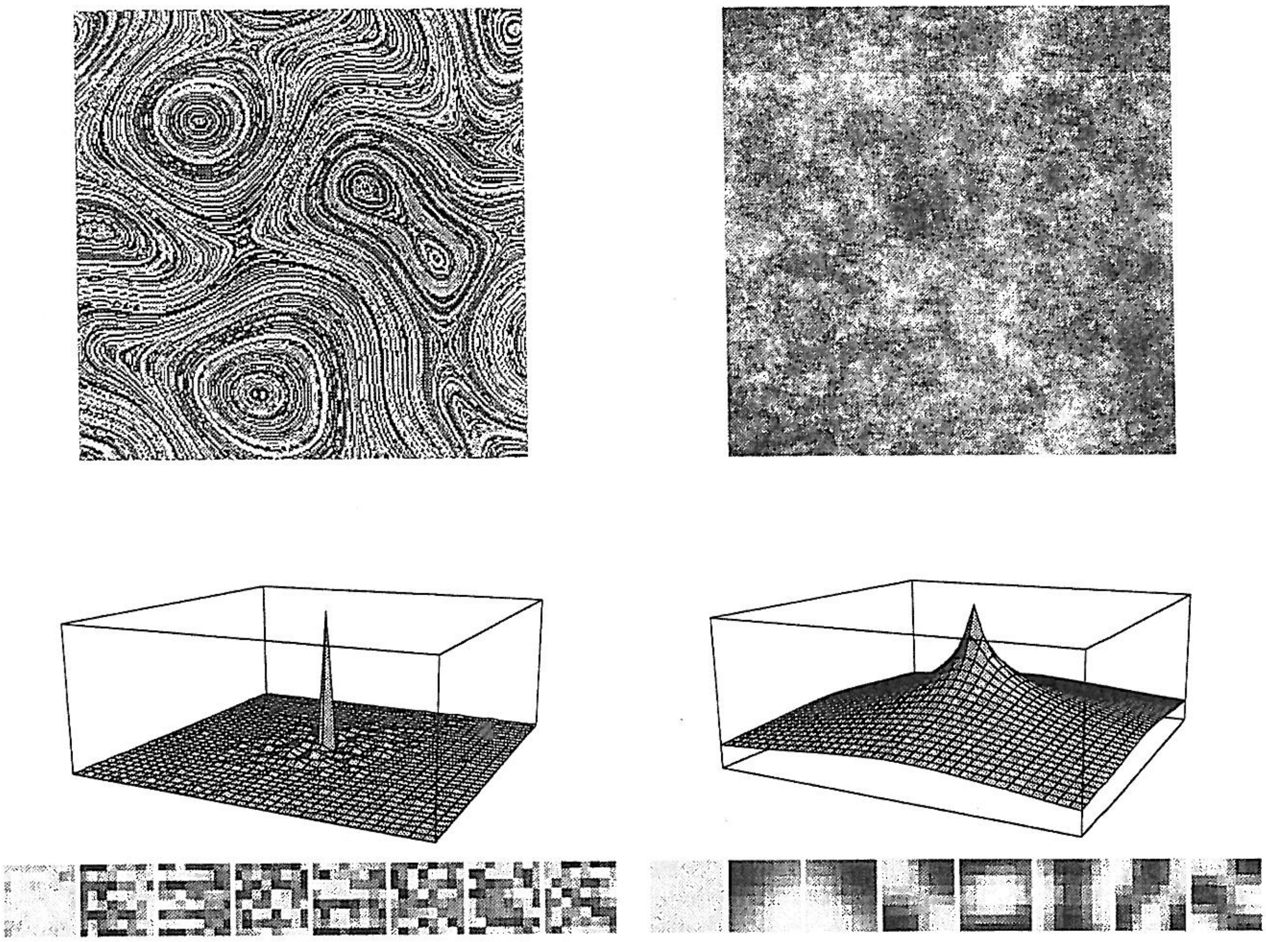

Fig. 1. Pictures from two hypothetical worlds (upper row), the corresponding autocorrelation functions (middle row) and the predicted receptive fields as resulting from the first eight principal components (lower row). According to decorrelation-based neural-network approaches these worlds should lead to the emergence of vision systems with substantially different receptive field structures. The 'orientation-only world' (left side), though consisting entirely of oriented features, is a world without any second-order correlations, as indicated by the isolated peak in the centre of the autocorrelation function (middle left). From the perspective of decorrelation approaches it appears completely random. Hence there should be neither an internal need nor an environmental pressure towards the development of receptive fields with an orientation-selective organization. This is confirmed by the structure of the predicted receptive fields which show a basically random organization (lower left). (It is possible to construct models that will produce oriented fields even in this case, but this result will then be independent of the type of environment, and is simply enforced by 'built-in' rules). The 'Gaussian world' (right side) is a world with the same correlations as found in our terrestrial environment (e.g. [6]) but without any of the oriented features, like edges or lines. While the decorrelation approaches could see no need for oriented receptive fields in the 'orientation-only world', they would predict that, despite the complete lack of oriented features in the 'Gaussian world', this world will cause a substantial pressure towards the development of oriented receptive fields (lower right). A further prediction is that the same set of receptive fields being optimally adapted to the requirements of the Gaussian world, will also be optimally suited for our terrestrial environment. In our simulations the 'orientation-only world' corresponds to a random field which is perfectly stationary and ergodic. It has been constructed by low-pass filtering of a white noise source, followed by a point-transform with a lookup-table containing random coefficients. The strong orientation structure has to be ascribed to dependencies higher than second-order (details can be found in [24]). The 'Gaussian world' consists of coloured Gaussian noise with a power spectrum falling off with $1/f^2$. This type of image is currently regarded as the statistical standard model for natural scenes, since its second-order correlations are equivalent to those of a wide variety of images from the natural environment (e.g. [6]). The receptive fields are computed from the principal components (eigenvectors) of the respective covariance matrices. These are the typical solutions towards which the receptive fields in decorrelation based models will finally converge (e.g. [16,19,20]).

one is a 'Gaussian world' which has the same second-order correlation as our environment [6] but lacks oriented features at all.

From the perspective of decorrelation-based neural network modelling, we are to expect, that organisms living on the nebulous Gaussian world (Fig. 1, right) would develop oriented receptive fields like those of the higher vertebrates existing in our terrestrial environment. By contrast, living in the 'orientation-only' world (Fig. 1, left), would not require oriented receptive fields (or even non-oriented lateral inhibition), simply because the sensory input would be already completely decorrelated. This demonstration reveals that, from the decorrelation perspective, the presence of oriented features is neither sufficient (Fig. 1, left) nor necessary (Fig. 1, right) for the formation of oriented receptive fields. From this we are to conclude, that decorrelation based models do not establish a causal relationship between the frequent occurrence of oriented features in natural images and the development of orientation selective neurones.

As to the explanatory power of the approach regarding receptive field structure, there are further critical arguments. The basic signal processing principle behind the decorrelation approaches is the removal of second-order correlations, either in form of a 'whitening' filter or as principal component analysis, or Karhunen–Loeve transform, respectively. In technical image processing, it is well known that the lack of image phase information in the autocorrelation function and in the corresponding spectral density renders them insufficient for the statistical modelling of natural images [18]. The relation between the statistical properties of images, suitable technical data compression strategies, and biological vision systems has recently received a detailed investigation [24]. This analysis revealed fundamental shortcomings of the second-order approach, in particular with respect to oriented image structures. A more recent critique of decorrelation based neural network approaches has suggested that the omission of phase information causes non-unique solutions and an inability to produce localized receptive fields [7].

To have demonstrated that decorrelation approaches cannot provide a causal explanation for the joint occurrence of a phenomenon like 'orientation' in the visual environment and in the mechanisms used for its representation, does not exclude the theoretical possibility of an 'accidental parallelism', i.e. of a coincidence without any causal relationship. However, such a hypothesis is in conflict with a number of investigations which revealed that orientation selective cortical processing does indeed yield information-theoretical advantages [22], especially a 'sparse' code [6,7,23,24,26] and an efficient use of associative memory [23]. We have repeatedly stressed that the 'sparsification' provided by oriented receptive fields is clearly superior to that of the circular fields of retinal ganglion cells [23,24,26], a difference that should not exist from a decorrelation perspective, since circular fields can already provide an almost perfect decorrelation [1,24]. Circular and oriented receptive fields being insofar more or less equivalent solutions, decorrelation based neural network models can be made to produce either sort. This would imply, however, that there are not specific environmental features which determine receptive field structure, but that some internal network rules cause a suitable 'symmetry-breaking'.

In conclusion, we are left with an alternative. Either, we accept the decorrelation hypothesis as a sufficient explanation for the observed orientation selectivity of cortical neurones. This implies to interpret the joint occurrence of orientation in the features of the visual environment and in the structure of receptive fields as an 'accidental parallelism', i.e. to give up the inspiring original thesis of Barlow [2], that the reason for the emergence of the different types of receptive fields has to be searched in their efficient adaptation to specific features of the environment. Or we concede that second-order statistics cannot be sufficient for capturing elementary aspects of our world [3] and that the research into receptive field structure should take into account statistical dependencies higher than second-order [7,24]. Some techniques capable of exploiting higher-order dependencies have recently been developed (e.g. [3–5,21]). In particular, a 'sparsification strategy' may give rise to quite realistic receptive fields [17]. Furthermore, exploitation of higher-order dependencies may also provide an adaptational explanation for the massive non-linearities found for other types of visual neurones like complex and end-stopped cells [11,24]. It can hence be envisaged that further research will finally provide us with an unambiguous description of the missing link between environmental structure and the orientation selectivity of simple cells. However, at present there exists neither a thorough differential analysis of the second-order vs. higher-order contributions in the respective network models, nor a statistical source model of the higher-order properties of natural images (see [24,25] for some suggestions). Once this has been achieved, we can look forward towards a conclusive explanation of the development of orientation-selective receptive fields in terms of higher-order statistics of the natural visual environment.

We thank I. Rentschler for helpful comments. Research was supported by a grant of the Deutsche Forschungsgeneinschaft to I. Rentschler and C.Z.

[1] Atick, J.J. and Redlich, A.N., Towards a theory of early visual processing, Neural Comput., 2 (1990) 308–320.
[2] Barlow, H.B., Single units and sensation: a neuron doctrine for perceptual psychology, Perception, 1 (1972) 371–394.
[3] Barlow, H.B., Unsupervised learning, Neural Comput., 1 (1989) 295–311.
[4] Bell, A. and Sejnowski, T., An information-maximization approach to blind separation and blind deconvolution, Neural Comput., 7 (1995) 1129–1159.
[5] Deco, G. and Brauer, W., Nonlinear higher order statistical decorellation by volume conserving neural architectures, Neural Networks, 8 (1995) 525–535.
[6] Field, D.J., Relations between the statistics of natural images and the

response properties of cortical cells, J. Opt. Soc. Am. A, A4 (1987) 2379–2394.
[7] Field, D.J., What is the goal of sensory coding?, Neural Comput., 6 (1994) 559–601.
[8] Földiak, B., Adaptive network for optimal linear feature extraction. In Proc. IEEE/INNS Int. Conf. Neural Networks, Vol. 1, IEEE Press, New York, 1989, pp. 401–405.
[9] Hubel, D.H. and Wiesel, T.N., Receptive fields of single neurones in the cat's striate cortex, J. Physiol., 148 (1959) 574–591.
[10] Kohonen, T., Self-Organization and Associative Memory, 2nd edn., Springer, Berlin, 1988.
[11] Krieger, G. and Zetzsche, C., Nonlinear image operators for the evaluation of local intrinsic dimensionality, IEEE Trans. Image Process. (Special Issue on Nonlinear Image Processing), 5 (1996) 1026–1042.
[12] Linsker, R., From basic network principles to neural architecture: emergence of orientation selective cells, Proc. Natl. Acad. Sci. USA, 83 (1986) 8390–8394.
[13] MacKay, D.M., Strife over visual cortical function, Nature, 289 (1981) 117–118.
[14] Mallat, S.G., A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell., 11 (1989) 674–693.
[15] Malsburg, von der C., Self-organization of orientation sensitive cells in striate cortex, Kybernetik, 14 (1973) 85–100.
[16] Oja, E., A simplified neuron as a principal component analyzer, J. Math. Biol., 15 (1982) 267–273.
[17] Olshausen, B. and Field, D., Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images, Nature, 381 (1996) 607–609.
[18] Oppenheim, A.V. and Lim, J.S., The importance of phase in signals, Proc. IEEE, 69 (1981) 529–541.
[19] Rubner, J. and Schulten, K., Development of feature detectors by self-organization, Biol. Cybern., 62 (1990) 193–199.
[20] Sanger, T.D., Optimal unsupervised learning in a single-layer feed-forward neural network, Neural Networks, 2 (1989) 459–473.
[21] Schmidhuber, J., Eldracher, M. and Foltin, B., Semilinear predictability minimization produces well-known feature detectors, Neural Comput., 8 (1996) 773–786.
[22] Watson, A.B., Efficiency of a model human image code, J. Opt. Soc. Am. A, A4 (1987) 2401–2417.
[23] Zetzsche, C., Sparse coding: the link between low level vision and associative memory. In R. Eckmiller, G. Hartmann and G. Hauske (Eds.), Parallel Processing in Neural Systems and Computers, Elsevier, Amsterdam, 1990, pp. 273–276.
[24] Zetzsche, C., Barth, E. and Wegmann, B., The importance of intrinsically two-dimensional image features in biological vision and picture coding. In A. Watson (Ed.), Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993, pp. 109–138.
[25] Zetzsche, C. and Krieger, G., Natural image statistics and the exploitation of local intrinsic dimensionality, Vis. Res., (1997) submitted.
[26] Zetzsche, C. and Schönecker, W., Orientation selective filters lead to entropy reduction in the processing of natural images, Perception, 16 (1987) 229.