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 - 75

ANTONIO BORSELLINO COLLEGE ON NEUROPHYSICS

(15 May - 9 June 1995)

"The role of features in structuring visual images"

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

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

Ciba Foundation Symposium 184

HIGHER-ORDER PROCESSING IN THE VISUAL SYSTEM

Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation

HIGHER-ORDER PROCESSING

1.

The role of features in structuring visual images

David C. Burr and M. Concetta Morrone

Istituto di Neurofisiologia del CNR, Via S. Zeno 51, 56127 Pisa and Department of Psychology, University of Rome, Italy

Abstract. Edges and lines carry much information about images and many models have been developed to explain how the human visual system may process them. One recent approach is the *local energy* model of Morrone and Burr. This model detects and locates both lines and edges simultaneously, by taking the Pythagorean sum of the output of pairs of matched filters (even- and odd-symmetric operators) to produce the all-positive local energy function. Maxima of this function signal the presence of all image features that are then classified as lines or edges (or both) and as positive or negative, depending on the strength of response of the even- and odd-symmetric operators. If the feature is an edge, it carries with it a brightness description that extends over space to the next edge. The model successfully explains many visual illusions, such as the Craik–O'Brien, Mach bands and a modified version of the Chevreul. Features can structure the visual image, often creating appearances quite contrary to the physical luminance distributions. In some examples the features dictate totally the image structure, 'capturing' all other information; in others the features are seen in transparence together with an alternate image. All cases can be predicted from the rules for combination of local energy at different scales.

1994 Higher-order processing in the visual system. Wiley, Chichester (Ciba Foundation Symposium 184) p 129–146

It is well accepted that the visual system does not simply transmit information passively from the eye to the brain, but actively analyses the image to produce meaningful forms and objects. Certain parts of the image, generally referred to as 'features', are more salient than others and to a great extent determine the perceived structure of an image (see, for example, Barlow 1959, Marr 1976). In previous publications, we have highlighted the importance of features in structuring images and have proposed the 'local energy' model of feature detection as a means of detecting and analysing features. In this paper, we briefly review the model and pursue the idea that features provide structure to an image, extending these ideas to transparency.

For the local energy model, features are lines, edges or combinations of both (ignoring for the moment possible higher-level features such as corners and

2.

intersections). Many algorithms have been devised for detecting lines and edges (e.g. Marr & Hildreth 1980, Marr 1982, Watt & Morgan 1985, Canny 1986, Kingdom & Moulden 1992), all of which work reasonably well under some conditions but encounter difficulties under others. For most algorithms (that use only one class of operator, such as a difference-of-Gaussian), combinations of lines and edges are particularly difficult features to localize accurately, as each will produce a zero or maximum, but at different positions (see Burr & Morrone 1992 for further discussion).

The local energy model of Morrone & Burr (1988, 1993, Burr & Morrone 1990, 1992) explicitly recognizes the importance of lines and edges as image features and encodes both feature types simultaneously with the same algorithm. We achieve this by convolving the image simultaneously with pairs of matched filters, with even- and odd-symmetric line-spread functions (tuned, respectively, to lines and edges), and combining the output of these filters by Pythagorean sum. Peaks in the output (local energy) mark all features—lines, edges and combinations of them. The feature is then coded as a line or an edge (or both), depending on the relative strengths of the odd- and even-symmetric operators. Furthermore, if the feature is coded as an edge, the edge is the signal for a change in brightness, which continues until the next edge.

A useful aspect of the local energy model is that it gives a physical definition of visual features. Peaks in local energy occur at the positions on an image where the Fourier components tend to come into phase. The absolute value of the phase of the harmonics depends on the type of feature, being 0° or 180° for bright and dark lines and ±90° for edges. Intermediate phases indicate the presence of both edges and lines. Thus, local energy can be considered to be a polar representation of the even- and odd-symmetric output, where the amplitude (or norm) defines feature strength and phase (or argument) defines feature type.

To model the known properties of the human visual system, we compute local energy separately over several orientations and spatial scales. The maxima of local energy at each scale and orientation produce 'feature maps' that provide independent descriptions of the brightness of the image (similar to the model of Kingdom & Moulden 1992). While this strategy finds firm support from physiological studies, it unfortunately creates the difficult problem of how to combine the information over scale and orientation. Here various possibilities exist, none of which is well justified by physiological or psychophysical data. Our (provisional) approach is simply to sum the separate feature maps (given by maxima in local energy) at each scale with an 'indeterminacy' factor proportional to scale size (see Morrone & Burr 1993). For most images, the features tend to be consistent at all scales, so no difficulties arise. However, in some instances (illustrated below), different scales may contain different features and brightness signals, leading to interesting results.

3.

Brightness

Like most models of edge detection, the local energy model assumes that edges induce in the image a brightness change that extends until there is information to the contrary (given by another edge). The best known example of this is the Craik–O'Brien illusion, illustrated in most vision texts (such as Cornsweet 1970; see also Burr 1987). The stimulus is usually a simple image, high-pass filtered so the luminance is identical everywhere except near the border of, for example, a central circle. The central circle seems to be of homogeneous brightness, much brighter than the background. We suggest that the abrupt change of luminance at the border signals an edge and hence a brightness change that extends throughout the entire region. Here, the information is provided only at the higher scales. As the lower scales indicate nothing to the contrary, there is no conflict, so the high-scale brightness signal prevails.

Figure 1 shows more complex examples of the influence of the phase of the image features at different scales of brightness. Mach bands, the light and dark lines seen where the luminance ramps meet plateaux (Fig. 1A), were first observed by the Austrian physicist Ernst Mach (Mach 1865). Inspection of the image reveals that at the points where the lines are seen, there appears to be an accompanying change in brightness. The region of the ramp, where luminance is increasing continuously, seems to be of relatively homogeneous brightness, especially when the image contrast is low.

Again, both the lines and the brightness change are predictable from the local energy model (illustrated under the image profile). The middle profile shows the local energy of a relatively high scale, the lower profile that of a lower scale. At both scales of analysis, the peaks in local energy occur at the position of the features. At the higher scale, the phase at the energy peaks (illustrated by the polar plots) is near 0° or 180°, the signal for a light or dark line. This explains the perception of lines (where none exist on the profile) and also predicts quantitatively the conditions under which they occur (see Ross et al 1989). However, at the lower scale, the phase is not exactly 0° or 180°, but 30° and 150°, a combination of line and edge. This correctly predicts that in addition to the line, an edge should be perceived, with an accompanying brightness change. Note that there are no peaks on the ramp (including the point where it crosses zero) and no features are perceived there. In this example, although the phase of the features varies with scale, both the line and brightness change are perceived concurrently as part of the same feature.

Figure 1B is a modification of the classic Chevreul illusion (Chevreul 1890), where a series of luminance steps do not seem of constant brightness, but take on a 'scallopy' appearance. We have added thin lines to each of the luminance steps, destroying the scallopy appearance of the Chevreul illusion, and creating a brightness step at the position of the line. Again, this brightness step can be predicted by the phase of the energy peaks. Whereas the phase at the higher

4.

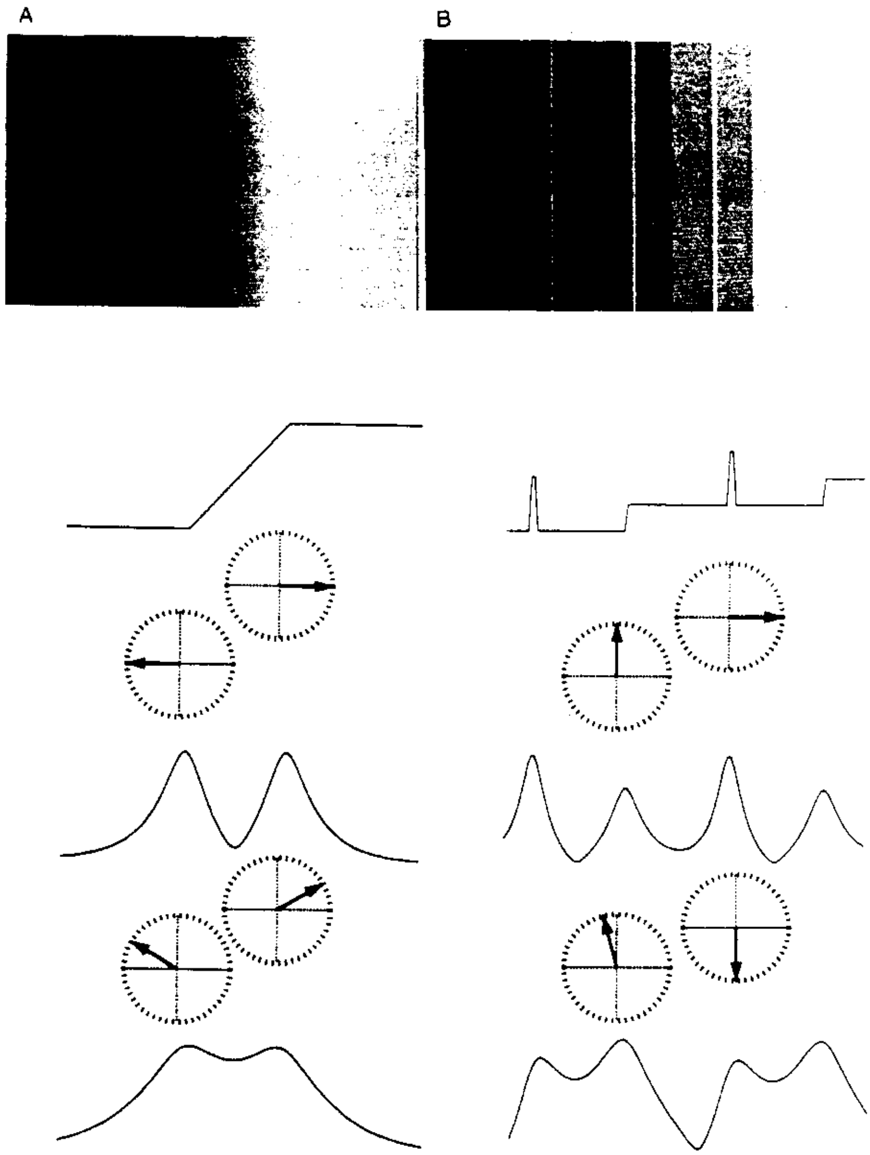

FIG. 1. Examples of Mach bands (A) and a modified version of the Chevreul illusion (B), with their illuminance profiles shown underneath (not to scale). The two lower curves show the local energy profiles computed at 2.4 cycles/ramp and 1.2 cycles/step (middle traces), and 1.2 cycles/ramp and 0.6 cycles/step (lower traces). The polar plots indicate schematically the phase of the local energy at the indicated peaks. At the lower scale, all the polar plots include a sine component, predicting a brightness change at the features.

5,

scale is near zero, at the lower scale it is $-90°$, the signal for a negative-going edge. This explanation works qualitatively, and also predicts quantitatively the magnitude of the induced brightness (Morrone et al 1994). Why the original Chevreul illusion should take on a scallopy appearance is less clear, but it is probably because of the continued series of edges of the same sign and a breakdown in the 'transitivity' of the brightness signalling mechanisms. The brightness signal accumulates across the series of same-signed edges, but the accumulation is not perfect (see Shapley & Reid 1985), leading to discrepant brightness signals about each edge.

The above images were chosen as examples where features can dictate brightness changes contrary to the luminance distribution of the image but predictable from the position and phase of the local energy peaks. They are interesting because the brightness information is different at different spatial scales. The edge information that produces the brightness change occurs only at the lower scale. However, the brightness generated (independently) at that scale is not confined by the low-scale features, but extends over the whole region up to the features of the finer scale, even though those features are lines. Brightness information may be generated at any scale, but if the features of different scales coincide (within a certain tolerance related to scale size), the high-scale features will dominate when the separate maps are summed (irrespective of feature type), and therefore will be most effective in defining the boundaries of the brightness region. However, as the following examples show, when the features of different scales are sufficiently distant to remain separate on summation, the high-scale features will not 'drag' the low-scale features, but both are perceived separately in transparency.

Transparency

One of the most powerful demonstrations of how features can structure an image is given by the 'blocked' image of Harmon & Julesz (1973) (Fig. 2A). The original image (not shown) has been 'sampled and coarsely quantized' by setting all the pixels within each square to the mean value of those of the original image. Although this sampling technique preserves sufficient low-frequency image information for face recognition (readily verified by blurring or distancing the image), this information cannot be extracted from the unfiltered blocked image.

The original explanation for this 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 & Julesz 1973). However, Fig. 2B casts some doubt on this explanation. There the phase of the spurious frequencies has been shifted by 90°, leaving their amplitude and distribution of maxima of local energy (and of maxima in local variance) untouched. The high-frequency features remain (as lines rather than edges), but are seen in transparency in front of a clearly recognizable face. Several other

6.

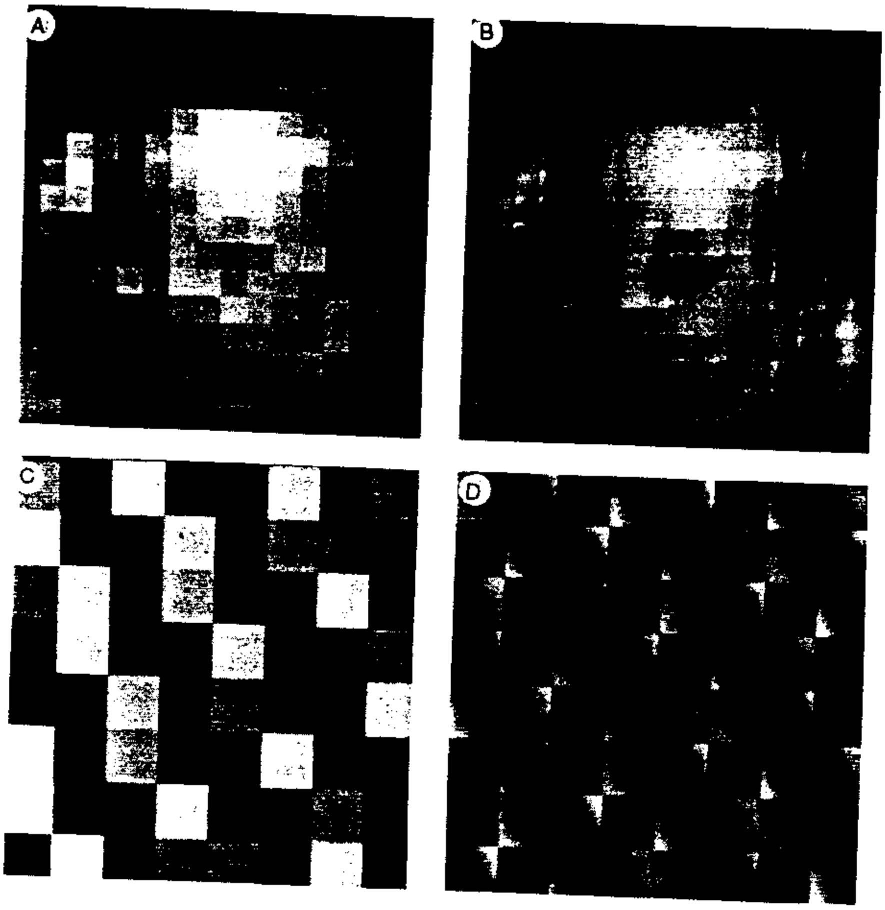

FIG. 2. Examples of coarse quantization of a natural image (A,B) and a simple sine wave (C,D). In the images on the left (A,C), all pixels within the blocks are set to the same mean luminance, preserving low-frequency information while introducing spurious high frequencies. In the images on the right (B,D), the phase of the spurious frequencies (higher than the Nyquist frequency) has been shifted by 90° (above) or 180° (below), leaving the amplitude untouched. After phase-shifting, the spurious harmonics are clearly far less detrimental to recognition than those in phase.

demonstrations, including the fact that adding more high-frequency noise can improve detection (Morrone et al 1983), all suggest that the blocks interfere with perception in ways other than simple critical-band masking (see also Hayes 1989).

Figures 2C and D illustrate the effect of blocking a more simple image, a sinusoid. Again, the sine wave is difficult to perceive in Fig. 2C, but becomes

7.

quite visible in Fig. 2D, when the high harmonics have been phase-shifted (this time by 180°). Figure 3 shows another blocked example that lends itself both to obtaining quantitative measurements and to modelling. The letter 'R' is obscured in Fig. 3A but quite visible (albeit somewhat blurred) in Fig. 3B. This effect is readily measurable quantitatively, as shown in Fig. 4. Letter detectability increases

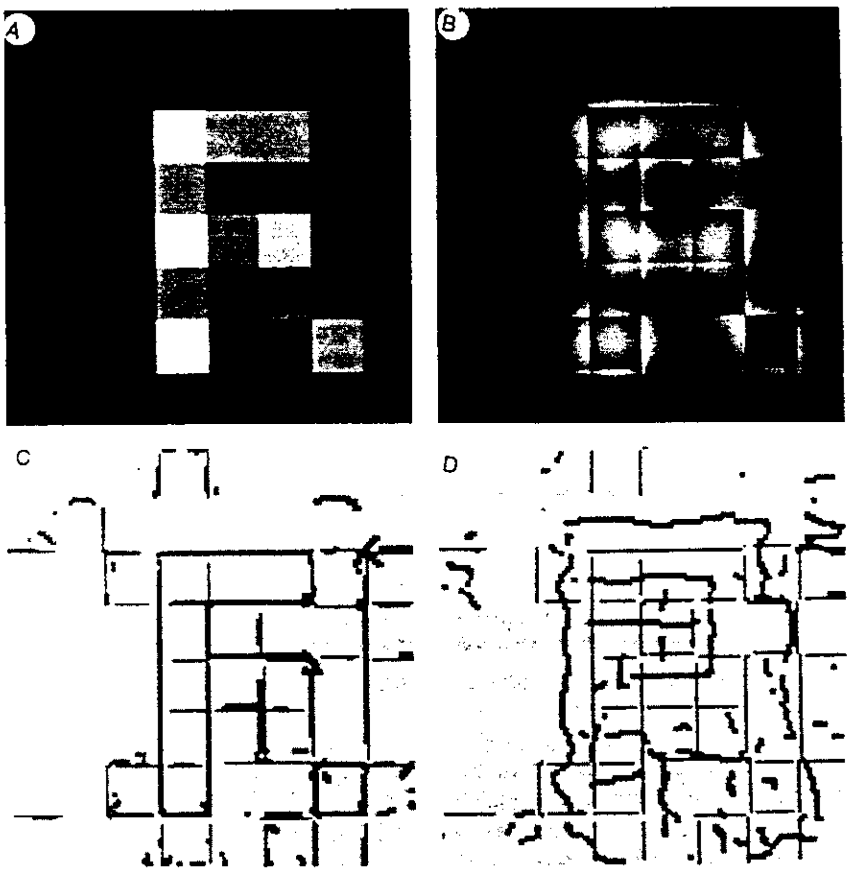

FIG. 3. An example of coarse quantization, with a simple letter form. The phases have been shifted by 180° in the right-hand image (B). The lower figures (C, D) show the peaks in local energy at a low scale (6 cycles/picture: thick grey lines) and higher scale (12 cycles/picture: thin black lines), computed with filters of four orientations of 60° bandwidth in orientation and 1.6 octaves in spatial frequency (full width, half height). When the harmonics are in phase, the maxima at the two scales are very similar, and both follow the block outlines (A, C). When out of phase, the maxima are quite different, with the lower scale tending to follow the letter outline, rather than the blocks.

8.

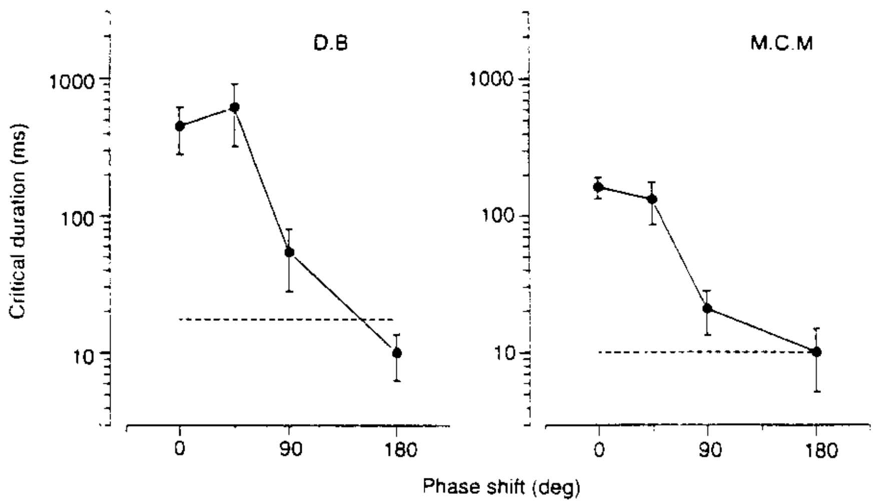

FIG. 4. Quantitative assessment of letter detectability after blocking. The letters were displayed on a monitor and after a certain interval a high-contrast noise mask was presented. Two subjects (D.B. and M.C.M.) were required to identify the letter (from a choice of six possibilities), at various stimulus offsets, to yield an estimate of *critical duration* (70% correct), plotted on the ordinate. The dotted line indicates detectability when the spurious harmonics were removed by blurring. When the spurious harmonics were in phase with the signal (0° on abscissa), the letters had to be presented for 200–500 ms to be detected, compared with 10–20 ms when the harmonics were out of phase, or blurred.

monotonically with the difference in phase of the signal and spurious frequencies, by more than an order of magnitude.

We have suggested that the high spatial frequencies of the blocked image do not mask or attenuate the low spatial frequency signal (by decreasing the gain of a hypothetical visual unit), but rather structure the way that the low-scale information is perceived (Burr & Morrone 1990, 1992). The blocking technique results in a high correlation in both position and phase of the low- and high-frequency content of the image. The high spatial frequencies all come into phase to form maxima of local energy along the borders of the blocks, so the feature maps at high and moderate scales follow the block pattern. When the feature maps are combined with the scale-related indeterminacy technique, the higher scales dominate and so the final feature map follows the block outlines. This feature map then structures the image, delineating it into discrete and separate objects, obscuring the lower-scale description of the face (Fig. 2A). However, when the phase of the spurious frequencies is changed, as in Fig. 2B, it breaks the phase-coherency between harmonics, creating features in distinct positions at the lower scales that are seen in transparency. The same discussion applies to the sinusoidal images of Figs 2C and 2D.

9.

Figures 3C and 3D illustrate the feature maps of the blocked letter at two scales (centred at 6 and 12 cycles/picture), obtained by searching for maxima along the preferred orientation, as suggested by Perona & Malik (1990). With the standard blocked image, the feature maps at these two scales coincide at the position of the blocks, so both will specify a blocked structure. However, shifting the phase of the spurious frequencies breaks the phase coherence in the pattern. This causes the peaks in local energy of the lower scale to occur at different positions from those of the upper scale, tending to follow the letter outline rather than the block structure. Summing these two maps will not result in a merger into one map where the blocked structure dominates, but will create two separate maps, signalling two distinct images to be perceived in transparency.

Figure 5 shows another example of how features may either structure the image (dragging low-scale information with them) or be seen in transparency. The patterns have been constructed so that the points of phase congruence align to create an arrow structure pointing left, while the lowest spatial frequencies (virtually a single harmonic) form an arrow pointing right. When the patterns are blurred, the upper and lower figures are identical, both showing a rightward-pointing arrow. However, with normal viewing, the top pattern seems to point left, when the lower pattern seems to contain both arrows in transparency. Again, the difference in the patterns is the relative phase of the low and high frequencies. In the upper pattern the low-frequency arrow is in phase with the higher harmonics, while in the lower pattern it is out of phase.

As with the blocked patterns, visual features have a profound influence on the perceived structure. In this example, the features change smoothly from line to edge, yet both classes of feature dictate the arrow structure. When all harmonics are in phase, the structure formed by the features is sufficiently strong and coherent to capture the low-frequency information and force the perceptual organization, against the organization dictated by the luminance information (a leftward-pointing arrow). When the low frequencies are out of phase, they break the phase coherence and, in a similar way to that illustrated in Fig. 3, allow two distinct feature maps to emerge to produce the sensation of transparency.

Motion

The concepts outlined for detection and classification of luminance contours may be extended readily to other visual domains, such as stereopsis, texture, colour and motion. Recently, we have applied the model to images in motion, with encouraging results (Morrone et al 1992, Del Viva & Morrone 1992, 1993). The local energy transform effectively performs the full-wave rectification required to detect second-order motion (see Chubb et al 1994, Wilson 1994 and Sperling et al 1994, this volume), and produces an all-positive output that greatly facilitates further elaboration. Image velocity is then given by the orientation

10.

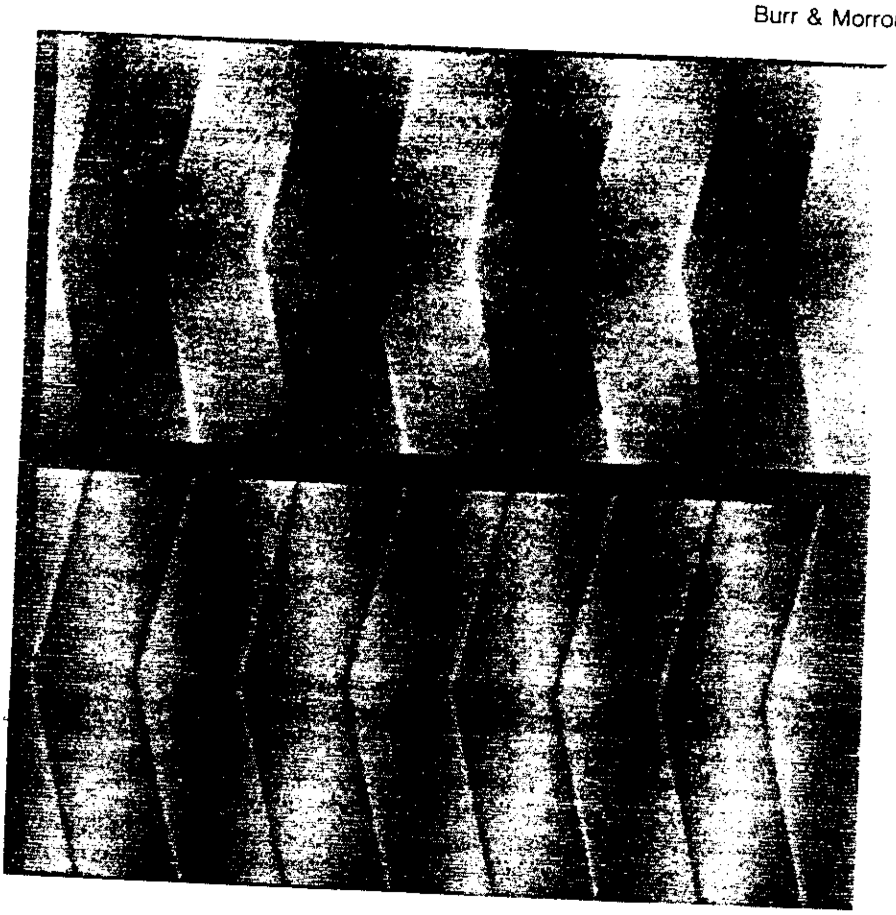

FIG. 5. An example of how features can structure an image. The two-dimensional luminance distribution L(x,y) of the two patterns is given by:

$$L(x,y) = L_0 + 4a/\pi \left\{ \cos(\phi + 2\pi)(y/4T - 3/8 + x/T) + \sum_{k=3}^{\infty} \cos\{2\pi [\,(y/2T - 1/2) + k(1/8 - y/4T) = k.x/T\,]\,\} \right\}$$

Where T is the spatial period of the fundamental. The harmonics are in phase for the upper pattern ($\phi = 0$) and out of phase for the lower pattern ($\phi = \pi$). The pattern is designed so that the higher harmonics come into phase to produce peaks in local energy, and hence features, that form an arrow pointing left. Although the features change systematically from line to edge, they structure the image compellingly when the lower harmonic is in phase, but allow it to be seen in transparency when out of phase.

11.

in three dimensions, defined as the direction of minimum curvature. This has proven successful over a wide range of conditions, including difficult images such as motion transparency.

Interestingly, many of the illusions demonstrated here in 2D space can be observed with image motion. For example, a pattern similar to the arrows of Fig. 5 (comprising only the upper half) can be generated by letting the ordinate represent time and the abscissa horizontal space. This produces a series of vertical gratings whose profile changes continuously over time. The impression of motion given by orientation in space–time is similar to the impression of orientation in Fig. 5. When the low spatial frequency is in phase with the higher frequencies, motion of the entire image is dictated by the motion of the features, even though the motion of average luminance is in the other direction (verified by blurring the image). However, when the low frequency is out of phase with the rest of the harmonics, transparent motion is seen simultaneously in both directions. The other illusions based on blocking can also be demonstrated for motion and are currently being investigated.

Conclusions

The examples shown here and in previous publications suggest that the local energy model can predict image appearance over a wide range of conditions. It is particularly successful with 'broad-band' images containing many Fourier harmonics that create strong peaks in local energy when they come into phase. However, the model is less successful with low-pass filtered images. In the extreme case, it fails completely to locate any features on a pure sinusoid (as its energy is uniform everywhere) and the motion algorithm does not detect drifting sine wave (simple first-order motion). Similarly, with heavily blurred images, like the 'R' of Fig. 3 or the patterns described by Georgeson (1994, this volume), the output is far from perfect.

There are several indications that low-pass images behave differently to high-pass or broad-band images, including the fact that negative faces are virtually unrecognizable in low-pass images, while they are as recognizable as positive faces in high-pass images (Hayes et al 1986). All this would suggest that a separate set of mechanisms may be implicated in encoding low-pass images. These mechanisms may respond more to the physical luminance levels of the image, rather than attempting to code 'features' or other symbolic extractions. However, when images comprise many harmonics (as do most natural images), the perceived structure is determined to a large extent not by the physical distribution of luminance, but by the visually salient features, formed when the harmonics come into phase with each other. The structure formed by the features is often sufficiently powerful to capture the gradual variation in image luminance.

12.

Acknowledgement

Supported by targeted grant from the CNR, *Robottica 93.00926.PF67.*

References

Barlow HB 1959 Possible principles underlying the transformations of sensory messages. In: Rosenblith WA (ed) Sensory communication. MIT Press, Cambridge, MA

Burr DC 1987 Implications of the Craik–O'Brien illusion for brightness perception. Vision Res 27:1903–1913

Burr DC, Morrone MC 1990 Edge detection in biological and artificial visual systems. In: Blakemore C (ed) Vision: coding and efficiency. Cambridge University Press, Cambridge, p 185–194

Burr DC, Morrone MC 1992 A non-linear model of feature detection. In: Pinter RB, Nabet B (eds) Non-linear vision. CRC Press, Boca Raton, FL, p 309–328

Canny J 1986 A computational approach to edge detection. IEEE (Inst Electr Electron Eng) Trans Pattern Anal Mach Intell 8:679–698

Chevreul ME 1890 The principles of harmony and contrast of colours (Martel C: transl). Bell, London

Chubb C, McGowan J, Sperling G, Werkhoven P 1994 Non-Fourier motion analysis. In: Higher-order processing in the visual system. Wiley, Chichester (Ciba Found Symp 184) p 193–210

Cornsweet TN 1970 Visual perception. Academic Press, New York

Del Viva M, Morrone MC 1992 Feature detection and non-Fourier motion. Perception 21(suppl 2):41

Del Viva MM, Morrone MC 1993 The importance of phase congruency for image capture. Invest Ophthalmol & Visual Sci (suppl) 34:1598(abstr)

Georgeson MA 1994 From filters to features: location, orientation, contrast and blur. In: Higher-order processing in the visual system. Wiley, Chichester (Ciba Found Symp 184) p 147–169

Harmon LD, Julesz B 1973 Masking in visual recognition: effect of two-dimensional filtered noise. Science 180:1194–1197

Hayes A 1989 Representation by images restricted in resolution and intensity level. PhD thesis, University of Western Australia, Perth, WA, Australia

Hayes T, Morrone MC, Burr DC 1986 Recognition of positive and negative bandpass-filtered images. Perception 15:595–602

Kingdom F, Moulden B 1992 A multi-channel approach to brightness coding. Vision Res 32:1565–1582

Mach E 1865 Über die Wirkung der räumlichen Vertheilung des Lichtreizes auf di Neztzhaut, I. Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenchaften 52:303–322

Marr D 1976 Early processing of visual information. Philos Trans R Soc Lond B Biol Sci 275:485–526

Marr D 1982 Vision. WH Freeman, San Francisco, CA

Marr D, Hildreth E 1980 Theory of edge detection. Proc R Soc Lond Ser B Biol Sci 207:187–217

Morrone MC, Burr DC 1988 Feature detection in human vision: a phase-dependent energy model. Proc R Soc Lond Ser B Biol Sci 235:221–245

Morrone MC, Burr DC 1993 A model of human feature detection based on matched filters. In: Dario P, Sandini G, Aebischer P (eds) Robots and biological systems: towards a new bionics? Springer-Verlag, Berlin, p 43–64

13.

Morrone MC, Burr DC, Ross J 1983 Added noise restores recognition of coarse quantised images. Nature 305:226–228
Morrone MC, Del Viva M, Burr DC 1992 A visual model for simultaneous analysis of form and motion. Invest Ophthalmol & Visual Sci (suppl) 33:1441(abstr)
Morrone MC, Burr DC, Ross J 1994 Illusory brightness step in the Chevreul illusion. Vision Res 34:1567–1574
Perona P, Malik J 1990 Detecting and localizing edges composed of steps, peaks and roofs. In: Proceedings of the international conference on computer vision, Osaka, Japan, 1990. Institute of Electrical and Electronics Engineers, New York, p 52–57
Ross J, Morrone MC, Burr DC 1989 The conditions for the appearance of Mach bands. Vision Res 29:699–715
Shapley R, Reid RC 1985 Contrast and assimilation in the perception of brightness. Proc Natl Acad Sci USA 82:5983–5986
Sperling G, Chubb C, Solomon JA, Lu Z-L 1994 Full-wave and half-wave rectification in motion and texture. In: Higher-order processing in the visual system. Wiley, Chichester (Ciba Found Symp 184) p 287–308
Watt RJ, Morgan MJ 1985 A theory of the primitive spatial code in human vision. Vision Res 25:1661–1674
Wilson HR 1994 Role of second-order motion signals in coherence and transparency. In: Higher-order processing in the visual system. Wiley, Chichester (Ciba Found Symp 184) p 227–244

DISCUSSION

Morgan: Does the local energy model make any key predictions that differ from those of other models? A peak in local energy at an edge or a bar corresponds to a peak in the second spatial derivative or to a zero crossing in the first derivative; both these features have been used by other models. What are the specific advantages of detecting local energy?

Burr: In the case of a very simple feature of known type, such as an isolated edge or line, there is no advantage at all, as specific filters matched to the feature type will be more efficient. The problems arise when the feature type is not known in advance, particularly if it happens to be a superimposed line and edge, because both trigger linear features in different ways: for an even-symmetric operator, the line will cause a peak and the edge a zero crossing. The sum of these is a displaced peak and zero crossing, neither corresponding in position to the feature. The local energy operator, however, responds positively to both feature types, so the two different feature types will not interfere with each other.

Parker: Take a cube that is rotating relative to a light source. As the cube rotates, the luminance distribution across the edge of the cube can reverse its sign: that is, a face of the cube that is at one position brighter than its neighbour along the edge can become darker than its neighbour if the cube rotates appropriately. If you want a computation to be consistent under such changes of edges under illumination, then you might want to do that kind of local energy computation. In a sense, you may be gathering information that's useful for some shape-related properties, rather than simply detecting things in a flat image.

14.

Malik: There's always the question of whether a visual task can be carried out by linear filtering or whether it requires an essential non-linearity. Perona and I proved that whenever you have a mixed feature, such as a combination of an edge and a bar (and it is an arbitrarily chosen combination that is not known in advance), it is not possible for a finite family of linear filters to locate the edge (Malik & Perona 1990). However, an energy filter can. An important point is that it's not necessary that the even and odd filters of the pair be in strict quadrature.

Morgan: It may be relevant that Vernier acuity for edges is reduced when they are of opposite contrast polarity (Mather & Morgan 1986). The situation for bars is more complicated: when the gap between them is smaller than about 10 arc min, Vernier thresholds are higher for opposite- than for same-polarity bars. At greater separations there is no difference (Morgan 1990). So the local energy model may be making some predictions that are not borne out by the psychophysics.

Wilson: David, one of the key things that you glossed over was the issue of capture. For example, you put lines in the centre of the bars in the Chevreul illusion (Fig. 1B) and you said that the high-frequency scale localizes the lines, but the low-frequency scales show a luminance difference. In the illusions, sometimes there was capture, sometimes there wasn't. Do you have a mechanistic model that predicts conditions under which capture will occur?

Burr: One of the key ideas to the energy model is coherence in arrival phase of the Fourier harmonics. When the phases of the harmonics within a channel are coherent, they produce a single feature that will 'capture' other nearby features at higher or lower scales. When the coherence is broken, as in Figs 2, 3 and 5, another set of quite distinct features is created and transparency is seen.

If you consider a motion illusion, where the ordinate is time and the abscissa space, the patterns create a set of feature-defined gradients, a form of second-order motion, moving in one direction and first-order motion (of a single sinusoid) in the other direction. When the sinusoid is in phase with the other components, it contributes to the same peaks in spatiotemporal energy and the first-order motion will be 'captured' by the feature motion; the whole pattern of drift is in one direction. When the sinusoid is out of phase, it is seen moving *transparently* over the features in the opposite direction (Del Viva & Morrone 1993).

Sperling: There have been some displays in which it seems to me that the quadrature is very like the second-order component. In other schemes, where we divide things up into first- and second-order components, it is possible to make displays in which there is no first-order stimulus and you see the second-order component all by itself. As I looked at some of your demonstrations, I thought about the possibility of different second-order computations that might give you the capture results directly.

15.

Watt: The differences between this model and the Watt & Morgan (1985) model (expanded in Watt 1988, 1991) are diminishing as the years go on. If your model combines information from different scales, particularly in ways that keep edge information segregated from line information, it is extremely close to the Watt & Morgan model. It doesn't matter much whether you detect peaks or centroids in response waveforms. There is, however, one very strong, logical difference that is not just a difference in detail: we never supposed that our account was intended to mark the location of a feature in a direct and completely linear geometrical mapping of the image. We supposed that our model was providing the locations in some sample version of the image that could be used to create spatial relations between different features—that's very different. We weren't concerned with having a transformation of the image that produced a discrete peak bang on where subjects saw an edge. We were quite happy to go for something that found response either side and then inferred, subsequently, that the edge lay in between. It seems to me that one could create a very strong distinction between two classes of model: those that are firmly rooted in mappings that are strictly within the image and produce edge maps, and those that are in some sense far more abstract (which is where I would put ours). This seems to be the major functional difference between the two classes of model: the rest seem increasingly less important.

Burr: Yes, I agree that despite the fact that the two models evolved to explain different aspects of vision, they are in fact quite similar in many respects. One difference, perhaps, is the fact that we do not analyse lines and edges separately, as you do, but combine them to form energy. Only later is the distinction made (after the possibility of 'feature capture').

Watt: But in terms of sheets of cells in V1 (or wherever), are you really supposing that each cell knows exactly where its receptive field lies with respect to the coordinate framework of the optical image? That seems to me to be another critical difference.

Bergen: Most models of what things look like—that is, models of perception—make a single prediction for each stimulus. Most perceptual illusions, however, display some degree of perceptual instability. Certainly, in some of the cases that you showed, such as the Chevreul illusion, you can see the effect (in this case a brightness illusion) or not depending on how you look at it. This is also evident in some of the work that I have done on texture in which the strength of the segregation seems to depend on the general level of visual attentiveness (see, for example, Bergen 1991). Even if the underlying processing is taken to be largely deterministic, when it comes to a model of the percept as such there needs to be some recognition of the fact that it is governed by a dynamic process.

Watt: That has always worried me. In looking at all these phenomena, my impression has been that if the visual system had some degree of selectivity concerning which band of spatial scales it was going to attend to, it would then

16.

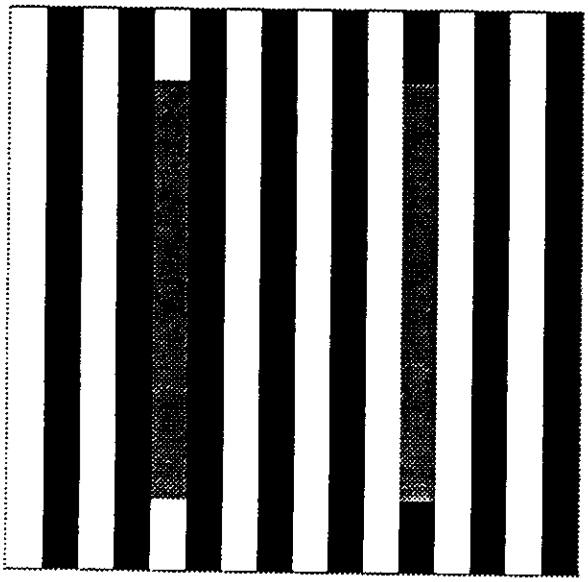

FIG. 1. (*Chubb*) A simple version of White's illusion. Although the grey bars in this figure are identical, the one on the left appears darker than that on the right. (See text for discussion.)

have the option to determine, through that control process, which pattern it would see at any one time. If, on top of that, the process that's choosing bands of scales to look at is hunting to establish where there are coherent patterns that can be used to organize information at other scales, then you would have a system that could oscillate quite widely between different organizations.

Bergen: There are some illusions that are very hard not to see, whereas others can easily be seen in more than one way. I think there is some information in this difference that most theories ignore.

Chubb: What are the illusions that you classify as hard not to see?

Bergen: Mach bands are a good example of an illusion that is relatively stable. For me, they just don't go away. Some people have a different experience of this, however.

Burr: Illusions such as Mach bands are easy to measure, with good agreement between observers. For many other illusions, I agree that there can be a 'bistability' that our model does predict, when more than one process is operating, with mutual inhibition between the two.

Bergen: I haven't heard anything yet that predicts very well which aspects of illusions are going to be more stable and which ones less.

Burr: Often there may be very simple explanations. For example, low spatial frequencies are heavily attenuated when stationary, but become more visible after blinks or eye movements; this can produce a transient change in appearance.

17.

The blocking illusion is interesting. It has always been regarded as one of the more robust visual illusions, guaranteed to work even for classroom demonstrations. When we first measured the magnitude of the effect (as a function of the signal:noise ratio), the blocked images required 1.5 log units more signal than the unblocked images to be recognizable. However, after a day's practice, this effect almost vanished. We incorporated several variants, randomizing every conceivable dimension and tried using brief presentations, all of which initially made the task virtually impossible, followed by rapid improvement. All this implies an immense plasticity in the system, perhaps of the type discussed by Charles Gilbert previously (Gilbert 1994, this volume).

Gilbert: Individual cells are capable of operating over different spatial scales, under both attentional and contextual control.

Bergen: We have to make a distinction between two very different things that can have the same effect on the outcome of a psychophysical experiment. The first is learning to do a recognition or a discrimination task with some unfamiliar stimulus; that is, something that doesn't look like the usual appearance of the thing that we are trying to recognize. The second is perceptual plasticity which causes the unfamiliar stimulus actually to start to look like the things that we are trying to recognize.

Shapley: If you create Mach bands with, say, a triangle wave, at moderate contrast and you fixate your eyes, the lines that you see at the peaks of the triangle wave can fade and the pattern be transformed into an apparent square wave.

Georgeson: That effect is entirely due to the interaction between afterimages and changes of fixation. The negative afterimage is shifted slightly and is then added to the stimulus waveform, effectively transforming it into a completely new wave form. If you do the experiments with brief presentations, all those instabilities go away (Georgeson & Turner 1984). Demonstrations that use prolonged inspection of this kind are rather dangerous in that low-level effects like the interaction between fixation and afterimages can make unstable the percepts that, with controlled observations, are much more stable (Georgeson 1984).

Chubb: One illusion that you did not discuss, but that seems relevant for distinguishing your model from, say, Grossberg & Todorovic's (1988) model, is White's illusion. There are several interesting variations of this illusion, but a simple version is shown in Fig. 1 (*Chubb*). The grey bars in this figure are physically identical. Nonetheless, the bar on the left appears darker than that on the right. This effect is problematic for models, such as Grossberg & Todorovic's (1988), that attempt to explain lightness perception on the basis of some sort of lateral inhibitory interaction across boundaries. The difficulty is that the region immediately bounding the left-hand bar has a lower mean luminance than the corresponding region bounding the right-hand bar. Thus, under models of lightness perception based on lateral inhibition, we should

18.

expect the left-hand bar to appear lighter than the right-hand bar. What does your model predict about the appearance of this figure?

Burr: We have not yet modelled White's effect in detail, but I agree it is a good illusion to study. I expect our model will predict the result, at least qualitatively. With the standard White's stimulus, one grey square falls on a light and the other on a dark bar of a grating, so each square has two white and two black sides. However, at the lower scales (lower than the grating periodicity) the edges within the bar will be stronger than those between adjacent bars, accounting for the brightness induction. The same explanation can probably be extended to the chequerboard version of this illusion, bearing in mind that the low frequencies of a chequerboard are the diagonals.

References

Bergen JR 1991 Theories of visual texture perception. In: Regan DM (ed) Spatial vision. Macmillan, London, p 114–134

Del Viva M, Morrone MC 1993 The importance of phase congruency for image capture. Invest Ophthalmol & Visual Sci (suppl) 34:1598(abstr)

Georgeson MA 1984 Eye movements, afterimages and monocular rivalry. Vision Res 24:1311–1319

Georgeson MA, Turner RSE 1984 Stability of phase recognition in complex spatial waveforms. Vision Res 27:1343–1360

Gilbert CD 1994 Circuitry, architecture and functional dynamics of visual cortex. In: Higher-order processing in the visual system. Wiley, Chichester (Ciba Found Symp 184) p 35–62

Grossberg S, Todorovic D 1988 Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. Percept Psychophys 43:241–277

Malik J, Perona P 1990 Preattentive texture discrimination with early visual mechanisms. J Opt Soc Am A Opt Image Sci 7:923–932

Mather G, Morgan MJ 1986 Irradiation: implications for theories of edge localization. Vision Res 26:1007–1015

Morgan MJ 1990 Hyperacuity. In: Regan DM (ed) Spatial vision. Macmillan, London, p 87–113

Watt RJ 1988 Visual processing: computational, psychophysical and cognitive research. Lawrence Earlbaum, London (Essays Cognit Psychol Ser)

Watt RJ 1991 Understanding vision. Academic Press, London

Watt RJ, Morgan MJ 1985 A theory of the primitive spatial code in human vision. Vision Res 25:1661–1674

19