Relations between the statistics of natural images and the response properties of cortical cells

David J. Field

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

Received May 15, 1987; accepted August 14, 1987

The relative efficiency of any particular image-coding scheme should be defined only in relation to the class of images that the code is likely to encounter. To understand the representation of images by the mammalian visual system, it might therefore be useful to consider the statistics of images from the natural environment (i.e., images with trees, rocks, bushes, etc). In this study, various coding schemes are compared in relation to how they represent the information in such natural images. The coefficients of such codes are represented by arrays of mechanisms that respond to local regions of space, spatial frequency, and orientation (Gabor-like transforms). For many classes of image, such codes will not be an efficient means of representing information. However, the results obtained with six natural images suggest that the orientation and the spatial-frequency tuning of mammalian simple cells are well suited for coding the information in such images if the goal of the code is to convert higher-order redundancy (e.g., correlation between the intensities of neighboring pixels) into first-order redundancy (i.e., the response distribution of the coefficients). Such coding produces a relatively high signal-to-noise ratio and permits information to be transmitted with only a subset of the total number of cells. These results support Barlow's theory that the goal of natural vision is to represent the information in the natural environment with minimal redundancy.

INTRODUCTION

Since Hubel and Weisel's[1] classic experiments on neurons in the visual cortex, we have moved a great deal closer to an understanding of the behavior and connections of visual cortical neurons. A number of recent models of early visual processing have been quite effective in accounting for a wide range of physiological and psychophysical observations.[2-4]

However, although we know much more about *how* the early stages of the visual system process information, there is still a great deal of disagreement about the reasons *why* the visual system works as it does. Theories of why cortical neurons behave as they do have varied widely from Fourier analysis[5,6] to edge detection.[7] However, no general theory has emerged as a clear favorite. Edge detection has proved to be an effective means of coding many types of images, but the evidence that cortical neurons can generally be classified as edge detectors is lacking (e.g., Refs. 8 and 9).

The notion that the visual cortex performs a global Fourier transform is no longer given serious consideration. The relatively broad spatial-frequency bandwidths and local spatial properties of cortical neurons make them unsuitable for extracting Fourier coefficients. But that leaves the question of what the visual system achieves with frequency-selective mechanisms if it does not perform a Fourier analysis. In a number of recent papers[4,9-13] an alternative to the strict Fourier approach was proposed. This new approach is based on Gabor's[14] theory of communication. Gabor showed how to represent time-varying signals in terms of functions that are localized in both time and frequency (the functions in time are represented by the product of a Gaussian and a sinusoid). These functions, now referred to as Gabor functions, have been used to describe the behavior of cortical simple cells that extend in both space and spatial frequency.

However, the approach provides no further insight into why cortical neurons might behave in this way. Daugman[12] points out that the Gabor code represents an effective means of filling the information space with functions that extend in both space and frequency. However, this does not necessarily imply that such a code must be an efficient means of representing the information in any image. As we shall see, the efficiency of a code will depend on the statistics of the input (i.e., the images). For a wide variety of images, a Gabor code will be quite an inefficient means of representing information.

Clearly, the definition of an efficient, or optimal, code depends on two parameters: the goal of the code and the statistics of the input.

With few exceptions (e.g., Refs. 15–18), theories of why visual neurons behave as they do have failed to give serious consideration to the properties of the natural environment. Our present theories about the function of cortical neurons are based primarily on the response of such neurons to stimuli such as checkerboards, sine-wave gratings, long straight edges, and random dot patterns.

Gibson[19] always stressed that one must understand the nature of the environment before one can understand the nature of visual processing. However, his comments have gone largely unheeded in the mainstream of vision research. There seems to be a belief that images from the natural environment vary so widely from scene to scene that a general description would be impossible. Thus an analysis of such images is presumed to give little insight into visual function.

The main thrust of this paper is that images from the natural environment should not be presumed to be random patterns. Such images show a number of consistent statistical properties. In this paper we suggest that a knowledge of these statistics can lead to a better understanding of why the mammalian visual system codes information as it does.

0740-3232/87/122379-16$02.00

Barlow[20–23] has stressed the need to understand visual processing in terms of the redundancy of visual images. Barlow suggests that the purpose of natural image processing is "to represent visual scenes by activity of a sparse selection of reliable and nonredundant (i.e., independent) elements." (See Ref. 23, p. 12.) In the following sections we shall consider ways in which an efficient visual system might take advantage of this redundancy.

In this paper a model will be presented that codes images into arrays of band-limited functions representing the response properties of cortical simple cells. However, this paper is not about some particular detail that makes the model unique. Indeed, the general theme of the model is much the same as that of the models of Sakitt and Barlow,[24] Watson,[4] and Daugman.[12] Instead, this paper is about why such models are effective in coding the information in natural images. Indeed, there are many ways of coding the information in an image. The image-processing literature is filled with codes that serve a wide variety of purposes.[25] In this paper we shall try to show why the response properties of cortical cells might provide an effective means of representing spatial information in the mammalian visual world.

The purpose of this paper is not to provide a proof of any form. That would be impossible with the small sample of images used in this study. Rather, the purpose here is to show that it is indeed possible to relate the behavior of cortical cells to the statistics of the natural environment.

THE MODEL

This model shares a number of features with several recently developed models.[4,11,12] The important feature of this particular model is that it provides the freedom to vary the parameters of the components (e.g., the spatial-frequency bandwidth or the orientation bandwidth of the theoretical cells) without losing information or adding free parameters.

The foundation of the model is derived from principles discussed by Gabor[14] and the principles of information theory.[26] Some of the basic ideas are represented in the information diagram shown in Fig. 1.

Consider a simple one-dimensional pattern consisting of 18 equally spaced pixels. In such an image, all the information is represented by the amplitudes of the 18 pixels. The information represented by such a pixel code can be represented by the information diagram shown in Fig. 1(a). A pixel is a function that is localized in space but extends in frequency (i.e., broadband). This property is represented by each rectangle in the diagram.

It is also possible to represent the information in this simple image by means of a discrete Fourier transform. Again, all the information is represented by 18 independent coefficients (i.e., the amplitudes of 9 sine functions and 9 cosine functions). Each of these functions is localized in frequency but extends in space [Fig. 1(b)].

Gabor's theory implies that one can also represent the information in terms of the amplitudes of functions that are localized in both space and frequency. In Figures 1(c) and 1(d) such coefficients are represented by rectangles, where the area of each rectangle is constant. Hence, if the function is large in space, then it is small in frequency and vice versa.

Such functions need not all be the same size in space or frequency. Figure 1(d) shows a common representation in which the low-frequency functions are relatively large in space and small in frequency. Again, the number of functions (i.e., independent coefficients) is constant. The functions that Gabor proposed (now called Gabor functions) have a number of mathematically elegant properties, which have been discussed elsewhere.[10–12] However, it should be noted that an array of such functions will not be orthogonal and that the transform based on such functions will not be reversible in the same way as the Fourier transform. Nonetheless each function is selective to a different region of the space–frequency diagram, and with appropriate spacing the functions may be considered quasi-orthogonal.

For the sake of clarity we will use the following terminology to describe the Gabor coefficients.[4] Each individual function will be referred to as a sensor. A spatial array of sensors tuned to the same spatial frequency will be referred to as a channel. For example, in Fig. 1(d) each of the individual rectangles is a sensor. A row of rectangles is a channel.

Therefore a narrow-band channel consists of a small number of spatially large sensors, and a broadband channel consists of a large number of spatially restricted sensors. In the next section we consider the more general case of two-dimensional frequency and space-limited functions; a channel

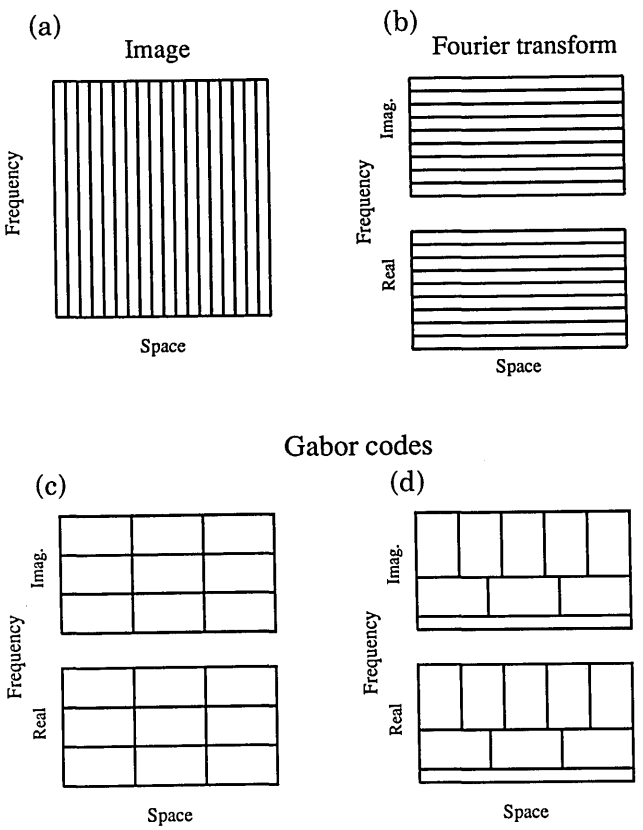


Fig. 1. Information diagrams, as proposed by Gabor.[14] These diagrams represent the information carried by a one-dimensional array of elements. For example, (a) represents the information carried by 18 pixels. Each pixel is localized in space but extends in frequency. The Fourier transform (b) represents the same amount of information with 18 sines and cosines. Each element is localized in frequency but extends in space. (c) and (d) Show Gabor codes, in which the information is represented by elements that are localized in both space and frequency. The area represented by any element is constant, and the total number of elements is also constant.

will then be defined as an array of sensors tuned to the same spatial frequency and orientation.

Finally, we use the term code to refer to the entire collection of channels (and therefore sensors) required to represent the image. All the codes to be compared in this paper have roughly the same total number of sensors. This total will equal the number of pixels in the original image.

Two-Dimensional Patterns

For a two-dimensional image, the trade-off in space and frequency is a bit more complicated than in the one-dimensional case. Such two-dimensional codes have been discussed elsewhere.[4,12] However, a few points should be made about two-dimensional sampling and the trade-off between space and spatial frequency. A two-dimensional, vertically oriented Gabor function (i.e., a sensor) is defined by the following equation:

$$g(x, y) = \exp\{-[x^2/2(\Delta W)^2 + y^2/2(\Delta L)^2]\}\cos(2\pi f x + \theta), \quad (1)$$

where ΔL denotes the spatial size of the sensor along its length (i.e., the axis parallel to the preferred orientation) and ΔW represents the width of the sensor in space (orthogonal to the preferred orientation). As noted, the sampling distance in space (i.e., spacing between sensors) and the frequency (i.e., between channels) is proportional to the size of the function in space and frequency. In the two-dimensional case, the length of the sensor determines the spacing in the length direction. Furthermore, the orientation bandwidth of the function is inversely proportional to length, which in turn determines the spacing in the frequency domain. In other words, the orientation bandwidth determines the following parameters of the code: the distance (in the frequency domain) between neighboring orientation channels, the number of orientation channels, the length of the sensor, and the separation between neighboring sensors in the length direction. Thus, if the code consists of channels that are narrowly tuned to orientation, then it must sample at a relatively larger number of orientations. However, since the sensors of such a channel are relatively long, the sampling in space will be coarse.

In a similar manner, the width of the sensor (ΔW) determines the spacing along the width. The width also determines the spatial-frequency bandwidth and hence the number of channels required at different spatial frequencies. These points are also developed in greater detail in Figs. 2 and 3.

Relative Bandwidths

Another feature of this model is that, on a linear scale, the bandwidths of the different sensors are proportional to their optimal frequencies. The overall design may be best described in terms of a rosettelike pattern as shown in Fig. 3. This pattern shows how the two-dimensional frequency plane is divided. The center of this rosette represents a frequency of zero ($f = 0$). Moving out from the center represents an increase in frequency. The preferred orientation of the channel is represented by the angle from horizontal. As in Fig. 2, a single channel is represented by a pair of ellipses that are diagonally opposed. In both of the codes represented in Fig. 3, the orientation bandwidth and the spatial-frequency bandwidth of the channel increase as a function of the frequency to which the channel is tuned. This produces a code in which the orientation bandwidth is constant in degrees and the spatial-frequency bandwidth is constant in octaves.

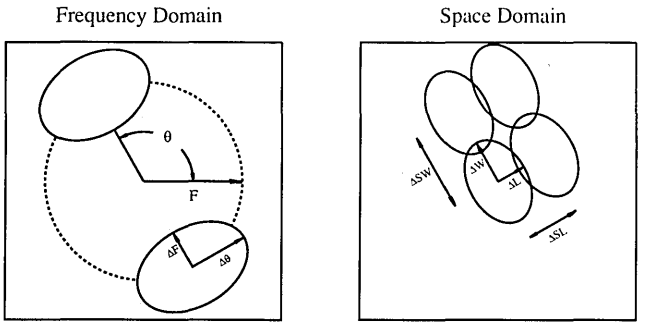


Fig. 2. Relations between the size of the channel in the frequency domain and the size and spacing in the space domain. A channel with a bandwidth ΔF (frequency domain) consists of an array of sensors with a width ΔW (space domain). The spacing along the width ΔSW will be proportional to the width ΔW. A channel with an orientation bandwidth $\Delta\theta$ (frequency domain) consists of any array of sensors with a length ΔL (space domain). The spacing along this length (ΔSL) will be proportional to the length.

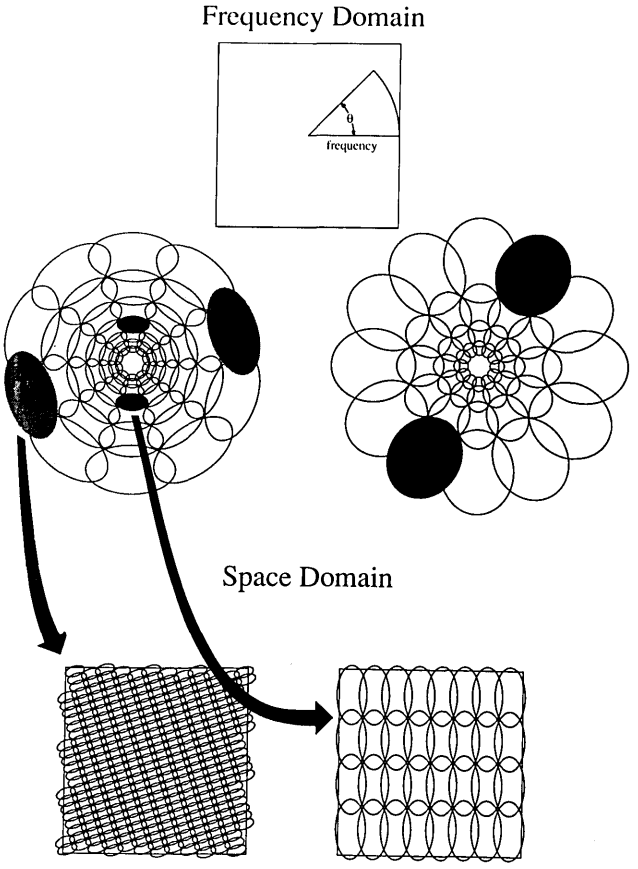


Fig. 3. Relations between the spacing of the channels in the frequency domain and the spacing of the sensors in the space domain. The upper two rosettes represent two possible coding schemes. Both codes consist of channels that have constant spatial-frequency bandwidths in octaves and constant orientation tuning in degrees. The rosette on the left consists of channels that are more broadly tuned to orientation and more narrowly tuned to frequency than those of the rosette on the right. The bottom two diagrams give a rough idea of the relative spacing of the sensors for two of the channels.

The diagrams at the bottom of Fig. 3 give a rough idea of the spacing of the individual sensors within each channel. Sensors tuned to low frequencies are large in space, and therefore sampling in space is relatively coarse. Sensors tuned to high frequencies are relatively small in space, and therefore sampling is relatively fine. Each ellipse in space represents the relative size of a particular sensor. Again, the bandwidth of the channel is inversely related to the size of the sensors associated with the channel:

$$\Delta F = k_1/\Delta W = k_2/\Delta SW, \tag{2}$$

$$\Delta\theta = k_1/\Delta L = k_2/\Delta SL. \tag{3}$$

The number of sensors associated with a particular channel ($N_{s/c}$) is therefore proportional to the size of the channel:

$$N_{s/c} = 1/(\Delta SW\Delta SL) = (\Delta F\Delta\theta)/k. \tag{4}$$

Also, since the number of channels N_c required to cover the two-dimensional frequency domain is inversely proportional to the bandwidths of the channels (i.e., $\Delta F\Delta\theta$), the total number of sensors N_{st} will be constant:

$$N_{st} = N_c N_{s/c} = \alpha(\Delta F\Delta\theta)/(\Delta F\Delta\theta) = \text{constant}. \tag{5}$$

The precise details of the sampling scheme are not discussed here, since they are not critical to the conclusions of this paper. However, it must be reemphasized that the total number of sensors is constant and is equal to the number of free parameters in the input (i.e., the number of pixels).

Phase

Finally, we must consider the relative phase of the sensors. In this version of the model, we will assume that at each sensor location there exist two orthogonal phase-selective sensors (i.e., in quadrature). Pollen and Ronner[27] have shown that adjacent simple cells demonstrate such a property. As has been noted,[8,12] their results do not suggest that the receptive field profiles must necessarily be even and odd symmetric [$\theta = 0, \pi/2, \pi, 3\pi/2$ in Eq. (1)]. Indeed, the evidence supports a wide range of symmetries.[8] However, as long as the two phase relations are in quadrature (i.e., they differ by 90 deg), it is not critical what phases are involved.

The information provided by such a pair of cells is best described in terms of a two-dimensional vector whose length provides information about the contrast energy at any given point and whose direction determines the phase of that energy. For example, consider the response of a particular pair of such sensors to a sine-wave grating. The relative outputs of the two orthogonal sensors will depend on the positions of the pair of sensors relative to the grating. If we consider the vector sum of the outputs of the two sensors, then the direction of the vector will oscillate as we move across the grating, indicating a change in phase, but the length of the vector will remain roughly constant. The magnitude of this constant response will depend on the relationship between the spatial frequency of the grating and the frequency response of the channel.

This description will prove important for our considerations of response variability discussed in later sections.

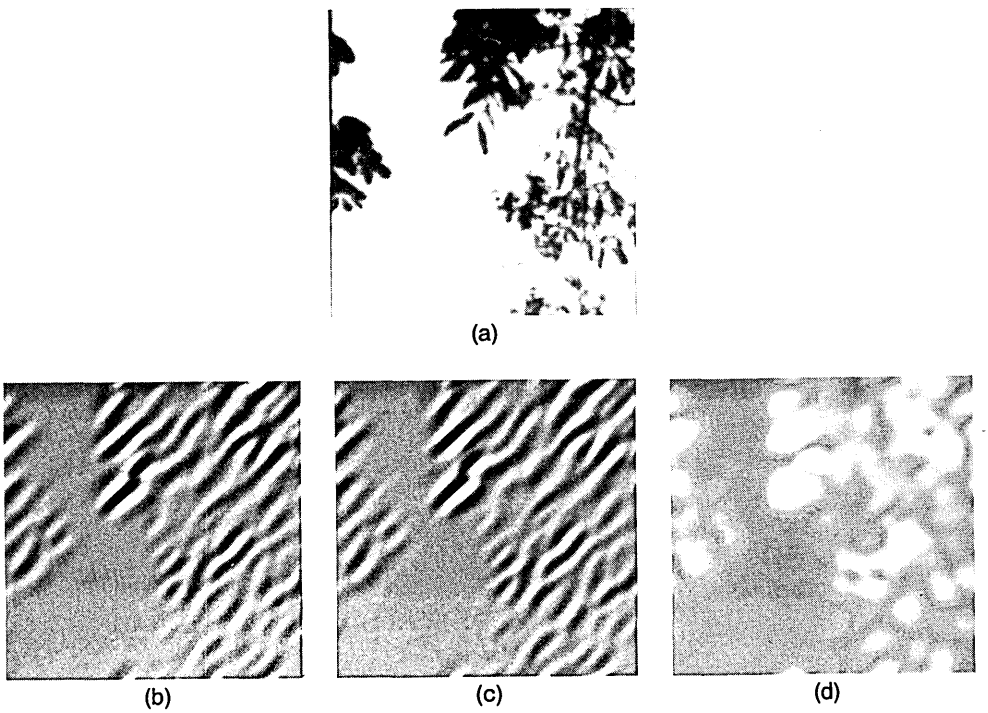


Fig. 4. The envelope of the response of one channel. Some of the calculations in this paper are based on the envelope of the response of the sensors. This envelope is sometimes referred to as the amplitude of the analytic signal[28] or the energy-density waveform.[29] (b) The response of an even symmetric channel to the image in (a). (c) The response of an odd symmetric channel tuned to the same frequencies and orientations. (d) The envelope, representing the root mean square [Eq. (6)] of the images in (b) and (c). The envelope takes into account both the real energy and the reactive energy of the image[28] and thus permits a measure of the energy at a given location within a particular orientation and frequency band, independent of the phase of the individual sensors.

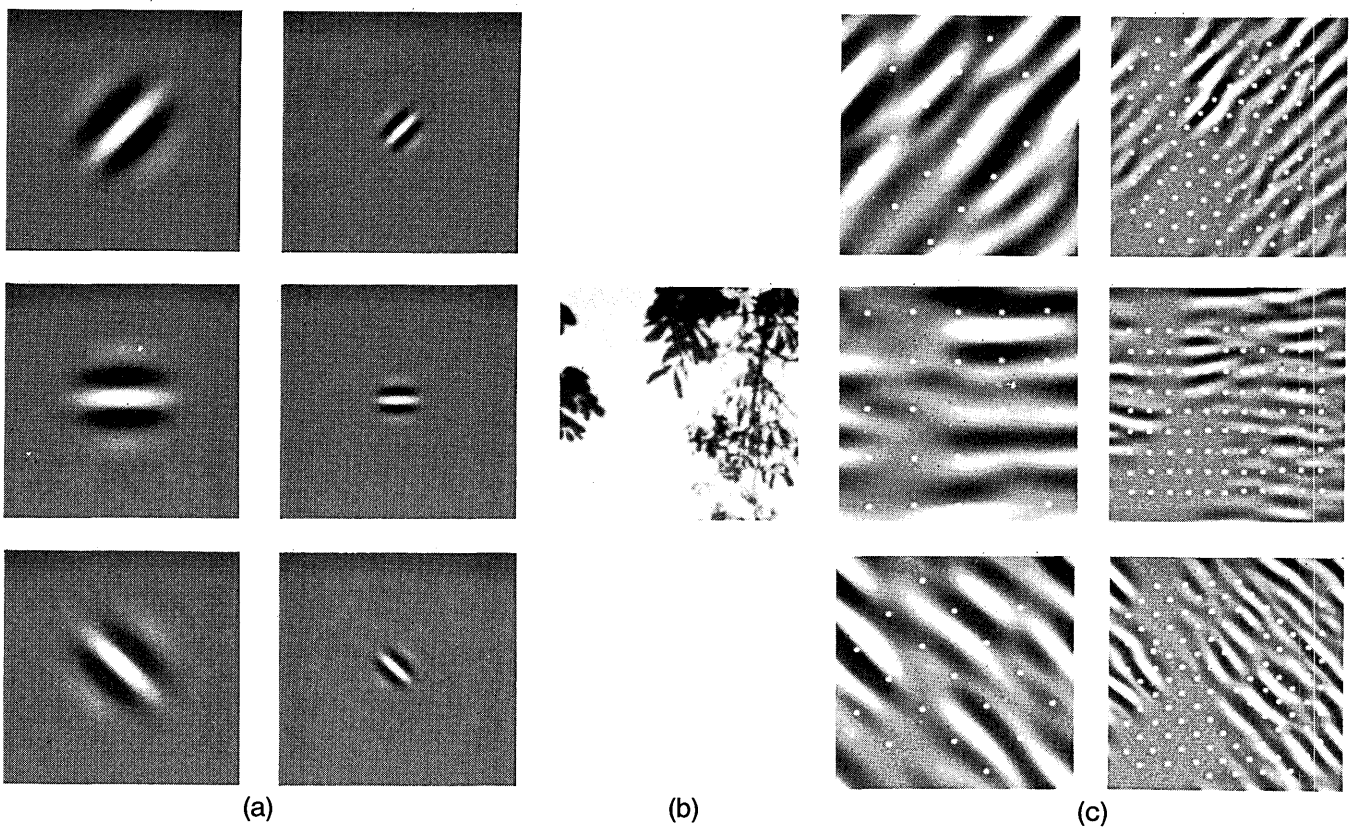


Fig. 5. An example of coding with six different channels. (a) Examples of the six types of sensor associated with each channel. (c) Convolution of the image in (b) with the six sensors shown in (a). The response of the individual sensors is determined by sampling these filtered images at a distance proportional to the size of the sensor (shown with dots). This diagram shows the response of only the even symmetric sensors. A complete code would involve the odd symmetric responses as well as the full distribution of orientations and spatial frequencies.

Our method of measuring the vector amplitude is shown in Fig. 4. Figure 4 shows an example of a particular image [Fig. 4(a)] and the image when filtered through an orientation- and frequency-selective channel with even [Fig. 4(b)] and odd [Fig. 4(c)] symmetry. Figure 4(d) shows the envelope $E(x, y)$ of these two orthogonal images, where

$$E(x, y) = [R(x, y)^2 + I(x, y)^2]^{1/2}. \quad (6)$$

Such an image shows how the length of the vector (the energy envelope) varies as a function of the position of the pair of sensors. Just as the amplitude spectrum shows the response as a function of frequency that is independent of the phase, the envelope E shows the response as a function of position that is independent of the phase. Such a function is not unique to this study. It is generally described as the amplitude of the analytic signal[28] or as the square root of the energy-density waveform.[29] (The function also shows interesting similarities to the behavior of cortical complex cells.)

In Section 3 we shall compare the variability of the responses of different types of channels to a particular image. However, the periodicity introduced by bandpass filtering an image can create a spurious contribution to this variability. Consider the filtered images in Figs. 4(b) and 4(c) resulting from filtering with the even and the odd symmetric channels. The periodicity in the filtered image is due to a rotation of the response vector (i.e., a change in the local phase) rather than to a change in the magnitude of the vector. The envelope provides a measure of the channel's response that is independent of this local phase change.

Overview

The main points of the model are described below and illustrated in Fig. 5:

(1) In the frequency domain, the distance between neighboring frequency channels is determined by the spatial-frequency bandwidth, which also determines the width of the individual sensors and hence the spacing along the width of the sensors

(2) In the frequency domain, the distance between neighboring 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, but there is freedom to choose the absolute magnitudes of these bandwidths.

(4) At each position there are two orthogonal sensors with phase relations in quadrature. A response envelope can be determined from each pair of sensors.

(5) The total number of sensors is independent of the particular choice of spatial-frequency or orientation bandwidth (i.e., an image consisting of 65,536 pixels is represented by 65,536 sensors).

In the following sections we look at the advantages of using such codes to represent the various types of images.

IMAGE ANALYSIS

Methods

The six scenes used in this study were photographed with a Keystone 3572 camera (35 mm) using XP1 Kodak monochrome film. The scenes were taken from various places around England and Greece. No attempt was made to select particular types of scene, but images were chosen that had no artificial objects (buildings, roads, etc.). Although it was hoped that these scenes were typical natural scenes, no effort was made to ensure this, and they may therefore represent biased samples.

The negatives were digitized on a laser densitometer (Joyce Loebel) into 256 × 256 pixels with a depth of 8 bits/pixel (256 density levels). The images were analyzed on a Sun Workstation computer using conventional software developed by the author.

Calibration

The modulation transfer function (MTF) of the optical system (lens and developing process) was determined from the response of the system to a point source. A photograph of a point source was taken with the same camera and film, and the negative was developed in the same manner as the six natural scenes. The results described below were corrected in accordance with this MTF.

IMAGE ANALYSIS: AMPLITUDE SPECTRA OF NATURAL IMAGES

In this section we discuss a particular property of natural images as illustrated by their amplitude or power spectra. This topic is discussed in greater detail in another paper. However, since the conclusions of this section play an important part in the next section, it is discussed briefly here.

Natural images, on the whole, appear to be rather complex. They are filled with objects and shadows and various surfaces containing various patterns at a wide range of orientations. Amid this complexity, it may seem surprising that such images share any consistent statistical features. Consider the six images shown in Fig. 6. Such images may seem widely different, but as a group they can be easily distinguished from a variety of other classes of image. For example, random-dot patterns are statistically different from all six of these natural images. This difference is best described in terms of the amplitude spectra or power spectra of the images, where the amplitude spectrum is defined as the square root of the power spectrum.

The two-dimensional amplitude spectra for two of the six images are shown in Fig. 7. The spectra of these images are quite characteristic and are quite different from that of white noise, which is by definition flat. They show greatest amplitude at low frequencies (i.e., at the center of the plot) and decreasing amplitude as the frequency increases. The

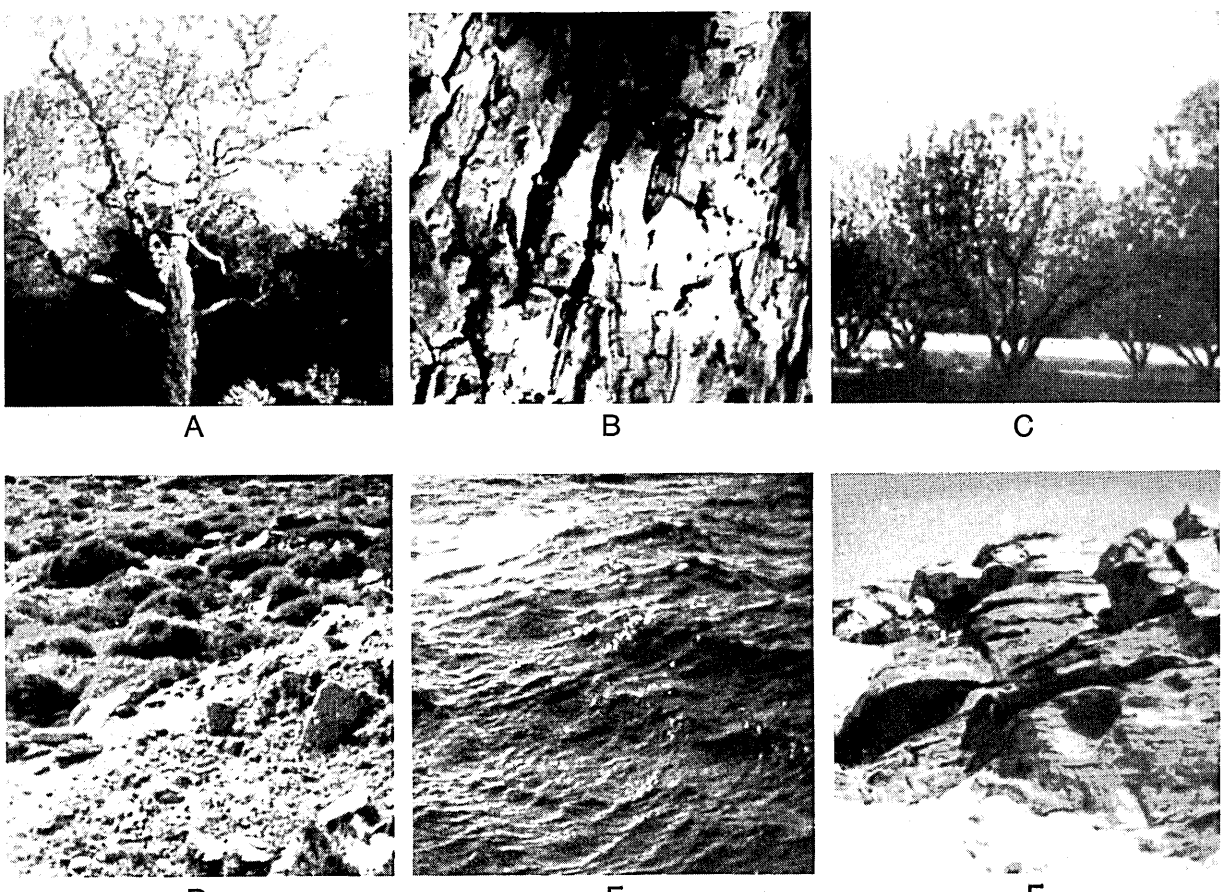


Fig. 6. Examples of the six images (A–F) in this study. Each image consists of 256 × 256 pixels with 256 gray levels (8 bits). However, only the central region was directly analyzed (160 × 160). See the text or details.

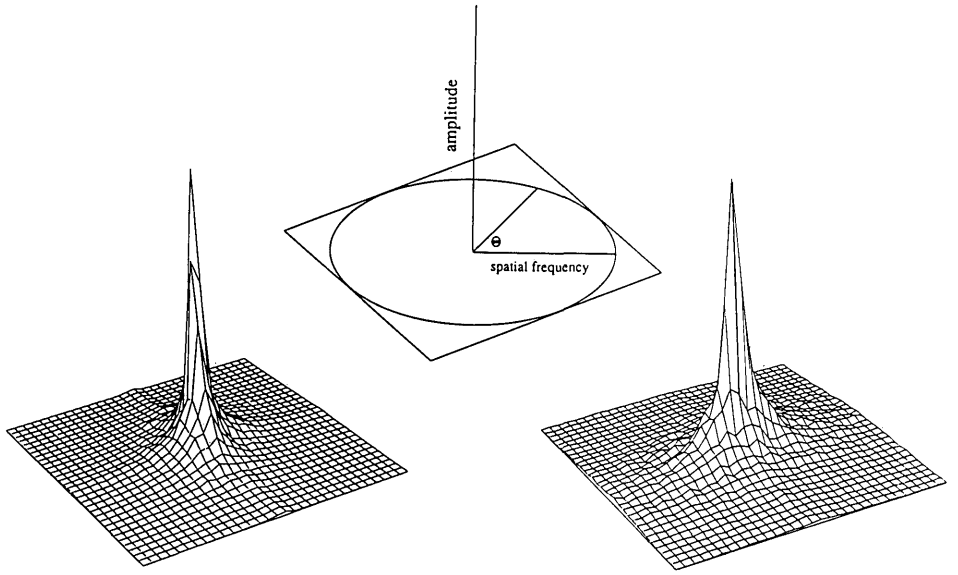


Fig. 7. Two-dimensional amplitude spectra for two images from Fig. 6 (A and D). The center of such a plot represents 0 spatial frequency. Frequency increases as a function of the distance from the center, and orientation is represented by the angle from the horizontal. For the sake of the clarity, each 256 × 256 amplitude spectrum has been reduced to 32 × 32. Thus each point in this plot represents an average of an 8 × 8 region of the spectrum. Such plots show that amplitude decreases sharply with increasing frequency at all orientations.

amplitude falls off quickly by a factor of roughly $1/f$ (i.e., the power falls at $1/f^2$). Figure 8 shows the amplitude spectra averaged across all orientations and plotted on log–log coordinates.

Although the description is by no means perfect, these amplitude spectra are all roughly described by a slope of -1. This is not to say that all scenes from the natural world would be expected to show a $1/f$ falloff; there are certainly scenes that do not show this property (i.e., a field of grass, the night sky, etc.). However, there are several reasons why this $1/f$ falloff in amplitude should be expected to be a rough average.

A $1/f$ falloff in the amplitude spectrum is what we would expect if the relative contrast energy of the image were scale invariant (i.e., independent of viewing distance). For example, consider an image of a surface with an amount of energy E between frequency f and frequency nf when viewed at a distance d. Increasing the distance by a factor a will shift the energy to the frequency range of af and anf. If we let the energy at any frequency equal

$$E(f) = g(f) * (2\pi f), \tag{7}$$

then to keep the energy constant in the range of all f to nf for all f requires

$$\int_{f_0}^{nf_0} g(f) * (2\pi f) df = K, \tag{8}$$

and it follows that

$$g(f) = k/f^2. \tag{9}$$

In other words, if the power falls off as $1/f^2$, there will be equal energy in equal octaves. For example, the total energy between 2 and 4 cycles/deg will equal the energy between 4 and 8 cycles/deg. (On a two-dimensional plot the area covered by an octave band is proportional to f^2.) This falloff in power can also be related to the fractal nature of the lumi-

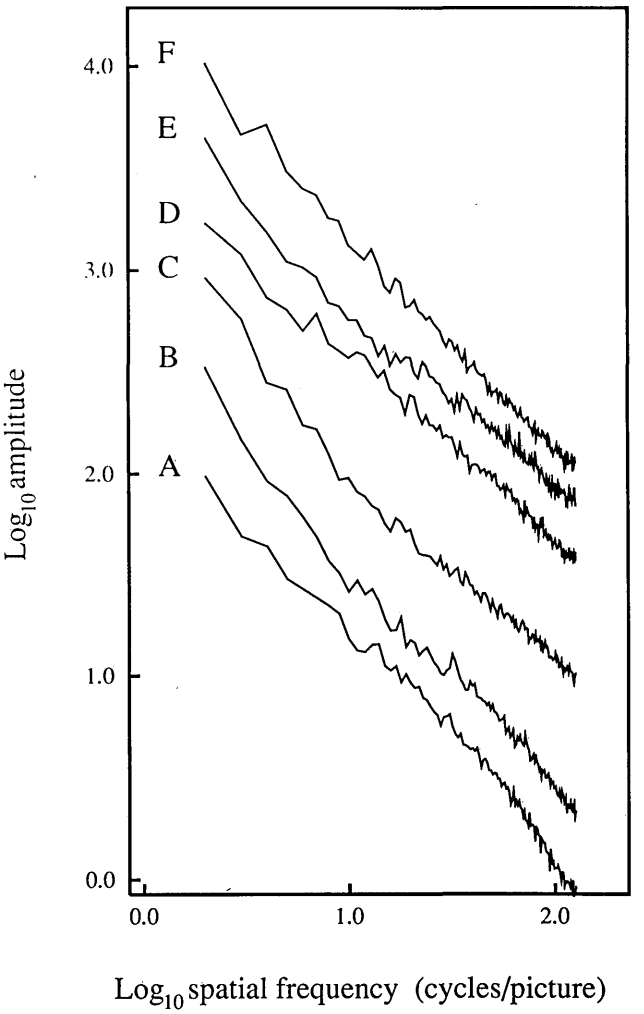


Fig. 8. Amplitude spectra for the six images A–F, averaged across all orientations. The spectra have been shifted up for clarity. On these log–log coordinates the spectra fall off by a factor of roughly $1/f$ (a slope of -1). Therefore the power spectra fall off as $1/f^2$.

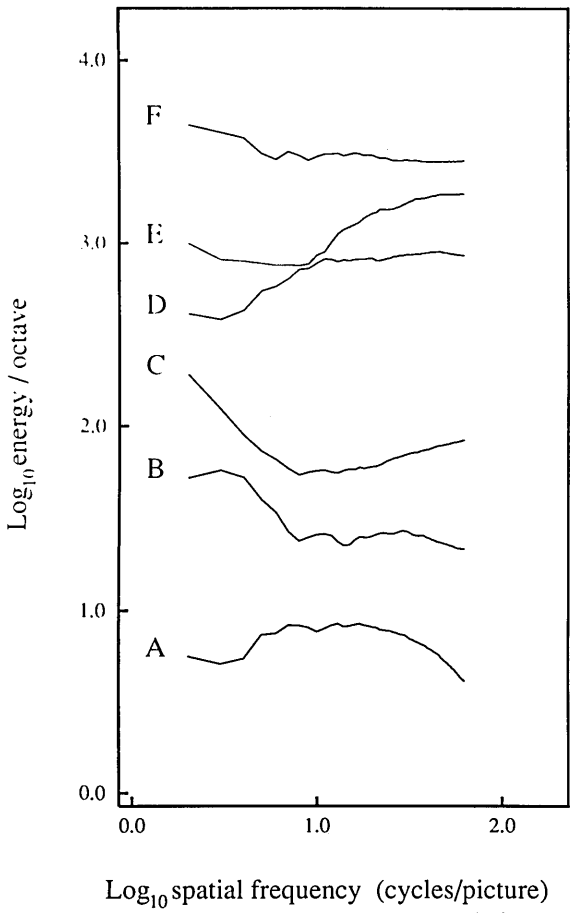


Fig. 9. Plot of energy per octave for each of the six images, for which energy is defined as the integral of the power spectrum between f and $2f$. For example, the data at 10 cycles/picture represents the total energy between 10 and 20 cycles/picture across all orientations. Although the results vary from image to image, these data suggest that, in contrast to the amplitude spectra, there is roughly equal energy in any given octave. By Parseval's formula [Eq. (10)] this implies that the variance of the filtered image (filtered through a given octave) will be roughly constant. See the text for details.

nance profiles of the images, where the $1/f^2$ falloff gives a fractal dimension of 2.5 (see Ref. 30).

Figure 9 shows the total energy in each 1-octave-wide band as a function of the spatial frquency of the low end of the octave. These data suggest that, despite the variability from image to image, no particular band contains consistently more energy.

In terms of our model, this falloff in power proves quite important. Since, the channels have constant bandwidths in octaves, as shown in Fig. 5, an image with a $1/f^2$ falloff in power will give roughly equal energy in each of the different channels. Furthermore, by Parseval's theorem[28] the variance of the image is equal to the integral of the power spectrum, or, in the discrete case,

$$\frac{\sum_{x=0}^{n-1}|g(x)|^2}{n} = \sum_{f=0}^{n-1}|G(f)|^2, \tag{10}$$

where $|G(f)|^2$ represents the power spectrum.

Since the different frequency-selective channels in a given code have constant bandwidths (in octaves), the outputs of the different channels will have roughly the same variance. If the statistics of natural images are also stationary (i.e., the statistics at any location in the visual field are no different from those at any other location), then all the different sensors will statistically have the same probability distribution and therefore carry equivalent amounts of information.

Hence, with a $1/f^2$ power spectrum and an array of channels with constant octave bandwidths and constant orientation tuning (in degrees), the information provided by each type of sensor will be roughly equal. Therefore the rosette-like codes described in Fig. 3 permit the information in the image to be distributed evenly across the array of sensors.

To obtain some intuition of why this is so, consider that the profiles of sensors with the same bandwidth in octaves are simply scaled versions of one another (i.e., they have the same number of ripples, etc.). Images with $1/f$ amplitude spectra are also invariant in terms of their power at different scales; that is, the amount of energy in any particular octave band is independent of the scale on which the image is viewed. Hence coding a scale-invariant image into an array of scale-invariant sensors produces an even distribution of the information.

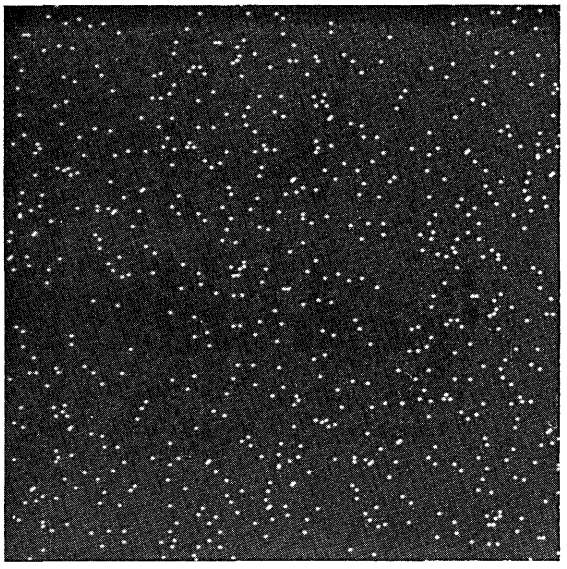
An even distribution may not seem to be an efficient means of representing information in an image. Indeed, a Karhounen–Loeve transform,[25] which is often described as an optimally efficient code, will result in an uneven distribution for images such as these. This is a point that we shall discuss below.

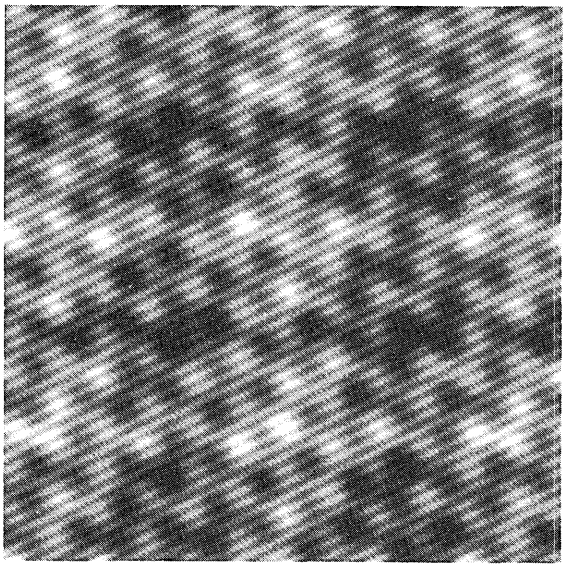
However, it is clear that having an even distribution is not by itself a sufficient condition for an efficient code. Indeed, the information is evenly distributed with a pixel code if the image statistics are stationary. A code using channels with constant bandwidths in octaves will permit the different sensors to provide roughly equal information. However, the efficiency of the code will depend on other parameters of these channels. A constant bandwidth does not restrict us to any particular bandwidth (e.g., 1 versus 2 octaves). For example, both of the codes shown in Fig. 3 have constant octave bandwidths. In the next section we consider how the values of these bandwidths affect the efficiency of the representation.

IMAGE ANALYSIS: ENERGY DISTRIBUTION

In this section we compare various coding schemes in terms of the way that they represent the six natural images shown in Fig. 6. The codes that will be compared will all fall within the general constraints of the model described earlier. However, within these constraints, it is possible to compare the advantages and disadvantages of codes that involve various bandwidths.

The main parameter that we will consider in this section is the variability of the responses of the arrays of sensors in the various channels. It was noted in the previous section that the variance in the outputs of the different channels will be roughly constant, independent of the channel that is selected. What, then, are the possible advantages of coding with a particular bandwidth? First, let us consider the rather unnatural images shown in Fig. 10. As was noted previously,[21] these two images are coded most efficiently with two quite different codes. Image 1 consists of a number of points





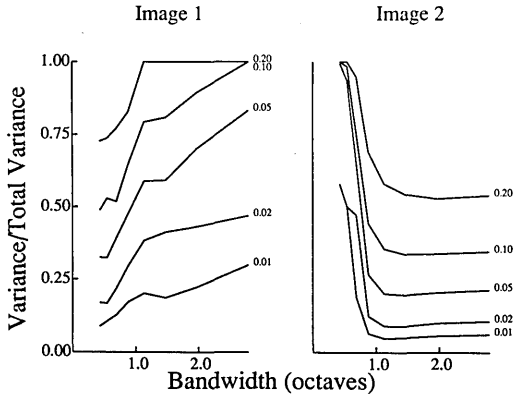


Fig. 10. Energy distribution for two artificial images. Image 1 consists of random points with a probability of 0.01. Image 2 consists of 10 sinusoids. The plots shown at the bottom of the figure describe the energy (as measured from the variance in the responses) represented by different subsets of sensors in relation to the total energy. Consider the data from Image 1 when the bandwidth is 1 octave. To get the data for this plot, the image was filtered through a rosette of 1-octave-wide channels (circular in the frequency domain: $\Delta F/\Delta\theta = 1.0$). The responses of the individual sensors and the envelope were then determined by appropriate sampling. The lowest curve labeled (0.01) shows the relative energy of the top 1% of the most active sensors. For Image 1, 1% of the sensors code about 20% of the total variance when the bandwidths of the channels are 1 octave, while 2% of the sensors code about 35% of the variance. The data for Image 1 show that more of the total variance can be represented with a small subset of the total sensors when the bandwidths are broadest. In contrast, for Image 2 more variance can be represented by a small subset of the total sensors when the bandwidths are narrowest.

scattered sparsely across the field. The probability of a point's existing at any given location is 0.01. What would be an efficient code for a class of images such as this?

Consider a code in which only the nonzero coefficients are transmitted. If we code the image in terms of pixel values, then for this type of image only 1% of the coefficients are needed to represent the image completely. Of course, information must be provided about the location of each particular nonzero coefficient, but as long as there are relatively few, there are a number of ways of efficiently representing the sparse set (e.g., run-length coding[25]).

Now, suppose that we perform a Fourier transform of this dot pattern. The spectrum of such a pattern will be quite broad. Almost all the coefficients will be nonzero. Therefore, to code such an image, it would be necessary to transmit the amplitude of most of the Fourier coefficients. In this sense, the pixel code can be described as an efficient code for such images, whereas the Fourier transform would be an inefficient code.

The opposite is the case for Image 2 of Fig. 10. This image consists of a small subset of sine-wave gratings of random orientation and frequency. To represent this image by using pixels would require a high proportion of the available pixels. However, a Fourier transform of such an image will result in only 10 nonzero coefficients of a possible 65,536. Clearly, for the class of images consisting of randomly selected sine-wave gratings, the Fourier transform is the most efficient code.

But what of our natural images? Within the constraints of our model, how can we design our code so that most of the information is packed into the smallest number of active sensors? The analyses described below were designed to test this.

Methods

All analyses described in this section were performed on the six 256 × 256 images described earlier and the two unnatural images described above. In all cases, the image was coded by arrays of sensors tuned to a range of frequencies and orientations within the constraints of the model described earlier. The procedure involved computing a discrete Fourier transform (DFT) of the image, multiplying the DFT by the appropriate channel, performing an inverse DFT, and sampling this convolved image in proportion to the sensor size (inversely proportional to the channel size). Both the even and the odd symmetric responses are determined, providing a measure of the evelope [Eq. (6)]. Because the edges of the image can produce spurious output in a filtered image,

only the central 160 × 160 pixels were directly analyzed; that is, the center of any particular sensor did not exceed this central region. However, depending on the size of this region, the sensors may extend out of it.

Even with this constraint, though, it was necessary to limit the analysis to channels centered above a particular frequency. This is because we wish to compare the behavior of codes with different bandwidths, including rather narrow bandwidths. A sensor tuned to low frequencies with a relatively narrow bandwidth has a large size relative to the image. Even with the border described above, the edges can have a significant effect. Because of this, the range of frequencies to which the channels were tuned was limited to the range of 24–128 cycles/picture width (128 is the Nyquist frequency for the 256 × 256 images). Thus, from the 25,600 possible sensors (160 × 160), roughly 20,800 sensors were used in the analysis described below.

The spatial-frequency bandwidths of the channels will be defined in relation to the width at half-height. In particular, the spatial-frequency bandwidth is defined as

$$B_{\text{oct}} = \log_2(f_a/f_b), \tag{11}$$

where f_a and f_b represent the frequencies at half-height above and below the peak spatial frequency.

With the exception of that for Fig. 13 below, the width/length aspect ratio of the channels ($\Delta F/\Delta\theta$) will remain fixed at 1.0. Thus the channels will be circular in the two-dimensional frequency domain.

The results described below are derived from the variability (i.e., the variance) of responses of the different sensors in a given code. To estimate this variability, it is important to have a sufficiently large sample of the different sensors. Beyond this, the precise nature of the sampling is unimportant to estimating the total variance. The estimate is not dependent on the precise degree of overlap between neighboring channels or neighboring sensors. It is important only that the number of sensors within a particular channel be inversely proportional to the channel size.

Results

To help to understand the analyses described in this section, let us first consider the images in Fig. 10. It is possible to represent such images by using the computational model described above. The amplitude spectra of such images do not fall off as $1/f$, and so different channels do not provide equal information. However, the general point can still be made.

The response of a particular channel can be defined in terms of the variance of the filtered image. As noted earlier, for the rosettelike codes the variance of the different channels will be roughly constant. However, a given variance can be produced by a range of different response distributions. For example, when a pixel code is used with Image 1, most of the response variance of the channels comes from the large responses of relatively few sensors. With the Fourier code, the overall variance (i.e., the power) comes from the generally low activity of a large number of coefficients.

The pixel code is efficient because most of the variance in the population of sensors is due to the response of relatively few sensors. Figure 10 provides a measure of the proportion of the variance represented by the most active sensors, as a function of the bandwidth (in octaves) of the code. For example, a particular code is chosen (within the constraints of the model) with a particular spatial frequency and orientation bandwidth (e.g., 1 octave, 20 deg). The code is then applied to the image, which results in a distribution of responses in the sensors. We can then consider the responses of the most active, say, 20% of the sensors and determine how much of the variance is accounted for by this subset. The plots in Fig. 10 show how much of the total energy is accounted for by the top 20, 10, 5, and 1% of the sensors.

Figure 10 shows that for Image 1, as the bandwidth of the code increases, the most active sensors account for higher and higher percentages of the total energy. In contrast, for Image 2, the most energy is packed into the fewest number of cells when the bandwidths of the channels are narrowest.

With this definition, the efficiency of the code is a function of both the parameters of the code and the statistics of the image. But what is the optimal solution for natural images? Figure 11 shows the results for our six images.

The results for each of the six images in Fig. 11, with the possible exception of Image E, show the same general trends. The optimal bandwidths are neither very narrow or very broad but are in the range of 0.5 to 1.5 octaves. In other words, in order to represent the maximum energy with the fewest sensors, the optimal solution within the constraints of the model is to use sensors in the range of 0.5 to 1.5 octaves.

If the energy of the image is represented primarily by a few of the total sensors, we would expect the response of these few sensors to be high relative to the average. Figure 12 shows the average response of different subsets relative to the average response of the population. These data show that when the bandwidths are about 1 octave, the most

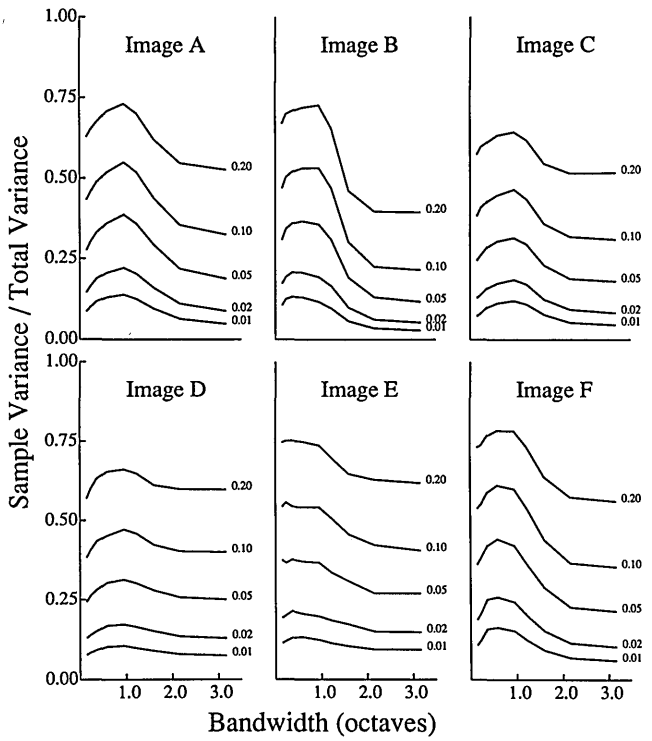


Fig. 11. Energy distribution for the six natural images. Results show the energy represented by different subsets of sensors. Results are shown for the range of 0.12–3.2 octaves. The ratio of orientation bandwidth to spatial-frequency bandwidth ($\Delta F/\Delta\theta$) is fixed at 1.0 (i.e., circular channels).

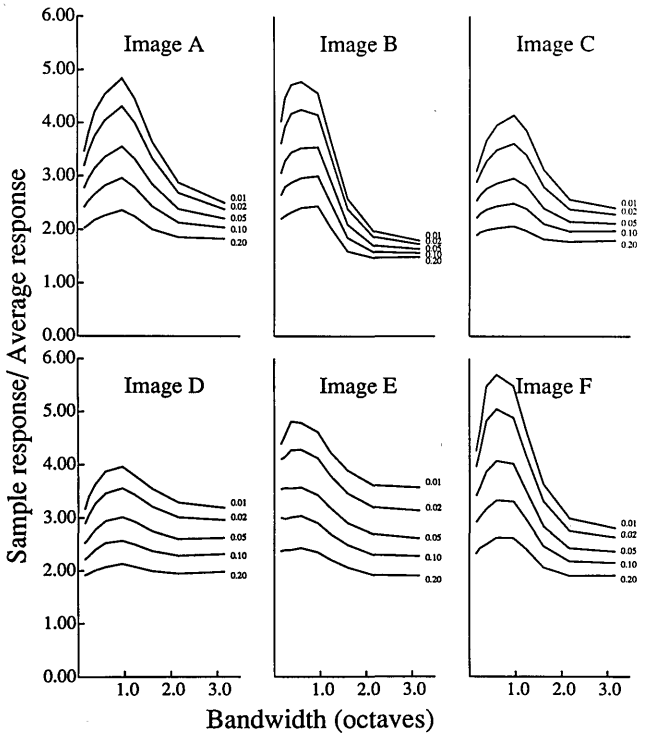


Fig. 12. Relative response of the different subsets of sensors as a function of the bandwidth. For example, the plot labeled 0.01 represents the average response of the top 1% of the sensors relative to the average response of all the sensors. If we consider the sensor response to be subject to noise, then this plot can be related to the signal-to-noise ratio. It suggests that spatial-frequency bandwidths in the range of 1 octave produce the highest signal-to-noise ratio.

active sensors have a high response to the average. If we consider the fact that cortical neurones are inherently rather noisy in their response to a stimulus,[31] this plot can be considered a measure of the signal-to-noise ratio of different types of sensors. Coding information into channels with approximately 1-octave bandwidths produces a representation in which a small proportion of the cells represents a large proportion of the information with a high signal-to-noise ratio.

We have so far considered only channels for which the ratio of the spatial-frequency bandwidth to the orientation bandwidth is constant ($\Delta F/\Delta\theta = 1.0$). Figure 13 shows results with various different aspect ratios. One of the difficulties of such an analysis is that the two-dimensional Gabor functions are not polar separable. That is, the spatial-frequency tuning is not independent of the orientation tuning (in degrees). Extending the orientation bandwidth actually extends the response of the channel to higher frequencies. With the $1/f$ amplitude spectrum, the response of the channel will be dependent primarily on the lower frequencies, with little effect produced by the extension. Nonetheless, the results of such an analysis are shown in Fig. 13. As can be seen, an aspect ratio of about 0.5–1.0 is somewhat optimal, although the effects are small.

LOG GABOR CHANNELS

In the previous section we compared a variety of Gabor codes in terms of their ability to represent the information in natural images. Gabor codes have a number of interesting mathematical properties. As described by Gabor[14] and more recently by Daugman,[12] a Gabor function represents a minimum in terms of the spread of uncertainty in space and spatial frequency (actually time and frequency in Gabor's description). However, the Gabor code is mathematically pure in only the Cartesian coordinates where all the Gabor channels are the same size in frequency and hence have sensors that are all the same size in space (i.e., all the rectangles in the diagrams in Fig. 1 are the same size). In such a case, the Gabor code represents the most effective means of packing the information space with a minimum of spread and hence a minimum of overlap between neighboring units in both space and frequency.

However, modifying the basic structure of the code to permit a polar distribution such as that shown in our rosettes (Fig. 3) alters the relative spread and overlap between neighbors. In this section some results are described that were obtained with a function that partially restores some of the destructive effects of the polar mapping. This function will be called the log Gabor function. It has a frequency response described by

$$G(f) = \exp\{-[\log(f/f_0)]^2/2[\log(\sigma/f_0)]^2\}, \quad (12)$$

that is, the frequency response is a Gaussian on a log frequency axis. Figure 14 provides a comparison between the

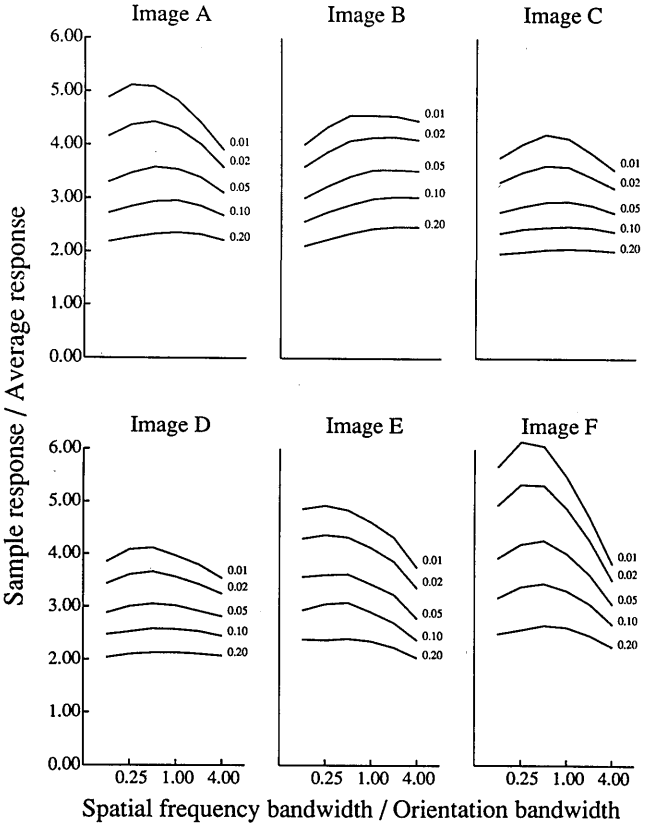


Fig. 13. Relative response of the different subsets as a function of the ratio of the orientation bandwidth to the spatial-frequency bandwidth ($\Delta F/\Delta\theta$). A ratio of 1.0 represents a channel that is circular in frequency space. A ratio of 0.25 represents a channel that is narrowly tuned to orientation realtive to the spatial-frequency tuning. For all plots shown here, the spatial-frequency bandwidth is fixed at 1.2 octaves.

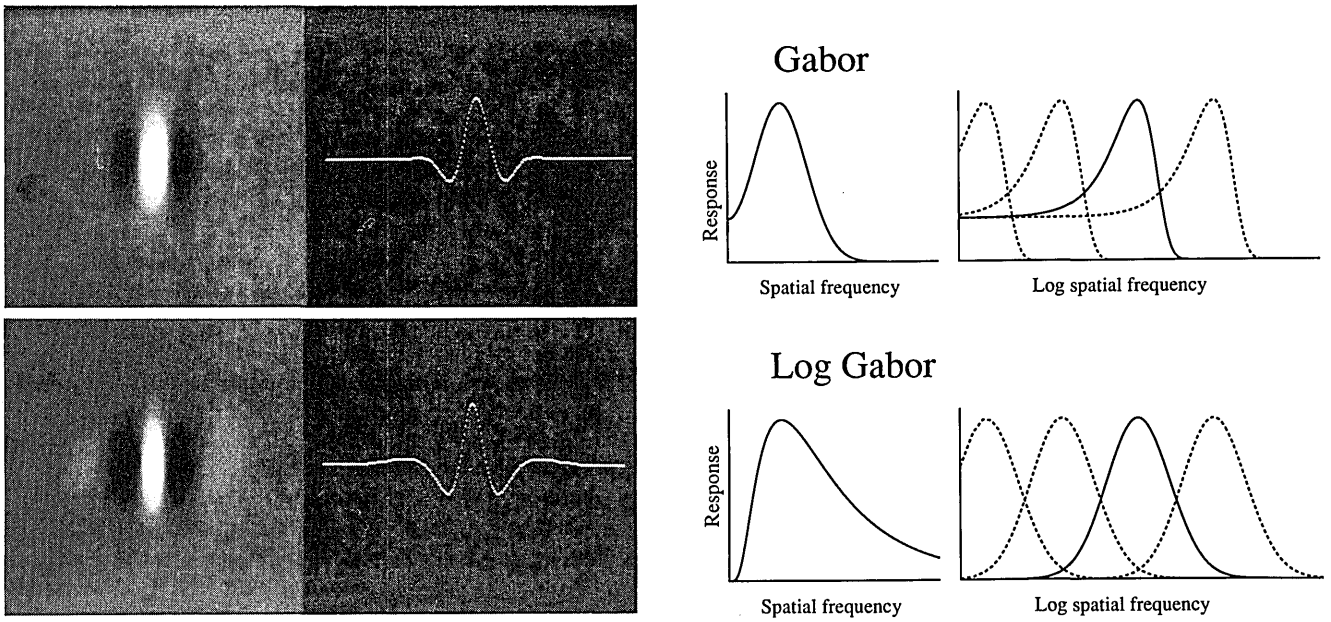


Fig. 14. Comparison of the Gabor and the log Gabor functions. The far left-hand images show examples of the two functions in their even symmetric form along with the response amplitude across the midline. The amplitude spectrum of a Gabor function is represented by a sum of Gaussians (positive plus negative frequencies) and will be roughly symmetric on a linear axis. The log Gabor is a Gaussian on a log frequency axis. The channels in our rosettes have constant bandwidths in octaves, and therefore the amplitude spectra of the different channels will the the same scale on the log frequency axis. As shown by the plots on the right, the Gabor functions overrepresent the low frequencies with this type of sampling. This is not the case for the log Gabor.

Gabor function and this log Gabor function. The important aspect of this function is that, unlike the Gabor function, the frequency response of the log Gabor is symmetric on a log axis. Indeed, the log axis is the standard method for representing the spatial-frequency response of visual neurones. The results of several studies seem to imply that such symmetry is strong possibility. Recent work by Hawken and Parker[32] suggests that the Gabor function fails to capture the precise form of the spatial-frequency tuning curves in monkey cortical cells. In their detailed study, a number of models are compared for their ability to predict the form of the spatial-frequency tuning curve. Gabor functions miss in this fit primarily because they fail to capture the relative symmetry of the tuning curves on a log axis. This does not mean that the log Gabor is the best-fitting function (Hawken and Parker actually suggest a model based on a sum of weighted Gaussians), but it may well provide a better description than the Gabor function.

One of the advantages of the log Gabor is its use with codes in which the bandwidths increase with frequency (i.e., are constant in octaves). The right-hand side of Fig. 14 shows a one-dimensional representation of the spacings of the Gabor function on a log axis. With the bandwidths constant in octaves, the Gabor functions overrepresent the low frequencies. Furthermore, with a $1/f$ falloff in the amplitude, most of the input to the Gabors will be provided by the low-frequency tails of the functions. This will, in essence, produce a correlated and redundant response to the low frequencies. In contrast, mapping the information into the log Gabors spreads the information equally across the channels.

Figure 15 shows the results of the log Gabors (solid lines) compared with the Gabors (dashed lines) obtained by using the methods described in the previous section (Figs. 11 and 12). For bandwidths of less than 1 octave, the two functions produce similar results. At bandwidths of >1 octave, the redundancy at the low frequencies becomes apparent. Since all the Gabor sensors receive a significant and redundant input from the low frequencies, the responses of any given sensor represents a smaller fraction of the total energy.

The frequency response of the log Gabor code permits a more compact representation than the Gabor code when the bandwidths are >1 octave. The log Gabor may not be the ideal function for coding images, and it is probably not the ideal function for representing cortical simple cells. However, its success with broad bandwidths may help to explain why the frequency-response functions of many cortical cells are symmetric on a log axis.

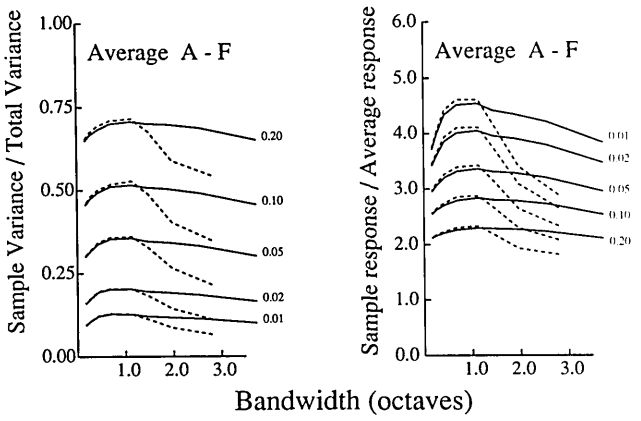


Fig. 15. Comparisons of the Gabor and log Gabor functions for coding the six natural images. Each plot shows the average result for the Gabor functions from Figs. 11 and 12 (dashed lines) and for log Gabors (solid lines). With bandwidths greater than 1 octave, the log Gabor shows the potential of producing a more compact code.

DISCUSSION

Over the past two decades, a number of attempts have been made to explain the purpose of the rather mysterious behavior of cortical neurons. The evidence that cortical neurons are selective to spatial frequency as well as orientation directed a number of researchers to suggest that such neurons must be producing something like a Fourier transform and not performing feature detection.[33,34] Features were assumed to be things such as edges, bars, and corners. The frequency selectivity of cortical cells seemed to be in opposition to the notion of feature detection because frequency selectivity seemed to have little to do with the properties of the natural environment.

We have tried to show in this paper that the response properties of cortical neurons are well suited to the statistics of natural images. The frequency selectivity allows the images to be represented by a few active cells. One might say that sensors with the tuning of cortical cells provide the best chance of giving a large response or no response at all. However, one should be hesitant in describing such behavior as feature detection. No mention has been made of what statistics of the environment might be biologically significant to the animal. No effort has been made to give preference to any particular object or event in the environment. We suggest instead that the code allows, on the average, the most information to be represented with a small proportion of cells. However, information is defined in relation to the variability of the images, not any specific feature.

We can also relate this description of information in terms of the redundancy of the images. Barlow[20–23] has provided a thorough discussion of the relations between redundancy and visual codes. Indeed, many of the results described here may be discussed best in terms of Barlow's theories of redundancy.

The redundancy in a set of images is usually defined in terms of the nth-order conditional probabilities of the coefficients (e.g., the amplitudes of the pixels; see Ref. 18 for an excellent discussion). Consider an array of pixels with a range of possible intensity levels. First-order statistics relate to the probability of individual pixels' (e.g., pixel i) taking on particular intensity levels (m) – $p(m_i)$.

There is redundancy in the first-order statistics when the distribution of intensities is not uniform. A nonuniform distribution implies that there is some degree of predictability or order in the intensity values.

Second-order statistics refer to the conditional probability of pairs of pixels. It is a measure of the probability that a pixel will take on a particular value given the value of another pixel, $p(m_i|\ n_j)$. Most considerations of redundancy include only the second-order redundancy portrayed in the power spectrum and the autocorrelation function. However, higher-order redundancy (e.g., third order) may provide a significant contribution to the total redundancy of an image. Consider an image consisting of small line segments of random orientation. In such an image, if we find two neighboring points with the same intensity, then it is likely that there will be a third point along the line with the same intensity. This correlation can be described as a third-order statistic, since it concerns the relation between triplets of points $p(m_i|\ n_j, o_k)$. The power spectrum and the autocorrelation function provide no information about this third-order statistic.

How should we represent these various forms of redundancy? Barlow suggests that the ultimate goal is to reduce the redundancy of the code. Does the model proposed here reduce redundancy? The answer is clearly no. The total number of sensors has remained fixed, and the amount of information represented by the code is constant (for a complete code). Thus the order in the code has been maintained, and therefore the total entropy or redundancy is constant.[35]

Instead, the code presented here provides a means of converting higher-order redundancy (correlations between pairs of pixels, triplets of pixels, etc.) into first-order redundancy (i.e., the response distribution of the sensors). Theoretically the total redundancy will be unchanged.

For example, consider Image 2 of Fig. 10. If we code such an image in terms of pixels, then there will be a fairly even distribution of responses. An even distribution implies that the entropy is relatively high and that the redundancy or predictability is low; that is, no particular response is more or less likely than any other. The redundancy of such an image lies in the correlations between neighboring pixels. A Fourier transform of this image transforms this second-order redundancy into redundancy in the response distribution (first order). The Fourier coefficients are not evenly distributed. The most likely state is 0 with a small probability of a large response. Hence the first-order statistics are more predictable or redundant.

The response distributions represented in Figs. 11 and 12 can be interpreted in terms of the redundancy of the different codes. With a 1-octave bandwidth, the information is packed into the smallest number of sensors, giving a highly skewed distribution and therefore a redundant code. In other words, the most efficient code by our terminology is the code with most redundant first-order statistics.

The next stage of processing can make efficient use of these first-order statistics by coding only the nonredundant elements (i.e., the highly active sensors). It is therefore the next stage of processing that has the potential for removing redundancy. The codes described here should be no more and no less redundant than the input.

Karhounen–Loeve Transforms

Bossomaier and Snyder[36] recently discussed some of the similarities between the response properties of simple cells and the solution of a statistical analysis called a Karhounen–Loeve transform (KLT; e.g., see Ref. 25). Bossomaier and Snyder suggest that, in the same way that the KLT can be used to reduce redundancy, local spatial-frequency analysis may be the optimal procedure for removing statistical redundancy in real images. However, what they presume to be the goal of the visual code is not the same as that proposed here. The KLT computes the eigenvalues of the covariance matrix (i.e., the covariance between pairs of pixels). The corresponding eigenvectors represent a set of orthogonal coefficients for which the vector with the greatest eigenvalue accounts for the greatest part of the covariance.

The KLT is used to represent the information from a class of images (or blocks within an image) as a hierarchy of orthogonal coefficients in which most of the image energy is represented consistently by the same small subset of coefficients. By using only this small subset, it is possible to code most of the energy in the images with a large reduction in the number of free parameters. Redundancy is reduced by re-

ducing the dimensions (i.e., the number of sensors) of the code. As was noted by Snyder and Bossomaier,[36] if the statistics of the images are stationary (i.e., across all the images, the statistics at any one location are no different from those at any other location), then the KLT produces coefficients similar to those of a Fourier transform. Indeed, if we consider the amplitude spectra shown in Figs. 5 and 6, it is clear that most of the energy can be conserved by transmitting only the low frequencies. For images such as these with power spectra that fall as a function of frequency, using the KLT to eliminate weak coefficients will remove the higher frequencies.

The code proposed in this paper behaves quite differently. The goal is not to discard any type of channel or sensor. The coefficients are matched to the image energy ($1/f^2$) so that, on the average, any particular coefficient (channel or sensor) is just as likely to be active as any other. No particular coefficient is favored by such a code. Hence there can be no reduction in the number of free parameters. This code is efficient because the redundancy of the first-order statistics (i.e., the response distribution of the sensors) has increased. For any particular image, only a small subset of the total number is active. Since the particular members of the subset will vary from image to image, it is not possible to eliminate any particular type of coefficient. If the number of free parameters (i.e., the dimensionality) is constant, then the total redundancy should remain constant.

It is proposed here that this coding scheme (which is not a local Fourier transform) represents a good method for transforming higher-order redundancy into first-order redundancy. It is presumed that this redundancy is used in later stages to produce a nonredundant code. Indeed, if the later stages analyze the outputs of only the very active cells (i.e., impose a threshold), then redundancy can be reduced. However, this model of simple-cell behavior does not achieve this. These results can only indirectly support Barlow's theory[20-23] that the ultimate goal of visual processing is to represent visual information in a nonredundant form. By this model, cortical simple cells do not represent information with less redundancy. Instead they transform the information to permit later stages of the visual system to be less redundant.

We must limit ourselves to the claim that such a code is only a good method for transforming redundancy. We have only searched through two types of code (Gabor and log Gabor). There might well be some other function that provides a more efficient means of transforming redundancy and that might also provide a better description of the neurophysiology.

Physiological Reality

A number of points must be addressed regarding the differences between the idealized model proposed here and the coding by the mammalian visual system. Some differences (e.g., scaling factors from fovea to periphery) are not directly relevant to the model. However, there are at least four areas in which the known properties of the visual system appear to conflict with the model.

1. *Precise Bandwidths and Spacing*

Our results suggest that channels with a ~1-octave bandwidth and an orientation/frequency bandwidth ratio of around 0.8 will produce a relatively efficient representation of the information in these six natural images. This is in good agreement with the general neurophysiological findings for spatial-frequency bandwidths in cortical simple cells (e.g., Refs. 37 and 38) and with ratios averaging around 0.6.[39,40] However, the various codes described here are based on an even distribution of channels and sensors (e.g., Fig. 3). The restrictions that were imposed permitted a comparison of different coding schemes. There is little reason to expect such precise coding in the visual cortex. Indeed, a number of authors[37,38] have pointed out the wide distribution in the spatial-frequency and orientation bandwidths of cortical neurons. Furthermore, it is unlikely that the spacing between cortical neurons is anything like the rigid grid proposed here. It remains unclear what effect variability will have on the code. A wide distribution in the bandwidths may even produce a more efficient code by providing a wider selection of potential matches between the sensors and the image.

2. *Noise*

Actual cortical neurons are quite noisy, providing a rather inconsistent response to a constant stimulus. Although noise is not explicitly attached to this model, one might expect the model to work well with noisy cells (i.e., sensors). As is shown in Figs. 11 and 12, when the bandwidths are in the region of 1 octave, the most active cells give the highest response relative to the average. Thus, if there is noise associated with this average response, then this range of bandwidths will give the optimal signal-to-noise ratio.

3. *Optics*

The photographs that were analyzed in this study were each corrected for the optics of the camera. However, one might question whether the analyses should reflect the optics of the mammalian eye. Indeed, it was noted above that the constant bandwidth (in octaves) was well suited to the falloff in energy of the images. A reduction in the amplitude of the higher spatial frequencies will certainly effect the equal distribution of information through the system. However, this should not necessarily imply that the optimal bandwidths should vary as a function of frequency.[41] As long as the falloff is not too steep, then each channel's input will be modified only by an overall reduction in contrast. This reduction may decrease the signal-to-noise ratio for the high-frequency channels but should have little effect on which bandwidths produce the optimal ratio. However, it does imply that a threshold applied equally across the different channels is more likely to effect the channels reduced by the relative effects of the optical system.

4. *Redundancy*

So far, we have discussed redundancy as something to avoid in a visual code. However, it is well known that redundancy can be an advantage if the coding process is subject to noise. However, this is a second form of redundancy. The redundancy that we have discussed so far refers to the statistics and correlations of the input. In this sense, it is redundant to code images of a point in terms of sinusoids because the different sinusoids are redundant. A second type of redundancy refers to overrepresenting the information in a stimulus by using more free parameters (i.e., coefficients) than are

required. For an image consisting of random points, this form of redundancy would be reflected in a code that had more than one point detector at each position. In this paper we have dealt with the first type of redundancy. We suggest that spatial-frequency bandwidths in the range of 1 octave are good for converting such redundancy into the response activity of cortical neurons. But this does not mean that the second type of redundancy would be harmful to such a code. It might prove to be useful to provide additional sensors to make certain that the signal gets transmitted.

SUMMARY

As Laughlin[16] pointed out, an efficient sensory system should match its analyzers to the nature of the signals it processes. If images of the natural environment were truly random, then a statistical match would not be possible. However, natural images are not random, and the redundant properties can be exploited to provide an efficient code.

In this paper an attempt was made to show that the receptive-field properties of mammalian cortical cells are well suited for representing the information contained in natural images. This conclusion is based on an extremely small sample consisting of six images. Clearly, such a sample is insufficient to serve as the basis of any form of proof. A proper analysis based on a much larger population of images would be required. Rather, the purpose of this paper is to provide suggestions for how to relate the statistics of the natural environment to cortical-cell behavior. The general approach is not limited to the spatial properties of cortical simple cells. Stereo vision, motion, and color might all be aided by a better understanding of the constraints imposed by the environment.

ACKNOWLEDGMENTS

A number of people have provided helpful discussions and assistance throughout the development of this project. These include many of the participants of the Rank Symposium on Efficiency held in December 1986 in Cambridge. I also thank D. J. Tolhurst and L. Ling for their invaluable advice and Jacob Nachmias and John Robson for their critical comments on early drafts of this paper. I also thank Dan Kersten for our discussions on redundancy. I am especially grateful to Horace Barlow and Graeme Mitchison for their comments and for the use of their Sun Workstation, which made these formidable analyses possible. This research was partly supported by National Eye Institute Research Fellowship Award F32 EY05737-01.

REFERENCES AND NOTES

1. D. G. Hubel and T. N. Weisel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," J. Physiol. **160,** 106–154 (1962).
2. D. Marr, *Vision* (Freeman, San Francisco, 1981).
3. H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," Vision Res. **19,** 19–33 (1979).
4. A. B. Watson, "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images*, O. J. Braddick and A. C. Slade, eds. (Springer-Verlag, Berlin, 1983).
5. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Physiol. **197,** 551–556 (1968).
6. C. Blakemore and F. W. Campbell "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images," J. Physiol. **203,** 237–260 (1969).
7. D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. London Ser. B **207,** 187–217 (1980).
8. D. J. Field and D. J. Tolhurst, "The structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex," Proc. R. Soc. London Ser. B **228,** 379–400 (1986).
9. L. A. Palmer, J. P. Jones, and W. H. Mullikin, "Functional organization of simple cell receptive fields," in *Models of the Visual Cortex*, D. Rose and V. G. Dobson, eds. (Wiley, New York, 1985), pp. 273–280.
10. S. Marcelja, "Mathematical description of the responses of simple cortical cells," J. Opt. Soc. Am. **70,** 1297–1300 (1980).
11. J. J. Kulikowski, S. Marcelja, and P. O. Bishop, "Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex," Biol. Cybern. **43,** 187–198 (1982).
12. 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).
13. M. R. Turner, "Texture discrimination by Gabor functions," Biol. Cybern. **55,** 71–83 (1986).
14. D. Gabor, "Theory of communication," J. Inst. Electr. Eng. **93,** 429–457 (1946).
15. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina," Proc. R. Soc. London Ser. B **216,** 427–459 (1982).
16. S. B. Laughlin, "Matching coding scenes to enhance efficiency," in *Physical and Biological Processing of Images*, O. J. Braddick and A. C. Sleigh, eds. (Springer, Berlin, 1983), pp. 42–72.
17. E. Switkes, M. J. Mayer, and J. A. Sloan, "Spatial frequency analysis of the visual environment: anistropy and the carpentered environment hypothesis," Vision Res. **18,** 1393–1399 (1978).
18. D. Kersten, "Predictability and redundancy of natural images," J. Opt. Soc. Am. A **4,** 2395–2400 (1987).
19. J. J. Gibson, *The Perception of the Visual World* (Houghton Mifflin, Boston, 1966).
20. H. B. Barlow, "The coding of sensory messages," in *Current Problems in Animal Behavior*, W. H. Thorpe and O. L. Zangwill, eds. (Cambridge U. Press, Cambridge, 1961), pp. 331–360.
21. H. B. Barlow, "Three theories of cortical function," in *Development Neurobiology of Vision*, R. D. Freeman, ed. (Plenum, New York, 1979), pp. 1–16.
22. H. B. Barlow, "The Ferrier lecture," Proc. R. Soc. London Ser. B **212,** 1–34 (1981).
23. H. B. Barlow, "Understanding natural vision," in *Physical and Biological Processing of Images*, O. J. Braddick and A. C. Sleigh, eds., Vol. 11 of Springer Series in Information Sciences (Springer-Verlag, Berlin, 1983), pp. 2–14.
24. B. Sakitt and H. B. Barlow, "A model for the economical encoding of the visual image in cerebral cortex," Biol. Cybern, **43,** 97–108 (1982).
25. A. N. Netraveli, and J. O. Limb, "Picture coding: a review," Proc. IEEE **68,** 366–406 (1980).
26. C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication* (U. Illinois Press, Champaign, Ill. 1949).
27. D. A. Pollen and S. F. Ronner, "Phase relationships between adjacent simple cells in the cat," Science **212,** 1409–1411 (1981).
28. R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1965).
29. A. W. Rihaczek, "Signal energy distribution in time and frequency," IEEE Trans. Inf. Theory **IT-14,** 369–374 (1968).
30. R. F. Voss, "Random fractal forgeries," in *Fundamental Algorithms for Compuer Graphics*, R. A. Earnshaw, ed. (Springer-Verlag, Berlin, 1985), pp. 805–829.
31. D. J. Tolhurst, J. A. Movshon, and A. F. Dean, "The statistical reliability of single neurons in act and monkey visual cortex," Vision Res. **23,** 775–785 (1983).
32. M. Hawken and A. Parker, "Spatial properties of neurons in the monkey striate cortex," Proc. R. Soc. London Ser. B **231,** 251–288 (1987).
33. K. K. DeValois, R. L. DeValois, and E. W. Yund, "Responses of

striate cortical cells to grating and checkerboard patterns," J. Physiol. **291,** 483–505 (1980).
34. D. G. Albrecht, R. L. DeValois, and L. G. Thorell, "Visual cortical neurons: are bars or gratings the optimal stimuli?" Science **207,** 88–90 (1981).
35. The redundancy of any particular representation must be defined over all *n*th-order statistics (i.e., over all the possible conditional probabilities). Transforming the coordinate system from one set of coordinates to another (e.g., rotation) will not alter the total redundancy of the representation but may transform redundancy of one form (e.g., third order) into that of another (e.g., first order).
36. T. Bossomaier and A. W. Synder, "Why spatial frequency processing in the visual cortex?" Vision Res. **26,** 1307–1309 (1986).
37. D. J. Tolhurst and I. D. Thompson, "On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat," Proc. R. Soc. London Ser. B **213,** 183–199 (1982).
38. R. L. DeValois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," Vision Res. **22,** 545–559 (1982).
39. J. A. Movshon, "Two-dimensional spatial frequency tuning of cat striate cortical neurons," in *Society for Neuroscience Abstracts* (Society for Neuroscience, Atlanta, Ga., 1979), p. 799.
40. M. A. Webster and R. L. DeValois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," J. Opt. Soc. Am. A **2,** 1124–1132 (1985).
41. It should be noted that the neurophysiological data do not support the notion of constant bandwidths in octaves.[37,38] Not only is there a wide range of bandwidths but the average bandwidth shows some decrease with increasing frequency. The present model has yet to account for this decrease.