Some aspects of information processing in biological vision

Y. V. Venkatesh

Computer Vision and Artificial Intelligence Laboratory, Department of Electrical Engineering, Indian Institute of Science, Bangalore 560 012, India

An attempt is made to present some challenging problems (mainly to the technically minded researchers) in the development of computational models for certain (visual) processes which are executed with, apparently, deceptive ease by the human visual system. However, in the interest of simplicity (and with a nonmathematical audience in mind), the presentation is almost completely devoid of mathematical formalism. Some of the findings in biological vision are presented in order to provoke some approaches to their computational models. The development of ideas is not complete, and the vast literature on biological and computational vision cannot be reviewed here. A related but rather specific aspect of computational vision (namely, detection of edges) has been discussed by Zucker[1], who brings out some of the difficulties experienced in the classical approaches.

Space limitations here preclude any detailed analysis of even the elementary aspects of information processing in biological vision. However, the main purpose of the present paper is to highlight some of the fascinating problems in the frontier area of modelling mathematically the human vision system.

'Facts not yet accounted for by available theories are of particular value for science, since it is on them that its development primarily depends.'

A. Butlerov[2]

SEEING is obvious to us, as human beings. However, on careful examination, we discover that the problem of how we identify objects – how we are able to tell cats from dogs, chairs from tables – is a fundamental one. Hebb[3] was one of the early workers to have made a serious attempt to analyse it.

Recognizing patterns (and taking action on the basis of the recognition) is the principal thing that most living systems do. It appears that one can learn much from a study of the way biological systems operate. Perhaps for this reason, vision has always been a paradigm problem for artificial intelligence (AI). Of course, there exists a considerable knowledge of the physiology of nerve cells[4] and the neuroanatomy of certain insects and mammals (including man)[5], but this has not yet provided us with an implementable logical design for the brain, not even for one aspect like vision. One constantly wonders whether we can treat the eye as the 'window' through which we can examine more efficiently the workings of the brain!!

In this context, it is appropriate to look back and analyse Turing's classic essay[6], which provides an affirmative answer to the question 'Can machines think?' He suggests that the machines (namely, digital computers) can be built which will be able to 'compete with men in all purely intellectual fields'. Turing begins with an attempt to define the meaning of the terms 'machine' and 'think', and decides, in the interest of avoiding ambiguity, to replace the question by another which is relatively unambiguous, involving an 'imitation game', in which three objects participate: a computer, a human being and an interrogator. The goal of the imitation game is to be able to distinguish between the human being and the machine, by querying both of them *appropriately*.

It is also known that the operations of the nervous system in processing information are fundamentally different from those a computer would perform under a similar situation[7]. Moreover, the operations the nervous system performs apparently lack logical precision and arithmetical depth. Nevertheless, the brain executes complicated tasks, using its own logic and arithmetic, with an efficiency unmatched by any known automaton. A comparison between the brain and the computer should refer to their (hardware or physical) components and their organization, as well as the representation and transmission of information:

- In the digital computer, the paradigm (which dominates computation) is that information must be digitized to guard against noise and degradation. Note that digitization imposes precision on an inherently imprecise physical system. A neuron, in contrast, is an analog device: its computations are based on smoothly varying ion currents rather than on bits (on–off) representing discrete ones and zeros. Yet, neural systems are generally accepted to be superbly efficient information processors.

- The connections among the neurons are numerous – any neuron may receive many thousands of inputs in a three-dimensional distribution. In contrast, the connections in a digital computer are limited to a small number, in a two-dimensional distribution.

- While the neurons operate about a million times more slowly than silicon chips, they are capable of doing millions of operations simultaneously, while the computer operates in a serial, one-step-at-a-time fashion.
- The individual neurons in the retina consume one ten-millionth as much power as the digital counterparts do.
- Neurons operate with far less precision than the digital computer.

Therefore, biological computation is very different from its digital counterpart, and any modelling of the former is to be made only in terms of the tasks of information processing that they perform. In other words, the visual processing tasks are to be decoupled from the hardware that performs them (see note 1).

Turing-type question for vision

The Turing question of relevance to vision is: Can machines see (see notes 2 and 3)?

If our goal is to understand the information processing aspects of vision, we need to distinguish, according to Marr[9], among three levels of analysis:

1. *Computational theory.* What is the problem of vision, and in what manner can (and do) the physical constraints enable a *unique* solution to be determined?
2. *Algorithm.* A clear procedure to implement the computational theory.
3. *Implementation.* Hardware or neural-ware realization of the algorithm.

As far as a theoretical answer is concerned, it should be added here that there are many perceptual processes (like illusions) whose problem specification (in the sense of Marr) itself is not clear.

In order to provide a practically (as distinct from logically) satisfying answer to this question (see note 4), the words 'machine' and 'see' are to be defined appropriately. As far as the equivalent 'imitation game' question is concerned, it has to be framed in such a way as to bring in a robot with cameras as eyes, as a participant. The human (H) and the two-eyed robot (R) are to be presented with questions by the interrogator who displays on the television screen, for both H and R, the same images of the various scenes. The Turing-type interrogation can be proposed. Both H and R are to answer questions related to the contents of the images displayed on the screen.

In fact, one wonders whether the question and answer method (in the imitation game) proposed by Turing is appropriate for pitting the human against the machine for 'seeing'. Interestingly, in Turing's imitation game the slowness and inaccuracy in arithmetic are the human weaknesses, and hence betray the human if he were to act as a machine. But not so in 'seeing' (see note 5)!

In the manner of Turing's disclaimers, it should be added that we should not ask whether the machines at present available would be able to see, but whether there are imaginable machines which would. The formulation of test questions for the imitation game in 'seeing' appears to be an interesting research problem by itself. A typical question could, for instance, be: Given an image of a scene as an input (for example, the image of Figure 1), can the machine print out on a typewriter a statement about what objects the scene contains and where they are?

Recognition of patterns and machine vision

In any attempt to automate recognition capabilities of biological vision, we should analyse how we (as humans) recognize the difference between, say, a square and a circle? Or, in general, how does a biological organism abstract the attribute of shape of an object (see note 6)? Observe that a square is recognizable as such, by humans, in different sizes and orientations and in different parts of the visual field. (In terms of biology, the recognition of a square is obviously independent of the particular groups of retinal cells excited.)

A visual image on the retina is nothing more than a pattern of light, and patterns are a collection of contours and edges, which in turn are defined as regions of sharp changes in intensity of light falling on the retina. One of the basic properties of the visual system is its sensitivity to contrast, the ability to detect an edge or a contour, which is determined by a change in brightness and darkness across the visual field. It is astonishing that our eyes execute complicated visual tasks (like distinguishing between a shadow and a pothole) with phenomenal ease. In fact, the human retina is estimated to perform more than 10 billion calculations per second before an image even reaches the optic nerve.

The major goal in machine vision is automatic recognition of objects, which implies that a correspondence be found between elements of the image and an *a priori representation of objects in the world.* It is not yet clear what type of prior world knowledge is really the heart of the matter. A recent paper by Pavlidis[11] offers some *subjective* comments on the reasons for the relatively slow rate of progress in machine vision in the last quarter of the century.

If the literature is any guide, whatever systems have been designed and built in the various laboratories of the world work in specialized domains only, requiring careful lighting and imaging conditions. For instance, bin-picking (i.e. picking objects jumbled together in a bin, not laid out flat without occlusion) is still beyond our reach. There appears to be nothing of significance in

Figure 1*a*. Aerial image of a part of the Earth's surface.

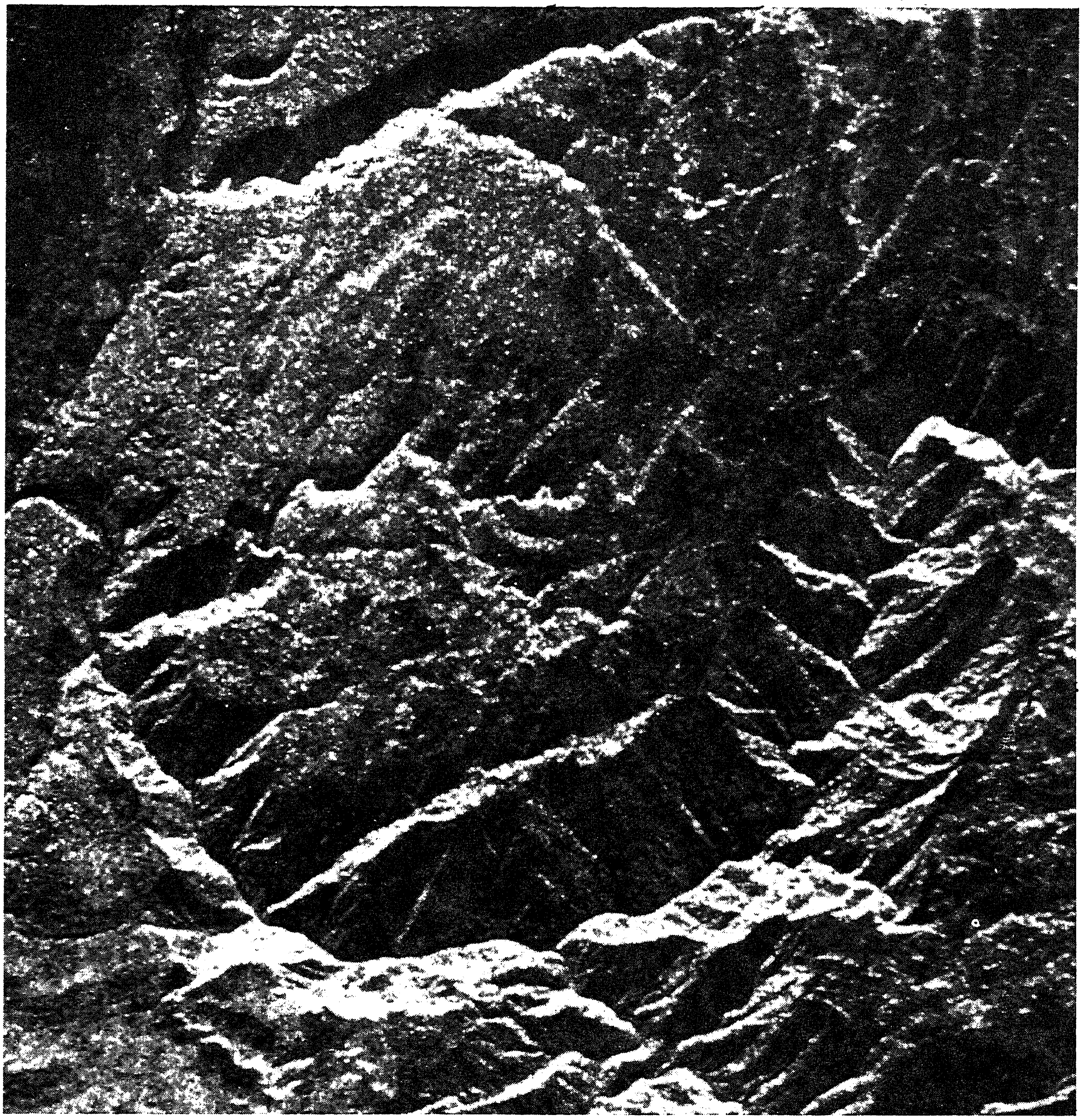


Figure 1*b*. Satellite image of a hilly terrain on the Earth's surface.

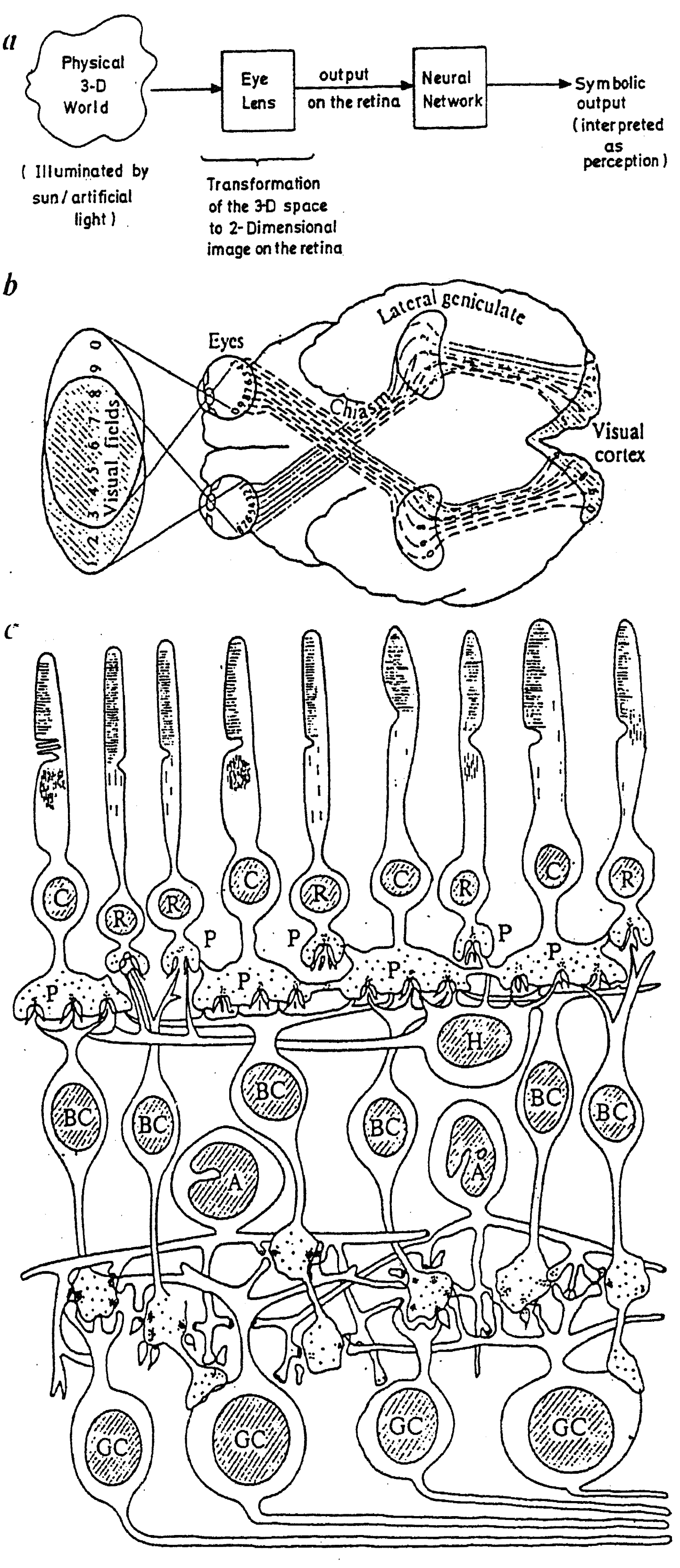
machine vision except for some very special applications in *controlled* environments like the OCR (optical character recognition) and industrial inspection of specific parts or objects.

However, we read in journals and newspapers that robots aided by image processors are doing wonderful things: car manufacture, packaging chips, and inspecting medicines, to name a few. Yet, it is to be emphasized that these automatons accomplish these jobs only under controlled conditions: lighting must be strictly modulated and objects must be carefully positioned to avoid ambiguities. If robots are to replace humans in every-day-life – on farms, construction sites and the like – they must be able to cope with a far more bewildering array of sensory data.

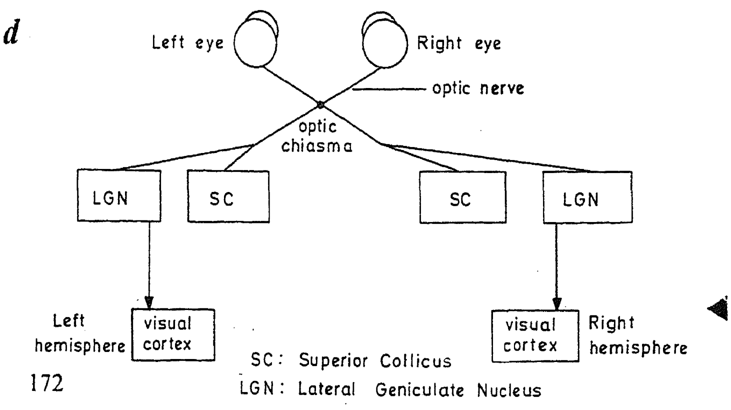
Biological vision

The human visual mechanism operates at an incredible speed of 3–6 ms to form a coherent image during the day, and 100 ms to form an image at night. Therefore, to the human eye, vision seems instantaneous, and we are usually unaware of the visual processing that is taking place.

In some primitive creatures, much of the information processing is done by the eye itself, and hence the eye can be called a mini-brain! However, more advanced creatures actually 'see' with their brains, since the trigger cells that respond to specific shapes and colours are found in the visual cortex itself. But even the human eye edits the visual information that comes to it. Each



The retina: C—cone; R—rod; P—pedicel; BC—bipolar cell; A—amacrine cell; GC—ganglion cell



eye has approximately 125 million photoreceptors, but there are only one million ganglion cells leading to the optic nerve that forwards the visual signals to the brain (Figure 2). Much of what the eye sees never reaches the brain.

From the physics of the human vision system, it is evident that the images formed on our retinas are upside down. From two such inverted two-dimensional projections of the three-dimensional world, the brain interprets the image 'right-side-up' and in three dimensions! Of course, the eye also uses visual cues, such as distance and texture, to make sense of the images it receives.

Neural architecture

In humans, the cerebral cortex almost completely envelops the rest of the brain. It is very complex, not only structurally but also in its functions. Vision is supposed to provide the brain with nine-tenths (?) of the information coming from all the senses. Humans perform recognition and other related activities (including locomotion) in a highly complex world of nonstationary objects under variable lighting conditions. The two special characteristics of the human visual system are:

- Visual processing is done not in a single stage but in multiple stages, not just hierarchically (see note 7) but rather heterarchically (Figure 3). This is a consequence of the physiological structuring (or architecture) of the retina – visual pathway – visual cortex mechanisms (Figure 2 *d*).
- Massively parallel computation.

The hierarchy refers to the representation of visual data in increasing abstraction – from the image intensity on the retina to changes in (i) intensity (used for stereo perception), (ii) surface composition, (iii) location, (iv) orientation, (v) end points of features, (vi) distance, (vii) reflectance and the like to the symbolic interpretation of the visual scene. Some intermediate representations are believed to be necessary before object recognition takes place. This seems to be one of the ways to achieve reliable correlation between our perceptions and the objects in the world, in spite of varying illumination and other factors. However, the contribution of high-level knowledge and inferential procedures to the vision process is not clear.

In this multilayered structure, two types of flow of information are involved:

- The (direct) flow of information is from the receptors (in the retina) through early processing to higher-level vision. These computations are organized as

Figure 2. ***a***, Block schematic representation of the human visual system; ***b***, essential elements of the visual pathway; ***c***, some details of the retinal structure; ***d***, block schematic of the visual pathway.

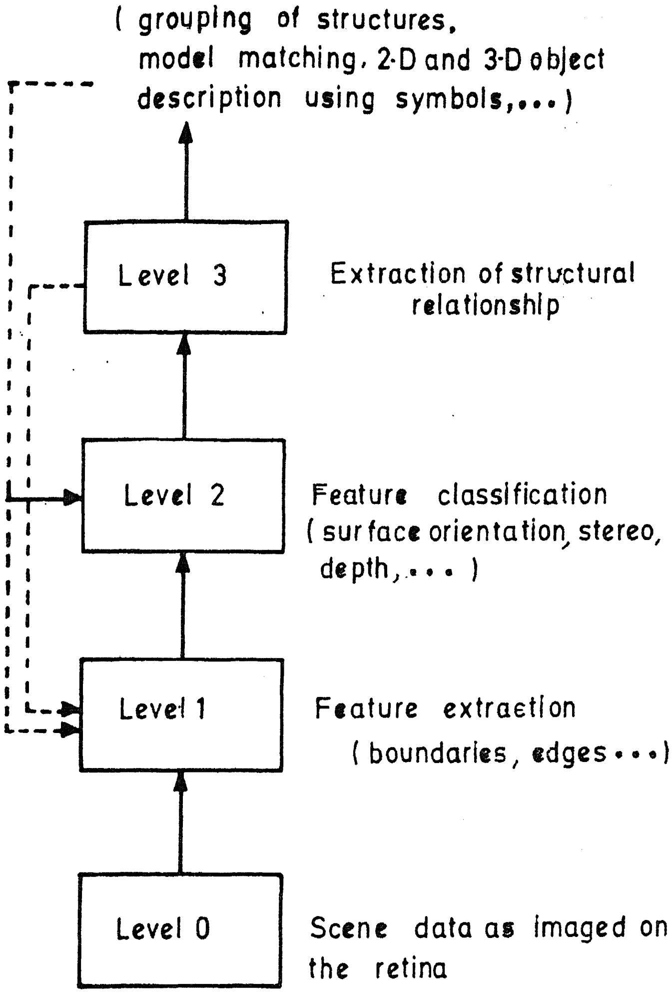


Figure 3. Hierarchical analysis of the image data.

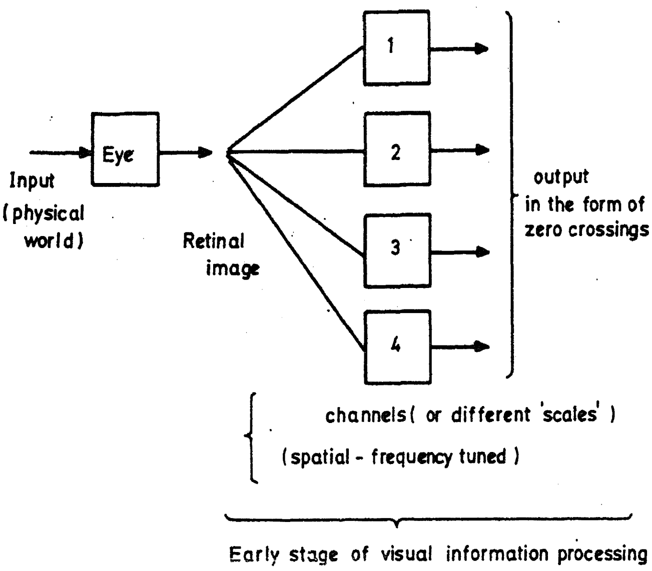


Figure 4. Multichannel processing of the retinal image.

sequences of filter-like operations within a set of parallel channels (Figure 4). At this level (or levels), signal processing operations can be handled in terms of linear (and nonlinear) systems and information theory.

- The feedback flow of information is from the later stages of processing, which modify and control the earlier stages (Figure 3).

Interactions between direct (or afferent) and feedback (or efferent) processes can be regarded as a form of *cooperative computation*, which forms the basis for some of the present-day models of visual processes.

Structure of the visual pathway

The output from each eye is conveyed to the brain by about a million nerve fibres grouped together in the optic nerve. These fibres are the axons of the ganglion cells of the retina (Figure 2 *c*). The messages from the light-sensitive elements, the rods and cones, pass through two to four synapses, and involve four other types of retinal cells before they reach the ganglion cells. In this path, a certain amount of analysis of retinal image takes place.

A large fraction of the optic-nerve fibres pass uninterrupted to two nests of cells deep in the brain called the lateral geniculate nuclei, where they make synapses (Figure 2 *d*). The lateral geniculate cells, in turn, send their axons directly to the primary visual cortex. However, the primary visual cortex is, probably, an early stage in processing the signals in terms of the degree of abstraction involved. Each geniculate and each cortex receives input from both the eyes, and each is concerned with the opposite half of the visual world. One subdivision of the cortex is the primary visual cortex (also known as the striate cortex or area 17), the most elementary of the cortical regions concerned with vision. The cortex is subdivided into areas having widely different but, in general, ordered functions. The flow of visual information is first to a primary cortical area and from there, either directly or via the thalamus, to a succession of other more specialized and complex cortical areas.

Neural processing

The retina of the (human or a biological) eye converts raw light into the nerve signals that the brain interprets as visual images. It is by comparing the signals from an array of visual cells that the retina and the brain build up a picture of the colour, shape and brightness of the outside world.

Rods are sensitive to low light levels and cones to higher ones. Furthermore, the cones themselves can alter the range of light intensities to which they respond, depending upon the average long-term brightness in a scene. This is an adaptive mechanism and accounts for the fact that when we step into bright sunlight from semidarkness, we experience the scene as washed out and overexposed.

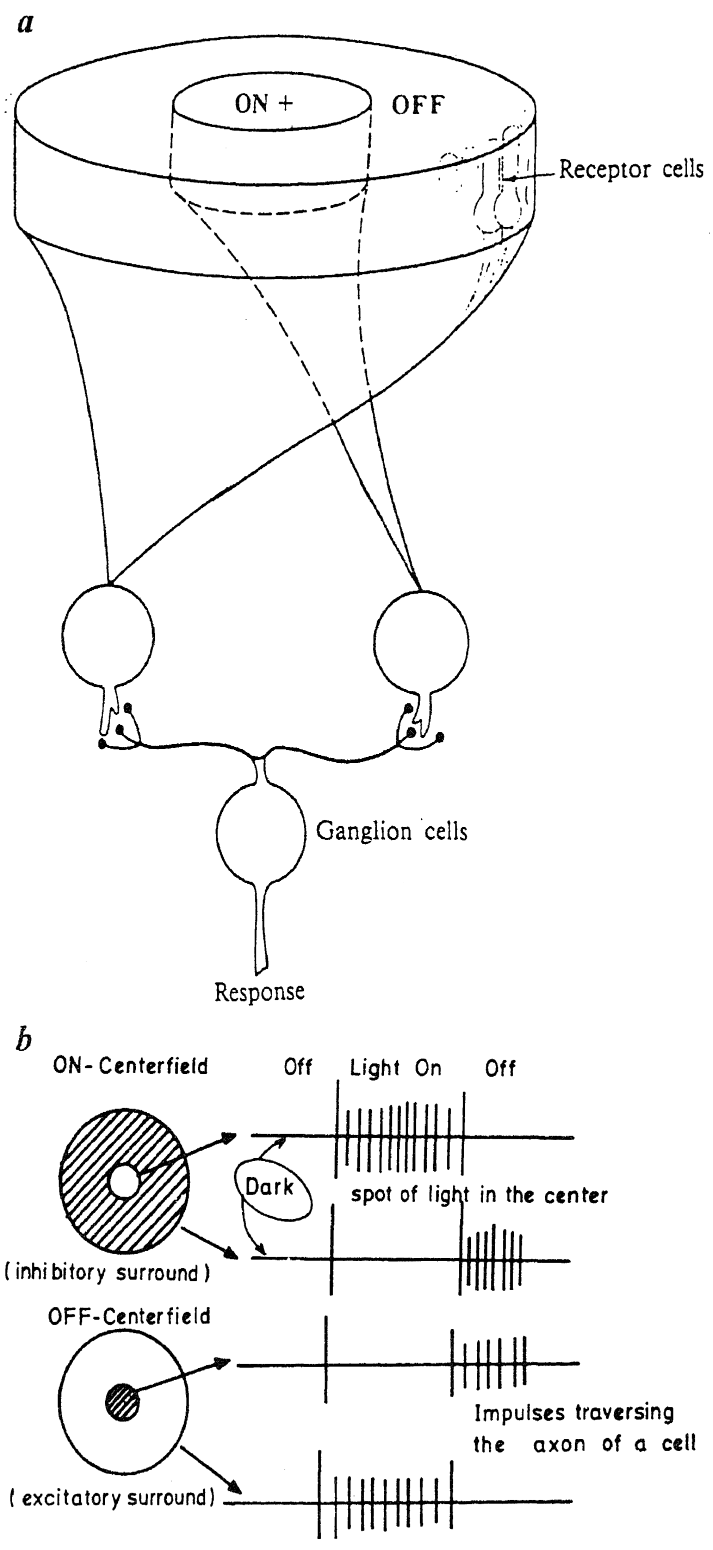


Figure 5. Centre-surround configuration: on–off, off–on cells.

The bipolar cells do not respond to the absolute brightness of the scene but only to the difference between the photoreceptor signal and the local average signal as computed by the horizontal cell network.

Real vision needs movement-sensitive and edge-enhancing functions of the amacrine and ganglion cells. Additional neural circuits are needed to recognize the patterns that the retina generates.

It has been found by electrophysiological experiments that both retinal ganglion and geniculate cells respond best to a roughly circular spot of light of a particular size in a particular part of the visual field. Each cell's receptive field (the patch of retinal receptor cells supplying the cell) is divided, with an excitatory centre and an inhibitory surround (an 'on-centre' cell) or exactly the reverse configuration (an 'off-centre' cell). This is the centre-surround configuration (Figure 5). Because these cells have circular symmetry, they respond to a line stimulus (a bar of light), whatever its orientation.

The signals from neighbouring visual cells may interact with each other; they may be added together so that their sensitivity is the sum of their receptive areas or they may inhibit each other. The signal from a group of cells or facets may be inhibited when cells surrounding their receptive field receive the same kind of stimulus (phenomenon of lateral inhibition). The signal produced by a patch of light falling on the central field will increase as the area of the patch increases until it fills exactly the receptive area. Henceforth, as the patch is increased further, the signal will actually decrease when the light begins to impinge on the inhibitory surround. The centre-surround organization is believed to enhance contrast perception, but to leave visual acuity unaffected.

The various salient features of the pattern – moving edges, either dark or light, the presence of moving dark spots, characteristic spatial frequencies in the pattern, and edges or spots moving in particular directions – are abstracted by the neural retina and transmitted inwards to the brain. The task of the cortex for the processing of visual information is different from that of the eye as an optical system. The eye, retina and lateral geniculate nucleus preserve essentially the spatial arrangement of the retinal image; the cortex transforms this geometry into abstract concepts.

Interestingly, it was found, based on the work of Cajal and de No (see reference 5), using the Golgi method, that the operations the cortex performs on the information it receives are local. Further, the brain's wiring pattern ensures that closely related information is mapped onto neighbouring groups of neurons. As an example, the cortical areas that perform the early processing of visual information preserve the spatial relations of the image.

The information on vision is relayed from one cortical area to the next, and the mapping of information becomes progressively more blurred, and the information carried more abstract.

There is a wide variety of cell types in the cortex, some simpler and some more complex in their response properties, the simpler cells feeding the more complex ones. Simple cells are orientation-specific and sensitive to the location of the oriented line. According to some researchers, these simple cells have receptive fields composed of several parallel elongated excitatory and inhibitory regions, modelled mathematically by the so-called Gabor function. Each receptive field can be

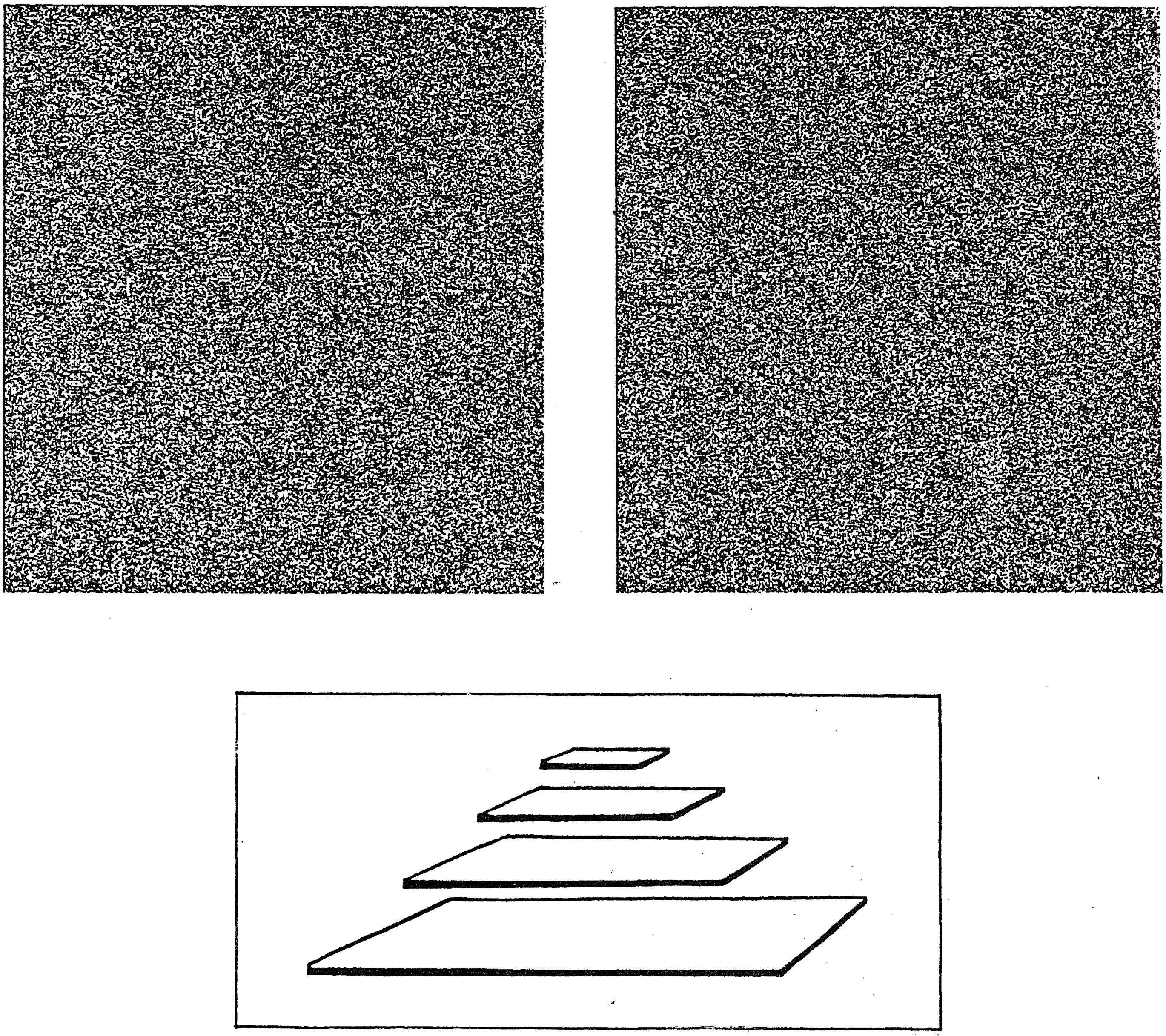


Figure 6. Random-dot stereo pair of images with the multiplanes as a three-dimensional percept.

characterized by a (radial) spatial frequency, corresponding to the inverse of the distance between bright bars, and by an orientation. The cell responds to a range of frequencies and orientations about its centre values. An important discovery is that the receptive field of each cell occupies a patch that is much smaller than the total frequency region to which the ensemble of cells is sensitive (see note 8).

More numerous are the orientation-specific complex cells, which are less particular about the exact position of a line. Complex cells behave as though they received input from a number of simple cells, all with the same receptive-field orientation but differing slightly in the exact location of their fields (see note 9) [see Reference 5b for the results of experiments (using microelectrodes) on the visual brain of the cat]. Certain brain cells respond to specific patterns at the eye, other brain cells to other patterns; some cells respond to movement in one direction but not the opposite or any very different direction; other cells respond to lines oriented at a certain angle and others to corners.

Hubel and Wiesel[5] have also found that pattern information of various kinds is brought together in 'columns' arranged at right angles to the clearly visible layers of the striate cortex. Such a structure seems to solve the problem of how the brain relates together not only three spatial dimensions but also colour, movement and other object characteristics.

Stereo vision

The visual cortex also combines the inputs from the two eyes. If two objects are separated in depth from the viewer, then the relative positions of their images will differ in the two eyes. The process of stereo vision, in

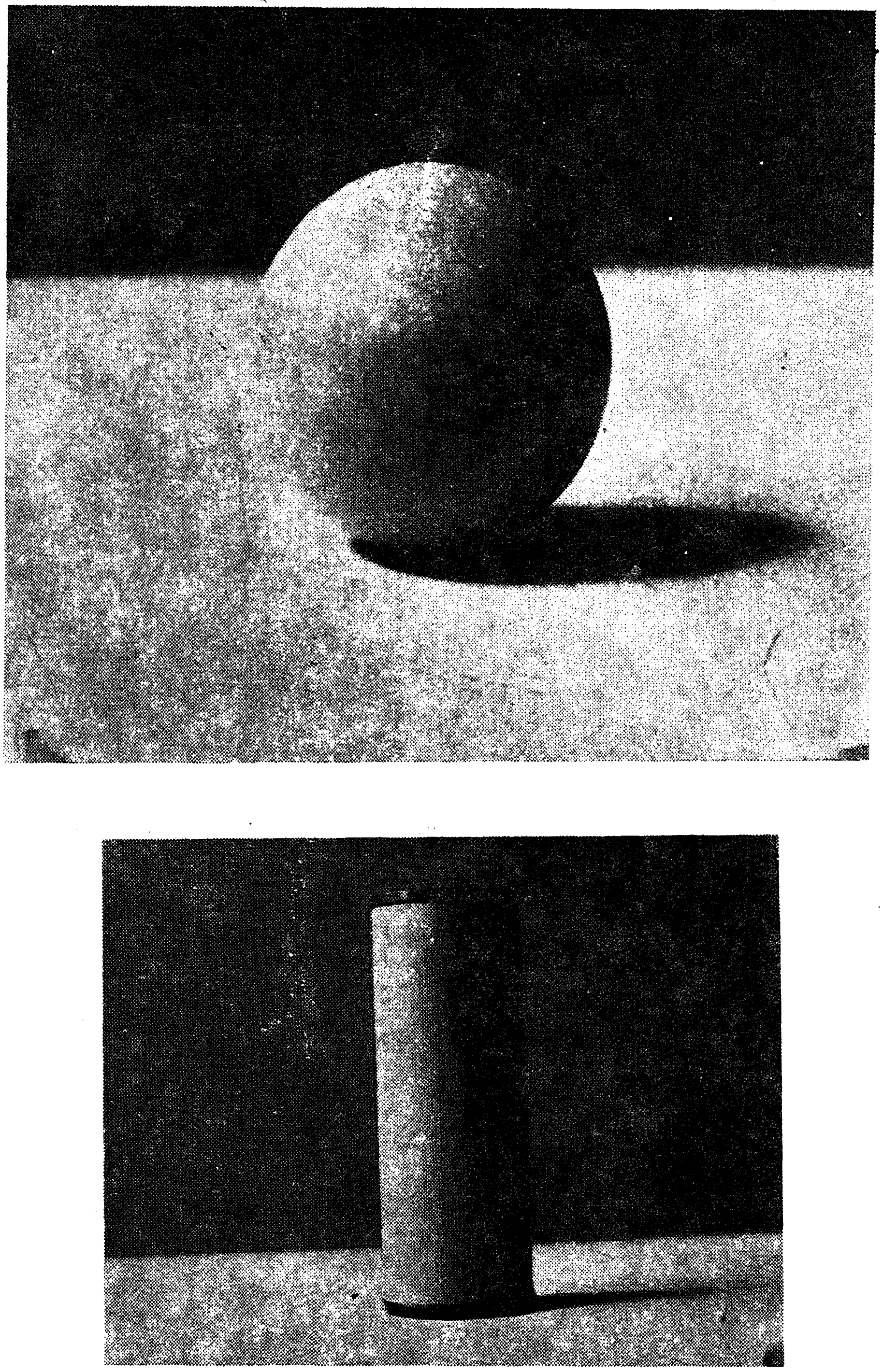


Figure 7. Natural images of objects.

essence, measures this difference in relative positions, called the *disparity*, and uses it to compute depth information for surfaces in the scene.

In primates and other species, binocular vision leads to depth perception, which is achieved in part by comparing the images from the two eyes. It has been discovered by Julesz[14], using some ingenious experiments, that form perception does not precede depth perception and that only 'local' contrast or texture comparisons are needed. Julesz achieved this by presenting each eye of a human subject with the same pattern of random dots, except that a region of dots in one image was displaced horizontally. The region so displaced is perceived by the subject as a surface at a different depth than its surround. See Figure 6 for the stereo pair which has to be seen by a stereoscope for perception of the three-dimensional multiple plane structure shown in the lower half of the figure.

Mathematical modelling of vision

The realization among researchers in vision now is that an understanding of biological vision is important in the attempt to create robots endowed with vision. Therefore, it is not surprising that vision system builders are looking for inspiration in biology. Mathematical modelling of vision is based on psychophysical/neurophysiological experiments on a biological system.

The scientific results emanating from neurophysiology and psychology usually *do not* address themselves directly to the issues arising out of the analysis of the so-called real and natural scenes (of the type shown in Figures 1 and 7). In neurophysiology, the input stimulus signals generally used are rectangular on–off pulse trains of light, simple periodic sinusoidal variations or moving versions of these. Such inputs present a very confused picture to anyone wishing to specify a computer model for perception. Here the emphasis is on monochromatic stimuli such as small objects (for example, circles or squares) against a uniform background, lights, characters and line drawings. Rarely is an actual picture of a real environment, such as an office scene or suburban street scene, ever employed.

Hubel and Wiesel[5] and others have experimentally discovered, in the visual system (in the cat and other mammals), networks of cells which are selectively responsive to such features as the direction of a contour, rate of change of brightness, and curvature of the contour. Any (perceptual) model, if it has to be mathematically complete, will have to assume the existence of modules (or *banks of filters*) operating on the input in order to extract (at least in a preliminary way) cues which are passed on to the next stage for further processing/abstraction.

In the attempt to model the visual system, it would be helpful to treat the process of vision as occurring in three (four) stages (Figure 2 *a*):

1. *Optical stage*, when an image of the outside world is projected onto the retina.
2. *Transduction stage*, when the light-sensitive visual cells convert the light energy into electrical signals.
3. *Physiological stage*, when these primary signals are analysed.
4. Possibly, the (last) stage that marks the conscious awareness of a visual display.

The neural net modelling of various aspects of visual perception started with Pitts and McCulloch's paper[15] on pattern recognition. Land and McCann[16] were among the first ones to demonstrate that edges in the retinal image play a key role in the process. They proposed that edge information is extracted in part in the retina, via lateral inhibition between retinal neurons, and that in the cortex there is an inverse process – lateral excitation – whereby the brightness of a patch of the image (bounded by

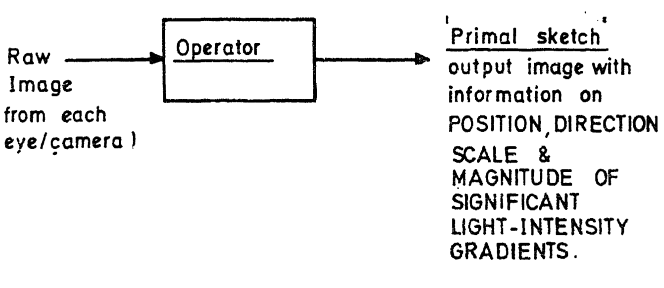


Figure 8. General schematic diagram for the derivation of the 'primal sketch'.

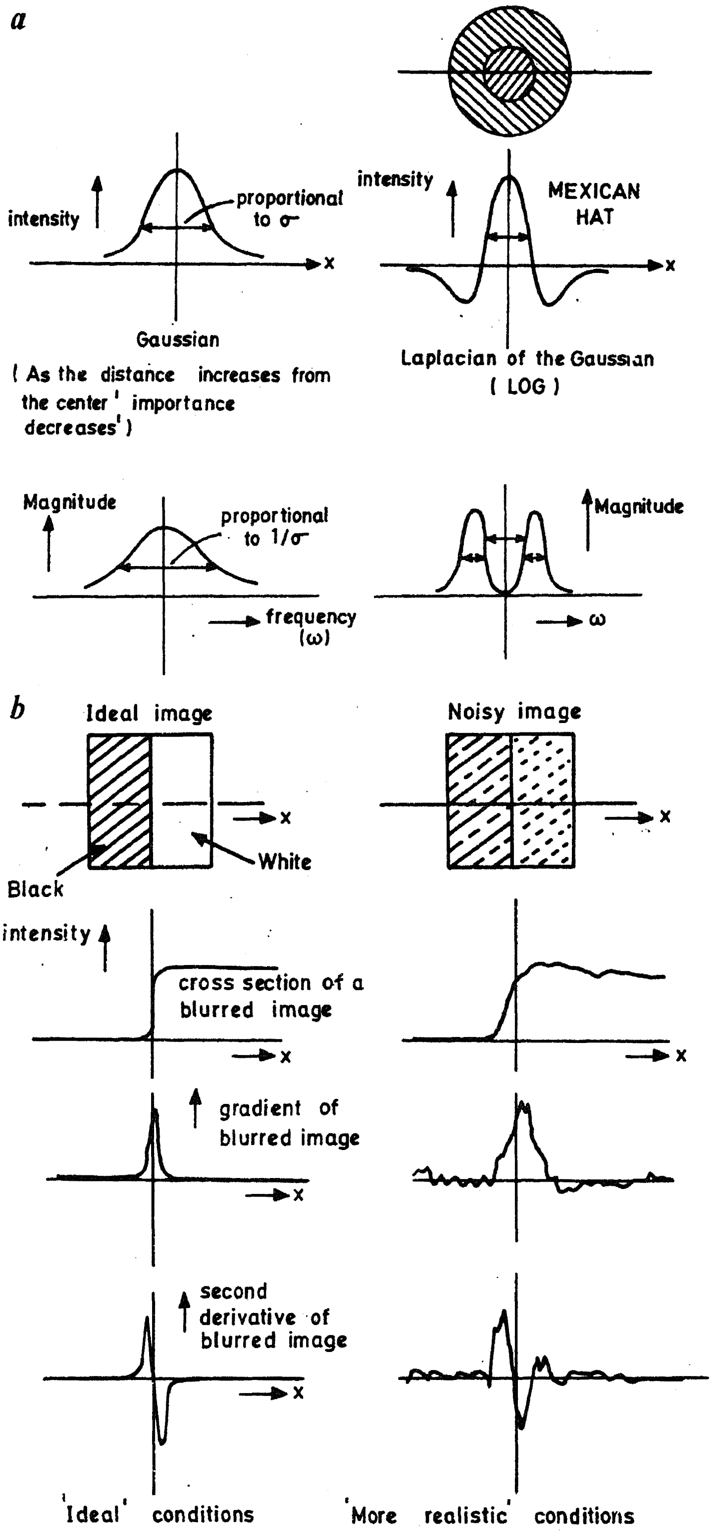


Figure 9. ***a***, One-dimensional Gaussian and its second derivative (i.e. a cross-section of the Laplacian of the Gaussian): spatial and frequency characteristics at one 'scale'; ***b***, consequences of applying the second derivative of the Gaussian to the (cross-section) of an ideal and a realistic 'step edge'.

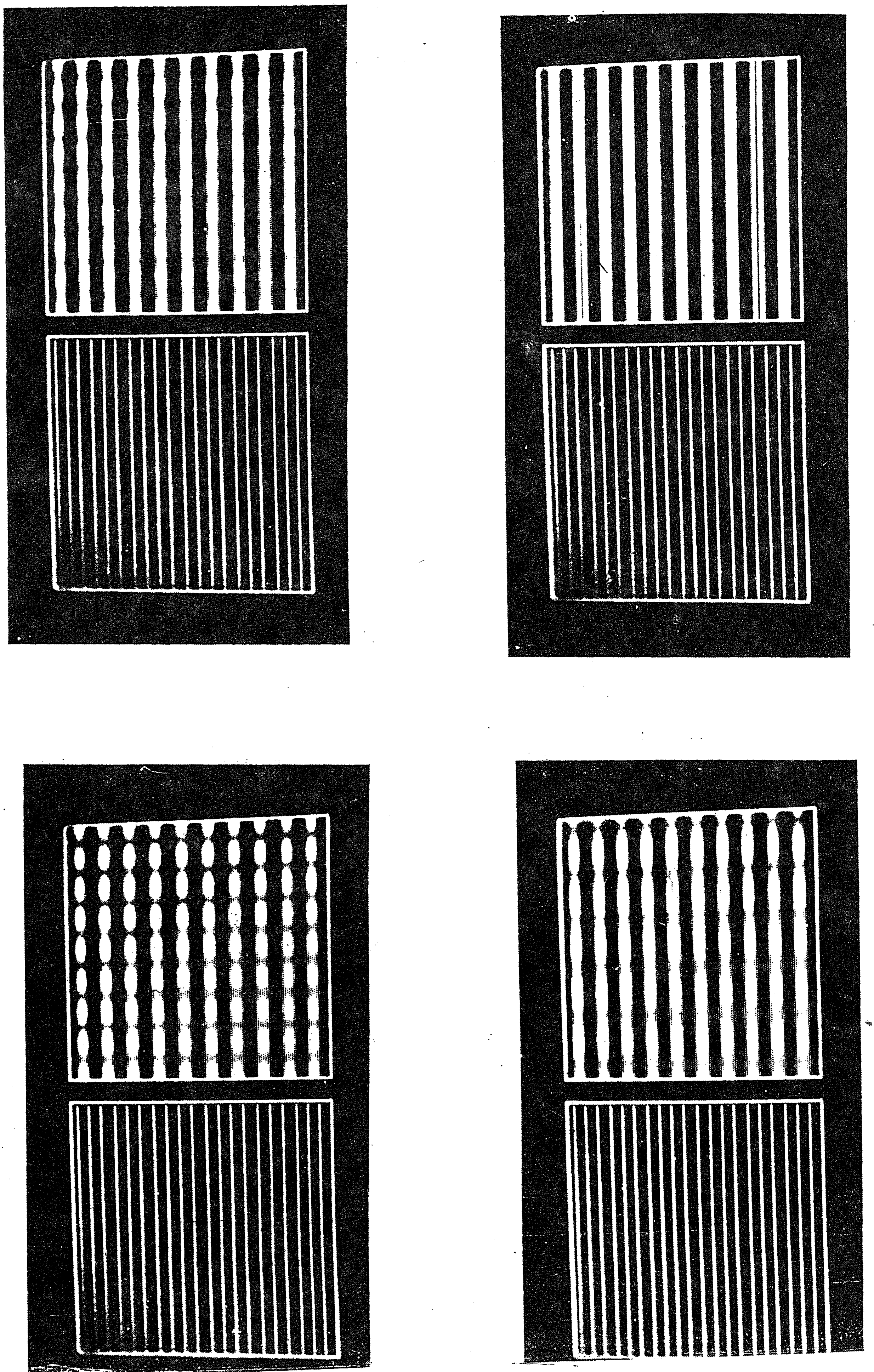


Figure 10. Daugman's synthetic images to demonstrate the inadequacy of the Marr model of human vision based on zero crossings.

edges) is computed. The principal hypothesis is that a major function of the visual cortex is to compute invariants of the scene at different levels of abstraction.

Explicit modelling results

Perhaps as a first step towards answering the question of the Turing type, Marr's book[9] investigates the question 'What does it mean, to see?' His research has provided quite an interesting framework for understanding how humans interpret visual information (see note 10).

In his formulation of the model of early vision[19], the vision problem begins with a large grey-level intensity array and culminates in a description that depends on that array and on the purpose for which it is being viewed. The first step of consequence is to compute a primitive but restrictive description of the grey-level changes present in the image, and then to implement all subsequent computations as manipulations of that description. The description itself is called the primal sketch (as indicated in a grossly simplified schematic diagram in Figure 8). According to Marr and Hildreth[20], in their attempt to model the early processing of information by the human visual system, the image is first processed independently through a set of different size operators (see note 11) whose shape is the Laplacian of a Gaussian $\Delta^2 g(x, y)$ (which is a *rotationally invariant* second-derivative operator), where the function $g(x, y)$ is the Gaussian function of two variables whose size is controlled by the variance parameter σ. The loci, along which the convolution outputs cross zero, mark the positions of intensity changes at different resolutions. These changes often correspond to what we intuitively call 'edges', and are the consequence of a change in some physical property of a surface: reflectance, geometry or incident illumination. See Figure 9 for the one-dimensional representation of the Laplacian of the Gaussian and for the consequences (also in one-dimension) of its operation on an ideal and a realistic 'step' edge.

A sharp change in the intensity will give rise to a peak in the output of a first-derivative operator (see note 12), or, equivalently, zero crossing in the output of the second derivative. These zero crossings can be described by their position, slope of the convolution output across zero, and the two-dimensional orientation. The set of descriptions from different operator sizes forms the input for later visual processes, such as stereopsis and motion analysis.

Distinct from the above approach, Zucker[1] presents the new trends, involving the 'tangent field' for the detection of boundaries of objects in a scene. According to Zucker, these are supposed to open up 'intriguing' biological and mathematical connections (see note 13).

However, there does not seem to exist a definitive account of the way in which the cortex actually processes the edge information.

Consider the two *distinguishable* patterns[18] in Figure 10, which are synthetically generated. The essence of this generation is that a sinusoidal pattern (of frequency ω_1) in the horizontal direction is modulated in the vertical direction by another sinusoidal pattern of frequency ω_2, the amount of modulation being restricted by the frequencies ω_1 and ω_2. The patterns so generated are clearly distinguishable. But their zero crossings (or contours obtained according to the Marr model of blurring by a gaussian function, and taking the second derivative of the resulting image) are the same vertical lines.

More generally, a drawing of a scene adequately represents the scene, despite the very different grey-level image to which it gives rise. The results of many of the edge detection algorithms which have been proposed for extracting line drawings from natural images have proved to be generally unsatisfactory. Doesn't this imply that even an adequate line drawing of a scene cannot be computed unless hypotheses about what is present are allowed to influence processing strategies?

Stereo-perception modelling

Marr[9] and Marr and Poggio[22] made explicit the steps involved in measuring stereo-disparity from the stereo pair of images: (i) a particular location on a surface in the scene is selected from one image; (ii) the same location is identified in the other image; and (iii) the disparity between the two corresponding image points is measured. The difficulty of the problem lies in steps (i) and (ii), that is, in matching the images of the same location, the so-called *correspondence problem*.

Dev[23] and Marr and Poggio[22] proposed a neural net implementation of the Julesz model. In this model, neurons in the visual cortex (excitatory interactions of neighbouring units signalling the same depth, and inhibitory interactions otherwise) are monocularly driven, and can signal the depth of a local patch of the surface of an object.

In this context, random-dot stereograms (of the type shown in Figure 6) are particularly interesting because when one tries to set up a correspondence between two arrays of dots, false targets occur in profusion. A *false target* refers to a possible but incorrect match between elements of the two views. In spite of such false targets, and in the absence of any monocular or high-level cues, we are able to determine the correct correspondence. Thus, the computational problem of human stereopsis reduces to that of obtaining primitive descriptions of locations to be matched from the images and solving the correspondence problem for these descriptions.

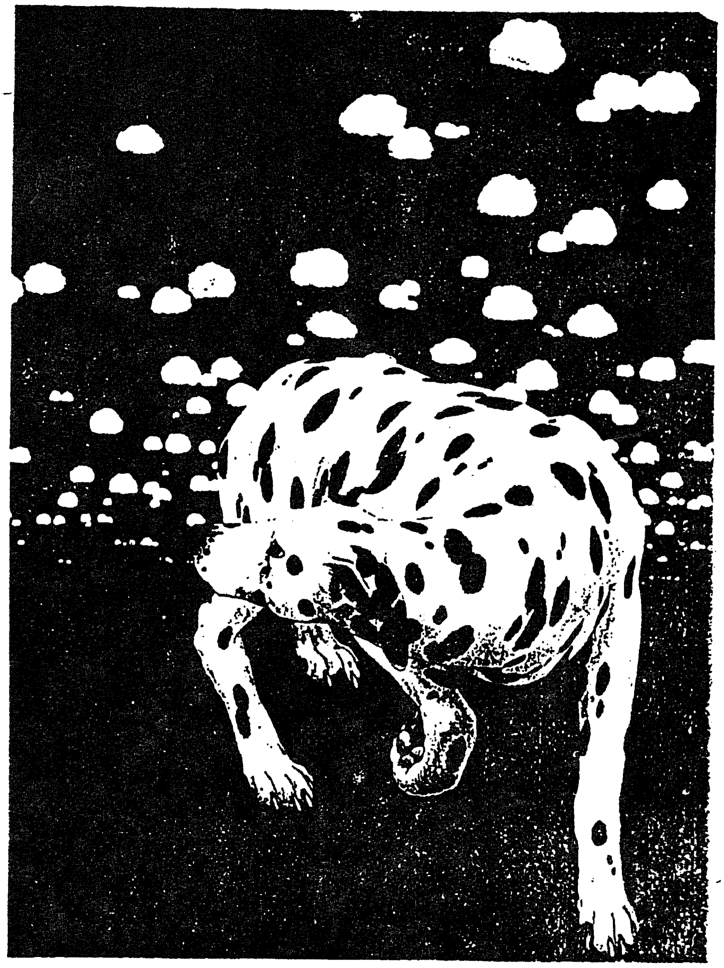


Figure 11. Patches of black and white giving rise to the perception of a dog.

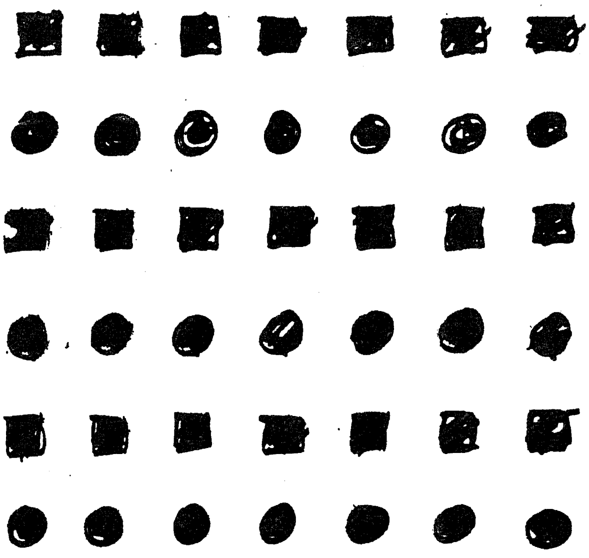


Figure 12. Pattern of circles and squares.

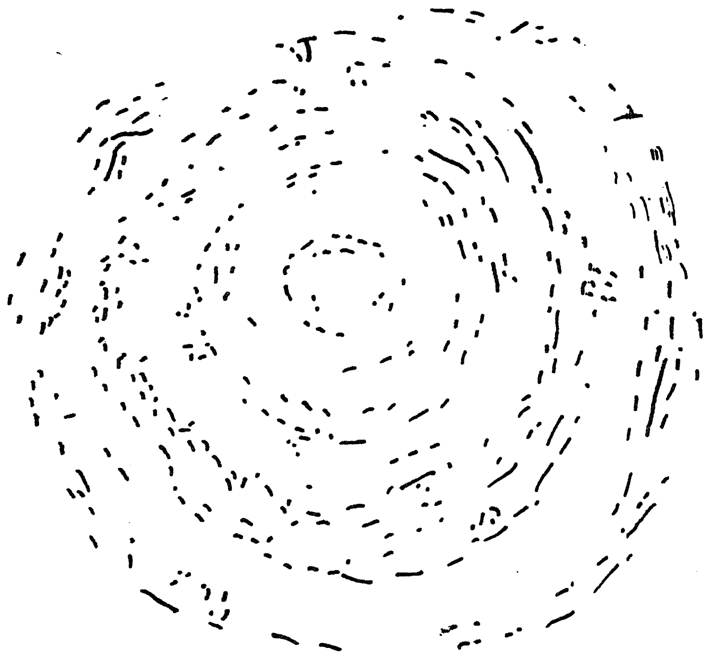


Figure 13. The so-called glass pattern.

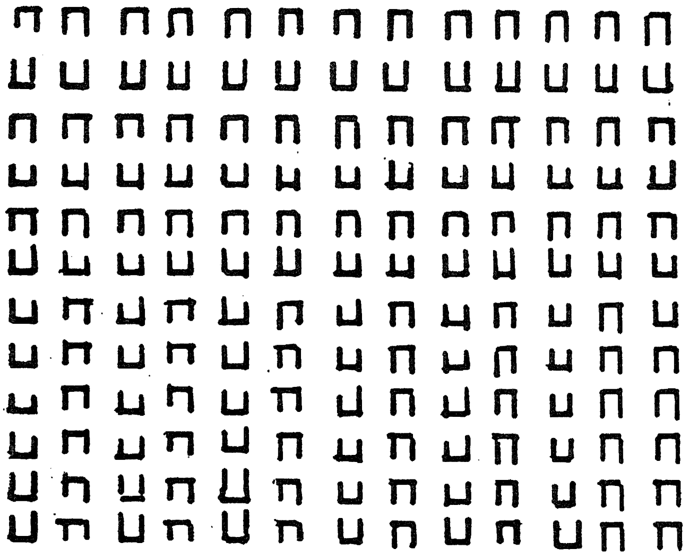


Figure 14. Beck's textural image.

Mathematically difficult perceptual processes

The human visual system has a highly developed capability for detecting many classes of patterns having properties of, for instance, symmetry, collinearity, parallelism and connectivity.

A major role is played by the process of perceptual organization (see note 14), in which groupings are formed directly from the spatial structure of the image without prior knowledge of its contents. Refer to Figure 11, containing mere patches of white and black, but still leading to the perception of an object (dog) in the scene. Under this category, one includes figure-background phenomena, segmentation, Gestalt perception and others.

The Gestalt psychologists, in the early part of the century, made much play of perceptual organization, in which stimulus patterns are organized into 'wholes'

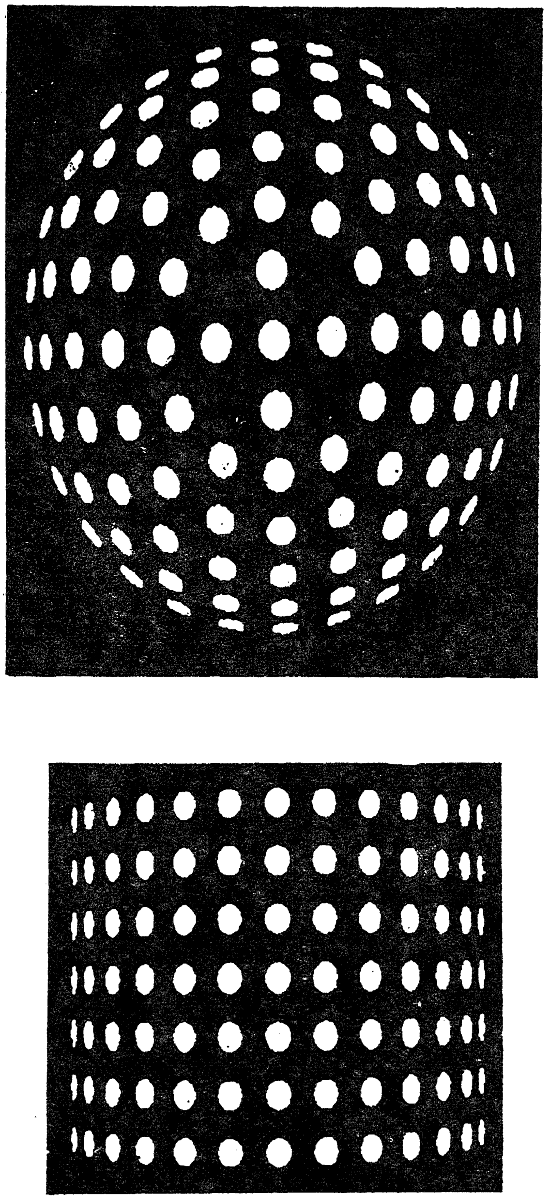


Figure 15. Pattern of white circles (of gradually decreasing sizes) against a dark background, which are perceived as surfaces.

(Gestalten). Patterns of dots were used in an attempt to establish the laws of organization[24,25]. A typical principle is shown in Figure 12: The circles and the squares are seen separately, each forming rows. This demonstrates that *similarity* is a factor in perceptual grouping.

In the glass pattern (Figure 13), we see and recognize many black dots on a white paper, but we also recognize circular groupings.

The textural image of Figure 14 (ref. 26) contains simply u's and inverted u's. We can distinguish between the upper and the lower halves in this image. This is apparently accomplished by the long horizontal groupings (or segmentation).

A human observer can easily understand the shapes of the surfaces whose images are shown in Figures 15 *a*, *b*. Even though some computational theories for the description of the above perceptual processes have been proposed[27], they are not satisfactory when applied to natural images.

Illusions as counterexamples (to current theories)?

There are also instances in which we *see* things not in the image as such; these are called *illusions*. See, for example, Figure 16, wherein we perceive a bright disk within the lines/regions, although the luminance is the same throughout the image.

Optical illusions seem to provide an important insight into the biological retina's role in reducing the bandwidth of visual information and extracting only the essential features of the image (see note 15). The illusions are created because the retina encodes the visual information selectively.

Perceptual illusions involve a wide range of stimulus properties, including luminance, colour, size, shape, orientation, distance and velocity. They are not yet fully understood, and the number of theories proposed to explain them is almost as great as the number of effects so far discovered.

The first type of illusion concerns the adaptive response of the visual system to fields of parallel lines. Figure 17 *a*, for example, has a high degree of contour-directional redundancy. After looking at it for a while and looking away, one sees a curious pattern of wavy circles.

With Figure 17 *b*, on the other hand, one sees wavy radial streamers as an afterimage. A general description could perhaps be that after exposure to near-parallel lines in one direction, the visual system seems to be hypersensitized to a direction roughly orthogonal to the direction of the stimulus. This leads to the possible existence of networks, in the (human) visual system, specifically sensitive to direction and capable of adapting to direction in such a way that our colour-sensitive mechanisms tend to adapt to colour, so that complementary directions tend to be seen after exposure to a given direction just as complementary colours tend to be seen after exposure to a given colour.

The Hermann grid (Figure 18) is an array of black squares. Although the white grid has equal luminance throughout, small grey circles can be seen at each of the intersections. These differences in perceived luminance cannot be measured by a photometer; they are not present in the physical stimulus. We see the grey circles because of the way the visual system processes adjacent areas of dark and light.

Artificial neuron modelling

Extending the results of McCulloch and Pitts[15] and Hebb[3], the present-day *connectionists* hold that the abstraction from the given image of a scene can be

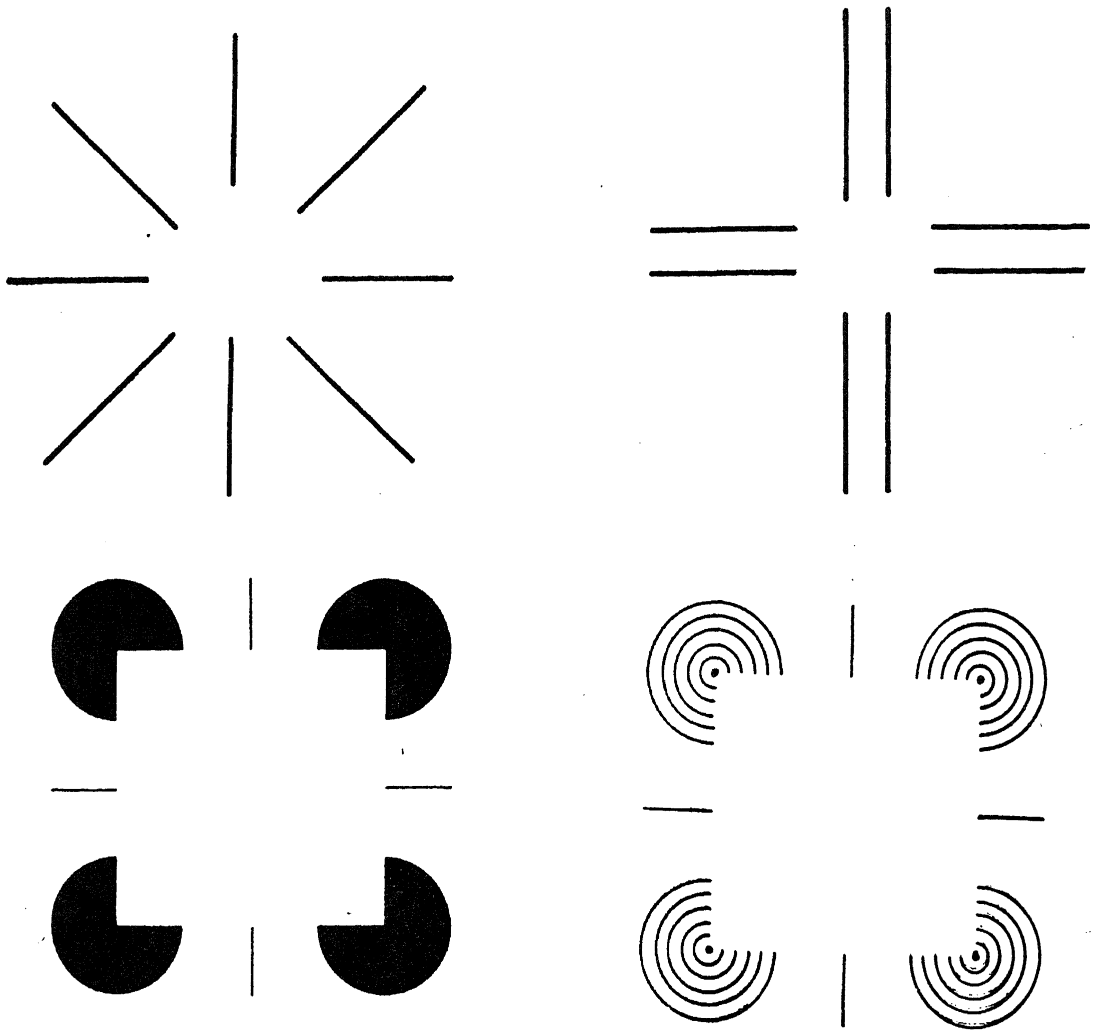


Figure 16. Some types of illusions.

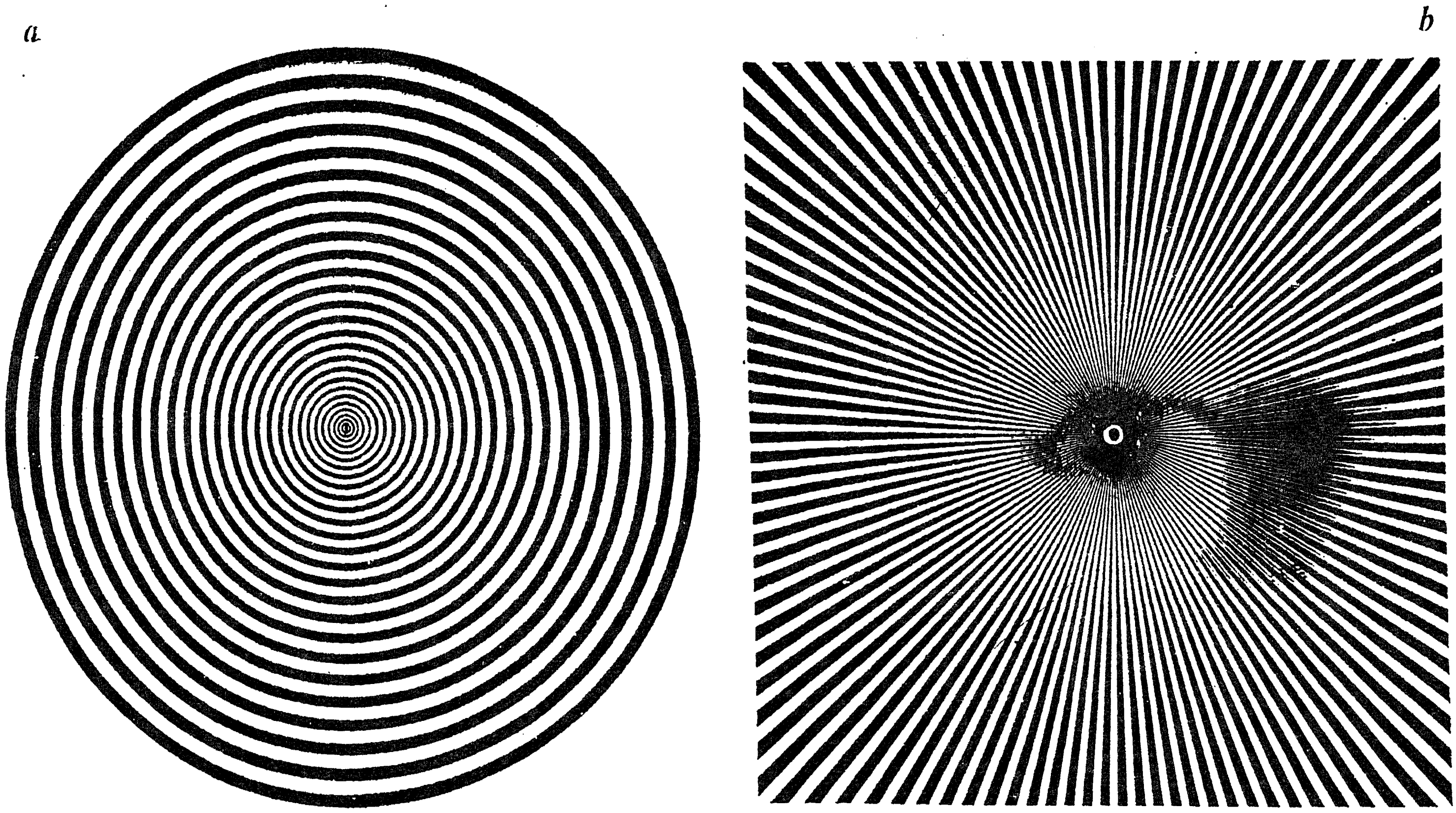


Figure 17. *a*, Ray pattern giving rise to complementary images; *b*, circular pattern giving rise to similar complementary images.

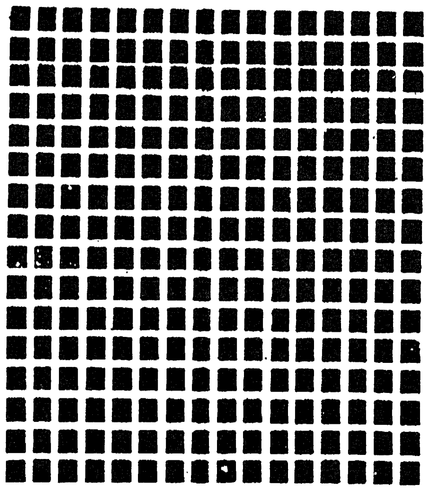


Figure 18. Hermann grid.

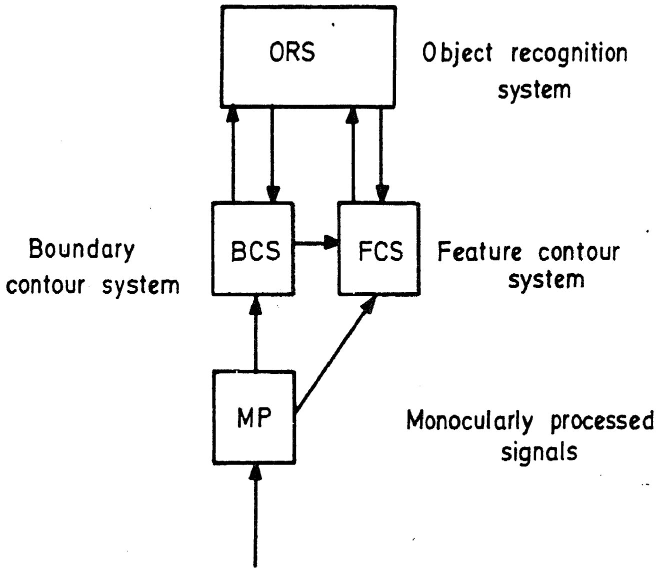


Figure 19. Grossberg model of human perception.

described in terms of the synaptic strengths in the neural network[29]. Given that the entire behavioural response is realized in a few hundred milliseconds (using the slow, also millisecond) neural circuitry of the brain, it appears that the connectionist models of (visual) perception seem to be the only way to achieve these response times.

Perhaps, motivated by such a consideration, some attempts have been made to create a '*connectionistic*' neural model for the visual cortex. For instance, von der Malsburg[30] considers a nerve net model (for the visual cortex of higher vertebrates) which consists of 338 neurons forming a sheet analogous to the cortex. These neurons are connected randomly to a 'retina' of 19 cells. With the stimuli in the form of light bars, a learning procedure is given for the organization of orientation selectivity of single units.

In a different framework, Grossberg *et al.*[31, 32] contend that some perceptual phenomena can be explained by their new model, consisting of three subsystems: boundary contour system (BCS), feature contour system (FCS) and object recognition system (ORG). These subsystems are supposed to interact to generate a representation of the physical world as the output of a sequence of operations on the human retinal image. This model is supposed to explain how our visual systems are designed to detect relatively invariant object boundaries in spite of noise and to recognize objects in an environment subjected to occlusion. See Figure 19 for a schematic of the Grossberg model.

For pre-attentive vision, Grossberg *et al.*[31,32] report that BCS and FCS lead to *emergent boundary segmentation and featural filling-in*. This architecture is claimed to clarify how scenic data about boundaries, textures, shading, depth, motion and other types of information can be cooperatively synthesized in real time into a coherent representation of three-dimensional form.

We have recently presented the actual results of some computational experiments on the Grossberg neural network model for brightness perception[33]. However, on implementation of the brightness perception and texture boundary detection Grossberg models, it has been found that the outputs do not match the predicted results. The model definitely *needs* to be changed, but what changes are to be made are not known.

Fukushima[13] has attempted to create an *artificial* neural network architecture which is ostensibly similar to the neuronal structure as described by Hubel and Wiesel[5], to recognize two-dimensional patterns. It has been found that the performance of this is far from satisfactory[34].

Conclusions

The success of the venture to create a machine vision system can lead to a bridge between neurobiology and information sciences. One of the new views of information processing is that analog systems may be able to solve problems found intractable by conventional digital methods.

Many of the results in the literature on modelling of perceptual processes involved in the recognition of patterns/objects in natural scenes appear to be unsatisfactory. Some of the images which are not adequately analysed by the current theories of perception have been presented in this paper.

With the invasion of computers into human activities, a topical question is: Can one design a machine to replace the human visual system and, more significantly, give '*artificial sight*' to the blind people? Conversely, can the mathematical analysis of the human vision system help us to understand the visual cortex behaviour itself? For related studies, see References 35–37.

1. Zucker, S. W., *Curr. Sci.*, 1993, **64**, 407–418.
2. Demidov, V., *How We See*, Mir Publishers, Moscow, 1986.

3. Hebb, D. O., *The Organization of Behaviour*, Wiley, New York, 1949.
4. Eccles, J. C., *The Physiology of Nerve Cells*, Johns Hopkins University Press, Baltimore, 1957.
5. Hubel, D. H. and Wiesel, T. N., *J. Physiol.*, 1963, **165**, 559–567 (a), and *J. Physiol.*, 1962, **160**, 106–154 (b).
6. Turing, A. M., reprinted in *The World of Mathematics*, presented with Commentaries and Notes by J. R. Newman, Tempus Books of Microsoft Press, Washington, 1988, vol. 4, pp. 2075–2098.
7. von Neumann, John, *The Computer and the Brain*, Yale University Press, New Haven, 1958.
8. Mahowald, M. A. and Mead, C., *Vigyan, Scientific American B*, 1991 (June), 82–90.
9. Marr, D., *Vision*, Freeman, New York, 1982.
10. Ullmann, J. R., *Nature*, 1975, **257**, 547–549.
11. Pavlidis, T., *Pattern Recog. Lett.*, 1992, **13**, 221–225.
12. Watson, A. B., *Comput. Vision, Graphics and Image Process.*, 1987, **39**, 311–327.
13. Fukushima, K., *Biol. Cybern.*, 1980, **36**, 193–202 (a), and *Neural Networks*, 1989, **2**, 413–420 (b).
14. Julesz, B., *Bell Systems Tech. J.*, 1960, **39**, 1125–1162 (a), and *Handbook of Sensory Physiology*, Springer Verlag, Heidelberg, 1978, Ch. 7, pp. 215–256.
15. Pitts, W. and McCulloch, W. S., *Bull. Math. Biophys.*, 1947, **9**, 127–147.
16. Land, E. H. and McCann, J. J., *J. Opt. Soc. Am.*, 1971, **61**, 1–11.
17. Westheimer, G., *J. Social Biol. Struct.*, 1987, **10**, 395–424.
18. Daugman, J., *J. Opt. Soc. Am.*, 1985, **2**, 1160–1169.
19. Marr, D., *Phil. Trans. R. Soc. London*, 1976, **B275**, 483–534.
20. Marr, D. and Hildreth, E., *Proc. R. Soc. London*, 1980, **B207**, 187–217.
21. Wilson, H. R. and Bergen, J. R., *Vision Res.*, 1979, **19**, 19–32.
22. Marr, D. and Poggio, T., Cooperative computation of sterodisparity, *Science*, 1976, **194**, 283–287.
23. Dev, P., *Int. J. Man-Machine Studies*, 1975, **7**, 511–528.
24. Wertheimer, M., in *Readings in Perception* (eds. Beardslee, D. and Wertheimer, M.), Princeton, USA, 1958, pp. 115–135.
25. Lowe, D., Perceptual Organization and Visual Recognition, Report No. STAN-CS-84-1020, Department of Computer Science, Stanford University, Stanford, September 1984.
26. Beck, J. and Prazdny, K. and Rosenfeld, A., in *Human and Machine Vision* (eds. Beck, J., Hope, B. and Rosenfeld, A.), Academic Press, New York, 1983.
27. Aloimonos, J. and Shulman, D., *J. Opt. Soc. of Am. A*, 1989, **6**, 908–919.
28. Day, R. H., *Interdisciplinary Sci. Rev.*, 1984, **9**, 47–58.
29. Feldman, J. A., *Byte*, 1985, 277–284.
30. von der Malsburg, Chr., *Kybernetik*, 1973, **14**, 85–100.
31. Grossberg, S. and Mingolla, E., *Psychol. Rev.*, 1985, **92**, 173–211 (a), and *Perception and Psychophys.*, 1985, **38**, 141–171.
32. Grossberg, S., Mingolla, E. and Todorovic, D., *IEEE Trans. Bio-Medical Eng.*, 1989, **36**, 65–84.
33. Venkatesh, Y. V. and Sai Ram, K., Technical Report, Department of Electrical Engineering, Indian Institute of Science, Bangalore, 1993, pp. 1–17, the Asian Conference on Computer Vision, Tokyo, November 1993.
34. Ganesh Murthy, C. N. S. and Venkatesh, Y. V., International Conference on Information Processing, Singapore, December 1991 (a), and Technical Report, Department of Electrical Engineering, Indian Institute of Science, Bangalore, 1993 (b).
35. Lythgoe, J. N., *The Ecology of Vision*, Clarendon Press, Oxford, 1979, Chapter 2.
36. Sinclair, Sandra, *How Animals See: Other Visions of Our World*, Croom Helm, London, 1985.
37. Cowan, J. D. and Sharp, D. H., *Q. Rev. Biophys.*, 1988, **21**, 365–427.

Notes

1. This *separability* is the foundation of the science of artificial intelligence.
2. Some fundamental questions come up for analysis: Are 'seeing' and 'thinking' mathematically the same, or is one a subset of the other? Can 'seeing' also be reduced to the 'digital' paradigm as thinking, the way Turing does?
3. It is generally accepted that the retina's nerve cells function as complex analog processors. The connections between these cells, the characteristics of their connections and the shape of the cells themselves all seem to play a major role in determining the basic parameters of analog processing. In this context, see reference 8 for the analog paradigm suggested as a model for retinal processing.
4. Unless, of course, one merely asserts that this question is equivalent to Turing's original question.
5. Ullmann[10]: 'If you consult a heart specialist you can be assured that he knows fairly accurately how the heart works, but a person who consults a psychiatrist has no such assurance concerning the brain.' Can't one here replace the word psychiatrist by psychophysicist or even a neurobiologist?
6. Object identification is for the most part a matter of form perception. Form may perhaps be described as a set of properties that are invariant over certain transformations (like brightness and colour). Sometimes visual form is restricted to properties of contours which may be generated by discontinuities in psychological properties of light other than brightness, i.e. hue, saturation and texture.
7. In the so-called early vision, information is represented in a spatial rather than symbolic form. This corresponds to the first few layers of the hierarchy.
8. This is interpreted as implying the existence of multiple channels in the visual system. See reference 12.
9. This is, possibly, the motivation for Fukushima's[13] structure of the 'neo-cognitron' for pattern recognition. Discussed later and in Reference 34.
10. It is curious that Marr has been quoted as having said[17]. 'He did not much mind if any of his theories failed to find experimental verification.' True enough, patterns can be generated which are discriminated by humans but not by the Marr model of computing the zero-crossings. See Reference 18 and the discussion below.
11. These findings are consistent with the discovery of Hubel and Wiesel[5] that at each cortical location there exist many neurons with differing receptive field sizes. See also Reference 21.
12. Edge detection amounts to numerical differentiation, which leads to (mathematically) difficult problems – the so-called ill-posedness.
13. However, the illusory contour problem, which was used by Zucker[1] as an example to show the inadequacy of the classical approaches, is also surprisingly left out of the new framework.
14. Any visual task performed spontaneously without effort or deliberation is regarded as a perceptual task.
15. As also for isolating and investigating the sensory and cognitive processes associated with normal perception[28].

ACKNOWLEDGEMENTS. The author expresses his grateful thanks to Professor P. N. Tandon and Professor M. A. Viswamitra for the invitation to deliver a lecture on biological information processing to the distinguished gathering of neurobiologists, biochemists, biophysicists, neuroscientists, mathematicians and others, at the Indian Institute of Science, Bangalore.