ARTICLE Communicated by Ted Adelson

What Is the Goal of Sensory Coding?

David J. Field
Department of Psychology, Cornell University, Ithaca, NY 14850 USA

A number of recent attempts have been made to describe early sensory coding in terms of a general information processing strategy. In this paper, two strategies are contrasted. Both strategies take advantage of the redundancy in the environment to produce more effective representations. The first is described as a "compact" coding scheme. A compact code performs a transform that allows the input to be represented with a reduced number of vectors (cells) with minimal RMS error. This approach has recently become popular in the neural network literature and is related to a process called Principal Components Analysis (PCA). A number of recent papers have suggested that the optimal "compact" code for representing natural scenes will have units with receptive field profiles much like those found in the retina and primary visual cortex. However, in this paper, it is proposed that compact coding schemes are insufficient to account for the receptive field properties of cells in the mammalian visual pathway. In contrast, it is proposed that the visual system is near to optimal in representing natural scenes only if optimality is defined in terms of "sparse distributed" coding. In a sparse distributed code, all cells in the code have an equal response probability across the class of images but have a low response probability for any single image. In such a code, the dimensionality is not reduced. Rather, the redundancy of the input is transformed into the redundancy of the firing pattern of cells. It is proposed that the signature for a sparse code is found in the fourth moment of the response distribution (i.e., the kurtosis). In measurements with 55 calibrated natural scenes, the kurtosis was found to peak when the bandwidths of the visual code matched those of cells in the mammalian visual cortex. Codes resembling "wavelet transforms" are proposed to be effective because the response histograms of such codes are sparse (i.e., show high kurtosis) when presented with natural scenes. It is proposed that the structure of the image that allows sparse coding is found in the phase spectrum of the image. It is suggested that natural scenes, to a first approximation, can be considered as a sum of self-similar local functions (the inverse of a wavelet). Possible reasons for why sensory systems would evolve toward sparse coding are presented.

Neural Computation **6, 559–601** (1994)

1 Introduction

Although we know a great deal about how sensory systems code information, there remains considerable debate regarding the goal of this coding. In many studies, there is an implicit assumption that there is no single goal. It is assumed that sensory systems solve a wide range of tasks important to the animal and since the range of tasks varies widely, one would not expect to see any common "theme" across the different coding strategies. A second approach, which serves as the basis for the ideas presented in this paper, proposes that it is possible to describe sensory coding in terms of a general information processing strategy. By this tradition, it is presumed that redundancy in different sensory environments can be represented within a single framework and that the goal of sensory coding is to transform the redundancy to provide some advantage to later stages of processing. In this paper, two information processing strategies are contrasted. Both of these approaches take advantage of the redundancy in the input to produce more effective representations of the environment. However, the two approaches achieve different goals and depend on different forms of redundancy.

The first of these, which will be described as "compact coding," has gained considerable attention in the neural network literature and serves as the basis of much of the work in image compression. This approach suggests that the principal goal of visual coding is to reduce the redundancy of the visual representation. Many of these ideas can be traced backed to Barlow's theories of redundancy reduction (e.g., Barlow 1961). Recently, a number of studies have proposed that spatial coding by the mammalian visual system is well described by codes that make use of the correlations to reduce the redundancy of the sensory representation (e.g., Atick and Redlich 1990, 1992; Atick 1992; Barlow and Foldiak 1989; Daugman 1988, 1991; Linsker 1988; Foldiak 1990; Sanger 1989).

This approach to coding is illustrated in Figure 1A. In a compact code, the goal is to represent all the likely inputs with a relatively small number of vectors (e.g., cells) with minimal loss in the description of the input. In such a code, the dimensionality of the representation is reduced, resulting in a code where only a subset of the possible inputs can be accurately represented. The code is effective when this subset is capable of representing the probable inputs to the code. In the next section, we will see how this approach reduces redundancy.

The second approach suggests that the principal goal of sensory coding is to produce a sparse-distributed representation of the sensory input. This approach has been specifically proposed with respect to visual code and the representations of natural scenes (Field 1987, 1989, 1993; Zetzsche 1990). However, several authors have noted that codes that produce sparse outputs may provide several advantages for representing sensory information (e.g., Barlow 1972, 1985; Palm 1980; Baum *et al.* 1988). Unfortunately, in much of this work, the distinction between sparse and

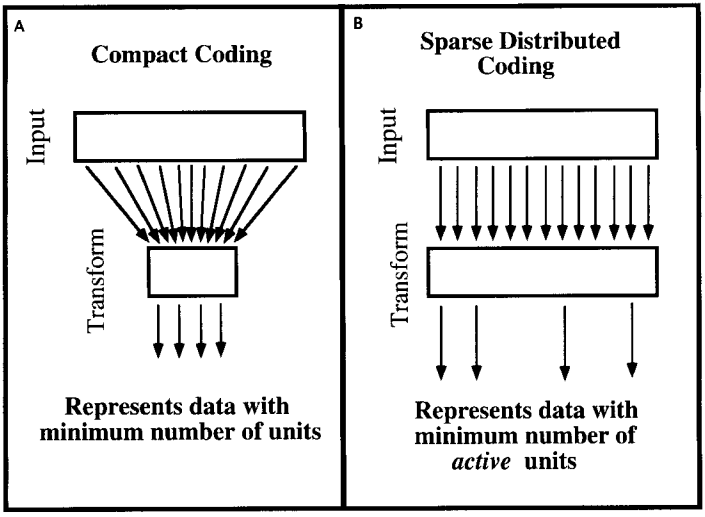


Figure 1: Two methods of taking advantage of redundancy in a sensory environment. In the compact coding approach (A), the code transforms the vector space to allow the data to be represented by a smaller number of vectors with only minimal loss in the representation (i.e., dimensionality is reduced). In a sparse coding scheme (B), the code transforms the vector space to allow the input to be represented with a minimum number of active cells. In a sparse coding scheme, the dimensionality is not reduced. Rather, the redundancy in the input is transformed into the redundancy in the firing rate of the cells (i.e., the response histogram) to produce a code where the response probability for any particular cell is relatively low.

compact codes has not been clear. Therefore, in this paper, much of the discussion will be devoted to the differences and requirements for each type of coding.

In a sparse-distributed code, the dimensionality of the representation is maintained (the number of cells remains roughly constant and may even increase). However, the number of cells responding to any particular instance of the input is minimized. Over the population of likely inputs, every cell has the same probability of producing a response (i.e., distributed) but that probability is low for any given cell (i.e., sparse). In sparse-distributed coding, the goal is to obtain a code where only a few cells respond to any given input. However, across the population of images the information is distributed across all of the cells. As we will see in the next section, this coding scheme does not reduce redundancy.

Rather, high-order redundancy is transformed into redundancy of the firing patterns of the cells (i.e., the response histograms) as was proposed previously (Field 1987). By this approach, the goal of the coding is to maximize the redundancy of the response histograms by minimizing the statistical dependencies between units.

In the following sections, these two approaches will be analyzed and contrasted. We will begin by looking at the type of redundancy required for producing a compact code and discuss the relations between this code and Principal Components Analysis (PCA). We will then look at the type of redundancy and the type of transform required for a sparse code. This will be followed by an experiment that models the response of visual neurons to natural scenes. Throughout this paper, we will concentrate on the response properties of cells in the mammalian visual system and the statistical relations found in the natural environment. However, the general ideas can be applied to any sensory system and any sensory environment.

2 State Space

One effective method for describing redundancy in a data set is to consider the "state space" of possible inputs. The state space (sometimes called vector space or phase space) describes the space of all possible states. This approach provides a useful means of showing how different coding schemes take advantage of the various forms of redundancy present in a population of inputs. Although the notion of state space has proved popular in descriptions of chaotic interactions and learning in neural networks (e.g., Churchland and Sejnowski 1992 for an introduction), it is not widely discussed in theories of visual coding.

To describe the state space of an image set, one can use the pixel amplitudes to represent the coordinate axes of the space. For any n-pixel image, one requires an n-dimensional space to represent the set of all possible images. Every possible image (e.g., a particular face, tree, etc.) is represented in terms of its unique location in the space. For example, white noise patterns (i.e., patterns with random pixel intensities) represent random locations in that n-dimensional space. Therefore, if we let the space become filled with examples of such white noise patterns, then all regions of the space would be filled equally—that is, the probability density will be uniform throughout the state space.

However, most naturally occurring phenomena are not random. Any redundancy that occurs in a data population implies that the population does not fill the state space uniformly. Natural scenes, for example, are statistically very different from white noise patterns and we would therefore not expect the state space of natural scenes to have the same probability density. Consider the case of a 256×256 pixel image. The state space of all possible images at this resolution requires a 65,536-

dimensional space where the amplitudes of each of the pixels represent the axes of the space. As one can imagine, the probability of seeing something resembling a natural scene when presented with white noise patterns is extremely low. This implies that in the state space of all possible 256×256 images, natural scenes occupy an extremely small area of the space.

To develop an understanding of the functional properties of the mammalian visual system, a number of recent studies have proposed that one must understand the statistical regularity of the mammalian visual environment and its relation to the visual code (e.g., Field 1987, 1989, 1993; Atick and Redlich 1990, 1992; Bialek *et al.* 1991; Eckert and Buchsbaum 1993; Hancock *et al.* 1992; van Hateren 1992; Kersten 1987). In terms of the state space, every statistical regularity that one finds in the visual environment provides a clue about the location and shape describing the probability density of natural scenes within the state space. Our visual environment is highly structured. The physics of how objects and surfaces reflect light forces the probability density into highly constrained forms. We will see that much of the debate about the goal of visual coding depends on the particular shape that is produced with data from the sensory system's typical environment. It will be argued that the state space describing the probability density of natural scenes is highly predictable but does not have the shape that is widely presumed.

Similarly, the spatial response properties of a cell (e.g., the receptive field) can also be described in terms of locations in the state space that will produce a response. For example, a cell with a linear response can be represented by a vector extending from the origin of the state space in a direction that depends uniquely on the particular receptive field profile. Every distinguishable receptive field profile is represented in terms of a unique direction in the state space. Thus, an array of cells can be considered as an array of vectors each pointing to unique locations in the state space.

If the local amplitudes of a waveform (e.g., the pixels) represent the coordinate axes of the state space, then an ortho-normal transform (e.g., a Fourier transform) can be represented by a rotation and/or translation of the coordinate axes.[1] This concept of treating a transform as a rotation of the coordinate axes will prove important in the discussions below. Ortho-normal transforms represent the special class of transforms where the vectors remain of normal length

$$|V_i| = 1$$

and the vectors remain orthogonal.

$$V_i \cdot V_j = 0$$

[1]Technically, an ortho-normal transform performs an "isometric" transform on the state space (i.e., it is form preserving).

It is unlikely that the visual system's transform is either completely orthogonal or normal, but this will not affect the main points of the discussion presented in this paper. To a first approximation, a sensory code can be thought of as rotation of the coordinate axes of the state space. The response properties of any particular cell (e.g., the receptive field) describe the direction of the vector in the state space. The collection of cells mapping out the visual field forms an array of vectors mapping out the state space.

To understand the goal of sensory coding, it is proposed that one must determine the relation between the directions of the vectors (i.e., the response properties of the cells) and the "state space" of typical inputs for that sensory system. Throughout this paper, terms such as "the form of the state space" or "the state space of input" will be used. This represents a shorthand for "the regions within the state space where images have a relatively high probability." The state space always represents the space of all possible images. However, "state space of natural scenes" will refer to the region within this space that describes where natural scenes are likely to fall.

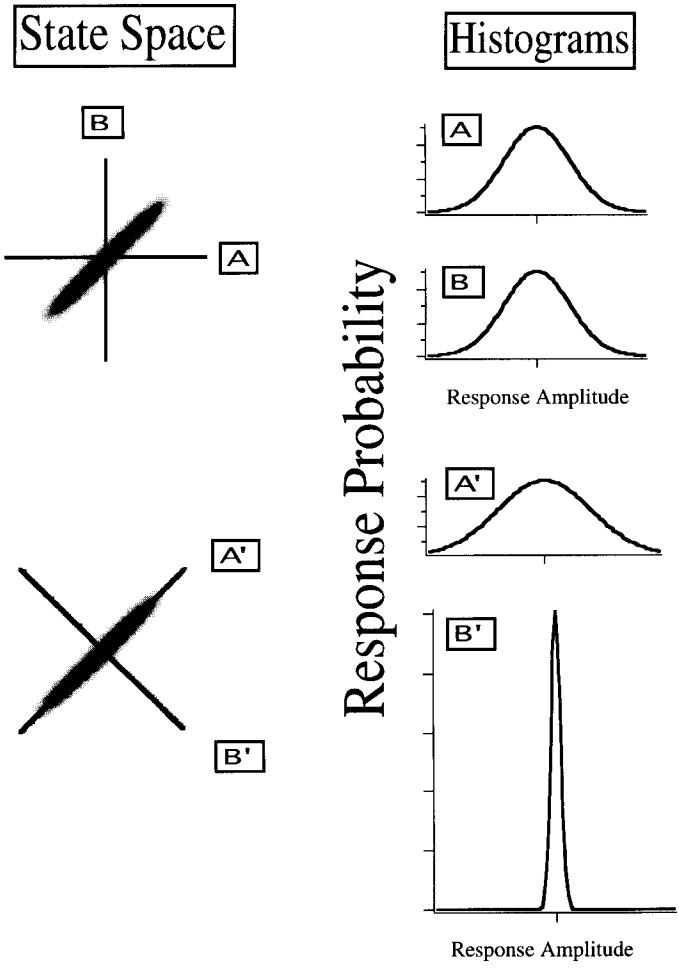
This study will investigate the relation between the spatial response properties of some of the cells in the visual pathway (e.g., the vectors described by the receptive field profiles) and the state space populated by natural scenes. The redundancy in the environment can take several forms. In the next sections, two forms of redundancy will be considered and we will explore two types of transform that can take advantage of this redundancy.

2.1 Correlations in State-Space. Let us consider a very simple example where a data set consists of only two independent variables (e.g., the intensities of two neighboring pixels). Figure 2 shows an example of the state space of a collection of two pixel images where the horizontal

Figure 2: *Facing page.* (a) An example of a two-dimensional state space. In this example, we have assumed that the axes represent the intensities of two pixels. The ellipse represents a population of probable inputs showing that Pixel A and Pixel B are highly correlated. However, in this example, there is little redundancy in the response histograms of the two pixels. Each shows a gaussian distribution with a relatively large variance. The histogram of activity for the two pixels is found by projecting the state space onto the two axes. (b) The transform where the new basis functions are represented by a rotation of the coordinate axes. In this new coordinate system, most of the variance in the data can be represented with only a single coefficient (A'). Removing B' from the code produces only minimal loss. This rotation of the coordinate system to allow the vectors to be aligned with the principal axes of the data is what is achieved with the process called Principal Component Analysis (PCA).

axis represents the intensity of Pixel A and the vertical axis represents the intensity of Pixel B. Any two-pixel image is represented as a unique point in the two-dimensional state space.

Figure 2a provides an example of one form of redundancy. In this case, the population of images produces a correlation in the outputs of the two pixels. Although the pixels are correlated, each pixel is contributing equal information about the image. In this case, it has been assumed that



the response behavior is normally distributed about the two axes. The histogram of activity (i.e., the probability distribution function) for the two pixels is found by projecting the state space onto the two axes. In this case, the activity of each pixel is normally distributed and the two have equal variance.

It is possible to transform the coordinate system to take advantage of the redundancy. Figure 2b shows the transform where the new basis functions are represented by a simple rotation of the coordinate axes where $A' = A + B$ and $B' = A - B$. To keep the basis vectors of normal length, the transform is

$$\begin{bmatrix} A' \\ B' \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} \tag{2.1}$$

In this new coordinate system, most of the variance in the data can be represented with only a single coefficient (A'). Removing B' from the code produces only minimal loss in the description of the input. This rotation of the coordinate system to allow the vectors to be aligned with the principal axes of the data is what is achieved with a process called Principal Component Analysis (PCA)—sometimes called the Karhounen–Loeve transform. Principal component analysis computes the eigenvalues of the covariance matrix (e.g., the covariance of the pixels in an image). The corresponding eigenvectors represent a hierarchy of orthogonal coefficients where the highest valued vectors account for the greatest part of the covariance.

By using only these highest valued vectors, the state space can be represented with a subset of the vectors and only minimal loss in the representation as measured in RMS error. As shown in Figure 2, the variance in A' has increased and the variance in B' has decreased markedly. Now, most of the variance of the input is coded in A'. In general, by removing those vectors with low variance, we end up with a state space that is "more packed." Thus, the new state space has less redundancy. Redundancy reduction is achieved by removing regions of the state space where the probability density is low. One can either remove low variance vectors from the representation or reduce the range that a vector can cover (i.e., reduce the dynamic range of the basis vector).

If we think of the data as forming an ellipse then PCA provides a method of finding the axes of the ellipse. Of course, this two-point transform is a rather restricted example. If one hopes to model the redundancy of natural scenes then one needs a high-dimensional space to describe the possible states. An n-pixel image requires an n-dimensional state space. Although the different forms of redundancy can become quite complex, it is not necessarily difficult to describe. It is possible to generalize the ellipse shown in Figure 3 to high dimensions where the region of high-probability images (i.e., high density) is described by

$$ax_0^2 + bx_1^2 + cx_2^2 + dx_3^2 + ex_4^2 + \cdots <= k \tag{2.2}$$

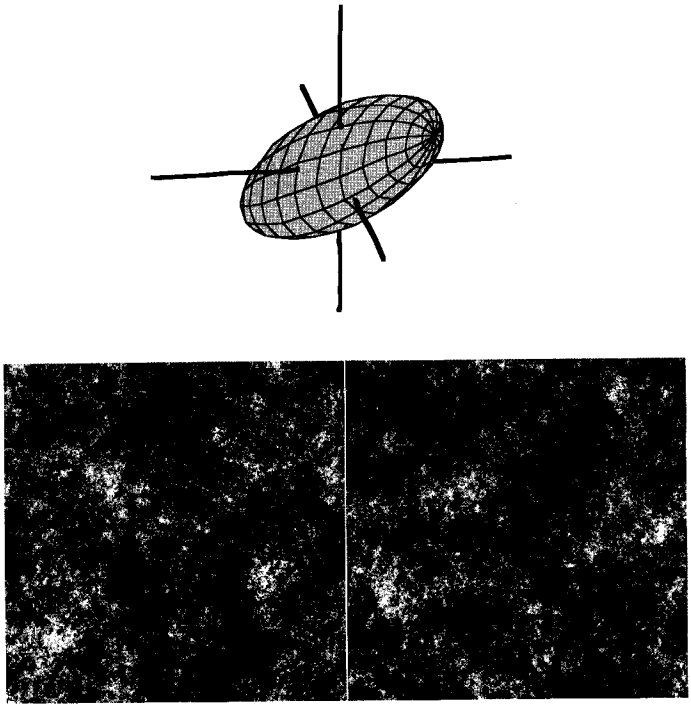


Figure 3: Ellipse and two filtered images. Examples of two images created by multiplying the spectrum of white noise (flat spectrum) by $1/f$. With images like this, all the redundancy is captured in the Fourier amplitude spectrum. The state space of images created in this fashion form a high-dimensional ellipse as described in equation 2.2. The axes of the ellipse are aligned with the vectors in the Fourier transform. Thus, the "principal components" and the vectors of Fourier transform are equivalent.

where

$$a \geq b \geq c \geq d \geq e \cdots$$

and $x_1, x_2, x_3 \ldots$ represent the axes of the ellipsoid. Whatever direction the axes are pointing, PCA can use the correlations in the data to find the directions of these axes and produce the vectors that correspond to the axes of the ellipsoid. The advantage of these particular vectors is that they provide a means of compressing image data with minimal RMS error. Just as the axes of the ellipse are ordered in equation 2.2, the eigenvectors are ordered in a sequence where the first few vectors account

for the highest proportion of the variance. If the goal is to compress an n-pixel image set using only m vectors where $m < n$, then the principal components (i.e., the principal axes) represent the optimal vectors for describing the data. As will be suggested in the next section, the state space representing natural scenes may not be well described by a high-dimensional ellipse. Nevertheless, no matter what the actual form of the state space, if the goal is to transmit a data set with a reduced number of vectors and with minimal RMS error, then PCA provides a means of finding the optimal set of vectors.[2]

2.2 Principal Components and the Amplitude Spectra of Natural Scenes. An interesting and important idea involves PCA when the statistics of a data set are stationary. Stationarity implies that over the population of images in the data set (e.g., all natural scenes), the statistics at one location are no different than at any other location.

Across all images $P(x_i \mid x_{i+1}, x_{i+2}, \ldots) = P(x_j \mid x_{j+1}, x_{j+2}, \ldots)$ for all i and j.

This is a fairly reasonable assumption with natural scenes since it implies that there are no "special" locations in an image where the statistics tend to be different (e.g., the camera does not have a preferred direction). It should be noted that stationarity is not a description of the presence or lack of local features in an image. Rather, stationarity implies that over the population, all features have the same probability of occurring in one location versus another.

When the statistics of an image set are stationary, the amplitudes of the Fourier coefficients of the image must be uncorrelated (e.g., Field 1989). Indeed, any two filters that are orthogonal over translation:

$$g(x) \cdot h(x - x_0) = 0 \qquad \text{for all } x_0$$

will have uncorrelated outputs in the presence of data with stationary statistics (Field 1993). Since this holds for the Fourier coefficients, the amplitudes of the Fourier coefficients will be uncorrelated.

This means that when the statistics of a data set are stationary then all the redundancy reflected in the correlations between pixels is captured by the amplitude spectra of the data. This should not be surprising since the Fourier transform of the autocorrelation function is equal to the power spectrum. Therefore, with stationary statistics, the amplitude spectrum describes the principal axes (i.e., the principal components) of the data in the state space (Pratt 1978). With stationary data, the phase spectra of the data are irrelevant to the directions of the principal axes.

As noted previously (Field 1987), an image that is scale invariant will have a well-ordered amplitude spectrum. For a two-dimensional image,

[2]It should be noted that PCA finds only the optimal linear solution. It is always possible that for a given data set there exists a nonlinear transform that will provide better compression than the vectors provided by PCA.

the amplitudes will fall inversely with frequency (i.e., a $1/f$ spectrum). Natural scenes have been shown to have spectra that fall as roughly $1/f$ (Burton and Moorhead 1987; Field 1987, 1993; Tolhurst *et al.* 1992). If we accept that the statistics of natural images are stationary, then the $1/f$ amplitude spectrum provides a complete description of the correlations in natural scenes. The amplitude spectrum certainly does not provide a complete description of the redundancy in natural scenes, but it does describe the relative amplitudes of the principal axes.

We can ask what an image would look like if all the redundancy in the image set was described by a $1/f$ amplitude spectrum. Or in terms of the state space, we can ask what images would look like if the probability density was described by a high-dimensional ellipsoid where the principal axes of the ellipsoid are the Fourier transform. Figure 3 shows examples of two such images. They are created by multiplying the spectrum of white noise (flat spectrum) by $1/f$. Although these images have amplitude spectra similar to natural scenes (and therefore the same principal components), they clearly do not look like real scenes. They do not contain any of the local structure like edges and lines found in the natural environment. As proposed in the next section, images that have such local structure are not described well by a high-dimensional ellipsoid.

A number of recent studies have discussed the similarities between the principal components of natural scenes and the receptive fields of cells in the visual pathway (Bossomaier and Snyder 1986; Atick and Redlich 1990, 1992; Atick 1992; Hancock *et al.* 1992; Baddeley and Hancock 1991; MacKay and Miller 1990; Derrico and Buchsbaum 1991). And there have been a number of studies that have shown that under the right constraints, units in competitive networks can develop large oriented receptive fields (e.g., Lehky and Sejnowski 1990; Linsker 1988; Sanger 1989). Indeed, it has been noted that networks like Linsker's work off the induced correlations caused by the overlapping receptive fields in the network and the network produces results similar to the principal components (MacKay and Miller 1990). And it has been noted that Hebbian learning can, under the right conditions, find the principal components (Oja 1982; Foldiak 1989; Sanger 1989).

This appears to pose a dilemma. If the principal components of images with stationary statistics are equivalent to the Fourier transform, shouldn't the derived receptive field profiles of the appropriate Hebbian network look like the Fourier transform if presented with natural scenes? Not necessarily. The two-dimensional Fourier transform of natural scenes shows considerable symmetry. The amplitude spectra of natural scenes fall as approximately $1/f$ at all orientations. Therefore, at any frequency there exists a range of orientations that are likely to have similar amplitude. When a number of Fourier coefficients have the same amplitude,

there will exist a wide range of linear combinations that will account for equal amounts of the covariance (i.e., the solution is degenerate). Therefore, linear combinations of these equivariant vectors will also account for equal amounts of the covariance.

Recently, Baddeley and Hancock (1991) and Hancock *et al.* (1992) calculated the principal components of a number of natural scenes directly. When gaussian windowed, the first few coefficients do show some resemblance to cortical receptive fields—but this is expected since a gaussian-windowed low-frequency sinusoid (the first Fourier coefficients with the greatest amplitude) will produce the popular "Gabor function" shown to provide good models of cortical receptive fields (e.g., Field and Tolhurst 1986). Past the first 3 or 4 Fourier coefficients, the receptive field profiles no longer look like cortical receptive field profiles. Instead, the profiles are substantially different from those found in the mammalian visual system.

Figure 4B shows an example of the two-dimensional amplitude spectrum of a natural scene. On average, a ring of frequencies around the origin will have the same amplitude. This produces degeneracy in the set of possible solutions. The principal components can consist of any linear combination of these frequencies. That is, an equivalent solution for the principal components may consist of an ortho-normal transformation of these vectors. Figure 4C shows four spatial frequencies that would be expected to account for equal amounts of the covariance in natural scenes. The four lower "receptive fields" represent a rotation of the other receptive fields. For a natural scene with a $1/f$ spectrum and stationary statistics, the vectors in either 4C or 4D will serve equally well as principal components. Thus, even if the statistics of the input are stationary, the Fourier transform is not necessarily the only solution.

It is important to recognize that if the statistics of the input are stationary then there is no reason for Principal Components Analysis to produce localized receptive fields. In Fourier terms, a function is localized because the phases of the different frequencies are aligned at the point where the function is localized. Cortical simple cells, for example, have been shown to have a high degree of phase alignment at the center of the receptive field (Field and Tolhurst 1986). The large low-frequency receptive fields may contain only a few spectral components. However, the small high-frequency receptive fields have relatively broad bandwidths and hence have a large number of spectral components, all aligned near the center of the receptive field. Randomizing the phases of a localized function will distribute the energy across the image. Since the covariance matrix does not capture the phases (if the statistics are stationary), Principal Components Analysis cannot produce localized receptive fields.

If the primary constraints of the code are to represent the input with reduced dimensionality, then the codes should not be capable of producing the small high-frequency orientation selective receptive fields. Atick

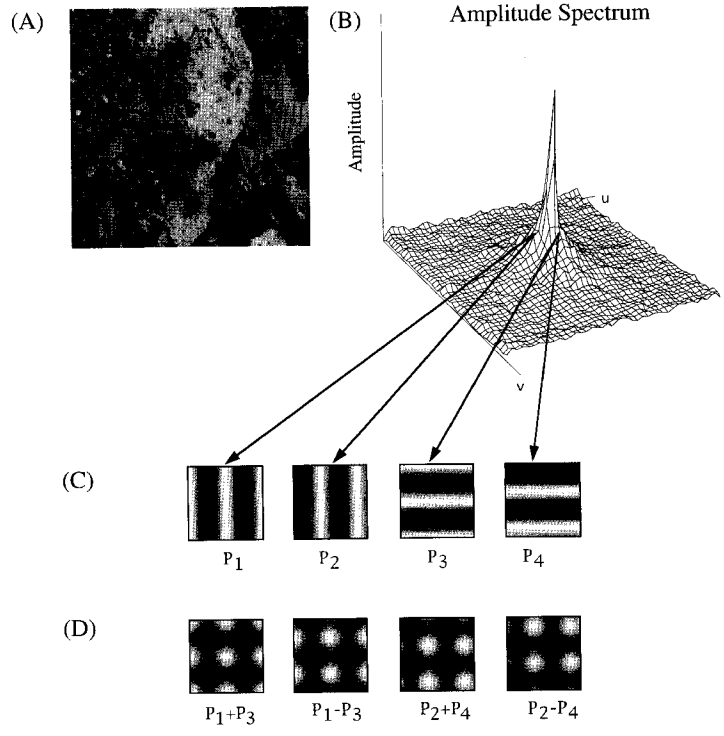


Figure 4: An example of a scene (A) and its two-dimensional amplitude spectrum (B). For a population of images scenes with stationary statistics, the amplitude spectrum describes the principal components. However, in the two-dimensional amplitude spectrum of natural scenes there is considerable symmetry with similar frequencies at different orientations likely to have the same amplitude (C). This means that appropriate combinations of the Fourier coefficients with equal amplitude will account for equal amounts of the covariance. (D) An example of a rotation of the Fourier vectors that should account for equal amounts of the covariance. With stationary statistics, the phase of the phase spectrum of the input is irrelevant to the principal components. Thus, Principal Components Analysis (PCA) cannot result in localized receptive fields without forcing constraints on the phase.

and Redlich (1990, 1992) have suggested that the spatial frequency tuning of retinal ganglion cells is well matched to the combination of amplitude spectra of natural scenes and high-frequency quantal limitations found in the natural environment. This is an important finding with regards to

the amplitude spectrum of the receptive field profile. However, a receptive field will be localized only if its phase spectrum is aligned, so this approach cannot explain why the receptive fields are localized. It has been proposed that it is the phase spectrum of natural scenes that must be considered if one is to account for the localized nature of receptive fields (Field 1989, 1993). Indeed, although there are numerous reports of networks that produce hidden units with oriented receptive fields, I have found no published report of a network that can produce the small localized receptive fields without specifically forcing these locality constraints on the network (e.g., forcing a small gaussian window). It has even been suggested that competitive learning is, in general, inappropriate for producing the small high-frequency receptive fields (Webber 1991). The receptive field profiles of the hidden units are always the same size as the window of the network.

In the next section it will be proposed that the principal components of natural images do not describe the important forms of redundancy. To summarize, PCA and compact coding have several problems in accounting for receptive field profiles of cells in the retina and primary visual cortex.

1. If one accepts that the statistics of natural scenes are stationary, then the principal components are dependent on only the amplitude spectra of the input. The phase spectra of the inputs are irrelevant.

2. Because of the symmetry in the amplitude spectra of natural scenes, the principal components are likely to have a degenerate solution. The resulting "receptive fields" will consist of an unconstrained collection of orientation components with random phase. This will be most apparent at higher frequencies where there exists a wide range of orientation components at each frequency.

3. A function is localized only when the phases of the Fourier coefficients are aligned. Thus, when the statistics are stationary, the Principal Components Analysis will not produce localized receptive fields because the phase spectrum is not constrained.

4. Since the principal components are not dependent on the phase spectra of the input, the principal components will not reflect the presence of local structure in the data.

In the next section, it will be proposed that the shape of the state space describing natural scenes is not elliptical and this nonelliptical state space provides a clue as to why the mammalian visual system represents spatial data as it does. To account for receptive field profiles in primary visual cortex, we consider a different theory regarding the goal of visual coding and we consider a different form of redundancy found in natural scenes.

However, before discussing sparse coding techniques, it should be noted that the PCA represents an important technique for reducing the

number of vectors describing the input—when it is possible. For example, in the color domain, it has been found that the first three principal components can account for much of the variance in the spectra of naturally occurring reflectances. This has led investigators to suggest that the three cone types found in primate vision provide an efficient description of our chromatic environment (e.g., Buchsbaum and Gottschalk 1983; Maloney 1986).

In general, the principal components can be used to determine which dimensions of the state space are needed to represent the data. If a reduced dimensionality is possible, the principal components will be able to tell you that. However, once the dimensions have been decided, it is proposed that other factors must be considered to determine the best choice of vectors to describe that space.

2.3 State Spaces That Allow Sparse Coding. Consider a two-pixel data set with redundancy like that shown in Figure 5. The data are redundant since the state space is not filled uniformly. However, this data set has some interesting properties. We can think of this data set as a collection of two types of images: one set with pixels that have high positive correlation and one set with high negative correlation.

For this data set, there will be no correlation between Pixel A and Pixel B. Furthermore, there is no "high-order" redundancy since there are only two vectors. This is an example of a type of second-order redundancy that is not captured by correlations. However, the same transformation performed as before (i.e., a rotation) produces a marked change in the histograms of the basis functions A' and B'. This particular data set allows a "sparse" response output. Although the variance of each basis function remains constant, the histogram describing the output of each basis function has changed considerably. After the transformation, vector A' is high or vector B' is high but they are never high at the same time. The histograms of each vector show a dramatic change. Relative to a normal distribution, there is a higher probability of no response and a higher probability of a high response, but a reduction in the probability of a mid-level response.

This change in shape can be represented in terms of the kurtosis of the distribution where the kurtosis is defined as the fourth moment according to

$$K = 1/n \sum[(x - x)^4/\sigma^4] - 3 \tag{2.3}$$

Figure 6 provides an example of distributions with various degrees of kurtosis. In a sparse code, any given input can be described by only a subset of cells, but that subset changes from input to input. Since only a small number of vectors describe any given image, any particular vector should have a high probability of no activity (when other vectors describe the image) and a higher probability of a large response (when the vector is part of the family of vectors describing the image). Thus, a sparse code

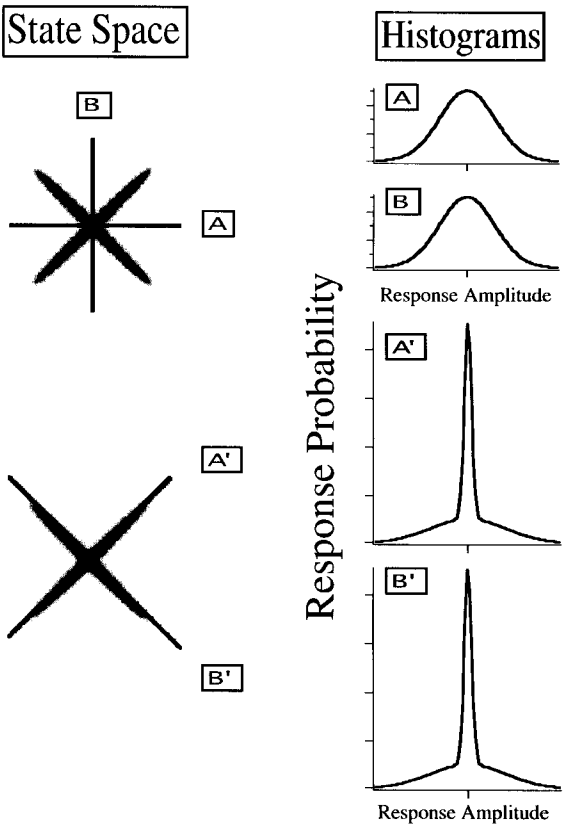


Figure 5: As in Figure 2, this shows an example of the state space of a population of two-pixel images. The data set are redundant since the state space is not filled uniformly. However, with these data, there are no correlations in the data and, therefore, there are no principal axes that can account for a major component of the variance. In such a data set, there is no way to take advantage of the redundancy by reducing the dimensionality. The right of the figure shows how the response distribution changes if the vector space is transformed to allow the vectors to be aligned with the axes of the data. In this case, each of the vectors maintains the same variance. However, the shape of the distribution is no longer gaussian (high entropy). Instead, the response distribution shows a high degree of kurtosis (lower entropy–higher redundancy).

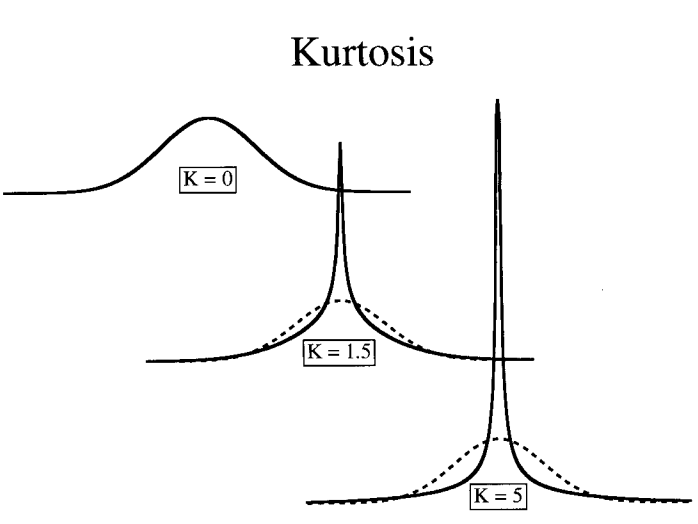


Figure 6: Examples of three levels of kurtosis. Each of the distributions has the same variance. A gaussian distribution has minimal redundancy (highest entropy) for a fixed variance. The higher the kurtosis, the higher the redundancy. With high kurtosis, there is a higher probability of a low response or a high response with a reduced probability of a mid-level response.

should have response distributions with high kurtosis. Although kurtosis appears to capture this property of a distribution with both high and low variances, one should not presume that kurtosis necessarily represents the optimal measure for defining a sparse code. At this time, we consider it a "useful measure." We will return to this point later.

As we move to higher dimensions (e.g., images with a larger number of pixels), we might consider the case where only one basis vector is active at a time (e.g., vector 1 or vector 2 or vector 3 ...):

$$ax_1 \cup ax_2 \cup ax_3 \cup ax_4 \ldots \tag{2.4}$$

In this case, each image can be described by a single vector and the number of images equals the number of vectors. However, this is a rather extreme case and is certainly an unreasonable description of most data sets, especially natural scenes.

When we go to higher dimensions, there exist a wide range of possible shapes that allow sparse coding. Overall, the shape must require a large

set of vectors to describe entire population of possible inputs, but require a subset of vectors to describe any particular input.

$$\text{Image} = \sum_{i}^{n} aV_i \qquad \text{where } n < m \tag{2.5}$$

where m is the number of dimensions required to represent all images in the population (e.g., all natural scenes).

For example, with a three-pixel image where only two pixels are nonzero at a time, it is possible to have:

$$ax_1 + bx_2 \cup ax_2 + bx_3 \cup ax_1 + bx_3 \tag{2.6}$$

This state space consists of three orthogonal planes. If the data fall in these three planes with equal probability, then there will be no correlations in the data and therefore the principal components will not provide the axes of the planes (indeed, there are no *principal* axes of the state space). However, by choosing vectors aligned with the planes (e.g., x_1, x_2, x_3), it is possible to have a code in which only two vectors are nonzero for any input. In some situations, the principal components may even point in the wrong direction for achieving a sparse code. Figure 7A shows an example where the data fall along slightly nonorthogonal lines. In this case, there is a positive correlation in pixels. As with the ellipse shown in 7B, the principal components lie along the diagonals rather than the axes of the data. However, selecting vectors aligned with the data can produce histograms with positive kurtosis even though these are nonorthogonal. Indeed, it is important to recognize that the optimal sparse representation of a data set is not necessarily an ortho-normal representation.

Figure 7C shows a three-dimensional variation that has some interesting properties. If the state space forms a hollow three-dimensional cone like that shown, the first principal component will fall along the major axis of the cone. Indeed, Figure 7D shows an ellipsoid with the same principal axes as 7C. However, in the case of the ellipsoid, all the redundancy is captured by the principal components (i.e., the principal axes define the ellipse). In the case of the cone, the principal components fail to exploit some of the most interesting aspects of the data.

The reason for using the cone as an analogy is that the vectors along the tangents of the cone extending from the origin can allow a sparse representation but the probability distribution is locally continuous. Also, we have found that the cone provides a reasonable description of how a localized function (e.g., a gaussian blob) will be distributed within the state space when the feature is varied in position and amplitude. However, a three-dimensional state space cannot begin to describe the full complexity of the redundancy in natural scenes, so the cone should be considered as only a crude example. The precise form within the state space is not critical to the argument. For a sparse code to be possible, a

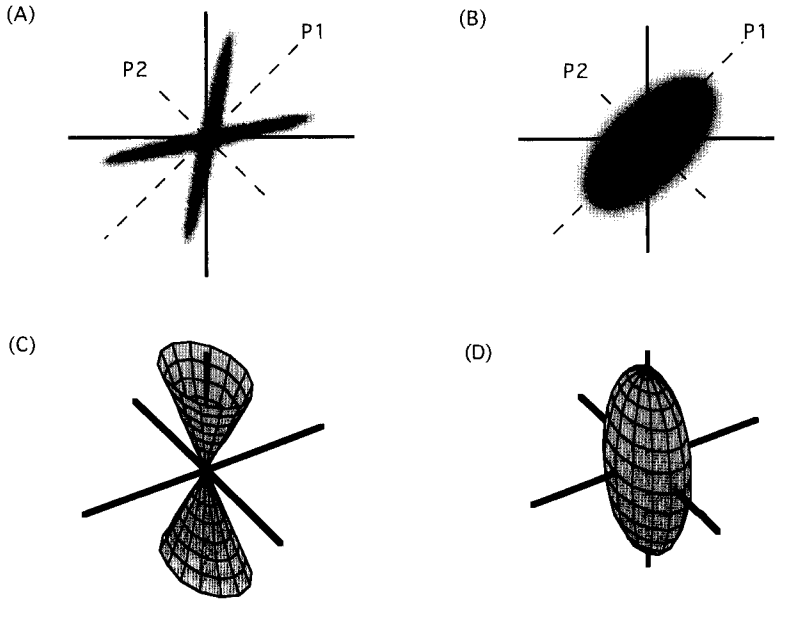


Figure 7: (A) An example of a data distribution where the principal components lie along the diagonal but where such components are ineffective at producing a sparse code. (B) A data set with the same principal components (an ellipse). (C, D) Two three-dimensional examples of state spaces that have the same principal components. State spaces like that shown in A and C allow sparse coding. State spaces like B and D do not. It is suggested that the local structure available in natural scenes produces a high-dimensional state space analogous to this cone rather than the ellipse.

large set of vectors must be required to describe all points on the form (all possible images) but the form must require only a subset of vectors to describe any particular point in the space (any particular image)—equation 2.5.

There will always exist a wide range of shapes of the probability density function that can result in the same principal components. The possibility of finding a sparse code depends on this shape. For an ellipsoid, it is not possible to produce a sparse output. In this paper, it is proposed that the signature of a sparse code is found in the kurtosis of the response distribution. A high kurtosis signifies that a large proportion of the cells is inactive (low variance) with a small proportion of the cells describing the contents of the image (high variance). However, an effective sparse code is not determined solely by the data or solely by

the vectors but by the relation between the data and the vectors. In the next section we take a closer look at the vectors described by the receptive field profiles of two types of visual cells and look at their relation to natural scenes.

Before we begin this description, it should be noted that the results of several studies suggest that cells with properties similar to those in the mammalian visual cortex will show high kurtosis in response to natural scenes. In Field (1987), visual codes with a range of different bandwidths were studied to determine how populations of cells would respond when presented with natural scenes. It was found that when the parameters of the visual code matched the properties of simple cells in the mammalian visual cortex, a small proportion of cells could describe a high proportion of the variance in a given image. When the parameters of the code differed from those of the mammalian visual system, the response histograms for any given image were more equally distributed. The published response histograms by both Zetzsche (1990) and Daugman (1988) also suggest that codes based on the properties of the mammalian visual system will show positive kurtosis in response to natural scenes. Burt and Adelson (1983) noted that the histograms of their "Laplacian pyramids" showed a concentration near zero when presented with their images and suggested that this property could be used for an efficient coding strategy. In this paper, it is emphasized that these histograms have these shapes because of the particular relation between the code and the properties of the images. In particular, when the input has stationary statistics, it is the phase spectrum that describes the redundancy required for sparse coding.

In previous work by the author (Field 1993) complex correlations in natural scenes were studied to determine the extent to which the phases were aligned across the different frequency bands. The extent of the phase alignment across neighboring frequency bands was found to be proportional to frequency (i.e., at high frequencies, the phases were aligned across a broader range of frequencies). In other words, at low frequencies, local structure tends to be spatially extended while at high frequencies the local structure tends to be more spatially limited. It was noted that this particular alignment in phases would be expected if these scenes were scale invariant with regards to both their amplitude spectra and their phase spectra. It was proposed that, to a first approximation, natural scenes should be considered as a sum of bandlimited, self-similar "phase structures" (points of phase alignment). The spectral extent of this phase alignment was found to be well matched to the bandwidths of cortical simple/complex cells (Field 1989, 1993) and it was suggested this property is what allowed the visual system to produce a sparse output. We will return to this discussion when we consider synthetic images that produce high kurtosis in wavelet transforms. First, however, this study considers a more direct test of sparse coding by looking at the kurtosis

of the response distributions of two types of visual cells when presented with natural scenes.

3 Receptive Fields in the Mammalian Visual Pathway as Vectors

Just as it is possible to rotate the coordinate axes in a wide variety of ways, there exist a wide range of transforms that are capable of representing the information in an n-dimensional data space. The Fourier transform represents one example of a rotation. Gabor (1946) described a family of transforms that were capable of representing one-dimensional waveforms. These ideas were extended by Kulikowski *et al.* (1982) to include transforms in which the bandwidths increase proportionally to frequency. Such transforms, which have recently become known as "wavelet transforms" (e.g., Mallatt 1989), consist of arrays of self-similar basis functions that differ only by translations, dilations, and rotations of a single function. Although much of the work on wavelet transforms has been devoted to the development of ortho-normal bases (e.g., Adelson *et al.* 1987; Daubechies 1988) the coding with these transforms has found a wide variety of applications (e.g., Farge *et al.* 1993) from image processing to the representation of turbulence. For our purposes, however, it is important to recognize that such transforms can be considered as rotations of the coordinate system.

Figure 8 shows one part of a two-dimensional implementation of a wavelet transform. Wavelets based on the gaussian modulated sinusoid (i.e., Gabor functions) have proved to be popular models of the visual cortex (Watson 1983; Daugman 1985; Field 1987). Although the Gabor function has found some support in the physiology (Webster and DeValois 1985; Field and Tolhurst 1986; Jones and Palmer 1987), other functions have been proposed such as the Cauchy (Klein and Levi 1985) and the log-Gabor used in this study (Field 1987). Transforms based on these functions capture some of the basic properties of cells in the mammalian visual cortex. (1) The receptive fields are localized in space and are bandpass in frequency but overlap in both space and frequency. (2) The spatial frequency bandwidths are constant when measured on logarithmic axes (in octaves) resulting in a set of self-similar receptive fields. (3) They are orientation selective. These basic properties provide the basis for a number of models of visual coding and the model used in this study does not have any major components that differ from previous models (e.g., Watson 1983; Daugman 1985; Field 1987).

It is important to recognize that these transforms are not ortho-normal. First, the basis functions are not quite orthogonal. Second, in this study (as in Field 1987) the vector length of the transform increases with frequency (i.e., the peak of the spatial frequency response is constant and the bandwidth increases). As previously noted, this results in a code with distributed activity in the presence of images with $1/f$ spectra like

that found in natural scenes (Field 1987; see Brady and Field 1993 for a discussion of vector length). This means that the variance of the filtered images will be roughly the same magnitude at the different frequencies.

It must also be emphasized that the functions used in this study to model visual neurons are highly idealized versions of actual cells. Both the models of retinal neurons and cortical simple cells involve codes in which all the cells have the same logarithmic bandwidth. Visual neurons are known to show a range of bandwidths which average around 1.5 octaves but become somewhat narrower at higher frequencies (e.g., Tolhurst and Thompson 1982; DeValois *et al.* 1982). However, the most important difference is that the "cells" in this study do not have many of the known nonlinearities found in cortical simple cells (e.g., end-stopping, cross-orientation inhibition, etc.). Indeed, the response histograms are allowed to go negative in our modeled cells while actual cortical cells can produce only the positive component of the histogram. We will return to this discussion of nonlinearities later. However, it should be remembered that the simple cells that are modeled are only rough approximations to the cells that are actually found in the visual cortex.

The codes in this study represent images with arrays of basis functions that are localized in the two-dimensional frequency plane as well as the two-dimensional image plane. Figure 8 shows one example of the division of the two-dimensional frequency plane and the corresponding

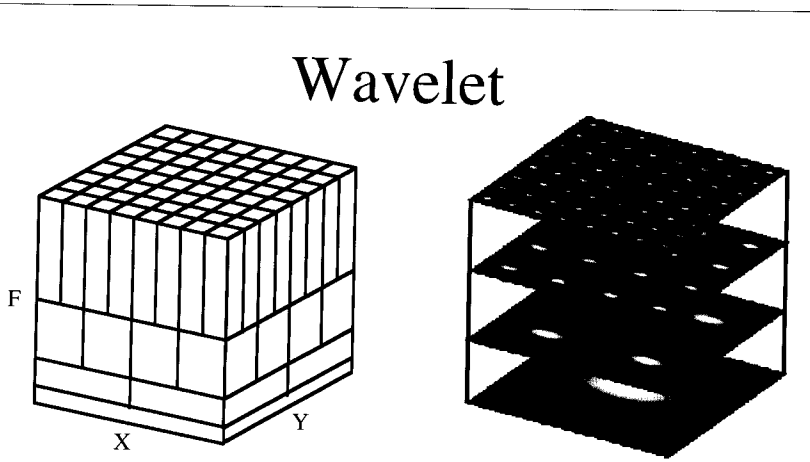


Figure 8: Two-dimensional wavelet. Three-dimensional information diagrams for a two-dimensional wavelet. The information diagrams are actually four-dimensional (u, v, x, y) but we have limited the diagram to the representation of a single orientation, to allow a graphic description.

representation in space for one orientation and one phase. Each basis function is thus localized in the four dimensions of x, y, u, v. In this particular description, the spatial sampling grid is rectangular, but this is not a requirement. One requires a four-dimensional plot to cover the full 2-D space by 2-D frequency trade-off, but this is difficult to depict graphically. If we consider only a single orientation for the purposes of the display, it is possible to show the space-frequency trade-off using the representations shown in the figure.

A comment should also be made with regards to the phase spectra of these filters. In line with previous studies (Field 1987, 1989, 1993), the oriented wavelet transform uses a pair of even- and odd-symmetric filters (i.e., filters in quadrature) at each location. Since these receptive fields are localized, they can be defined in Fourier terms by an alignment in the phase at the center of the receptive field (e.g., Field and Tolhurst 1986; Field 1993). Even complex cells have localized receptive fields, so to some extent, they must also be phase selective. One recent model of complex cells describes the response of a complex cell as a vector sum of two quadrature-phase cortical cells (Field 1987; Morrone and Burr 1988). By this model, complex cells detect an alignment of phases at particular locations of the visual field but the response does not depend on the absolute phase of that alignment (e.g., sine versus cosine). However, in this study the intention is not to model all of the various forms of nonlinearity found in the mammalian visual system. Rather, the goal is to show that the basic properties of visual neurons (i.e., orientation tuning, spatial frequency tuning, and position selectivity) produce a sparse representation of natural scenes.

4 Kurtosis in the Response to Natural Scenes

In the previous sections, it was proposed that if the visual system is producing a sparse representation of the environment then the histograms describing the response of the visual system should have a high kurtosis. In this section, these two approaches will be contrasted by investigating the histograms of the cells in several visual codes.

4.1 Method.

Images. The images used in the following sections consist of digitized photographs of the natural environment. They consist of 55 scenes, six of which are shown in Figure 9. The only photographic restriction was that the images have no man-made features (buildings, signs, etc.) since these tend to have different statistical structures (e.g., a higher probability of long straight edges). Images were photographed with Ilford XP1 film using a 35-mm camera. Photographic negatives were scanned using a Barneyscan digitizer that provided a resolution of 512×512 pixels per

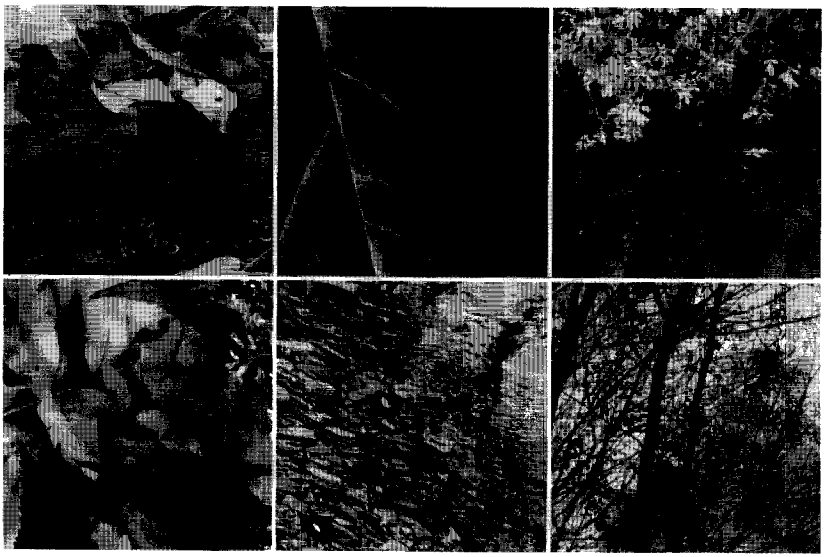


Figure 9: Examples of six images used in this study. Each of the images was calibrated for the intensity response of the film.

picture with 256 gray levels. The images were calibrated for luminance using Munsell swatches allowing the pixel values to have a linear relation to the image intensities in the original scene. The optics of the camera were taken into account by determining the response to thin lines and correcting for the changes produced in the amplitude spectrum (i.e., the modulation transfer function). Before analysis, the log of the image was determined and the calculations were based on this simple nonlinear transform of the image. This allows the units to respond in terms of contrast (i.e., ratios of intensities) rather than to intensity differences (i.e., amplitude) since:

$$\log(a) - \log(b) = \log(a/b) \tag{4.1}$$

This is believed to produce a more accurate representation of cells in the visual pathway since cells are know to produce a more linear response to contrast (e.g., see Shapley and Lennie 1985) rather than amplitude. It was also found that the pixel histograms of some of the images had spuriously high kurtosis values (several images had $K > 20$ before the log transform) because of a few bright points in the image such as bright sky filtering through a tree. One should note film normally uses a compressive gamma, so studies that do not calibrate their film end up with

similar high-intensity compression without intending to do so. Also, this collection of images did not contain large blank regions (e.g., an image of half sky). Such images result in a large number of inactive cells that will also increase the kurtosis for any code using local operators. These efforts were made to minimize any biases that might have been present in our image collection and allow a more accurate comparison of the different codes.

4.2 Population Activity. Population measures were based on the outputs of arrays of cells determined by convolving the images with the appropriate filters. The methodology is similar to that of Field (1987). However, in this study results are provided for a fixed number of spatial frequencies and orientations. Two types of filters are compared. The first is a center surround operator made from a difference of gaussians (DOG) where the surround has three times the width of the center.

$$g(x, y) = 9e^{-(x^2+y^2)/2(\sigma)^2} - e^{-(x^2+y^2)/2(3\sigma)^2} \tag{4.2}$$

For the DOG calculations, each image was convolved with two filter sizes (the spectrum peaked at 20 and 40 cycles/picture). The second type is the oriented "log-Gabor" described previously (Field 1987). Radially the function has a log-normal spectrum:

$$G(k) = e^{-\ln(k/k_0)^2/2[\ln(\sigma/k_0)]^2} \tag{4.3}$$

and is gaussian along the orthogonal axis.

As in Field (1987) the local phase was represented by a pair of filters in quadrature (i.e., even-symmetric and odd-symmetric). Thus, each image was convolved with filters at two spatial frequencies (k_0 = 20 and 40 cycles/picture), four orientations (10, 55, 100, and 145°), and two phases. For each bandwidth selected, this requires $2 \times 2 \times 4 = 16$ convolutions per image to obtain the total response histograms at the 8 spectral locations. Although these codes do not provide a complete representation of the image, the results provide a relatively direct method of comparing the population responses of different codes and eliminate problems of filters near the edges of the spectrum.

With the wavelet codes, different spatial frequency bandwidths are compared. With the orientation bandwidth set to 40° orientation tuning (full width at half height), the spatial frequency bandwidth was set to one of 8 spatial frequency bandwidths ranging from 0.5 to 8.0 octaves in logarithmic steps where the bandwidth is defined as

$$B_{\text{oct}} = \ln_2(k_2/k_1) \tag{4.4}$$

where k_1 and k_2 define the lower and upper frequencies defining the width at half-height.

For each image and each bandwidth, the kurtosis of the distribution was determined by calculating the histogram of the pixels from the 16 filtered images. Near the edges of each of the filtered images, the response

of any basis function can produce spurious results. To remove these effects, only the data from the central 180×180 region was used in the analysis. Thus for each image, the histograms were based on a total of $16 \times 180 \times 180 = 518,400$ samples for the wavelet and $2 \times 180 \times 180 = 97,200$ samples for the DOG. One should note that these samples are not completely independent samples since the images are not sampled in proportion to the size of the basis function as in Field (1987). However, this will have minimal effect on the overall histograms but implies that the low-frequency channel (e.g., at 20 cycles/picture) will have histograms based on fewer independent samples than the high-frequency channels with smaller receptive fields (e.g., 40 cycles/degree).

4.3 Results. Figure 10A shows the histogram for image 1 with a single condition (spatial frequency bandwidth: 1.4 octave–20° orientation bandwidths). With these filters, the response of any function is as likely to be positive as negative. One can see that the general form of the histogram is much like that shown in the kurtosis plots described previously. Indeed, this pattern has a kurtosis of 6.2. Figure 10B shows the results describing the kurtosis for the original image, the DOG function, and the wavelet when the bandwidth was fixed at 1.4 octaves. These results show that both the DOG function and the oriented wavelet show increased kurtosis. Figure 10C shows the results for 20 $1/f$ noise patterns like those shown in Figure 3. These images have random phase spectra but similar amplitude spectra to natural scenes (and therefore the same principal components).

Figure 11A shows the results describing the kurtosis of the response histograms for the wavelet as a function of the spatial frequency bandwidth. Results are shown for the six images shown in Figure 3. Figure 11B provides a histogram of the bandwidths that produce peak kurtosis for all 55 of our natural scenes. These results show that the bandwidth that produces the maximum kurtosis typically falls in the range of 1.0 to 3.0 octaves. This falls within the range of bandwidths that are most commonly found in the mammalian visual system (e.g., Tolhurst and Thompson 1982).

5 Discussion

The results shown above support the notion that codes that have similar properties to that found in the mammalian visual system are effective at producing a sparse representation of the natural scenes. The results in Figure 10 suggest that as one moves from the retina to the cortex, the kurtosis of the distribution increases. Higher levels appear to be more capable of taking advantage of the redundancy in natural scenes.

The redundancy that is captured by these codes is not due to the amplitude spectra of the images. Figure 10B shows that images with

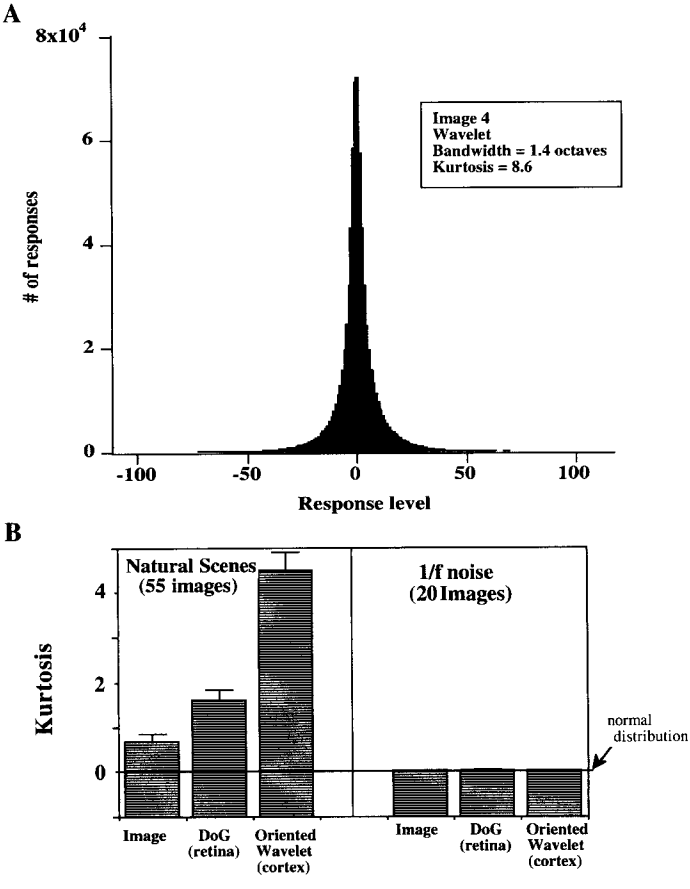


Figure 10: (A) The response histogram for a single image using the wavelet. The histogram is based on the response at four orientations at spatial frequencies of 20 and 40 cycles/picture and two phases (even- and odd-symmetric). (B) The mean kurtosis for the images calculated from the pixels of the original image, after convolution with the difference of gaussians and after convolution with the wavelet. The bottom right shows the same processes with 20 $1/f$ noise patterns that have approximately the same amplitude spectra as that in the natural scenes.

similar amplitude spectra as natural scenes but random phase spectra produce histograms with kurtosis values of 0.0 (normal distributions). Since these images differ only in their phase spectra, it is clear that the

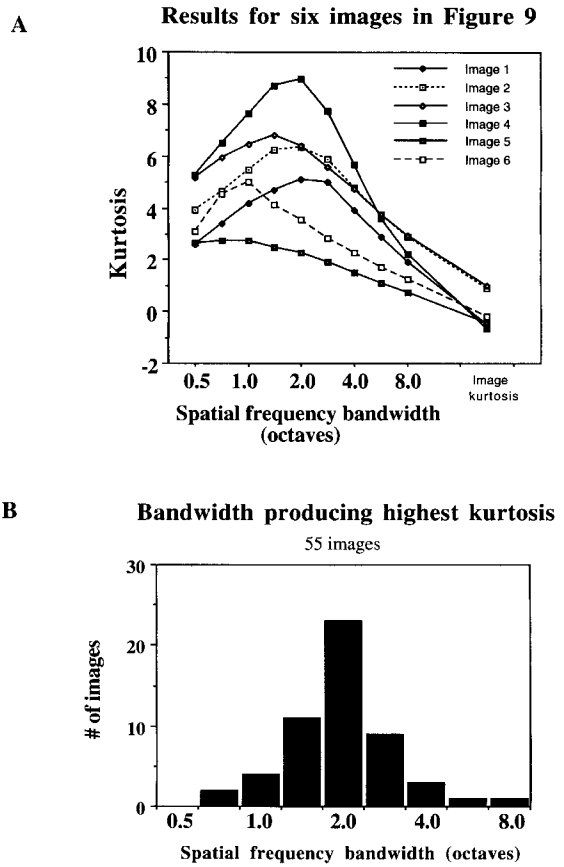


Figure 11: Distribution of peak kurtosis as a function of bandwidth. (A) The kurtosis of the response distributions for the six images shown in Figure 9 as a function of the spatial frequency bandwidth (orientation bandwidth is fixed at 40°). For each of the 55 images, the kurtosis of the response distribution was determined and the bandwidth that produced the highest kurtosis was calculated. (B) The distribution of spatial frequency bandwidths that produced the highest kurtosis.

phase spectra of natural scenes play a major role in determining whether the wavelet transform will produce a sparse output.

5.1 Synthetic Images That Produce High Kurtosis. In Field (1993), it was proposed that self-similar transforms are effective at producing sparse representations with natural scenes because such scenes have sparse, scale-invariant phase spectra. That is, the phases in such scenes are aligned at a relatively small number of locations in space/frequency and the spectral extent of the alignment is proportional to frequency. It is a relatively simple process to synthesize images that have such properties. Figure 12 provides examples of two such images. These particular images were created using the equation:

$$\sum_{m=0}^{a} \sum_{n=1}^{\beta\sigma^2 m} \omega_{nm} g[\tau\sigma^m x - x_{nm}, \tau\sigma^m y - y_{nm}] \tag{5.1}$$

where a is the number of scales in the image, β controls the density of the elements at each scale, σ controls the spectral distance between scales, and τ controls the relative spatial extent of the function at each scale. Figure 12A and B shows two densities where $g(x, y)$ is a gaussian modulated sinusoid. Each of the functions is added with the same average spatial amplitude independent of scale (ω_{nm} has the same average amplitude at all scales). If each function is treated as a vector, then the vector length of each function decreases as the scale increases (proportional to $1/f$). By increasing the number of vectors in proportion to the square of the frequency and adding them in random positions, this technique produces a scale-invariant $1/f$ amplitude spectrum but in which the phases are locally aligned (See Appendix).

Each of these images can be thought of as an "inversion" of the visual code. In a sparse-distributed code, only a subset of the cells is active but each cell has the same probability of activity. The method described in equation 5.1 is roughly equivalent to assigning a low probability to each vector in a wavelet code and summing the sparse set of vectors together. The state space of these images is analogous to the state space shown in Figure 4 but, of course, in higher dimensions. We are currently attempting to provide a better description of the high-dimensional state space and believe that a set of high-dimensional cones of different diameters is a useful analogy. A subset of vectors from the wavelet code can represent this image because it is created by a subset of vectors from the wavelet code. Although these images do not have all the redundancy of natural scenes, they provide a good first approximation and certainly provide a better approximation than the $1/f$ noise patterns shown in Figure 3.[3]

The results in Figure 10 suggest that although the DOG patterns show a significant increase in kurtosis, the oriented wavelet results in higher kurtosis. This suggests, that in natural scenes, when local structure is present (i.e., the phases are aligned), the structure tends to be oriented. This is not mathematically necessary. It is possible to create images

[3]As the density approaches 0.5, the images show the same structure as the $1/f$ noise patterns shown earlier in Figure 3.

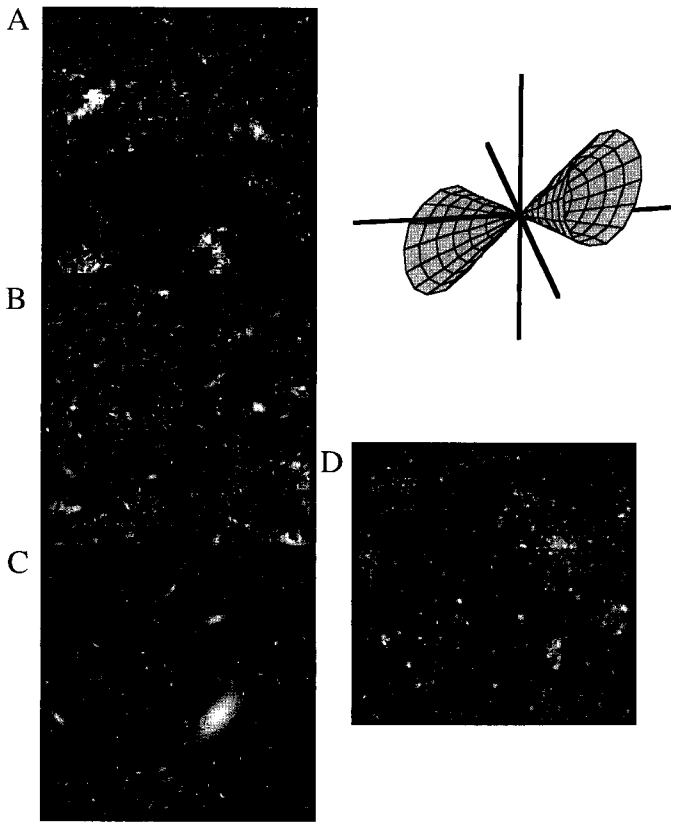


Figure 12: Images that produce high kurtosis can be created by "inverting" the visual code. The top images were created by distributing a set of self-similar functions as described in equation 5.1. The images were created by producing a random sum of the oriented wavelet basis functions. The two images shown were created by assigning a probability of either (A) 0.01, (B) 0.08, or (C) 0.3 to each of the members of the wavelet basis set (e.g., Fig. 8). The lower right image (D) was created using nonoriented difference of gaussians (DOG). It is proposed that images such as these provide a first approximation to the statistical structure of natural scenes. The state space is analogous to the cone (E) with the principal axes aligned with the Fourier vectors.

like that shown in Figure 12 using DOG functions instead of oriented wavelets. Under such circumstances, an oriented wavelet will not produce an increase in kurtosis over the DOG.

It is proposed that because natural scenes have regions where the phases are locally aligned (i.e., the images have local structure), both ganglion cell receptive fields and cortical receptive fields have the shape that they do. The amplitude spectra of natural scenes and the limits in quantal catch may certainly be important in understanding the spatial frequency tuning curves of ganglion cells and the contrast sensitivity function as suggested by Atick and Redlich (1991, 1992). However, it is proposed here that it is the phase spectrum of natural scenes that one must understand to account for the phase spectra of receptive fields in the visual pathway. Indeed, it is likely that many natural phenomena show this self-similar phase structure and may be why the wavelet transform has found so much success in applied mathematics.

5.2 Higher-Order Redundancy in Natural Scenes. Although the images in Figure 12 may provide a good first approximation to the statistics of the environment, these images certainly do not capture all the redundancy that is present in real scenes and do not look that much like real scenes. Consider the analogy of looking at the redundancy in musical symphonies. One might try to find the properties of the typical notes that are played. One might then try to create synthetic symphonies by playing those notes randomly. Although these sounds would be more like symphonies than random noise, they would not sound like symphonies because they would not contain any of the rules of music. Similarly, the images in Figure 12 are more like natural scenes than $1/f$ noise, but they do not contain any of the combinatorial rules found in real scenes.

As previously noted (Field 1993), natural scenes show a significant degree of continuity across space and frequency. Figure 13 shows a simple example of this continuity. A "fractal" edge is split into three frequency bands. Across the different frequency bands and across the length of the edge, the orientation of the edge twists and turns. Locally, the shift in orientation is continuous. The wavelet code does not directly take advantage of this continuity. However, it has recently been proposed (Field *et al.* 1993) that the visual system may make use of this redundancy by using connections between locally oriented units along the lines of the lateral connections found in the primary visual cortex (e.g., Rockland and Lund 1983; Blasdel *et al.* 1985; see Gilbert 1992 for a review). Computational studies by Zucker *et al.* (1989) suggest that such an approach may be an effective strategy in representing continuous features in natural scenes.

It should also be emphasized that the "cells" modeled in this study do not have the spatial nonlinearities that are often found in visual neurons. Such nonlinearities come in various forms. First, in primary visual cortex, "simple cells" do not represent the majority of cells. Both complex and hypercomplex cells certainly represent a major component to visual coding and show highly nonlinear behavior. However, the basic findings described here will at least apply to complex cells. The bandwidths

and spatial extents of complex cells are similar to those of simple cells. As noted earlier, one model of complex cells suggests that they detect an

alignment of phases, like simple cells, but do not differentiate the absolute phase of that alignment. Since they are localized in space/frequency in a similar manner to simple cells, it would be expected that complex cells will show a similar sparse representation to the modeled simple cells studied here.

It should also be noted that even classically defined simple cells exhibit a number of important nonlinearities (e.g., end-stopping and cross-orientation inhibition). We currently believe that these nonlinearities may help to increase the "sparseness" of the code by allowing the code to play a local "winner take all" strategy. This is beyond the scope of this paper, but it should be mentioned that the kurtosis values shown in Figures 10 and 11 are likely to be significantly lower than the kurtosis values actually found in the mammalian visual cortex.

It is important to treat the results described in this study with some caution. We have not added the nonlinearities known to exist in the mammalian visual system and we have only searched through a small variety of linear transforms. There may well be linear transforms that produce higher kurtosis than the wavelets tested here and it is quite likely that there exist more effective nonlinear codes. Our main goal in this paper is to demonstrate that the bandpass, localized receptive fields found in the mammalian visual system can take advantage of the redundancy in the environment without compressing the representation (reducing redundancy) and without using the principal components.

5.3 Why Sparse Coding? In this paper, evidence was provided that codes that share properties with the mammalian visual system produce a sparse-distributed output to natural scenes. But the question remains as to why a sensory system would evolve a strategy like sparse coding. As was noted, there are a number of reasons that one might want to produce a compact code. A compact code allows the data to be stored and transmitted with a smaller population of total cells. But what are the advantages of sparse codes?

Under some conditions, it is possible to use the sparse output to produce a compressed image. Techniques such as run-length coding are designed to do just that. There is also a considerable literature on compression techniques with sparse matrices (e.g., Evans 1985; Schendel

Figure 13: *Facing page.* Unlike the scenes shown in Figure 12, natural scenes have structures that are continuous across space and frequency. The orientation of the edge shifts across the length of the edge as well as across the different scales. The edge has a well-defined orientation only at a given scale and position along the length of the edge. This oriented structure that is localized in space/frequency makes the wavelet code an effective representation. However, the continuity between scales and across the length of the edge at each scale represents a form of redundancy that is not directly captured by the wavelet and requires more complex processing.

1989). However, it is proposed here that sparse coding serves purposes other than compression. Barlow (1972, 1985) has proposed that codes that minimize the number of active neurons can be useful to the resentation and detection of "suspicious coincidences." Zetzsche (1990) has suggested that it is the work on associative memories that provides the most biologically plausible reasons that sensory systems would use sparse coding schemes.

Three interrelated advantages are described below.

Signal-to-Noise. First, a sparse coding scheme can increase the signal-to-noise ratio. As was noted in Field (1987), if most of the variance of a data set is represented by a small subset of cells, then that subset must have a high response relative to the cells that are not part of the subset. The smaller the subset, the higher the response of each member of the subset, given an image with constant variance. If we consider the response of the subset as "the signal" and if all the cells in the population are subject to additive noise, then by increasing the response of the subset of cells relative to the population, it is possible to increase the probability of detecting the correct subset of cells that represent the image. However, it should be emphasized that arguments of this form depend critically on the properties of the noise (e.g., correlated versus uncorrelated) and the location of the noise (e.g., photon versus neural transduction noise).

Correspondence and Feature Detection. Although signal-to-noise considerations may be important, it is proposed that the main reason for sparse coding is that it assists in the process of recognition and generalization. In an ideal sparse code, the activity of any particular basis function has a low probability. Because the response of each cell is relatively rare, tasks that require matching of features should be more successful. Consider the problem of identifying a corresponding structure (e.g., an edge) across two frames of a movie or across two images of a stereo-pair. If the probability of a cell's response has probability p and there are n comparable cells of that type within some region of the visual field then, assuming independent probability of response, the probability of detecting the correct correspondence depends on the probability that only the corresponding cell responds, which is

$$P(\text{only corresponding cell responds}) = (1 - p)^n$$

The lower the probability p (i.e., the more sparse the code), the more likely a correct correspondence. As noted above, we would not expect cortical cells to be completely independent of their neighbors. Nonetheless, the general rule should still hold. As a code becomes more sparse, the probability of detecting a correct correspondence increases. This suggests that a sparse code should be of assistance in tasks requiring solutions to a correspondence problem and can be related to what Barlow calls a suspicious coincidence (e.g., Barlow 1989). If the probability that any cell

responds is low ($p \ll 0.5$), then the probability of two cells responding is p^2 ($p \ll 0.25$) assuming response independence. Higher-order relations (pairwise, triad, etc.) become increasingly rare and therefore more informative when they are present in the stimulus.

This implies that the unique pattern of activity found with sparse codes may also be of assistance with the general problem of recognition. If relations among units are used to recognize a particular view of an object (e.g., a face), then with a sparse code, any particular high-order relation is relatively rare. In a compact code, a few cells have relatively high probability of response. Therefore, any particular high-order relation among this group is relatively common. Overall, in a compact code different objects are represented in terms of the differential firing of the same subset of cells. With a sparse-distributed code, a large population has a relatively low probability of response. Different objects are represented by a unique subset of cells. That is, different objects are represented by which cells are active rather than how much they are active.

The primary difficulty with this line of thinking is that implementing a process that detects sparse structure in neural architecture may not be so straightforward. Higher-order relations may be relatively rare in a sparse code, but detecting all possible nth-order relations among m cells requires m^n detectors. It is likely that the visual system looks only for "probable" relations by looking at combinations only within local regions and looking only for probable structure (e.g., continuity). This is certainly an interesting problem, but further study of the higher-order structure of natural scenes would be required to determine whether the visual system is using an effective strategy.

Storage and Retrieval with Associative Memory. The third advantage of sparse coding comes from work on networks with associative memories and is related to the above discussion. Several authors have noted that when the inputs to the networks are sparse, the networks can store more memories and provide more effective retrieval with partial information (Palm 1980; Baum *et al.* 1988). Baum *et al.* (1988) suggest that the advantages of sparsity for cell efficiency are "so great that it can be useful to artificially 'sparsify' data or responses which are not sparse to begin with." Indeed, it is not surprising that many types of networks will solve problems more efficiently if the inputs are first "sparsified." Since the sparse representation will have fewer higher-order relations, learning to classify or discriminate inputs should require less computation. Therefore, "sparsifying" the input should help to simplify many of the problems that the network is designed to face. This work with associative memory will hopefully lead to a better understanding of the advantages of sparse codes and help us to understand whether transforming the data to create a sparse input is generally a useful strategy to help networks solve problems.

In this study, there is no attempt to provide a complete description of the possible uses of sparse codes. Rather, the goal is to demonstrate that sparse coding represents one important method for taking advantage of redundancy and that sensory systems show evidence that they make use of this method.

5.4 Factorial Codes and Projection Pursuit. A gaussian distribution has the lowest redundancy (highest entropy) of any distribution given a fixed variance. Thus, a distribution with high kurtosis has a lower entropy than a gaussian distribution. The results shown in Figure 10 show that the visual code has relatively redundant first-order histograms. A sparse-distributed code converts high-order redundancy (relations between units) into first-order redundancy (the response distributions of the basis vectors). Therefore, a transform that produces highly redundant histograms has decreased the higher-order redundancy.

Is it possible to find a code that completely removes higher-order redundancy? In such a case, the responses of the vectors would be independent of one another. For example, if the responses of two vectors are independent then

$$P(V_i \cap V_j) = P(V_i) \cdot P(V_j)$$

and

$$P(V_i \mid V_j) = P(V_i)$$

and in general, if all the vectors are independent then the probability of any given image can be determined by multiplying the probabilities of each of the vectors.

$$P(\text{image}) = \prod P(V_i)$$

In this case, the image is described as having a "factorial code" (Barlow *et al.* 1989; Barlow 1987; Schmidhuber 1992; Atick *et al.* 1993).

A population of images like those shown in Figure 12 can have nearly factorial codes since these were actually generated by combining nearly orthogonal vectors with independent probabilities. That is, the probability that any function was added to the image did not depend on the probability that any other function was added to the image. The differences between the images in Figures 12 and 13 point out the importance of sources of redundancy that remain after coding by the wavelet. Indeed, the fact that the natural scenes do not look like the images in Figure 12 demonstrates that after natural scenes are coded by the wavelet, the responses of wavelet basis functions will not be independent. Thus the individual probabilities of units are unlikely to predict the probability of a particular natural scene. Whether one is searching for a sparse code or a factorial code, the goal is to find a set of units that are as independent

from each other as possible. However, the sparse coding approach provides a guide for achieving this goal. By maximizing the kurtosis (i.e., the redundancy) of the response histograms, one effectively minimizes statistical dependencies between units (Field 1987).

In this paper, there has been no discussion of how one might find the optimal sparse code for a given data set. An effective sparse code must have two properties. It must span the space of inputs (i.e., preserve information) and show high kurtosis in the response histograms. At this time, there do not appear to be any learning rules that can achieve these goals. There is a problem related to this, described as "projection pursuit" (e.g., Friedman 1987; Huber 1985; Intrator 1992, 1993). In many domains, where one must deal with high-dimensional data, one is interested in finding interesting projections of the data. The "projections" refer to the response histograms of the vectors describing the data. Intrator (1992, 1993) has noted that one should look for projections (histograms) that are as far from gaussian as possible. Indeed, in line with the above discussion, the more non-gaussian the histogram, the more independent the units should be. High kurtosis represents just one way to deviate from a gaussian distribution. It is not clear whether natural scenes have redundancy that will allow other forms of non-gaussian behavior in the response histograms. It is also not clear that the visual system can take advantage of other forms of non-gaussian behavior. In this paper, it is noted only that the visual system appears to have found one type of non-gaussian distribution (high kurtosis) and this particular distribution results in a sparse representation.

However, at this time, there is no known technique for finding the optimal sparse code. Techniques such as those of Intrator (1992, 1993), Foldiak (1990), Schmidhuber (1992), and Linsker (1993) may ultimately provide insights into this problem, but their work demonstrates that the solution may not be a simple one.

6 Summary

In this paper, we have compared two approaches to sensory coding: compact and sparse-distributed coding. When the statistics of the inputs are stationary, it was noted that compact coding depends primarily on the amplitude spectra of the data (i.e., the correlations). When effective dimensionality reduction is possible, compact coding provides a good first step. Indeed, some aspects of sensory coding (e.g., trichromacy) require a consideration of the efficiency achieved by dimensionality reduction. However, many redundant data sets do not allow effective compression. Furthermore, even if compact coding is possible, there will be a variety of codes that will produce equal compression.

If the goal of a code is to assist higher level processing (e.g., recognition), then other coding strategies and other forms of redundancy must

be considered. To account for the primary receptive field properties of cells found in the mammalian visual system (i.e., localized, bandpass, self-similar), it was proposed that one must consider a sparse-distributed representation of natural scenes. It was noted that when the statistics of the data are stationary, sparse coding depends primarily on the phase spectra of the data. In this paper, we have concentrated on the visual system and the statistics of natural scenes. However, the main ideas presented here can be applied to any sensory system. Natural sounds, for example, are likely to have local structure similar to that found in natural scenes. We are currently working on the question of whether the sparse coding approach to natural sounds can provide an account of the frequency selectivity found in auditory neurons.

It is proposed that the evidence for this selectivity will be found in the kurtosis of the response distribution. If a sensory system is designed for sparse coding, then we would expect cells in that sensory system to show high kurtosis in the presence of the typical sensory environment. If the general goal of sensory coding is to produce a sparse representation of the environment, then we would expect that recording from any sensory neuron in any animal will produce a histogram with high kurtosis—as long as the recording is performed in an awake, behaving animal in its natural environment. This is a general claim that is left for future work to answer.

Appendix

An image consisting of the appropriate functions distributed across the image in the appropriate manner will result in an image that is scale invariant in both contrast and structure. As noted previously (Field 1987), a two-dimensional image that is scale invariant in contrast will have an amplitude spectrum that falls with frequency as $1/f$. In this appendix, it is demonstrated that images that obey the rules described in equation 5.1 will be scale-invariant in their contrast (have $1/f$ amplitude spectra) and also have a scale-invariant structure. Consider a simple function localized in space and assigned a scale k. In equation 5.1, the scale $k = \sigma^m$. By the scaling theorem, it follows that

$$g(kx.ky) \leftrightarrow G(u/k, v/k)/k^2$$

Now consider a sum of scaled functions that are placed in random positions relative to one another. Functions that are in random position will, on average, have phases that are orthogonal. In line with the sum of intensities in incoherent optics, the power spectrum (square of the am-

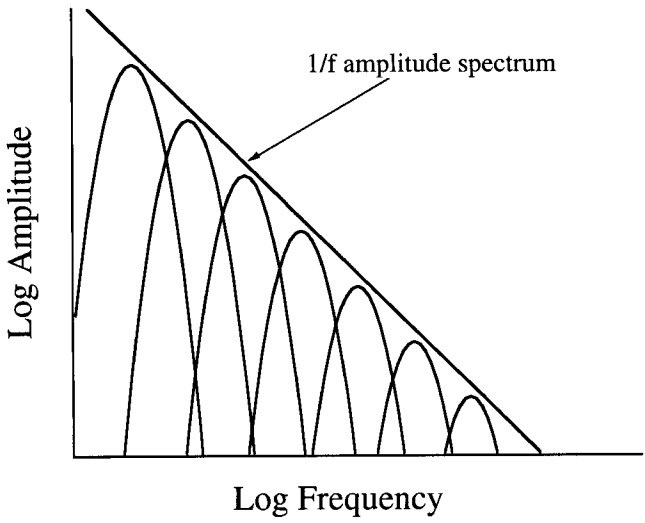


Figure 14: See text.

plitude spectrum) is equal to the sum of the power spectra of each of the functions, thus

$$\begin{array}{lll} & \text{Space} & \text{Amplitude spectra} \\ \sum_i^n g(kx + x_i, ky + y_i) & \leftrightarrow & \sqrt{n[G(u/k, v/k)k^2]^2} \\ & & = \sqrt{n}|G(u/k, v/k)|/k^2 \end{array}$$

In equation 5.1, the images are created by summing $\beta \cdot k^2$ functions at each scale where $k \propto \sigma^m$. Thus at each scale, the spectrum is proportional to

$$\begin{aligned} G(u, v, k) &= \sqrt{k^2} G(u/k, v/k)/k^2 \\ &= G(u/k, v/k)/k \end{aligned}$$

Thus at each scale, the peak of the spectrum falls by a factor of $1/k$. Figure 14 shows an example of the spectra at different scales. The spacing in the frequency domain is proportional to σ^m $(m = 1, 2, 3, \ldots)$. On a log frequency axis, the spacing is

$$\log(\sigma^m) \propto m \quad (m = 1, 2, 3, \ldots)$$

With this sum of scaled functions, the amplitude spectra of the synthetic image will fall as $1/f$, that is, the contrast is scale-invariant. The

local structure is also scale invariant. At each scale k (i.e., within each frequency band), the image consists of $\beta \cdot k^2$ elements of amplitude w_{nm} and size $1/\tau k$. If we magnify the image by some factor r then we shift to a scale with $\beta \cdot (r \cdot k)^2$ elements of size $1/\tau k$ and amplitude w_{nm}. However, if we scale our window with the magnification then the area is reduced by a factor of $1/r^2$. Thus with a window of constant angular size, the number of elements at any scale is independent of magnification.

$$\beta \cdot (r \cdot k)^2 / r^2 = \beta \cdot k^2$$

Thus, at all scales there remain βk^2 randomly positioned elements of size $1/\tau k$. The images created using the method described in Figure 12 have local structure that is scale-invariant. As noted in the text, this provides a good first approximation to natural scenes but fails to capture combinatorial rules found in real scenes.

Acknowledgments

This work was supported by NIH Grant R29MR50588. I would like to thank Peter Foldiak, Nuala Brady, and the six reviewers for their helpful comments.

References

Adelson, E. H., Simoncelli, E., and Hingorani, R. 1987. Orthogonal pyramid transforms for image coding. *SPIE Visual Commun. Image Process.* **II**, 845.

Atick, J. J., and Redlich, A. N. 1990. Towards a theory of early visual processing. *Neural Comp.* **4**, 196–210.

Atick, J. J. 1992. Could information theory provide an ecological theory of sensory processing? *Network* **3**, 213–251.

Atick, J. J., and Redlich, A. N. 1992. What does the retina know about natural scenes? *Neural Comp.* **4**, 449–572.

Atick, J. J., Li, Zhaoping, and Redlich, N. 1993. What does post-adaptation color appearance reveal about cortical color coding? *Vision Res.* **33**, 123–129.

Baddeley, R. J., and Hancock, P. J. 1991. A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proc. Roy. Soc. London B* **246**, 219–223.

Barlow, H. B. 1961. The coding of sensory messages. *Current Problems in Animal Behavior*. Cambridge, Cambridge University Press.

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

Barlow, H. B. 1985. The Twelfth Bartlett Memorial Lecture: The role of single neurons in the psychology of perception. *Q. J. Exp. Psychol.* **37A**, 121–145.

Barlow, H. B. 1989. Unsupervised learning. *Neural Comp.* **1**, 295–311.

Barlow, H. B., and Foldiak, P. 1989. Adaptation and decorrelation in the cortex. In *The Computing Neuron*, R. Durbin, C. Miall, and G. Mitchison, eds., pp. 54–72. Addison-Wesley, Reading, MA.

Barlow, H. B., Kaushal, T. P., and Mitchison, G. J. 1989. Finding minimum entropy codes. *Neural Comp.* **1**, 412–423.

Baum, E. B., Moody, J., and Wilczek, F. 1988. Internal representations for associative memory. *Biol. Cybern.* **59**, 217–228.

Bialek, W., Ruderman, D. L., and Zee, A. 1991. Optimal sampling of natural images: A design principle for the visual system? In *Advances in Neural Information Processing 3*, R. Lippmann, J. Moody, and D. Touretzky, eds., pp. 363–369. Morgan Kaufmann, San Mateo, CA.

Blasdel, G. G., Lund, J. S., and Fitzpatrick, D. 1985. Intrinsic connections of macaque striate cortex: Axonal projections of cells outside lamina 4C. *J. Neurosci.* **5**, 3350–3369.

Bossomaier, T., and Snyder, A. W. 1986. Why spatial frequency processing in the visual cortex? *Vision Res.* **26**, 1307–1309.

Brady, N., and Field, D. J. 1993. What's constant in contrast constancy?: A vector length model of suprathreshold sensitivity. *Vision Res.*, submitted.

Buchsbaum, G., and Gottschalk, A. 1983. Trichromacy, opponent colors coding and optimum color information transmission in the retina. *Proc. Roy. Soc. London B* **220**, 89–113.

Burt, P. J., and Adelson, E. H. 1983. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications* **31**, 532–540.

Burton, G. J., and Moorhead, I. R. 1987. Color and spatial structure in natural scenes. *Appl. Opt.* **26**, 157–170.

Churchland, P. S., and Sejnowski, T. J. 1992. *The Computational Brain.* MIT Press, Cambridge, MA.

Daubechies, I. 1988. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41**, 909–996.

Daugman, J. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Amer.* **2(7)**, 1160–1169.

Daugman, J. G. 1988. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transact. Acoustics, Speech Signal Process.* **36**(7), 1169–1179.

Daugman, J. G. 1991. Self-similar oriented wavelet pyramids: Conjectures about neural non-orthogonality. In A. Gorea, ed., *Representations of Vision*. Cambridge University Press, Cambridge.

Derrico, J. B., and Buchsbaum, G. 1991. A computational model of spatiochromatic coding in early vision. *J. Visual Commun. Image Process.* **2**, 31–38.

DeValois, R. L., Albrecht, D. G., and Thorell, L. G. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559.

Eckert, M. P., and Buchsbaum, G. 1993. Efficient coding of natural time varying images in the early visual system. *Phil. Trans. R. Soc. London B* **339**, 385–395.

Evans, D. 1985. *Sparsity and Its Applications.* Cambridge University Press, Cambridge.

Farge, M., Hunt, J., and Vassilicos, J. C., eds. 1992. *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*. Oxford University Press, Oxford.

Field, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Amer.* **4**, 2379–2394.

Field, D. J. 1989. What the statistics of natural images tell us about visual coding. *Proc. SPIE* **1077**, 269–276.

Field, D. J. 1993. Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems. In *Wavelets, Fractals and Fourier Transforms*, M. Farge, J. Hunt, and J. C. Vassilicos, eds. Oxford University Press, Oxford.

Field, D. J., and Tolhurst, D. J. 1986. The structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex. *Proc. Roy. Soc. London B* **228**, 379–400.

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

Foldiak, P. 1989. Adaptive network for optimal linear feature extraction. In *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks*, Vol. 1, pp. 401–405. IEEE Press, New York.

Foldiak, P. 1990. Forming sparse representations by local anti-Hebbian learning. *Biol. Cybern.* **64**, 165–170.

Friedman, J. H. 1987. Exploratory projection pursuit. *J. Amer. Statist. Assoc.* **82**, 249–266.

Gabor, D. 1946. Theory of Communication. *J. IEE London* **93**(III), 429–457.

Gilbert, C. D. 1992. Horizontal integration and cortical dynamics. *Neuron* **9**, 1–13.

Hancock, P. J., Baddeley, R. J., and Smith, L. S. 1992. The principal components of natural images. *Network* **3**, 61–70.

Huber, P. J. 1985. Projection pursuit. *Ann. Statist.* **13**, 435–475.

Intrator, N. 1992. Feature extraction using an unsupervised neural network. *Neural Comp.* **4**, 98–107.

Intrator, N. 1993. Combining exploratory projection pursuit and projection pursuit regression with application to neural networks. *Neural Comp.* **5**, 443–455.

Jones, J., and Palmer, L. 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**(6), 1233–1258.

Kersten, D. 1987. Predictability and redundancy of natural images. *J. Opt. Soc. Amer.* **4**, 2395–2400.

Klein, S. A., and Levi, D. M. 1985. Hyperacuity thresholds of 1 sec: Theoretical predictions and empirical validation. *J. Opt. Soc. Amer.* **2**(7), 1170–1190.

Kulikowski, J. J., Marcelja, S., and Bishop, P. O. 1982. Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biol. Cybern.* **43**, 187–198.

Lehky, S. R., and Sejnowski, T. J. 1990. Network model of shape-from-shading: Neural function arises from both receptive and projective receptive fields. *Nature (London)* **333**, 452–454.

Linsker, R. 1988. Self-organization in a perceptual network. *Computer* **21**, 105–117.

Linsker, R. 1993. Deriving receptive fields using an optimal encoding crite-

rion. In *Advances in Neural Information Processing Systems 5*, S. J. Hanson, J. D. Cowan, and C. L. Giles, eds., pp. 953–960. Morgan Kaufmann, San Mateo, CA.

MacKay, D. J., and Miller, K. D. 1990. Analysis of Linsker's simulation of Hebbian rules. *Neural Comp.* **1**, 173–187.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transact. Pattern Anal. Machine Intelligence* **11**(7), 674–693.

Maloney, L. T. 1986. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Amer. A* **3**, 1673–1683.

Morrone, M. C., and Burr, D. C. 1988. Feature detection in human vision: A phase-dependent energy model. *Proc. Roy. Soc. London B* **235**, 221–245.

Oja, E. 1982. A simplified neuron model as a principal component analyzer. *J. Math. Biol.* **15**, 267–273.

Palm, G. 1980. On associative memory. *Biol. Cybern.* **36**, 19–31.

Pratt, W. K. 1978. *Digital Image Processing.* Wiley, New York.

Rockland, K., and Lund, J. S. 1983. Intrinsic laminar lattice connections in primary visual cortex. *J. Comp. Neurol.* **216**, 303–318.

Sanger, T. D. 1989. Optimal unsupervised learning in a single layer network. *Neural Networks* **2**, 459–473.

Schendel, U. 1989. *Sparse Matrices.* Wiley, New York.

Schmidhuber, J. 1992. Learning factorial codes by predictability minimization. *Neural Comp.* **4**, 863–879.

Shapley, R. M., and Lennie, P. 1985. Spatial frequency analysis in the visual system. *Annu. Rev. Neurosci.* **8**, 547–583.

Tolhurst, D. J., and Thompson, I. D. 1982. On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proc. Roy. Soc. London Ser. B* **213**, 183–199.

Tolhurst, D. J., Tadmor, Y., and Tang Chao 1992. The amplitude spectra of natural images. *Ophthal. Physiol. Opt.* **12**, 229–232.

van Hateren, J. H. 1992. Real and optimal neural images in early vision. *Nature (London)* **360**, 68–69.

Webber, C. J. St. C. 1991. Competitive learning, natural images and cortical cells. *Network* **2**, 169–187.

Watson, A. B. 1983. 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.

Webster, M. A., and DeValois, R. L. 1985. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J. Opt. Soc. Amer.* **2**(2), 1124–1132.

Zetzsche, C. 1990. Sparse coding: The link between low level vision and associative memory. In *Parallel Processing in Neural Systems and Computers*, R. Eckmiller, G. Hartmann, and G. Hauske, eds. North-Holland, Amsterdam.

Zucker, S. W., Dobbins, A., Iverson, L. 1989. Two stages of curve detection suggest two styles of visual computation. *Neural Comp.* **1**, 68–81.

Received January 8, 1993; accepted October 20, 1993.