ELSEVIER

Neurocomputing 44–46 (2002) 549–553

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Formation of pinwheels of preferred orientation by learning sparse neural representations of natural images

A. Moukovski[a,*], D.M. Gorinevski[a,b], M.A. Giese[a,c,d], W. von Seelen[a]

[a] *Institut fur Neuroinformatik, RuhrUniversität Bochum, Universitatstr. 150, Gebaude ND/04, D44780 Bochum, Germany*

[b] *Financial cybernetics GmbH, Bochumerstr 123, 45661 Recklinghausen, Germany*

[c] *Center for Biological and Computational Learning, Massachusetts Institute of Technology, 45, Carlton Street, Cambridge, MA 02139-4307, USA*

[d] *Honda R&D, Boston, USA*

Abstract

We devise self-organizing model of the striate cortex that learns orientation maps by sparse coding of natural images. The model assumes the existence of oriented receptive fields and of retinotopic mapping. We demonstrate that learning sparse representations of natural images leads to the formation of spatially periodic orientation maps. If and only if the sparseness of the representation is sufficiently high, these orientation maps reproduce different critical parameters of experimentally measured maps in the striate cortex. We conclude the functional topology of the visual cortex that may be tailored to optimize the encoding of natural stimuli with minimal redundancy of the underlying representation. © 2002 Published by Elsevier Science B.V.

Keywords: Sparse coding; Cortex orientation maps

1. Introduction

Orientation maps in the striate cortex have been studied extensively in the experimental literature using optical recording techniques (e.g. [2,3]). The preferred

* Corresponding author.

E-mail addresses: mukovski@neuroinformatik.ruhr-uni-bochum.de (A. Moukovski), giese@mit.edu (M.A. Giese), dg@financial-cybernetics.de (D.M. Gorinevski), vs@neuroinformatik.ruhr-uni-bochum.de (W. von Seelen).

0925-2312/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0925-2312(02)00421-6

orientations of the cortical columns vary smoothly along the cortical surface in a regular hypercolumnar structure corresponding to spatial bandpass characteristics.

A number of self-organization models were developed to explain the emergence of these highly structured patterns. Most of them realize correlation-based learning with feedback via lateral connection [6,7,10]. These models do not clarify the role of these ordered structures for information processing.

Recent theoretical works provided support for the hypothesis of a relationship between redundancy reduction in cortical information encoding [1,4] and the structure of the visual cortex by demonstrating that learning algorithms enforcing sparse neural coding of images can generate localized receptive fields with similar properties of the cells in the primary visual cortex of mammals [11]. As shown recently in [5], the low redundancy coding in the presence of weak probabilistic couplings in the topological neighborhood of neural units can also account for the topological ordering of learning oriented receptive fields.

2. Model

We devise self-organizing model which is not based on explicit correlations between the neural inputs. The model accounts for the emergence of ordered orientation maps from random initial states by low redundancy encoding of natural images through sparse neural representations. The model assumes the existence of retinotopy and oriented receptive fields, but not the existence of prewired structured lateral connections. Structured interactions must, therefore, arise solely from the statistical properties of the image data and the information theoretical strategy for its encoding.

The modeled neural network represents the stimuli $g(x, y)$ with small approximation errors and with a small number of active neurons (their activities denoted by c_i). The receptive field $R_i(x, y)$ with the form of a Gabor function has size $0.6^\circ \times 0.6^\circ$, while distance between the centers of receptive fields of neighboring neurons is 0.2° of the visual field.

It has been shown (e.g. [11]) that the approximations with sparse activation distribution (c^*) can be obtained by minimizing a cost functional:

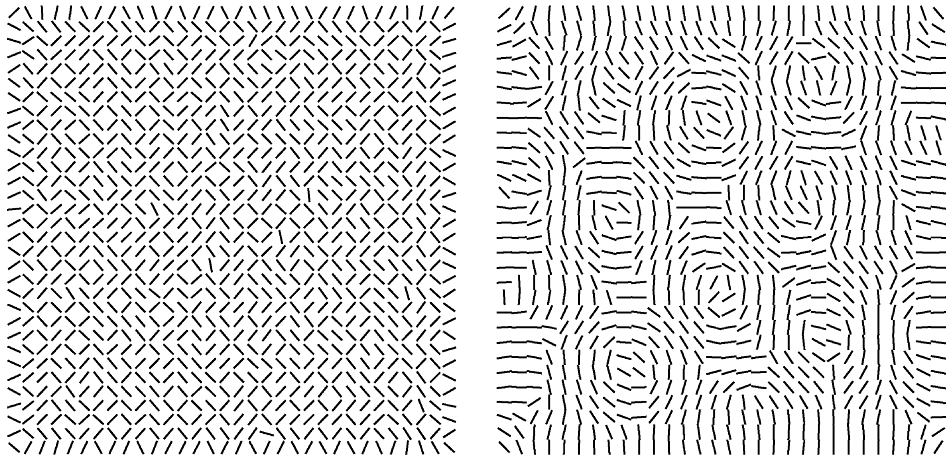
$$E(g) = \int \left[g(x, y) - \sum R_i(x, y, \phi_i)c_i\right]^2 \mathrm{d}x\,\mathrm{d}y + \mu \sum |c_i| + \lambda \sum c_i^2. \tag{1}$$

During the learning process, a sequence of natural images is presented. The learning process aims at finding an orientation map $\phi = (\phi_1, \phi_2, \ldots, \phi_N)$, that minimizes the cost functional given by Eq. (1) over the time series of training images:

$$\phi = \arg\min_{\phi} \langle E[g, c^*, \phi] \rangle, \tag{2}$$

where $\langle . \rangle$ denotes averaging over the image set series.

In order to find the optimal orientation map, we adjusted the orientations sequentially using stochastic gradient descent procedure.



(a) (b)

Fig. 1. Two trained orientation maps. The map (a) was learned with small sparseness enforcement term in the error functional ($\mu = 0.8$). The map (b) emerges for $\mu = 1.6$.

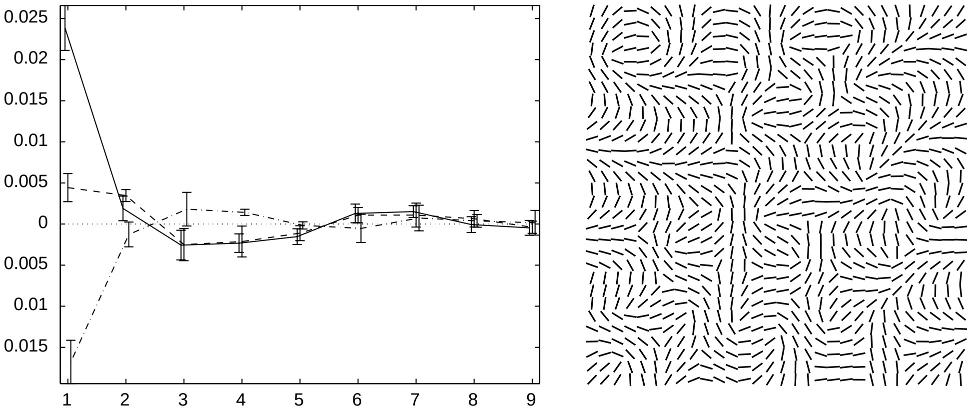
3. Results

We demonstrate that, by enforcement of sparse coding, the preferred orientations of the model neurons self-organize into spatially periodic orientation maps. Only when the sparseness of the representation is sufficiently high, the structure of the maps resembles experimentally measured cortical orientation maps in the primary visual cortex of mammals (cf. Fig. 1).

1. In the case of high sparseness, learned maps have isoorientation domains and orientation singularities with vorticities of $\pm 180^\circ$.
2. Spatial radial autocorrelation function of orientation map is an oscillating function that decays exponentially (cf. [8]).
3. Two-dimensional power spectrum of orientation map is symmetrical and has a band-pass structure (cf. [8]).
4. Pinwheels of opposite signs have a tendency to be coupled, as estimated from the distribution of the first, nearest neighboring pinwheels of the same and opposite signs. (cf. [9]).
5. The topology of the learned patterns is reproducible for changes of the neuron density and for different geometries of the simulated neural grid.

4. Linear regression dynamics estimation and modeling

Our previous results suggest that cortical orientation maps seem to be optimal for the efficient coding of visual information in the sense of a redundancy minimization principle. The question arises as to how the dynamics that is underlying this pattern



(a) (b)

Fig. 2. (a) Swindale Kernels binned by distance ($\mu = 0.8(-. - .)$, $1.2(- \; -)$, $1.6(\_)$), as integrated for 4 maps for each μ. (b) Resulting map produced using interaction kernel obtained from formation dynamics of the maps with high sparsity ($\mu = 1.6$).

formation process is related to other generative models of cortical maps. To investigate this relationship, we fitted a simple dynamic model to the learning dynamics of our learning algorithm using the standard least-squares estimation procedure. We used the dynamic model that takes into account the periodicity of the orientations and the assumption that the interactions between the preferred orientations of different neurons should be dependent only on the distances between the receptive field centers. A generative model of this type has been proposed by Swindale [12]. The fitted model has the form

$$\partial/\partial t[\phi_i(t)] = \sum_{d} \sum_{j \in N(d(i,j))} W(d(i,j)) \sin(\phi_i(t) - \phi_j(t)), \tag{3}$$

where kernel $W(d(i,j))$ determines the interaction between the preferred orientations of different neurons in the learning dynamics and depends only on the distance between neurons i and j. $j \in N(d(i,j))$ indicates indices of neurons which is centered at distance d from neuron i.

Fig. 2 shows the results of the estimation of the interaction kernel $W(d(i,j))$ from the learning dynamics. For high sparseness, the interaction kernel is positive for small distances d. For medium distances, it is negative and the interaction strength decays to zero for large distances. This interaction, derived from our learning algorithm and the statistical properties of the image data corresponds exactly to the interaction kernel that has been proposed in the classical model by Swindale [12]. This is not the case for low sparseness, where the interactions between neighboring neurons are always negative. We simulated the generative model by Swindale using the estimated interaction kernels. For low and high sparseness, the maps generated by the Swindale model have the same distance autocorrelation function and power spectrum as the maps learned directly from the image data. While other models realize correlation-based learning with feedback via lateral connections [5,7,10], we have discovered such interactions by analyzing the dynamics of map formation driven only by the image statistics and the sparse coding principle.

5. Conclusion

Our results indicate that the principle of redundancy reduction might determine the structure of the primary visual cortex not only based on the level of individual receptive fields, but also on larger anatomical scales, such as on the level of orientation columns. Learning optimal representations of natural images in this information theoretical sense is sufficient to account for the formation of cortical maps with realistic parameters, even without the assumption of specific topologically prestructured lateral connections. The functional topology of the striate cortex on multiple anatomical scales may therefore be tailored for an encoding of natural visual stimuli in representations with minimal redundancy.

Acknowledgements

This work has been supported by SFB 509 founded by the Deutsche Forschungsgemeinschaft (DFG), HONDA R & D Americas, Inc., and the Volkswagen Stiftung.

The simulations were performed on Fujitsu VPP300/8 vector-parallel supercomputer. The supercomputer time was presented by Rechenzentrum der RWTH Aachen (project P158).

References

[1] H.B. Barlow, Unsupervised learning, Neural Comput. 1 (1989) 295–311.
[2] G.G. Blasdel, G. Salama, Voltage sensitive dyes reveals a modular organization in the monkey striate cortex, Nature 321 (1986) 579–585.
[3] T. Bonhoeffer, A. Grinvald, Orientation columns in cat are organized in pinwheel like patterns, Nature 353 (1991) 429–431.
[4] D.J. Field, Visual coding, redundancy, and feature detection, in: M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, The MIT Press, Cambridge, London, 1995, pp. 1012–1016.
[5] A. Hyvärinen, P.O. Hoyer, M. Inki, Topographic independent component analysis, Neural Comput. 13 (2001) 1525–1558.
[6] C. von der Malsburg, Self-organisation of orientation sensitive cells in the striate cortex, Kybernetik 14 (1973) 85–100.
[7] H. Nakagama, T. Saito, S. Tanaka, Effect of imbalance in activities between ON and OFF center LGN cells on orientation map formation, Biol. Cybernet. 83 (2000) 85–92.
[8] E. Niebur, F. Wörgötter, Design principles of column organization in visual cortex, Neural Comput. 6 (1994) 602–614.
[9] K. Obermayer, G.G. Blasdel, Singularities in primate orientation maps, Neural Comput. 9 (1997) 555–575.
[10] K. Obermayer, G.G. Blasdel, K. Schulten, A neural network model for the formation and for the spatial structure of retinotopic maps, orientation and ocular dominance columns, Proceedings of the ICANN91, Helsinki, 1991.
[11] B.A. Olshausen, D.J. Field, Emergence of simple cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.
[12] N.V. Swindale, A model for the formation of orientation columns, Proc. R. Soc. B 215 (1982) 211–230.