INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR.853 - 74

ANTONIO BORSELLINO COLLEGE ON NEUROPHYSICS

(15 May - 9 June 1995)

"A nonlinear model of feature detection"

David C. Burr
Istituto di Neurofisiologia
Consiglio Nazionale delle Ricerche
56100 Pisa
Italy

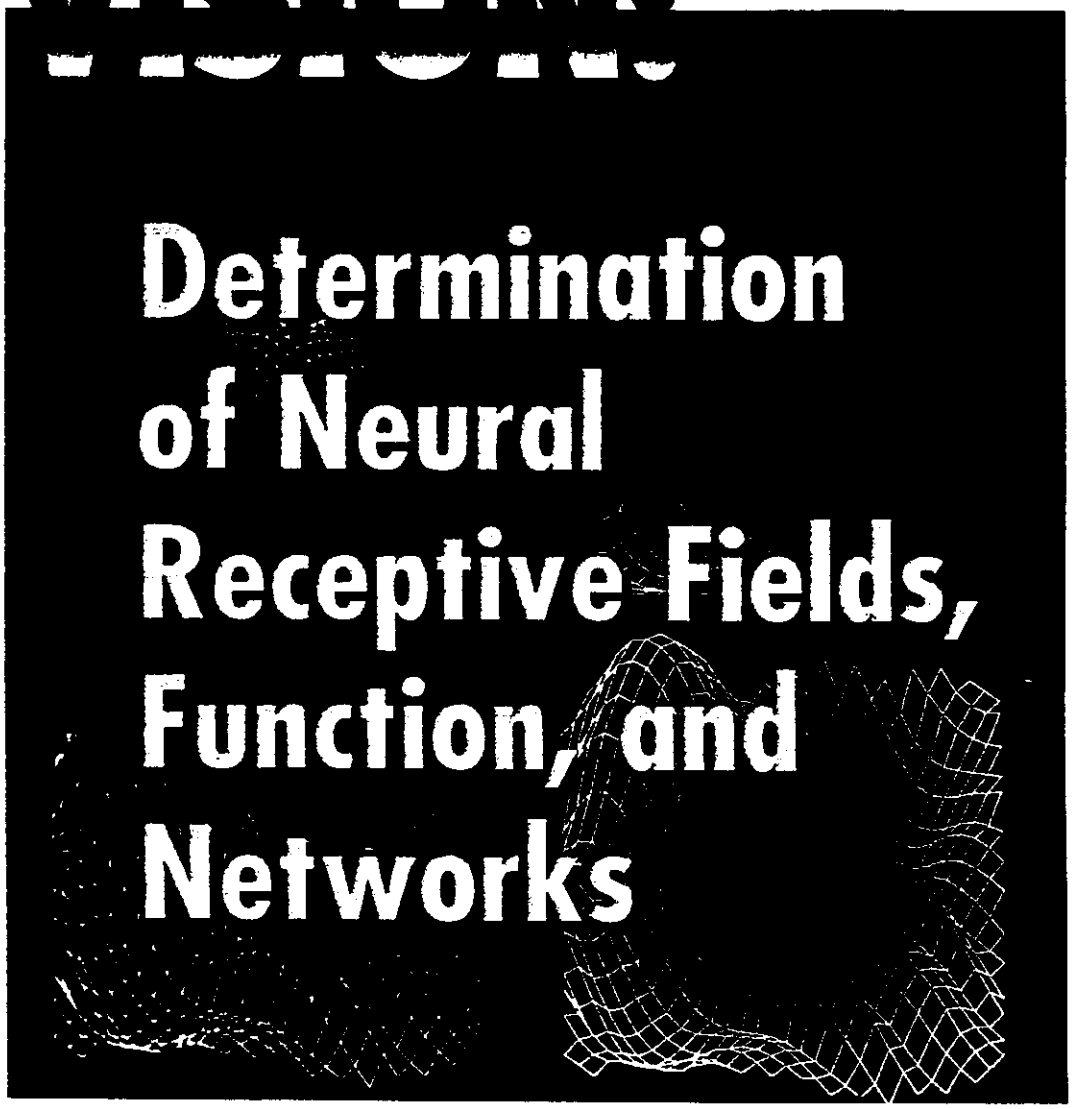
These are preliminary lecture notes, intended only for distribution to participants.

NTER
ABET

NONLINEAR VISION: Determination of

NONLINEAR VISION:

Determination of Neural Receptive Fields, Function, and Networks



Robert B. Pinter
Bahram Nabet

Chapter 11

A NONLINEAR MODEL OF FEATURE DETECTION

David C. Burr and M. Concetta Morrone

TABLE OF CONTENTS

INTRODUCTION

One of the major tasks of the human visual system is to extract quickly and effortlessly the most salient information from an image to form a symbolic representation of the scene. This argument was made most forcefully by Marr (1976, 1982) with the concept of the "primal sketch", an early and grossly simplified abstract representation of the original scene. The primal sketch is not simply a reduced or sampled version of the original image, but a highly nonlinear caricature comprising lines, cylinders, and other basic forms that capture the essential essence of the scene, while ignoring much detail, as well as gradual variation of luminance. Figure 1 depicts a well-known example of how economically an artistic sketch may convey the essential features of an image. A few well-placed strokes are sufficient to represent the image without ambiguity.

Two of the more common visual features are lines and edges, both being rich sources of visual information. The potential of lines to convey information is well illustrated in the sketch of Figure 1, and edges are obviously important in delimiting the boundaries of objects. Indeed, there is a sense in which edges and lines are almost interchangeable, illustrated by the fact that sketch-artists often use lines to symbolize edges (see Figure 1). However, mathematically speaking, the two types of features are quite distinct and, indeed, orthogonal: edges can be described as odd-symmetric functions and lines are even-symmetric functions, orthogonal with respect to each other in L^2.

It is because these two equally important features are orthogonal, defining a two-dimensional feature space, that we propose that a two-dimensional basis set is required for their detection. We further argue that nonlinear combination of the two bases is the most efficient way of detecting lines and edges.

Many models have been proposed for line and edge detection, all similar in one major respect: they convolve the image with simple linear operators of various sizes and search for maxima or zero-crossings in the output. However, this approach runs into several difficulties, particularly when the image features are adjacent and when the features are not simply edges or lines but combinations of both (see Figure 2).

Our approach to line and edge detection differs from that of others in that we have attempted to understand why lines and edges are important to vision, and have been able to provide a robust operational definition of what constitutes a visual feature. Our model, like most others, convolves the image with linear operators of limited bandwidth. But it differs from others in that it requires two sets of linear operators, orthogonal to each other (and related by the Hilbert transform). It also incorporates two major nonlinearities: a second-order (squaring) nonlinearity to combine the output of the matched filters, and a higher-order "nonmaximal suppression" nonlinearity, whereby only the points producing maxima in the combined output are considered as features effective in structuring visual information. This chapter outlines the details of the model and illustrates its application to several interesting images. These demonstrations show how the model successfully predicts human perception, both qualitatively and quantitatively, in conditions where most previous models fail.

II. THE LOCAL ENERGY MODEL OF FEATURE DETECTION

As mentioned above, our model requires two sets of matched operators, one even-symmetric, the other odd-symmetric, hence, one the Hilbert transform of the other. The image is convolved separately by the two sets of operators, and the outputs combined by "Phythagorean sum" (square root of the sum of squares) to give what we term the "local energy" profile. The operation is usually performed separately over several scales and

FIGURE 1. A sketch by Matisse, illustrating how a few well-placed strokes may provide an adequate symbolic representation of an image. Note that the lines have been positioned to correspond to luminance borders, or edges, of the original image.

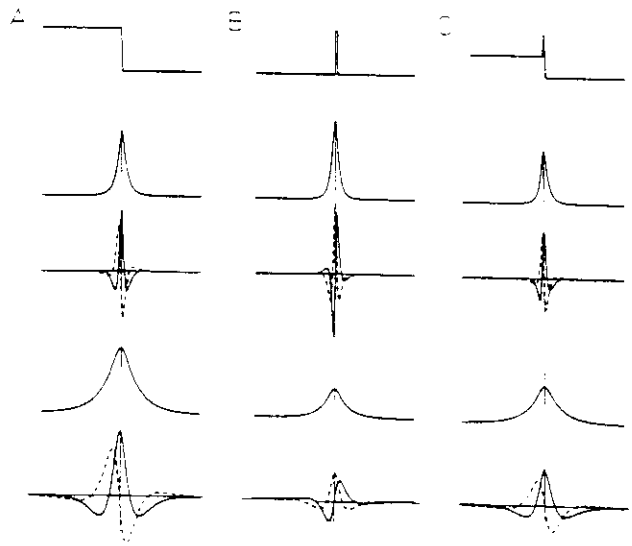


FIGURE 2. Examples of the operation of the model for an edge, a line, and a combination edge-line. Under each profile is shown the output of the even-symmetric (dashed lines) and odd-symmetric (continuous lines) at two spatial scales. Immediately above these traces is the local-energy profile, the square root of the sums of the squared output from the even- and odd-symmetric operators. Whereas the linear operators produce many minima, maxima, and zero-crossings, only some of which fall at the position of the feature, the local-energy functions are always positive, with one clear peak at the position of the feature. The three different features are all marked unambiguously with the same operator. To decide which type of feature caused the local energy peak, it is sufficient to evaluate the response of the linear operators at that point: for the edge, only the odd-symmetric operators respond at that point; for the line, only the even-symmetric operators; for the edge and line, both operators respond, so the feature is recognized veridically.

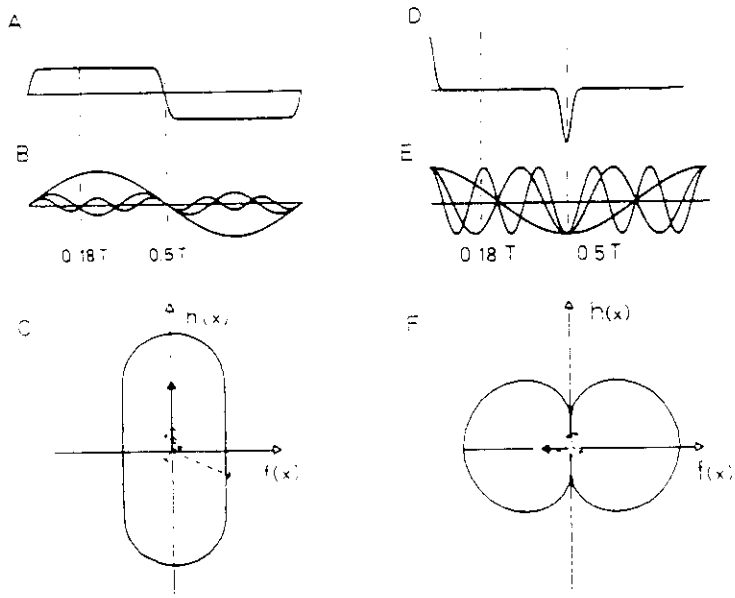


FIGURE 3. Illustration of the relationship between features, phase congruence, and local energy. See text for explanation.

orientations, modeling the action of cortical visual detectors (e.g., Blakemore and Campbell, 1969), but this is, in fact, not an essential feature of the model (see Morrone and Owens, 1987).

Figure 2 illustrates the action of the model in one dimension for three profiles, an edge, a line, and a combination edge and line. For clarity we illustrate the response at only two spatial scales, although at least four are required to model human vision (Wilson and Bergen, 1979). The superimposed curves under the profile show the responses of the even-symmetric operators (dashed curves) and odd-symmetric operators (continuous curves). These linear responses are similar to those predicted by most models of feature detection that use only one class of operator (either even-symmetric: Marr and Hildreth, 1982; Watt and Morgan, 1985; or odd-symmetric: Canny, 1983, 1986) and illustrate some of the difficulties inherent with this approach. Although the odd-symmetric operators peak at the position of the edge, there are also spurious negative peaks at either side of the true edge, and two very large peaks near the bar. The even-symmetric operators face similar problems as line detectors, and the zero-crossings mark several spurious features. The feature that presents most difficulties for models with a single class of detector is the combination edge and line. For this feature, neither the peaks nor the zero-crossings of the even- or the odd-symmetric detectors correspond to the perceived position of the feature. This class of feature, in fact, occurs very frequently in natural scenes, particularly under conditions of oblique lighting (see Horn, 1977; Perona and Malik, 1990; and Figure 4).

Several strategies have been devised to minimize spurious feature marking, usually involving thresholding and comparison across spatial scales (Marr, 1982; Watt and Morgan, 1985; Yuille and Poggio, 1985). However, many of these strategies have been criticized as being computationally expensive, biologically implausible, and often ineffective, particularly with more complicated images with closely adjacent features. And none of these strategies is effective with the combination of line and edge features (like Figure 2C).

The local energy output (Pythagorean sum of even- and odd-symmetric output) is shown directly above the linear outputs. One obvious advantage of this operation is that at each

scale the profile has only one peak, and that peak corresponds in position to the feature. Furthermore, the same function signals both types of features with a positive peak (rather than relying on the difficult process of searching for zero-crossings). After the feature has been detected, it can easily be identified by examining the amplitude of the linear operators at that point: a response of odd-symmetric operators signals an edge, and even-symmetric operators a line. If both classes of operators respond at the position of the feature (such as in Figure 2C), then both a line and an edge are signaled.

What were the theoretical considerations that led us to choose local energy as an indicator of visual features? It has long been known that the Fourier phase spectrum is important for vision, indeed, far more important than the Fourier amplitude spectrum. If the phase spectra of two images are interchanged, leaving the amplitude spectra intact, the appearance of the hybrid pictures is determined almost completely by the phase spectra (Oppenheim and Lim, 1981; Piotrowski and Campbell, 1982). We have recently shown that the importance of the phase to vision lies not in the Fourier phase spectrum per se, but in how phase relationships of the various harmonics cause them to interact to create visual features (Morrone and Burr, 1988; Burr and Morrone, 1990). Specifically, we have demonstrated that visually salient features occur at those points of an image where the Fourier components come into phase with each other. Thus, they can be detected by searching for points of congruence in arrival phase.

Figure 3 illustrates the concept of phase congruence of Fourier harmonics and shows how this congruence creates maxima in the local-energy function. The upper traces (A and D) show two periodic waveforms, a squarewave and a series of delta functions, both slightly blurred with a Gaussian filter (to attenuate the infinite series of harmonics). The sinusoids under each waveform represent the first three components in the Fourier expansion. For both waveforms, these harmonics (and all higher harmonics) come into phase periodically, at twice the frequency of the fundamental; and the point where the harmonics come into phase is where the feature is seen, be it edge or line. For an edge, all cosine harmonics have arrival phases of $\pm \pi/2$ (depending on the polarity of the edge). For a bar, the arrival phases are all 0 or π. If the feature were a combination edge and line (like Figure 2C), the arrival phases would take on an intermediate value at the point of phase congruence. From these and other observations, we have proposed the generalization that visually salient features always occur at the point of maximum phase congruence (Morrone and Burr, 1988).

The relationship between local energy and phase congruence is illustrated in the lower graphs of Figure 3. These curves, parametric in x, were obtained by plotting the input functions $f(x)$ (Figure 3 A and D) against their Hilbert transforms $h(x)$. In this representation, local energy is given by the distance of the curves from the origin (the norm).*

The arrows inside the graphs depict the vectors associated in this space with the first four individual harmonics of the periodic waveforms, calculated at two positions: $x = 0.5T$ the point where the feature is seen, and $x = 0.18T$, an arbitrary point on the plateau. The norm of the vectors for each harmonic does not vary with position (x) on the waveform [as $\sin^2(2\pi x/T) + \cos^2(2\pi x/T) = 1$], but the argument is proportional to the product of position and spatial frequency ($2\pi x/T$). At $x = 0.5T$ (the position of the edge or line), the arrival phases of all the harmonics are the same ($\pi/2$ in Figure 3C and π in Figure 3F), so the vectors are all aligned. For $x = 0.18T$, however, the arrival phases of the harmonics are very different, so the vectors (indicated by dashed arrows) will all point in different

* The functions f(x) can be considered to be the ouptut of an even-symmetric operator of broad bandwidth, and its Hilbert transform h(x) the output of an odd-symmetric operator of identical bandwidth. In practice the model does not evaluate local energy for the whole pattern, but separately at different scales. However, the concepts illustrated here for the physical image can readily be extended to the output of our linear operators by considering band-pass filtered versions of f(x) and h(x).

directions. Each point of the parametric curves is the resultant of all the vectors associated with the harmonics. Clearly, the norm of the resultant vector will be greatest when the individual vectors are most closely aligned; that is, when the arrival phases of the component harmonics are most similar. Hence, peaks in local energy will mark points of maximal arrival-phase congruence. Although this example uses periodic patterns, the same argument can readily be extended to aperiodic patterns, by considering a patch-wise Fourier analysis.

Operators similar to local energy have been used in several applications, including image enhancement (Granlund, 1978; Wilson et al., 1983), motion detection (Adelson and Bergen, 1985, Heeger, 1987) and in modeling hyperacuity (Klein and Levi, 1985). However, the motivation for using this class of operation has always been to eliminate phase information, rather than to capitalize on it. These seemingly contradictory assertions can be reconciled by the understanding that although local energy is, in fact, insensitive to absolute phase, it is highly sensitive to relative phase organization. For example, the local energy of a sinewave is identical to that of cosinewave; and any L_2 function will have the same local-energy profile as its Hilbert transform, or any other phase-shifted version of it. However, local energy does depend on the relationships between phases of the various Fourier harmonics, as the above example clearly shows. It is for both of these reasons that we find the function so useful. It is sensitive to phase relationships and, therefore, detects points of phase congruence; but it responds equally well, irrespective of the value of the phase at those points of phase congruence, so the same operator can detect both lines and edges. Distinguishing between lines and edges is a simple process, readily established by evaluating the phase (or relative contribution of the even- and odd-symmetric operators) at the maxima of local energy.

III. THE MODEL IN TWO DIMENSIONS

The model has recently been extended to two dimensions (Burr and Morrone, 1990; Morrone and Burr, 1992). This step is not as routine as may be imagined, as the operators must be Hilbert transforms of each other, and the Hilbert transform is not defined in two dimensions. Therefore the operators must be essentially one dimensional and oriented in space. It is interesting that the oriented operators that we were forced to use in order to obtain a Hilbert transform in fact resemble quite closely the receptive field properties of visual neurones. In both cat and monkey, the most striking property of cortical neurons (as distinct from retinal and thalamic neurons) is that they respond very selectively to stimuli of given orientations (Hubel and Wiesel, 1962, 1977). We use four sets of oriented operators, oriented at 0, 90, and $\pm 45°$. Along one axis the operators are modulated with the even- or odd-symmetric profile depicted in Figure 2, and along the other with a simple Gaussian of about the same size.

Another difficulty for the two-dimensional model is in detecting maxima in local energy. If only the absolute maxima of the plane are considered, only a few isolated points would be marked. On the other hand, if maxima are marked line by line (at each orientation), too many features would be marked, producing a tendency for "overshoot", making lines appear longer than they are and corners terminated incorrectly. Our strategy is to establish first the direction of maximal change of energy (within a small window) and search for maxima only along this direction. Again this process finds a neurophysiological analogue in human and animal vision called "cross-orientation inhibition": mutual inhibition between neurons of differing orientation preference (Morrone, Burr, and Maffei, 1982; Morrone and Burr, 1986; Burr and Morrone, 1987).

Finally, there is the difficulty of how information at various scales and various orientations is combined to give an overall feature map of the scene. For this we find little guidance either from physiological or from psychophysical research. Although there is

compelling evidence that information is decomposed by filters selective for both orientation and spatial frequency, there is as yet no information of how this information may be resynthesized to yield a unified percept. Indeed, for the human visual system, there may be no need for resynthesis of information at all. However, to gain some insight into how our model performs, we must be able to combine our separate representations to form a single feature map. To achieve this, each map is given a scale-dependent "uncertainty weighting" by blurring the feature maps with Gaussians of space constantly proportional to the scale of the operators, and then summed together. The effect of the uncertainty weighting is to privilege-attach positional information from the higher scales, which we may expect to be more precise.

An example of the operation of the model in two dimensions is shown in Figure 4. The image was convolved separately by even- and odd-symmetric operators of four spatial scales and four orientations. (In practice this operation is achieved using the pyramid technique of Burt and Adelson [1983] for efficiency of computation.) At each scale and at each orientation, the even- and odd-symmetric outputs were combined by Pythagorean sum to produce the local energy profile, from which the local maxima were calculated (along the direction of maximal energy gradient) to yield the independent feature maps. The feature maps were blurred in proportion to scale size (again achieved by "expanding" the pyramid) and summed to give the image of Figure 4B.

The first obvious point to make is that although the feature map may lack some of the artistic merit expressed in Figure 1, it is in fact a quite good sketch of the original image. Furthermore, evey relevant feature, line, edge, and specularity, has been marked, with no glaring false positives. Given that there has been no thresholding or postprocessing, this result is more than acceptable.

The histograms of Figure 4C give an idea of the proportions of various features types by plotting the distributions of phase at the maxima of local energy, separately for three different scales. All three histograms have clear peaks at $\pm\pi/2$, corresponding to positive- and negative-going edges, suggesting that these features are the most common. However, it is important to note that while these phases do predominate, there remains a large proportion of marked features of intermediate phases (like the feature of Figure 2C), features that often result from oblique lighting and shadowing effects. For our model, this class of features presents no difficulty and is veridically marked as combination edge bars. But for any model based on simple linear convolution, these types of features are difficult to localize veridically (see Figure 2C).

IV. STRUCTURING THE IMAGE

One of the clearest illustrations that vision is not strictly linear is provided by Harmon and Julesz's (1973) demonstration of coarse-quantization, illustrated in Figure 5A. The original image of Figure 4A has been "quantized" by setting all the pixels within each square to the mean value of those of the original image. This sampling technique preserves nearly all low-frequency image information (below the Nyquist sampling rate), while introducing the spurious high spatial frequency components that shape the blocks. What is interesting is that although blocking preserves sufficient low frequency information to allow face recognition (readily verified by blurring or distancing the image), by no effort of will can one extract this information from the unfiltered blocked image.

Whatever mechanisms underlie the phenomenon of blocking, they demonstrate a clear nonlinearity in the human visual system. The low-frequency information about the face cannot be extracted when viewed together with the high spurious spatial frequencies, clearly violating the additivity principle of linear systems. The original explanation for the effect was that the high spurious frequencies introduced by blocking mask the lower spatial frequencies that contain the image information, rendering them effectively invisible (Harmon

FIGURE 4. An example of results of the model in two dimensions. B is the output of the model applied to A (see text for details). The histograms of C show the distribution of arrival phases at the marked features, for high (upper curve), medium (middle curve), and low (lower curve) scales. Although there is a clear tendency for the phases to group at $+\pi/2$ (corresponding to edges), all phases are represented in the histograms.

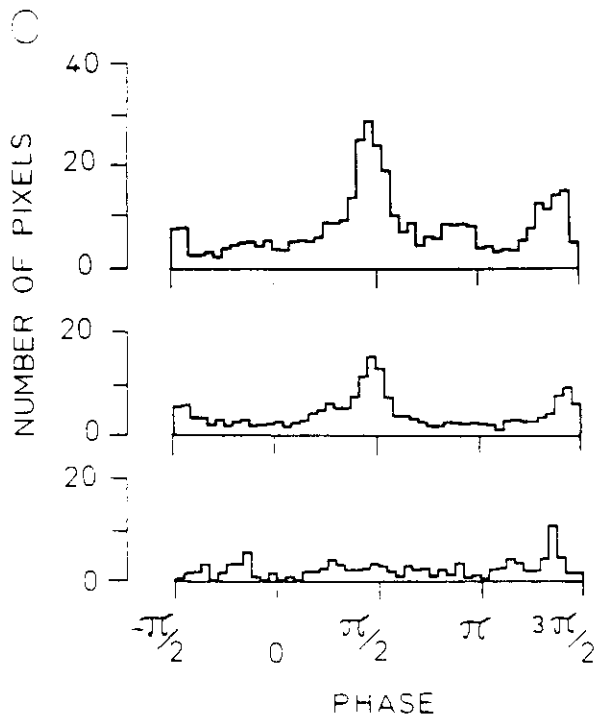


FIGURE 4C.

and Julesz, 1973). However, although nonlinear compression of this sort certainly occurs (e.g., Stromeyer and Julesz, 1972; Legge and Foley, 1980), calculations of the strength of masking (for example, from the data of Legge and Foley, 1980 or Anderson and Burr, 1985) show that the spurious spatial frequencies could not have sufficient power to suppress the low-frequency signals below detection threshold. Perhaps the best demonstration that the low spatial frequencies are not merely "masked" is that addition of further high-frequency energy (in the form of high-pass noise) causes the blocked image to be recognizable (Morrone et al., 1983).

Our explanation for the blocking phenomenon is that the high spatial frequencies do not remove the low spatial frequency information from vision, but structure the way that it is perceived. According to our model, only maxima in local energy are perceived as features, and in this image it is the high spatial frequencies that dictate where the maxima of local energy fall (partly because of the scale-dependent uncertainty weighting). Figure 4C shows the output of the model, produced by the procedure described above. The marked features fall in a grid-like pattern along the borders of the blocks; and according to our model, these features provide an abstract description of the image, causing it to be perceived as blocks. The information contained in the low spatial frequencies is not masked or suppressed, but becomes "compartmentalized" by the structure provided by the maxima of local energy. After the high spatial frequencies of image have been removed by digital blurring (Figure 5B), the low spatial frequencies contribute more to the position of the maxima of local energy, creating a feature map more consistent with the features of the face (Figure 5D), allowing a blurred face to be perceived.

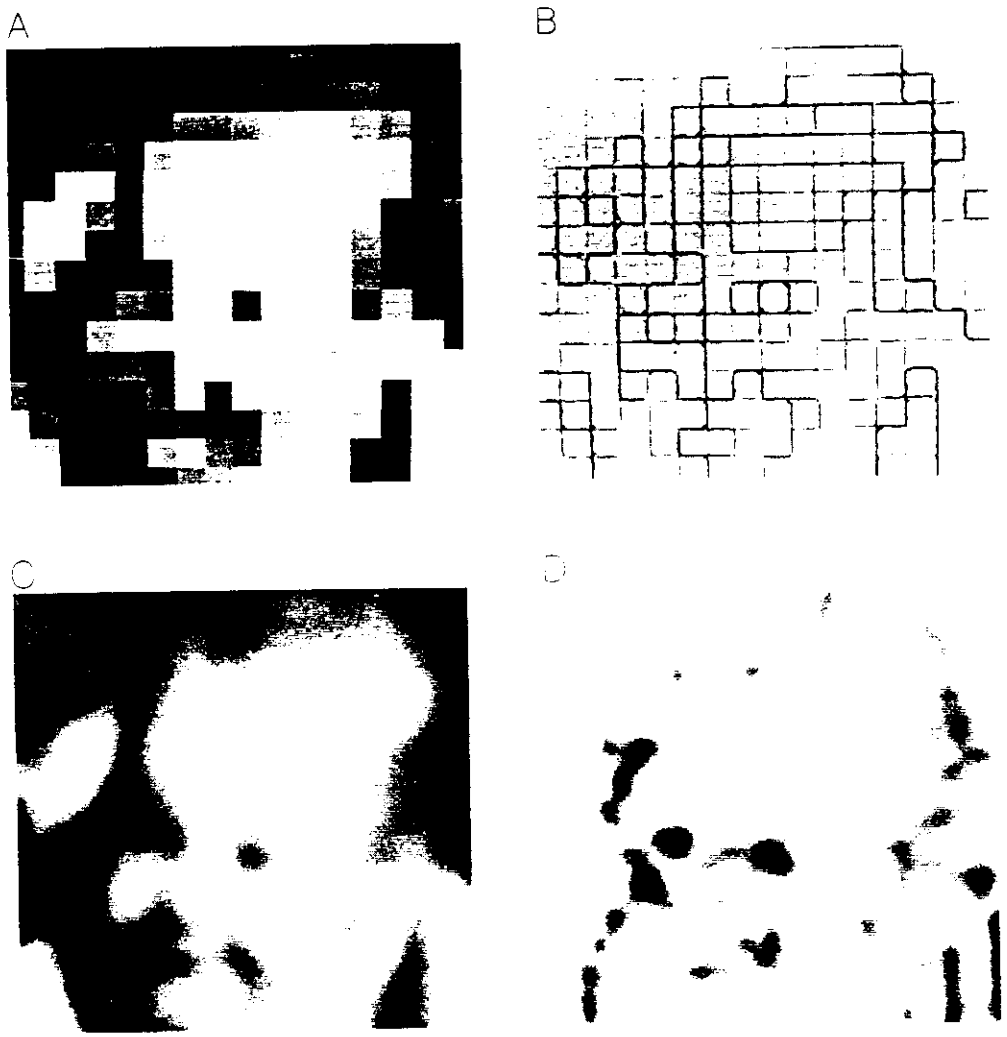


FIGURE 5. Illustration of how the local energy model predicts the "coarse quantization" phenomenon, first described by Harmon and Julesz (1973). The image of 4A has been quantized by setting all pixels within each block to the average level (A). This process preserves most low spatial frequency information (below the *Nyquist* rate, half the sampling frequency), readily verified by viewing the picture from a distance, or blurring the image (C). B and D show the output of the model to the blocked and blurred images (see text for further explanation).

V. BRIGHTNESS ILLUSIONS

Visual illusions have long been invoked to demonstrate that the visual system does not simply transpose the external scene into a veridical internal copy, but encodes important information into a symbolic representation. In doing so it sometimes errs, to produce a so-called visual illusion.

One of the best known and most powerful illusions is the phenomenon of Mach bands, the paradoxical light and dark bands seen where a luminance ramp meets a plateau (Mach, 1865, 1906; Ratliff, 1965). Figure 6 provides two illustrations of the phenomenon, in one and in two dimensions. The figure on the left is a trapezoidal waveform, but is not perceived as such. There are conspicuous spikes of brightness, where there are none in the luminance

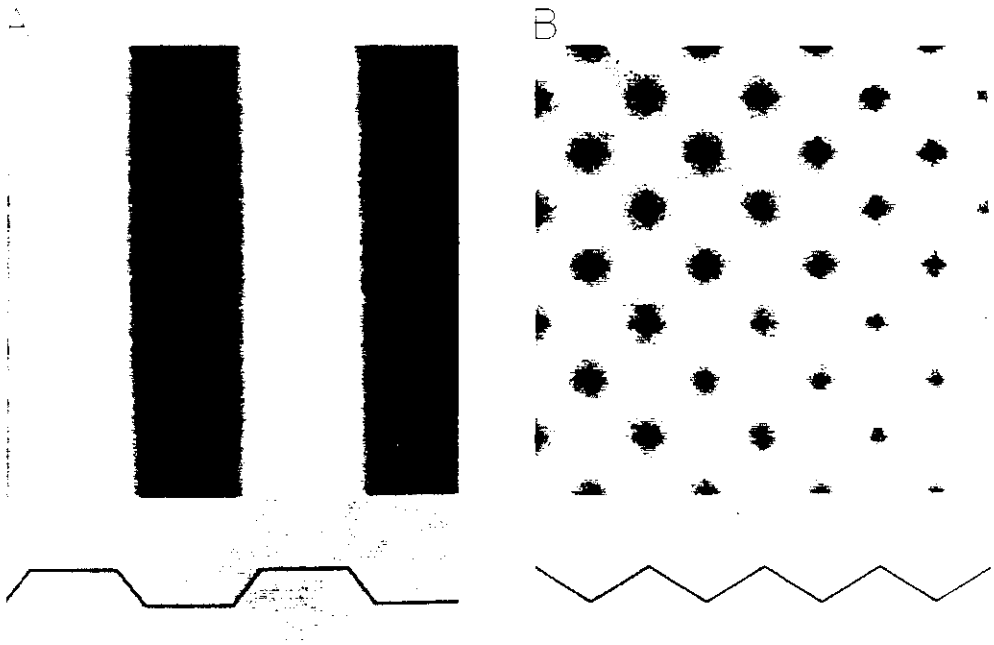


FIGURE 6. Two examples of the well-known brightness illusion of Mach bands. On the left the luminance profile is trapezoidal, but we perceive sharp bright and dark lines where the ramps meet plateaus. The pattern on the right was created by multiplying a vertical with a horizontal triangle wave to form a pyramid-like luminance distribution; yet we perceive sharp bright and dark crosses, not present in the luminance distribution.

profile. Note also that at the point where the bands are seen, there is a change in brightness (partly disguised by the Mach bands), so that the luminance ramp between the bands appears of uniform brightness, with no particular feature at the point where the luminance profile crosses zero. Figure 6B shows the same effect in two dimensions (Morrone et al., 1986). The luminance distribution is, in fact, pyramidal, the product of a horizontal and vertical triangular waveform; yet we perceive the pattern as bright and dark star-like structures.

The explanation for the Mach bands that appears in most textbooks (e.g., Cornsweet, 1970; Ratliff, 1965) was first advanced by Mach himself (Mach, 1906): that lateral inhibition in the retina effectively differentiates the image, producing overshoot and undershoot at the points where the bands are seen. However, a major difficulty with global differentiation (or any other linear high-pass filter operation) is that it implies that as the ramp is made steeper that bands should become stronger and stronger, to be maximal for a steep square edge, which would severely distort our everyday perception. However, the bands, in fact, become weaker as the ramp is steepened and do not occur at all if the ramp extends over less than 30'' (Ross et al., 1981; see also Fiorentini, 1973; Ratliff, 1984).

Our explanation for Mach bands does not rely solely on differentiation or other linear operations, but on the fact that the energy model marks line-like features at the point where the bands are seen (Ross et al., 1989). Figure 7 shows the local energy profile at two scales. At both scales the energy peaks at the points where the bands are seen, so the summed feature map will predict features to be there and there alone. At the higher scale there is a strong response from even-symmetric operators at these peaks, signaling the presence of sharp lines. At lower scales the odd-symmetric operators also respond at the energy peaks, and this response signals that a brightness change should accompany the Mach bands. As

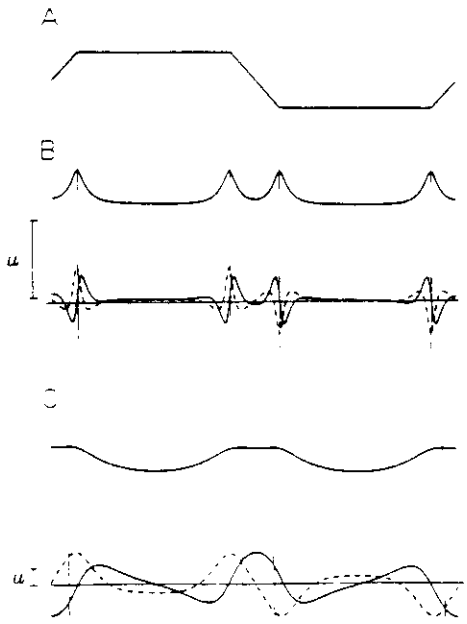


FIGURE 7. Illustration of the local-energy model applied to a trapezoidal waveform. Under the waveform are the local energy profiles of a high (B) and a low scale (C), together with the even-symmetric (dashed lines) and odd-symmetric (continuous lines) output. The energy at the higher scale peaks at the postion where the Mach bands are seen; and at that position there is a strong response from the even-symmetric operators, indicating that the feature should be a line. At the lower scale, the position of the feature is not positioned exactly on the kneepoint of the luminance distribution, but (as discussed elsewhere), the lower scales contribute little to localization of the features. They do, however, contribute to the brightness map; and because the odd-symmetric operators (as well as the even-symmetric operators) respond at the marked feature, they signal a brightness change

in the previous example, the change in brightness occurs at the feature defined by the local energy peaks, and the high spatial scales dominate in determining the position of these peaks.

Local energy predicts not only the occurrence of Mach bands and where they should occur, but also the perceived strength of the bands. We have measured the strength of the Mach bands, by adjusting the contrast of a trapezoidal waveform until the bands were just visible. The inverse of contrast at threshold gives an estimate of contrast sensitivity, a unitless index greater than 1. It turned out that the contrast thresholds were very easy to set, and three observers agreed closely on the threshold settings. Contrast sensitivity for Mach bands were measured over a wide range of spatial frequencies and for a variety of stimuli filtered in the spatial frequency domain (Morrone et al., 1986; Ross et al., 1989).

Figure 8 shows an example of the results. Contrast thresholds for seeing Mach bands on blurred stimuli were measured as a function of the low-pass filter cutoff frequency (triangular symbols). As the pattern is blurred, and bands become progressively weaker, so more contrast is required to see them. For comparison we also measured contrast thresholds for detecting a high-pass trapezoid stimulus (circular symbols), deliberately chosen to have similar thresholds to those of the Mach bands.

The lines passing through the symbols of Figure 8 are not simply best fits of the data, but quantitative predictions from the local-energy model. To predict the detection thresholds for the high-pass stimulus, we first calculated the local energy at each scale. The energy was then combined by the standard procedure of probability summation across scales and

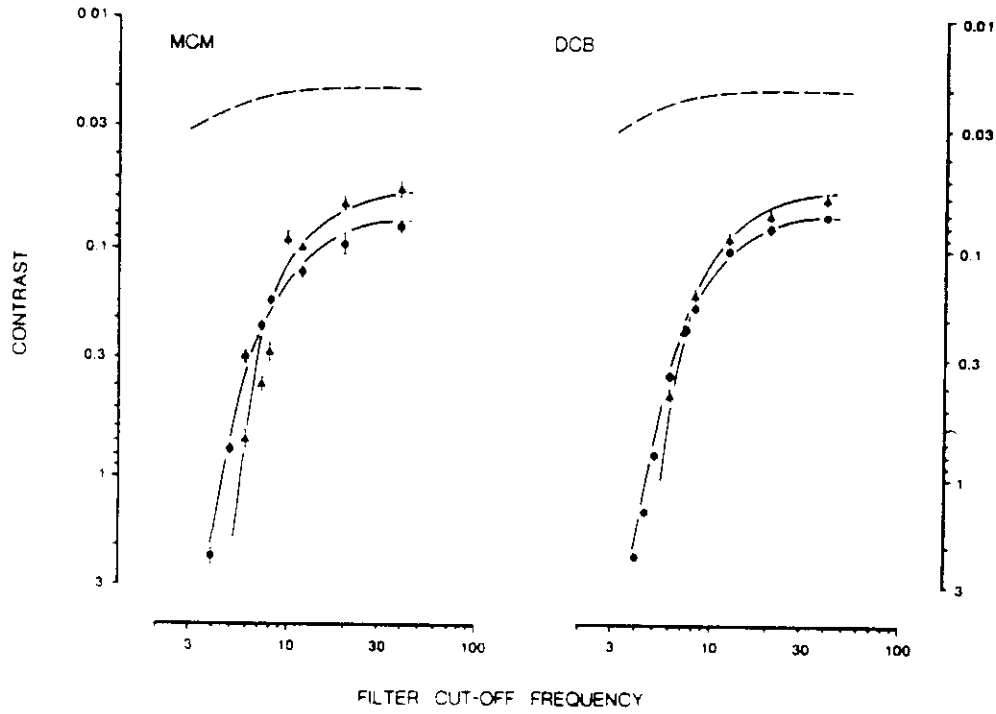


FIGURE 8. Contrast thresholds for seeing Mach bands (triangles) on a low-pass filtered trapezoidal waveform (circle), as a function of the filter cutoff frequency. The circles indicate detection thresholds of a high-pass trapezoid, measured for comparison. The curves passing through the symbols are predictions from the local energy model, while the dashed curves above are predictions derived from lateral inhibition (see Ross et al., 1989 for details).

across space (a form of nonlinear summation that takes account of the fact that increasing the number of visual units that respond to a stimulus increases the probability that at least one will respond and detect it: see Sachs et al., 1971; Ross et al., 1989). The results of these predictions (as a function of filter cutoff frequency) are represented by the continuous curves passing through the circles. To predict the strength of the Mach bands, we considered only energy that contributed to the peaks at the position where the Mach bands are seen. Again, this energy was summed probabilistically across space and across scale to yield the predictions represented by the other set of solid lines in Figure 8.

It shoud be stressed that the predictive curves of Figure 8 were derived with only one degree of freedom: absolute sensitivity determining the height of the curves. Once this was fixed for the detection results, the prediction curve of the Mach band thresholds was entirely determined. Given that there are no other variable parameters (or "fudge factors") in the model, the predictions are surprisingly good. Not only do the curves follow the general shape of the data, but they even predict subtle differences in the data, including the crossover around 7 c/deg.

The dashed curves above the figure are predictions of the strength of Mach bands for a model based solely on differentiation or high-pass filtering (assuming even-symmetric filters). It is clear that such a model does not come close to predicting quantitatively the actual data.

Many illusions illustrate how brightness* does not depend entirely on the physical luminance of the stimuli. One of the clearest demonstrations of this is the Craik-Cornsweet-

* A technical term meaning the perceived or apparent luminance of a surface, often misused to refer to physical properties of images

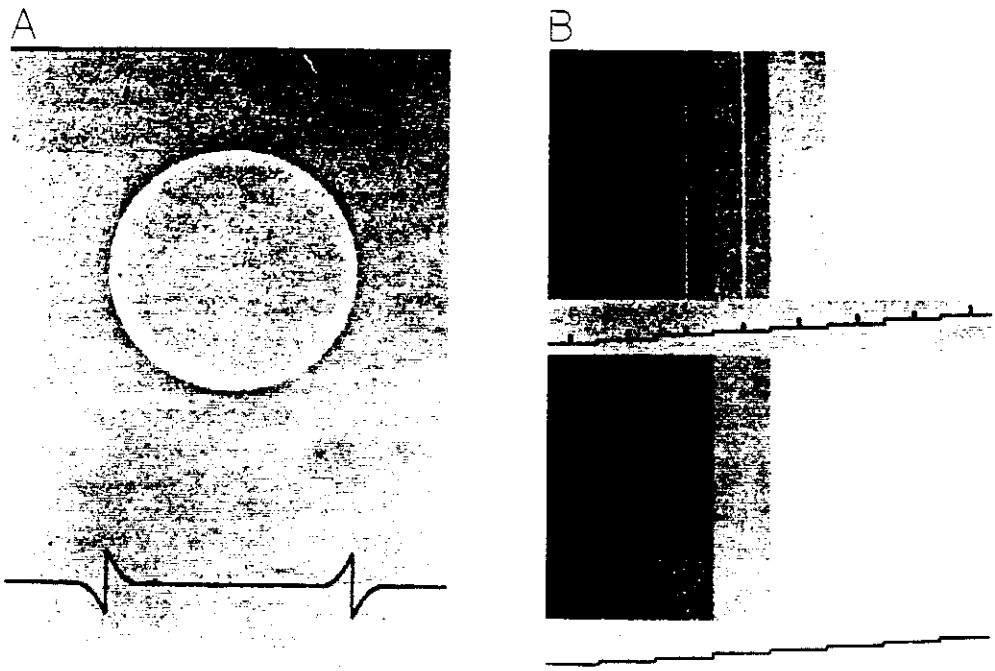


FIGURE 9. (A) An example fo the Craik-O'Brien-Cornsweet illusion. The pattern has identical luminance everywhere except at the border, yet the circle seems to be lighter than the background. (B) The lower figure shows the classic Chevreul illusion: the steps are all of uniform luminance, but do not appear so. When thin lines are added to edge step, they appear to divide each step into two regions of quite different brightness.

O'Brien illusion (Craik, 1966; O'Brien, 1958; Cornsweet, 1970), illustrated in Figure 9A. The luminance is, in fact, identical everywhere except at the border of the perceived circle, yet the circle seems brighter than the background. Original explanations for the phenomenon relied on high-pass filtering of the visual system (Cornsweet, 1970; Campbell et al., 1978), but as for the explanations for Mach bands, the filtering cannot predict quantitatively the results. The illusion holds even at very high spatial frequencies, well outside the range where vision may be expected to perform high-pass filtering (Burr, 1987).

The local-energy model readily accounts for the illusion. Local energy peaks at the borders of the inner circle, so features are seen there. As the profile is odd-symmetric, only odd-symmetric operators respond at the marked feature, so the feature is perceived as an edge, and perceiving an edge means perceiving a brightness step. The step in brightness caused by the edge extends to the whole region, as there are no other features to provide contradictory information.

Other brightness illusions are illustrated in Figure 9B. A staircase luminance profile is not seen veridically, but takes on the "scallopy" appearance of the well-known Chevreul illusion (Chevreul, 1890). For this illusion we do not as yet have a satisfactory explanation, but expect that it reflects a nontransitivity in visual computations. Each edge in the series has the same polarity and should, therefore, signal successive brightness increments. The second region should appear brighter than the first, the third brighter than the second, etc. If complete transitivity were observed, the difference between the brightness of the first and third region should be twice that of the difference between adjacent panels. However, there is good evidence that transitivity is not preserved totally, even for two successive edges of the same polarity (Shapley and Reid, 1985). Failure in preserving transitivity would create

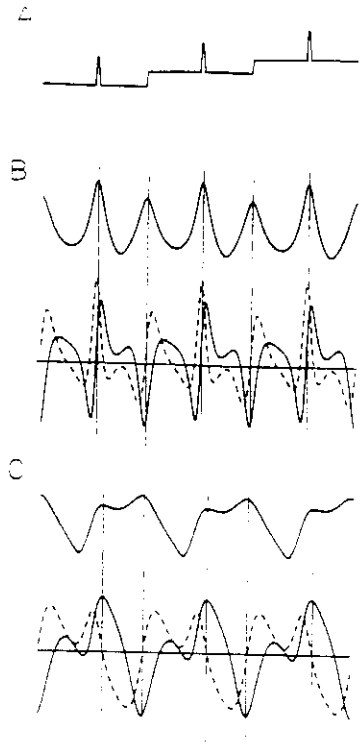


FIGURE 10. Illustration of how local energy predicts the brightness illusion of Figure 9. A shows the input waveform, and B and C the local-energy profiles, together with the even- (dashed line) and odd- (continuous lines) symmetric outputs at two scales. At the high scale (B), the local-energy peaks correspond in position to the lines and edges, acurately predicting the position of these features. At this scale it is mainly the even-symmetric operators that respond at the position of the lines, correctly predicting their appearance. At the lower scale, however, odd-symmetric detectors respond at the peak of local energy caused by the lines, predicting the observed brightness change.

an incompatibility between the local brightness steps and the global brightness of the whole pattern, and this incompatibility may produce the scallopy pattern we observe.

Unfortunately, the above explanation is not yet in a form where it can be tested quantitatively. However, a new illusion, more amenable to quantitative measurement can be created by adding a thin line to each step (Morrone et al., 1991). The series of lines changes the appearance of the pattern completely, causing it to appear "square-wave-like" with the regions bounded by lines and edges seen with uniform brighness, with a large brightness step at the position of each line.

Again, the local-energy model can explain readily this illusion. The profiles of Figure 10 show the response of the local energy and of the matched filters at two spatial scales. The energy profile peaks at the positions of the lines and edges, predicting that all features should be seen in those positions. The even-symmetric response at the high spatial scales predicts the presence of lines at those positions. But at the lower scales the odd-symmetric operators also respond at the peaks of local energy, predicting an edge, with an accompanying brightness change. We see both sharp lines and a change in brightness, in agreement with the predictions of the model.

As we saw in the previous example, the real test of a model is how well it predicts quantitatively human performance. We have measured the apparent contrast of the illusion under various conditions. Figure 9B shows how apparent contrast varies as the pattern is

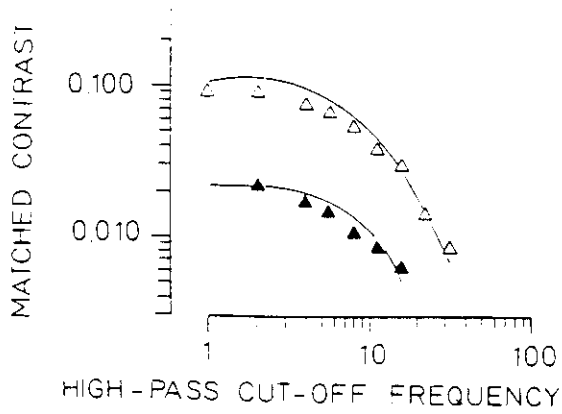


FIGURE 11. Apparent contrast of the illusory edge of Figure 9, measured at high (open triangles) and low (filled triangels) contrast. The lines passing through the symbols are predictions from the local-energy model, obtained from the strength of response of the odd-symmetric operators (see Figure 10).

high-pass filtered. Again, the solid curves are predictions from the model, based on the strength of the odd-symmetric response at the feature. As the low frequencies of the pattern are reduced, the odd-symmetric response becomes progressively diminished, until there should be no illusion at all. And, indeed, the experimental data follow closely the predictions of the model.

VI. DISCUSSION

To conclude, we have presented a model for early human vision whereby the retinal image is not simply transmitted through a system of linear filters, but encoded by nonlinear mechanisms to form an abstract, symbolic representation. In general, this representation reflects the physical reality. However, there exist several examples of visual illusions where humans perceive an image as different from that defined by its physical luminance distribution. In all the examples examined so far, our model performs like human vision in "seeing" the illusions and predicts accurately the strength of the illusions under various conditions.

In general, we were guided by the known physiology of mammalian vision in choosing parameters for the model. Our filters all have a limited amplitude spectrum of about 1.5 octaves, agreeing with most physiological and psychophysical data (e.g., Maffei and Fiorentini, 1973; Movshon et al, 1978a,b; Blakemore and Campbell, 1969; Legge and Foley, 1980; Anderson and Burr, 1991). But unlike many models of visual perception (Marr, 1982; Robson, 1980), the limited bandwidth of the operators (forming so-called spatial frequency channels) does not perform a major role in our model. Indeed, we have demonstrated reasonable success by applying the operation with broad-band matched filters of 3 octaves bandwidth (Morrone and Owens, 1987).

However, once the decision has been made for independent analysis of the image at various scales, the problem arises as to how the results of the independent analyses should be combined, and for this problem we find little guidance in the literature. Our strategies of relying more on the higher than lower scales for feature localization (but not in determining brightness) seems intuitively logical and does receive some support from the literature: phase judgments vary little with spatial frequency, implying an improvement in absolute positional judgments at higher scales (Burr, 1980; Tyler and Gorea, 1986). However, we would prefer that this important strategy be based on firmer data and have commenced some studies to this end.

A different strategy for synthesis has been suggested by Perona and Malik (1990). Rather than evaluate local maxima separately over all scales, Perona and Malik search all energy maps (of different scale and orientations) concurrently, and choose the absolute maxima, thereby producing a single feature map. The strategy has the advantage of not requiring reintegration of separate feature maps and is not at all biologically implausible. Hopefully, it will be possible to devise a quantitative test to distinguish between the two strategies.

Another difference between Perona and Malik's approach and ours is that they use a different strategy to evaluate local energy. Rather than nonlinear combination of two orthogonal filters, they convolve the image with many matched filters with differing phase response and choose those that give the greater response. This process is formally equivalent to our approach (consider the parametric plots of Figure 3), but is, of course, computationally much more expensive. Moreover, although the physiological evidence remains somewhat ambiguous (Field and Tolhurst, 1986), there is now clear psychophysical evidence that in human vision, detectors tend to have phase responses of 0, $\pi/2$ and $\pm\pi$, implying that receptive fields are eight even-symmetric or odd-symmetric (Burr et al., 1989). Thus, both the experimental evidence and considerations of computational efficiency imply that the nonlinear technique of taking the Pythagorean sum of orthogonal operators is reasonable.

Most of the operations that we suggest for our model are readily implementable by physiologial "hardware". There exist two major classes of neurons in the mammalian primary visual cortex, "simple cells" and "complex cells" (Hubel and Wiesel, 1962, 1977). Both simple and complex cells respond selectively to stimulus orientation, and the preferred orientation of the stimulus changes progressively with position in the cortex. The major difference between the two cell classes is that simple cells are quasilinear (except for a half-wave rectification resulting from the fact that their firing rate cannot fall below zero), while complex cells exhibit a clear second-order (squaring) nonlinearity (Movshon et al., 1978a,b; Maffei et al., 1979; Spitzer and Hochstein, 1985a,b). Furthermore, the linear simple-cells tend to be grouped so that adjacent cells have similar orientation and spatial frequency tuning, but differ in phase response by $\pi/2$ (Pollen and Ronner, 1981). The simple cells are ideally suited to act like the matched filters of our model. Complex cells, on the other hand, with their second-order nonlinearity, are ideal candidates to extract local energy, either from input supplied by the simple cells or by similar operations performed within their own subunits. If this suggestion were correct, then the complex cells would be primarily responsible for location of visual features, while simple cells were responsible for their identification.

An important nonlinearity in our model is the nonmaximum suppression, where only local peaks in the energy functions are considered to provide brightness information. Although no such mechanism has yet been observed neurophysiologically, it is by no means unreasonable. One of the primary mechanisms of elaboration of sensory information is mutual inhibition between neurons (Eccles, 1969). Lateral inhibition helps shape the receptive-field structure of visual neurons (Hartline, 1949), and inhibition between cone types enhances color sensitivity and helps maintain color constancy (Wiesel and Hubel, 1966). More recently, "cross-orientation inhibition" has been demonstrated between orientation-selective cortical neurons, a mechanism which should tend to enhance the response of the strongest firing neurons (Morrone et al., 1982; Morrone and Burr, 1986; Burr and Morrone, 1987; see also Chapter 12 of this volume). Similar sorts of mechanisms could well lead to suppression of weaker responses, ensuring that only local maxima of energy contribute toward the feature map. However, whether this operation is actually implemented neurophysiologically, and if so, at what level, remain open questions.

REFERENCES

Adelson, E. H. and Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion, *J. Opt. Soc. Am.*, **A2**, 284–299.

Anderson, S. J. and Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system, *Vision Res.*, **25**, 1147–1154.

Anderson, S. J. and Burr, D. C., (1991). The Two-dimensional spatial and spatial frequency properties of motion sensitive mechanisms in human vision, *J. Opt. Soc. Am.*, **A8**, 1340–1351.

Blakemore, C. and Campbell, F. W. (1969). On the existence of neurones in the visual system selectively sensitive to the orientation and size of retinal images, *J. Physiol. (London)*, **225**, 437–455.

Burr, D. C. (1980). Sensitivity to spatial phase, *Vision Res.*, **20**, 391–396.

Burr, D. C. (1987). Implications of the Craik-O'Brien illusion for brightness perception, *Vision Res.*, **27**, 1903–1913.

Burr, D. C. and Morrone, M. C. (1987). Inhibitory interactions in the human visual system revealed in pattern visual evoked potentials, *J. Physiol. (London)*, **389**, 1–21.

Burr, D. C. and Morrone, M. C. (1990). Edge detection in biological and artificial visual systems, in *Vision: Coding and Efficiency*, Blakemore, C., Ed., Cambridge, University Press, London.

Burr, D. C., Morrone, M. C., and Spinelli, D. (1989). Evidence for edge and bar detectors in human vision, *Vision Res.*, **29**, 419–431.

Burt, P. J. and Adelson, E. H. (1983). The laplacian pyramid as a compact image code, *IEEE Trans. COM*, **31**, 532–540.

Campbell, F. W., Howell, E. R., and Johnstone, J. R. (1978). A comparison of threshold and suprathreshold appearance of gratings with components in the low and high spatial frequency range, *J. Physiol. (London)*, **284**, 193–201.

Canny, J. F. (1983). Finding edges and lines in images, *M I T AI Lab. Tech. Rep.*, **720**.

Canny, J. (1986). A computational approach to edge detection, *IEEE Trans. PAMI*, **8**, 679–698.

Chevreul, M. E. (1890) *The Principles of Harmony and Contrast of Colours*, (Translated by C. Martel), Bell, London.

Cornsweet, T. N. (1970). *Visual Perception*, Academic Press, New York.

Craik, K. (1966). *The Nature of Psychology*, Cambridge University Press, London.

Eccles, J. (1969). *The Inhibitory Pathways of the Central Nervous System*, Liverpool University Press, Liverpool.

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. R. Soc. London*, **B228**, 379–399.

Fiorentini, A. (1972). Mach band phenomena, in *Handbook of Sensory Physiology, Vol VII/4*, Jameson, D. and Huvich, L. M., Eds., Springer-Verlag, Berlin.

Granlund, G. (1978). In search of a general picture operator, *Comput. Graphics Image Process.*, **8**, 155–173.

Harmon, L. D. and Julesz, B. (1973). Masking in visual recognition: effect of two-dimensional filtered noise, *Science*, **180**, 1194–1197.

Hartline, H. K. (1949). Inhibition of activity of visual receptors by illuminating nearby retinal elements in the Limulus eye, *Fed. Proc.*, **8**, 69.

Heeger, D. J. (1987). Model for the extraction of image flow, *J. Opt. Soc. Am.*, **4A**, 1455–1471.

Horn, B. K. P. (1977). Image intensity understanding, *Artif. Intell.*, **8**, 201–231.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol. (London)*, **160**, 106–154.

Hubel, D. H. and Wiesel, T. N. (1977). Architecture of macaque monkey visual cortex, *Proc. R. Soc. London*, **B198**, 1–59.

Klein, S. A. and Levi, D. M. (1985). Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation, *J. Opt. Soc. Am.*, **A2**, 1170–1190.

Legge, G. E. and Foley, J. M. (1980). Contrast masking in human vision, *J. Opt. Soc. Am.*, **70**, 1458–1471.

Mach, E. (1865). Uber die Wirkung der raumlichen Vertheilung des Lichreizes auf di Neztzhaut, *I.S.-B. Akad. Wiss. Wien Math.*, **54**, 303–322.

Mach, E. (1906). Uber den Einfluss Raumlich und zeitlich variierender Lichtreize auf die Gesichtswahrnehmung, *Sitzungsber. Akad. Wiss. Wien Math.*, **115**, 633–648.

Maffei, L. and Fiorentini, A. (1973). The visual cortex as a spatial frequency analyzer, *Vision Res.*, **13**, 1255–1267.

Maffei, L., Morrone, M. C., Pirchio, M., and Sandini, G. (1979). Response of visual cortical cells to periodic and nonperiodic stimuli, *J. Physiol. (London)*, **296**, 27–47.

Marr, D. (1976). Early processing of visual information, *Philos. Trans. R. Soc. London*, **B275**, 485–526.

Marr, D. (1982). *Vision*, Freeman, San Francisco.

Marr, D. and Hildreth, E. (1980). Theory of edge detection, *Proc. R. Soc. London*, **B207**, 187–217.

Morrone, M. C. and Burr, D. C. (1986). Evidence for the existence and development of visual inhibition in humans, *Nature (London)*, **321**, 235–237.

Morrone, M. C. and Burr, D. C. (1988). Feature detection in human vision: a phase dependent energy model, *Proc. R. Soc. (London)*, **B235**, 221–245.

Morrone, M. C. and Burr, D. C. (1992). A model of human feature detection based on matched features, in *Robots and Biological Systems*, Dario, P., Sandini, G., and Aebischer, P., Eds., Springer-Verlag, Berlin.

Morrone, M. C., Burr, D. C., and Maffei, L. (1982). Functional significance of cross-orientational inhibition. I. Neurophysiological evidence, *Proc. R. Soc. (London)*, **B216**, 335–354.

Morrone, M. C., Burr, D. C., and Ross, J. (1983). Added noise restores recognition of coarse quantised images, *Nature (London)*, **305**, 226–228.

Morrone, M. C., Burr, D. C., Ross, J., and Moulden, B. (1991). Illusory squarewaves from staircase plus lines, *Perception*, **20**, A41.

Morrone, M. C. and Owens, R. (1987). Feature detection from local energy, *Pattern Rec. Lett.*, **1**, 103–113.

Morrone, M. C., Ross, J., Burr, D. C., and Owens, R. (1986). Mach bands depend on spatial phase, *Nature (London)*, 250–253.

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978a). Spatial summation in the receptive fields of simple cells in the cat's striate cortex, *J. Physiol. (London)*, **283**, 53–77.

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978b). Receptive field organization of complex cells in the cat's striate cortex, *J. Physiol. (London)*, **283**, 79–99.

O'Brien, V. (1958). Contour perception, illusion and reality, *J. Opt. Soc. Am.*, **48**, 112–119.

Oppenheim, A. V. and Lim, J. S. (1981). The importance of phase in signals, *Proc. IEEE*, **69**, 529–541.

Perona, P. and Malik, J. (1990). Detecting and localizing edges composed of steps, peaks and roofs, *Berkeley Tech. Rep.* 90/590.

Piotrowski, L. N. and Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase, *Perception*, **11**, 337–346.

Pollen, D. A. and Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex, *Science*, **212**, 1409–141.

Ratliff, F. (1965). *Mach Bands: Quantitative Studies on Neural Networks in the Retina*, Holden-Day, San Francisco.

Ratliff, F. (1984). Why Mach bands are not seen at the edges of a step, *Vision. Res.*, **24**, 163–166.

Robson, J. G. (1980). Neural images, the physiological basis of spatial vision, in *Visual Coding and Adaptability*, Harris, L. S. and Erlbaum, L., Eds., Hillsborough, NJ, 177–214.

Ross, J., Holt, J. J., and Johnstone, J. R. (1981). High frequency limitations on Mach bands, *Vision Res.*, **21**, 1165–116.

Ross, J., Morrone, M. C., and Burr, D. C. (1989). The conditions for the appearance of Mach bands, *Vision Res.*, **29**, 699–715.

Sachs, M. B., Nachmias, J., and Robson, J. G. (1971). Spatial frequency channels in human vision, *J. Opt. Soc. Am.*, **61**, 1176–1186.

Shapley, R. and Reid, R. C. (1985). Contrast and assimilation in the perception of brightness, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 5983–598.

Spitzer, H. and Hochstein, S. (1985a). Simple- and complex-cell response dependences on stimulus parameters, *J. Neurophysiol.*, **53**, 1244–1265.

Spitzer, H. and Hochstein, S. (1985b). A complex-cell receptive field model, *J. Neurophysiol.*, **53**, 1266–1286.

Stromeyer, C. F. and Julesz, B. (1972). Spatial frequency masking in vision: critical bands and spread of masking, *J. Opt. Soc. Am.*, **62**, 1221–1232.

Tyler, C. W. and Gorea, A. (1986). Different encoding mechanisms for phase and contrast, *Vision Res.*, **26**, 1073–1082.

Watt, R. J. and Morgan, M. J. (1985). A theory of the primitive spatial code in human vision, *Vision Res.*, **25**, 1661–167.

Wiesel, T. N. and Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey, *J. Neurophysiol.*, **29**, 1115–1156.

Wilson, H. R. and Bergen, J. R. (1979). A four mechanism model for threshold spatial vision, *Vision Res.*, **19**, 19–32.

Wilson, R., Knutsson, H. E., Granlund, G. H. (1983). Anisotropic nonstationary image estimation and its application, IEEE Trans. COM, **31**, 388–397.

Yuille, A. L. and Poggio, T. (1985). Fingerprint theorems for zero-crossings, *J. Opt. Soc. Am.*, **A2**, 683–692.