In: Wavelets, Fractals and Fourier Transforms: New Developments and New Applications. Farge, M., Hunt, J. and Vassilicos., eds. Oxford University Press. (1993)

Scale-invariance and Self-similar 'Wavelet' Transforms: an Analysis of Natural Scenes and Mammalian Visual Systems

D.J. Field

Cornell University Ithaca, New York

Abstract The processing of spatial patterns by the mammalian visual system shows a number of similarities to the 'wavelet transforms' which have recently attracted considerable interest outside of the study of sensory systems. At the level of the primary visual cortex, these visual systems consist of arrays of neurons selective to local regions of space, spatial frequency and orientation. The spatial frequency bandwidths of these neurons increase with frequency resulting in a set of approximately self-similar "receptive fields". In this paper, we look at the question of why this strategy of representing the visual environment would evolve. The question is approached by looking at the statistical structure of natural scenes and observing how this structure relates to the visual system's representation of spatial patterns. It is proposed that natural scenes are approximately scale invariant with regards to both their power spectra and their phase spectra. Principally because of the phase spectra, wavelet-like transforms are capable of producing a sparse, informative representation of these images. It is suggested that self-similar codes like the wavelet are effective for so many natural phenomena because such phenomena show similar structures to those found in these natural scenes.

1 Introduction

For the last two decades, many of the researchers investigating mammalian visual processing have been developing transforms derived from the properties of the mammalian visual system. Following from Gabor's original notion of time-frequency representations (Gabor, 1946), these models have proved to be an effective means of exploring how spatial information is represented and processed by the visual system. Many of these transforms are of a general class referred to as self-similar (e.g., Watson, 1983; Field, 1987; Daugman, 1988). The basis functions of these self-similar codes consist of arrays of localized and band-pass basis functions which only vary by a di-

lation, a translation or a rotation. However, these self-similar transforms have found applications in a wide range of image coding situations including compact coding (Watson, 1987; Daugman, 1988; Adelson, Simoncelli & Freeman, 1991), identification systems (Daugman, 1991), texture segregation (Bovik, Clark & Geisler, 1990) and methods for efficiently transmitting the time-varying chromatic signal required for HDTV (Watson, 1990; Watson, 1991).

Although much of the background for this work was derived from basic ideas in information theory and image compression, much of our current knowledge of these two-dimensional self-similar codes comes directly from work directed towards an understanding of how the visual system processes spatial information. It is this historical background which provides the foundations for many of the ideas discussed in this paper (Kulikowski, Marcelja & Bishop, 1982; Sakitt & Barlow, 1982; Watson, 1983; Daugman, 1985).

In the next sections, we will discuss some of the spatial response properties of cells in the mammalian visual cortex. The transforms that have been used to model the mammalian visual system have come in a variety of forms and many of these would now come under the heading of "wavelet transforms". Basis functions have ranged from Gaussian derivatives (Young, 1987) to functions derived from hexagonal latices (Watson & Ahumada, 1989). One of the more popular transforms was derived from the Gaussian modulated sinusoid (i.e., the "Gabor functions" -equations 1 and 2) and we will return to this code in the next section. Unfortunately, the basis functions of such transforms are not orthogonal making it difficult to invert and difficult to use in modelling.

To make the modelling and inversion easier, two approaches have been developed. One approach is to derive codes which share many of the properties of the visual system but which use orthogonal basis sets (e.g., Adelson, Simoncelli & Hingorani, 1987; Watson and Ahumada, 1989; Simoncelli & Adelson, 1990; Adelson et al., 1991). A second approach developed by Daugman (1988) uses a two-layered 'neural network' to find the coefficients required to invert the code. However, there is little evidence that the mammalian visual system uses either technique. Furthermore, there is little reason to expect that the transform employed by the visual system would need to be inverted. Nonetheless, these techniques have been extremely useful in modelling and image analysis.

The developments in "Wavelet theory" (e.g., Combes et al., 1989; Meyer and Paul, 1991) have extended our knowledge of self-similar representations and have extended our understanding of the limitations and capabilities of such transforms. They have shown clearly that the mathematics underlying these transforms have applications to a wide range of phenomena outside of image representation. This volume provides an important insight into the variety of applications we are likely to see in the next few years.

Although these transforms are finding a wider and wider range of applications, a number of important questions remain unanswered. Most important: Why are these transforms so effective in representing natural phenomena? What aspects of natural phenomena make them useful? What types of phenomena are not well described by these self-similar transforms? When do they perform poorly? Is there an optimal transform for every class of data? If so, then how does one find it?

In this paper, we use the mammalian visual system as a model and look at the relationship between this transform and statistical properties of the visual environment. Following from previous work (Field, 1987, Field, 1989), it will be suggested that for the mammalian visual system, the particular transform employed by the visual system conforms to the particular self-similar structures found in natural scenes. The visual system is efficient because it is well matched to the statistical redundancy of the visual environment.

The self-similar structure of these scenes will be described using two different measures. The first is the power spectrum and the related autocorrelation function. The second is a measure of the local phase structure. By understanding both of these structures but mostly by understanding how the "local phase spectrum" is aligned across neighbouring frequency bands, we hope to gain a better insight into the goal of coding by the mammalian visual system.

1.1 The mammalian visual system

Although there are number of differences between the visual systems of different mammals, there are a considerable number of similarities, especially in the representation of spatial information. The most extensively studied systems are those of the cat and monkey and it is studies on these animals that provide the basis of much of our knowledge about visual coding.

The acuity of the cat is significantly lower than that of the monkey, but within the range of sensitivities covered by these visual systems (i.e., the spatial frequency range), the methods by which spatial information is processed follows a number of similar rules. The area that we will be considering is a region at the back of the brain referred to as primary visual cortex (Area 17). This area is the principle projection area for visual information and consists of cells (i.e., neurons) that receive input from neurons in the eye (via an area called the lateral geniculate nucleus - LGN). The neurons in the primary visual cortex are selective to a number of stimulus parameters.

By placing an electrode near one of these neurons and recording small voltage changes in the neurons membrane, one finds that the neurons will "respond" to a selective range of stimuli. Hubel and Weisel (1962) were the first to provide a spatial mapping of the response properties of these

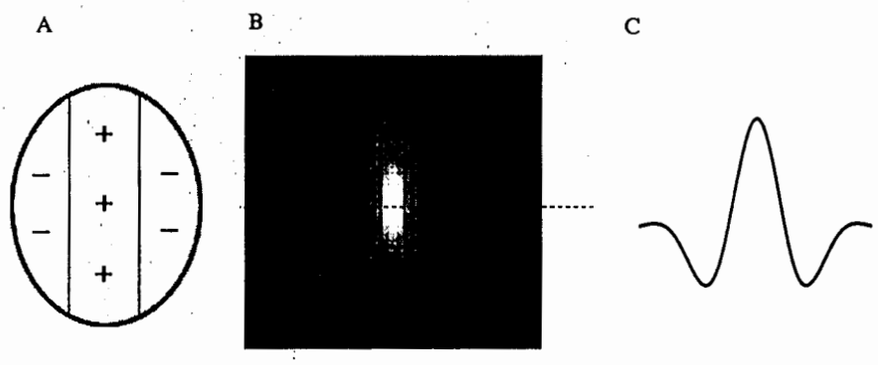


Figure 1. A two dimensional receptive field is defined by the region of the visual field that produces either excitation (+) or inhibition (-) in the activity of the neuron. (a) shows the polarity of the response in a particular receptive field while the intensity in (b) shows the magnitude of the response in terms of intensity. (c) shows the one dimensional profile of this receptive field

neurons. By moving small spots and lines in front of the animal, they found that the neurons would respond when an appropriate stimulus was presented at particular region in the visual field of the animal. The map describing the response region of the cell is referred to as the "receptive field". Figure 1 shows the type of response that is often obtained from one of these neurons. If a spot of light is shown within the receptive field, then the cell will either increase its firing rate (excitation) or decrease its firing rate (inhibition).

Figure 1a shows the response to a bright spot as a function of position across the visual field where (+) refers to excitation (a bright spot increases the activity of the cell and a dark spot decreases the activity). The (-) refers to inhibition (a bright spot decreases the activity of the cell and a dark spot increases the activity). If we allow intensity to represent the magnitude of excitation or inhibition we get the profile something like that shown in Figure 1b. Figure 1c shows a cross section of the centre of this receptive field. As one can probably guess, this particular cell would respond well to a vertical line placed at the centre of the receptive field. With diffuse illumination or a line placed horizontally across the receptive field, the excitation and inhibition will typically cancel and therefore the cell will not respond.

Different neurons respond to different positions within the visual field. Furthermore, at any given position in the visual field, different neurons have receptive fields oriented at different angles and show a variety of sizes.

Thus, the entire visual field is covered by receptive fields that vary in size and orientation. [1] Neurons with receptive fields like the one above were described by Hubel and Weisel as "simple cells" and were distinguished from other types of neurons in primary visual cortex referred to as "complex" and "hyper-complex". (The principle difference is that these neurons show a high degree of spatial non-linearity.)

Throughout the 1960s and 70s there was considerable discussion of how to describe these receptive field profiles and what the function of these neurons might be. One approach (Campbell, & Robson, 1968) was derived from linear systems theory. The same cortical neurons described above were found to be selective to a limited band of orientations and spatial frequencies (Campbell et al., 1968; Campbell et al. 1969; Blakemore, & Campbell, 1969) and this led to the notion that the visual system was providing a description similar to that of the Fourier transform. At the time of this work, such results appeared to contradict the notion that these neurons were selective local features like edges and lines. Indeed, the response of these neurons to edges and lines led other investigators to suggest that the analysis was similar to algorithms that extracted information from an image by performing a more local operation like edge detection (Marr & Hildreth, 1980).

It was not until 1980 (Marcelja, 1980), that the functions describing these receptive fields were considered in terms of Gabor's "theory of communication" (Gabor, 1946). Marcelja noted that the functions proposed by Gabor to analyze time-varying signals showed a number of interesting similarities to the receptive fields of cortical neurons.

Marcelja's suggestion was that the profile described by the line weighting function (Figure 1c) might be well described by a Gaussian modulated sinusoid:

$$f(x) = \sin(2\pi kx + \theta)e^{-x^2/2\sigma^2} \tag{1.1}$$

This function, now referred to as a "Gabor function" has served as a model of cortical neurons by a wide variety of visual scientists. Early tests of this notion showed that such functions did indeed provide an excellent fit to the receptive fields of cortical neurons (Webster & DeValois, 1985; Field & Tolhurst, 1986). Figure 2 provides an idea of the variety of receptive field profiles that could be described by this method along with results showing the best fit to a collection of cortical simple cell receptive fields.

Gabor (1947) proposed these functions because they are well localized in both time and frequency. Daugman (1985) and Watson (1983) generalized

[1] A neuron in primary visual cortex is typically selective to other parameters as well—e.g., retinal disparity, temporal frequency and direction of motion. These parameters as well as the change in the receptive field size in the visual periphery will not be factors in the models presented here

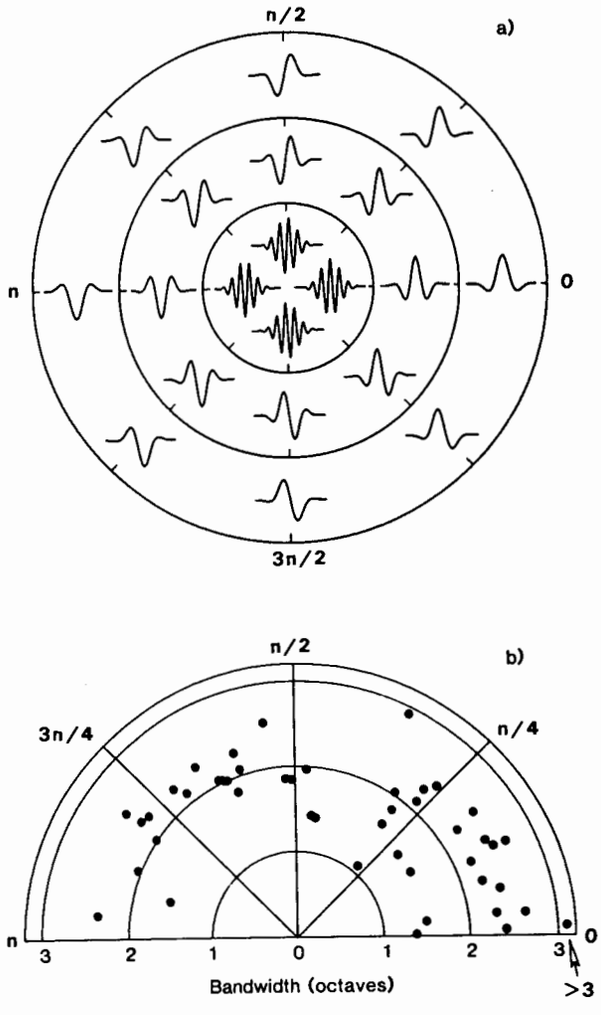


Figure 2. The figure shows results for 46 simple cells from Field and Tolhurst (1986). The receptive field profiles were normalized to the same size and the best fitting "Gabor function" was determined (equation 1.1). The lower plot represents the variety of receptive field profiles as a function of their bandwidth and symmetry. Only half of the plot is represented because a 180 degree flip in orientation is equivalent to a flip in symmetry

Gabor's notion to the two-dimensions of space where the two-dimensional basis function is described as the produce of a two dimensional Gaussian and a sinusoid. For example,

$$f(x,y) = \sin(2\pi kx + \theta)e^{-(x^2/2\sigma_1^2 + y^2/2\sigma_2^2)} \tag{1.2}$$

Although Jones and Palmer (Jones & Palmer, 1987b) found that the full two-dimensional receptive field profiles were well described by this two-dimensional "Gabor function", other studies (Hawken & Parker, 1987) have found that other types of functions (e.g., sum of Gaussians) may provide a better fit and may also achieve better localization (Stark and Wilson, 1990). Although some of the differences between these various models may prove to be important, the differences are not large. All of the basis functions proposed involve descriptions in terms of oriented functions that are well localized in both space and frequency. Figure 3 shows results from two different laboratories showing examples of four 2-D receptive fields from the cat (a) and the amplitude spectra of various cortical cells in the macaque (b).

In the next section, we will be discussing transforms which describe the "average" receptive field properties of cortical cells. However, no single basis set will be capable of describing all of the receptive field types that are found in the mammalian visual cortex. As shown in Figures 2 and 3, there is significant variability in receptive field profiles and their spectra. For example, the bandwidths of cortical cells average around 1.4 octaves (equation 3) but bandwidths less than 1.0 or greater than 2.0 octaves are not uncommon (Tolhurst and Thompson, 1981; DeValois, 1982). It is not clear whether this variability is important to the visual system or whether it should be considered as a limitation in design. For the purposes of this paper, however, we will concentrate only on this average behaviour, not the variability.

1.2 Transform families

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 increased proportionally to frequency. Figure 5 describes a range of possible transforms on a one dimensional image. This space of transforms show bandwidths which vary with frequency according to the equation:

$$\sigma(k) = Bk^{-\alpha} \tag{1.3}$$

where B varies from left to right and α varies from top to bottom. These information diagrams portray a range of techniques of dividing the information space consisting of 18 values into functions localized in space (x) and frequency (k).

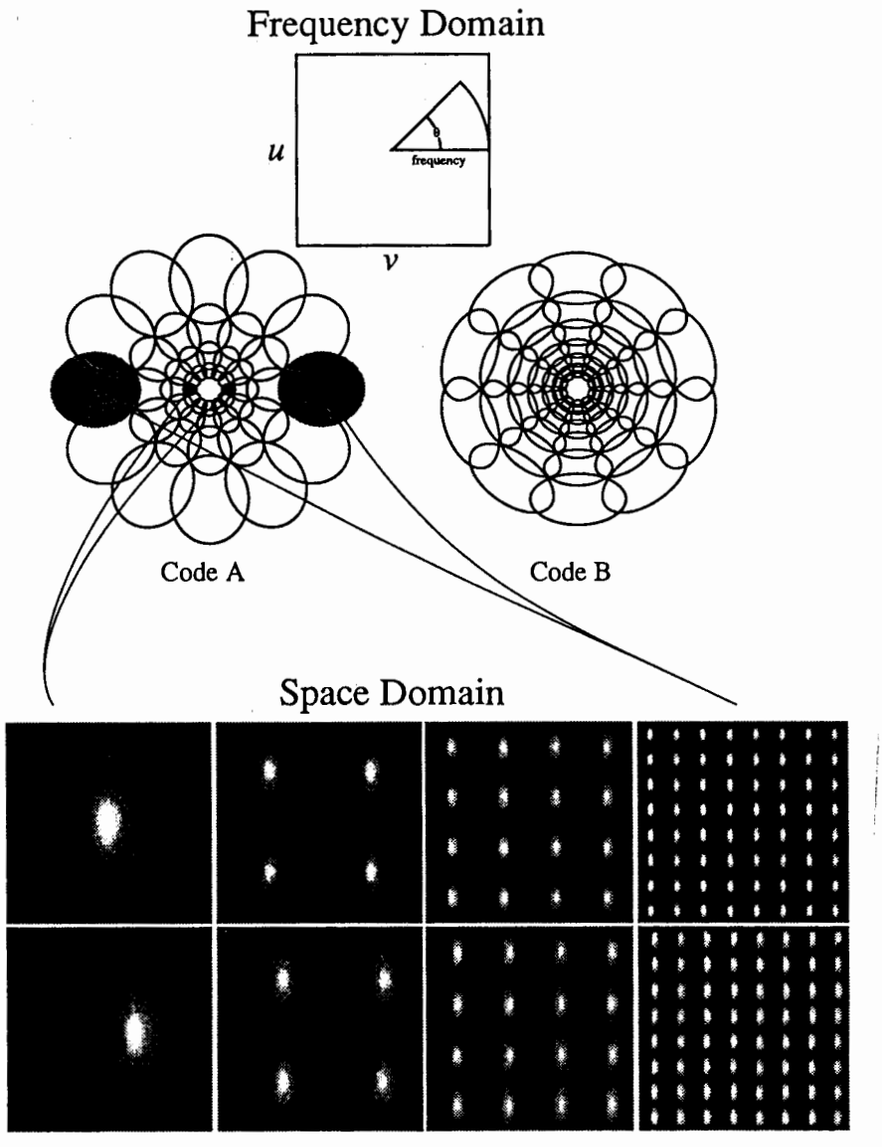


Figure 4. Relations between the two-dimensional spectra and some of their basis functions. The two 'rosettes' represent two different self-similar codes. For both codes, the bandwidths increase with frequency and the orientation tuning is constant in degrees. The spatial basis functions shown below represent the spatial sampling grid for four of the filters from the code (the distance between functions has been increased for clarity). Each two-dimensional wavelet is present in both even and odd symmetric forms as shown in the two lower arrays

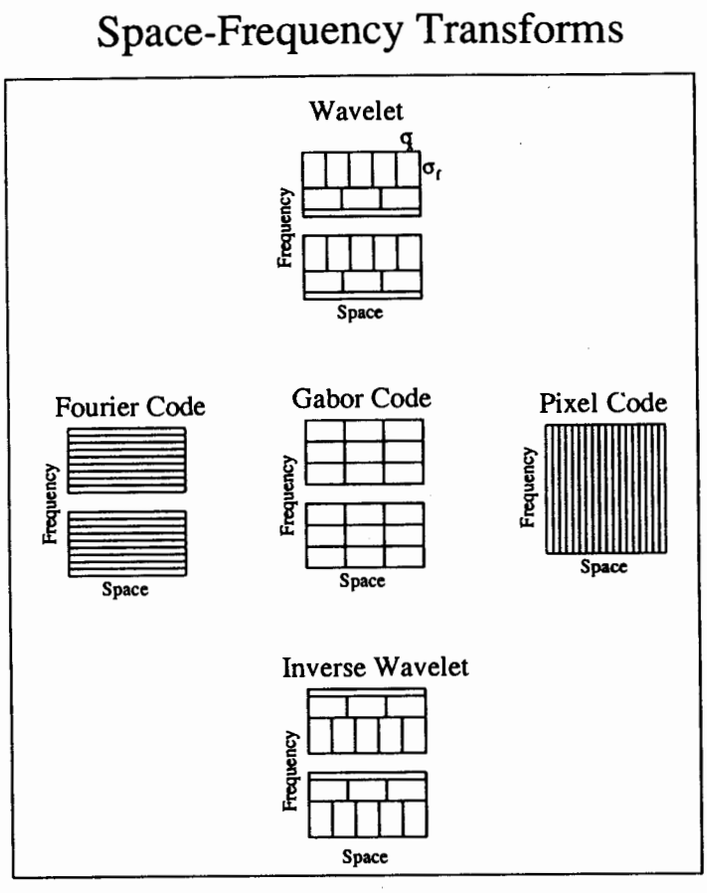


Figure 5. Information diagrams, in line with that originally proposed by Gabor (1946) and Kullikowski (1982). Each diagram represents a different method of representing an 18 point, one-dimensional waveform. A Fourier transform represents the information space with 18 sines and cosines. Each element in the Fourier code (a sine or cosine) is localized in frequency and extends in space. The Gabor code represents information with elements that are localized in both space and frequency. The wavelet code represents information with basis functions that decrease in size and increase in bandwidth with increasing frequency. These five represent a small set of an entire family of such codes described by equation (1.3). All of these codes are capable of representing the information in a data set without loss

The right of Figure 5 shows a transform based on a spatially local "pixel code". With such a code, the 18 values in space are described in terms of "pixels" that are localized in space and extend in frequency. Figure 5 also shows that these 18 values can also be represented with 18 Fourier coefficients that are localized in frequency and extend in space. As Gabor noted, there are many other ways to represent these 18 values. The middle column of Figure 5 shows examples of a transform where the functions used to represent the image are localized in both space and frequency. Following Gabor, the area represented by any function in this information space is limited by:

$$\sigma_k \sigma_x \geq 1/2 \tag{1.4}$$

where the functions that achieve this minimum are the "Gabor functions" described above.[2]

There exist a large variety of possible transforms that are capable of producing a complete representation without loss. Equation 1.3 describes a range of codes other than the five shown here, and there are many others not well described by equation 1.3. There is nothing about any of the transforms in Figure 5 which would make them inherently better or worse than any other. To correctly specify the effectiveness of a particular code, one must define (1) the goal of the code, (2) the type of data that is likely to be encountered, and (3) the conditions under which the transform must function. A Fourier transform, for example, may produce an effective representation for some types of images (e.g., sinewaves) but may be far from optimal for other classes of images (e.g., star charts).

All the representations shown in Figure 5 are capable of completely representing the information in an 18 value data set. The middle line shows transforms which have the same bandwidths at all frequencies ($\alpha = 0$ in equation 1.3). However, Kulikowski et al. suggested that the representation of the visual system followed a code more like that described by the "wavelet code" shown in Figure 5. In transforms like this, the bandwidths (σ_k) increase proportionally to frequency (i.e., $\alpha = 1$ in equation 1.3). In such a code, the basis functions are simply scaled versions of each other (i.e., self-similar).

1.3 Terminology

Wavelet: We will be using the following terminology to describe these transforms. Most important, *we will use the term "wavelet" to apply to transforms in which the basis functions are self-similar (i.e., only differ*

[2]Throughout this paper, we use the term σ_k and σ_x to refer to the size of the functions in space and frequency, respectively. The term Δk and Δx will refer to the sampling distance between neighbouring functions in frequency and space.

by dilations, translations and rotations of a single function). However, to avoid any offence to the vision community, it must be re-emphasized that according to this definition, the "wavelet transform" was developed several years before the important contributions of Grossman, Meyer and Morlet. For example, besides the work in audition, continuous versions of the wavelet (continuous in space and frequency) have been used in visual modelling by Klein and Levi (1985) to help explain hyperacuity thresholds. And for two-dimensional transforms, there were a number of important contributions by Daugman, Watson, Adelson and others in the early 1980s as discussed above.

There appears to be some effort to use the term to apply only to self-similar transforms in which the basis functions are orthogonal or only to those with basis functions which integrate to zero (as in the Log-Gabor transforms (Field, 1987) and Cauchy (Klein and Levi, 1985)), but these requirements are not accepted by everyone.

Gabor transform: In the vision community, the term "Gabor transform" has been applied to all transforms which have basis functions derived from a Gaussian modulated sinusoid (equations 1.1 and 1.2). However, in this paper we will restrict the definition to apply to only those transforms in which the basis functions have the same spatial extent $-\sigma_x$ is constant— (e.g., see Figure 5) and thus the same bandwidth $-\sigma_k$ is constant. This may seem like an odd definition to those in the vision community, but it is a definition which appears to be gaining acceptance (e.g., see Farge in this volume).

Spatial frequency measured in cycles/degree: A receptive field of a neuron in the visual system typically covers some local region of the visual field. Its size can be measured in terms of the angular extent (e.g., 2 degrees of visual angle). That neuron will typically "respond" to a stimulus with a sinusoidal luminance profile when the the period of the sinusoid matches the excitatory and inhibitory regions (Figure 1) of the receptive field. The selectivity of the response is predicted from the Fourier transform of the receptive field profile. Since the response of the cell is determined from the angular region covered by the receptive field, researchers have settled on the term cycles/degree to describe the spatial frequency. For example, in the monkey, it is common to find cells that are optimally tuned anywhere from 1/2 to 30 cycles/degree.

Bandwidth measured in octaves: Since much of the work in vision has its roots in the auditory literature, much of the terminology is in common. One such term is the measure of bandwidth described in terms of octaves. Such a measure is quite useful for a self-similar transform since dilations of a function do not alter the bandwidth when measured in octaves. The spatial-frequency bandwidth is measured as the full-width at half-height, where the bandwidth in octaves is defined as:

$$B_{\text{oct}} = \ln_2(k_2/k_1) \tag{1.5}$$

where k_1 and k_2 are defined as the lower and upper frequencies defining the width at half-height. As noted below, the spatial frequency bandwidths of cortical cells show a fair degree of scatter but appear to average around 1.4 octaves. Although we will treat the visual system as a self-similar code throughout this paper, it should be noted that there is substantial evidence that high frequency cells have somewhat narrower bandwidths (in octaves) at higher spatial frequencies (e.g., Tolhurst and Thompson, 1982). The ramifications of this finding are not yet clear.

1.4 Two-dimensional transforms

The one-dimensional space-frequency transforms were generalized to two-dimensions by several investigators (Watson, 1983; Daugman, 1985; Adelson et al., 1987; Murenzi, 1989). Although many of the main points are described below, the details of the two-dimensional coding schemes used in this paper can be found in more detail in previous papers by the author (Field, 1987,1989).

In general, these codes 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 4 shows an example of the division of the two-dimensional frequency plane and the corresponding representation in space for one orientation. 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. This is useful for the measures we wish to perform. However, there is no evidence that the mammalian visual system uses a rectangular grid and indeed, other grids have been proposed (Watson and Ahumada, 1989). However, as long as the sampling distance is roughly proportional to the size of the receptive fields as discussed below, the conclusions of this paper remain unaffected.

The division of the information space with these two-dimensional transforms is analogous to the 1-dimensional transforms described in Figure 5. Unfortunately, one requires a four-dimensional plot to cover the full 2-D space by 2-D frequency trade-off. If we consider only a single orientation for the purposes of the display, we may show the space-frequency trade-off using the representations shown in Figure 6a and b. Figure 6a shows a representation where the bandwidths are constant at all frequencies (Gabor). Figure 6b shows a representation where the bandwidths are proportional to frequency (wavelet).

As with the one-dimensional codes, the sampling in space and frequency is proportional to size of the functions in space and frequency. In particular, for the transforms described in this paper the sampling distance is set to:

$$\Delta k = \sqrt{2\pi}\sigma_k \tag{1.6}$$

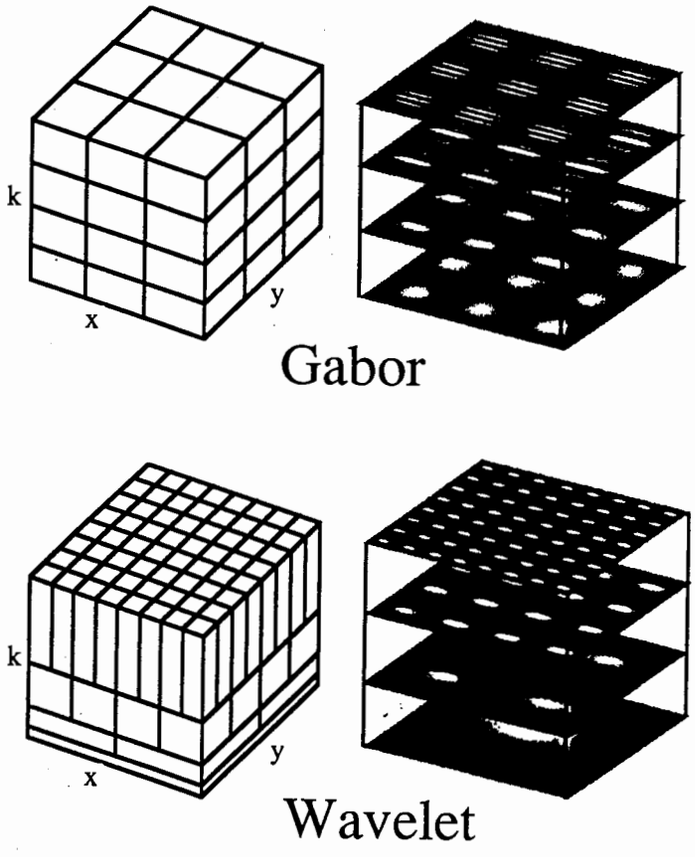


Figure 6. Three-dimensional information diagrams and their corresponding basis functions. The information diagrams are actually four-dimensional (u, v, x, y) but we have limited the diagram to the representation of single orientation, to allow a graphical description. The Gabor code has constant size and constant bandwidths while the Wavelet has bandwidths which increase with frequency as shown and results in a set of self-similar basis functions

$$\Delta x = \sqrt{2\pi}\sigma_x \tag{1.7}$$

For a Gaussian or a Gabor function defined by equation 1.1, the relation between the width in space and frequency is

$$\sigma_k = \frac{1}{2\pi\sigma_x} \tag{1.8}$$

By selecting the sampling schemes like that described above, it is is possible to produce a family of codes with the same overlap in both space and frequency and where the extent of the overlap does not depend on the particular bandwidths selected. In other words, by following the correct sampling it is possible to produce a family of 1:1 transforms where the overlap between neighbouring basis functions is the same in both space and frequency.

With this constraint, both the Wavelet and the Gabor transforms are complete, in that they represent all the information in the original image, and they are critically sampled (1 : 1) since this coding scheme represents each image with a constant number of functions (e.g., 65536) independent of the parameters that are chosen. However, it must be emphasized that these two types of transforms represent only a tiny fraction of the possible set of complete codes.

If the only goal is to represent the original data completely, then it will not be possible to show any selective advantage for any of the transforms described above. To understand why the mammalian visual system would evolve its particular "wavelet-like" strategy, one must know more about the task that the visual system must face. The answer proposed in this paper is that the visual environment has a particular structure that allows such a representation to be efficient for the natural environment. In the next, section we will consider some of the statistics of natural scenes which appear to be relevant to our understanding of the mammalian visual system.

2 The Statistical Structure of Natural Scenes

In this section, we will be looking at some of the statistical properties of natural scenes. For many people, this may seem like an odd notion. At first glance, it may not seem possible to classify something like a "natural image". It may seem that the images that are 'typically' focused on our retina must differ so widely from scene to scene that such classification would be either impossible or meaningless.

However, our visual environment is not at all random. We live in a highly structured visual world where objects and surfaces reflect light in ways that are determined by specific physical laws. Such structure limits the range of images that we are likely to encounter. There exists a degree of predictability or redundancy in our environment and it is proposed that

the mammalian visual system has taken advantage of this redundancy to produce a representation which is efficient at representing these images.

Images: In the next sections, we will be showing that our environment is not random. The data we will be considering consist of digitized photographs of the natural environment. They consist of a variety of scenes some of which are shown in Figure 7. The only photographic restriction was that the images have no manmade features (buildings, signs etc.) since these tend to have different statistical structures (e.g., a higher probability of long straight lines). The images used in this study were taken from the environment around Ithaca, NewYork. Images were photographed with Kodak XP1 film using a 35mm camera . Photographic negatives were scanned using a Barneyscan digitizer that provided a resolution of 512x512 pixels per 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).

2.1 Scale invariance

There are a number of statistical properties that we might consider when looking at natural scenes. Any structures that occur with some degree of predictability may help in understanding why the mammalian visual system uses the coding scheme described above. However, in this paper we will focus on the types of structures that would be expected if the statistics of the environment are scale invariant.

That is, suppose we assume that the global statistics of our images do not change as we "zoom" into any particular region of the image. As with most 'fractals' we can describe this invariance as:

$$G(x,y) = a^p G(x/a, y/a) \tag{2.1}$$

and if we force $p = 0$, which forces the variance (or contrast in our case) to also be maintained across scale, then

$$G(x,y) = G(x/a, y/a) \tag{2.2}$$

What sorts of statistical structures do we expect from this simple restriction? In this paper, we will be considering two forms of statistics: invariance in contrast across scale (reflected in the power spectrum) and invariance in the local structure (reflected in the phase spectrum). It is suggested that the constraint of scale invariance produces, (a) several of specific limitations on the range of possible images that can occur, (b) produces some accurate predictions of the statistical structures of natural

Figure 7. Eight examples of the types of images that were used in this study. All of the images were calibrated to take account of the optics of the camera and the non-linearities of the film

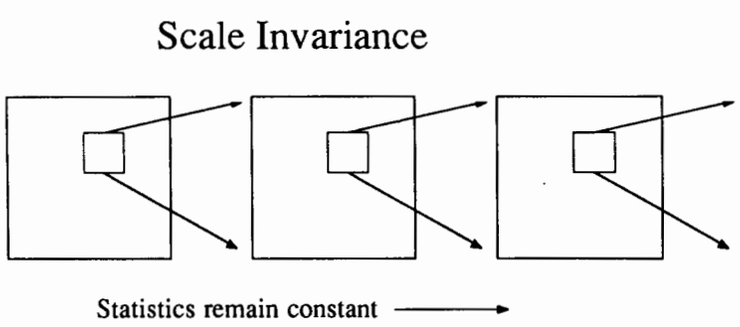


Figure 8. For an image to be scale invariant, the statistics of the image must remain constant as one magnifies any local region of the image. In this paper, we wish to consider both how the amplitude and phase spectra scale in natural scenes

scenes, and (c) provides us with a possible explanation of why the mammalian visual system would use the self-similar 'wavelet' strategy .

Invariance in contrast over scales

Consider the images shown in Figure 7. Each of these has information at a variety of scales. In a real scene, if one magnifies any particular region one is not likely to find an increase or decrease in the amount of contrast at any scale. In a photograph, the contrast as measured by the variance in pixel intensities should remain roughly constant independent of the viewing distance. On average, one should not expect to find either more or less variance in the image as one magnifies a particular region.

This invariance in "contrast" across scale is best described in terms of the amplitude or power spectrum of the image. Since the variance of the two-dimensional image is proportional to the volume under the power spectrum (Parseval's theorem), it is possible to determine the power spectrum that results in constant contrast at all scales.

Consider an image with an amount of energy E between frequency k and frequency nk. A magnification will shift the range of frequencies to the range from ak to ank. The total energy between frequencies k and nk is also shifted into the range of ak to ank.

Assume that we have an image in which the energy density at any frequency is $g(k)$. For a two-dimensional spectrum, the total energy at frequency (k) summed across all orientations will be:

$$E(k) = (2\pi k)g(k) \tag{2.3}$$

For a band of frequencies to have constant energy at all magnification requires:

$$\int_a^{na} (2\pi k) g(k) dk = C \tag{2.4}$$

where C is a constant. To maintain constant energy at all scales requires that:

$$g(k) \propto 1/k^2 \tag{2.5}$$

Assuming that the resolution is proportional to the window size, if the power spectrum falls off as $1/k^2$ this implies that the two-dimensional image will have constant variance at all scales. That is, it would be possible to expand any local region of the image and maintain the same variance.

Previously, it was noted (Field, 1987) that this notion of scale invariance provides a good description of the amplitude spectra of natural scenes. The amplitude spectra of six images were shown to fall off as a function of frequency by a factor of approximately $1/k$ (the power spectra fall by $1/k^2$).

Figure 9 shows an image along with a portion of its two dimensional amplitude spectrum. When the falloff averaged across orientations is plotted on log-log coordinates we get plots like that shown in Figure 10a. Figure 10b shows the falloffs for 85 scenes collected from a variety of natural environments. Although the results are not identical for all these scenes, the slopes of the amplitude spectra average approximately 1.1 or a power spectrum of $1/k^{2.2}$. This implies a fractal dimension of 2.9 (almost space-filling) and are therefore near to scale invariant.[3]

Such patterns show approximately equal energy in equal octaves. As previously noted (Field, 1987), this form of redundancy may allow a visual system to calibrate its gain to match the expected distribution of energy. However, as we will see in the next section, knowledge of the power spectrum is not sufficient to predict that wavelets will be effective codes.

Scale-invariance in the phase spectrum

Although the amplitude spectra represents one property of images that follows from their scale-invariance, there exists other structure which is more relevant to the understanding of visual codes. Consider the two images

[3]Although these images were calibrated for the cameras optics and film nonlinearities, a number of parameters can produce slopes steeper than $1/f$. An actual optical system has limited depth of field which makes it impossible to have all points in a three-dimensional environment in perfect focus (these images were collected with a small aperture). Any motion in the scene or camera will also move the spectra to steeper slopes. Finally, if an image contains significant blank regions (e.g. regions of sky), then the image will not be scale invariant at all locations, and the slopes will be steeper.

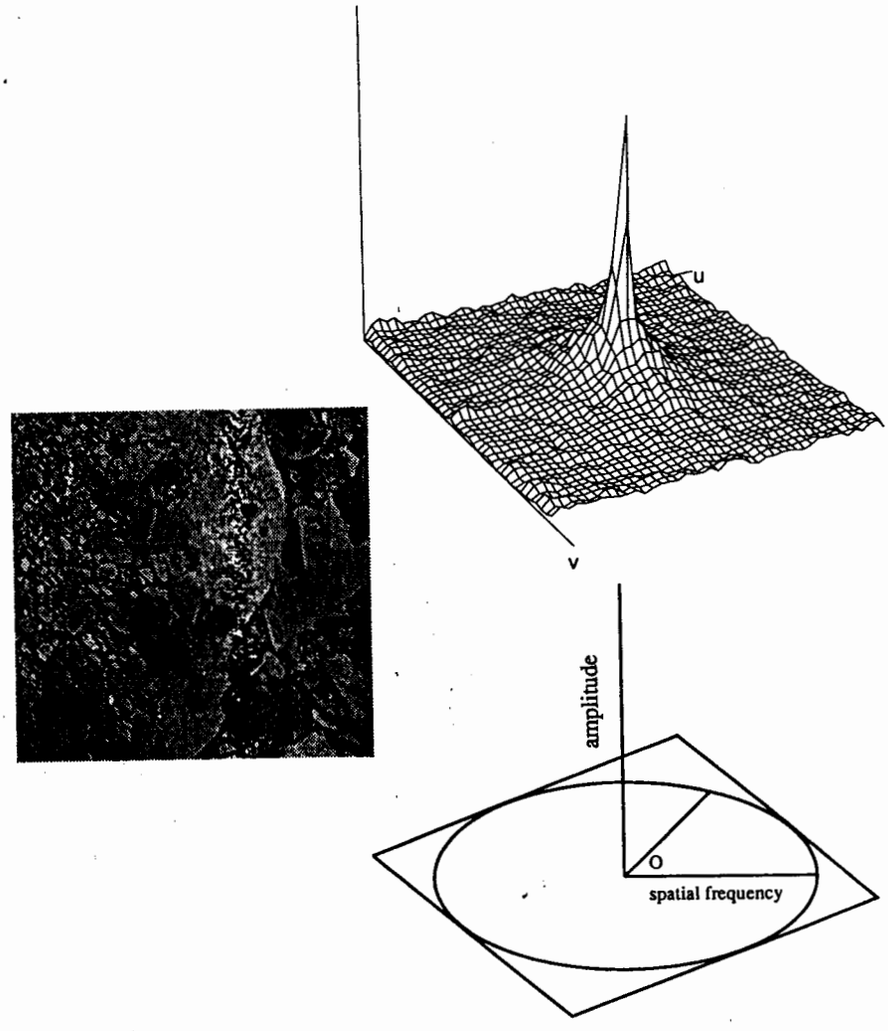


Figure 9. A natural scene along with its two-dimensional amplitude spectra

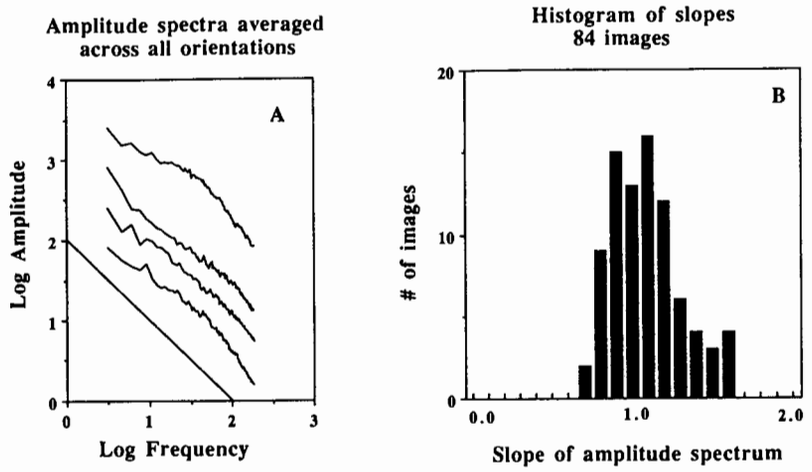
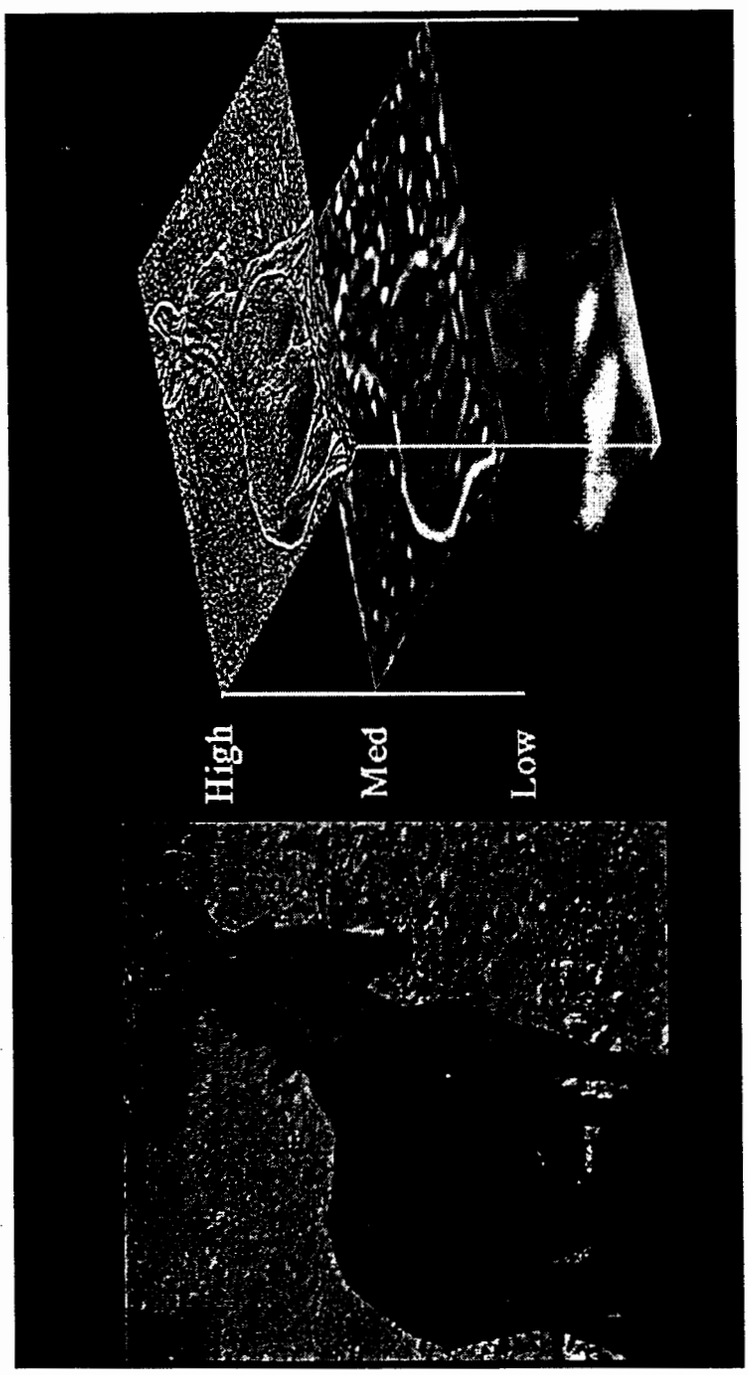


Figure 10. Part (a) shows the falloffs for the top four images in Figure 7 (each shifted up for clarity). On these log-log coordinates, the amplitude spectra have a slope of around -1. Graph (b) shows a histogram of the slopes for 85 images. The mean slope of the amplitude spectra is -1.1 (power spectra fall as $k^{-2.2}$)

shown in Figure 11a and b. Each of these images has the same amplitude spectrum and therefore the same auto-correlation function. The image in 11b was created by filtering white noise to the same amplitude spectrum as in 11a. This is equivalent to randomizing the phases of the Fourier coefficients while keeping the amplitudes constant. The two images are considerably different. The image on the right does not contain any of the lines or edges shown in the image on the left.

The presence of edges and lines in an image corresponds to a type redundancy between the different scales of the image which is destroyed when the phases are randomized. Figure 11 also shows the two images sliced into low-middle and high frequency bands. An image with random phases, (Figure 11b) will show no redundancy across scales. However, one can see that in Figure 11a , there exists a number of structures which extend across the different frequency bands. Many of the edges found in the low frequency bands are also found in the medium and high frequency bands.


High
Med
Low

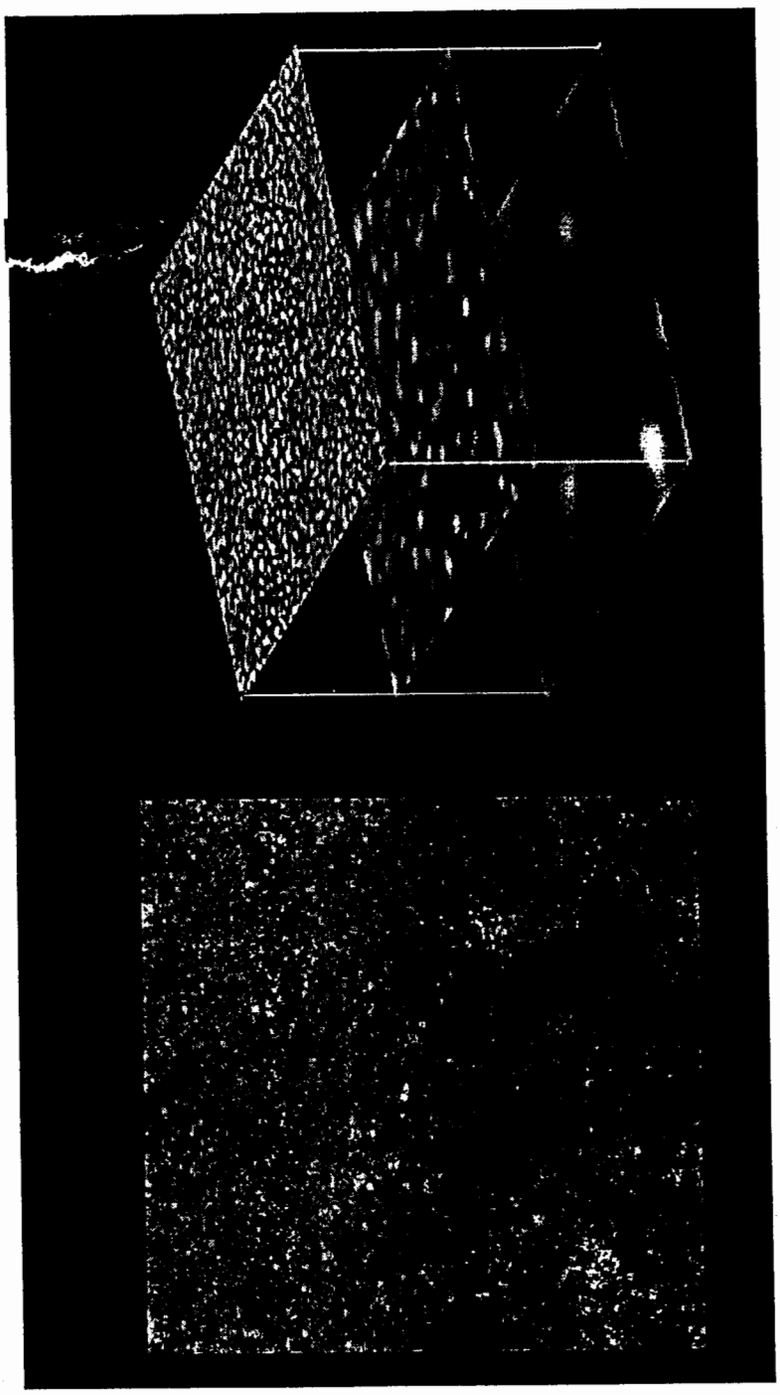


Figure 11. This figure shows the result of splitting an image into three frequency bands. For a natural scene (a), there is likely to be a variety of features which extend across different bands. The second image (b) was created by filtering white noise to produce a power spectrum that falls as k^{-2} . In this image, there are no structures that extend across the different bands. It is proposed that a wavelet code is effective when the phases are aligned across neighbouring different bands in a manner predicted from scale-invariance

These features exist because some degree of alignment exists between the phases at different frequencies. However, the problem of identifying and measuring this alignment is not a simple one. When there exists more than one location where these broadband features are present, the phases of the global Fourier coefficients become extremely complex. To detect this alignment of phases, one must define the relative phases locally in space as well as locally in spatial frequency.

To understand what we must look for, consider how a simple function like that shown in the top of Figure 12 will change when the function is scaled in size (i.e., width). The figure shows examples of four one-dimensional functions that vary in size and three ways of representing them. The left of Figure 12 shows the spectra of these four functions. For functions with constant amplitude, when the width is decreased:

$$G(x) \rightarrow G(ax) \tag{2.6}$$

the spectrum of each function will drop in amplitude, shift to higher frequencies and increase in bandwidth.

$$F(k) \rightarrow F(k/a)/a \tag{2.7}$$

The middle of Figure 12 shows a space-frequency plot of these four functions. The representation demonstrates the trade-off between space and frequency described earlier with the Gabor transforms.

All of the functions shown in Figure 12 have similarities in their phase spectra. The Fourier coefficients are all in cosine phase when the spectrum is defined at the peak of the function. Indeed, in terms of the phase spectra, a feature like an edge or a line marks locations where the phases are locally aligned. The middle of Figure 12 can be considered to show the regions in space-frequency where the alignment occurs. This extent over which the phases are aligned will be referred in general terms as a local **"phase structure."**

An image which consists of arrays of scaled functions will show a greater range of alignment at the high frequencies relative to the low frequencies. A waveform that is invariant over scale requires that the bandwidth of the phase structures (the frequency window over which the phases are aligned) be proportional to frequency.

The concept of phase alignment becomes more difficult to define in an image with a large number of phase structures. Consider a waveform in which we distribute the functions like that shown in Figure 12 to make a waveform like that shown in Figure 13a. To make this type of waveform, one takes a large function $g(x)$ and places it randomly on the waveform $g(x - x_n)$. Then one takes two such functions of half the size $g(2(x - x_n)$ and adds those to the waveform at random positions. Halve the size again

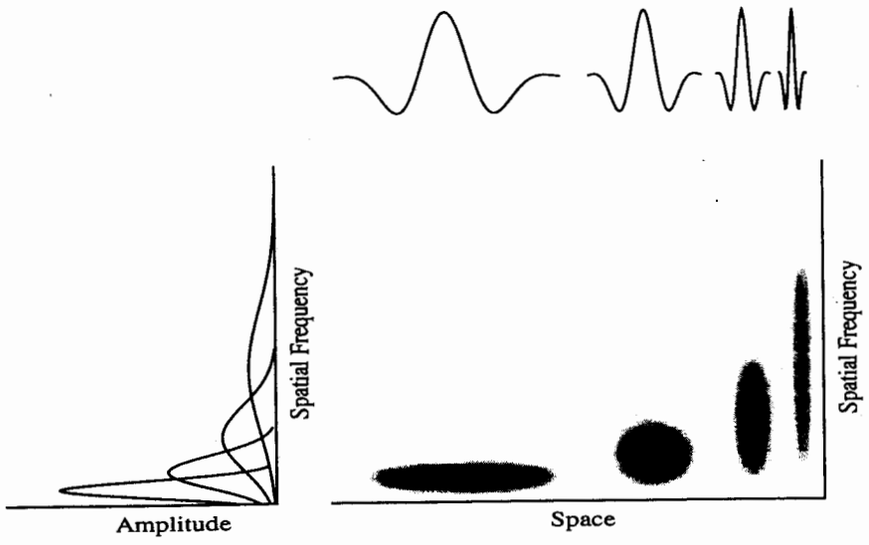


Figure 12. Space-frequency demonstration of how localized functions scale in space and frequency. The graph on the left shows the amplitude spectra of these four functions and the middle graph provides a representation in both space and frequency. For each of the functions shown, the phases are aligned across a local region of space and frequency. By adding a large number of functions like this, one obtains the graph shown in Figure 13

and add four of these to the waveform. By continuing this process one ends up with the waveform like that shown at the bottom of 13a.

The waveform is described by the equation:

$$\sum_{m=0}^{a}\sum_{n-1}^{2^m} g(2^m(x - x_{nm})) \tag{2.8}$$

where x_{nm} is randomly and independently distributed across the waveform. In Figure 13a:

$$g(x) = sin(2\pi x/\sigma)e^{-x2/2\sigma^2} \tag{2.9}$$

For the waveform shown, there were only 9 iterations ($a = 9$) but theoretically one can design a scale invariant image by allowing m to be unbounded. A similar waveform will be created by taking a random collection of basis functions from the wavelet code (e.g., one out of four randomly) like that shown in Figure 5 and summing them together.

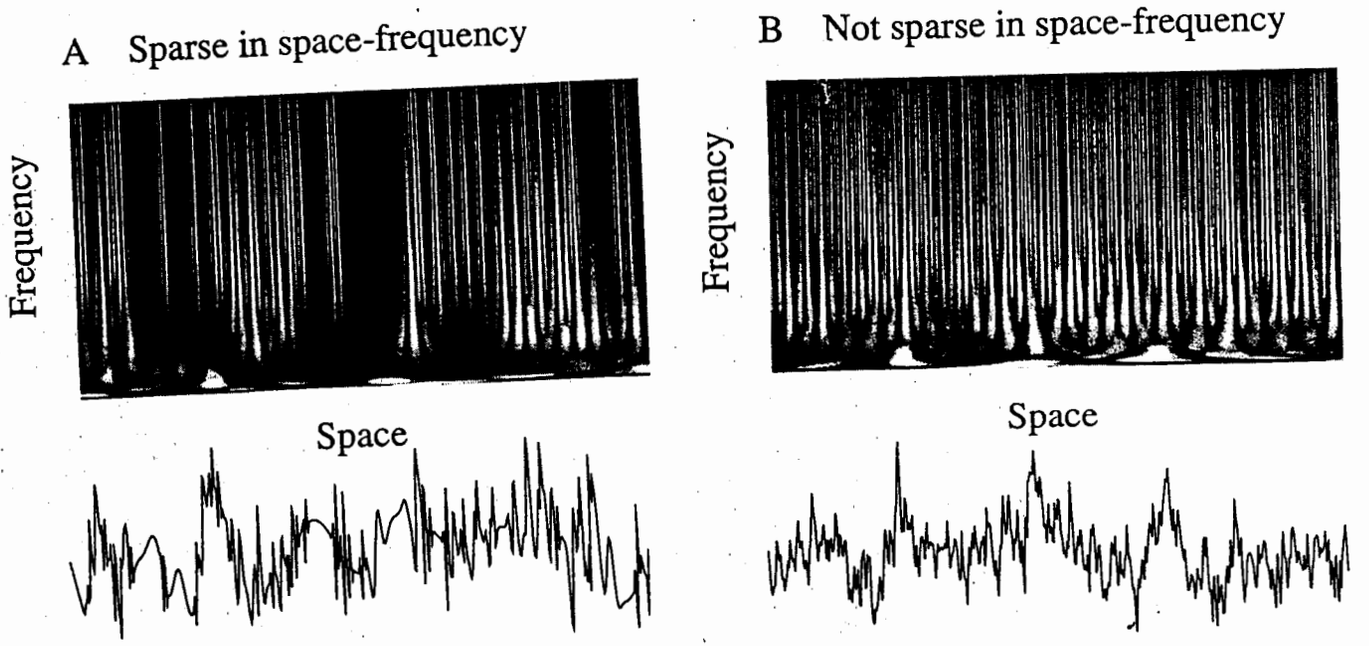


Figure 13. By distributing a family of self-similar functions according to equation (16), it is possible to create a waveform that is scale-invariant with regards to both the power spectrum and the phase spectrum. In such an image, the local phases have a greater range of alignment at the high frequencies and this is captured in the wavelet representation. When described by a continuous wavelet which matches the structure, one can see that the representation is relatively sparse. One will not see this redundancy, if its described by the points in space or described by the Fourier coefficients. The waveform in (b) has the approximately the same power spectrum (k^{-1}) but with a random phase spectrum. When described by the wavelet, almost all basis

Such an image will be scale invariant with regards to both the power spectrum and the phase spectrum. But more important, when such a waveform is represented by a wavelet code, the representation will be sparse. By sparse, we mean that the waveform can be represented by a relatively small subset of the total set of basis functions (If one out of four basis functions went in to make the waveform, then one out of four will be needed to represent it). Figure 13a also shows an example of a continuous wavelet transform of this waveform. One can see that there are a number of 'inactive' regions. The wavelet produces a sparse representation. The waveform is not sparse in space (a small subset of points in space cannot describe the waveform). It is also not sparse in frequency (a small subset of Fourier coefficients cannot describe the waveform). Only when represented in both space and frequency as with the wavelet does this waveform appear sparse.

The waveform in Figure 13b is an image with a similar power spectrum (k^{-1}) but has a random phase spectrum (pink noise). The waveform in 13a is sparse when represented by the wavelet, while the waveform in 13b is not sparse by any representation. In this paper, it is proposed that wavelets are effective codes when waveforms show this scale-invariant and sparse distribution of features. Indeed, it has been previously shown (Field, 1987) that wavelet codes can produce a sparse representation of natural scenes.

Does this imply that natural scenes have scale-invariant phase spectra? In the next section, a method is described for measuring these structures in complex images.

3 Measuring Redundancy in Two-dimensional Scenes

In Figures 11a and 13a, one expects to find some redundancy across the different scales. It might seem possible to determine the degree of alignment by correlating the images in different frequency bands. Unfortunately, the simple measure of correlation will not provide any meaningful results. Consider two filters that are orthogonal, independent of their relative position (e.g., selective to different frequency bands).

$$\int f_1(x) f_2(x - x_0) dx = 0 \quad \text{for all } x_0. \tag{3.1}$$

Then the convolution of those two filters with any waveform will also be orthogonal.

$$\int dx_0 \int f_1(x - x_0) g(x) dx \int f_2(y - x_0) g(y) dy = 0 \tag{3.2}$$

Since the correlation is simply a function of the product of the filtered images, the correlations will always be zero, to the extent that the two filters are orthogonal. In other words, if one attempts to determine the degree

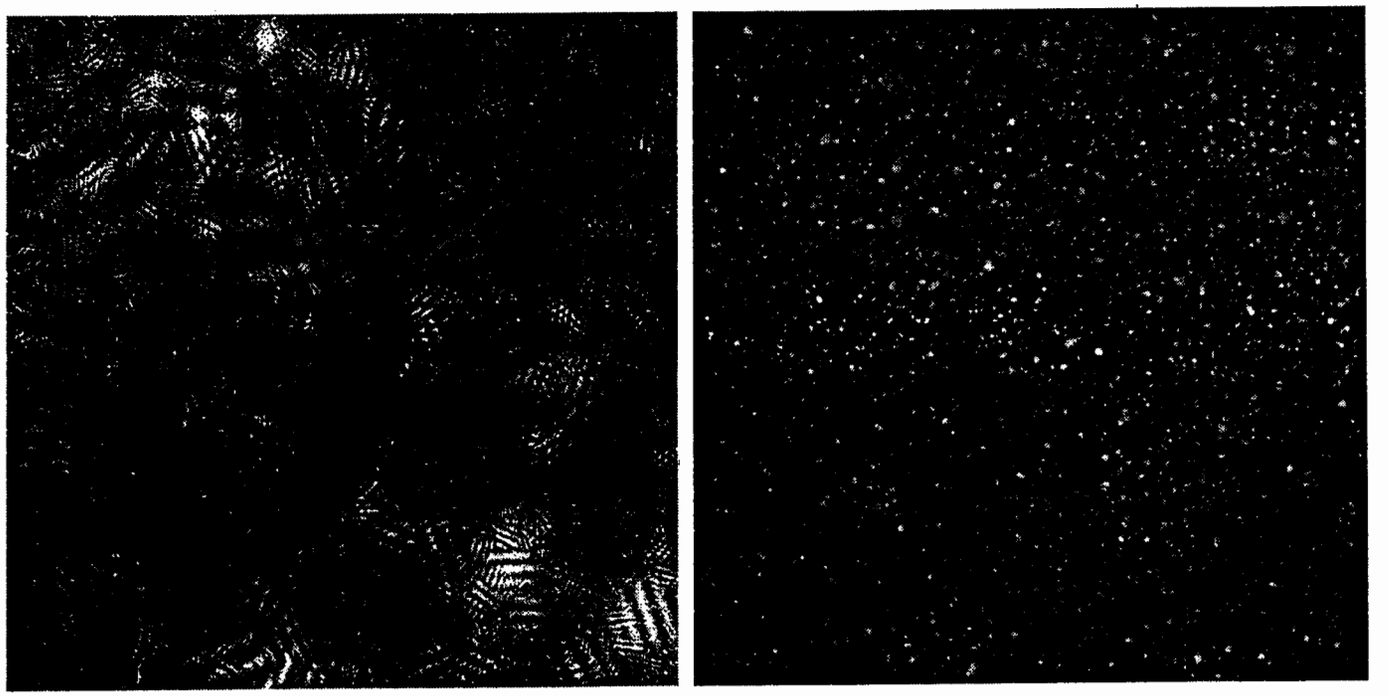


Figure 14. Two examples of images with scale invariant power spectra and phase spectra (the images are derived from equation (2.8) but sums over 4^m functions rather than 2^m and each function is two-dimensional). Unlike the image shown in 11b, these images have a greater phase alignment at high frequency and will produce a sparse output with a wavelet code. The image on the left is derived from functions in which the bandwidths are narrower than for the functions used in the image on the right

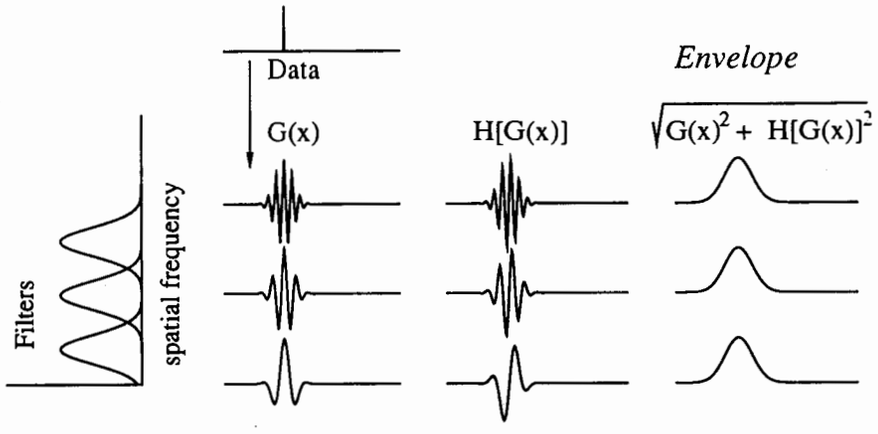


Figure 15. The correlation between any two frequency bands was determined by means of the 'envelope' or complex amplitude of the filtered response. The figure shows a waveform (a single point) filtered into three frequency bands. If the correlations had been determined by means of this filtered output alone, the correlations would not be significant (equation (3.2)). Correlations between the envelopes (vector sum of the filtered output and its Hilbert transform) provides a measure of the extent to which different features extend through frequency

of similarity across scale by measuring correlations between the filtered images like that shown in Figure 11b and 13b, one will obtain a correlation of 0.

Consider an image of a single point like that shown at the top Figure 15. When the image is filtered into different frequency bands, the resultant bands will be uncorrelated as long as the filters are orthogonal. To show that there exists redundancy across the scales requires a different measure.

If we consider that the filtered images represent a vector which is rotating in phase and changing slowly in amplitude, it becomes clear that it is the amplitude of the vectors that are correlated. One measure of this complex amplitude is described by the vector sum of the waveform and its Hilbert transform (Field, 1987; Morrone and Burr, 1988; Bracewell, 1989; Field, 1989)

$$E(x) = \sqrt{g(x)^2 + H(g(x))^2}. \tag{3.3}$$

The two dimensional analogue is shown in Figure 16. Although the filtered images will not be correlated, this is not the case for the envelopes derived from the filtered images. The envelopes of these filtered images

will be correlated to the extent the phases are aligned at similar locations in the two frequency bands.

Correlations were determined from the two-dimensional envelopes described above. The results we describe will show correlations between filters across both space and frequency. In both cases, the correlations are determined at distances that correspond to the filter size in space and frequency. To help understand this, consider the representation shown in Figure 12. Like in Figures 5 and 6, this representation shows how a two-dimensional image can be divided into functions that are localized in space and frequency. As with the representations in Figure 5 and 6, the spacing between basis functions is a function of the bandwidth. For these codes the spacing is based on a square grid where:

$$\Delta x/\sigma_x = \Delta k/\sigma_k = \sqrt{2\pi}. \tag{3.4}$$

The particular factor of $\sqrt{2\pi}$ is not crucial to the results presented in the next section. What is important for our analysis is that the spacing be proportional to the bandwidth and that the overlap between neighbours in space ($\Delta x/\sigma_x$) be equal to the overlap in frequency ($\Delta k/\sigma_k$). By comparing correlations across space and frequency using this sampling scheme, this representation allows us to compare the relative redundancy in space and frequency.

For the results discussed in this paper we will consider only the effect of the frequency bandwidth on the measured redundancy of the code. Since we are dealing with two-dimensional functions, it is also possible to discuss role of orientation tuning as well. However, that is beyond the scope of this paper. The orientation tuning will not be independently varied. Rather, the orientation tuning will be linked to the frequency tuning. Thus the equation of the tuning in frequency for the Gabor code is:

$$G(k_u, k_v) = e^{-((k_u - k_{u_o})(k_v - k_{v_o}))^2/2\sigma^2}. \tag{3.5}$$

We will be measuring the redundancy of our natural scenes using both a "Gabor code" and a "wavelet code." For the Gabor code, σ is constant for all frequencies.

$$\text{Gabor} \quad \sigma = \text{constant} \tag{3.6}$$

However, for the self-similar 'wavelet' code, σ is dependent on k. The basis functions have a similar spectrum but which conforms to the logarithmic mapping as shown in Figure 4. As described previously (Field, 1987), radially the function has the spectrum:

$$G(k) = e^{-(\ln(k/k_0))^2/2\ln(\sigma/k_o)^2} \tag{3.7}$$

and is Gaussian along the orthogonal axis.

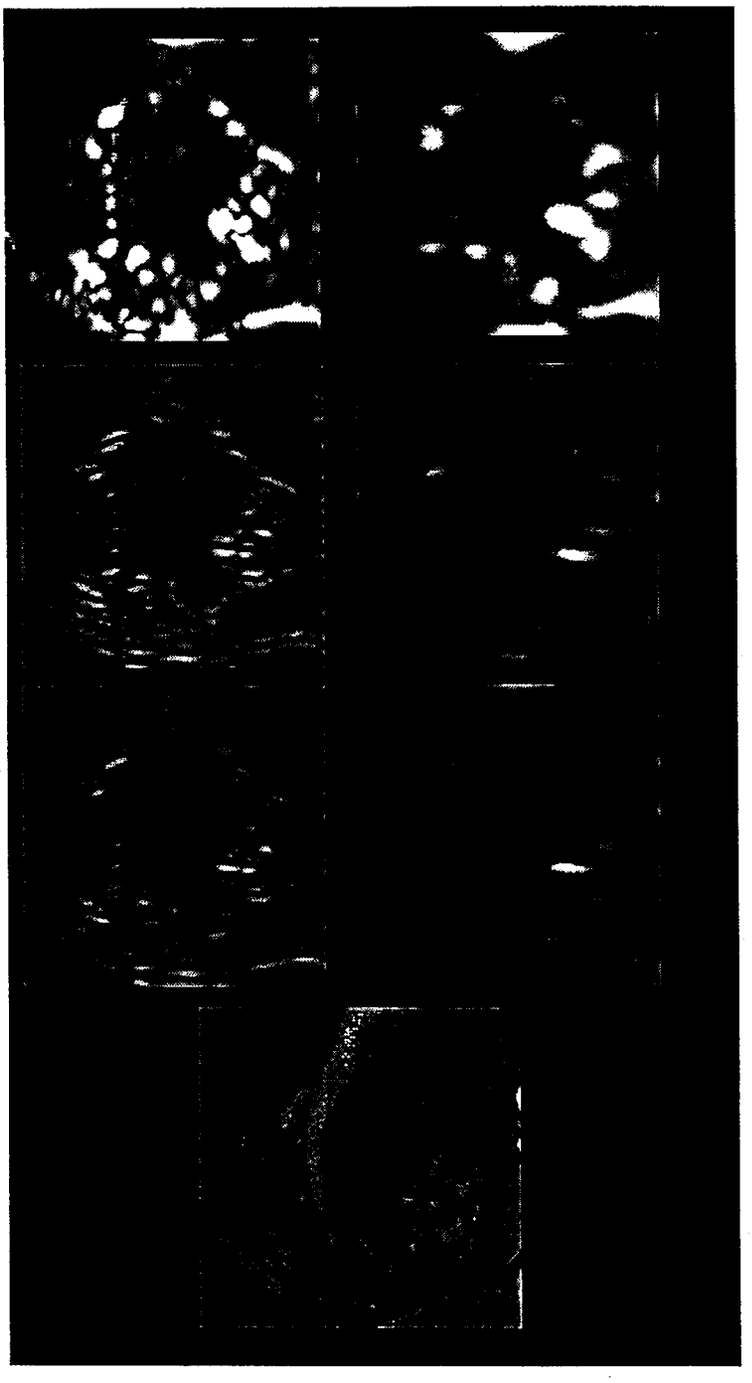


Figure 16. The two-dimensional version of Figure 15. The figure shows an image convolved with oriented wavelets at two frequencies along with their Hilbert transforms. The vector sums of these two orthogonal waveforms produces the envelopes shown on the right. Correlations between these two images provides a measure of the spectral distance over which the phases are aligned

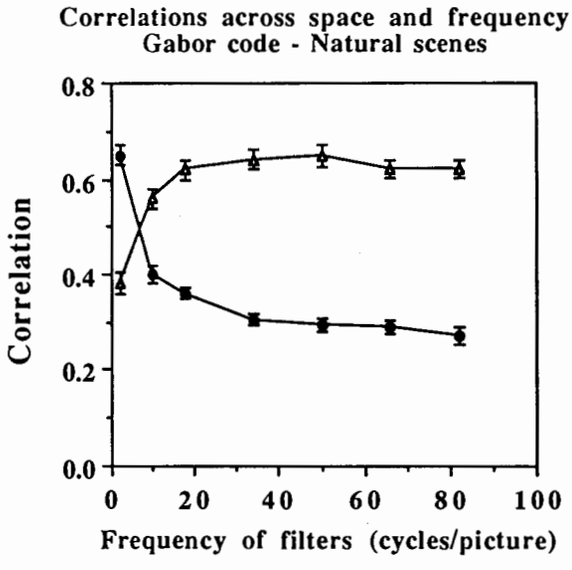


Figure 17. Results when measuring correlations between neighbouring Gabor filters for 24 natural scenes. The image is represented by arrays of basis functions that are the same same size in space and the same bandwidths in frequency (e.g., Figure 6a). The 'distance' between neighbours in space is proportional to their size and the 'distance' between the neighbours in frequency is proportional to their bandwidth. With increasing frequency, neighbours in frequency become more highly correlated while neighbours in space become less correlated. This would be expected if the structures in these images become more localized and increase in bandwidth with increasing frequency

3.1 Correlations with 2D Gabor

In the first set of results, we consider the correlations as a function of frequency. The representation of information in the image will be like that shown in Figure 6a. The results of these analyses are show in Figure 17. First, consider the correlations marked "frequency". This shows the correlations between data filtered at two neighbouring frequency bands (centred at k and Δk). One can see that at the low frequencies the correlations in frequency are relatively low while at high frequencies the correlations are high. The redundancy between frequency bands is higher at the high frequencies.

The curve labelled "space" shows the correlations between neighbouring Gabor functions in space. Separated by a distance of Δx as described

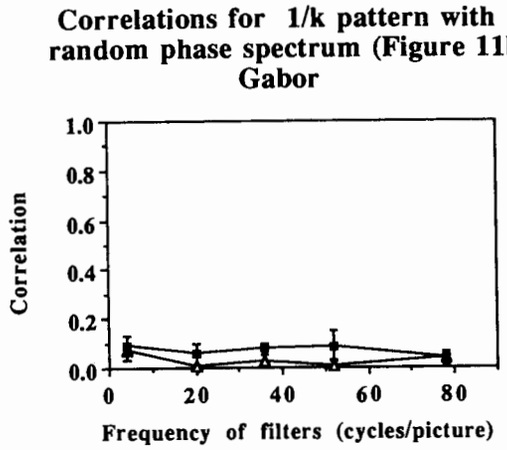


Figure 18. Results when measuring correlations between neighbouring Gabor filters for the image shown in Figure 11b. The image has a random phase spectrum. Since this image has no structures which extend across space or frequency, the correlations are near 0

above. This measure was determined by shifting the filtered image by a distance Δx (orthogonal to the orientation axis) and cross correlating.

$$C_k = \mathrm{Corr}(F_k(x,y), F_{k+\Delta k}(x,y)) \tag{3.8}$$

$$C_x = \mathrm{Corr}(F_k(x,y), F_k(x+ax, y+ay)) \tag{3.9}$$

$$\text{where } \sqrt{ax^2+ay^2} = \Delta x. \tag{3.10}$$

These results demonstrate that the correlation between neighbouring filters increases with increasing frequency. This is what would be expected if the phase alignment is scale invariant. That is, if one begins with a structure which is localized in space and frequency, scaling the structure down in size shifts the structure to higher frequencies and increases the range of frequencies over which the phases are aligned.

To demonstrate that this correlation is due to the phase spectra and not to either the power spectra or some unusual property of the code, the same correlations were determined for an image in which the phase spectrum was randomized (e.g., Figure 11b). The results are shown in Figure 18. As expected, the correlations fall to near 0.

These results suggest that our natural scenes contain an alignment of the phases that extends over greater ranges with increasing frequency. Do these results suggest that the phase spectra of our natural scenes are scale

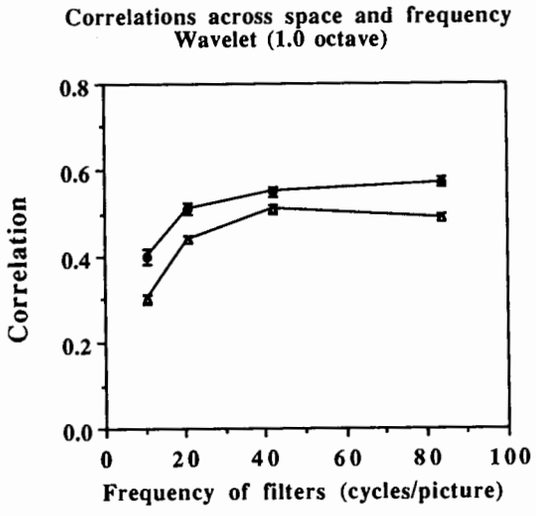


Figure 19. Results when measuring correlations between neighbouring wavelet filters for 24 natural scenes. The image is represented by arrays of self-similar basis functions in which the bandwidths are proportional to frequency (e.g., Figure 6b). The 'distance' between neighbours in space is proportional to their size and the 'distance' between the neighbours in frequency is proportional to their bandwidth. The results suggest that structures in these images follow the parameters of the wavelet and are expected if the images are roughly scale invariant. That is, structures at high frequencies tend to be smaller and have broader bandwidths

invariant? If so, a wavelet code should produce substantially different results.

3.2 Correlations with the wavelet

In the wavelet transform, the basis functions become smaller and increase their bandwidth with increasing frequency. If these phase structures also increase with frequency then the correlations should be the roughly constant across the spectrum.

Figure 19 shows the results of measuring correlations when the basis functions are wavelets like those in Figure 12. The correlations were determined in the same way where the sampling distance is proportional to the bandwidth. Although the correlations show some increase at higher frequencies, the difference between correlations in space and frequency remains roughly constant at different frequencies.

The results suggest that with the wavelet code, the extent of the phase alignment increases proportional to frequency supporting the notion that the phase spectra are scale invariant.

4 Discussion

In this paper, we have discussed two forms of scale invariance. The first form of scale invariance is reflected in the power spectra. For two-dimensional images, scale invariance implies a power spectrum which falls in the range of $1/k^2$.

The second form of scale invariance relates to the local structure of the image and is dependent on the phase spectrum. Knowing that the spectrum falls as $1/k^2$ does not imply that the phase spectrum is also scale invariant. Figure 20 shows examples of two images which have $1/k^2$ power spectrum but do not consist of scale invariant structures.

The concept of phase alignment in local structures was discussed and it was noted that the bandwidth of these structures should be proportional to frequency in a scale invariant image. However, the scale invariance criterion does not determine what this bandwidth must be. Indeed, Figure 14 showed two examples of scale invariant images with quite different bandwidths.

Although the images in Figure 14 may appear somewhat naturalistic they do not appear anything like our natural scenes. One of the important differences between the structures in these images and those in a natural scene is that in these synthetic images the structures are selected and positioned quite independently of each other.

It is also possible to have phase structures which appear to have finite lengths when measured by the correlation techniques, but which are actually continuous and drifting across the space-frequency spectrum. Figure 21 provides an example of such an image. Fractal edges like these have structures which shift and change orientation as one moves from low to high frequencies. The phases are not aligned across the different frequency bands.

It is certain that natural scenes do not consist of randomly positioned and randomly oriented phase structures. There is certainly some continuity across scale, frequency and position. However, our present methods do not allow us to measure the extent of this continuity. Nonetheless, I would like to suggest that natural scenes, like many natural phenomena and like the fractal edge shown in Figure 21, consist of a variety of continuous phase structures that twist and shift as they move across scales. A wavelet code with orientation and frequency selective mechanisms provides a good match to the shifting and twisting exhibited by these structures.

The mammalian visual system has evolved a strategy efficiently of mapping these phase tracks. Consider the analogy of walking down a curved

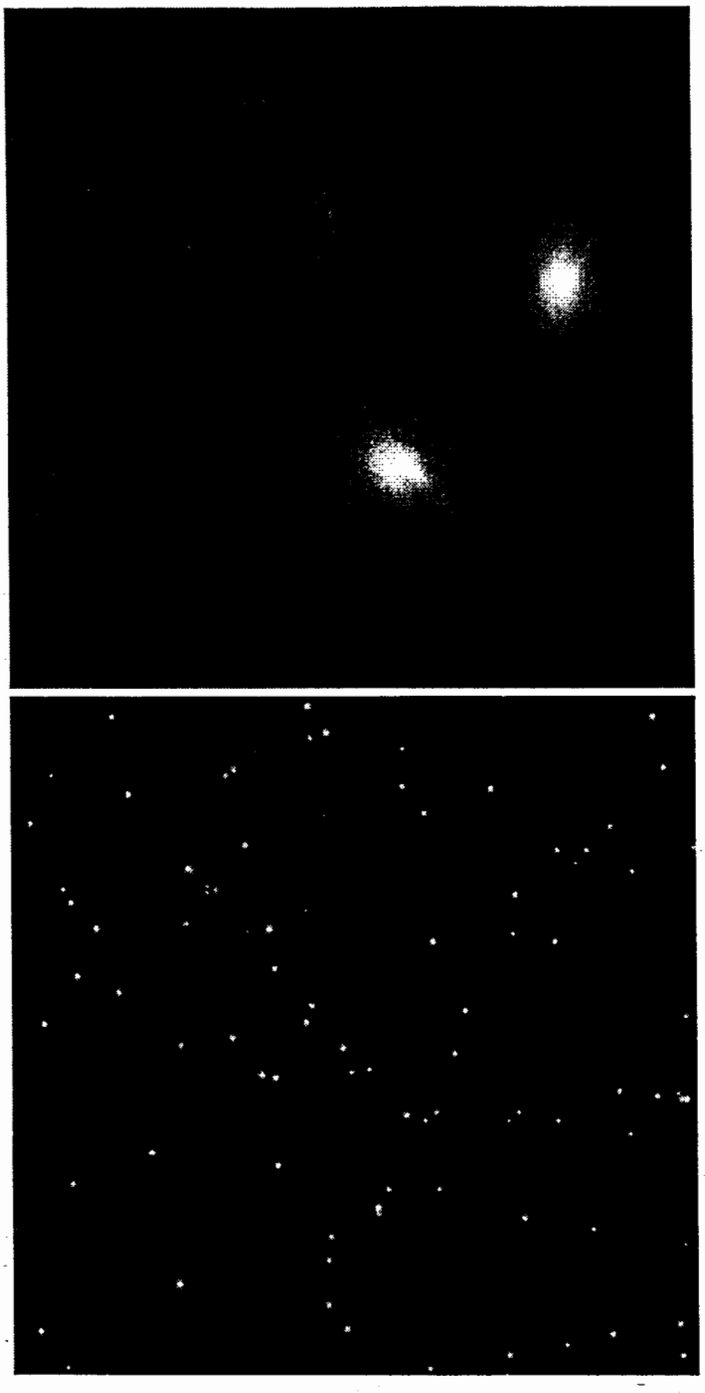


Figure 20. Two images in which the bandwidth of structures do not scale with frequency. Both have amplitude spectra which fall as approximately k^{-1} (power spectra of k^{-2}) and both have non-random phase spectra, but a wavelet representation is not ideal for either, if the goal is to produce a sparse code. Image (a) consists of a sparse array of points filtered to achieve a k^{-2} spectrum. Image (b) consists of a sum of randomly oriented Gabor functions with constant σ and amplitude decreasing with increasing frequency

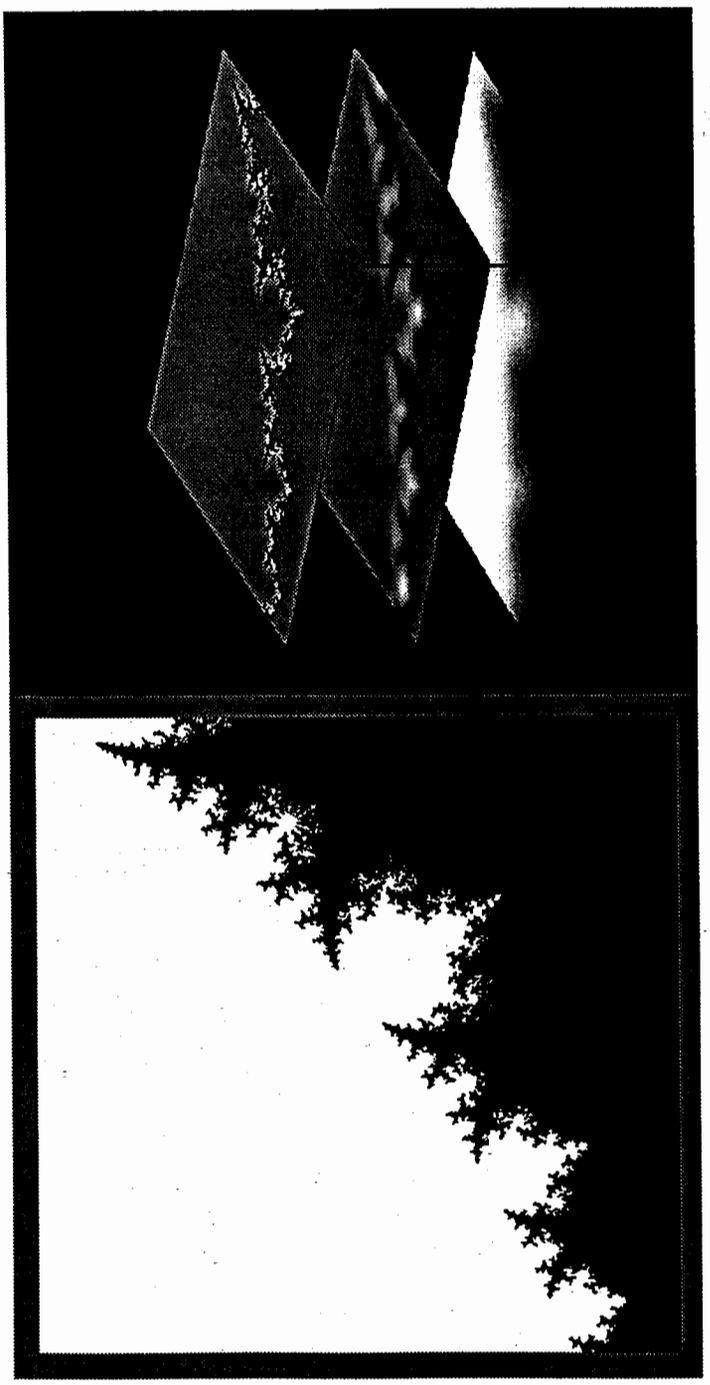


Figure 21. An image with structures which scale with frequency but which are not at independent locations at different frequencies. Notice that local regions of the edge shift and rotate across scale

path. How does one walk down a path with the minimum number of steps? By adjusting one's step size to match the typical curvature of the path (e.g., long steps when the path tends to be straight), one can produce an optimal number of steps to traverse the path. If one makes steps that are longer than the optimal size, then one will find that a step often takes you off the path requiring a step back to return to the path. Step sizes that are too small result in a larger number of steps than required to traverse the entire path. Most of the steps will be redundant since they will be in the same direction as the previous step. By this analogy, the visual system code achieves two goals. First, it produces a sparse representation by producing the minimum number of active neurons (steps) to represent natural scenes. This idea is supported by the results of Field (1987), Daugman (1989), and Zetzsche (1990) that these codes do indeed produce a sparse representation of natural scenes and is in line with proposals by Barlow (e.g., 1961) that visual systems should take advantage of the redundancy of their environment.

It must be emphasized what is meant by sparse. For any particular scene, only a small subset of the population of cortical cells needs to be active to represent a scene. However, across all scenes, every cell is as likely to be active as any other. No coefficient is unnecessary and therefore no coefficient can be removed from the transform without loss. Thus, this coding does not reduce the dimensionality.

For a visual system, a sparse representation will have a number of advantages. Most of the image energy is concentrated into a few active neurons providing a good signal to noise ratio if there exists some degree of background noise in these neurons. Probably more important, however, is that the information in the image is concentrated. Each active cell is providing a more complete description of where features (i.e., points of phase alignment) are present. This is likely to allow later stages of the visual system an easier time of analyzing and identifying the structure in the world since the number of possible images that can be produced from this sparse set has been reduced considerably.

The second achievement of this representation is that if there exists structures which are locally continuous like that shown in Figure 21, then the activity of a cell provides information regarding the likely locations of other features. The locations of features at one scale can provide a guide for the search for features at other scales. Indeed, there exists psychophysical evidence that these 'phase tracks' are used by human observers (Hayes, 1988). Following these tracks may provide a method for efficiently searching through the vast array of data to allow the identification of meaningful and relevant structure.

5 Conclusion

Although there has been considerable recent interest in self-similar transforms, there remains the question of when they are effective and when other types of transform would be more appropriate given the particular goal in mind. In this paper we have looked at some of the scale invariant properties of natural scenes and conclude that the mammalian visual system has evolved a code which is particularly effective at representing the particular scale invariant properties of the environment. For natural scenes and for many scale invariant synthetic scenes, this code produces a sparse, informative representation.

However, it is NOT possible to conclude that a self-similar code is an effective representation for all natural phenomena. Scale invariant codes can be efficient at representing scale invariant phase structures, if efficiency is defined in terms of sparseness. If the goal is not to produce a sparse representation or if the data do not show scale invariant properties, the wavelet representation may be quite inappropriate. For many forms of data, other transforms may be more useful. However, for the mammalian visual system and its relation to the mammalian visual environment, the self-similar solution may be the most effective for what it needs to achieve.

6 Acknowledgements

I would like to thank C. Vascillicos and A.B. Watson for their comments on an early draft of this chapter. I would also like to acknowledge Nuala Brady for her help in collecting and digitizing many of the images and for her work behind Figure 10.

References

Adelson, E.H., Simoncelli, E., and Hingorani, R. (1987) Orthogonal pyramid transforms for image coding, SPIE Visual Communications and Image Processing II, 845.

Adelson, E., Simoncelli, E., and Freeman, W.T. (1991) Pyramids and multiscale representations. In A. Gorea (Ed.), Representations of vision (pp. 3-16), Cambridge: Cambridge University Press.

Barlow, H.B. (1961) The coding of sensory messages. Current Problems in Animal Behavior, Cambridge, Cambridge U. Press.

Blakemore, C., and Campbell, F.W. (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images, Journal of Physiology, 203, 237-260.

Bovik, A.C., Clark, M., and Geisler, W. (1990) Multichannel texture analysis using localized spatial filters, IEEE Pattern Analysis and Machine Intelligence, 12(1), 55-73.

Bracewell, R.N. (1989) The Fourier Transform and Its Applications, 2nd edition, New York: McGraw-Hill.

Combes, J.M., Grossman, A., Tchamitchian, P., (editors) (1989) Wavelets, Time-Frequency Methods and Phase Space, 1st Int. Wavelets. Conf., Marseille, Dec. 1987, Inverse Probl. Theoret. Imaging, Springer.

Campbell, F.W., Cooper, G.F., Robson, J.G., and Sachs, M.B. (1969) The spatial selectivity of visual cells of the cat and the squirrel monkey, Proceedings of the Physiological Society, 120-121.

Campbell, F.W., Cleland, B.G., Cooper, G.F., and Enroth-Cugell, C. (1968) The angular selectivity of visual cortical cells to moving gratings, Journal of Physiology 198, 237-250.

Campbell, F.W., Cooper, G.F., and Enroth-Cugell, C. (1969) The spatial selectivity of the visual cells of the cat, Journal of Physiology, 203, 223-235.

Campbell, F.W., and Robson, J.G. (1968) Application of Fourier analysis to the visibility of gratings, Journal of Physiology, 197, 551-556.

Daugman, J. (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America, 2(7), 1160-1169.

Daugman, J.G. (1988) Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, IEEE Transactions on Acoustics, Speech and Signal Processing, 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: Cambridge University Press.

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.

Field, D.J. (1987) Relations between the statistics of natural images and the response properties of cortical cells, Journal of the Optical Society of America, 4, 2379-2394.

Field, D.J., and Tolhurst, D.J. (1986) The structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex, Proceedings of the Royal Society. London. B, 228, 379-400.

Field, D.J. (1989) What the statistics of natural images tell us about visual coding, Proceedings of SPIE 1077 (269-276), Los Angeles, California.

Gabor, D. (1946) Theory of Communication, JIEE London, 93(III): 429-457.

Hawken, M.J., and Parker, A.J. (1987) Spatial properties of neurons in the monkey striate cortex, Proc. Roy. Soc. Lond. B. 231: 251-288.

Hayes, A. (1989) "Representation by images restricted in resolution and intensity range" PhD Thesis, Department of Psychology, University of Western Australia.

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

Jones, J., and Palmer, L. (1987a) The two-dimensional spatial structure of simple receptive fields in cat striate cortex, Journal of Neurophysiology, 58(6), 1187-1211.

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

Klein, S. A., and Levi, D.M. (1985) Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation, Journal of the Optical Society of America. 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, Biological Cybernetics. 43: 187-198.

Mallat, S.G. (1989) A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and machine intelligence, 11(7), 674-693.

Marcelja, S. (1980) Mathematical description of the responses of simple cortical cells, J. Opt. Soc. Am. 70: 1297-1300.

Marr, D., and Hildreth E. (1980) Theory of edge detection, Proceedings of the Royal Society of London. B 207: 187-217.

dimension. In Wavelets, Time-Frequency Methods and Phase Space. 1st Int. Wavelets. Conf., Marseille, Dec. 1987, Inverse Probl. Theoret. Imaging. Combes, J.M., Grossman, A., Tchamitchian, P (editors), Springer.

Sakitt, B., and Barlow, H.B. (1982) A model for the economical encoding of the visual image in cerebral cortex, Biological Cybernetics. 43: 97-108.

Simoncelli, E.P., and Adelson, E.H. (1990) Subband transforms. In J.W.Woods (Ed.), Subband image coding Norwell, MA: Kluwer Academic Publishers.

Stork, D.G., and Wilson, H.R. (1990) Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? J. Opt. Soc. Am. A. 7: 1362-1373.

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.

Watson, A.B. (1983) Detection and recognition of simple spatial forms. In O.J. Braddick, and A.C. Slade (Ed.), Physical and Biological Processing of Images Berlin: Springer-Verlag.

Watson, A.B. (1987) Efficiency of an image code based on human vision, Journal of the Optical Society of America A. 4(12): 2401-2417.

Watson, A.B., and Ahumada, A.J., Jr. (1989) A hexagonal orthogonal oriented pyramid as a model of image representation in visual cortex, IEEE Trans. Biomed. Eng. 36(1), 97-106.

Watson, A.B. (1990) Perceptual components architecture for digital video. Journal of the Optical Society of America, 7, 1943-1954.

Watson, A.B. (1991) Multidimensional pyramids in vision and video. In A. Gorea (Ed.), Representations of vision (pp. 17-26), Cambridge: Cambridge University Press.

Webster, M.A., and DeValois, R.L. (1985) Relationship between spatial-frequency and orientation tuning of striate-cortex cells, Journal of the Optical Society of America 2, 2, 1124-1132.

Young, R.A. (1986) The gaussian derivative model for spatial vision: I. Retinal mechanisms, Spatial Vision, 2, 273-293.

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 (Ed.) Amsterdan: North-Holland.