Biol. Cybern. 91, 408–416 (2004)
DOI 10.1007/s00422-004-0524-0

Biological Cybernetics

Is sparse and distributed the coding goal of simple cells?

Li Zhao

Brain and Cognitive Sciences Institute, Henan University of Science and Techology, 6 An Hui Road, Luoyang, Henan 471003, China

Received: 7 January 2004 / Accepted: 28 September 2004 / Published online: 30 November 2004

Abstract. The question of why the receptive fields of simple cells in the primary visual cortex are Gabor-like is a crucial one in vision research. Many research efforts (Olshausen and Field 1996, 1997; van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998) that yield a set of localized, oriented, and bandpass Gabor-like receptive fields believe that sparse and distributed is the coding goal of simple cells. This paper investigates a more general coding strategy that measures equally any departure from normality in the simple cells' responses. That is, we investigate the possibility that highly kurtotic response histograms may result if simple cells explicitly seek, not maximally kurtotic, but rather maximally non-Gaussian response histograms to natural images. It is found that, under this coding strategy, the simulations produce a majority of localized, oriented, bandpass (Gabor-like) receptive fields. Some receptive fields, however, are spatially distributed and show little oriented structure. Nearly all receptive fields, *regardless of whether they are Gabor-like or non-Gabor-like*, yield highly kurtotic response histograms to natural images. Thus, in seeking maximally non-Gaussian response histograms, receptive fields spontaneously yield highly kurtotic histograms. The presence in our ensemble of nonlocalized, nonoriented receptive fields may be due to the artificial requirement that receptive fields be orthonormal. We conclude that the high kurtoses observed in the response histograms of simple-cell receptive fields to natural images may reflect a property of natural images themselves rather than an explicit coding goal used to structure simple-cell receptive fields.

1 Introduction

The question of why simple cells in primary visual cortex (V1) possess Gabor-like receptive fields is an active topic in vision research. The approach applied by many researchers is to consider the properties of the inputs to these cells together with other constraints. Evolutionary considerations have led researchers to focus on natural images (i.e., images of rocks, trees, mountains and other non-man-made structures). Among the constraints, some researchers suggest that the purpose of coding is to try to achieve a sparse and distributed coding of the input images (Field 1994; Olshausen and Field 1996, 1997).

Field (1994, 1987) found that the projections of a population of natural images to a simple cell conforms to a highly kurtotic distribution (Fig. 1). That is, many responses are near 0, but the nonzero responses tend to be large in absolute value. This led Field (1994) to propose that the ensemble of simple cells collectively achieves a sparse and distributed code for the set of natural images. A code is sparse if most neurons are silent in response to any given input image, while those that are active are strongly active. A code is distributed if all neurons are activated equally often across the set of all input images. Under Field's view, the visual system explicitly builds sparse and distributed representations into the code achieved by simple cells. It sets out to construct a code that is as sparse and well distributed as it can be. Thus one should explicitly seek receptive fields whose distributions of responses to natural images are highly kurtotic.

If this is true, then the kurtosis

$$K = \frac{E((X - E(X))^4)}{\sigma^4} - 3 \tag{1}$$

is a good measure since any departure from normality at the center and the tail give much larger kurtosis than any other part does.

Although this consideration fits the empirical study of response distribution of simple cells to natural images very well (Field 1994), it is important to note that, even if some sort of "sparseness" were the crucial factor, it is not clear that kurtosis would be the most adequate measure. The possibility remains that human vision seeks simply a code in which each receptive field produces a response distribution deviating as strongly as possible from normality. In this case the high kurtosis of simple-cell response distribution may reflect an empirical property of natural images

Correspondence to: L. Zhao
(e-mail: lzhao@mail.haust.edu.cn)

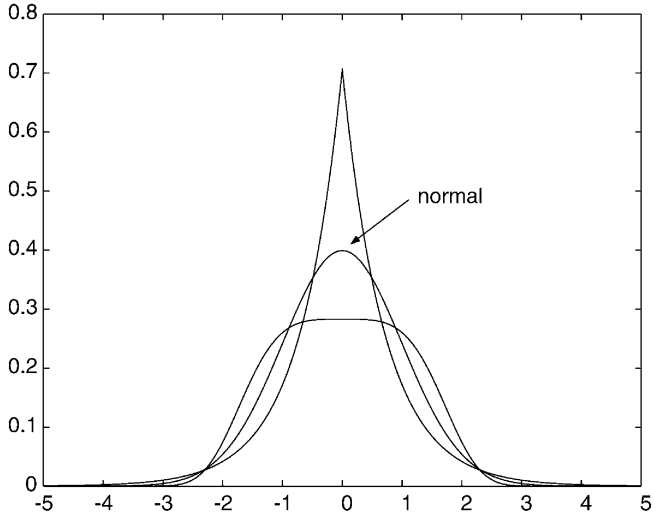

Fig. 1. Distributions with different kurtoses

instead of a coding principle of human vision. Moreover, there is no theoretical reason why one should assume that the tails in the distribution should be given a special weight in the measure.

In (1), σ is the standard deviation of random variable X. In this paper, X is the response of a simple cell to a natural image. If X is normally distributed, K is zero. For a distribution with a positive K, the distribution has a large peak at zero and heavy tails. The larger the K, the larger the peak and the heavier the tails. For a distribution with a negative K, the distribution becomes flatter than Gaussian. The smaller the K, the flatter the distributions if variance is held constant. (Figure 1 is an illustration of different kurtoses.)

Following the sparse and distributed argument, Olshausen and Field (1996, 1997) demonstrated a neural network that can generate Gabor-like receptive fields by minimizing

$$E = \sum_{x,y} \left[\mathbf{I}(x, y) - \sum_i a_i \boldsymbol{\phi}_i(x, y) \right]^2 + \lambda \sum_i S\left(\frac{a_i}{\sigma}\right) . \qquad (2)$$

Here $\mathbf{I}$ is the input image, (x, y) are the coordinates to index a pixel of image $\mathbf{I}$, and i is the index to all the receptive fields.

The second part of (2) is a measurement of the sparseness. In implementation, S is selected as a function that is unimodal and peaks at zero (and hence penalizes any nonzero responses/coefficients a_i) such as $\log(1+x^2)$.

The first part of (2) is the measurement of reconstruction of the input image $\mathbf{I}$ by receptive fields $\boldsymbol{\phi}_i$ and their responses a_i. The smaller the better for this part. An error-free reconstruction (this term is zero) implies that for every input image $\mathbf{I}$, there is a perfect internal representation by both the receptive fields $\boldsymbol{\phi}_i$ and their responses a_i.

The receptive fields generated by this neural network method are compelling – they are localized, oriented, and bandpass Gabor-like patches.

Although the research by Olshausen and Field (1996, 1997) generated compelling receptive fields, it still can only assert that sparse and distributed coding is sufficient for the generation of Gabor-like receptive fields. It remains unclear whether sparse and distributed coding is *necessary* for producing Gabor-like receptive fields.

2 A new measure that seeks maximally non-Gaussian simple-cell responses

If $\boldsymbol{\psi}$ is the receptive field of a simple cell in V1 and $\mathbf{I}$ is the input image, dot product $\boldsymbol{\psi}\mathbf{I}$ is the response of the simple cell. In this research, natural images $\mathbf{I}$ are used as the inputs in the light of evolutionary theory. According to the theory of projection pursuit (Friedman 1987), the statistical distribution of $\boldsymbol{\psi}\mathbf{I}$ is almost always Gaussian. Therefore, as a statistical heuristics based on projection pursuit, a receptive field $\boldsymbol{\psi}$ should make the distribution of $\boldsymbol{\psi}\mathbf{I}$ as non-Gaussian as possible. Then the problem becomes how to measure the nonnormality of $\boldsymbol{\psi}\mathbf{I}$.

As discussed in the last section, the simulations of Olshausen and Field (1996, 1997) explicitly seek receptive fields whose distributions of responses to natural images are highly kurtotic. However, it is unclear why the tails in the distribution should be given a special weight in the measure.

An alternative with minimum assumptions is that every departure from normality in the distribution of responses should be weighted equally. Therefore, clusters between peak and tail are treated as identical to tails. If the generated receptive fields are not similar to the receptive fields in V1, this indicates that nature most likely uses a more special measure, such as kurtosis, to favor some special designs. These designs must have special biological advantages. For a high kurtosis distribution, it has a spikelike peak at zero and a fat tail. This indicates that it is very unlikely to fire in response to a randomly chosen natural image, yet when it does fire, it fires strongly. This is very much like a digital device given the fact that neural cells only function on the positive part. This can give a robust design in V1. However, if the receptive fields generated by this new measure are similar to those of simple cells, this will indicate that the coding goal may not be to achieve a sparse and distributed code as assumed by previous research. Even if the receptive fields generated using the new measure do produce sparse and distributed responses, one cannot say the goal is to produce a sparse and distributed code. In other words, sparse and distributed coding may be *sufficient* a priori to produce a sparse and distributed coding, but it may be not *necessary* to produce a sparse and distributed coding. Following this argument, we give a measure with minimum assumptions that weights all departures from normality equally.

Consider a receptive field $\boldsymbol{\psi}$ and a randomly chosen input image $\mathbf{I}$. Then the response of the simple cell with receptive field $\boldsymbol{\psi}$ is

$$X = \boldsymbol{\psi}\mathbf{I} . \qquad (3)$$

If X has a standard normal distribution, then one can obtain a random variable R uniformly distributed on $(-1, 1)$ by setting

$$R = 2\Phi(X) - 1\,, \tag{4}$$

where Φ is the standard normal cumulative distribution function. That is, for any $x \in \mathbb{R}$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt\,. \tag{5}$$

Thus, if X is standard normal, then the probability density function of R is

$$p(r) = \frac{1}{2}, \quad r \in (-1, 1)\,. \tag{6}$$

Therefore, an equally weighted measure of departure from normality (Friedman 1987) is

$$C(\boldsymbol{\psi}) = \int_{-1}^{1} \left(p(r) - \frac{1}{2} \right)^2 dr = \int_{-1}^{1} p^2(r) dr - \frac{1}{2}\,. \tag{7}$$

This measure may be only one of many possible measures. However, it is also possible that all these measures turned out be equivalent. Nonetheless, this topic is an interesting question worth pursuing in future research, but it is beyond the scope of this paper.

Equation (7) cannot be calculated in general since the probability density function $p(r)$ is not known. To get around this, it can be approximated as follows. First, $p(r)$ can be expanded as a linear sum of Legendre polynomials. (We would like to point out that other orthonormal polynomials may also be used. The result, we think, may not be changed since the crucial thing is the order of the polynomials.):

$$p(r) = \sum_{j=0}^{\infty} a_j P_j(r)\,, \tag{8}$$

where P_j is the jth-order normalized Legendre polynomial, and

$$a_j = \int_{-1}^{1} P_j(r) p(r) dr = E(P_j(R))\,. \tag{9}$$

Hence (7) becomes

$$C(\boldsymbol{\psi}) = \int_{-1}^{1} \left(\sum_{j=0}^{\infty} a_j P_j(r) \right) p_R(r) dr - \frac{1}{2}\,. \tag{10}$$

Noting that $E^2(P_0(R)) = 1/2$, one then obtains

$$C(\boldsymbol{\psi}) = \sum_{j=1}^{\infty} E^2(P_j(R))\,. \tag{11}$$

In practice it suffices to keep only the first several items of C. To overcome the shortcomings of the kurtosis (e.g., a small difference in the tail yields a large difference in the kurtosis), it is found that $J = 5$ (which is also one order greater than the order of the kurtosis) is enough for this purpose (Friedman 1987). A larger J yields no fundamental difference (Friedman 1987). (In all the simulations carried out in this research, $J = 5$. We also tried for $J = 10$ or 20. However, there were no fundamental differences in the results.) Hence C can be approximated as

$$C(\boldsymbol{\psi}) = \int_{-1}^{1} p_R^2(r) dr - \frac{1}{2} \approx \sum_{j=1}^{J} E_R^2(P_j(R))\,. \tag{12}$$

C can be readily estimated given a large population of $\mathbf{I}$.

Specifically, for a set comprising N randomly selected images $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_N$, one can estimate the left side of (12) by

$$C(\boldsymbol{\psi}) = \sum_{j=1}^{J} \left(\frac{1}{N} \sum_{i=1}^{N} P_j(2\Phi(X) - 1) \right)^2\,. \tag{13}$$

3 A functional model and its whitening procedure

A functional model of the simple cells in the V1 can be formulated based on the objective function given in the last section. According to this model, simple cells in the V1 function in such a way as to catch any departure from normality in their responses to the input images. Mathematically, a simple cell is a biological optimizer of (12).

Since the data have large different variances in different directions, this makes searching for $\boldsymbol{\psi}$ difficult (Friedman 1987). To overcome this, in Friedman's projection pursuit (Friedman 1987), the data are first processed by principal component analysis. This is the whitening procedure in projection pursuit. This procedure makes the processed data uncorrelated and with equal unit variances. In this procedure, no principal components are removed (Friedman 1987). However, removing these principal components of small eigenvalues affects little the reconstruction of the input images. Therefore, whitening by principal component transformation with principal components of large eigenvalues is also used in the later simulations.

In the following section, a systematic simulation study with different whitening procedures that retain 100%, 99%, 98%, 96.5%, 95%, and 90% of the energy of the eigenvalues was carried. The percent energy of eigenvalues is the ratio of the sum of the largest eigenvalues (corresponding to the retained principal components) to the sum of all eigenvalues. The estimated frequency cutoffs are about 6.7 (99%), 6 (98%), 5 (96.5%), 4 (95%), and 2.5 (90%) circles/image.

4 Simulations for natural images

In the discrete domain, the input image $\mathbf{I}$ is a column vector in $\mathbb{R}^n$ (i.e., $n \times 1$ matrix), and the receptive field $\boldsymbol{\psi}$ is a row vector in $\mathbb{R}^n$ (i.e., $1 \times n$ matrix). The objective function is generalized to m receptive fields as follows. Let $\boldsymbol{\Psi}$ be the $m \times n$ matrix whose rows are receptive fields $\boldsymbol{\psi}_i$ ($i = 1, 2, \ldots, m$). Then

$$C(\Psi)=\sum_{i=1}^{m}\sum_{j=1}^{J}E^2(P_j(R_i))\,, \tag{14}$$

where R_i (corresponding to the ith receptive field) is as follows:

$$R_i=2\Phi(\psi_i\mathbf{I})-1\,. \tag{15}$$

In the simulation, Ψ is an $n\times n$ matrix.

For this study, the input images and receptive fields are all 12×12 matrices. These input images are raster scanned to form 144-dimensional column vectors (i.e., 144×1 matrix), and the receptive fields are raster scanned to form 144-dimensional row vectors (i.e., 1×144 matrices). Accordingly, the dot products $\psi_i\mathbf{I}$ ($i=1,2,\ldots,144$) can be construed as responses of simple cells in the V1. These small 12×12 images are generated from the same five natural images (Fig. 2) as were used in (Bell and Sejnowski 1996). A set of 6000 (see Fig. 3 for a sample) small 12×12 images is used in the simulation.

The images are first whitened. To this end, a principal component analysis is carried out. The covariance matrix of the input images (denoted as COV) is calculated by

$$COV=E((\mathbf{I}-E(\mathbf{I}))(\mathbf{I}-E(\mathbf{I}))^T)\,. \tag{16}$$

The eigenvectors of the covariance are the principal components. These vectors form an $n\times n$ (n is the number of pixels in an image I) matrix V. The eigenvalues indicate the variance magnitudes of the corresponding principal components. These eigenvalues can form a diagonal matrix D. Assume the eigenvalues (thus the corresponding eigenvectors) in the diagonal matrix D are arranged in descending order, i.e., $D(1,1)$ is the largest eigenvalue, $D(2,2)$ is the second largest eigenvalue, ... , and $D(n,n)$ is the smallest eigenvalue. Therefore, one has the following equation:

$$COV=VDV^T\,. \tag{17}$$

Then the input images are processed as follows.

$$\tilde{\mathbf{I}}=D^{-1/2}V^T(\mathbf{I}-E(\mathbf{I}))\,, \tag{18}$$

Fig. 2. A natural image used in the simulation

Fig. 3. A sample collection of 12×12 images of the total 6000 used in the simulation

where D is a $w\times w$ diagonal matrix formed by the first w largest eigenvalues. V is an $n\times w$ matrix formed by the corresponding eigenvectors. w is chosen as the number of eigenvalues, which accounts for 100% ($w=144$), 99% ($w=101$), 98% ($w=75$), 96.5% ($w=55$), 95% ($w=44$), and 90% ($w=22$) of the sum of all eigenvalues. Figure 4 shows some examples of the original image for 100% and processed images for 98% and 90%.

Thus $D^{-1/2}V^T$ is the whitening matrix. It is easy to show that after this processing, $\tilde{\mathbf{I}}$'s elements are uncorrelated, i.e.,

$$E(\tilde{\mathbf{I}}\tilde{\mathbf{I}}^T)=\mathbf{1}\,, \tag{19}$$

where $\mathbf{1}$ is the $w\times w$ identity matrix. The simulation is carried out in the space spanned by the column vectors of $\tilde{\mathbf{I}}$. To bring the results into the original image space, one can perform a dewhitening process as follows:

$$\mathbf{I}=VD^{1/2}\tilde{\mathbf{I}}\,, \tag{20}$$

where $VD^{1/2}$ is the dewhitening matrix.

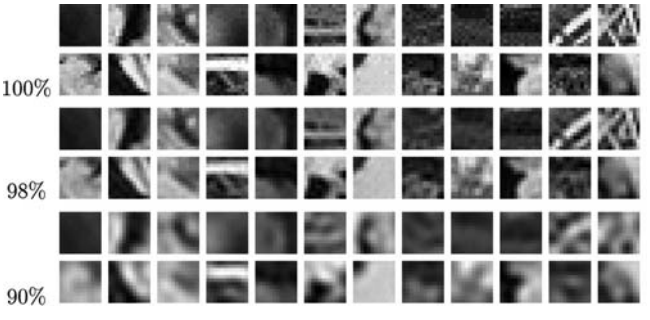

Fig. 4. Some examples of the original image for 100% and processed images for 98% and 90%

It is easy to show that the mean of each element of $\tilde{\mathbf{I}}$ is 0. Furthermore, in whitened input image space, for any unit vector $\boldsymbol{\psi}$, the dot product $\boldsymbol{\psi}\tilde{\mathbf{I}}$ has zero mean and unitary variance, i.e.,

$$E(\boldsymbol{\psi}\tilde{\mathbf{I}}) = 0, \quad E((\boldsymbol{\psi}\tilde{\mathbf{I}})^2) = 1\,. \tag{21}$$

Any orthogonal vectors $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}_j$ result in uncorrelated receptive field outputs in the whitened image space $\tilde{\mathbf{I}}$, i.e.,

$$E((\boldsymbol{\psi}_i\tilde{\mathbf{I}})(\boldsymbol{\psi}_j\tilde{\mathbf{I}})) = 0\,. \tag{22}$$

Therefore, by enforcing the orthonormality of the receptive fields $\boldsymbol{\psi}_j$ (where $1 \le j \le w$), the unitary variance and uncorrelated receptive field outputs are ensured. Thus we seek a set of receptive fields $\boldsymbol{\psi}_j$ $(1 \le j \le w)$ such that (a) they are orthonormal and (b) Eq. (14) is maximal over all sets of receptive fields that satisfy (a). By maximizing (14), we obtain a set of receptive fields such that the histograms of the output vectors $\boldsymbol{\psi}_j\tilde{\mathbf{I}}$ $(1 \le j \le w)$ are maximally non-Gaussian. Mathematically, combining (a) and (b) is equivalent to seeking the maximal solution of the following equation:

$$C(\boldsymbol{\Psi}) = \sum_{i=1}^{m}\sum_{j=1}^{J} E^2(P_j(R_i)) - \lambda\|\boldsymbol{\Psi}\boldsymbol{\Psi}^T - \mathbf{1}\|^2\,, \tag{23}$$

where $\|\cdot\|$ is the square norm of a matrix, λ is a free parameter, and $\mathbf{1}$ is the identity matrix.

The orthonormal constraint is artificial. It is mainly used to ensure that the generated receptive fields are different. It certainly can be replaced by other more reasonable constraints. Nonetheless, the alternative constraint still awaits further research. Moreover, it is worth pointing out that there is *only one free parameter* λ used in our final objective function (which corresponds to this orthonormality constraint). The larger this free parameter, the closer the receptive fields are to orthonormal.

There are many methods for solving the optimization problem (Press et al. 1988). It has been found that the conjugate gradient method fits the problem very well. The simulation results yielded by this method using 6000 12×12 images with six different whitening procedures are presented as follows.

5 Simulation results

The first simulation is carried out with all eigenvectors retained. Figures 5 and 6 are the results of the simulation. This is the convergent result, although it may take 1 month on a Linux machine or a supercomputer. Note that the receptive fields are dewhitened and the full whitening used here boosts the noise in the images. It looks like optimization is more dependent on the noise than on the signal. These receptive fields do not look like the receptive fields in the V1, which are localized, oriented, and bandpass. However, interestingly, nearly all (except two) receptive fields produce responses with a sparse distribution. The mean of the kurtoses is 3. This clearly indicates

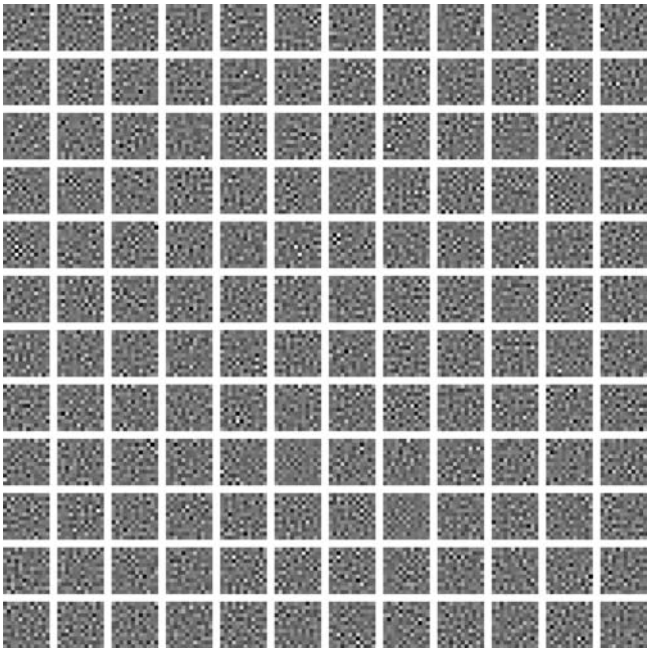

Fig. 5. The simulated 144 receptive fields

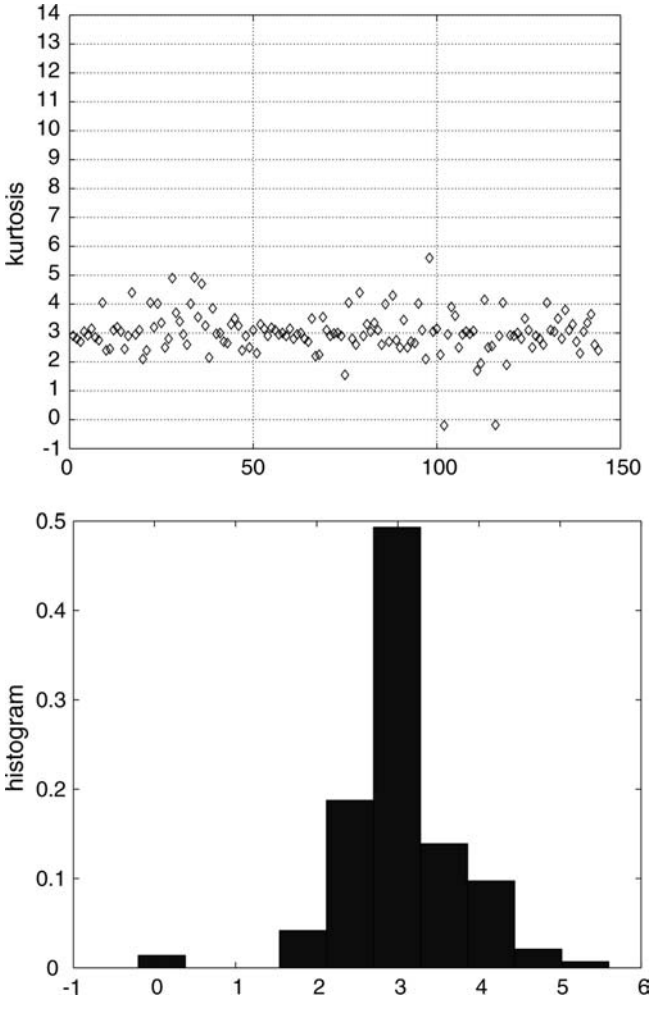

Fig. 6. Kurtoses of responses of the simulated 144 receptive fields to the input images; the histogram of these kurtoses

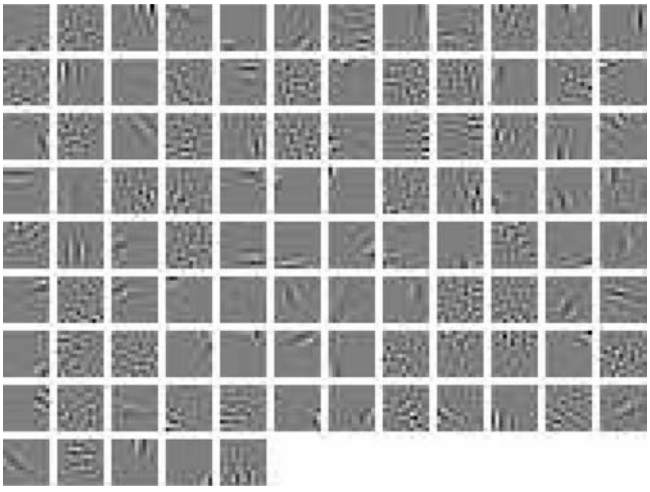

Fig. 7. The simulated 101 receptive fields

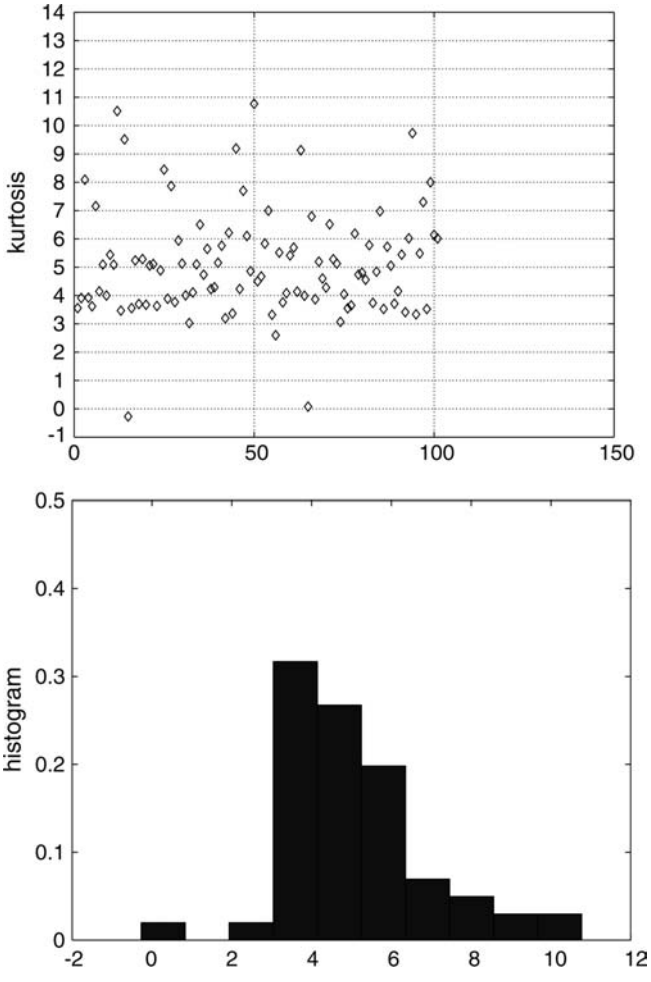

Fig. 8. Kurtoses of the responses of the simulated 101 receptive fields to the input images; the histogram of these kurtoses

that, compared with the goal of explicitly seeking sparse and distributed coding, maximized non-Gaussianity as an alternative goal can also produce sparse and distributed coding even though we did not explicitly seek sparse and distributed coding.

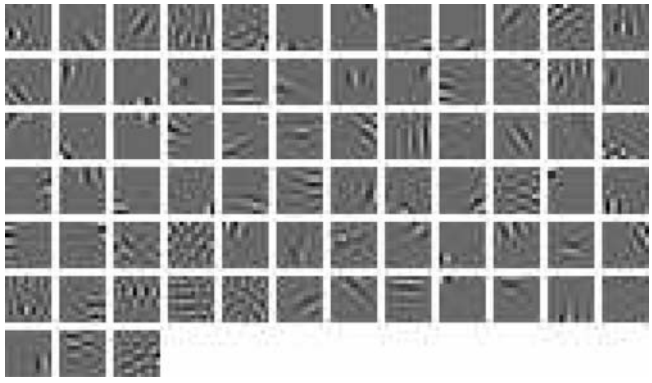

Fig. 9. The simulated 75 receptive fields

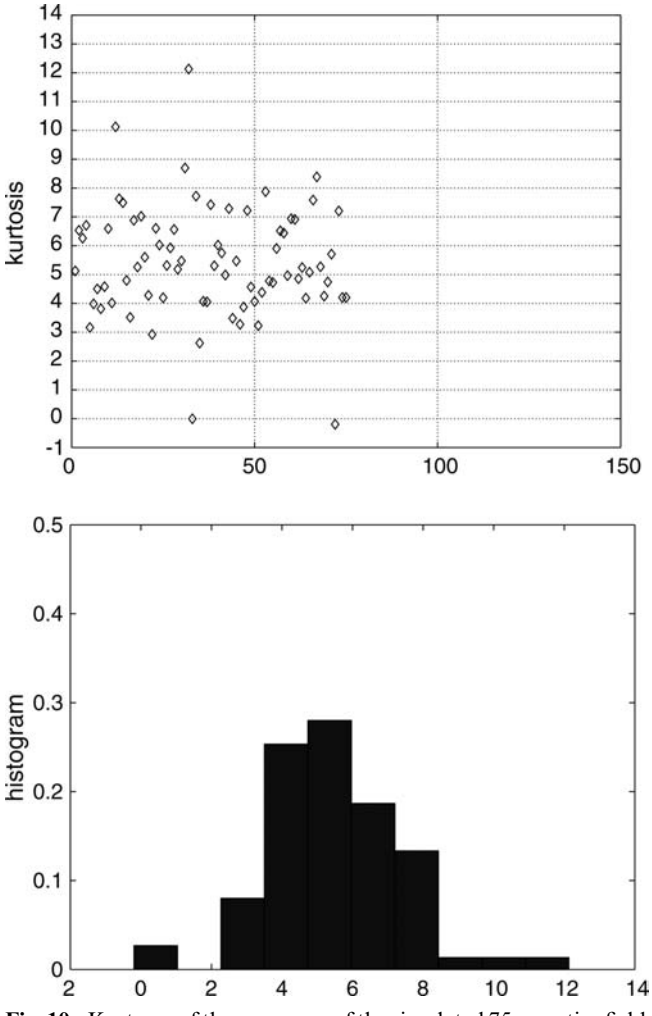

Fig. 10. Kurtoses of the responses of the simulated 75 receptive fields to the input images; the histogram of these kurtoses

In the following five simulations, 101, 75, 55, 44, and 22 eigenvectors are reserved to form the whitening matrix. This corresponds, respectively, to 99%, 98%, 96.5%, 95%, and 90% of the sum of all eigenvalues. The resulting receptive fields and the corresponding kurtosis responses are shown in Figs. 7–16. Interestingly, by removing just 1% of the total energy, many Gabor-like receptive fields appear, although some receptive fields still do not have

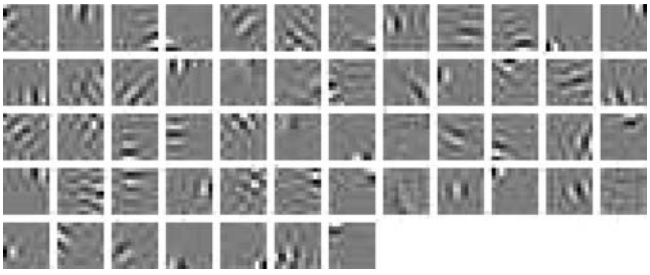

Fig. 11. The simulated 55 receptive fields

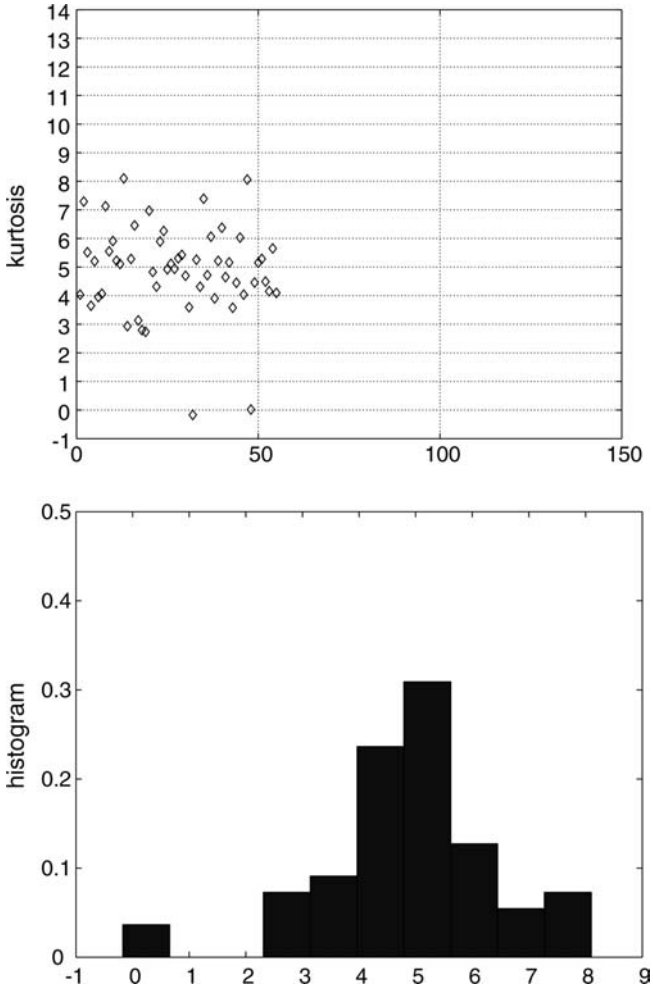

Fig. 12. Kurtoses of the responses of the simulated 55 receptive fields to the input images; the histogram of these kurtoses

any obvious structure. When the percentage of energy decreases, more and more Gabor-like receptive fields appear.

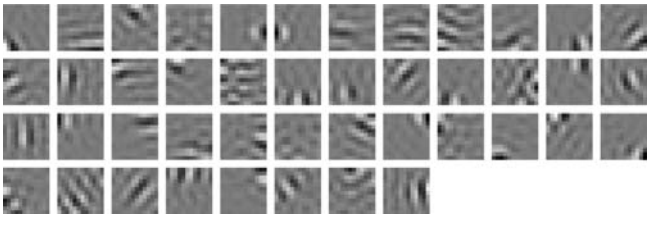

Fig. 13. The simulated 44 receptive fields

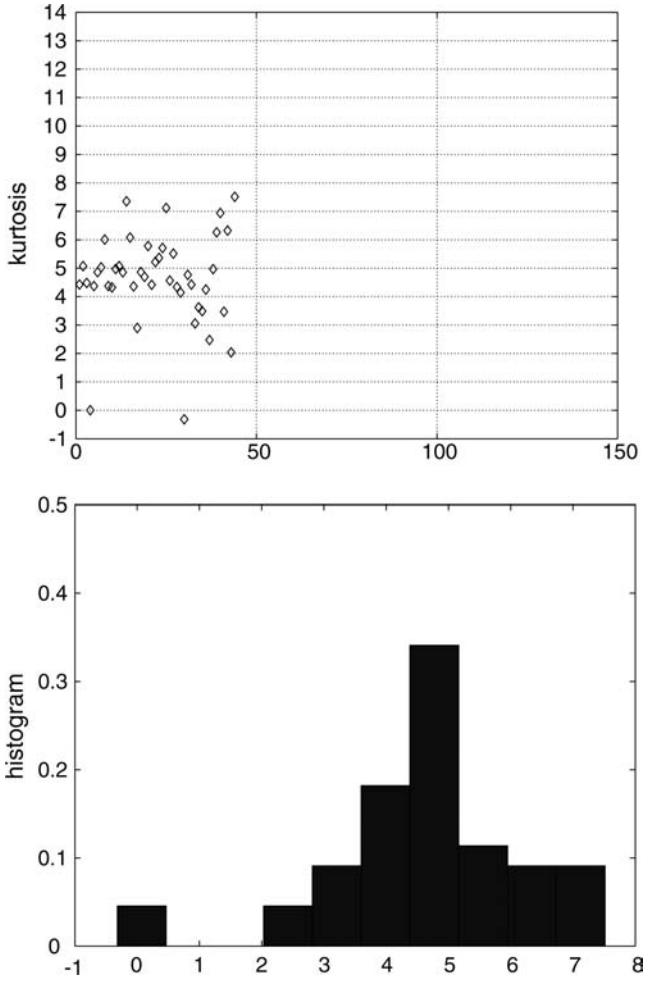

Fig. 14. Kurtoses of the responses of the simulated 44 receptive fields to the input images; the histogram of these kurtoses

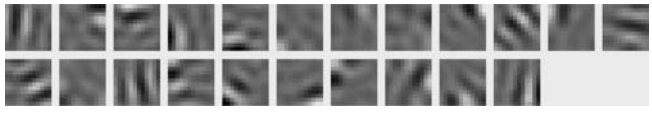

Fig. 15. The simulated 22 receptive fields

Nonetheless, even if the total energy is decreased, some receptive fields still do not have any evident structure. This may be due to the fact that we applied an artificial orthonormal constraint whose main purpose was to ensure that the generated receptive fields were different.

There is a clear trend in the receptive field results: As the band limitations increase in strength, a clear majority of Gabor-like receptive fields appear. The reason for this is that as band limitations increase, the high-frequency parts in the input images are removed. This makes the algorithm easy to converge. Moreover, there are fewer high-frequency parts in the converged results.

It should be noted that other algorithms that have been used to seek sparse and distributed codes for natural images do not require orthonormality of receptive fields. One might well ask what would happen if we relaxed the requirement that receptive fields be orthonormal. Specifically, it seems likely that we might obtain a reduction in the

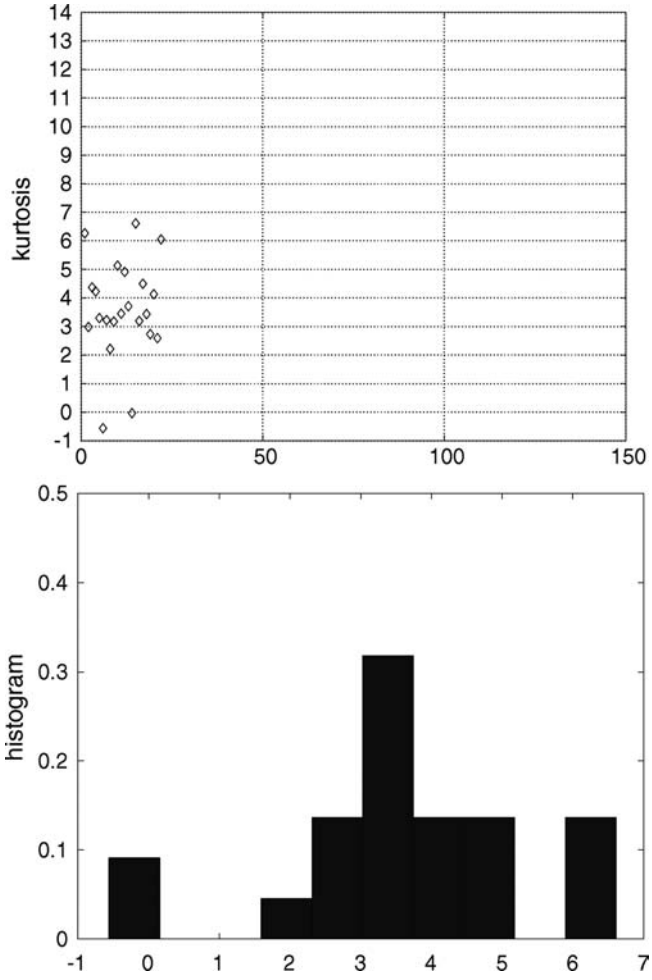

Fig. 16. Kurtoses of the responses of the simulated 22 receptive fields to the input images; the histogram of these kurtoses

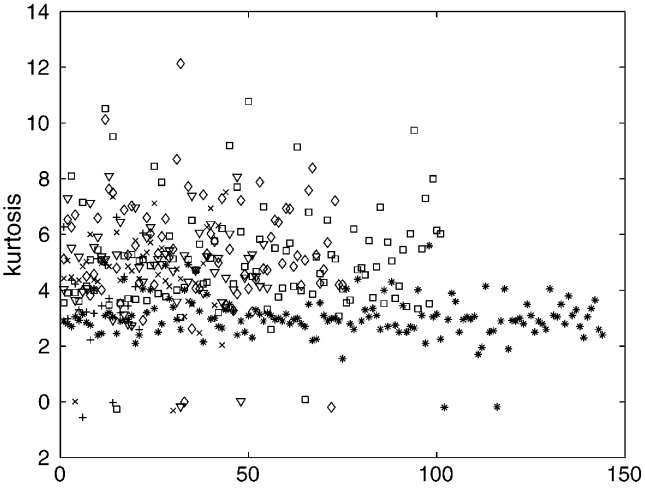

Fig. 17. Comparsion of the kurtoses from simulations of 144, 101, 75, 55, 44, and 22 receptive fields. "*": 144 RFs, square: 101 RFs, diamond: 75 RFs, triangle: 55 RFs, "x": 44 RFs, "+": 22 RFs

Table 1. Mean, median, and variance of the kurtoses

Number of RF	Mean	Median	Variance
144	3.0208	2.9700	0.7374
101	5.0951	4.8595	1.8664
75	5.4329	5.2603	1.9198
55	4.8849	5.1066	1.5669
44	4.6252	4.7298	1.5935
22	3.6191	3.4420	1.7390

number of receptive fields showing no obvious structure while increasing the number of Gabor-like receptive fields since a group of Gabor-like receptive fields may simply not satisfy the orthonormal requirement, e.g., observing the result of Olshausen and Field (1996, 1997). Furthermore, Gabor functions cannot form orthonormal bases.

The kurtoses yielded by these receptive fields are also shown in Figs. 8, 10, 12, 14, and 16. (A comparison of these kurtoses for different numbers of receptive fields is shown in Fig. 17.) Although the objective function (23) treats any departure from normality equally without explicitly seeking high kurtotic responses, nearly all these receptive fields (except two in every figure) have highly kurtotic response histograms. The means, medians, and standard variances of these kurtoses are summarized in Table 1 for comparison.

6 Discussion

This paper tries to answer the question: Why do simple cells in V1 possess Gabor-like receptive fields? Much research effort has gone into addressing this problem. Here we discuss some of this related research. One popular related method is independent component analysis. However, although this method may be a good model for audio perception where sounds can be transparently mixed, in vision, this may not be true since visual composing procedures such as occlusion are nonlinear. Therefore, the application of independent component analysis to vision research may be limited. One of the best results in this kind of research is by van Hateren and colleagues (van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998). They used a fast fixed-point method proposed by Hyvärinen and Oja (Hyvärinen and Oja 1997) to find the extreme (maximum or minimum) solution of the kurtosis K – because if the simple cells really use kurtosis as the measure, the receptive fields give the maximum or minimum of K. The receptive fields found in this research are localized, oriented, and bandpass. One interesting finding is that only by maximizing kurtosis can one produce simple cell-like receptive fields, while by minimizing kurtosis one cannot. It should be pointed out that, although the authors discuss their work under the framework of independent component analysis, the method they applied is technically equivalent to a method using kurtosis as the measure of the departure from normality. Bell and Sejnowski (1996) also proposed an independent component analysis algorithm. However, as was pointed out by Olshausen and Field, this method is technically equivalent to the method by Olshausen and Field (cf. discussion and appendix of Olshausen and Field 1997). However, as was pointed out by one reviewer, we do not have a general

argument against statistical independence. We have only pointed out the limitations of several implementations of statistical independence. Whether statistical independence can be the ultimate goal of biological vision systems is beyond the scope of this paper.

Another related study was by Chubb et al. (Chubb et al. 1997). They propose that the simple-cell receptive fields in V1 are collectively optimized to reject the null hypothesis that the visual input is devoid of spatial structure. Although their method is more general than those explicitly seeking sparse and distributed code method, it still can only catch some special cases of departure from normality. The reason is that the functional goal of their method only includes some special correlate functions that can only correlate with some special departures from normality; although these departures can be in places other than the center and the tails, this method still gives these places special positions like that the sparse and distributed coding method gives special positions to the center and the tails.

One interesting question is, Comparing our method with other methods, can we consider the deviation from normality the most favorable goal of biological vision system? With the evidence we have thus far, we are not sure this is true. We can only assert that the goal proposed in this paper is a possible alternative for simple cells. Therefore, the question is still an open one.

7 Conclusion

In this paper, we try to give an alternative to the sparse and distributed coding goal. We argue that the coding for simple cells is to find any departures from normality equally. Under this coding strategy, we found that most receptive fields generated are Gabor-like and are localized, oriented, and bandpass. The responses of these receptive fields are highly kurtotic regardless of the receptive fields are Gabor-like or non-Gabor-like. Thus, in seeking maximally non-Gaussian response histograms, receptive fields spontaneously yield highly kurtotic histograms. Hence sparse and distributed could be only a byproduct of a more general coding strategy. However, some structureless receptive fields always exist. We suspect that the receptive fields without structure are due to the artificial orthonormal constraint. We thus conclude that the high kurtosis observed in the response histograms of simple-cell receptive fields to natural images may reflect a property of natural images themselves, rather than an explicit coding goal used to structure simple-cell receptive fields.

References

Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? Network 2:213–251

Bell AJ, Sejnowski TJ (1996) Edges are the "independent components" of natural scenes. Adv Neural Inf Process Sys 9:831–837

Chubb CF, Lu ZL, Sperling G (1997) Structure detection: a statistically certified unsupervised learning procedure. Vis Res 37:3343–3365

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4(12):2379–2394

Field DJ (1994) What is the goal of sensory coding? Neural Comput 6:559–601

Friedman JH (1987) Exploratory projection pursuit. J Am Stat Assoc 82:249–266

Hyvärinen A, Oja E (1997) A fast fixed-Point algorithm for independent component analysis. Neural Comput 9(7):1483–1492

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? Vis Res 37:3311–3325

Press WH, Flannery BP, Tueukolsky SA, Vetterling WT (1988) Numerical recipes in C. Cambridge University Press, Cambridge, UK

van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequence yields spatiotemporal filters similar to simple cells in primary visual cortex. Proc R Soc Lond B 265:2315–2320

van Hateren JH, van der Schaaf A (1998) Independent component analysis of natural image compared with simple cells in primary visual cortex. Proc R Soc Lond B 265:359–366