A non-negative sparse coding network learns contour coding and integration from natural images

Patrik O. Hoyer* and Aapo Hyvärinen

Neural Networks Research Centre, Helsinki University of Technology

P.O. Box 5400, FIN-02015 HUT, Finland

*) Corresponding author, email: patrik.hoyer@hut.fi

April 11, 2001

Abstract

An important approach in visual neuroscience considers how the function of the early visual system relates to the statistics of its natural input. Previous studies have shown how many basic properties of the primary visual cortex, such as the receptive fields of simple and complex cells and the spatial organization (topography) of the cells, can be understood as efficient coding of natural images. Here we extend the framework by considering how the responses of complex cells could be efficiently coded by a higher-order neural layer. This leads to contour coding and end-stopped receptive fields. Interestingly, contour integration can in this framework be seen as a direct result of top-down noise reduction, suggesting such a role for cortico-cortical feedback connections in the visual cortex.

Keywords: natural images, neural networks, contours, cortex, independent component analysis

Introduction

After Hubel and Wiesel (1962) first showed that neurons in mammalian primary visual cortex (V1) are optimally stimulated by bars and edges, a large part of visual neuroscience has been concerned with exploring the response characteristics of neurons in V1 and in higher visual areas. However, such studies do not directly answer the question of why the neurons respond in the way that they do. Why does it make sense to filter the incoming visual signals with V1 simple-cell like receptive fields? What is the goal of the neural code? Not only are such questions interesting in their own right, but finding answers would give us a deeper understanding of the information processing in the visual system and could even give predictions for neuronal receptive fields in higher visual areas.

One important approach for answering such questions is to consider how the function of the visual system relates to the properties of natural images. It has long been hypothesized that the early visual system is adapted to the input statistics (Attneave, 1954; Barlow, 1961). Such an adaptation is thought to be the result of the combined forces of evolution and neural learning during development. This hypothesis has lately been gaining ground as information-theoretically efficient coding strategies have recently been used to explain much of the early processing of visual sensory data, including response properties of neurons in the retina (Srinivasan et al., 1982; Atick and Redlich, 1992), LGN (Dan et al., 1996), and V1 (Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998; van Hateren and Ruderman, 1998; Hoyer and Hyvärinen, 2000; Tailor et al., 2000; Wachtler et al., 2001; Rao and Ballard, 1999; Hyvärinen and Hoyer, 2000; Simoncelli and Schwartz, 1999; Zetzsche and Krieger, 1999; Hyvärinen and Hoyer, in press).

In an influential paper, Olshausen and Field (1996) showed how the classical receptive fields of simple cells in V1 can be understood in a sparse coding framework. The basic idea is to model the observed multi-dimensional data (random vector) $\mathbf{x}$ by a linear generative model

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i s_i + \mathbf{n}, \tag{1}$$

where the s_i are mutually independent sparse 'source' variables, and $\mathbf{n}$ is gaussian noise. (Bold letters indicate vector quantities.) Sparseness is a property independent of scale, and implies that the sources s_i have probability densities which are highly peaked at zero and have heavy tails. Essentially, the idea is that any single typical input pattern $\mathbf{x}$ can be described accurately using only a few active (significantly non-zero) units s_i. However, all of the basis patterns $\mathbf{a}_i$ are needed because the set of active units changes from input to input.

When this model is estimated from patches of natural image data (with the elements of $\mathbf{x}$ representing pixel gray-scale values), the basis patterns $\mathbf{a}_i$ resemble Gabor functions and V1 simple cell classical receptive fields, and the corresponding s_i can be interpreted as the activations (firing rates) of the cells (Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998). This basic result has already been extended to explain spatiotemporal (van Hateren and Ruderman, 1998), chromatic (Hoyer and Hyvärinen, 2000; Tailor et al., 2000; Wachtler et al., 2001), and binocular (Hoyer and Hyvärinen, 2000) properties of simple cells, simply by estimating the model from input data consisting of image sequences, colour images, and stereo images, respectively. In addition, the sparse coding framework can be extended to show emergence of complex cell properties (Hyvärinen and Hoyer, 2000), as well as V1-like topography (Hyvärinen and Hoyer, in press).

An important question is how to extend these models to account for response properties of neurons higher in the processing hierarchy. Perhaps the most straightforward approach is to assume the feedforward processing up to V1 approximately known, and fit the original sparse coding model (Eq. 1) to the responses of model V1 neurons. Such an approach allows us to analyze the dependencies of V1 responses in the same way that the model has been used to analyze pixel dependencies in natural images (Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998).

Note that it is crucial that our model V1 outputs are nonlinear functions of the data; if not, there would be no hope of a second layer learning anything that could not be accomplished by a single layer. (This is because a linear transform followed by another linear transform is still a linear transform, so there is then no need for a multi-layer network.) However, all cells in V1 are to some degree nonlinear; even simple cells show significant nonlinearities (such as half-rectification or -squaring) (Heeger, 1992). Thus, any reasonably realistic model of V1 responses will be nonlinear, and this enables our network to learn higher-order features.

In this contribution we model V1 outputs by a classic complex cell model, and estimate the linear model in Eq. (1) assuming sparse, non-negative sources. Thus we attempt to model further layers in the visual hierarchy. We show how our network learns contour coding from natural images in an unsupervised fashion, and how contour integration can be understood as resulting from top-down noise-reducing feedback. Thus we obtain an elegant information-theoretical explanation of contour coding and integration in early vision.

Methods and results

Model complex cells

Numerous models of V1 receptive fields exist, any of which could be used to provide the simulated V1 output that is needed here. Ideally, one should include all kinds of receptive fields (simple and complex, disparity-selective, motion-selective, and so on) to allow our sparse coding model to analyze the entire set of dependencies between the firing rates of such neurons. Such an overly general analysis is, however, intractable for reasons of computational complexity. A more restricted analysis is, on the other hand, quite

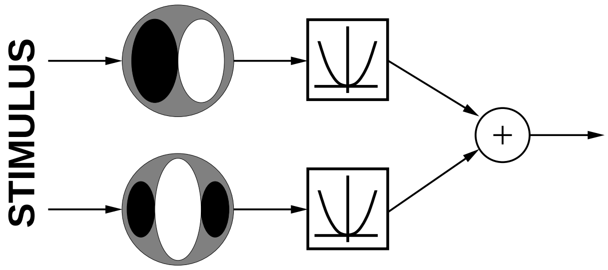

Figure 1: Complex cell model used. The response of a complex cell was obtained by linearly filtering with quadrature gabor filters, taking squares, and summing.

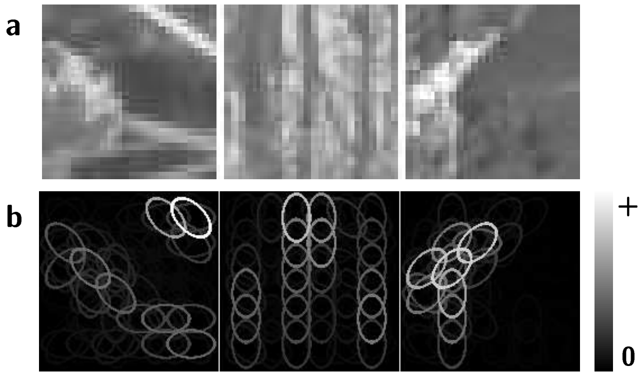

Figure 2: Model complex cell responses to natural image patches. **(a)** Three patches from the set of natural images. **(b)** Responses of the model complex cells to the patches. The output of each complex cell is depicted as a small ellipse (showing the orientation and the approximate extent of the receptive field), the brightness of which gives the response strength.

feasible, and in addition could be much simpler to understand and analyze.

We thus chose to consider here only a very basic model in which V1 outputs are given by 'complex cells' represented by a classic energy model (Pollen and Ronner, 1983; Heeger, 1992) detecting spatially localized oriented Fourier energy in static, monocular, achromatic images. The response of a model complex cell was calculated by summing the squared responses of a pair of quadrature gabor filters. (For details, see the Appendix.) This is depicted in Figure 1.

For simplicity of interpretation and for computational reasons, we restricted our analysis to a single spatial scale, and the cells were placed on a rectangular 6-by-6 grid and exhibited 4 different orientations. In Figure 2 we illustrate the behaviour of our model complex cells by showing their responses to a few natural image patches.

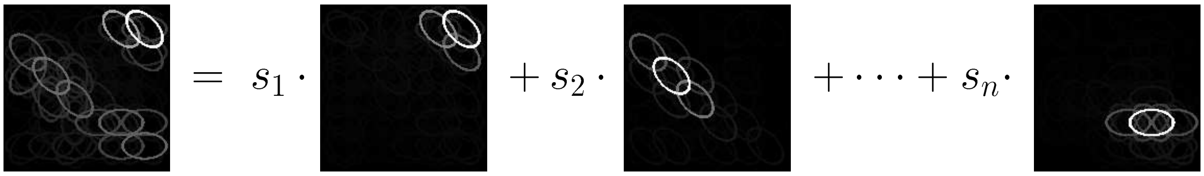

Figure 3: The linear model for complex-cell responses. Each activity pattern is represented as a linear combination of basis patterns $\mathbf{a}_i$. The goal in sparse coding is to find basis patterns such that the coefficients s_i are as 'sparse' as possible, meaning that for most input patterns only a few of them are needed to represent the pattern accurately. Cf. Equation 1.

Sparse coding of complex cell responses

Having sampled a large number of activation patterns (such as those shown in Figure 2b), we then fitted the sparse coding model (Eq. 1) to this data. Each complex cell activation pattern gives one data vector $\mathbf{x}$ (with each element representing the firing rate of one neuron). Each source s_i represents the activation level of one higher-order neuron, whose 'receptive field' is closely related to the corresponding $\mathbf{a}_i$. The goal is to find basis patterns $\mathbf{a}_i$ such that typical input patterns $\mathbf{x}$ can be described accurately by only a few significantly active higher-order neurons. This 'higher-order' sparse coding model is graphically illustrated in Figure 3.

Because firing rates cannot go negative it is natural to require our representation to be non-negative. Thus both the $\mathbf{a}_i$ and the s_i were restricted to non-negative values. Arguments for non-negative representations have previously been convincingly presented by Lee and Seung (1999). However, in contrast to their approach, we emphasize the importance of sparseness in addition to good reconstruction. Such emphasis on sparseness has previously been forcefully argued for by Barlow (1972) and Field (1994). Thus, we combine sparse coding and the constraint of non-negativity into a single model. Note that neither assumption is arbitary; both follow from the properties of complex cell responses, which are sparse and non-negative.

Properties of the found representation

A representative subset of the 288 estimated basis patterns $\mathbf{a}_i$ is shown in Figure 4. (The details of the learning procedure are given in the Appendix.) Note that most basis patterns consist of a variable number of active complex cells arranged collinearly. This makes intuitive sense, as collinearity is a strong feature of the visual world (Krüger, 1998; Sigman et al., 2001). In addition, analyzing images in terms of smooth

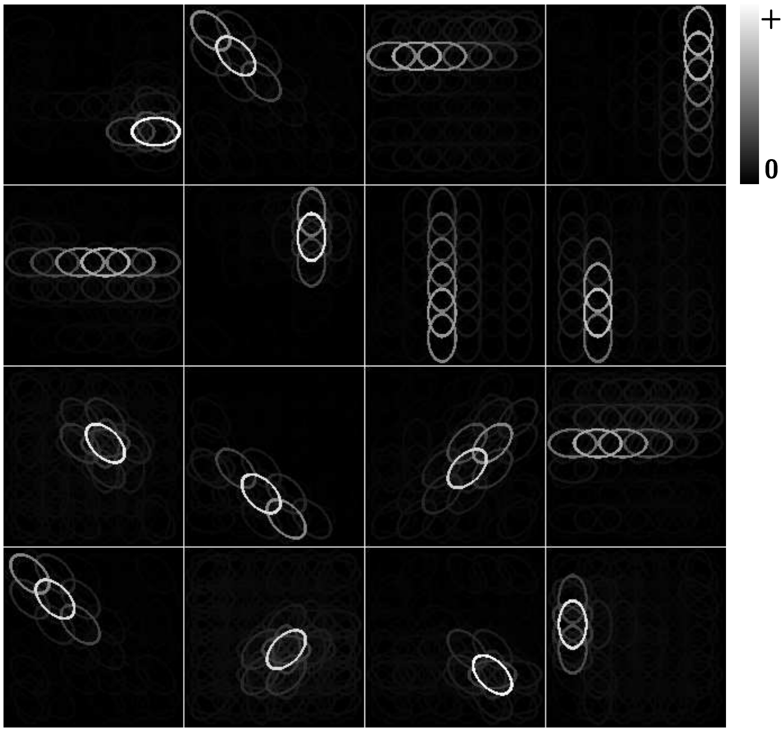

Figure 4: A representative set of basis functions from the learned basis. The majority of units code the simultaneous activation of collinear complex cells, indicating a smooth contour in the image.

contours is supported by evidence from both psychophysics (Field et al., 1993; Polat and Sagi, 1993) and physiology (Kapadia et al., 1995; Polat et al., 1998; Kapadia et al., 2000), and is incorporated in many models of contour integration, see e.g. (Grossberg and Mingolla, 1985; Li, 1999; Neumann and Sepp, 1999). To our knowledge, ours is the first model to learn this type of a representation from the statistics of natural images.

It is easy to understand why basis patterns should consist of collinear complex cell activity patterns: Such patterns are typical in the data set, and can be coded efficiently if a whole contour (or at least a large part of one) can be represented by a single higher-level unit. The necessity for different length basis patterns comes from the fact that long basis patterns simply cannot code short (or curved) contours, and short basis patterns are inefficient at representing long, straight contours.

It should be noted that those higher-order neurons which represent long contours bear many similarities to 'collator' (or 'collector') units, proposed in the psychophysical litterature (Mussap and Levi, 1996; Moulden, 1994). Such units are thought to integrate the responses of smaller, collinear filters, to give a more robust estimate of global orientation than could be achieved with elongated linear mechanisms.

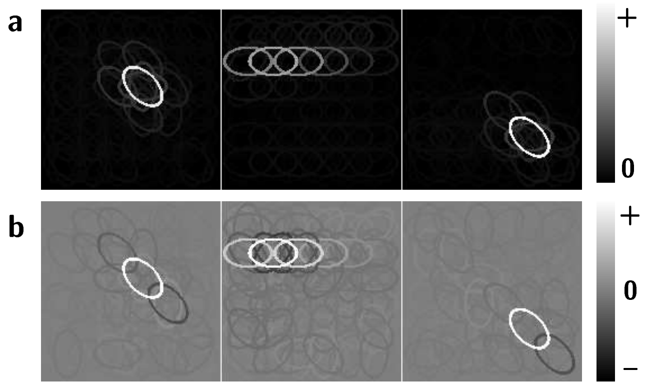

Figure 5: **(a)** Three basis patterns from the estimated basis. **(b)** Optimal approximating linear filters for the units in (a). These are the filters that minimize the mean squared error between the linear response (followed by half-rectification) and the optimal activations.

Nonlinear responses of higher-order neurons

Although the model is linear from the sources s_i to the data $\mathbf{x}$, the optimal (most likely) s_i are a nonlinear function of the data $\mathbf{x}$, due to the noise and the *overcompleteness* of the basis (Olshausen and Field, 1997). In other words, the higher-order neurons respond to the complex cell activity patterns in a nonlinear fashion. In particular, there is competition between the neurons (Olshausen and Field, 1997), so that they respond only when they are better than competing units at representing the stimulus. As a prominent feature of the found representation is the existence of different-length patterns, this leads to units being selective for contour length, in addition to being tuned to position and orientation: 'Collator' units do not respond to short contours, whereas those units representing short contours exhibit *end-stopping* (Hubel and Wiesel, 1965).

To illustrate the nonlinear transform from complex cell activities $\mathbf{x}$ to higher-order activities s_i we can make a linear approximation. Optimal approximating linear filters are shown in Figure 5b, for the units whose basis patterns are depicted in Figure 5a. Note that units representing short contour segments tend to have inhibitory regions at one (or both) of the ends of their 'receptive fields', illustrating the end-stopping effect. On the other hand, units which code longer contours have inhibitory weights from complex cells which are positioned on the contour but are of the wrong orientation. This enhances the selectivity of these units so that they don't respond to contours that only partly overlap the receptive field. Thus all higher-level units have extra-classical properties that make them clearly distinct from standard complex cells.

Contour integration through feedback noise reduction

A central question in visual neuroscience concerns the computational role of feedback connections. It has been suggested that the purpose of feedback is that of using information from higher-order units to modulate lower-level outputs, so as to selectively enhance responses which are consistent with the broader visual context (Lamme, 1995). In our sparse coding model, this is naturally interpreted as the reduction of noise in the bottom-up signals (Hupé et al., 1998). Note that 'noise' in this context refers to any activity that is not consistent with the learned statistical model and is thus not (only) neural or dark noise.

Why would feedback noise reduction be useful? In many cases, there can be multiple conflicting interpretations of the stimulus even on the lowest level, and top-down feedback is needed to resolve such conflicts. In essence, feedback noise reduction computes the most likely interpretation of the scene (Knill and Richards, 1996), combining bottom-up information with top-down priors.

Feedback noise reduction would also be a useful way of communicating between separate pathways. For example, a motion pathway could use motion cues to infer a contour, and subsequently use noise-reducing feedback to activate contour-coding neurons. (This because luminance contours and motion contour cues are probably statistically quite dependent, and one often implies the other.) Shape recognition could then proceed as usual, giving rise to shape-from-motion.

Noise reduction is very straightforward in the sparse coding model (Eq. 1). Having found the optimal (most likely) s_i (i.e. the higher-order neuron activations), we can obtain the noise-reduced complex cell activity directly as $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{a}_i s_i$. Such noise-reduction (Hyvärinen, 1999; Lewicki and Olshausen, 1999; Simoncelli and Adelson, 1996) essentially suppresses responses which are not consistent with the statistics of the training data, while retaining responses that do fit the learned statistical model. (See the Appendix for details.)

In Figure 6, we show how feedback noise reduction in our model results in the emphasis of smooth contours. We generated image patches by placing Gabor filters at random locations and orientations. In one case, there was a collinear alignment of three consecutive Gabors; in the other these same Gabors had random orientations. These image patches are shown in Figure 6a. Next, we processed these by our model complex cells, as we had processed the natural image patches in our main experiments. The result is shown in Figure 6b. Finally, we estimated the higher-level activities s_i, and plotted the noise-reduced complex cell activity $\sum_{i=1}^{n} \mathbf{a}_i s_i$ in Figure 6c.

Note how the noise reduction step supresses responses to 'spurious' edges that do not form part of the

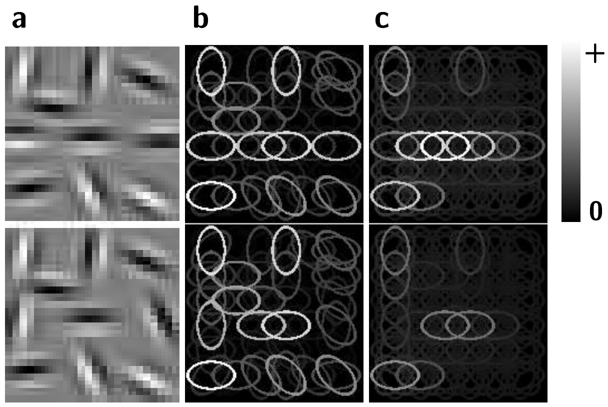

Figure 6: Noise reduction performs contour integration. **(a)** Two image patches containing Gabors at random locations and orientations. In the top patch there is a collinear set of three Gabors, whereas in the bottom patch these same Gabors had random orientations. **(b)** The response of the model complex cells to the images in (a). **(c)** The response of the complex cells after feedback noise reduction using the learned higher-order representation. Note that the reduction of noise has left the activations of the collinear stimuli but suppressed activity that did not well fit the learned sparse coding model.

collinear arrangement. Many computational models of contour integration have been proposed (see e.g. (Grossberg and Mingolla, 1985; Li, 1999; Neumann and Sepp, 1999)), that contain some form of process by which responses of cells forming part of a contour are amplified relative to others. In particular, it has been argued that such contextual effects should be seen not at the initial burst of activity but rather through the sustained response of the cell after a significant time delay (Hess et al., 1998). Exactly that kind of dynamics is predicted by our model, which thus provides an information-theoretical explanation of such contour integration.

Comparing the denoised responses (Figure 6c) with each other one can also see collinear contextual interactions in the model. The response to the central gabor is stronger when it is flanked by collinear gabors (upper row) than when the flankers have random orientations (bottom row), even though the flankers fall well outside the receptive field of the central neuron. This type of contextual interactions have been the subject of much study recently (see e.g. (Kapadia et al., 1995; Polat et al., 1998; Kapadia et al., 2000)), and are hypothesized to be somehow related to contour integration, although such a relation is not certain.

Discussion

Several authors (Krüger, 1998; Sigman et al., 2001; Geisler et al., 2001) have studied the mutual dependencies of 'complex-cell' responses from natural images. The main result from these studies is that complex-cell responses are not mutually independent, and in particular they are highly correlated when the complex-cells are arranged collinearly (Krüger, 1998) or on a common circle (Sigman et al., 2001). However, these investigations only considered pairwise statistics of responses, not higher-order statistics. From previous work in the area of natural image statistics, we know that higher-order statistics are of fundamental importance, e.g. in explaining oriented receptive fields (Zetzsche and Krieger, 1999; Field, 1994; Olshausen and Field, 1996). Thus it is probably important to consider higher than second-order statistics also when analyzing dependencies between complex-cell outputs.

An approach related to ours is given by the predictive coding model of Rao and Ballard (1999). They proposed a multi-layered network where the higher-level units provide predictive feedback to the lower units. The lower-level neurons then send forward residuals between the observed data (the stimulus) and the prediction. There are no 'fixed weights' (such as the complex-cell filtering of our model) in their network. Although the generality of their approach can be a benefit, it is at the same time also a clear drawback in that the analysis and interpretation of the results is much less straightforward than in more simple, restricted models.

Another recent study related to ours is that given by Park et al. (2000). They show how receptive fields sensitive to translation, contraction, expansion, and rotation (such as those observed in area MSTd) emerge when learning a sparse code for optical flow fields. Although the inputs to their model are vector fields, not simulated neuron outputs, their work nevertheless shows that even quite complex neuronal response properties could be learned from the visual input statistics in a completely unsupervised manner.

Note that although contour integration in our model is explicitly handled by feedback connections from higher-order units, this does not rule out the role of horizontal ('intrinsic') connections between complex-cells in V1. On the contrary, in a neural implementation, it would probably be quite useful to implement at least part of the noise reduction by the use of horizontal connections. This paper suggests that the contour integration performed by these connections can be viewed as noise reduction, thus implicating a possible general purpose of horizontal connections in the visual cortex.

It should also be noted that our model did not show emergence of any units tuned to curved contours. This might be caused by the limited sampling in space and orientation by our model complex cells, which thus cannot represent all curves equally well; future larger-scale experiments are needed to explore this issue. Another possibility is that curved contours are best represented by standard complex cells (bound by the horizontal connections in V1), and only straight contours are coded by 'collator' units. Or, perhaps curved contours should really be represented at the next hierarchical layer. In such a representation, curved contours would be described as the activation of a suitable combination of end-stopped units such as those found in this study.

In conclusion, we have extended the sparse coding framework to work on the outputs of model V1 cells. This provides a way towards predicting and understanding neuronal response properties higher in the visual processing hierarchy. In addition, the framework suggests to view feedback pathways as performing noise reduction on the lower level activities. Specifically, we have shown how contour coding and end-stopped receptive fields, as well as contour integration, can be understood in this efficient coding framework.

Appendix

To facilitate the reproduction of the results we provide the complete MATLAB code for all experiments at `http://www.cis.hut.fi/phoyer/contournet.tar.gz`.

A sample of 100,000 image patches of 24×24 pixels were sampled from natural images available on the WWW (`http://www.cis.hut.fi/projects/ica/data/images`). The output of a model complex cell was given by

$$C_{\{x_c,y_c,\theta\}} = \left(\sum_{x,y} G_{\{e,x_c,y_c,\theta\}}(x,y)I(x,y)\right)^2 + \left(\sum_{x,y} G_{\{o,x_c,y_c,\theta\}}(x,y)I(x,y)\right)^2 \qquad (2)$$

when fed with the image patch $I(x,y)$. Here $G_{\{e,x_c,y_c,\theta\}}$ and $G_{\{o,x_c,y_c,\theta\}}$ are even- and odd-symmetric (respectively) Gabor filters, centered on (x_c, y_c) and of orientation θ. The gaussian envelope had an aspect ratio of 1.5 and the spatial frequency of the sinusoid was chosen to give $2-3$ main excitatory/inhibitory regions, in accordance with measurements of receptive fields in V1 (DeAngelis et al., 1993).

The model (Eq. 1) was estimated by maximizing the log-likelihood of the data. This is equivalent to minimizing

$$C(\mathbf{A}, \mathbf{s}^{(n)}, n = 1 \dots N) = \sum_n \left[\|\mathbf{x}^{(n)} - \mathbf{A}\mathbf{s}^{(n)}\|^2 + \lambda \sum_i s_i^{(n)} \right] \qquad (3)$$

where $\mathbf{A}$ denotes the matrix containing the $\mathbf{a}_i$ as its columns, $\mathbf{x}^{(n)}$ represents the n:th data vector, and $\mathbf{s}^{(n)}$ is the vector containing the sources $s_i^{(n)}$ corresponding to the n:th data vector. The linear activation penalty term comes from assuming the sources to be exponentially distributed, i.e. $p(s_i) = \exp(-s_i)$, while the quadratic term derives from the Gaussian prior on the noise. The constant λ defines the tradeoff between representation error and sparseness, and is obtained directly from the noise level (selected manually) assumed in the model. The objective was minimized (under the non-negativity constraint on both $\mathbf{A}$ and the $\mathbf{s}^{(n)}$) by gradient descent with respect to the $\mathbf{s}^{(n)}$ in the short run and with respect to $\mathbf{A}$ under a longer timescale, as in (Olshausen and Field, 1996). An additional constraint was that each source was assumed to contribute equally to the data, i.e. $\|\mathbf{a}_i\| = \|\mathbf{a}_j\|$ for all i, j.

The weights shown in Figure 5 were obtained by first finding the optimal s_i for all data points by minimizing the objective (Eq. 3) and then iteratively seeking weights $\mathbf{w}_i$ that minimize $\sum_n \{(s_i^{(n)} - \mathrm{rect}(\mathbf{w}_i^T \mathbf{x}^{(n)}))^2\}$, where rect() denotes half-rectification.

The denoising was done by, for a given data vector $\mathbf{x}$, finding the most likely sources by minimizing $\|\mathbf{x} - \mathbf{A}\mathbf{s}\|^2 + \lambda \sum_i s_i$ with respect to $\mathbf{s}$. The denoised data vector is then given by $\mathbf{A}\mathbf{s}$.

References

Atick, J. J. and Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, 4:196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61:183–193.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. A., editor, *Sensory Communication*, pages 217–234. MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Bell, A. J. and Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338.

Dan, Y., Atick, J. J., and Reid, R. C. (1996). Efficient coding of natural scences in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, 16:3351–3362.

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69:1091–1117.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, 6:559–601.

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local 'association field'. *Vision Research*, 33(2):173–193.

Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724.

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries and emergent segmentations. *Perception and psychophysics*, 38(2):141–171.

Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198.

Hess, R. F., Dakin, S. C., and Field, D. J. (1998). The role of 'contrast enhancement' in the detection and appearance of visual contours. *Vision Research*, 38(6):783–787.

Hoyer, P. O. and Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, 11(3):191–210.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154.

Hubel, D. H. and Wiesel, T. N. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28:229–289.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394:784–787.

Hyvärinen, A. (1999). Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11(7):1739–1768.

Hyvärinen, A. and Hoyer, P. O. (2000). Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12(7):1705–1720.

Hyvärinen, A. and Hoyer, P. O. (in press). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*.

Kapadia, M. K., Ito, M., Gilbert, C. D., and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15(4):843–856.

Kapadia, M. K., Westheimer, G., and Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84:2048–2062.

Knill, D. C. and Richards, W., editors (1996). *Perception as Bayesian Inference*. Cambridge University Press.

Krüger, N. (1998). Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8:117–129.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15(2):1605–1615.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lewicki, M. and Olshausen, B. (1999). Probabilistic framework for the adaption and comparison of image codes. *J. Opt. Soc. Am. A: Optics, Image Science, and Vision*, 16(7):1587–1601.

Li, Z. (1999). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13(1):25–50.

Moulden, B. (1994). Collator units: Second-stage orientational filters. In *Higher-order processing in the visual system. Ciba Foundation Symposium 184*, pages 170–192. John Wiley, Chichester, U.K.

Mussap, A. J. and Levi, D. M. (1996). Spatial properties of filters underlying vernier acuity revealed by masking: Evidence for collator mechanisms. *Vision Research*, 36(16):2459–2473.

Neumann, H. and Sepp, W. (1999). Recurrent V1-V2 interaction in early visual boundary processing. *Biological Cybernetics*, 81:425–444.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325.

Park, K.-Y., Jabri, M., Lee, S.-Y., and Sejnowski, T. (2000). Independent components of optical flows have MSTd-like receptive fields. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 597–602.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., and Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 391:580–584.

Polat, U. and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33:993–999.

Pollen, D. A. and Ronner, S. F. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. on Systems, Man, and Cybernetics*, 13:907–916.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, 2(1):79–87.

Sigman, M., Cecchi, G. A., Gilbert, C. D., and Magnasco, M. O. (2001). On a common circle: Natural scenes and gestalt rules. *Proceedings of the National Academy of Science, USA*, 98:1935–1940.

Simoncelli, E. P. and Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *Proc. Third IEEE International Conference on Image Processing*, pages 379–382, Lausanne, Switzerland.

Simoncelli, E. P. and Schwartz, O. (1999). Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In *Advances in Neural Information Processing Systems 11*, pages 153–159. MIT Press.

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proc. Royal Society ser. B*, 216:427–459.

Tailor, D. R., Finkel, L. H., and Buchsbaum, G. (2000). Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40:2671–2676.

van Hateren, J. H. and Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:2315–2320.

van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:359–366.

Wachtler, T., Lee, T.-W., and Sejnowski, T. J. (2001). The chromatic structure of natural scenes. *Journal of the Optical Society of America A*, 18(1):65–77.

Zetzsche, C. and Krieger, G. (1999). Nonlinear neurons and higher-order statistics: new approaches to human vision and electronic image processing. *Proc. SPIE*, 3644:2–33.

Acknowledgements

The authors wish to thank Bruno Olshausen for making available the sparsenet source code (on which our code is based) and Harri Valpola for useful suggestions and valuable comments on the manuscript. In addition, thanks are due to Jarmo Hurri, Mika Inki and Pentti Laurinen for interesting discussions and Erkki Oja for a stimulating work environment. P.O.H. was funded by the Helsinki Graduate School in Computer Science and Engineering. A.H. was funded by the Academy of Finland.