

What the statistics of natural images tell us about visual coding.

David J. Field

Physiological Laboratory, University of Cambridge
Cambridge England, CB2 3EG

ABSTRACT

Why does the mammalian visual system represent information as it does? If we assume that visual systems have evolved to cope with the natural environment then we might expect the coding properties of the visual system to be related to the statistical structure of our environment. Indeed, images of the natural environment do not have random statistics. The first-order statistics (e.g., distribution of pixel values) and second-order statistics (e.g., power spectra) of natural images have been discussed previously and they bear important relations to visual coding. Statistics higher than second-order are difficult to measure but provide crucial information about the image. For example, it can be shown that the lines and edges found in natural images are a function of these higher-order statistics. In this paper, these higher-order statistics will be discussed in relation to the coding properties of the mammalian visual system. It is suggested that the spatial parameters of the cortical 'filters' (e.g., bandwidths of simple and complex cells) are closely related to these higher-order statistics. In particular, it will be shown that the spatial non-linearities shown by cortical complex cells provide the early visual system with the information required to learn about these statistics.

1. INTRODUCTION

The precise function of cells in the mammalian visual cortex have puzzled researchers since they were first mapped out by Hubel and Wiesel¹ in 1962. Theories as to why these cells behave as they do have varied from edge and bar detection to Fourier analysis. More recently, it has been suggested that insight into the behaviour of cells in the mammalian visual pathway may be gained by a better understanding of the statistics of the visual environment^{2,3}. This line of approach will be continued in this paper. We will consider a particular property of natural images, and how this property relates to the behaviour of cortical cells. In particular, we wish to emphasize a type of cell described as 'complex'^{1,4}. Although such cells are suggested to represent as much as half of primary visual cortex⁵, no widely accepted theory of their function is available. Complex cells show a particular type of spatial nonlinearity which distinguishes them from the more linear and more widely studied 'simple' cells.

There are two main points to this paper. First, natural images show a particular form of redundancy which does not show up in measures of second-order statistics. This redundancy results in local features like edges and lines. In the frequency domain, an edge or bar can be described as a local correlation in phase: at the location of the edge, the Fourier components in the same phase. However, when the image contains a large number of edges and lines at different positions, the correlation between global Fourier coefficients is lost. There may still be redundancy, but it is now in the form of a higher-order statistic. It might loosely be described as a redundancy across different scales of the image; the information present in different frequency bands is not independent.

In the second part of this paper, a visual model will be described which represents images with 'filters' that share many properties of cortical receptive fields. The most important property that will be investigated will be the spatial non-linearity shown by cortical complex cells. By manipulating the bandwidths in this model, we will attempt to get a better idea of why cortical cells have their particular parameters and how these parameters relate to the statistics described above.

1.1 The statistics of natural images

Randomly selected images from the natural world do not have random statistics. Such images do not fit a description of randomly generated pixels. Many of the regularities of natural images are best described in terms of "nth-order" statistics. For example, first-order statistics refer to the probability distributions of the values of each of the pixels (i.e., the luminances at different points). If we assume that all pixels in an image represent samples from the same distribution, then it is possible to pool the pixels from a single image to get an estimate of first order statistics.

Second order statistics refer to the relations between pairs of pixels. The autocorrelation function, for example, is a measure of the correlations between pixels as a function of the distance between pixels. The power spectrum is also a measure of second order statistics (it is the Fourier transform of the autocorrelation function) and therefore provides an alternative measure. First and second-order statistics provide important information about an image and provide important clues about visual coding. Second-order statistics of natural scenes show rather consistent behaviour³ (i.e., the power spectrum falls off as a function of frequency (f) by a factor of roughly $1/f^2$). And such statistics are crucial to understanding visual coding.



Figure 1. Two images with the same power spectrum.

However, there are other consistent properties of natural images other than that provided by first and second-order statistics. Figure 1a shows a natural scene (Professor F.W. Campbell). Figure 1b shows an image with the same power spectrum. This figure was created by randomizing the phases of the Fourier coefficients of the original. Thus the two images have the same correlational structure (second-order statistics). However, these two images are clearly different. One of the important differences is the lack of any clear edges and lines in the image when the phases are randomized. The presence of these edges and lines cannot be measured in terms of second order statistics. To understand the differences between these two images, we must consider higher-order statistics.

1.2 Stationary statistics and local energy.

The general assumption that will be made in this section is that the statistics of natural scenes are 'stationary'. This means that if we consider a very large collection of images, the statistics at one position are no different than any other. For example, if vertical edges tended to occur only in the top half of all the images then the statistics would not be stationary. If the statistics of natural scenes are stationary then it can be shown that the amplitudes of all the real and imaginary coefficients (sines and cosines) will be uncorrelated. If all positions of an image have the same statistics then all phases of any particular frequency are equally likely. Thus, the correlation of any two frequencies (frequency n and m) can be determined by integrating the product of the amplitudes at each position. The correlation is proportional to the sum of the products, thus:

$$r = k \int_0^{2\pi} e^{inx} e^{imx} dx \quad (1)$$

$$= k \int_0^{2\pi} e^{i(n+m)x} dx \quad (2)$$

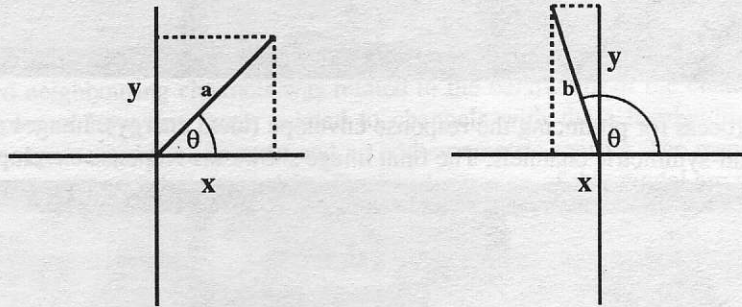
$$= k [e^{i(n+m)2\pi-1}] / i(n+m) \quad n \text{ and } m \text{ must be integers for the discrete Fourier transform, thus (3)}$$

$$e^{i(n+m)2\pi} = 1 \quad \text{and therefore: (4)}$$

$$r = 0 \quad (5)$$

Thus if all phases are equally likely then the sines and cosines of all the frequencies will be uncorrelated. However, the fact that the two frequencies are uncorrelated does not mean that the coefficients are independent. The problem can be reduced to two vectors. Consider the two vectors shown in Figure 2, representing the complex amplitude at two frequencies. Let us assume that the two vectors are perfectly correlated in length while their direction (θ) is random.

Figure 2. Two-dimensional vectors with the same length ($a=b$) but with uncorrelated projections



Under such conditions, the following projections will all be uncorrelated with one another:

$$x_1 = a \cdot \cos(\theta_1) \quad y_1 = a \cdot \sin(\theta_1) \quad x_2 = b \cdot \cos(\theta_2) \quad y_2 = b \cdot \sin(\theta_2) \quad (6)$$

In other words, the second-order statistics do not provide information that the two-dimensional vectors are redundant in any way. One might say that the information is provided by a fourth-order statistic since the fourth dimension can be predicted with knowledge of the other three. However, it is possible to convert this information into a second-order statistic by using a non-linear transformation. The transform into polar coordinates where $z = (x^2 + y^2)^{1/2}$ is the obvious choice. Under such a transformation, one would discover that z_1 and z_2 are correlated.

A similar process can be applied to two-dimensional spatial filters.^{3,6,7} Consider an image filtered through a linear filter selective to a limited band of frequencies and orientations. An example is shown in Figure 3. Figure 3a shows the original image. Figure 3b shows the result of the convolving the image with a linear filter which produces no phase shift in the spectrum (an even-symmetric filter). Figure 3c is the result of convolving the image with a similar filter that shifts the phase of each coefficient by 90° . These two filters are orthogonal and are described as 'in quadrature'. As with the vector above, the vector amplitude at each point in space can be computed as the vector sum of these two orthogonal vectors. The result is shown in Figure 3b. Figure 4 shows the same process when the filter involves a different band of frequencies. The main point is that the outputs of linear filters like those of Figure 3b, 3c, 4b and 4c will be uncorrelated independent of the statistics of the original image. However, the correlations between the vector amplitudes shown in Figures 3d and 4d (response envelope) will provide important information about the redundancy of the original image.

1.3 Complex cells

This similarities between this response envelope and behavior of "complex" cells have previously been noted^{3,7}. Like the functions shown above, such cells are sensitive to a limited band of orientations and frequencies in a local region of space. Within this region such cells show a spatial non-linearity which appears to be quite similar to the function described above. Furthermore, work of Pollen and Ronner⁸ has shown that adjacent simple cells in the cortex may differ by the 90° required to be quadrature pairs. It might therefore be possible to produce a cell showing many of the properties described above with some form of non-linear sum of adjacent simple cells. But this does not tell us why a visual system would require such cells. With such cells representing as much as half of primary visual cortex⁵, it seems likely that they play an important role in visual coding. In the following sections we will suggest one possible account.

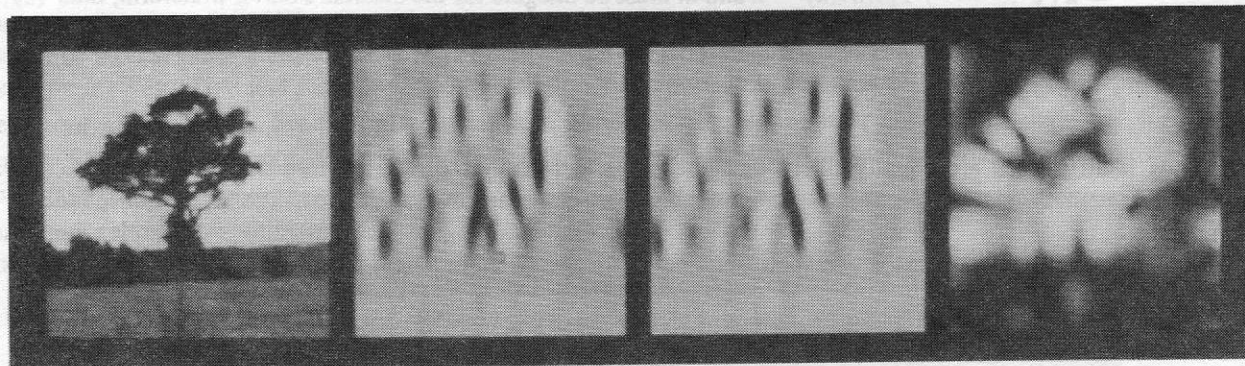


Figure 3. Process for producing the response envelope (local energy). Images series shows the original image followed by even and odd-symmetric channels. The final image shows the response envelope.

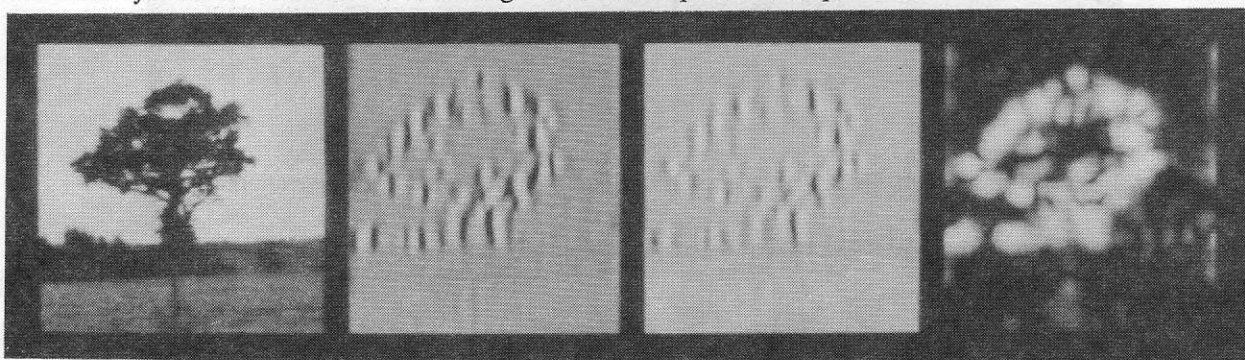


Figure 4. Same as Figure 3 but the filter is selective to a higher band of frequencies.

2. THE VISUAL MODEL

In the next section we will be comparing the correlations between model neurons. To do this, we use a code which attempts to model many of the properties of the visual cortex. However, it is by no means a complete model of the visual cortex. Rather, it is a method of representing information in an image which will allow us to explore various parameters of the visual system as well as to indirectly test various properties of images. The general form of the model has been developed by a number of previous investigators^{9,10} and discussed in greater detail in a previous paper of the author³. The main features are outlined below. Two terms must be defined. A **sensor** refers to a single filter localized in space and representing a single hypothetical cell. The term **channel** refers to the spatial array of sensors tuned to the same frequencies and orientations.

- 1) In the frequency domain, the distance between neighbouring channels is determined by the spatial frequency bandwidth, which also determines the width of the individual sensors and hence the distance between neighbouring sensors in space.
- 2) In the frequency domain, the distance between neighbouring orientation channels is determined by the orientation bandwidth, which also determines the length of the individual sensors and hence the spacing along the length of the sensors.
- 3) Spatial frequency bandwidths are constant in octaves, and orientation bandwidths are constant in degrees for all the sensors of a given code. However, there is freedom to change these bandwidths for each code.
- 4) At each position, there are two orthogonal sensors with phase relations in quadrature. The local energy (i.e., the hypothetical complex cell) is represented as the vector sum of these two sensors.
- 5) The total number of sensors is independent of the particular choice of spatial frequency or orientation bandwidth (i.e., the image consisting of 65,536 pixels is represented by 65,536 sensors).

The particular function which will be used is similar to the Gabor representation but differs in that the frequency response is log normal. The merits of this function have been discussed previously. For a vertically oriented sensor like that shown in Figure 7, the response is described by the function:

$$G(f,O) = \exp(-(\ln(f/f_0)^2/2\sigma^2)) * \exp(-(O-O_0)^2/2\rho^2) \quad (7)$$

The sampling distance between different sensors is the same as the sampling distance between different channels. For all codes described in this paper, the sampling distance corresponded to a distance of:

$$D = \sigma * (2\pi)^{1/2} \quad (8)$$

The sampling distance between two neighbouring channels was related to the bandwidth of the channels by the same factor. Figure 5 provides three examples of how this factor related to channels with three bandwidths. At each position, two orthogonal phase relations were represented (i.e., an even and an odd-symmetric sensor). The response of the complex cell (local energy) was simply the vector sum of these two units. Further discussion of this model are provided in an earlier paper³.

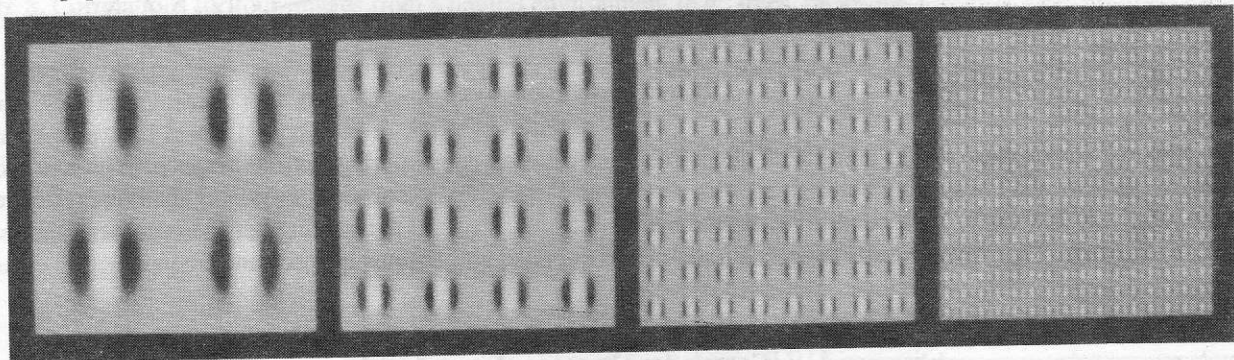


Figure 5. An example of the vertically oriented sensors (even-symmetric) in a code where the frequency bandwidth is 1 octave. For the sake of clarity, the sampling distance between the sensors was increased by a factor of roughly 2.

3 METHODS OF ESTIMATING CORRELATIONS

A model of this type allows us to determine correlations between neighbouring sensors in space as well as correlations between different channels (i.e., different frequency bands). For each bandwidth, the estimates of the correlations in frequency involved four steps:

- (1) The image was filtered by two channels (equation 1) separated by the distances described in equation 2 and demonstrated in Figures 3 and 4.
- (2) The complex amplitude was determined as described above.
- (3) Correlation between the two complex amplitudes was determined for all the central region (the edges of the image will produce spurious correlations).
- (4) Repeat steps (1) - (3) for all neighbouring channels.

The estimates of correlation in space uses a similar procedure:

- (1) The image was filtered by a single channel (equation 1).
- (2) The complex amplitude was determined as described above.
- (3) Correlation between the complex channel and a displaced version of the same channel was determined. The extent of the displacement was a function of the size of the sensors in that particular channel (i.e., the distance corresponds to distance to the next sensor in that channel - as in Figure 5).
- (4) Repeat steps (1) - (3) for all neighbouring channels.

4 RESULTS

Figure 6 shows the results for the correlation between neighbouring channels for the image in Figure 1. When the channels are narrow-band and the spacing between neighbouring channels is relatively small, the correlations between channels is large. As the bandwidth of the channels and the spacing between channels is increased, the correlation drops. This implies that for the image in Figure 1, the correlations extend across different frequencies. However, the correlation is not complete and the farther away in frequency the lower the correlation.

Figure 6. Correlations between neighbouring channels.

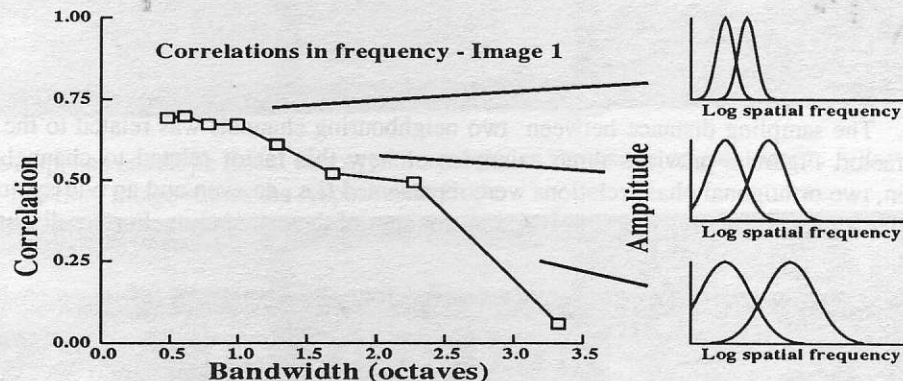


Figure 7 shows these results along with the correlations between neighbouring sensors in space. With narrow-band channels, the sensors and the spacing between sensors is relatively large. It may not seem surprising that the correlations in space are relatively low under such conditions. As the bandwidths increase and the spacing between sensors decrease, the correlations increase. Thus, these results suggest a trade-off between the correlations in space and frequency.

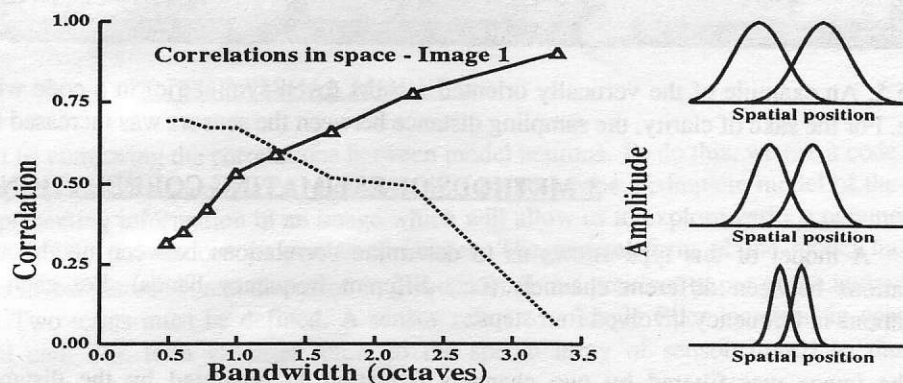


Figure 7. This figure shows the correlations between neighbouring sensors of the code. In all cases, the sampling distance was proportional to the sensor size which is inversely proportional to channel size.

Figure 8 shows the results for four other images. The trade-off seems to be a rather consistent feature. However, what is most interesting is that the region of the trade-off are roughly in the range of cortical cell bandwidths which cover the range of 1-2 octaves. The consistency in the results may lead one to question whether the tradeoff is simply a function of the code and not the images. It is certainly true that correlations in space and frequency must trade off to some extent but the point where they cross is indeed a function of the correlational structure of the images. To demonstrate this point, the image in Figure 1b was analysed. In that image, the phases of the Fourier coefficients were intentionally randomized. In this image one would not expect any high correlations between channels. The results are shown in Figure 9. In these data, one can see that the spatial correlations remain unchanged (because the power spectrum was unchanged), but the frequency correlations were removed (the residual correlation are due to the fact that the two functions overlap in frequency to some extent).

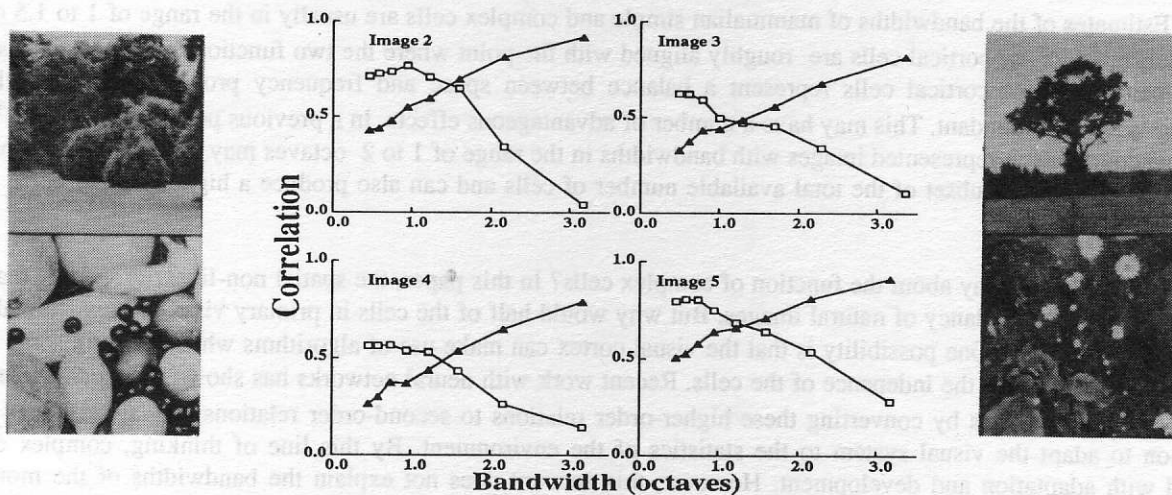
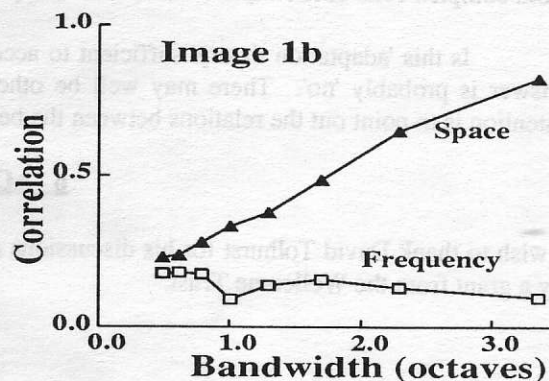
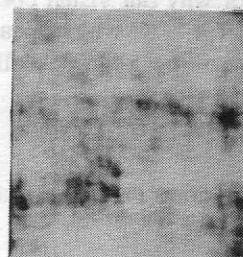


Figure 8. Correlations for four images from a natural environment (e.g., trees, rocks, etc.)

Figure 9. Correlations for the image shown in Figure 1b. This image has random phases which is reflected in the results which show no significant correlations in frequency.



5 DISCUSSION

As stated earlier, there are two parts to this paper. In the first section, two type of redundancy were discussed. The first is represented by correlations across space and is measured in terms of second order statistics like the auto-correlation function and power spectrum. The second type of redundancy is rarely considered in discussions of natural scenes. This redundancy is between different scales of the image (i.e., different frequency bands). This type of redundancy is not second-order and therefore does not show up in correlations. However, it was shown that a particular non-linear transform is capable of converting these higher-order statistics to second-order relations between sensors and can then be measured in terms of correlation. A model was described which allowed an image to be represented in terms of linear arrays of theoretical cells or 'sensors' that are selective to local bands of different frequencies and orientations. From these linear 'sensors', it was shown how one could obtain an estimate of the local energy (a measure which shows interesting similarities to the behaviour of cortical 'complex' cells).

In the second section, we investigated the relations between a model of visual coding and these two forms of redundancy. The results demonstrate that both of these two forms of redundancy are present in natural scenes. Different codes, appear to promote different amounts of these two types of redundancy. If one represents natural scenes with narrow bandwidth functions then there will be high redundancy between the neighbouring channels. If one represents natural scenes with broad bandwidth function then there will be high correlations in space. Any code which attempts to maximize the independence of the units of the code (i.e., the sensors or cells in the case of the visual system) must take into account both types of redundancy.

Estimates of the bandwidths of mammalian simple and complex cells are usually in the range of 1 to 1.5 octaves. Thus, the bandwidths of cortical cells are roughly aligned with the point where the two functions cross^{5,11}. This suggests that the bandwidths of cortical cells represent a balance between space and frequency producing a state where no neighbour is highly redundant. This may have a number of advantageous effects. In a previous paper by the author³, it was shown that codes which represented images with bandwidths in the range of 1 to 2 octaves may allow natural scenes to be represented with only a subset of the total available number of cells and can also produce a high signal to noise level in those cells.

What does this say about the function of complex cells? In this paper, the spatial non-linearity was used as a tool to demonstrate the redundancy of natural images. But why would half of the cells in primary visual cortex show this kind of spatial non-linearity? One possibility is that the visual cortex can make use of algorithms which perform some form of de-correlation to increase the independence of the cells. Recent work with neural networks has shown that such decorrelation is easily possible¹². Thus by converting these higher-order relations to second-order relations, visual cells may use de-correlation to adapt the visual system to the statistics of the environment. By this line of thinking, complex cells are involved with adaptation and development. However, this account does not explain the bandwidths of the more linear simple cells. Without the spatial non-linearity, only spatial correlations would be apparent. Thus, with narrow-band channels correlations would appear to be minimized. To explain the bandwidths of simple cells, a model involving feedback from complex cells onto simple cells would be required.

Is this 'adaptation theory' sufficient to account for why our visual cortex contains so many complex cells? The answer is probably 'no'. There may well be other uses for cells which measure local energy. In this paper, the only intention is to point out the relations between the behaviour of these cortical cells and the statistics of the environment.

6 ACKNOWLEDGEMENTS

I wish to thank David Tolhurst for his discussions and comments on earlier drafts of this paper. This work was supported by a grant from the Wellcome Trust.

7 REFERENCES

1. D. G. Hubel and T.N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *J. Physiol.* 160, 106-154 (1962).
2. S.B. Laughlin, "Matching coding scenes to enhance efficiency" in *Physical and Biological Processing of Images*, O.J. Braddick and A.C. Sleight, eds. (Springer, Berlin, 1983), pp. 42-72.
3. D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4 2379-2394 (1987).
4. A.F. Dean and D.J. Tolhurst. On the distinctness of simple and complex cells in the visual cortex of the cat. *J. Physiol.*, London 344, 305-325 (1983).
5. D.J. Tolhurst, and I.D. Thompson. On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proc. Roy. Soc. London B* 213, 183-189 (1982);
6. E.H. Adelson, and J.B. Bergen, Spatio-temporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2 284-299. (1985).
7. M.C. Morrone and D.C. Burr, "Feature Detection in human vision: A phase-dependent energy model". *Proc. Roy. Soc London Ser B* 235, 221-245 (1988);
8. D.A. Pollen and S.F. Ronner, "Phase relationships between adjacent simple cells in the cat," *Science* 212, 1409-1411 (1981).
9. J.G. Daugman. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters" *J. Opt. Soc. Am. A* 2, 1160-1169 (1985).
10. A.B. Watson. Efficiency of a model human image code *J. Opt. Soc. Am. A* 4 -2378 (1987).
11. R.L. DeValois, D.G. Albrecht, and L.G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, 22, 545-559 (1982).
12. Barlow, H.B. and Foldiak, P.Z., *Adaptation and de-correlation in the cortex*, "To appear in the computing Neuron (Miall, Durbin and Mitchison, eds.) Addison Wesley.